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 $\Omega$) be a class of subsets of $\Omega$. Later, $\Omega$ 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 $\Omega$ 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 $\pi$-system if $A \cap B \in \mathcal{A}$ whenever $A, B \in \mathcal{A}$,
- $\sigma$-$\cap$-closed (closed under countable 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 $A, B \in \mathcal{A}$,
- $\sigma$-$\cup$-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.
Definition 1.2 (\(\sigma\)-algebra). A class of sets \(\mathcal{A} \subset 2^\Omega\) is called a \(\sigma\)-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 \(\sigma\)-algebra is also named a \(\sigma\)-field. As we will see, we can define probabilities on \(\sigma\)-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

\[
\mathcal{A} \text{ is } \cap\text{-closed } \iff \mathcal{A} \text{ is } \cup\text{-closed},
\]

\[
\mathcal{A} \text{ is } \sigma\cap\text{-closed } \iff \mathcal{A} \text{ is } \sigma\cup\text{-closed}.
\]

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 \(\sigma\cap\text{-closed}\) and let \(A_1, A_2, \ldots \in \mathcal{A}\). Hence

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

Thus \(\mathcal{A}\) is \(\sigma\cup\text{-closed}\). The other cases can be proved similarly.

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

(i) \(\mathcal{A}\) is \(\cap\text{-closed}\).

(ii) If in addition \(\mathcal{A}\) is \(\sigma\cup\text{-closed}\), then \(\mathcal{A}\) is \(\sigma\cap\text{-closed}\).

(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

\[
\bigcup_{n=1}^{\infty} A_n = A_1 \uplus (A_2 \setminus A_1) \uplus ((A_3 \setminus A_1) \setminus A_2) \uplus (((A_4 \setminus A_1) \setminus A_2) \setminus A_3) \uplus \ldots .
\]
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 $\sigma$-ring if it is also $\sigma$-$\cup$-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 $A \subset 2^\Omega$ is called a \textbf{$\lambda$-system} (or Dynkin’s $\lambda$-system) if

(i) $\Omega \in A$,

(ii) for any two sets $A, B \in A$ with $A \subset B$, the difference set $B \setminus A$ is in $A$, and

(iii) $\bigcup_{n=1}^{\infty} A_n \in A$ for any choice of countably many pairwise disjoint sets $A_1, A_2, \ldots \in A$.

Example 1.11. (i) For any nonempty set $\Omega$, the classes $A = \{\emptyset, \Omega\}$ and $A = 2^\Omega$ are the trivial examples of algebras, $\sigma$-algebras and $\lambda$-systems. On the other hand, $A = \{\emptyset\}$ and $A = 2^\Omega$ are the trivial examples of semirings, rings and $\sigma$-rings.

(ii) Let $\Omega = \mathbb{R}$. Then $A = \{A \subset \mathbb{R} : A$ is countable$\}$ is a $\sigma$-ring.

(iii) $A = \{(a, b) : a, b \in \mathbb{R}, a \leq b\}$ 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 $\sigma$-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

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

be the set of all sequences whose first $n$ values are $\omega_1, \ldots, \omega_n$. Let $A_0 = \{\emptyset\}$. For $n \in \mathbb{N}$, define

$$A_n := \{[\omega_1, \ldots, \omega_n] : \omega_1, \ldots, \omega_n \in E\}. \quad (1.1)$$

Hence $A := \bigcup_{n=0}^{\infty} A_n$ is a semiring but is not a ring (if $\#E > 1$).

(vii) Let $\Omega$ be an arbitrary nonempty set. Then

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

is an algebra. However, if $\#\Omega = \infty$, then $A$ is not a $\sigma$-algebra.

(viii) Let $\Omega$ be an arbitrary nonempty set. Then

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

is a $\sigma$-algebra.

(ix) Every $\sigma$-algebra is a $\lambda$-system.

(x) Let $\Omega = \{1, 2, 3, 4\}$ and $A = \{\emptyset, \{1, 2\}, \{1, 4\}, \{2, 3\}, \{3, 4\}, \{1, 2, 3, 4\}\}$. Hence $A$ is a $\lambda$-system but is not an algebra. \hfill $\Diamond$
Theorem 1.12 (Relations between classes of sets).

(i) Every \(\sigma\)-algebra also is a \(\lambda\)-system, an algebra and a \(\sigma\)-ring.

(ii) Every \(\sigma\)-ring is a ring, and every ring is a semiring.

(iii) Every algebra is a ring. An algebra on a finite set \(\Omega\) is a \(\sigma\)-algebra.

Proof. (i) This is obvious.

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

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

\[\Box\]

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

\[\liminf_{n \to \infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m\]

and

\[\limsup_{n \to \infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\]

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

Remark 1.14. (i) \(\liminf\) and \(\limsup\) can be rewritten as

\[\liminf_{n \to \infty} A_n = \{\omega \in \Omega : \#\{n \in \mathbb{N} : \omega \notin A_n\} < \infty\},\]

\[\limsup_{n \to \infty} A_n = \{\omega \in \Omega : \#\{n \in \mathbb{N} : \omega \in A_n\} = \infty\}.

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

\[1_A(x) := \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases} \quad (1.2)\]

With this notation,

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

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

\[\Box\]

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 $\sigma$-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 $\sigma$-algebra. The analogous statement holds for rings, $\sigma$-rings, algebras and $\lambda$-systems. However, it fails for semirings.

Proof. We give the proof for $\sigma$-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 $\Omega \in \mathcal{A}$.

(ii) Assume $A \in \mathcal{A}$. 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 $A^c \in \mathcal{A}$.

(iii) Assume $A_1, A_2, \ldots \in \mathcal{A}$. 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 $A \in \mathcal{A}$.

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. \qed

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

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

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

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

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 $\sigma$-algebra if and only if $\sigma(\mathcal{A}) = \mathcal{A}$.

The same statements hold for $\lambda$-systems. Furthermore, $\delta(\mathcal{E}) \subset \sigma(\mathcal{E})$. \diamond
Theorem 1.18 (∩-closed λ-system). Let $\mathcal{D} \subset 2^\Omega$ be a λ-system. Then

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

Proof. “$\iff$” This is obvious.

“$\Rightarrow$” 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) ($\sigma\cup$-closedness) Let $A, B \in \mathcal{D}$. By assumption, $A \cap B \in \mathcal{D}$, and hence trivially $A \cap B \subset A$. Thus $A \setminus B = A \setminus (A \cap B) \in \mathcal{D}$. This implies that $\mathcal{D}$ is $\cap$-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 $\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty B_n \in \mathcal{D}$.

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

Theorem 1.19 (Dynkin’s $\pi$-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

$$\mathcal{D}_B := \{A \in \delta(\mathcal{E}) : A \cap B \in \delta(\mathcal{E})\}.$$
In order to show that $\delta(\mathcal{E})$ is a $\pi$-system, it is enough to show that

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

In order to show that $\mathcal{D}_E$ is a $\lambda$-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

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

By assumption, $A \cap E \in \mathcal{E}$ if $A \in \mathcal{E}$; 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.

We are particularly interested in $\sigma$-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 $\tau \subset \Omega$ is called a **topology** on $\Omega$ 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) $\left( \bigcup_{A \in \mathcal{F}} A \right) \in \tau$ for any $\mathcal{F} \subset \tau$.

The pair $(\Omega, \tau)$ 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 $\sigma$-algebras, topologies are closed under finite intersections only, but they are also closed under arbitrary unions.

Let $d$ be a metric on $\Omega$, and denote the open ball with radius $r > 0$ centred 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

$$\tau = \left\{ \bigcup_{(x, r) \in \mathcal{F}} B_r(x) : \mathcal{F} \subset \Omega \times (0, \infty) \right\}.$$
Definition 1.21 (Borel σ-algebra). Let \((\Omega, \tau)\) be a topological space. The σ-algebra
\[ B(\Omega) := B(\Omega, \tau) := \sigma(\tau) \]
that is generated by the open sets is called the Borel σ-algebra on \(\Omega\). The elements \(A \in B(\Omega, \tau)\) are called Borel sets or Borel measurable sets.

Remark 1.22. In many cases, we are interested in \(B(\mathbb{R}^n)\), 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 [35, 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 \(B(\mathbb{R}^n)\) and hence \(C\) is a Borel set. In particular, \(\{x\} \in B(\mathbb{R}^n)\) for every \(x \in \mathbb{R}^n\).

(iii) \(B(\mathbb{R}^n)\) is not a topology. To show this, let \(V \subset \mathbb{R}^n\) such that \(V \notin B(\mathbb{R}^n)\). If \(B(\mathbb{R}^n)\) were a topology, then it would be closed under arbitrary unions. As \(\{x\} \in B(\mathbb{R}^n)\) for all \(x \in \mathbb{R}^n\), we would get the contradiction \(V = \bigcup_{x \in V} \{x\} \in B(\mathbb{R}^n)\).

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
\[ a < b \quad \text{if} \quad a_i < b_i \quad \text{for all} \quad i = 1, \ldots, n. \] (1.4)

For \(a < b\), we define the open rectangle as the Cartesian product
\[ (a, b) := \bigtimes_{i=1}^{n} (a_i, b_i) := (a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_n, b_n). \] (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:
\[ \mathcal{E}_1 := \{ A \subset \mathbb{R}^n : A \text{ is open} \}, \quad \mathcal{E}_2 := \{ A \subset \mathbb{R}^n : A \text{ is closed} \}, \]
\[ \mathcal{E}_3 := \{ A \subset \mathbb{R}^n : A \text{ is compact} \}, \quad \mathcal{E}_4 := \{ B_r(x) : x \in \mathbb{Q}^n, r \in \mathbb{Q}^+ \}, \]
\[ \mathcal{E}_5 := \{ (a, b) : a, b \in \mathbb{Q}^n, a < b \}, \quad \mathcal{E}_6 := \{ [a, b) : a, b \in \mathbb{Q}^n, a < b \}, \]
\[ \mathcal{E}_7 := \{ (a, b] : a, b \in \mathbb{Q}^n, a < b \}, \quad \mathcal{E}_8 := \{ [a, b] : a, b \in \mathbb{Q}^n, a < b \}, \]
\[ \mathcal{E}_9 := \{ (\infty, b) : b \in \mathbb{Q}^n \}, \quad \mathcal{E}_{10} := \{ (\infty, b] : b \in \mathbb{Q}^n \}, \]
\[ \mathcal{E}_{11} := \{ (a, \infty) : a \in \mathbb{Q}^n \}, \quad \mathcal{E}_{12} := \{ [a, \infty) : a \in \mathbb{Q}^n \}. \]

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

**Proof.** We only show 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 \( r(x) > \frac{1}{2} R(y), \) 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

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

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

**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 \( \pi \)-system. Hence, the Borel \( \sigma \)-algebra equals the generated \( \lambda \)-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. \( \diamond \)
Definition 1.25 (Trace of a class of sets). Let \( A \subset 2^\Omega \) be an arbitrary class of subsets of \( \Omega \) and let \( A \in 2^\Omega \setminus \{\emptyset\} \). The class
\[
A|_A := \{ A \cap B : B \in A \} \subset 2^A
\]
(1.6)
is called the \textit{trace} of \( A \) on \( A \) or the \textit{restriction} of \( A \) to \( A \).

Theorem 1.26. Let \( A \subset \Omega \) be a nonempty set and let \( A \) be a \( \sigma \)-algebra on \( \Omega \) or any of the classes of Definitions 1.6–1.10. Then \( A|_A \) is a class of sets of the same type as \( A \); however, on \( A \) instead of \( \Omega \).

Proof. This is left as an exercise. \( \square \)

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

Exercise 1.1.2. Give a counterexample that shows that, in general, the union \( A \cup A' \) of two \( \sigma \)-algebras need not be a \( \sigma \)-algebra.

Exercise 1.1.3. Let \( (\Omega_1, d_1) \) and \( (\Omega_2, d_2) \) 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 \( U_f \in B(\Omega_1) \).

Hint: First show that for any \( \varepsilon > 0 \) and \( \delta > 0 \) the set
\[
U_f^{\delta,\varepsilon} := \{ x \in \Omega_1 : \text{ there are } y, z \in B_\varepsilon(x) \text{ with } d_2(f(y), f(z)) > \delta \}
\]
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 \( \Omega \) be an uncountably infinite set and \( A = \sigma(\{\omega\} : \omega \in \Omega) \). Show that
\[
A = \{ A \subset \Omega : \text{A is countable or } A^c \text{ is countable} \}.
\]

Exercise 1.1.5. Let \( A \) be a ring on the set \( \Omega \). Show that \( A \) 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 $\mu$ is

(i) **monotone** if $\mu(A) \leq \mu(B)$ for any two sets $A, B \in \mathcal{A}$ with $A \subset B$,

(ii) **additive** if $\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i)$ 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) $\sigma$-**additive** if $\mu\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty \mu(A_i)$ 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 $\mu(A) \leq \sum_{i=1}^n \mu(A_i)$, and

(v) $\sigma$-**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 $\mu(A) \leq \sum_{i=1}^\infty \mu(A_i)$.

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

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

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

(i) **finite** if $\mu(A) < \infty$ for every $A \in \mathcal{A}$ and

(ii) $\sigma$-**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 $\delta_\omega(A) = 1_A(\omega)$ (see (1.2)). Then $\delta_\omega$ is a probability measure on any $\sigma$-algebra $\mathcal{A} \subset 2^\Omega$. $\delta_\omega$ is called the **Dirac measure** for the point $\omega$. 
(ii) Let \( \Omega \) be a finite nonempty set. By 
\[
\mu(A) := \frac{\#A}{\#\Omega} \quad \text{for} \quad A \subset \Omega,
\]
we define a probability measure on \( A = 2^\Omega \). This \( \mu \) is called the uniform distribution on \( \Omega \). For this distribution, we introduce the symbol \( U_\Omega := \mu \). The resulting triple \( (\Omega, A, U_\Omega) \) is called a Laplace space.

(iii) Let \( \Omega \) be countably infinite and let 
\[
A := \{ A \subset \Omega : \#A < \infty \text{ or } \#A^c < \infty \}.
\]
Then \( A \) is an algebra. The set function \( \mu \) on \( A \) defined by 
\[
\mu(A) = \begin{cases} 
0, & \text{if } A \text{ is finite}, \\
\infty, & \text{if } A^c \text{ is finite},
\end{cases}
\]
is a content but is not a premeasure. Indeed, \( \mu \left( \bigcup_{\omega \in \Omega} \{ \omega \} \right) = \mu(\Omega) = \infty \), 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 \( \Omega \) be an (at most) countable nonempty set and let \( A = 2^\Omega \). Further, let \( (p_\omega)_{\omega \in \Omega} \) be nonnegative numbers. Then \( A \mapsto \mu(A) := \sum_{\omega \in A} p_\omega \) defines a \( \sigma \)-finite measure on \( 2^\Omega \). We call \( p = (p_\omega)_{\omega \in \Omega} \) the weight function of \( \mu \). The number \( p_\omega \) is called the weight of \( \mu \) at point \( \omega \).

(vi) If in (v) the sum \( \sum_{\omega \in \Omega} p_\omega \) equals one, then \( \mu \) is a probability measure. In this case, we interpret \( p_\omega \) as the probability of the elementary event \( \omega \). 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 \( \mu \) is called counting measure on \( \Omega \). If \( \Omega \) is finite, then so is \( \mu \).

(viii) Let \( A \) 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 \( A = \bigcup_{i=1}^{n} (a_i, b_i] \), define 
\[
\mu(A) = \sum_{i=1}^{n} (b_i - a_i).
\]
Then \( \mu \) is a \( \sigma \)-finite content on \( A \) (even a premeasure) since \( \bigcup_{n=1}^{\infty} (-n, n] = \mathbb{R} \) and 
\[
\mu((-n, n]) = 2n < \infty \text{ for all } n \in \mathbb{N}.
\]
Lemma 1.31 (Properties of contents). Let \( \mathcal{A} \) be a semiring and let \( \mu \) 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) \( \mu \) 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) \( \mu \) is subadditive. If \( \mu \) is \( \sigma \)-additive, then \( \mu \) is also \( \sigma \)-subadditive.

(iv) If \( \mathcal{A} \) is a ring, then \( \sum_{n=1}^{\infty} \mu(A_n) \leq \mu \left( \bigcup_{n=1}^{\infty} A_n \right) \) 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 \( \mu \) 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 \( B \setminus A = \bigcup_{i=1}^{n} C_i \). Hence \( \mu(B) = \mu(A) + \sum_{i=1}^{n} \mu(C_i) \geq \mu(A) \) and thus \( \mu \) 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

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

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

\[
\bigcup_{i=1}^{c_k} C_{k,i} = B_k \subset A_k.
\]

Similarly, there exist \( d_k \in \mathbb{N} \) and \( D_{k,1}, \ldots, D_{k,d_k} \in \mathcal{A} \) such that

\[
A_k \setminus B_k = \bigcup_{i=1}^{d_k} D_{k,i}.
\]

Since \( \mu \) is additive, we have

\[
\mu(A_k) = \sum_{i=1}^{c_k} \mu(C_{k,i}) + \sum_{i=1}^{d_k} \mu(D_{k,i}) \geq \sum_{i=1}^{c_k} \mu(C_{k,i}).
\]
Again due to additivity and monotonicity, we get
\[
\mu(A) = \mu\left( \bigcup_{k=1}^{n} \bigcup_{i=1}^{c_k} (C_{k,i} \cap A) \right) = \sum_{k=1}^{n} \sum_{i=1}^{c_k} \mu(C_{k,i} \cap A) \leq \sum_{k=1}^{n} \sum_{i=1}^{c_k} \mu(C_{k,i}) \leq \sum_{k=1}^{n} \mu(A_k).
\]
Hence \(\mu\) is subadditive. By a similar argument, \(\sigma\)-subadditivity follows from \(\sigma\)-additivity.

(iv) Let \(\mathcal{A}\) be a ring and let \(A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}\). Since \(\mu\) is additive (and thus monotone), we have by (ii)
\[
\sum_{n=1}^{m} \mu(A_n) = \mu\left( \bigcup_{n=1}^{m} A_n \right) \leq \mu(A) \quad \text{for any } m \in \mathbb{N}.
\]
It follows that \(\sum_{n=1}^{\infty} \mu(A_n) \leq \mu(A)\). \(\square\)

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. \(\diamond\)

Theorem 1.33 (Inclusion-exclusion formula). Let \(\mathcal{A}\) be a ring and let \(\mu\) be a content on \(\mathcal{A}\). Let \(n \in \mathbb{N}\) and \(A_1, \ldots, A_n \in \mathcal{A}\). 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 \(n\). \(\square\)

The next goal is to characterise \(\sigma\)-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 $\mu$ be a content on the ring $\mathcal{A}$.

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

(ii) $\mu$ is called **upper semicontinuous** if $\mu(A_n) \xrightarrow{n \to \infty} \mu(A)$ 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) $\mu$ 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 $\mu$ 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 $\mu$ be a content on the ring $\mathcal{A}$. Consider the following five properties.

(i) $\mu$ is $\sigma$-additive (and hence a premeasure).

(ii) $\mu$ is $\sigma$-subadditive.

(iii) $\mu$ is lower semicontinuous.

(iv) $\mu$ is $\emptyset$-continuous.

(v) $\mu$ is upper semicontinuous.

Then the following implications hold:

(i) $\iff$ (ii) $\iff$ (iii) $\implies$ (iv) $\iff$ (v).

If $\mu$ is finite, then we also have (iv) $\implies$ (iii).

Proof. “(i) $\implies$ (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 $A = \bigcup_{n=1}^{\infty} (A \cap B_n)$. Since $\mu$ is monotone and $\sigma$-additive, we infer

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

Hence $\mu$ is $\sigma$-subadditive.

“(ii) $\implies$ (i)” This follows from Lemma 1.31(iv).

“(i) $\implies$ (iii)” Let $\mu$ 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) ⇒ (i)” Assume now that (iii) holds. Let \( B_1, B_2, \ldots \in \mathcal{A} \) be mutually disjoint, and assume that \( B = \bigcup_{n=1}^{\infty} B_n \in \mathcal{A} \). 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 \( \mu \) is \( \sigma \)-additive and therefore a premeasure.

“(iv) ⇒ (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 \( \mu(A_n) - \mu(A) = \mu(B_n) \xrightarrow{n \to \infty} 0 \).

“(v) ⇒ (iv)” This is evident.

“(iii) ⇒ (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) ⇒ (iii)” (for finite \( \mu \)) Assume that \( \mu(A) < \infty \) for every \( A \in \mathcal{A} \) and that \( \mu \) 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

\[
\mu(A) - \mu(A_n) = \mu(A \setminus A_n) \xrightarrow{n \to \infty} 0.
\]

Hence (iii) follows.

\[\Box\]

**Example 1.37.** (Compare Example 1.30(iii).) Let \( \Omega \) be a countable set, and define

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

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

Then \( \mu \) is an \( \emptyset \)-continuous content but not a premeasure.
Definition 1.38. (i) A pair \((\Omega, \mathcal{A})\) consisting of a nonempty set \(\Omega\) and a \(\sigma\)-algebra \(\mathcal{A} \subset 2^\Omega\) is called a measurable space. The sets \(A \in \mathcal{A}\) are called measurable sets. If \(\Omega\) 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 \(\mu\) 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 \(\sigma\)-finite measures on \((\Omega, \mathcal{A})\) is denoted by \(\mathcal{M}_{\sigma}(\Omega, \mathcal{A})\).

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

1.3 The Measure Extension Theorem

In this section, we construct measures \(\mu\) on \(\sigma\)-algebras. The starting point will be to define the values of \(\mu\) 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 \(\sigma\)-algebra.

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

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

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

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 \(\mu\) to a (uniquely determined) measure on the Borel \(\sigma\)-algebra \(B(\mathbb{R}^n) = \sigma(A)\)? We will see that this is indeed possible. The resulting measure is called Lebesgue measure (or sometimes Lebesgue-Borel measure) \(\lambda\) on \((\mathbb{R}^n, B(\mathbb{R}^n))\).

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 realisation 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 $\omega_1, \ldots, \omega_n$:

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

Let $A_0 = \{\emptyset\}$. For $n \in \mathbb{N}$, define the class of cylinder sets that depend only on the first $n$ coordinates

$$A_n := \{[\omega_1, \ldots, \omega_n] : \omega_1, \ldots, \omega_n \in E\}, \quad (1.8)$$

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

We interpret $[\omega_1, \ldots, \omega_n]$ as the event where the outcome of the first experiment is $\omega_1$, the outcome of the second experiment is $\omega_2$ and finally the outcome of the $n$th experiment is $\omega_n$. 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,

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

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

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

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

Hence $(\Omega, d)$ is a compact metric space. Clearly,

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

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

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

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

Exercise: Prove the statements made above. 

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

**Theorem 1.41 (Carathéodory).** Let $A \subset 2^\Omega$ be a ring and let $\mu$ be a $\sigma$-finite premeasure on $A$. There exists a unique measure $\tilde{\mu}$ on $\sigma(A)$ such that $\tilde{\mu}(A) = \mu(A)$ for all $A \in A$. Furthermore, $\tilde{\mu}$ is $\sigma$-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 ∩-closed generator).** Let \((\Omega, A, \mu)\) be a \(\sigma\)-finite measure space and let \(E \subset A\) be a \(\pi\)-system that generates \(A\). Assume that there exist sets \(E_1, E_2, \ldots \in \mathcal{E}\) such that \(E_n \uparrow \Omega\) and \(\mu(E_n) < \infty\) for all \(n \in \mathbb{N}\). Then \(\mu\) is uniquely determined by the values \(\mu(E), E \in \mathcal{E}\).

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

**Proof.** Let \(\nu\) be a (possibly different) \(\sigma\)-finite measure on \((\Omega, A)\) 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 
\[
\mathcal{D}_E = \{ A \in A : \mu(A \cap E) = \nu(A \cap E) \}.
\]

In order to show that \(\mathcal{D}_E\) is a \(\lambda\)-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 
\[
\mu((A \setminus B) \cap E) = \mu(A \cap E) - \mu(B \cap E)
= \nu(A \cap E) - \nu(B \cap E) = \nu((A \setminus B) \cap E).
\]

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 \(\pi\)-system, Theorem 1.19 yields 
\[
A \supset \mathcal{D}_E \supset \delta(\mathcal{E}) = \sigma(\mathcal{E}) = A.
\]

Hence \(\mathcal{D}_E = A\).

This implies \(\mu(A \cap E) = \nu(A \cap E)\) for any \(A \in A\) and \(E \in \mathcal{E}\) with \(\mu(E) < \infty\). Now let \(E_1, E_2, \ldots \in \mathcal{E}\) be a sequence such that \(E_n \uparrow \Omega\) and \(\mu(E_n) < \infty\) for all \(n \in \mathbb{N}\). Since \(\mu\) and \(\nu\) are lower semicontinuous, for all \(A \in A\), we have 
\[
\mu(A) = \lim_{n \to \infty} \mu(A \cap E_n) = \lim_{n \to \infty} \nu(A \cap E_n) = \nu(A).
\]
The additional statement is trivial as $\mathcal{E} := \mathcal{E} \cup \{\Omega\}$ is a $\pi$-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 $\mu$ 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 $\pi$-system and $\sigma(\mathcal{E}) = 2^\Omega$. Hence a finite measure $\mu$ on $(\Omega, 2^\Omega)$ is uniquely determined by the values $\mu(E_n)$, $n \in \mathbb{Z}$.

However, a $\sigma$-finite measure on $\mathbb{Z}$ is not uniquely determined by the values on $\mathcal{E}$: Let $\mu$ 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 $\mu$ and $\nu$ one needs a generator that contains sets of finite measure (of $\mu$). Do the sets $F_n = [-n, n] \cap \mathbb{Z}$, $n \in \mathbb{N}$ do the trick? Indeed, for any $\sigma$-finite measure $\mu$, we have $\mu(F_n) < \infty$ for all $n \in \mathbb{N}$. However, the sets $F_n$ do not generate $2^\Omega$ (but which $\sigma$-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 $\pi$-system that generates $2^\Omega$ and such that $\mu(F_n) < \infty$ for all $n \in \mathbb{N}$. The conditions of the theorem are fulfilled as $F_n \uparrow \Omega$. \hfill \Box

Example 1.44 (Distribution function). A probability measure $\mu$ 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 $\pi$-system that generates $\mathcal{B}(\mathbb{R}^n)$ (see Theorem 1.23). In particular, a probability measure $\mu$ on $\mathbb{R}$ is uniquely determined by its distribution function $F : \mathbb{R} \to [0, 1], x \mapsto \mu((-\infty, x])$. \hfill \Box

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 $\pi$-system. In fact, here a probability measure $\mu$ is not uniquely determined by the values, say $\mu(\{1, 2\}) = \mu(\{2, 3\}) = \frac{1}{2}$. We just give 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 $\pi$-system that generates $2^\Omega$. Hence a probability measure $\mu$ 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 $\mathcal{E}$. \hfill \Box
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) \( \mu^* \) is monotone,

(iii) \( \mu^* \) is \( \sigma \)-subadditive.

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

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

Define

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

where \( \inf \emptyset = \infty \). Then \( \mu^* \) is an outer measure. If in addition \( \mu \) is \( \sigma \)-subadditive, then \( \mu^*(A) = \mu(A) \) for all \( A \in A \).

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

(i) Since \( \emptyset \in A \), we have \( \{ \emptyset \} \in U(\emptyset) \); hence \( \mu^*(\emptyset) = 0 \).

(ii) If \( A \subset B \), then \( U(A) \supset U(B) \); hence \( \mu^*(A) \leq \mu^*(B) \).

(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 \( \mu^*(A) \leq \sum_{n=1}^{\infty} \mu^*(A_n) \). Without loss of generality, assume \( \mu^*(A_n) < \infty \) and hence \( U(A_n) \neq \emptyset \) for all \( n \in \mathbb{N} \). Fix \( \varepsilon > 0 \). For every \( n \in \mathbb{N} \), choose a covering \( F_n \in U(A_n) \) such that

\[
\sum_{F \in F_n} \mu(F) \leq \mu^*(A_n) + \varepsilon 2^{-n}.
\]

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

\[
\mu^*(A) \leq \sum_{F \in F} \mu(F) \leq \sum_{n=1}^{\infty} \sum_{F \in F_n} \mu(F) \leq \sum_{n=1}^{\infty} \mu^*(A_n) + \varepsilon.
\]

Let \( A \in A \). Since \( \{ A \} \in U(A) \), we have \( \mu^*(A) \leq \mu(A) \). If \( \mu \) is \( \sigma \)-subadditive, then for any \( F \in U(A) \), we have \( \sum_{F \in F} \mu(F) \geq \mu(A) \); hence \( \mu^*(A) \geq \mu(A) \).

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

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

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
\[
\mu^*(A \cap E) + \mu^*(A^c \cap E) \leq \mu^*(E) \quad \text{for any } E \in 2^\Omega.
\]

Proof. As \( \mu^* \) is subadditive, the other inequality is trivial. \( \square \)

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) (\( \pi \)-system) Let \( A, B \in \mathcal{M}(\mu^*) \) and \( E \in 2^\Omega \). Then
\[
\begin{align*}
\mu^* ((A \cap B) \cap E) + \mu^* ((A \cap B)^c \cap E) &= \mu^*(A \cap B \cap E) + \mu^*(A \cap B^c \cap E) \\
&\leq \mu^*(A \cap B \cap E) + \mu^*(A^c \cap B \cap E) \\
&\quad + \mu^*(A^c \cap B^c \cap E) + \mu^*(A \cap B^c \cap E) \\
&= \mu^*(B \cap E) + \mu^*(B^c \cap E) \\
&= \mu^*(E).
\end{align*}
\]

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

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

Proof. Let \( A, B \in \mathcal{M}(\mu^*) \) with \( A \cap B = \emptyset \). Then
\[
\mu^*(A \cup B) = \mu^*(A \cap (A \cup B)) + \mu^*(A^c \cap (A \cup B)) = \mu^*(A) + \mu^*(B).
\]
Inductively, we get (finite) additivity. By definition, \( \mu^* \) is \( \sigma \)-subadditive; hence we conclude by Theorem 1.36 that \( \mu^* \) is also \( \sigma \)-additive. \( \square \)

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

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

Hence, let \( A_1, A_2, \ldots \in \mathcal{M}(\mu^*) \) be mutually disjoint, and define \( A := \bigcup_{n=1}^\infty A_n \). We have to show \( A \in \mathcal{M}(\mu^*) \); that is,
\[
\mu^*(A \cap E) + \mu^*(A^c \cap E) \leq \mu^*(E) \quad \text{for any } E \in 2^\Omega. \tag{1.11}
\]
Let $B_n = \bigcup_{i=1}^{n} A_i$ for all $n \in \mathbb{N}$. For all $n \in \mathbb{N}$, we have
\[
\mu^*(E \cap B_{n+1}) = \mu^*((E \cap B_{n+1}) \cap B_n) + \mu^*((E \cap B_{n+1}) \cap B_n^c) \\
= \mu^*(E \cap B_n) + \mu^*(E \cap A_{n+1}).
\]
Inductively, we get $\mu^*(E \cap B_n) = \sum_{i=1}^{n} \mu^*(E \cap A_i)$. The monotonicity of $\mu^*$ now implies that
\[
\mu^*(E) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \geq \mu^*(E \cap B_n) + \mu^*(E \cap A^c) \\
= \sum_{i=1}^{n} \mu^*(E \cap A_i) + \mu^*(E \cap A^c).
\]
Letting $n \to \infty$ and using the $\sigma$-subadditivity of $\mu^*$, we conclude
\[
\mu^*(E) \geq \sum_{i=1}^{\infty} \mu^*(E \cap A_i) + \mu^*(E \cap A^c) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c).
\]
Hence (1.11) holds and the proof is complete. \(\square\)

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, $\sigma$-subadditive and $\sigma$-finite set function with $\mu(\emptyset) = 0$. Then there is a unique $\sigma$-finite measure $\tilde{\mu} : \sigma(\mathcal{A}) \to [0, \infty]$ such that $\tilde{\mu}(A) = \mu(A)$ for all $A \in \mathcal{A}$.

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

In order to establish the existence of $\tilde{\mu}$, 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 1.31(ii), $\mu^*$ is monotone. Hence $\mu^*$ is an outer measure by Lemma 1.47 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 $\sigma$-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
\[
E \subset \bigcup_{n=1}^{\infty} E_n \quad \text{and} \quad \sum_{n=1}^{\infty} \mu(E_n) \leq \mu^*(E) + \varepsilon.
\]
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 $E_n \setminus A = E_n \setminus B_n = \bigcup_{k=1}^{m_n} C_n^k$. Hence

$$E \cap A \subset \bigcup_{n=1}^{\infty} B_n, \quad E \cap A^c \subset \bigcup_{n=1}^{\infty} \bigcup_{k=1}^{m_n} C_n^k$$

and

$$E_n = B_n \bigcup \bigcup_{k=1}^{m_n} C_n^k.$$

$\mu^*$ is $\sigma$-subadditive and by assumption $\mu$ is additive. From $\mu^*|_A \leq \mu$ (we will see that even equality holds), we infer that

$$\mu^*(E \cap A) + \mu^*(E \cap A^c) \leq \sum_{n=1}^{\infty} \mu^*(B_n) + \sum_{n=1}^{\infty} \sum_{k=1}^{m_n} \mu(C_n^k)$$

$$= \sum_{n=1}^{\infty} \left( \mu(B_n) + \mu^*(C_n^k) \right)$$

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

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

**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 $\sigma$-algebra $\mathcal{B}(\mathbb{R}^n)$. In order to check the assumptions of Theorem 1.53, we only have to check that $\mu$ is $\sigma$-subadditive. To this end, let $(a, b], (a(1), b(1)], (a(2), b(2)], \ldots \in \mathcal{A}$ with

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

We show that

$$\mu((a, b]) \leq \sum_{k=1}^{\infty} \mu((a(k), b(k)]). \quad (1.12)$$

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

$$\mu((a(k), b_\varepsilon(k)]) \leq \mu((a(k), b(k)]) + \varepsilon 2^{-k-1}.$$
Further choose $a_\varepsilon \in (a, b)$ such that $\mu((a_\varepsilon, b]) \geq \mu((a, b]) - \varepsilon/2$. Now $[a_\varepsilon, b]$ is compact and
\[
\bigcup_{k=1}^{\infty} (a(k), b_\varepsilon(k)) \supset \bigcup_{k=1}^{\infty} (a(k), b(k)] \supset (a, b] \supset [a_\varepsilon, b],
\]
whence there exists a $K_0$ such that $\bigcup_{k=1}^{K_0} (a(k), b_\varepsilon(k)) \supset (a_\varepsilon, b]$. As $\mu$ is (finitely) subadditive (see Lemma 1.31(iii)), we obtain
\[
\mu((a, b]) \leq \varepsilon/2 + \mu((a_\varepsilon, b]) \leq \varepsilon/2 + \sum_{k=1}^{K_0} \mu((a(k), b_\varepsilon(k)]
\]
\[
\leq \varepsilon/2 + \sum_{k=1}^{K_0} (\varepsilon 2^{-k-1} + \mu((a(k), b(k)])) \leq \varepsilon + \sum_{k=1}^{\infty} \mu((a(k), b(k)]).
\]

Letting $\varepsilon \downarrow 0$ yields (1.12); hence $\mu$ is $\sigma$-subadditive.\hfill \diamond

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 $\lambda^n$ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ with the property that
\[
\lambda^n((a, b]) = \prod_{i=1}^{n} (b_i - a_i) \quad \text{for all } a, b \in \mathbb{R}^n \text{ with } a < b.
\]

$\lambda^n$ 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 $\mathcal{A} = \{(a, b) : a, b \in \mathbb{R}, a \leq b\}$. $\mathcal{A}$ is a semiring and $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$, where $\mathcal{B}(\mathbb{R})$ is the Borel $\sigma$-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
\[
F(b_\varepsilon(k)) - F(b(k)) < \varepsilon 2^{-k-1}.
\]

As in Example 1.54, it can be shown that $\tilde{\mu}_F((a, b]) \leq \varepsilon + \sum_{k=1}^{\infty} \tilde{\mu}_F((a(k), b(k)]).$ This implies that $\tilde{\mu}_F$ is $\sigma$-subadditive. By Theorem 1.53, we can extend $\tilde{\mu}_F$ uniquely to a $\sigma$-finite measure $\mu_F$ on $\mathcal{B}(\mathbb{R})$.\hfill \diamond
Definition 1.57 (Lebesgue-Stieltjes measure). The measure $\mu_F$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined by

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

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 $\mu_F$ 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 $\alpha_n \geq 0$ for all $n \in \mathbb{N}$ such that $\sum_{n=1}^{\infty} \alpha_n < \infty$. Then $F = \sum_{n=1}^{\infty} \alpha_n \mathbb{I}_{[x_n, \infty)}$ 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 $\sigma$-finite measure. Then $\mu$ 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 $x \geq 0$ and $F(x) = -\# \{n \in \mathbb{N} : x_n \in [x, 0]\}$, then $\mu = \mu_F$. On the other hand, if $\mu$ 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 $\mu_F$ is a probability measure.

We will now have a closer look at the case where $\mu_F$ 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 $F(\infty) \leq 1$ instead of $F(\infty) = 1$, then $F$ is called a (possibly) defective p.d.f. If $\mu$ 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 $\mu$.

Clearly, $F_{\mu}$ is right continuous and $F(-\infty) = 0$, since $\mu$ is upper semicontinuous and finite (Theorem 1.36). Since $\mu$ is lower semicontinuous, we have $F(\infty) = \mu(\mathbb{R})$; hence $F_{\mu}$ is indeed a (possibly defective) distribution function if $\mu$ 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 \( \sigma \)-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 \( \sigma \)-finite measure \( \mu \) on \((\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))\) such that

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

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 \( \mu_i \). 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 \( \sigma \)-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 \( \mu \) on \( \mathcal{A} = \{[\omega_1, \ldots, \omega_n] : \omega_1, \ldots, \omega_n \in E, n \in \mathbb{N}\} \) by

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

We aim at extending \( \mu \) to a measure on \( \sigma(\mathcal{A}) \). In order to check the assumptions of Theorem 1.53, we have to show that \( \mu \) is \( \sigma \)-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 \( \mu \) (see Lemma 1.31(iii)), this implies \( \mu(A) \leq \sum_{n=1}^{N} \mu(A_n) \leq \sum_{n=1}^{\infty} \mu(A_n) \); hence \( \mu \) is \( \sigma \)-subadditive.

We now give two different proofs for (1.13).

1. **Proof.** The metric \( d \) from (1.9) induces the product topology on \( \Omega \); hence, as remarked in Example 1.40, \((\Omega, d)\) 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.

2. **Proof.** We now show by elementary means the validity of (1.13). The procedure imitates the proof that \( \Omega \) 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. \( \diamond \)

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 non-empty set and \( \Omega = E^\mathbb{N} \). Let \((p_e)_{e \in E}\) be a probability vector. Then there exists a unique probability measure \( \mu \) on \( \sigma(A) = \mathcal{B}(\Omega) \) such that

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

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

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 \( \sigma(A) \) that were first defined on a semiring \( A \) only. The following theorem, however, shows that the measure of a set from \( \sigma(A) \) can be well approximated by finite and countable operations with sets from \( A \).

Denote by \( A \triangle B := (A \setminus B) \cup (B \setminus A) \) (1.14) the symmetric difference of the two sets \( A \) and \( B \).

Theorem 1.65 (Approximation theorem for measures). Let \( A \subset 2^\Omega \) be a semiring and let \( \mu \) be a measure on \( \sigma(A) \) that is \( \sigma \)-finite on \( A \).

(i) For any \( A \in \sigma(A) \) and \( \varepsilon > 0 \), there exist mutually disjoint sets \( A_1, A_2, \ldots \in A \) such that \( A \subset \bigcup_{n=1}^{\infty} A_n \) and \( \mu \left( \bigcup_{n=1}^{\infty} A_n \setminus A \right) < \varepsilon \).

(ii) For any \( A \in \sigma(A) \) with \( \mu(A) < \infty \) and any \( \varepsilon > 0 \), there exists an \( n \in \mathbb{N} \) and mutually disjoint sets \( A_1, \ldots, A_n \in A \) such that \( \mu \left( A \triangle \bigcup_{k=1}^{n} A_k \right) < \varepsilon \).

(iii) For any \( A \in \mathcal{M}(\mu^*) \), there are sets \( A_-, A_+ \in \sigma(A) \) 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 \( \mu^* \) instead of \( \mu \)). If \( A \) is an algebra, then in (ii) for any \( A \in \sigma(A) \), we even have \( \inf_{B \in A} \mu(A \triangle B) = 0 \). \( \diamond \)

Proof. (ii) As \( \mu \) and the outer measure \( \mu^* \) coincide on \( \sigma(A) \) and since \( \mu(A) \) is finite on \( \sigma(A) \), by the very definition of \( \mu^* \) (see Lemma 1.47) there exists a covering \( B_1, B_2, \ldots \in A \) 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
\[
C \triangle D = (D \setminus C) \cup (C \setminus D) \subset (D \setminus C) \cup (C \setminus (D \cup E)) \cup E \subset (C \triangle (D \cup E)) \cup E.
\]
Choosing \( C = A, D = \bigcup_{i=1}^{n} B_i \) and \( E = \bigcup_{i=n+1}^{\infty} B_i \), this yields
\[
\mu \left( A \triangle \bigcup_{i=1}^{n} B_i \right) \leq \mu \left( A \triangle \bigcup_{i=1}^{\infty} B_i \right) + \mu \left( \bigcup_{i=n+1}^{\infty} B_i \right)
\leq \mu \left( \bigcup_{i=1}^{\infty} B_i \right) - \mu(A) + \frac{\varepsilon}{2} \leq \varepsilon.
\]
As \( \mathcal{A} \) is a semiring, there exist a \( k \in \mathbb{N} \) and \( A_1, \ldots, A_k \in \mathcal{A} \) such that
\[
\bigcup_{i=1}^{n} B_i = B_1 \uplus \bigcup_{i=2}^{n} \left( \bigcap_{j=1}^{i-1} (B_i \setminus B_j) \right) =: \biguplus_{i=1}^{k} A_i.
\]

(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
\[
\mu(A \cap E_n) \geq \sum_{m=1}^{\infty} \mu(B_{n,m}) - 2^{-n} \varepsilon.
\]
(This is possible due to the definition of the outer measure \( \mu^* \), which coincides with \( \mu \) on \( \mathcal{A} \).) Let \( \bigcup_{m,n=1}^{\infty} B_{n,m} = \bigcup_{n=1}^{\infty} A_n \) for certain \( A_n \in \mathcal{A}, n \in \mathbb{N} \) (Exercise 1.1.1). Then
\[
\mu \left( \bigcup_{n=1}^{\infty} A_n \setminus A \right) = \mu \left( \bigcup_{n=1}^{\infty} \bigcup_{m=1}^{\infty} B_{n,m} \setminus A \right)
\leq \mu \left( \bigcup_{n=1}^{\infty} \bigcup_{m=1}^{\infty} (B_{n,m} \setminus (A \cap E_n)) \right)
\leq \sum_{n=1}^{\infty} \left( \sum_{m=1}^{\infty} \mu(B_{n,m}) - \mu(A \cap E_n) \right) \leq \varepsilon.
\]
(iii) Let \( A \in \mathcal{M}(\mu^*) \) and \((E_n)_{n\in\mathbb{N}}\) as above. For any \( m, n \in \mathbb{N} \), choose \( A_{n,m} \in \sigma(A) \) such that \( A_{n,m} \supset A \cap E_n \) and \( \mu^*(A_{n,m}) \leq \mu^*(A \cap E_n) + \frac{2^{-n}}{m} \).

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

**Remark 1.67 (Regularity of measures).** (Compare with Theorem 13.6.) Let \( \lambda^n \) 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

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

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 \( \lambda^n \)). 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 \( \lambda^n \) 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 \( \lambda^n \) 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 \( K := [-N, N]^n \setminus U \subset A \). \( \diamond \)

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

(i) A set \( A \in \mathcal{A} \) is called a \( \mu \)-null set, or briefly a null set, if \( \mu(A) = 0 \). By \( \mathcal{N}_\mu \) we denote the class of all subsets of \( \mu \)-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 \( \mu \)-almost everywhere (a.e.) or for almost all (a.a.) \( \omega \) 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 \( A, B \in \mathcal{A} \), and assume that there is a null set \( N \) such that \( A \triangle B \subset N \). Then we write \( A = B \pmod{\mu} \).
Definition 1.69. A measure space \((\Omega, \mathcal{A}, \mu)\) is called \textbf{complete} if \(\mathcal{N}_\mu \subseteq \mathcal{A}\).

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

\[
\left(\Omega, \mathcal{M}(\mu^*), \left.\mu^*\right|_{\mathcal{M}(\mu^*)}\right).
\]

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 sequel, we will not need these statements. Hence, instead of giving a proof, we refer to the textbooks on measure theory (e.g., [35]).

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

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

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

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 \(\sigma\)-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 \(\sigma\)-algebra, Dirac measures for different points \(\omega \in \Omega\) cannot be distinguished. \(\diamondsuit\)

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

\[
\mu|_{\Omega'}(A) := \mu(A) \quad \text{for} \ A \in \mathcal{A} \text{ with } A \subseteq \Omega'.
\]

This measure is called the \textbf{restriction} of \(\mu\) to \(\Omega'\).

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

Later we will see (Corollary 1.84) that \(\mathcal{B}(\mathbb{R})|_A = \mathcal{B}(A)\), where \(\mathcal{B}(A)\) is the Borel \(\sigma\)-algebra on \(A\) that is generated by the (relatively) open subsets of \(A\). \(\diamondsuit\)
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 \( \mathcal{B}(\mathbb{R}^n)|_A \) by
\[
\mu(B) := \frac{\lambda^n(B)}{\lambda^n(A)} \quad \text{for } B \in \mathcal{B}(\mathbb{R}^n) \text{ with } B \subset A.
\]
This measure \( \mu \) is called the uniform distribution on \( A \) and will be denoted by \( U_A := \mu \).

Exercise 1.3.1. Show the following generalisation 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 \( \sum_{n:|x_n| \leq K} \alpha_n < \infty \) for all \( K > 0 \).

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

(i) Give a characterisation of \( A \) as in Exercise 1.1.4 (page 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 \( \mu \) is a measure on \((\Omega, \mathcal{A})\).

*Hint:* In particular, one has to show that \( \mu \) 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
\[
X^{-1}(A') \in \mathcal{A} \quad \text{for any } A' \in \mathcal{A}'.
\]
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 \( \sigma \)-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 $\text{id} : \Omega \rightarrow \Omega$ is $\mathcal{A} - \mathcal{A}$-measurable.

(ii) If $\mathcal{A} = 2^\Omega$ or $\mathcal{A}' = \{\emptyset, \Omega'\}$, then any map $X : \Omega \rightarrow \Omega'$ is $\mathcal{A} - \mathcal{A}'$-measurable.

(iii) Let $A \subset \Omega$. The indicator function $\mathbb{1}_A : \Omega \rightarrow \{0, 1\}$ is $\mathcal{A} - 2^{\{0,1\}}$-measurable if and only if $A \in \mathcal{A}$. $\diamond$

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

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

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

Proof. This is left as an exercise. $\Box$

We now consider $\sigma$-algebras that are generated by more than one map.

Definition 1.79 (Generated $\sigma$-algebra). Let $\Omega$ 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 $X_i : \Omega \rightarrow \Omega_i$ 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 $\sigma$-algebra on $\Omega$ that is generated by $(X_i, i \in I)$. This is the smallest $\sigma$-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 $X : \Omega \rightarrow \Omega'$ and $X' : \Omega' \rightarrow \Omega''$ be measurable maps. Then the map $Y := X' \circ X : \Omega \rightarrow \Omega''$, $\omega \mapsto X'(X(\omega))$ is $\mathcal{A} - \mathcal{A}''$-measurable.

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

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

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

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 $\sigma$-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

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

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

(iii) ($\sigma$-$\cup$-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 $A' = \sigma(X_i : i \in I)$. Then the following holds: A map $Y : \Omega \to \Omega'$ is $A - \mathcal{A}'$-measurable if and only if $X_i \circ Y$ is $A - \mathcal{A}_i$-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 $A - \mathcal{A}_i$-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 \( \sigma \)-algebra).** Let \( \mathcal{E} \subset 2^\Omega \) and assume that \( A \subset \Omega \) is nonempty. Then \( \sigma (\mathcal{E}|_A) = \sigma (\mathcal{E})|_A \).

**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

\[
\sigma (\mathcal{E}|_A) = \sigma (\{ E \cap A : E \in \mathcal{E} \}) = \sigma (\{ X^{-1}(E) : E \in \mathcal{E} \}) = X^{-1}(\sigma (\mathcal{E})) = \{ A \cap B : B \in \sigma (\mathcal{E}) \} = \sigma (\mathcal{E})|_A.
\]

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

**Corollary 1.84 (Trace of the Borel \( \sigma \)-algebra).** Let \( (\Omega, \tau) \) be a topological space and let \( A \subset \Omega \) be a subset of \( \Omega \). Then

\[
B(\Omega, \tau)|_A = B(A, \tau|_A).
\]

**Example 1.85.** (i) Let \( \Omega' \) be countable. Then \( X : \Omega \to \Omega' \) is \( A - 2^{\Omega'} \)-measurable if and only if \( X^{-1}(\{ \omega' \}) \in A \) for all \( \omega' \in \Omega' \). If \( \Omega' \) is uncountably infinite, this is wrong in general. (For example, consider \( \Omega = \Omega' = \mathbb{R}, A = B(\mathbb{R}) \), and \( X(\omega) = \omega \) for all \( \omega \in \Omega \). Clearly, \( X^{-1}(\{ \omega \}) = \{ \omega \} \in B(\mathbb{R}) \). If, on the other hand, \( A \subset \mathbb{R} \) is not in \( B(\mathbb{R}) \), then \( A \in 2^\mathbb{R} \), but \( X^{-1}(A) \notin B(\mathbb{R}) \).)

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

\[
[x] := \max \{ k \in \mathbb{Z} : k \leq x \} \text{ and } \lceil x \rceil := \min \{ k \in \mathbb{Z} : k \geq x \}. \quad (1.16)
\]

The maps \( \mathbb{R} \to \mathbb{Z}, x \mapsto \lfloor x \rfloor \) and \( x \mapsto \lceil x \rceil \) are \( B(\mathbb{R}) - 2^\mathbb{Z} \)-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 \( B(\mathbb{R}) \). By the composition theorem (Theorem 1.80), for any measurable map \( f : (\Omega, A) \to (\mathbb{R}, B(\mathbb{R})) \) the maps \( \lfloor f \rfloor \) and \( \lceil f \rceil \) are also \( A - 2^\mathbb{Z} \)-measurable.
(iii) A map $X : \Omega \to \mathbb{R}^d$ is $\mathcal{A} - \mathcal{B}(\mathbb{R}^d)$-measurable if and only if
\[ X^{-1}((\infty, a]) \in \mathcal{A} \quad \text{for any} \quad a \in \mathbb{R}^d. \]
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 $\mathcal{E}_1, \ldots, \mathcal{E}_{12}$ from Theorem 1.23. \hfill \Box

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

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
\[ \varphi : [-1, 1] \to \overline{\mathbb{R}}, \quad x \mapsto \begin{cases} \tan(\pi x/2), & \text{if } x \in (-1, 1), \\ -\infty, & \text{if } x = -1, \\ +\infty, & \text{if } x = +1. \end{cases} \]
In fact, $\tilde{d}(x, y) = |\varphi^{-1}(x) - \varphi^{-1}(y)|$ for $x, y \in \overline{\mathbb{R}}$ defines a metric on $\overline{\mathbb{R}}$ such that $\varphi$ and $\varphi^{-1}$ are continuous. Hence $\varphi$ is a topological isomorphism. We denote by $\overline{\tau}$ the corresponding topology induced on $\overline{\mathbb{R}}$ and by $\tau$ the usual topology on $\mathbb{R}$.

Corollary 1.87. With the above notation, $\overline{\tau}|_{\mathbb{R}} = \tau$ and hence $\mathcal{B}(\overline{\mathbb{R}})|_{\mathbb{R}} = \mathcal{B}(\mathbb{R})$.

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 $(\Omega, \tau)$ and $(\Omega', \tau')$ 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 $A' \in \tau'$. \hfill \Box

For $x, y \in \overline{\mathbb{R}}$, we agree on the following notation.
\[
\begin{align*}
  x \lor y &= \max(x, y) & \text{(maximum)}, \\
  x \land y &= \min(x, y) & \text{(minimum)}, \\
  x^+ &= \max(x, 0) & \text{(positive part)}, \\
  x^- &= \max(-x, 0) & \text{(negative part)}, \\
  |x| &= \max(x, -x) = x^- + x^+ & \text{(modulus)}, \\
  \text{sign}(x) &= \mathbb{1}_{\{x > 0\}} - \mathbb{1}_{\{x < 0\}} & \text{(sign function)}. 
\end{align*}
\]
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 \( x \mapsto \text{sign}(x) \) also is measurable. Using Corollary 1.82, we thus get the following corollary.

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

\[ f \text{ is } \mathcal{A} - \mathcal{B}(\mathbb{R}^n)\text{-measurable} \iff \text{each } f_i \text{ is } \mathcal{A} - \mathcal{B}(\mathbb{R})\text{-measurable.} \]

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

**Proof.** For \( b \in \mathbb{R}^n \), we have \( f^{-1}((-\infty,b)) = \cap_{i=1}^n f_i^{-1}((-\infty,b_i)). \) 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 \( i \)th coordinate. Clearly, \( \pi_i \) 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 \( \frac{0}{0} := 0 \) 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 $(\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then the following maps are also measurable:

$$\inf_{n \in \mathbb{N}} X_n, \quad \sup_{n \in \mathbb{N}} X_n, \quad \lim_{n \rightarrow \infty} X_n, \quad \limsup_{n \rightarrow \infty} X_n.$$ 

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

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

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 $\liminf_{n \rightarrow \infty} X_n := \sup_{n \in \mathbb{N}} Y_n$ also is measurable. The proof for the limes superior is similar. \qed

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

Definition 1.93 (Simple function). Let $(\Omega, \mathcal{A})$ be a measurable space. A map $f : \Omega \rightarrow \mathbb{R}$ 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

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

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

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

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

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 \((\Omega, A)\) 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 \(A_1, A_2, \ldots \in A\) and numbers \(\alpha_1, \alpha_2, \ldots \geq 0\) 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 \(i = 1, \ldots, 2^n\). It is easy to check that \(\bigcup_{i=1}^{2^n} B_{n,i} = \Omega\). 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 factorisation lemma.

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

Proof. “\(\Leftarrow\)” If \(\varphi\) is measurable and \(g = \varphi \circ f\), then \(g\) is measurable by Theorem 1.80.

“\(\Rightarrow\)” Now assume that \(g\) is \(\sigma(f) - B(\mathbb{R})\)-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 \(B_n \in A'\) such that \(f^{-1}(B_n) = A_n\); that is, such that \(\mathbb{1}_{A_n} = \mathbb{1}_{B_n} \circ f\). Define \(\varphi : \Omega' \to \mathbb{R}\) by 
\[ \varphi = \sum_{n=1}^{\infty} \alpha_n \mathbb{1}_{B_n}. \]

Clearly, \(\varphi\) is \(A' - B(\mathbb{R})\)-measurable and \(g = \varphi \circ f\).
Now drop the assumption that \( g \) is nonnegative. Then there exist measurable maps \( \varphi^- \) and \( \varphi^+ \) 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 \(\mu\) be a measure on \((\Omega, \mathcal{A})\). Further, let \(X : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')\) be measurable. The **image measure** of \(\mu\) under the map \(X\) is the measure \(\mu \circ X^{-1}\) on \((\Omega', \mathcal{A}')\) that is defined by

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

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

\[
\mu \circ X^{-1}(\{x\}) = \sum_{y \in \mathbb{Z}} \mu(\{(x - y, y)\}). \quad \diamond
\]

**Example 1.100.** Let \(L : \mathbb{R}^n \to \mathbb{R}^n\) be a linear bijection and let \(\lambda\) 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 \(|\det(L^{-1})| \prod_{i=1}^n (b_i - a_i)\). \quad \diamond

As a generalisation 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 \(\mu\) be a measure on \(\mathbb{R}^n\) that has a continuous (or piecewise continuous) density \(f : \mathbb{R}^n \to [0, \infty)\). That is,

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

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 \(\varphi'\). 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. \quad \diamond
Exercise 1.4.2. Let \((\Omega, \mathcal{A}, \mu)\) be a measure space and let \(f : \Omega \to \mathbb{R}\) be measurable. Assume that \(g : \Omega \to \mathbb{R}\) fulfils \(g = f\) \(\mu\)-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 \(B(\mathbb{R}) - B(\mathbb{R})\)-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 \(\sigma\)-algebra generated by the cylinder sets
\[
\{[\omega_1, \ldots, \omega_n] : n \in \mathbb{N}, \omega_1, \ldots, \omega_n \in \{0, 1\}\}.
\]
Further, let \(\mu = (\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1)^\otimes \mathbb{N}\) be the Bernoulli measure on \(\Omega\) 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 \(n\)th coordinate map. Finally, let
\[
U(\omega) = \sum_{n=1}^{\infty} X_n(\omega) 2^{-n} \quad \text{for} \quad \omega \in \Omega.
\]
(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 \(f|_C\) 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 \(\lambda\) (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 \((\Omega, \mathcal{A}, \mathbb{P})\). The sets \(A \in \mathcal{A}\) are called events. In most cases, the events of \(\Omega\) are not observed directly. Rather, the observations are aspects of the single experiments that are coded as measurable maps from \(\Omega\) 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 $P$ under $X$. Hence we have to develop a calculus to determine the distributions of, e.g., sums of random variables.

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

(i) $X$ is called a random variable with values in $(\Omega', A')$. If $(\Omega', A') = (\mathbb{R}, B(\mathbb{R}))$, then $X$ is called a real random variable or simply a random variable.

(ii) For $A' \in A'$, we denote $\{X \in A'\} := X^{-1}(A')$ and $P[X \in A'] := P[X^{-1}(A')]$. 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 $P_X := P \circ X^{-1}$ is called the distribution of $X$.

(ii) For a real random variable $X$, the map $F_X : x \mapsto P[X \leq x]$ is called the distribution function of $X$ (or, more accurately, of $P_X$). We write $X \sim \mu$ if $\mu = P_X$ and say that $X$ has distribution $\mu$.

(iii) A family $(X_i)_{i \in I}$ of random variables is called identically distributed if $P_{X_i} = P_{X_j}$ for all $i, j \in I$. We write $X \overset{D}{=} Y$ if $P_X = P_Y$ (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 $(\Omega, A, P)$ and a random variable $X : \Omega \to \mathbb{R}$ such that $F_X = F$.

The simplest choice would be $(\Omega, A) = (\mathbb{R}, B(\mathbb{R}))$, $X : \mathbb{R} \to \mathbb{R}$ the identity map and $P$ 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 $\Omega := (0, 1)$, $A := B(\mathbb{R})|_\Omega$ and let $P$ be the Lebesgue measure on $(\Omega, A)$ (see Example 1.74). Define the left continuous inverse of $F$:

$$F^{-1}(t) := \inf\{x \in \mathbb{R} : F(x) \geq t\} \quad \text{for} \ t \in (0, 1).$$
Then
\[ F^{-1}(t) \leq x \iff t \leq F(x). \]
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
\[ \mathbb{P}[\{ t : F^{-1}(t) \leq x \}] = F(x). \]
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 \( \mathbb{P}[X = 1] = p, \mathbb{P}[X = 0] = 1 - p. \) Then \( \mathbb{P}_X =: \text{Ber}_p \) is called the **Bernoulli distribution** with parameter \( p \); formally
\[ \text{Ber}_p = (1 - p) \delta_0 + p \delta_1. \]
Its distribution function is
\[ F_X(x) = \begin{cases} 
0, & \text{if } x < 0, \\
1 - p, & \text{if } x \in [0, 1), \\
1, & \text{if } x \geq 1. 
\end{cases} \]
The distribution \( \mathbb{P}_Y \) of \( Y := 2X - 1 \) is sometimes called the **Rademacher distribution** with parameter \( p \); formally \( \text{Rad}_p = (1 - p) \delta_{-1} + p \delta_{+1} \). In particular, \( \text{Rad}_{1/2} \) 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
\[ \mathbb{P}[X = k] = \binom{n}{k} p^k (1 - p)^{n-k}. \]
Then \( \mathbb{P}_X =: b_{n,p} \) 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
\[ \mathbb{P}[X = n] = p (1 - p)^n \quad \text{for any } n \in \mathbb{N}_0. \]
Then \( \gamma_p := b_{1,p} := \mathbb{P}_X \) is called the **geometric distribution**\(^2\) with parameter \( p \); formally
\[ \gamma_p = \sum_{n=0}^{\infty} p (1 - p)^n \delta_n. \]

\(^2\) Warning: For some authors, the geometric distribution is shifted by one to the right; that is, it is a distribution on \( \mathbb{N} \).
Its distribution function is \( F(x) = 1 - (1 - p)^{|x+1|} \) 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 \( (1 - p)\delta_0 + p\delta_1 \) (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

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

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

\[
b_{r, p}^- := \sum_{k=0}^{\infty} \binom{-r}{k} (-1)^k p^r (1 - p)^k \delta_k
\]

the negative binomial distribution or Pascal distribution with parameters \( r \) and \( p \). (Here \( \binom{x}{k} = \frac{x(x-1)\cdots(x-k+1)}{k!} \) for \( x \in \mathbb{R} \) and \( k \in \mathbb{N} \) is the generalised 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 \( r \)th 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 \( \mathbf{P}_X = : \text{Po} \lambda \) is called the Poisson distribution with parameter \( \lambda \).

(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 \( B, W, n \in \mathbb{N} \):

\[
\text{Hy}_{B,W,n}(\{b\}) = \frac{\binom{B}{b} \binom{W}{n-b}}{\binom{B+W}{n}}, \quad b \in \{0, \ldots, n\}.
\]

(1.18)

This generalises easily to the situation of \( k \) colours and \( B_i \) balls of colour \( i = 1, \ldots, k \). As above, we get that the probability of drawing out of \( n \) balls exactly
$b_i$ balls of colour $i$ for each $i = 1, \ldots, k$ (with the restriction $b_1 + \ldots + b_k = n$ and $b_i \leq B_i$ for all $i$) is given by the \textbf{generalised hypergeometric distribution}

$$\text{Hyp}_{B_1, \ldots, B_k, n}(\{ (b_1, \ldots, b_k) \}) = \frac{(b_1) \ldots (b_k)}{(B_1 + \ldots + B_k)}.$$  \hspace{1cm} (1.19)

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

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

Then $\mathbb{P}_X =: \mathcal{N}_{\mu, \sigma^2}$ is called the Gaussian \textbf{normal distribution} with parameters $\mu$ and $\sigma^2$. 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

$$\mathbb{P}[X \leq x] = \mathbb{P}[X \in [0, x]] = \int_{0}^{x} \theta e^{-\theta t} \, dt \quad \text{for} \quad x \geq 0.$$  

Then $\mathbb{P}_X$ is called the \textbf{exponential distribution} with parameter $\theta$ (in shorthand, $\exp(\theta)$).

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

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

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

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

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

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

\textbf{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 $\Gamma$ denotes the gamma function.) Then $\Gamma_{\theta, r}$ is called the \textbf{Gamma distribution} with scale parameter $\theta$ and shape parameter $r$.\hfill\blacktriangle

(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 \textbf{Beta distribution} with parameters \( r \) and \( s \).

(iii) Let \( a > 0 \) and let \( \text{Cau}_a \) be the distribution on \( \mathbb{R} \) with density
\[
x \mapsto \frac{1}{a \pi} \frac{1}{1 + (x/a)^2}.
\]
Then \( \text{Cau}_a \) is called the \textbf{Cauchy distribution} with parameter \( a \).

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

\textbf{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.

\textbf{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 \( aX \sim \exp\theta/a \).
2

Independence

The measure theory from the preceding chapter is a linear theory that could not describe the dependence structure of events or random variables. We enter the realm of probability theory exactly at this point, where we define independence of events and random variables. Independence is a pivotal notion of probability theory, and the computation of dependencies is one of the theory’s major tasks.

In the sequel, \((\Omega, A, P)\) is a probability space and the sets \(A \in A\) are the events. As soon as constructing probability spaces has become routine, the concrete probability space will lose its importance and it will be only the random variables that will interest us. The bold font symbol \(P\) will then denote the universal object of a probability measure, and the probabilities \(P[\cdot]\) with respect to it will always be written in square brackets.

2.1 Independence of Events

We consider two events \(A\) and \(B\) as (stochastically) independent if the occurrence of \(A\) does not change the probability that \(B\) also occurs. Formally, we say that \(A\) and \(B\) are independent if

\[
P[A \cap B] = P[A] \cdot P[B].
\]

Example 2.1 (Rolling a die twice). Consider the random experiment of rolling a die twice. Hence \(\Omega = \{1, \ldots, 6\}^2\) endowed with the \(\sigma\)-algebra \(\mathcal{A} = 2^\Omega\) and the uniform distribution \(P = U_\Omega\) (see Example 1.30(ii)).

(i) Two events \(A\) and \(B\) should be independent, e.g., if \(A\) depends only on the outcome of the first roll and \(B\) depends only on the outcome of the second roll. Formally, we assume that there are sets \(\tilde{A}, \tilde{B} \subset \{1, \ldots, 6\}\) such that

\[
A = \tilde{A} \times \{1, \ldots, 6\} \quad \text{and} \quad B = \{1, \ldots, 6\} \times \tilde{B}.
\]
Now we check that $A$ and $B$ indeed fulfil (2.1). To this end, we compute $P[A] = \frac{\#A}{36} = \frac{\#\tilde{A}}{6}$ and $P[B] = \frac{\#B}{36} = \frac{\#\tilde{B}}{6}$. Furthermore,

$$P[A \cap B] = \frac{\#(\tilde{A} \times \tilde{B})}{36} = \frac{\#\tilde{A}}{6} \cdot \frac{\#\tilde{B}}{6} = P[A] \cdot P[B].$$

(ii) Stochastic independence can occur also in less obvious situations. For instance, let $A$ be the event where the sum of the two rolls is odd,

$$A = \{(\omega_1, \omega_2) \in \Omega : \omega_1 + \omega_2 \in \{3, 5, 7, 9, 11\}\},$$

and let $B$ be the event where the first roll gives at most a three

$$B = \{(\omega_1, \omega_2) \in \Omega : \omega_1 \in \{1, 2, 3\}\}.$$ 

Although it might seem that these two events are entangled in some way, they are stochastically independent. Indeed, it is easy to check that $P[A] = P[B] = \frac{1}{2}$ and $P[A \cap B] = \frac{1}{4}$. 

What is the condition for three events $A_1, A_2, A_3$ to be independent? Of course, any of the pairs $(A_1, A_2)$, $(A_1, A_3)$ and $(A_2, A_3)$ has to be independent. However, we have to make sure also that the simultaneous occurrence of $A_1$ and $A_2$ does not change the probability that $A_3$ occurs. Hence, it is not enough to consider pairs only.

Formally, we call three events $A_1, A_2$ and $A_3$ (stochastically) independent if

$$P[A_i \cap A_j] = P[A_i] \cdot P[A_j] \quad \text{for all } i, j \in \{1, 2, 3\}, \; i \neq j,$$

and

$$P[A_1 \cap A_2 \cap A_3] = P[A_1] \cdot P[A_2] \cdot P[A_3].$$

Note that (2.2) does not imply (2.3) (and (2.3) does not imply (2.2)).

**Example 2.2 (Rolling a die three times).** We roll a die three times. Hence $\Omega = \{1, \ldots, 6\}^3$ endowed with the discrete $\sigma$-algebra $\mathcal{A} = 2^\Omega$ and the uniform distribution $P = U_\Omega$ (see Example 1.30(ii)).

(i) If we assume that for any $i = 1, 2, 3$ the event $A_i$ depends only on the outcome of the $i$th roll, then the events $A_1$, $A_2$ and $A_3$ are independent. Indeed, as in the preceding example, there are sets $\tilde{A}_1, \tilde{A}_2, \tilde{A}_3 \subset \{1, \ldots, 6\}$ such that

$$A_1 = \tilde{A}_1 \times \{1, \ldots, 6\}^2,$$
$$A_2 = \{1, \ldots, 6\} \times \tilde{A}_2 \times \{1, \ldots, 6\},$$
$$A_3 = \{1, \ldots, 6\}^2 \times \tilde{A}_3.$$

The validity of (2.2) follows as in Example 2.1(i). In order to show (2.3), we compute

$$P[A_1 \cap A_2 \cap A_3] = \frac{\#(\tilde{A}_1 \times \tilde{A}_2 \times \tilde{A}_3)}{216} = \prod_{i=1}^{3} \frac{\#\tilde{A}_i}{6} = \prod_{i=1}^{3} P[A_i].$$
Consider now the events
\[
A_1 := \{ \omega \in \Omega : \omega_1 = \omega_2 \}, \\
A_2 := \{ \omega \in \Omega : \omega_2 = \omega_3 \}, \\
A_3 := \{ \omega \in \Omega : \omega_1 = \omega_3 \}.
\]
Then \(#A_1 = #A_2 = #A_3 = 36\); hence \(P[A_1] = P[A_2] = P[A_3] = \frac{1}{6}\). Furthermore, \(#(A_i \cap A_j) = 6\) if \(i \neq j\); hence \(P[A_i \cap A_j] = \frac{1}{36}\). Hence (2.2) holds. On the other hand, we have \(#(A_1 \cap A_2 \cap A_3) = 6\), thus \(P[A_1 \cap A_2 \cap A_3] = \frac{1}{36} \neq \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6}\). Thus (2.3) does not hold and so the events \(A_1, A_2, A_3\) are not independent.

In order to define independence of larger families of events, we have to request the validity of product formulas, such as (2.2) and (2.3), not only for pairs and triples but for all finite subfamilies of events. We thus make the following definition.

**Definition 2.3 (Independence of events).** Let \(I\) be an arbitrary index set and let \((A_i)_{i \in I}\) be an arbitrary family of events. The family \((A_i)_{i \in I}\) is called **independent** if for any finite subset \(J \subset I\) the product formula holds:
\[
P\left[ \bigcap_{j \in J} A_j \right] = \prod_{j \in J} P[A_j].
\]

The most prominent example of an independent family of infinitely many events is given by the perpetuated independent repetition of a random experiment.

**Example 2.4.** Let \(E\) be a finite set (the set of possible outcomes of the individual experiment) and let \((p_e)_{e \in E}\) be a probability vector on \(E\). Equip (as in Theorem 1.64) the probability space \(\Omega = E^N\) with the \(\sigma\)-algebra \(\mathcal{A} = \sigma(\{[\omega_1, \ldots, \omega_n] : \omega_1, \ldots, \omega_n \in E, n \in \mathbb{N}\})\) and with the product measure (or Bernoulli measure)
\[
P = (\sum_{e \in E} p_e \delta_e)^\otimes N; \text{ that is where } \P[[\omega_1, \ldots, \omega_n]] = \prod_{i=1}^{n} p_{\omega_i}.
\]
Let \(A_i \subset E\) for any \(i \in \mathbb{N}\), and let \(A_i\) be the event where \(\tilde{A}_i\) occurs in the \(i\)th experiment; that is,
\[
A_i = \{ \omega \in \Omega : \omega_i \in \tilde{A}_i \} = \bigcup_{(\omega_1, \ldots, \omega_i) \in E^{i-1} \times \tilde{A}_i} [\omega_1, \ldots, \omega_i].
\]
Intuitively, the family \((A_i)_{i \in \mathbb{N}}\) should be independent if the definition of independence makes any sense at all. We check that this is indeed the case. Let \(J \subset \mathbb{N}\) be finite with \(k := \#J\) and \(n := \max J\). Formally, we define \(B_j = A_j\) and \(\tilde{B}_j = \tilde{A}_j\) for \(j \in J\) and \(B_j = \Omega\) and \(\tilde{B}_j = E\) for \(j \in \{1, \ldots, n\} \setminus J\). Then
\[
P \left[ \bigcap_{j \in J} A_j \right] = P \left[ \bigcap_{j \in J} B_j \right] = P \left[ \bigcap_{j=1}^{n} B_j \right] = \sum_{e_{1} \in \tilde{B}_1} \cdots \sum_{e_{n} \in \tilde{B}_n} \prod_{j=1}^{n} p_{e_{j}} = \prod_{j=1}^{n} \left( \sum_{e \in \tilde{B}_j} p_{e} \right) = \prod_{j \in J} \left( \sum_{e \in \tilde{A}_j} p_{e} \right).
\]
This is true in particular for \( \# J = 1 \). Hence \( P[A_i] = \sum_{e \in A_i} p_e \) for all \( i \in \mathbb{N} \), whence
\[
P[\bigcap_{j \in J} A_j] = \prod_{j \in J} P[A_j]. \tag{2.4}
\]
Since this holds for all finite \( J \subset \mathbb{N} \), the family \( (A_i)_{i \in \mathbb{N}} \) is independent.

If \( A \) and \( B \) are independent, then \( A^c \) and \( B \) also are independent since
\[
\]
We generalise this observation in the following theorem.

**Theorem 2.5.** Let \( I \) be an arbitrary index set and let \( (A_i)_{i \in I} \) be a family of events. Define \( B^0_i = A_i \) and \( B^1_i = A^c_i \) for \( i \in I \). Then the following three statements are equivalent.

(i) The family \( (A_i)_{i \in I} \) is independent.

(ii) There is an \( \alpha \in \{0, 1\}^I \) such that the family \( (B^\alpha_i)_{i \in I} \) is independent.

(iii) For any \( \alpha \in \{0, 1\}^I \), the family \( (B^\alpha_i)_{i \in I} \) is independent.

**Proof.** This is left as an exercise. \( \square \)

**Example 2.6 (Euler’s prime number formula).** The **Riemann zeta function** is defined by the Dirichlet series
\[
\zeta(s) := \sum_{n=1}^{\infty} n^{-s} \quad \text{for } s \in (1, \infty).
\]
Euler’s prime number formula is a representation of the Riemann zeta function as an infinite product
\[
\zeta(s) = \prod_{p \in \mathcal{P}} (1 - p^{-s})^{-1}, \tag{2.5}
\]
where \( \mathcal{P} := \{p \in \mathbb{N} : p \text{ is prime}\} \).

We give a probabilistic proof for this formula. Let \( \Omega = \mathbb{N} \), and for fixed \( s > 1 \) define \( P \) on \( 2^\Omega \) by
\[
P[\{n\}] = \zeta(s)^{-1} n^{-s} \quad \text{for } n \in \mathbb{N}.
\]
Let \( p\mathbb{N} = \{pn : n \in \mathbb{N}\} \) and \( \mathcal{P}_n = \{p \in \mathcal{P} : p \leq n\} \). We consider \( p\mathbb{N} \subset \Omega \) as an event. Note that \( P[p\mathbb{N}] = p^{-s} \) and that \( (p\mathbb{N}, p \in \mathcal{P}) \) is independent. Indeed, for \( k \in \mathbb{N} \) and mutually distinct \( p_1, \ldots, p_k \in \mathcal{P} \), we have
\[
\bigcap_{i=1}^k (p_i \mathbb{N}) = (p_1 \cdots p_k) \mathbb{N}.
\]
2.1 Independence of Events

Thus

\[ P\left[ \bigcap_{i=1}^{k} (p_i \mathbb{N}) \right] = \sum_{n=1}^{\infty} P\left[ \{p_1 \cdots p_k n\} \right] \]

\[ = \zeta(s)^{-1} (p_1 \cdots p_k)^{-s} \sum_{n=1}^{\infty} n^{-s} \]

\[ = (p_1 \cdots p_k)^{-s} = \prod_{i=1}^{k} P[p_i \mathbb{N}]. \]

By Theorem 2.5, the family \(((p \mathbb{N})^c, p \in \mathcal{P})\) is also independent, whence

\[ \zeta(s)^{-1} = P[\{1\}] = P\left[ \bigcap_{p \in \mathcal{P}^c} (p \mathbb{N})^c \right] \]

\[ = \lim_{n \to \infty} P\left[ \bigcap_{p \in \mathcal{P}^c} (p \mathbb{N})^c \right] \]

\[ = \lim_{n \to \infty} \prod_{p \in \mathcal{P}^c} (1 - P[p \mathbb{N}]) = \prod_{p \in \mathcal{P}} (1 - p^{-s}). \]

This shows (2.5).

If we roll a die infinitely often, what is the chance that the face shows a six infinitely often? This probability should equal one. Otherwise there would be a last point in time when we see a six and after which the face only shows a number one to five. However, this is not very plausible.

Recall that we formalised the event where infinitely many of a series of events occur by means of the limes superior (see Definition 1.13). The following theorem confirms the conjecture mentioned above and also gives conditions under which we cannot expect that infinitely many of the events occur.

**Theorem 2.7 (Borel-Cantelli lemma).** Let \(A_1, A_2, \ldots\) be events and define \(A^* = \limsup_{n \to \infty} A_n\).

(i) If \(\sum_{n=1}^{\infty} P[A_n] < \infty\), then \(P[A^*] = 0\). (Here \(P\) could be an arbitrary measure on \((\Omega, \mathcal{A})\).)

(ii) If \((A_n)_{n \in \mathbb{N}}\) is independent and \(\sum_{n=1}^{\infty} P[A_n] = \infty\), then \(P[A^*] = 1\).

**Proof.** (i) \(P\) is upper semicontinuous and \(\sigma\)-subadditive; hence, by assumption,

\[ P[A^*] = \lim_{n \to \infty} P\left[ \bigcup_{m=n}^{\infty} A_m \right] \leq \lim_{n \to \infty} \sum_{m=n}^{\infty} P[A_m] = 0. \]
De Morgan’s rule and the lower semicontinuity of $P$ yield
\[ P[(A^*)^c] = P \left[ \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n^c \right] = \lim_{m \to \infty} P \left[ \bigcap_{n=m}^{\infty} A_n^c \right]. \]

However, for every $m \in \mathbb{N}$ (since $\log(1 - x) \leq -x$ for $x \in [0, 1]$),
\[ P \left[ \bigcap_{n=m}^{\infty} A_n^c \right] = \prod_{n=m}^{\infty} (1 - P[A_n]) \]
\[ = \exp \left( \sum_{n=m}^{\infty} \log (1 - P[A_n]) \right) \leq \exp \left( - \sum_{n=m}^{\infty} P[A_n] \right) = 0. \]

**Example 2.8.** We throw a die again and again and ask for the probability of seeing a six infinitely often. Hence $\Omega = \{1, \ldots, 6\}^\mathbb{N}$, $A = (2^{\{1, \ldots, 6\}})^\otimes \mathbb{N}$ is the product $\sigma$-algebra and $P = (\sum_{e \in \{1, \ldots, 6\}} \frac{1}{6} \delta_e \otimes \mathbb{N})$ is the Bernoulli measure (see Theorem 1.64).

Furthermore, let $A_n = \{\omega \in \Omega : \omega_n = 6\}$ be the event where the $n$th roll shows a six. Then $A^* = \limsup A_n$ is the event where we see a six infinitely often (see Example 1.14). Furthermore, $(A_n)_{n \in \mathbb{N}}$ is an independent family with the property
\[ \sum_{n=1}^{\infty} P[A_n] = \sum_{n=1}^{\infty} \frac{1}{6} = \infty. \]
Hence the Borel-Cantelli lemma yields $P[A^*] = 1$.

**Example 2.9.** We roll a die only once and define $A_n$ for any $n \in \mathbb{N}$ as the event where in this one roll the face showed a six. Note that $A_1 = A_2 = A_3 = \ldots$. Then $A^* = \limsup A_n$ is the event where we see a six infinitely often (see Example 1.14). Furthermore, $(A_n)_{n \in \mathbb{N}}$ is an independent family with the property
\[ \sum_{n=1}^{\infty} P[A_n] = \sum_{n=1}^{\infty} \frac{1}{6} = \infty. \]
Hence the Borel-Cantelli lemma, the assumption of independence is indispensable.

**Example 2.10.** Let $\Lambda \in (0, \infty)$ and $0 \leq \lambda_n \leq \Lambda$ for $n \in \mathbb{N}$. Let $X_n$, $n \in \mathbb{N}$, be Poisson random variables with parameters $\lambda_n$. Then
\[ P[X_n \geq n \text{ for infinitely many } n] = 0. \]

Indeed,
\[ \sum_{n=1}^{\infty} P[X_n \geq n] = \sum_{n=1}^{\infty} \sum_{m=n}^{\infty} P[X_n = m] = \sum_{m=1}^{\infty} \sum_{n=1}^{m} P[X_n = m] \]
\[ = \sum_{m=1}^{\infty} \sum_{n=1}^{m} e^{-\lambda_n} \frac{\lambda_n^m}{m!} \leq \sum_{m=1}^{\infty} m \frac{\Lambda^m}{m!} = \Lambda e^\Lambda < \infty. \]

Note that in Theorem 2.7 in the case of independent events, only the probabilities $P[A^*] = 0$ and $P[A^*] = 1$ could show up. Thus the Borel-Cantelli lemma belongs to the class of so-called 0-1 laws. Later we will encounter more 0-1 laws (see, for example, Theorem 2.37).
Now we extend the notion of independence from families of events to families of classes of events.

**Definition 2.11 (Independence of classes of events).** Let $I$ be an arbitrary index set and let $E_i \subset A$ for all $i \in I$. The family $(E_i)_{i \in I}$ is called independent if, for any finite subset $J \subset I$ and any choice of $E_j \in E_j$, $j \in J$, we have

$$P\left[ \bigcap_{j \in J} E_j \right] = \prod_{j \in J} P[E_j].$$  (2.6)

**Example 2.12.** As in Example 2.4, let $(\Omega, A, P)$ be the product space of infinitely many repetitions of a random experiment whose possible outcomes $e$ are the elements of the finite set $E$ and have probabilities $p = (p_e)_{e \in E}$. For $i \in \mathbb{N}$, define

$$E_i = \{ \{ \omega \in \Omega : \omega_i \in A \} : A \subset E \}.$$

For any choice of sets $A_i \in E_i$, $i \in \mathbb{N}$, the family $(A_i)_{i \in \mathbb{N}}$ is independent; hence $(E_i)_{i \in \mathbb{N}}$ is independent. \hfill \Diamond

**Theorem 2.13.** (i) Let $I$ be finite, and for any $i \in I$ let $E_i \subset A$ with $\Omega \in E_i$. Then

$$(E_i)_{i \in I} \text{ is independent } \iff (2.6) \text{ holds for } J = I.$$  

(ii) $(E_i)_{i \in I}$ is independent $\iff (E_j)_{j \in J}$ is independent for all finite $J \subset I$.

(iii) If $(E_i \cup \{\emptyset\})$ is $\cap$-stable, then

$$(E_i)_{i \in I} \text{ is independent } \iff (\sigma(E_i))_{i \in I} \text{ is independent.}$$

(iv) Let $K$ be an arbitrary set and let $(I_k)_{k \in K}$ be mutually disjoint subsets of $I$. If $(E_i)_{i \in I}$ is independent, then $(\bigcup_{i \in I_k} E_i)_{k \in K}$ also is independent.

**Proof.** (i) “$\implies$” This is trivial.

(i) “$\iff$” For $J \subset I$ and $j \in I \setminus J$, choose $E_j = \Omega$.

(ii) This is trivial.

(iii) “$\iff$” This is trivial.

(iii) “$\implies$” Let $J \subset I$ be finite. We will show that for any two finite sets $J$ and $J'$ with $J \subset J' \subset I$,

$$P\left[ \bigcap_{i \in J'} E_i \right] = \prod_{i \in J'} P[E_i] \text{ for any choice } \begin{cases} E_i \in \sigma(E_i), & \text{if } i \in J, \\ E_i \in E_i, & \text{if } i \in J' \setminus J. \end{cases}$$  (2.7)
The case \( J' = J \) is exactly the claim we have to show.

We carry out the proof of (2.7) by induction on \( \#J \). For \( \#J = 0 \), the statement (2.7) holds by assumption of this theorem.

Now assume that (2.7) holds for every \( J \) with \( \#J = n \) and for every finite \( J' \supset J \). Fix such a \( J \) and let \( j \in I \setminus J \). Choose \( J' \supset J := J \cup \{j\} \). We show the validity of (2.7) with \( J \) replaced by \( J' \).

Let \( E_i \in \sigma(\mathcal{E}_i) \) for any \( i \in J \), and let \( E_i \in \mathcal{E}_i \) for any \( i \in J' \setminus (J \cup \{j\}) \). Define two measures \( \mu \) and \( \nu \) on \((\Omega, \mathcal{A})\) by

\[
\mu : E_j \mapsto p \left[ \bigcap_{i \in J'} E_i \right] \quad \text{and} \quad \nu : E_j \mapsto \prod_{i \in J'} p[E_i].
\]

By the induction hypothesis (2.7), we have \( \mu(E_j) = \nu(E_j) \) for every \( E_j \in \mathcal{E}_j \cup \{\emptyset, \Omega\} \). Since \( \mathcal{E}_j \cup \{\emptyset\} \) is a \( \pi \)-system, Lemma 1.42 yields that \( \mu(E_j) = \nu(E_j) \) for all \( E_j \in \sigma(\mathcal{E}_j) \). That is, (2.7) holds with \( J \) replaced by \( J \cup \{j\} \).

(iv) This is trivial, as (2.6) has to be checked only for \( J \subset I \) with \( \#(J \cap I_k) \leq 1 \) for any \( k \in K \). \( \square \)

2.2 Independent Random Variables

Now that we have studied independence of events, we want to study independence of random variables. Here also the definition ends up with a product formula. Formally, however, we can also define independence of random variables via independence of the \( \sigma \)-algebras they generate. This is the reason why we studied independence of classes of events in the last section.

Independent random variables allow for a rich calculus. For example, we can compute the distribution of a sum of two independent random variables by a simple convolution formula. Since we do not have a general notion of an integral at hand at this point, for the time being we restrict ourselves to presenting the convolution formula for integer-valued random variables only.

Let \( I \) be an arbitrary index set. For each \( i \in I \), let \((\Omega_i, \mathcal{A}_i)\) be a measurable space and let \( X_i : (\Omega, \mathcal{A}) \rightarrow (\Omega_i, \mathcal{A}_i) \) be a random variable with generated \( \sigma \)-algebra \( \sigma(X_i) = X_i^{-1}(\mathcal{A}_i) \).

**Definition 2.14 (Independent random variables).** The family \((X_i)_{i \in I}\) of random variables is called independent if the family \((\sigma(X_i))_{i \in I}\) of \( \sigma \)-algebras is independent.

As a shorthand, we say that a family \((X_i)_{i \in I}\) is “i.i.d.” (for “independent and identically distributed”) if \((X_i)_{i \in I}\) is independent and if \( P_{X_i} = P_{X_j} \) for all \( i, j \in I \).
Remark 2.15. (i) Clearly, the family \((X_i)_{i \in I}\) is independent if and only if, for any finite set \(J \subset I\) and any choice of \(A_j \in \mathcal{A}_j, j \in J\), we have
\[
P\left[\bigcap_{j \in J} A_j \right] = \prod_{j \in J} P[A_j].
\]
The next theorem will show that it is enough to request the validity of such a product formula for \(A_j\) from an \(\cap\)-stable generator of \(\mathcal{A}_j\) only.

(ii) If \((\bar{A}_i)_{i \in I}\) is an independent family of \(\sigma\)-algebras and if each \(X_i\) is \(\bar{A}_i - \mathcal{A}_i\)-measurable, then \((X_i)_{i \in I}\) is independent. This is a direct consequence of the fact that \(\sigma(X_i) \subset \bar{A}_i\).

(iii) For each \(i \in I\), let \((\Omega'_i, \mathcal{A}'_i)\) be another measurable space and assume that \(f_i : (\Omega_i, \mathcal{A}_i) \to (\Omega'_i, \mathcal{A}'_i)\) is a measurable map. If \((X_i)_{i \in I}\) is independent, then \((f_i \circ X_i)_{i \in I}\) is independent. This statement is a special case of (i) since \(f_i \circ X_i\) is \(\sigma(X_i) - \mathcal{A}'_i\)-measurable (see Theorem 1.80).

\[\text{Theorem 2.16 (Independent generators).} \quad \text{For any } i \in I, \text{ let } \mathcal{E}_i \subset \mathcal{A}_i \text{ be a } \pi\text{-system that generates } \mathcal{A}_i. \text{ If } (X_i^{-1}(\mathcal{E}_i))_{i \in I} \text{ is independent, then } (X_i)_{i \in I} \text{ is independent.}\]

\[\text{Proof.} \quad \text{By Theorem 1.81(iii), } X_i^{-1}(\mathcal{E}_i) \text{ is a } \pi\text{-system that generates the } \sigma\text{-algebra} \]
\[X_i^{-1}(\mathcal{A}_i) = \sigma(X_i). \quad \text{Hence the statement follows from Theorem 2.13.} \]

Example 2.17. Let \(E\) be a countable set and let \((X_i)_{i \in I}\) be random variables with values in \((E, 2^E)\). In this case, \((X_i)_{i \in I}\) is independent if and only if, for any finite \(J \subset I\) and any choice of \(x_j \in E, j \in J,\)
\[
P\left[X_j = x_j \text{ for all } j \in J\right] = \prod_{j \in J} P[X_j = x_j].
\]
This is obvious since \(\{\{x\} : x \in E\} \cup \{\emptyset\}\) is a \(\pi\)-system that generates \(2^E\), thus \(X_i^{-1}(\{\{x_i\}\}) : x_i \in E\} \cup \{\emptyset\}\) is a \(\pi\)-system that generates \(\sigma(X_i)\) (Theorem 1.81).

Example 2.18. Let \(E\) be a finite set and let \(p = (p_e)_{e \in E}\) be a probability vector. Repeat a random experiment with possible outcomes \(e \in E\) and probabilities \(p_e\) for \(e \in E\) infinitely often (see Example 1.40 and Theorem 1.64). Let \(\Omega = E^\mathbb{N}\) be the infinite product space and let \(\mathcal{A}\) be the \(\sigma\)-algebra generated by the cylinder sets (see (1.8)). Let \(P = (\sum_{e \in E} p_e \delta_e)^{\otimes \mathbb{N}}\) be the Bernoulli measure. Further, for any \(n \in \mathbb{N}\), let \(X_n : \Omega \to E, (\omega_m)_{m \in \mathbb{N}} \mapsto \omega_n,\)
be the projection on the \( n \)th coordinate. In other words: For any simple event \( \omega \in \Omega \), 
\( X_n(\omega) \) yields the result of the \( n \)th experiment. Then, by (2.4) (in Example 2.4), for 
\( n \in \mathbb{N} \) and \( x \in E^n \), we have
\[
\mathbb{P}[X_j = x_j \text{ for all } j = 1, \ldots, n] = \mathbb{P}[x_1, \ldots, x_n] = \mathbb{P}\left[\bigcap_{j=1}^{n} X_j^{-1}(\{x_j\})\right] 
= \prod_{j=1}^{n} \mathbb{P}[X_j^{-1}(\{x_j\})] = \prod_{j=1}^{n} \mathbb{P}[X_j = x_j],
\]
and \( \mathbb{P}[X_j = x_j] = p_{x_j} \). By virtue of Theorem 2.13(i), this implies that the family 
\( (X_1, \ldots, X_n) \) is independent and hence, by Theorem 2.13(ii), \( (X_n)_{n \in \mathbb{N}} \) is independent 
as well.

In particular, we have shown the following theorem.

**Theorem 2.19.** Let \( E \) be a finite set and let \( (p_e)_{e \in E} \) be a probability vector on \( E \). 
Then there exists a probability space \( (\Omega, \mathcal{A}, \mathbb{P}) \) and an independent family \( (X_n)_{n \in \mathbb{N}} \) 
of \( E \)-valued random variables on \( (\Omega, \mathcal{A}, \mathbb{P}) \) such that \( \mathbb{P}[X_n = e] = p_e \) for any 
\( e \in E \).

Later we will see that the assumption that \( E \) is finite can be dropped. Also one can allow for different distributions in the respective factors. For the time being, however, this theorem gives us enough examples of interesting families of independent random variables.

Our next goal is to deduce simple criteria in terms of distribution functions and densities for checking whether a family of random variables is independent or not.

**Definition 2.20.** For any \( i \in I \), let \( X_i \) be a real random variable. For any finite 
subset \( J \subset I \), let 
\[
F_J := F_{(X_j)_{j \in J}} : \mathbb{R}^J \to [0, 1],
\]
\[
x \mapsto \mathbb{P}[X_j \leq x_j \text{ for all } j \in J] = \mathbb{P}\left[\bigcap_{j \in J} X_j^{-1}((\infty, x_j])\right].
\]

Then \( F_J \) is called the joint distribution function of \( (X_j)_{j \in J} \). The probability measure 
\( \mathbb{P}_{(X_j)_{j \in J}} \) on \( \mathbb{R}^J \) is called the joint distribution of \( (X_j)_{j \in J} \).

**Theorem 2.21.** A family \( (X_i)_{i \in I} \) of real random variables is independent if and only if, for every finite \( J \subset I \) and every \( x = (x_j)_{j \in J} \in \mathbb{R}^J \),
\[
F_J(x) = \prod_{j \in J} F_{\{j\}}(x_j). \tag{2.8} \]
Proof. The class of sets \( \{(-\infty, b], b \in \mathbb{R}\} \) is an \( \cap \)-stable generator of the Borel \( \sigma \)-algebra \( \mathcal{B}(\mathbb{R}) \) (see Theorem 1.23). Equation (2.8) says that, for any choice of real numbers \((x_i)_{i \in I}\), the events \((X^{-1}(((-\infty, x_i]))_{i \in I}\) are independent. Hence Theorem 2.16 yields the claim. \(\square\)

**Corollary 2.22.** In addition to the assumptions of Theorem 2.21, we assume that any \(F_j\) has a continuous density \(f_j = f(x_j)_{j \in J}\). That is, there exists a continuous map \(f_j : \mathbb{R}^j \to [0, \infty)\) such that
\[
F_j(x) = \int_{-\infty}^{x_{j_1}} dt_1 \cdots \int_{-\infty}^{x_{j_n}} dt_n f_j(t_1, \ldots, t_n) \quad \text{for all } x \in \mathbb{R}^d
\]
(where \(J = \{j_1, \ldots, j_n\}\)). In this case, the family \((X_i)_{i \in I}\) is independent if and only if, for any finite \(J \subset I\)
\[
f_j(x) = \prod_{j \in J} f_j(x_j) \quad \text{for all } x \in \mathbb{R}^d. \tag{2.9}
\]

**Corollary 2.23.** Let \(n \in \mathbb{N}\) and let \(\mu_1, \ldots, \mu_n\) be probability measures on \((\mathbb{R}, \mathcal{B}(\mathbb{R}))\). Then there exists a probability space \((\Omega, \mathcal{A}, \mathbf{P})\) and an independent family of random variables \((X_i)_{i=1,\ldots,n}\) on \((\Omega, \mathcal{A}, \mathbf{P})\) with \(\mathbf{P}_{X_i} = \mu_i\) for each \(i = 1, \ldots, n\).

**Proof.** Let \(\Omega = \mathbb{R}^n\) and \(\mathcal{A} = \mathcal{B}(\mathbb{R}^n)\). Let \(\mathbf{P} = \bigotimes_{i=1}^n \mu_i\) be the product measure of the \(\mu_i\) (see Theorem 1.61). Further, let \(X_i : \mathbb{R}^n \to \mathbb{R}, (x_1, \ldots, x_n) \mapsto x_i\) be the projection on the \(i\)th coordinate for each \(i = 1, \ldots, n\). Then, for any \(i = 1, \ldots, n\),
\[
F_{\{i\}}(x) = \mathbf{P}[X_i \leq x] = \mathbf{P}[\mathbb{R}^{i-1} \times (-\infty, x] \times \mathbb{R}^{n-i}]
\]
\[
= \mu_i((-\infty, x]) \cdot \prod_{j \neq i} \mu_j(\mathbb{R}) = \mu_i((-\infty, x]).
\]

Hence indeed \(\mathbf{P}_{X_i} = \mu_i\). Furthermore, for all \(x_1, \ldots, x_n \in \mathbb{R}\),
\[
F_{\{1,\ldots,n\}}((x_1, \ldots, x_n)) = \mathbf{P}\left[ \bigotimes_{i=1}^n (-\infty, x_i] \right] = \prod_{i=1}^n \mu_i((-\infty, x_i]) = \prod_{i=1}^n F_{\{i\}}(x_i).
\]

Hence Theorem 2.21 (and Theorem 2.13(i)) yields the independence of \((X_i)_{i=1,\ldots,n}\). \(\square\)

**Example 2.24.** Let \(X_1, \ldots, X_n\) be independent exponentially distributed random variables with parameters \(\theta_1, \ldots, \theta_n \in (0, \infty)\). Then
\[
F_{\{i\}}(x) = \int_0^x \theta_i e^{-\theta_i t} dt = 1 - e^{-\theta_i x} \quad \text{for } x \geq 0
\]
and hence
\[
F_{\{1,\ldots,n\}}((x_1, \ldots, x_n)) = \prod_{i=1}^n (1 - e^{-\theta_i x_i}).
\]
Consider now the random variable $Y = \max(X_1, \ldots, X_n)$. Then
\[
F_Y(x) = \mathbf{P}[X_i \leq x \text{ for all } i = 1, \ldots, n]
= F_{\{1, \ldots, n\}}((x, \ldots, x)) = \prod_{i=1}^{n} (1 - e^{-\theta_i x}).
\]
The distribution function of the random variable $Z := \min(X_1, \ldots, X_n)$ has a nice closed form:
\[
F_Z(x) = 1 - \mathbf{P}[Z > x]
= 1 - \mathbf{P}[X_i > x \text{ for all } i = 1, \ldots, n]
= 1 - \prod_{i=1}^{n} e^{-\theta_i x} = 1 - \exp\left(- (\theta_1 + \ldots + \theta_n) x\right).
\]
In other words, $Z$ is exponentially distributed with parameter $\theta_1 + \ldots + \theta_n$.

**Example 2.25.** Let $\mu_i \in \mathbb{R}$ and $\sigma_i^2 > 0$ for $i \in I$. Let $(X_i)_{i \in I}$ be real random variables with joint density functions (for finite $J \subset I$)
\[
f_J(x) = \prod_{j \in J} \left(2\pi \sigma_j^2\right)^{-\frac{1}{2}} \exp\left(-\sum_{j \in J} \frac{(x_j - \mu_j)^2}{2\sigma_j^2}\right) \text{ for } x \in \mathbb{R}^J.
\]
Then $(X_i)_{i \in I}$ is independent and $X_i$ is normally distributed with parameters $(\mu_i, \sigma_i^2)$.

For any finite $I = \{i_1, \ldots, i_n\}$ (with mutually distinct $i_1, \ldots, i_n$), the vector $Y = (X_{i_1}, \ldots, X_{i_n})$ has the $n$-dimensional normal distribution with $\mu = (\mu_{i_1}, \ldots, \mu_{i_n})$ and with $\Sigma = \Sigma^T$ the diagonal matrix with entries $\sigma_1^2, \ldots, \sigma_n^2$ (see Example 1.105(ix)).

**Theorem 2.26.** Let $K$ be an arbitrary set and $I_k, k \in K$, arbitrary mutually disjoint index sets. Define $I = \bigcup_{k \in K} I_k$.

If the family $(X_i)_{i \in I}$ is independent, then the family of $\sigma$-algebras $\{\sigma(X_j, j \in I_k)\}_{k \in K}$ is independent.

**Proof.** For $k \in K$, let
\[
Z_k = \left\{ \bigcap_{j \in I_k} A_j : A_j \in \sigma(X_j), \# \{j \in I_k : A_j \neq \Omega\} < \infty \right\}
\]
be the ring of finite-dimensional cylinder sets. Clearly, $Z_k$ is a $\pi$-system and $\sigma(Z_k) = \sigma(X_j, j \in I_k)$. Hence, by Theorem 2.13(iii), it is enough to show that $(Z_k)_{k \in K}$ is independent. By Theorem 2.13(ii), we can even assume that $K$ is finite.
For $k \in K$, let $B_k \in \mathcal{Z}_k$ and $J_k \subset I_k$ be finite with $B_k = \bigcap_{j \in J_k} A_j$ for certain $A_j \in \sigma(X_j)$. Define $J = \bigcup_{k \in K} J_k$. Then

$$
P\left[\bigcap_{k \in K} B_k\right] = \prod_{j \in J} P[A_j] = \prod_{k \in K} \prod_{j \in J_k} P[A_j] = \prod_{k \in K} P[B_k].$$

\[\square\]

Example 2.27. If $(X_n)_{n \in \mathbb{N}}$ is an independent family of real random variables, then also $(Y_n)_{n \in \mathbb{N}} = (X_{2n} - X_{2n-1})_{n \in \mathbb{N}}$ is independent. Indeed, for any $n \in \mathbb{N}$, the random variable $Y_n$ is $\sigma(X_{2n}, X_{2n-1})$-measurable by Theorem 1.91, and $(\sigma(X_{2n}, X_{2n-1}))_{n \in \mathbb{N}}$ is independent by Theorem 2.26.

Example 2.28. Let $(X_{m,n})_{(m,n) \in \mathbb{N}^2}$ be an independent family of Bernoulli random variables with parameter $p \in (0, 1)$. Define the waiting time for the first “success” in the $m$th row of the matrix $(X_{m,n})_{m,n}$ by

$$Y_m := \inf \{ n \in \mathbb{N} : X_{m,n} = 1 \} - 1.$$ 

Then $(Y_m)_{m \in \mathbb{N}}$ are independent geometric random variables with parameter $p$ (see Example 1.105(iii)). Indeed,

$$\{Y_m \leq k\} = \bigcup_{l=1}^{k+1} \{X_{m,l} = 1\} \in \sigma(X_{m,l}, l = 1, \ldots, k + 1) \subset \sigma(X_{m,l}, l \in \mathbb{N}).$$

Hence $Y_m$ is $\sigma(X_{m,l}, l \in \mathbb{N})$-measurable and thus $(Y_m)_{m \in \mathbb{N}}$ is independent. Furthermore,

$$P[Y_m > k] = P[X_{m,l} = 0, l = 1, \ldots, k + 1] = \prod_{l=1}^{k+1} P[X_{m,l} = 0] = (1 - p)^{k+1}.$$ 

Concluding, we get $P[Y_m = k] = P[Y_m > k - 1] - P[Y_m > k] = p(1 - p)^k$.

\[\diamond\]

Definition 2.29 (Convolution). Let $\mu$ and $\nu$ be probability measures on $(\mathbb{Z}, 2^\mathbb{Z})$. The convolution $\mu \ast \nu$ is defined as the probability measure on $(\mathbb{Z}, 2^\mathbb{Z})$ such that

$$(\mu \ast \nu)(\{n\}) = \sum_{m=-\infty}^{\infty} \mu(\{m\}) \nu(\{n-m\}).$$

We define the $n$th convolution power recursively by $\mu^{*1} = \mu$ and

$$\mu^{*(n+1)} = \mu^{*n} \ast \mu.$$ 

Remark 2.30. The convolution is a symmetric operation: $\mu \ast \nu = \nu \ast \mu$.

\[\diamond\]

Theorem 2.31. If $X$ and $Y$ are independent $\mathbb{Z}$-valued random variables, then $P_{X+Y} = P_X \ast P_Y$. 

\[\square\]
Proof. For any \( n \in \mathbb{Z} \),
\[
P_{X+Y}(\{n\}) = P[X + Y = n]
\]
\[
= P\left[ \bigcup_{m \in \mathbb{Z}} \left( \{X = m\} \cap \{Y = n - m\} \right) \right]
\]
\[
= \sum_{m \in \mathbb{Z}} P[\{X = m\} \cap \{Y = n - m\}]
\]
\[
= \sum_{m \in \mathbb{Z}} P_X(\{m\}) P_Y(\{n - m\}) = (P_X \ast P_Y)(\{n\}). \quad \Box
\]

Owing to the last theorem, it is natural to define the convolution of two probability measures on \( \mathbb{R}^n \) (or more generally on an Abelian group) as the distribution of the sum of two independent random variables with the corresponding distributions. Later we will encounter a different (but equivalent) definition that will, however, rely on the notion of an integral that is not yet available to us at this point (see Definition 14.17).

**Definition 2.32 (Convolution of measures).** Let \( \mu \) and \( \nu \) be probability measures on \( \mathbb{R}^n \) and let \( X \) and \( Y \) be independent random variables with \( P_X = \mu \) and \( P_Y = \nu \). We define the convolution of \( \mu \) and \( \nu \) as \( \mu \ast \nu = P_{X+Y} \).

Recursively, we define the convolution powers \( \mu^{*k} \) for all \( k \in \mathbb{N} \) and let \( \mu^{*0} = \delta_0 \).

**Example 2.33.** Let \( X \) and \( Y \) be independent Poisson random variables with parameters \( \mu \) and \( \lambda \geq 0 \). Then
\[
P[X + Y = n] = e^{-\mu} e^{-\lambda} \sum_{m=0}^{n} \frac{\mu^m \lambda^{n-m}}{m! (n-m)!}
\]
\[
= e^{-(\mu+\lambda)} \frac{1}{n!} \sum_{m=0}^{n} \binom{n}{m} \mu^m \lambda^{n-m} = e^{-(\mu+\lambda)} \frac{(\mu + \lambda)^n}{n!}.
\]
Hence \( \text{Poi}_\mu \ast \text{Poi}_\lambda = \text{Poi}_{\mu+\lambda} \). \( \diamondsuit \)

**Exercise 2.2.1.** Let \( X \) and \( Y \) be independent random variables with \( X \sim \text{exp}_\theta \) and \( Y \sim \text{exp}_\rho \) for certain \( \theta, \rho > 0 \). Show that
\[
P[X < Y] = \frac{\theta}{\theta + \rho}. \quad \clubsuit
\]

**Exercise 2.2.2 (Box-Muller method).** Let \( U \) and \( V \) be independent random variables that are uniformly distributed on \([0, 1]\). Define
\[
X := \sqrt{-2 \log(U)} \cos(2\pi V) \quad \text{and} \quad Y := \sqrt{-2 \log(U)} \sin(2\pi V).
\]
Show that $X$ and $Y$ are independent and $N_{0,1}$-distributed.

*Hint:* First compute the distribution of $\sqrt{-2\log(U)}$ and then use the transformation formula (Theorem 1.101) as well as polar coordinates.

---

### 2.3 Kolmogorov’s 0-1 Law

With the Borel-Cantelli lemma, we have seen a first 0-1 law for independent events. We now come to another 0-1 law for independent events and for independent σ-algebras. To this end, we first introduce the notion of the tail σ-algebra.

**Definition 2.34 (Tail σ-algebra).** Let $I$ be a countably infinite index set and let $(A_i)_{i \in I}$ be a family of σ-algebras. Then

$$T((A_i)_{i \in I}) := \bigcap_{\substack{J \subseteq I \\# J < \infty}} \sigma \left( \bigcup_{j \in I \setminus J} A_j \right)$$

is called the **tail σ-algebra** of $(A_i)_{i \in I}$. If $(A_i)_{i \in I}$ is a family of events, then we define

$$T((A_i)_{i \in I}) := T((\emptyset, A_i, A_i^c, \Omega)_{i \in I}).$$

If $(X_i)_{i \in I}$ is a family of random variables, then we define $T((X_i)_{i \in I}) := T((\sigma(X_i))_{i \in I}).$

The tail σ-algebra contains those events $A$ whose occurrence is independent of any fixed finite subfamily of the $X_i$. To put it differently, for any finite subfamily of the $X_i$, we can change the values of the $X_i$ arbitrarily without changing whether $A$ occurs or not.

**Theorem 2.35.** Let $J_1, J_2, \ldots$ be finite sets with $J_n \uparrow I$. Then

$$T((A_i)_{i \in I}) = \bigcap_{n=1}^{\infty} \sigma \left( \bigcup_{m \in I \setminus J_n} A_m \right).$$

In the particular case $I = \mathbb{N}$, this reads $T((A_n)_{n \in \mathbb{N}}) = \bigcap_{n=1}^{\infty} \sigma \left( \bigcup_{m=n}^{\infty} A_m \right)$.

The name “tail σ-algebra” is due to the interpretation of $I = \mathbb{N}$ as a set of times. As is made clear in the theorem, any event in $T$ does not depend on the first finitely many time points.

**Proof.** “⊂” This is clear.
“…” Let \( J_n \subset I, n \in \mathbb{N} \), be finite sets with \( J_n \nrightarrow I \). Let \( J \subset I \) be finite. Then there exists an \( N \in \mathbb{N} \) with \( J \subset J_N \) and

\[
\bigcap_{n=1}^{\infty} \sigma\left( \bigcup_{m \in I \setminus J_n} A_m \right) \subset \bigcap_{n=1}^{N} \sigma\left( \bigcup_{m \in I \setminus J_n} A_m \right) = \sigma\left( \bigcup_{m \in I \setminus J_N} A_m \right) \subset \sigma\left( \bigcup_{m \in I \setminus J} A_m \right).
\]

The left hand side does not depend on \( J \). Hence we can form the intersection over all finite \( J \) and obtain

\[
\bigcap_{n=1}^{\infty} \sigma\left( \bigcup_{m \in I \setminus J_n} A_m \right) \subset T((A_i)_{i \in I}).
\]

Maybe at first glance it is not evident that there are any interesting events in the tail \( \sigma \)-algebra at all. It might not even be clear that we do not have \( T = \{\emptyset, \Omega\} \). Hence we now present simple examples of tail events and tail \( \sigma \)-algebra measurable random variables. In Section 2.4, we will study a more complex example.

**Example 2.36.** (i) Let \( A_1, A_2, \ldots \) be events. Then the events \( A_* := \liminf_{n \to \infty} A_n \) and \( A^* := \limsup_{n \to \infty} A_n \) are in \( T((A_n)_{n \in \mathbb{N}}) \). Indeed, if we define \( B_n := \bigcap_{m=n}^{\infty} A_m \) for \( n \in \mathbb{N} \), then \( B_n \nrightarrow A_* \) and \( B_n \in \sigma((A_m)_{m \geq N}) \) for any \( n \geq N \). Thus \( A_* \in \sigma((A_m)_{m \geq N}) \) for any \( N \in \mathbb{N} \) and hence \( A_* \in T((A_n)_{n \in \mathbb{N}}) \). The case \( A^* \) is similar.

(ii) Let \((X_n)_{n \in \mathbb{N}}\) be a family of \( \mathbb{R} \)-valued random variables. Then the maps \( X_* := \liminf_{n \to \infty} X_n \) and \( X^* := \limsup_{n \to \infty} X_n \) are \( T((X_n)_{n \in \mathbb{N}}) \)-measurable. Indeed, if we define \( Y_n := \sup_{m \geq N} X_m \), then for any \( N \in \mathbb{N} \), the random variable \( X^* = \inf_{n \geq 1} Y_n = \liminf_{n \geq N} Y_n \) is \( T_N := \sigma(X_n, n \geq N) \)-measurable and hence also measurable with respect to \( T((X_n)_{n \in \mathbb{N}}) = \bigcap_{n=1}^{\infty} T_n \).

The case \( X_* \) is similar.

(iii) Let \((X_n)_{n \in \mathbb{N}}\) be real random variables. Then the Cesàro limits

\[
\liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i \quad \text{and} \quad \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i
\]

are \( T((X_n)_{n \in \mathbb{N}}) \)-measurable. In order to show this, choose \( N \in \mathbb{N} \) and note that

\[
X_* := \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \liminf_{n \to \infty} \frac{1}{n} \sum_{i \geq N} X_i
\]

is \( \sigma((X_n)_{n \geq N}) \)-measurable. Since this holds for any \( N \), \( X_* \) is \( T((X_n)_{n \in \mathbb{N}}) \)-measurable. The case of the limes superior is similar. \( \diamond \)
Corollary 2.38. Let $(A_i)_{i \in I}$ be an independent family of $\sigma$-algebras. Then the tail $\sigma$-algebra is $P$-trivial, that is,

$$P[A] \in \{0, 1\} \text{ for any } A \in \mathcal{T}((A_i)_{i \in I}).$$

Proof. It is enough to consider the case $I = \mathbb{N}$. For $n \in \mathbb{N}$, let

$$\mathcal{F}_n := \left\{ \bigcap_{k=1}^{n} A_k : A_1 \in \mathcal{A}_1, \ldots, A_n \in \mathcal{A}_n \right\}.$$

Then $\mathcal{F} := \bigcup_{n=1}^{\infty} \mathcal{F}_n$ is a semiring and $\sigma(\mathcal{F}) = \sigma(\bigcup_{n \in \mathbb{N}} \mathcal{A}_n)$. Indeed, for any $n \in \mathbb{N}$ and $A_n \in \mathcal{A}_n$, we have $A_n \in \mathcal{F}$; hence $\sigma(\bigcup_{n \in \mathbb{N}} \mathcal{A}_n) \subset \sigma(\mathcal{F})$. On the other hand, we have $\mathcal{F}_m \subset \sigma(\bigcup_{n=1}^{m} \mathcal{A}_n) \subset \sigma(\bigcup_{n \in \mathbb{N}} \mathcal{A}_n)$ for any $m \in \mathbb{N}$; hence $\mathcal{F} \subset \sigma(\bigcup_{n \in \mathbb{N}} \mathcal{A}_n)$.

Let $A \in \mathcal{T}((\mathcal{A}_n)_{n \in \mathbb{N}})$ and $\varepsilon > 0$. By the approximation theorem for measures (Theorem 1.65), there exists an $N \in \mathbb{N}$ and mutually disjoint sets $F_1, \ldots, F_N \in \mathcal{F}$ such that $P[A \triangle (F_1 \cup \ldots \cup F_N)] < \varepsilon$. Clearly, there is an $n \in \mathbb{N}$ such that $F_1, \ldots, F_N \in \mathcal{F}_n$ and thus $F := F_1 \cup \ldots \cup F_N \in \sigma(\mathcal{A}_1 \cup \ldots \cup \mathcal{A}_n)$. Obviously, $A \in \sigma(\bigcup_{m=n+1}^{\infty} \mathcal{A}_m)$; hence $A$ is independent of $F$. Thus

$$\varepsilon > P[A \setminus F] = P[A \cap (\Omega \setminus F)] = P[A](1 - P[F]) \geq P[A](1 - P[A] - \varepsilon).$$

Letting $\varepsilon \downarrow 0$ yields $0 = P[A](1 - P[A]). \Box$

Corollary 2.39. Let $(A_n)_{n \in \mathbb{N}}$ be a sequence of independent events. Then

$$P\left[ \limsup_{n \to \infty} A_n \right] \in \{0, 1\} \text{ and } P\left[ \liminf_{n \to \infty} A_n \right] \in \{0, 1\}.$$

Proof. Essentially this is a simple conclusion of the Borel-Cantelli lemma. However, the statement can also be deduced from Kolmogorov’s 0-1 law as limes superior and limes inferior are in the tail $\sigma$-algebra. \Box

Corollary 2.38. Let $(X_n)_{n \in \mathbb{N}}$ be an independent family of $\mathbb{R}$-valued random variables. Then $X_* := \liminf_{n \to \infty} X_n$ and $X^* := \limsup_{n \to \infty} X_n$ are almost surely constant. That is, there exist $x_*, x^* \in \mathbb{R}$ such that $P[X_* = x_*] = 1$ and $P[X^* = x^*] = 1$.

If all $X_i$ are real-valued, then the Cesàro limits

$$\liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i \text{ and } \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i$$

are also almost surely constant.
Proof. Let $X_* := \liminf_{n \to \infty} X_n$. For any $x \in \mathbb{R}$, we have $\{X_* \leq x\} \in T((X_n)_{n \in \mathbb{N}})$; hence $P[X_* \leq x] \in \{0, 1\}$. Define

$$x_* := \inf\{x \in \mathbb{R} : P[X_* \leq x] = 1\} \in \mathbb{R}.$$ 

If $x_* = \infty$, then evidently

$$P[X_* < \infty] = \lim_{n \to \infty} P[X_* \leq n] = 0.$$ 

If $x_* \in \mathbb{R}$, then

$$P[X_* \leq x_*] = \lim_{n \to \infty} P\left[X_* \leq x_* + \frac{1}{n}\right] = 1$$

and

$$P[X_* < x_*] = \lim_{n \to \infty} P\left[X_* \leq x_* - \frac{1}{n}\right] = 0.$$ 

If $x_* = -\infty$, then

$$P[X_* > -\infty] = \lim_{n \to \infty} P[X_* > -n] = 0.$$ 

The cases of the limes superior and the Cesàro limits are similar. \square

Exercise 2.3.1. Let $(X_n)_{n \in \mathbb{N}}$ be an independent family of $\text{Rad}_{1/2}$ random variables (i.e., $P[X_n = -1] = P[X_n = +1] = \frac{1}{2}$) and let $S_n = X_1 + \ldots + X_n$ for any $n \in \mathbb{N}$. Show that $\limsup_{n \to \infty} S_n = \infty$ almost surely. ♣

2.4 Example: Percolation

Consider the $d$-dimensional integer lattice $\mathbb{Z}^d$, where any point is connected to any of its $2d$ nearest neighbours by an edge. If $x, y \in \mathbb{Z}^d$ are nearest neighbours (that is, $\|x - y\|_2 = 1$), then we denote by $e = \langle x, y \rangle = \langle y, x \rangle$ the edge that connects $x$ and $y$. Formally, the set of edges is a subset of the set of subsets of $\mathbb{Z}^d$ with two elements:

$$E = \{\{x, y\} : x, y \in \mathbb{Z}^d \text{ with } \|x - y\|_2 = 1\}.$$ 

Somewhat more generally, an undirected graph $G$ is a pair $G = (V, E)$, where $V$ is a set (the set of “vertices” or nodes) and $E \subset \{\{x, y\} : x, y \in V, x \neq y\}$ is a subset of the set of subsets of $V$ of cardinality two (the set of edges or bonds).

Our intuitive understanding of an edge is a connection between two points $x$ and $y$ and not an (unordered) pair $\{x, y\}$. To stress this notion of a connection, we use a different symbol from the set brackets. That is, we denote the edge that connects $x$ and $y$ by $\langle x, y \rangle = \langle y, x \rangle$ instead of $\{x, y\}$. 
Our graph \((V, E)\) is the starting point for a stochastic model of a porous medium. We interpret the edges as tubes along which water can flow. However, we want the medium not to have a homogeneous structure, such as \(\mathbb{Z}^d\), but an amorphous structure. In order to model this, we randomly destroy a certain fraction \(1 - p\) of the tubes (with \(p \in [0, 1]\) a parameter) and keep the others. Water can flow only through the remaining tubes. The destroyed tubes will be called “closed”, the others “open”. The fundamental question is: For which values of \(p\) is there a connected infinite system of tubes along which water can flow? The physical interpretation is that if we throw a block of the considered material into a bathtub, then the block will soak up water; that is, it will be wetted inside. If there is no infinite open component, then the water may wet only a thin layer at the surface.

We now come to a formal description of the model. Choose a parameter \(p \in [0, 1]\) and an independent family of identically distributed random variables \((X^p_e)_{e \in E}\) with \(X^p_e \sim \text{Ber}_p\); that is, \(\mathbf{P}[X^p_e = 1] = 1 - \mathbf{P}[X^p_e = 0] = p\) for any \(e \in E\). We define the set of open edges as

\[
E^p := \{e \in E : X^p_e = 1\}. \tag{2.10}
\]

Consequently, the edges in \(E \setminus E^p\) are called closed. Hence we have constructed a (random) subgraph \((\mathbb{Z}^d, E^p)\) of \((\mathbb{Z}^d, E)\). We call \((\mathbb{Z}^d, E^p)\) a percolation model (more precisely, a model for bond percolation, in contrast to site percolation, where vertices can be open or closed). An (open) path (of length \(n\)) in this subgraph is a sequence \(\pi = (x_0, x_1, \ldots, x_n)\) of points in \(\mathbb{Z}^d\) with \(\langle x_{i-1}, x_i \rangle \in E^p\) for all \(i = 1, \ldots, n\). We say that two points \(x, y \in \mathbb{Z}^d\) are connected by an open path if there is an \(n \in \mathbb{N}\) and an open path \((x_0, x_1, \ldots, x_n)\) with \(x_0 = x\) and \(x_n = y\). In this case, we write \(x \leftrightarrow^p y\). Note that “\(\leftrightarrow^p\)” is an equivalence relation; however, a random one, as it depends on the values of the random variables \((X^p_e)_{e \in E}\) for every \(x \in \mathbb{Z}^d\), we define the (random) open cluster of \(x\); that is, the connected component of \(x\) in the graph \((\mathbb{Z}^d, E^p)\):

\[
C^p(x) := \{y \in \mathbb{Z}^d : x \leftrightarrow^p y\}. \tag{2.11}
\]

**Lemma 2.40.** Let \(x, y \in \mathbb{Z}^d\). Then \(\mathbbm{1}_{\{x \leftrightarrow^p y\}}\) is a random variable. In particular, \(#C^p(x)\) is a random variable for any \(x \in \mathbb{Z}^d\).

**Proof.** We may assume \(x = 0\). Let \(f_n(y) = 1\) if there exists an open path of length at most \(n\) that connects 0 to \(y\), and \(f_n(y) = 0\) otherwise. Clearly, \(f_n(y) \uparrow \mathbbm{1}_{\{0 \leftrightarrow^p y\}}\); hence it suffices to show that each \(f_n\) is measurable. Let \(B_n := \{-n, -n+1, \ldots, n-1, n\}^d\) and \(E_n := \{e \in E : e \cap B_n \neq \emptyset\}\). Then \(Y_n := (X^p_e : e \in E_n) : \Omega \rightarrow \{0, 1\}^{E_n}\) is measurable (with respect to \(\mathcal{F}_{\{0, 1\}^{E_n}}\)) by Theorem 1.90. However, \(f_n\) is a function of \(Y_n\), say \(f_n = g_n \circ Y_n\) for some map \(g_n : \{0, 1\}^{E_n} \rightarrow \{0, 1\}\). By the composition theorem for maps (Theorem 1.80), \(f_n\) is measurable. The additional statement holds since \(#C^p(x) = \sum_{y \in \mathbb{Z}^d} \mathbbm{1}_{\{x \leftrightarrow^p y\}}\). \(\square\)
Definition 2.41. We say that percolation occurs if there exists an infinitely large open cluster. We call
\[ \psi(p) := \mathbb{P} \left[ \text{there exists an infinite open cluster} \right] \]
\[ = \mathbb{P} \left[ \bigcup_{x \in \mathbb{Z}^d} \{ #C^p(x) = \infty \} \right] \]
the probability of percolation. We define
\[ \theta(p) := \mathbb{P} \left[ #C^p(0) = \infty \right] \]
as the probability that the origin is in an infinite open cluster.

By the translation invariance of the lattice, we have
\[ \theta(p) = \mathbb{P} \left[ #C^p(y) = \infty \right] \quad \text{for any } y \in \mathbb{Z}^d. \quad (2.12) \]

The fundamental question is: How large are \( \theta(p) \) and \( \psi(p) \) depending on \( p \)?

We make the following simple observation.
Theorem 2.42. The map $[0, 1] \rightarrow [0, 1]$, $p \mapsto \theta(p)$ is monotone increasing.

**Proof.** Although the statement is intuitively so clear that it might not need a proof, we give a formal proof in order to introduce a technique called **coupling**. Let $p, p' \in [0, 1]$ with $p < p'$. Let $(Y_e)_{e \in E}$ be an independent family of random variables with $P[Y_e \leq q] = q$ for any $e \in E$ and $q \in \{p, p', 1\}$. At this point, we could, for example, assume that $Y_e \sim U_{[0, 1]}$ is uniformly distributed on $[0, 1]$. Since we have not yet shown the existence of an independent family with this distribution, we content ourselves with $Y_e$ that assume only three values $\{p, p', 1\}$. Hence

\[
P[Y_e = q] = \begin{cases} 
  p, & \text{if } q = p, \\
  p' - p, & \text{if } q = p', \\
  1 - p', & \text{if } q = 1.
\end{cases}
\]

Such a family $(Y_e)_{e \in E}$ exists by Theorem 2.19. For $q \in \{p, p'\}$ and $e \in E$, we define

\[
X^q_e := \begin{cases} 
  1, & \text{if } Y_e \leq q, \\
  0, & \text{else}.
\end{cases}
\]

Clearly, for any $q \in \{p, p'\}$, the family $(X^q_e)_{e \in E}$ is independent (see Remark 2.15(ii)) and $X^q_e \sim \text{Ber}_q$. Furthermore, $X^p_e \leq X^{p'}_e$ for any $e \in E$. The procedure of defining two families of random variables that are related in a specific way (here “≤”) on one probability space is called a **coupling**.

Clearly, $C^p(x) \subset C^{p'}(x)$ for any $x \in \mathbb{Z}^d$; hence $\theta(p) \leq \theta(p')$. \hfill \Box

With the aid of Kolmogorov’s 0-1 law, we can infer the following theorem.

Theorem 2.43. For any $p \in [0, 1]$, we have

\[
\Psi(p) = \begin{cases} 
  0, & \text{if } \theta(p) = 0, \\
  1, & \text{if } \theta(p) > 0.
\end{cases}
\]

**Proof.** If $\theta(p) = 0$, then by (2.12)

\[
\Psi(p) \leq \sum_{y \in \mathbb{Z}^d} P[\#C^p(y) = \infty] = \sum_{y \in \mathbb{Z}^d} \theta(p) = 0.
\]

Now let $A = \bigcup_{y \in \mathbb{Z}^d} \{\#C^p(y) = \infty\}$. Clearly, $A$ remains unchanged if we change the state of finitely many edges. That is, $A \in \sigma((X^p_F)_{e \in E\setminus F})$ for every finite $F \subset E$. Hence $A$ is in the tail $\sigma$-algebra $T((X^p_F)_{e \in E})$ by Theorem 2.35. Kolmogorov’s 0-1 law (Theorem 2.37) implies that $\Psi(p) = P[A] \in \{0, 1\}$. If $\theta(p) > 0$, then $\Psi(p) \geq \theta(p)$ implies $\Psi(p) = 1$. \hfill \Box

Due to the monotonicity, we can make the following definition.

**Definition 2.44.** The critical value $p_c$ for percolation is defined as

\[
p_c = \inf\{p \in [0, 1] : \theta(p) > 0\} = \sup\{p \in [0, 1] : \theta(p) = 0\} = \inf\{p \in [0, 1] : \psi(p) = 1\} = \sup\{p \in [0, 1] : \psi(p) = 0\}.
\]
We come to the main theorem of this section.

**Theorem 2.45.** For \( d = 1 \), we have \( p_c = 1 \). For \( d \geq 2 \), we have \( p_c(d) \in \left[ \frac{1}{2d-1}, \frac{2}{3} \right] \).

**Proof.** First consider \( d = 1 \) and \( p < 1 \). Let \( A^- := \{ X_{(n,n+1)}^p = 0 \text{ for some } n < 0 \} \) and \( A^+ := \{ X_{(n,n+1)}^p = 0 \text{ for some } n > 0 \} \). Let \( A = A^- \cap A^+ \). By the Borel-Cantelli lemma, we get \( \Pr[A^-] = \Pr[A^+] = 1 \). Hence \( \theta(p) = \Pr[A^c] = 0 \).

Now assume \( d \geq 2 \).

**Lower bound.** First we show \( p_c \geq \frac{1}{2d-1} \). Clearly, for any \( n \in \mathbb{N} \),

\[
\Pr[\#C^p(0) = \infty] \leq \Pr[\text{there is an } x \in C^p(0) \text{ with } \|x\|_\infty = n].
\]

We want to estimate the probability that there exists a point \( x \in C^p(0) \) with distance \( n \) from the origin. Any such point is connected to the origin by a path without self-intersections \( \pi \) that starts at \( 0 \) and has length \( m \geq n \). Let \( \Pi_{0,m} \) be the set of such paths. Clearly, \( \# \Pi_{0,m} \leq 2d \cdot (2d-1)^{m-1} \) since there are \( 2d \) choices for the first step and at most \( 2d - 1 \) choices for any further step. For any \( \pi \in \Pi_{0,m} \), the probability that \( \pi \) uses only open edges is

\[
\Pr[\pi \text{ is open}] = p^m.
\]

Hence, for \( p < \frac{1}{2d-1} \),

\[
\theta(p) \leq \sum_{m=n}^{\infty} \sum_{\pi \in \Pi_{0,m}} \Pr[\pi \text{ is open}] \\
\leq \frac{2d}{2d-1} \sum_{m=n}^{\infty} ((2d-1)p)^m \\
= \frac{2d}{(2d-1)(1 - (2d-1)p)} ((2d-1)p)^m \xrightarrow{n \to \infty} 0.
\]

We conclude that \( p_c \geq \frac{1}{2d-1} \).

**Upper bound.** We can consider \( \mathbb{Z}^d \) as a subset of \( \mathbb{Z}^d \times \{0\} \subset \mathbb{Z}^{d+1} \). Hence, if percolation occurs for \( p \) in \( \mathbb{Z}^d \), then it also occurs for \( p \) in \( \mathbb{Z}^{d+1} \). Hence the corresponding critical values are ordered \( p_c(d+1) \leq p_c(d) \).

Thus, it is enough to consider the case \( d = 2 \). Here we show \( p_c \leq \frac{2}{3} \) by using a contour argument due to Peierls ([122]), originally designed for the Ising model of a ferromagnet, see Example 18.21 and (18.13).

For \( N \in \mathbb{N} \), we define (compare (2.11) with \( x = (i,0) \))

\[
C_N := \bigcup_{i=0}^{N} C^p((i,0))
\]
as the set of points that are connected (along open edges) to at least one of the points in \( \{0, \ldots, N\} \times \{0\} \). Due to the subadditivity of probability (and since \( \mathbb{P}[\#C^p((i, 0)) = \infty] = \theta(p) \) for any \( i \in \mathbb{Z} \)), we have

\[
\theta(p) = \frac{1}{N + 1} \sum_{i=0}^{N} \mathbb{P}[\#C^p((i, 0)) = \infty] \geq \frac{1}{N + 1} \mathbb{P}[\#C_N = \infty].
\]

Now consider those closed contours in the dual graph \( (\mathbb{Z}^2, \tilde{E}) \) that surrounds \( C_N \) if \( \#C_N < \infty \). Here the dual graph is defined by

\[
\mathbb{Z}^2 = \left( \frac{1}{2}, \frac{1}{2} \right) + \mathbb{Z}^2,
\]

\[
\tilde{E} = \left\{ \{x, y\} : x, y \in \mathbb{Z}^2, \|x - y\|_2 = 1 \right\}.
\]

An edge \( \tilde{e} \) in the dual graph \( (\mathbb{Z}^2, \tilde{E}) \) crosses exactly one edge \( e \) in \( (\mathbb{Z}^2, E) \). We call \( \tilde{e} \) open if \( e \) is open and closed otherwise. A circle \( \gamma \) is a self-intersection free path in \( (\mathbb{Z}^2, \tilde{E}) \) that starts and ends at the same point. A contour of the set \( C_N \) is a minimal circle that surrounds \( C_N \). Minimal means that the enclosed area is minimal (see Fig. 2.2). For \( n \geq 2N \), let

\[
\Gamma_n = \left\{ \gamma : \gamma \text{ is a circle of length } n \text{ that surrounds } \{0, \ldots, N\} \times \{0\} \right\}.
\]
We want to deduce an upper bound for $\# \Gamma_n$. Let $\gamma \in \Gamma_n$ and fix one point of $\gamma$. For definiteness, choose the upper point $(m + \frac{1}{2}, \frac{1}{2})$ of the rightmost edge of $\gamma$ that crosses the horizontal axis (in Fig. 2.2 this is the point $(5 + \frac{1}{2}, \frac{1}{2})$). Clearly, $m \geq N$ and $m \leq n$ since $\gamma$ surrounds the origin. Starting from $(m + \frac{1}{2}, \frac{1}{2})$, for any further edge of $\gamma$, there are at most three possibilities. Hence

$$\# \Gamma_n \leq n \cdot 3^n.$$ 

We say that $\gamma$ is closed if it uses only closed edges (in $\tilde{E}$). A contour of $C_N$ is automatically closed and has a length of at least $2N$. Hence for $p > \frac{2}{3}$

$$\mathbb{P}[\# C_N < \infty] = \sum_{n=2N}^{\infty} \mathbb{P}[\text{there is a closed circle } \gamma \in \Gamma_n] \leq \sum_{n=2N}^{\infty} n \cdot (3(1-p))^n \xrightarrow{N \to \infty} 0.$$ 

We conclude $p_c \leq \frac{2}{3}$. \hfill \qedsymbol

In general, the value of $p_c$ is not known and is extremely hard to determine. In the case of bond percolation on $\mathbb{Z}^2$, however, the exact value of $p_c$ can be determined due to the self-duality of the planar graph $(\mathbb{Z}^2, E)$. (If $G = (V, E)$ is a planar graph; that is, a graph that can be embedded into $\mathbb{R}^2$ without self-intersections, then the vertex set of the dual graph is the set of faces of $G$. Two such vertices are connected by exactly one edge; that is, by the edge in $E$ that separates the two faces. Evidently, the two-dimensional integer lattice is isomorphic to its dual graph. Note that the contour in Fig. 2.2 can be considered as a closed path in the dual graph.) We cite a theorem of Kesten [92].

**Theorem 2.46 (Kesten (1980)).** For bond percolation in $\mathbb{Z}^2$, the critical value is $p_c = \frac{1}{2}$ and $\theta(p_c) = 0$.

**Proof.** See, for example, the book of Grimmett [60, pages 287ff]. \hfill \qedsymbol

It is conjectured that $\theta(p_c) = 0$ holds in any dimension $d \geq 2$. However, rigorous proofs are known only for $d = 2$ and $d \geq 19$ (see [64]).

**Uniqueness of the Infinite Open Cluster**

Fix a $p$ such that $\theta(p) > 0$. We saw that with probability one there is at least one infinite open cluster. Now we want to show that there is exactly one.

Denote by $N \in \{0, 1, \ldots, \infty\}$ the (random) number of infinite open clusters.
2.4 Example: Percolation 73

**Theorem 2.47 (Uniqueness of the infinite open cluster).** For any $p \in [0, 1]$, we have $P_p[N \leq 1] = 1$.

**Proof.** This theorem was first proved by Aizenman, Kesten and Newman [1, 2]. Here we follow the proof of Burton and Keane [21] as described in [60, Section 8.2].

The cases $p = 1$ and $\theta(p) = 0$ (hence in particular the case $p = 0$) are trivial. Hence we assume now that $p \in (0, 1)$ and $\theta(p) > 0$.

**Step 1.** We first show that

$$P_p[N = m] = 1 \quad \text{for some } m = 0, 1, \ldots, \infty. \quad (2.13)$$

We need a 0-1 law similar to that of Kolmogorov. However, $N$ is not measurable with respect to the tail $\sigma$-algebra. Hence we have to find a more subtle argument. Let $u_1 = (1, 0, \ldots, 0)$ be the first unit vector in $Z^d$. On the edge set $E$, define the translation $\tau : E \to E$ by $\tau(\langle x, y \rangle) = \langle x + u_1, y + u_1 \rangle$. Let

$$E_0 := \{(x_1, \ldots, x_d), (y_1, \ldots, y_d) \in E : x_1 = 0, y_1 \geq 0\}$$

be the set of all edges in $Z^d$ that either connect two points from $\{0\} \times Z^{d-1}$ or one point of $\{0\} \times Z^{d-1}$ with one point of $\{1\} \times Z^{d-1}$. Clearly, the sets $(\tau^n(E_0), n \in \mathbb{Z})$ are disjoint and $E = \bigcup_{n \in \mathbb{Z}} \tau^n(E_0)$. Hence the random variables $Y_n := (X^p_{\tau^n(e)})_{e \in E_0}, n \in \mathbb{Z}$, are independent and identically distributed (with values in $\{0, 1\}^{E_0}$). Define $Y = (Y_n)_{n \in \mathbb{Z}}$ and $\tau(Y) = (Y_{n+1})_{n \in \mathbb{Z}}$. Define $A_m \in \{0, 1\}^E$ by

$$\{Y \in A_m\} = \{N = m\}.$$ 

Clearly, the value of $N$ does not change if we shift *all* edges simultaneously. That is, $\{Y \in A_m\} = \{\tau(Y) \in A_m\}$. An event with this property is called *invariant* or shift invariant. Using an argument similar to that in the proof of Kolmogorov’s 0-1 law, one can show that invariant events (defined by i.i.d. random variables) have probability either 0 or 1 (see Example 20.26 for a proof).

**Step 2.** We will show that

$$P_p[N = m] = 0 \quad \text{for any } m \in \mathbb{N} \setminus \{1\}. \quad (2.14)$$

Accordingly, let $m = 2, 3, \ldots$. We assume that $P[N = m] = 1$ and show that this leads to a contradiction.

For $L \in \mathbb{N}$, let $B_L := \{-L, \ldots, L\}^d$ and denote by $E_L = \{e = \langle x, y \rangle \in E : x, y \in B_L\}$ the set of those edges with both vertices lying in $B_L$. For $i = 0, 1$, let $D^i_L := \{X^p_e = i \text{ for all } e \in E_L\}$. Let $N^i_L$ be the number of infinite open clusters if we consider all edges $e$ in $E_L$ as open (independently of the value of $X^p_e$). Similarly define $N^0_L$ where we consider all edges in $E_L$ as closed. Since $P_p[D^i_L] > 0$, and since $N = m$ almost surely, we have $N^i_L = m$ almost surely for $i = 0, 1$. 


Let
\[ A^2_L := \bigcup_{x^1, x^2 \in B_L \setminus B_{L-1}} \left\{ C^p(x^1) \cap C^p(x^2) = \emptyset \right\} \cap \left\{ \# C^p(x^1) = \# C^p(x^2) = \infty \right\} \]
be the event where there exist two points on the boundary of \( B_L \) that lie in different infinite open clusters. Clearly, \( A^2_L \uparrow \{ N \geq 2 \} \) for \( L \to \infty \).

Define \( A^2_{L,0} \) in a similarly way to \( A^2_L \); however, we now consider all edges \( e \in E_L \) as closed, irrespective of whether \( X^p_e = 1 \) or \( X^p_e = 0 \). If \( A^2_L \) occurs, then there are two points \( x^1, x^2 \) on the boundary of \( B_L \) such that for any \( i = 1, 2 \), there is an infinite self-intersection free open path \( \pi_{x^i} \) starting at \( x^i \) that avoids \( x^{3-i} \). Hence \( A^2_L \subset A^2_{L,0} \). Now choose \( L \) large enough for \( P[A^2_{L,0}] > 0 \).

If \( A^2_{L,0} \) occurs and if we open all edges in \( B_L \), then at least two of the infinite open clusters get connected by edges in \( B_L \). Hence the total number of infinite open clusters decreases by at least one. We infer \( P_p[N^1 \leq N^0_L - 1] \geq P_p[A^2_{L,0}] > 0 \), which leads to a contradiction.

**Step 3.** In Step 2, we have shown already that \( N \) does not assume a finite value larger than 1. Hence it remains to show that almost surely \( N \) does not assume the value \( \infty \). Indeed, we show that

\[ P_p[N \geq 3] = 0. \] (2.15)

This part of the proof is the most difficult one. We assume that \( P_p[N \geq 3] > 0 \) and show that this leads to a contradiction.

We say that a point \( x \in \mathbb{Z}^d \) is a **trifurcation point** if

- \( x \) is in an infinite open cluster \( C^p(x) \),
- there are exactly three open edges with endpoint \( x \), and
- removing all of these three edges splits \( C^p(x) \) into three mutually disjoint infinite open clusters.

By \( T \) we denote the set of trifurcation points, and let \( T_L := T \cap B_L \). Let \( r := P_p[0 \in T] \). Due to translation invariance, we have \( (\# B_L)^{-1} E_p[\# T_L] = r \) for any \( L \). (Here \( E_p[\# T_L] \) denotes the expected value of \( \# T_L \), which we define formally in Chapter 5.) Let

\[ A^3_L := \bigcup_{x^1, x^2, x^3 \in B_L \setminus B_{L-1}} \left( \bigcap_{i \neq j} \left\{ C^p(x^i) \cap C^p(x^j) = \emptyset \right\} \right) \cap \left( \bigcap_{i=1}^3 \left\{ \# C^p(x^i) = \infty \right\} \right) \]

be the event where there are three points on the boundary of \( B_L \) that lie in different infinite open clusters. Clearly, \( A^3_L \uparrow \{ N \geq 3 \} \) for \( L \to \infty \).

As for \( A^2_{L,0} \), we define \( A^3_{L,0} \) as the event where there are three distinct points on the boundary of \( B_L \) that lie in different infinite open clusters if we consider all edges in \( E_L \) as closed. As above, we have \( A^3_L \subset A^3_{L,0} \).
For three distinct points $x^1, x^2, x^3 \in B_L \setminus B_{L-1}$, let $F_{x^1,x^2,x^3}$ be the event where for any $i = 1, 2, 3$, there exists an infinite self-intersection free open path $\pi_{x^i}$ starting at $x^i$ that uses only edges in $E^p \setminus E_L$ and that avoids the points $x^j, j \neq i$. Then

$$A^3_{L,0} \subset \bigcup_{x^1,x^2,x^3 \in B_L \setminus B_{L-1}} F_{x^1,x^2,x^3},$$

Let $L$ be large enough for $P_p[A^3_{L,0}] \geq P_p[N \geq 3]/2 > 0$. Choose three pairwise distinct points $x^1, x^2, x^3 \in B_L \setminus B_{L-1}$ with $P_p[F_{x^1,x^2,x^3}] > 0$.

If $F_{x^1,x^2,x^3}$ occurs, then we can find a point $y \in B_L$ that is the starting point of three mutually disjoint (not necessarily open) paths $\pi_1, \pi_2$ and $\pi_3$ that end at $x^1, x^2$ and $x^3$. Let $G_{y,x^1,x^2,x^3}$ be the event where in $E_L$ exactly those edges are open that belong to these three paths (that is, all other edges in $E_L$ are closed). The events $F_{x^1,x^2,x^3}$ and $G_{y,x^1,x^2,x^3}$ are independent, and if both of them occur, then $y$ is a trifurcation point. Hence

$$r = P_p[y \in T] \geq P_p[F_{x^1,x^2,x^3}] \cdot (p \wedge (1-p)) \#E_L > 0.$$

Now we show that $r$ must equal 0, which contradicts the assumption $P_p[N \geq 3] > 0$. We consider $T_L$ as the vertex set of a graph by considering two points $x, y \in T_L$ as neighbours if there exists an open path connecting $x$ and $y$ that does not hit any other point in $T$. In this case, we write $x \sim y$. A circle is a self-avoiding (finite) path that ends at its starting point. Note that the graph $(T_L, \sim)$ has no circles. Indeed, if there was an $x \in T_L$ and a self-avoiding open path that hits two points, say $y, z \in T_L$, then by removing the three edges $e \in E^p$ adjacent to $x$, the cluster $C_p(x)$ would split into at most two infinite open clusters, one of which would contain $y$ and $z$.

Write $\deg_{T_L}(x)$ for the degree of $x$; that is, the number of neighbours of $x$ in $(T_L, \sim)$. As $T_L$ has no circles, $\#T_L - \frac{1}{2} \sum_{x \in T_L} \deg_{T_L}(x)$ is the number of connected components of $T_L$, thus it is in particular nonnegative. On the other hand, $3 - \deg_{T_L}(x)$ is the number of edges $e \in E^p$ adjacent to $x$ whose removal generates an infinite open cluster in which there is no further point of $T_L$. Let $M_L$ be the number of infinite open clusters that appear if we remove all three neighbouring open edges from all points in $T_L$. Then

$$M_L = \sum_{x \in T_L} (3 - \deg_{T_L}(x)) \geq \#T_L.$$

For any of these clusters, there is (at least) one point on the boundary $B_L \setminus B_{L-1}$. Hence

$$\frac{\#T_L}{\#B_L} \leq \frac{\#(B_L \setminus B_{L-1})}{\#B_L} \leq \frac{d}{L} L \rightarrow \infty 0.$$

Now $r = (\#B_L)^{-1}E_p[\#T_L] \leq d/L$ implies $r = 0$. (Note that in the argument we used the notion of the expected value $E_p[\#T_L]$ that will be formally introduced only in Chapter 5.)