The Language of Probability

Michael Hediger

University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

May 28, 2024

References: The given script is inspired by the books of Patrick Billingsley [1] and Jean-François Le Gall [2].

Contents

1 Introduction: Part I

The following gives a list of greek letters:

α	alpha	ν	nu
β	beta	ξ,Ξ	xi
γ,Γ	gamma	o	o, omicron
δ,Δ	delta	π,Π	pi
ϵ,ε	epsilon	ρ	rho
ζ	zeta	σ,Σ	sigma
η	eta	τ	tau
θ,Θ	theta	υ,Υ	upsilon
ι	iota	ϕ,φ,Φ	phi
κ	kappa	χ	chi
λ,Λ	lambda	ψ,Ψ	psi
μ	mu	ω,Ω	omega

We clarify some logical operations on statements. If S_1 and S_2 are two statements, we write $S_1 \Rightarrow S_2$ if S_1 implies S_2. S_1 if and only if S_2 ($S_1 \Leftrightarrow S_2$) means that S_1 implies S_2 and S_2 implies S_1. Further, the following logical operators are helpful:

$$\begin{aligned} &: \text{ means } \textit{such that}\\ &\forall \text{ means } \textit{for all}\\ &\exists \text{ means } \textit{there exists}. \end{aligned}$$

1.1 Sets

Sets can be defined by their elements $A = \{\omega_1, \omega_2, \dots, \omega_n\}$ or upon a certain property

$$A = \{\omega \colon \omega \text{ has property } P\}.$$

Let A be a set, then

$$\begin{aligned} &\omega \in A \text{ means that } \omega \text{ is an element of } A,\\ &\omega \notin A \text{ means that } \omega \text{ is not an element of } A. \end{aligned}$$

Example 1.1. *The set which contains the strictly positive integers* $1, 2, 3, \dots$ *is denoted with* $\mathbb{N}$. *If* $n \in \mathbb{N}$, *then so is* $n + 1$. *It is a matter of convention whether* $0 \in \mathbb{N}$ *or not. For us,* $0 \notin \mathbb{N}$.

Example 1.2. *The set of integers is denoted with* $\mathbb{Z}$, *it contains* 0, $\mathbb{N}$, *and the set of points* $\{-n \colon n \in \mathbb{N}\}$.

Example 1.3. $\mathbb{Q}$ *is the set of rational numbers:*

$$\mathbb{Q} = \big\{q \colon q = \frac{n}{m},\ n, m \in \mathbb{Z},\ m \neq 0\big\}.$$

Example 1.4. *It can be shown that there does not exists* $q \in \mathbb{Q}$ *s.t.* $q^2 = 2$. *This shows that* $\sqrt{2} \notin \mathbb{Q}$. *The same is true for* π *and Euler's number* e. *The latter numbers belong to the set of real numbers, denoted with* $\mathbb{R}$. *In particular,* $\mathbb{R}$ *contains all the integers and rational numbers.*

Example 1.5. *Let* $A_1, \dots, A_n$, $n \in \mathbb{N}$, *be a family of sets. The Cartesian product of* $A_1, \dots, A_n$ *is given by*

$$\prod_{i=1}^{n} A_i = A_1 \times \dots \times A_n = \{\omega \colon \omega = (\omega_1, \dots, \omega_n) \colon \omega_i \in A_i,\ i = 1, \dots, n\}.$$

An element ω of $A_1 \times \cdots \times A_n$ is referred to as a vector with coordinates $\omega_i \in A_i$, $i = 1, \ldots, n$. If $A_i = A$, $i = 1, \ldots, n$, we write $A_1 \times \cdots \times A_n = A^n$. The space $\mathbb{R}^k$ is referred to as the real coordinate space of dimension k.

Definition 1.1. *Let A and B be two sets. Then, we define the following set operations for A and B.*

Equality of sets: *$A = B$ if and only if A and B contain the same elements. That is, any element of A is also an element of B and any element if B is also an element of A.*

Inclusion: *$A \subset B$ if and only if $\omega \in A$ implies that $\omega \in B$. If $A \subset B$, we say that A is a subset of B.*

Intersection: *The intersection of A and B is the set*

$$A \cap B = \{\omega \colon \omega \in A \text{ and } \omega \in B\}.$$

Union: *The union of A and B is the set*

$$A \cup B = \{\omega \colon \omega \in A \text{ or } \omega \in B\}.$$

Set difference: *The difference between A and B is the set*

$$A \setminus B = \{\omega \colon \omega \in A \text{ and } \omega \notin B\}.$$

It is often of interest to consider the intersection and union of more than just two sets. Let $\{A_i \colon i \in I\}$ be a family of sets where I is some set. Then, the intersection of A_i, $i \in I$, is defined as the set

$$\bigcap_{i \in I} A_i = \{\omega \colon \omega \in A_i \ \forall\, i \in I\}. \tag{1}$$

The union of A_i, $i \in I$, is defined as

$$\bigcup_{i \in I} A_i = \{\omega \colon \exists\, i \in I \text{ s.t. } \omega \in A_i\}. \tag{2}$$

As an example one could think of $I = \mathbb{N}$ or $I = \{1, \ldots, N\}$, where $N \in \mathbb{N}$.

We can apply Definition 1.1 to derive several elementary properties of set operations.

Proposition 1.1. *Let A, B and C be some sets.*

Properties of the inclusion:

(1.1) *$A \subset A$, i.e., each set is a subset of itself;*

(1.2) *$A \subset B$ and $B \subset A$ if and only if $A = B$;*

(1.3) *$A \subset B$ and $B \subset C$ implies that $A \subset C$.*

Associativity:

(2.1) $(A \cup B) \cup C = A \cup (B \cup C)$;

(2.2) $(A \cap B) \cap C = A \cap (B \cap C)$.

Commutativity:

(3.1) $A \cup B = B \cup A$;

(3.2) $A \cap B = B \cap A$.

Distributive law:

(4.1) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$;

(4.2) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

As an example, let us proof two items of the latter proposition.

Proof of (1.2) in Proposition 1.1. The strategy is to show that $A \subset B$ and $B \subset A$ implies that $A = B$ and $A = B$ implies that $A \subset B$ and $B \subset A$. Suppose that $A \subset B$ and $B \subset A$. By Definition 1.1, we need to show that this implies that A and B contain the same elements. Let $\omega \in A$, then $\omega \in B$, since $A \subset B$. On the other hand, take any $\omega \in B$. Since $B \subset A$, $\omega \in A$. Thus, any $\omega \in A$ is also an element of B and vice versa. This shows that $A \subset B$ and $B \subset A$ implies that $A = B$. For the other direction, let us assume that $A = B$. Then, given $\omega \in A$, it is true that $\omega \in B$ as well, thus by Definition 1.1, $A \subset B$. Also if $\omega \in B$, since $A = B$, it follows that $\omega \in A$. Thus again by Definition 1.1, $B \subset A$. Hence $A = B$ implies that $A \subset B$ and $B \subset A$. □

Proof of (4.1) in Proposition 1.1. We need to show that $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$ and $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$. Let $\omega \in A \cap (B \cup C)$. This implies that ω is an element of A and an element of $B \cup C$. By the definition of $B \cup C$, that means that $\omega \in B$ or $\omega \in C$. Still, $\omega \in A$, thus $\omega \in A$ and $\omega \in B$ or $\omega \in A$ and $\omega \in C$. This shows that $\omega \in (A \cap B) \cup (A \cap C)$ and hence $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$. For the other inclusion, let $\omega \in (A \cap B) \cup (A \cap C)$. Then, by the definition of $(A \cap B) \cup (A \cap C)$, $\omega \in A$ and $\omega \in B$ or $\omega \in A$ and $\omega \in C$. This means that ω is a member of $B \cup C$ and A. Hence, $\omega \in A \cap (B \cup C)$. □

Exercise 1.1. *Show (1.3) of Proposition 1.1.*

Exercise 1.2. *Let A and B be two sets. Show that $A \cap B \subset A$ and conclude that $A \subset B$ implies that $A \cap B = A$.*

Exercise 1.3. *Let A and B be two sets. Show that $B \subset A \cup B$ and conclude that $A \subset B$ implies that $A \cup B = B$.*

Exercise 1.4. *Let A and B be two sets. Show that $A \cup B = A \cup (B \setminus A)$.*

Example 1.6. *The set which has no elements is called the empty set and denoted with $\emptyset$.*

Proposition 1.2. *Given any subset A, $\emptyset \subset A$.*

Proof. Notice that we have used the definition that two sets are equal $A = B$ if and only if each element of A is also an element of B and vice versa. Assume for the moment that it is not true that for all subsets A, $\emptyset \subset A$. That means that there exists a subset A such that $\emptyset$ is not a subset of A. By Definition 1.1 that means that the empty set must contain an element which does not belong to A. This gives a contradiction and hence it is not true that there exists a subset A such that $\emptyset$ is not a subset of A, i.e., for any subsets A, $\emptyset \subset A$. □

Using Exercises 1.2 and 1.3, the latter result shows that $\emptyset \cap A = \emptyset$ and $\emptyset \cup A = A$.

Definition 1.2. *Let A and B be two sets. A and B are said to be disjoint if $A \cap B = \emptyset$. More generally, let $\{A_i : i \in I\}$ be any family of sets. $\{A_i : i \in I\}$ is said to be disjoint if $A_i \cap A_j = \emptyset$ for any $i \neq j$.*

Exercise 1.5. *Let A and B be two sets. Is it possible that $A \subset B$ (or vice versa) and A and B are disjoint.*

In the following, we list some properties of the difference between sets.

Proposition 1.3. *Let A, B and C be sets. Then,*

(i) $C \setminus (A \cap B) = (C \setminus A) \cup (C \setminus B)$;

(ii) $C \setminus (A \cup B) = (C \setminus A) \cap (C \setminus B)$;

(iii) $(B \setminus A) \cap C = (B \cap C) \setminus A$;

(iv) $(B \setminus A) \cup C = (B \cup C) \setminus (A \setminus C)$.

Notice that Exercise 1.4 is just a special case of item (iv) in Proposition 1.3 if we set $C = A$.

In what follows, it is often the case that a particular set Ω is given and one only considers subsets A of Ω. In that case, we use the notation $A^c = \Omega \setminus A$ for the complement of A in Ω.

Proposition 1.4. *Let A and B be subsets of Ω. Then,*

- $A \cup A^c = \Omega$;
- $A \cap A^c = \emptyset$;
- $A \setminus B = A \cap B^c$;
- $\emptyset^c = \Omega$;
- $\Omega^c = \emptyset$;
- $(A \subset B) \Rightarrow (B^c \subset A^c)$;
- $(A^c)^c = A$.

Further, it is true that

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

With reference to (1) and (2), we remark that the last two properties can be extended to unions and intersections of arbitrary families $\{A_i : i \in I\}$ of subsets of Ω:

$$\left(\bigcap_{i \in I} A_i\right)^c = \bigcup_{i \in I} A_i^c,$$
$$\left(\bigcup_{i \in I} A_i\right)^c = \bigcap_{i \in I} A_i^c.$$

We remark that since $\{A_i : i \in I\}$ consists only of subsets of Ω,

$$\bigcap_{i \in I} A_i = \{\omega \in \Omega : \omega \in A_i \ \forall\, i \in I\} \text{ and } \bigcup_{i \in I} A_i = \{\omega \in \Omega : \exists\, i \in I \text{ s.t. } \omega \in A_i\}.$$

1.2 The principle of induction

Let for any $n \in \mathbb{N}$, $S(n)$ be a statement. In order to proof that $S(n)$ is true for any $n \in \mathbb{N}$ we can adapt the following strategy:

Principle of induction To verify that $S(n)$ is true for any $n \in \mathbb{N}$ we check (I) and (II):

(I) $S(n)$ is true for $n = 1$ (base case);

(II) $S(n) \Rightarrow S(n+1)$ for any $n \in \mathbb{N}$ (induction step).

We note that if one seeks to proof $S(n)$ for $n \geq N$, $N \in \mathbb{N}$, then one needs to verify

(I) $S(N)$ is true (base case);

(II) $S(n) \Rightarrow S(n+1)$ for any $n \geq N$ (induction step).

Example 1.7. *Let us use a proof by induction to verify that for any $n \in \mathbb{N}$,*

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}.$$

First we verify the statement in the base case. Clearly, if $n = 1$, $1 = (1(1+1))/2$. For the induction step, we let $n \in \mathbb{N}$ be arbitrary and assume that $1 + 2 + \cdots + n = (n(n+1))/2$. Therefore,

$$1 + 2 + \cdots n + n + 1 = \frac{n(n+1)}{2} + n + 1 = \frac{n(n+1) + 2(n+1)}{2} = \frac{(n+1)(n+1+1)}{2},$$

which verifies the induction step and completes the argument.

Definition 1.3. *Given $n \in \mathbb{N} \cup \{0\}$, The number $n!$ is called n-factorial and given by*

$$n! = n(n-1) \cdot \ldots \cdot 1.$$

We use the convention that $0! = 1$.

Example 1.8. *We want to verify the following statement: Let $n \in \mathbb{N}$, then,*

$$\forall N \in \mathbb{N}\ \exists C > 0 \text{ s.t. } n! \geq CN^n.$$

In words, the statement reads as follows: Given any $n \in \mathbb{N}$ it is true that for any $N \in \mathbb{N}$ there exists a positive constant C s.t. $n!$ is greater or equal to CN^n. To proof it, let $n \in \mathbb{N}$ and $N \in \mathbb{N}$ be given. There are two cases, either $n \leq N$ or $n > N$. In the first case, if $n \leq N$, we can pick $C = 1/N^N$. Then,

$$n! \geq 1 \geq \frac{N^n}{N^N}.$$

Thus, in this case, no induction is needed. For the case $n > N$, we use a proof by induction on the new statement at base $n = N$:

$$\exists C > 0 \text{ s.t. } n! \geq CN^n \text{ for all } n \geq N.$$

By the first case, we have already verified the base step. Thus it remains to verify the induction step. We thus assume that for any given $n \geq N$, the latter statement is true. Then,

$$(n+1)! = (n+1)n! \geq (n+1)CN^n \geq (N+1)CN^n = CN^{n+1} + CN^n \geq CN^{n+1},$$

which completes the induction step. In summary, we started with an arbitrary $N \in \mathbb{N}$ and have shown that for both cases, $n \leq N$ and $n > N$, there exists C s.t. $n! \geq CN^n$. This shows the original statement for any $n \in \mathbb{N}$.

Exercise 1.6. *Verify that for any $n \in \mathbb{N}$,*

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

1.3 Order structure of the real numbers

Definition 1.4. *Let $a < b$, $a, b \in \mathbb{R}$. Then,*

- $[a, b] = \{x \in \mathbb{R} \colon a \leq x \leq b\}$ *is called a closed interval;*
- $(a, b) = \{x \in \mathbb{R} \colon a < x < b\}$ *is called an open interval;*

- $[a, b) = \{x \in \mathbb{R} \colon a \leq x < b\}$ *is called a right-open interval;*
- $(a, b] = \{x \in \mathbb{R} \colon a < x \leq b\}$ *is called a left-open interval.*

A set $I \subset \mathbb{R}$ is said to be an interval if it is either closed, open, right-open or left-open.

Definition 1.5. *Let $a, b \in \mathbb{R}$. The unbounded real intervals are given by the sets:*

- $[a, \infty) = \{x \in \mathbb{R} \colon a \leq x < \infty\}$, $(a, \infty) = \{x \in \mathbb{R} \colon a < x < \infty\}$;
- $(-\infty, b] = \{x \in \mathbb{R} \colon -\infty < x \leq b\}$, $(-\infty, b) = \{x \in \mathbb{R} \colon -\infty < x < b\}$.

Definition 1.6. *Let $A \subset \mathbb{R}$. An element $s \in \mathbb{R}$ is called an upper (resp. lower) bound of A, if $x \leq s$ (resp. $x \geq s$) for all $x \in A$. If A has an upper (resp. lower) bound then we say that A is bounded from above (resp. below). If A is bounded from below and above, A is bounded.*

Definition 1.7. *Let $A \subset \mathbb{R}$ be a set. An element $s \in \mathbb{R}$ is called supremum of A (we write $s = \sup A$) if s is the smallest upper bound of A. That is, the following two items are satisfied:*

(i) s is an upper bound of A;

(ii) Every number $s' < s$ is not an upper bound of A.

Example 1.9. *Let $A = [0, 1)$. Then, 1 is an upper bound for $[0, 1)$, since $x < 1 \Rightarrow x \leq 1$ for any $x \in [0, 1)$. In order to show that $\sup[0, 1) = 1$, we need to verify that 1 is the smallest upper bound of $[0, 1)$, i.e., any $s' < 1$ can not be an upper bound of $[0, 1)$. To show this, it is sufficient to proof that there exists a real number q in $[0, 1) \cap (s', 1)$ (we provide an argument later). Then, $q \in [0, 1)$ with $q > s'$ and hence s' can not be an upper bound for $[0, 1)$. Thus, $\sup[0, 1) = 1$. Notice that $\sup[0, 1) \notin [0, 1)$, i.e., the supremum must not be an element of the set itself.*

Definition 1.8. *Let $A \subset \mathbb{R}$ be a set. An element $s \in \mathbb{R}$ is called infimum of A (we write $s = \inf A$) if s is the greatest lower bound of A. That is, the following two items are satisfied:*

(i) s is a lower bound of A;

(ii) Every number $s' > s$ is not a lower bound of A.

Example 1.10. *Let $A = [0, 1)$. Since $x \geq 0$ for any $x \in [0, 1)$, 0 is a lower bound of $[0, 1)$. As $0 \in [0, 1)$, $\inf[0, 1) = 0$.*

Definition 1.9. *Let $A \subset \mathbb{R}$. If $s = \sup A \in A$ (resp. $s = \inf A \in A$) we call s the maximum (resp. the minimum) of A.*

The following result is of general importance. It shows that for each nonempty subset of the real line which has an upper (resp. lower) bound, the supremum (resp. infimum) exists.

Proposition 1.5. *Let $A \subset \mathbb{R}$ s.t. $A \neq \emptyset$. Suppose that there exists an upper (resp. lower) bound for A. Then, $\sup A$ (resp. $\inf A$) exists.*

Example 1.11. *Let $A = [0, 1)$. We have seen that the minimum of $[0, 1)$ exists. However, the maximum of $[0, 1)$ does not exist, since $\sup[0, 1) = 1 \notin [0, 1)$. This makes sense, as $[0, 1)$ is right-open and there does not exist a maximal element of $[0, 1)$.*

In order to show that there exists a number in between any two distinct real numbers, the rational numbers $\mathbb{Q}$ are helpful.

Proposition 1.6 ($\mathbb{Q}$ is dense in $\mathbb{R}$)**.** *For any two real numbers $x_1, x_2 \in \mathbb{R}$ (say $x_1 < x_2$), there exists a rational number between x_1 and x_2, i.e., there exists $q \in \mathbb{Q}$ s.t. $x_1 < q < x_2$.*

To proof the above result, we rely on two fundamental results.

Proposition 1.7. *For any $x \in \mathbb{R}$ there exists $n \in \mathbb{N}$ s.t. $n > x$.*

Proposition 1.8. *Let $A \subset \mathbb{Z}$ s.t. $A \neq \emptyset$. If A has an upper (resp. lower) bound, then A has a maximum (resp. minimum).*

Proof of Proposition 1.6. By Proposition 1.7, since $x_2 \neq x_1$, let $n \in \mathbb{N}$ s.t.

$$n > \frac{1}{x_2 - x_1}.$$

Thus, $1/n < x_2 - x_1$. Let us define the following set

$$A = \{a \in \mathbb{Z} \colon a > nx_1\}.$$

By Proposition 1.7, $A \neq \emptyset$ Further, for any $a \in A$, $a > nx_1 \Rightarrow a \geq nx_1$. Hence, nx_1 is a lower bound for A. By Proposition 1.8, there exists a minimum m of A. Then, we must have that

$$m > nx_1 \text{ (since } m \in A) \text{ but } m - 1 \leq nx_1,$$

since otherwise we have $m - 1 \in A$ which contradicts that m is the smallest number which is strictly greater than nx_1. It follows that

$$x_1 < \frac{m}{n} = \frac{m-1}{n} + \frac{1}{n} \leq x_1 + \frac{1}{n} < x_1 + x_2 - x_1 = x_2.$$

Hence if we let $q = m/n$, the result follows. □

Example 1.12. *In Example 1.9, we have postponed an argument that there exists a real number in $[0,1) \cap (s',1)$. Clearly, only the case $s' > 0$ (i.e., $[0,1) \cap (s',1) = (s',1)$) is of interest since otherwise $s' \leq 0$ and then s' is clearly no upper bound for $[0,1)$. Using Proposition 1.6 we find $q \in \mathbb{Q} \subset \mathbb{R}$ s.t. $q \in (s',1)$ and the result follows.*

Definition 1.10. *Let $A \subset \mathbb{R}$ s.t. $A \neq \emptyset$. We define,*

(i) $\sup A = \infty$ *if A has no upper bound;*

(ii) $\inf A = -\infty$ *if A has no lower bound.*

Proposition 1.9. *Let $A, B \subset \mathbb{R}$ s.t. $A \subset B$ $(A, B \neq \emptyset)$. Then,*

$$\inf A \geq \inf B \quad \textit{and} \quad \sup A \leq \sup B.$$

Proof. Let $m = \inf B$. Then, m is s.t. $m \leq x$ for any $x \in B$. Since $A \subset B$, every element of A is an element of B and thus $m \leq x$ for any $x \in A$. Then, by definition, $\inf A$ is the greatest lower bound of A and m was found to be a lower bound of A, thus $m \leq \inf A$. Let $m = \sup M$. Then, $x \leq M$ for any $x \in B$. In particular, since $A \subset B$, $x \leq M$ for any $x \in A$. By definition, $\sup A$ is the smallest upper bound of A, thus $\sup A \leq M$. □

Proposition 1.10. *Let $A \subset \mathbb{R}$ be a nonempty set. Then,*

- *if* $\inf A > -\infty$, *for any $\delta > 0$, there exists $x \in A$ s.t. $x < \inf A + \delta$;*
- $\sup A < \infty$, *for any $\delta > 0$, there exists $x \in A$ s.t. $x > \sup A - \delta$;*

Proof. Suppose that there exists $\delta > 0$ s.t. for any $x \in A$, $x \geq \inf A + \delta$. Then, $\inf A + \delta$ is a lower bound for A which is greater than $\inf A$. This is not possible. Similarly, suppose that there exists $\delta > 0$ s.t. for any $x \in A$ $x \leq \sup A - \delta$. Then, $\sup A - \delta$ is an upper bound for A which is smaller than $\sup A$. Again, this is not possible. □

Remark 1.1. *Clearly $\mathbb{R}$ is not bounded and hence* $\inf \mathbb{R} = -\infty$ *and* $\sup \mathbb{R} = \infty$*. Sometimes, it is convenient to adjoin $\mathbb{R}$ with the objects $-\infty$ and ∞, i.e., consider the set*

$$\mathbb{R} \cup \{-\infty, \infty\}.$$

This set is referred to as the extended real numbers (sometimes also written as $[-\infty, \infty]$). We use the notation $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. It is important to note that by definition, $-\infty, \infty \notin \mathbb{R}$, i.e., these objects are not numbers. However, for any real number $x \in \mathbb{R}$, $x > -\infty$ and $x < \infty$. Therefore, we assume that $-\infty$ and ∞ satisfy the relation $-\infty < \infty$. With that relation between $-\infty$ and ∞, we have that for any $x \in \overline{\mathbb{R}}$, $-\infty \leq x \leq \infty$. By now we have seen that $-\infty$ and ∞ appear in the context of unbounded sets. We will see latter, that $-\infty$ and ∞ also appear in the definition of diverging sequences. For future references, we also write $\overline{\mathbb{R}}_+ = [0, \infty) \cup \{\infty\}$.

Example 1.13. *Let $a, b \in \mathbb{R}$ and assume that for any $\varepsilon > 0$, $a \leq b + \varepsilon$. Then, $a \leq b$. To see it, let $B = \{x + \varepsilon \colon \varepsilon > 0, x \geq b\}$ and $A = \{x \colon x \geq a\}$. We have that* $\inf B = b$ *and* $\inf A = a$*. Clearly, $B \subset A$. Hence, by the latter proposition,* $\inf A = a \leq b = \inf B$*. Similarly, if for any $\varepsilon > 0$, $a \geq b - \varepsilon$, we must have $a \geq b$. To see it, take $B = \{x - \varepsilon \colon \varepsilon > 0, x \leq b\}$ and $A = \{x \colon x \leq a\}$. Then* $\sup B = b$ *and* $\sup A = a$*. Hence, by the latter proposition, since $B \subset A$ we have that $b \leq a$. Notice, that the latter results do not change if we allow for $a = \infty$ or $b = \infty$, i.e., $a, b \in \overline{\mathbb{R}}$ (cf. Remark 1.1). For example, in the case where $a \leq b + \varepsilon$, if $a = \infty$, then $b = \infty$ and hence $a = b$. If $b = \infty$, then either $a = b$ or $a < b$. Finally, we remark that if for any $\varepsilon > 0$, $a < b + \varepsilon$, then, $a \leq b$ as well (resp. $a \geq b$ if for any $\varepsilon > 0$ $a > b - \varepsilon$). This is because the statement $a < b + \varepsilon$ (resp. $a > b - \varepsilon$) implies that $a \leq b + \varepsilon$ (resp. $a \geq b - \varepsilon$).*

Exercise 1.7. *For each of the following sets, identify its infimum and supremum. Deduce whether the sets have a minimum or maximum.*

(a) $A = \{1/n \colon n \in \mathbb{N}\}$;

(b) $B = \mathbb{Q} \cap [0, 2)$;

(c) $C = \mathbb{Z} \cap (-\infty, 0]$.

Up to now $\inf A$ and $\sup A$ were only defined for $A \subset \mathbb{R}$. We extend the notions of infimum and supremum to the extended real numbers as follows:

Definition 1.11.

- *If $A \subset \overline{\mathbb{R}}$ s.t. $A \subset \mathbb{R}$ and A is bounded, then* $\sup A$ *and* $\inf A$ *are defined as in Definitions 1.7 and 1.8, respectively.*

- *If $A \subset \overline{\mathbb{R}}$ s.t. there exists no real number s which is s.t. for any $x \in A$, $x \leq s$, then* $\sup A = \infty$*. In particular, this is the case if $\infty \in A$.*

- *If there exists $s \in \mathbb{R}$ s.t. $x \leq s$ for any $x \in A$, then* $\sup A$ *is defined as follows: If $A \subset \mathbb{R}$, then* $\sup A$ *is defined as in Definition 1.7. Otherwise, $A = \{-\infty\} \cup A^*$, $A^* \subset \mathbb{R}$, and we define*

$$\sup A = \begin{cases} \sup A^*, & \textit{if } A^* \neq \emptyset, \\ -\infty, & \textit{otherwise.} \end{cases}$$

  In particular, with this definition, $\sup A$ *is the smallest upper bound of A.*

- *If $A \subset \overline{\mathbb{R}}$ s.t. there exists no real number s which is s.t. for any $x \in A$, $x \geq s$, then* $\inf A = -\infty$*. In particular, this is the case if $-\infty \in A$.*

- *If there exists $s \in \mathbb{R}$ s.t. $x \geq s$ for any $x \in A$, then $\inf A$ is defined as follows: If $A \subset \mathbb{R}$, then $\inf A$ is defined as in Definition 1.8. Otherwise, $A = A^* \cup \{\infty\}$, $A^* \subset \mathbb{R}$, and we define*

$$\inf A = \begin{cases} \inf A^*, & \text{if } A^* \neq \emptyset, \\ \infty, & \text{otherwise.} \end{cases}$$

In particular, with this definition, $\inf A$ is the greatest lower bound of A.

Proposition 1.11. $\sup \emptyset = -\infty$ *and* $\inf \emptyset = \infty$.

Proof. Suppose that there exists $s \in \overline{\mathbb{R}}$ which is not an upper bound for $\emptyset$. Then, there exists $a \in \emptyset$, s.t. $a > s$. Which is clearly not possible, since $\emptyset$ contains not a single element. Therefore, any $s \in \overline{\mathbb{R}}$ is in fact an upper bound for $\emptyset$. Therefore, $\sup \emptyset = \min(\overline{\mathbb{R}}) = -\infty$, the smallest upper bound of $\emptyset$. Similarly, $\inf \emptyset = \max(\overline{\mathbb{R}}) = \infty$, the largest lower bound of $\emptyset$. □

1.4 Solution to exercises

Solution 1.1 (Solution to Exercise 1.1)**.** *We need to show that $\omega \in A \Rightarrow \omega \in C$. Thus, let $\omega \in A$. Since $A \subset B$, that implies that $\omega \in B$. Further, since $B \subset C$, any member of B is a member of C. In particular, $\omega \in C$. Since ω was an arbitrary element of A, this completes the argument.*

Solution 1.2 (Solution to Exercise 1.2)**.** *Let $\omega \in A \cap B$. By Definition 1.1, this means that $\omega \in A$ and $\omega \in B$. In particular, $\omega \in A$. Hence $A \cap B \subset A$. In order to verify the second claim, let us assume that $A \subset B$. We want to show that in this case $A \cap B = A$. Let $\omega \in A$, then since $A \subset B$, $\omega \in B$. In particular, $\omega \in A$ and $\omega \in B$. This shows that $A \subset A \cap B$. We already knwo that it is generally true that $A \cap B \subset A$. Thus, by (1.2) of Proposition 1.1, $A \subset B \Rightarrow A \cap B = A$.*

Solution 1.3 (Solution to Exercise 1.3)**.** *By Definition 1.1, $A \cup B$ must contain all the elements from B. To see this, suppose by contradiction that there exists $\omega \in B$ s.t. $\omega \notin A \cup B$. $\omega \notin A \cup B$ means ω is an element that is not in A and also not in B (otherwise it would be in one or the other). This is not true since $\omega \in B$. Hence $B \subset A \cup B$. Let us verify that if $A \subset B \Rightarrow A \cup B = B$. We have shown that in general $B \subset A \cup B$. Thus, it remains to show that if $A \subset B$, $A \cup B \subset B$. That is clear, since if $\omega \in A \cup B$, then $\omega \in A$ or $\omega \in B$. If $\omega \in B$, we are done. If $\omega \in A$, we know that $\omega \in B$ as well, since $A \subset B$ was assumed.*

Solution 1.4 (Solution to Exercise 1.4)**.** *We show that $A \cup B \subset A \cup (B \setminus A)$ and $A \cup (B \setminus A) \subset A \cup B$. Since $B \setminus A \subset B \subset A \cup B$ and $A \subset A \cup B$, it is clear that $A \cup (B \setminus A) \subset A \cup B$. Let $\omega \in A \cup B$, then $\omega \in A$ or $\omega \in B$. If $\omega \in A$, since $A \subset A \cup (B \setminus A)$, $\omega \in A \cup (B \setminus A)$. If $\omega \in B$, then either $\omega \notin A$, hence $\omega \in B \setminus A \subset A \cup (B \setminus A)$. Or $\omega \in A \cap B \subset A$ and hence $\omega \in A \cup (B \setminus A)$ as well.*

Solution 1.5 (Solution to Exercise 1.5)**.** *Yes, take $A = \emptyset$ and B any arbitrary set.*

Solution 1.6 (Solution to Exercise 1.6)**.** *We proof the claim be induction. The base step is clear. In order to verify the induction step, we choose $n \in \mathbb{N}$, and assume that*

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

We have that

$$\begin{aligned}
\sum_{k=1}^{n+1} k^2 &= \frac{n(n+1)(2n+1)}{6} + (n+1)^2 \\
&= \frac{n(n+1)(2n+1) + 6(n+1)^2}{6} \\
&= \frac{(n+1)(n(2n+1) + 6(n+1))}{6} \\
&= \frac{(n+1)(2n^2 + 7n + 6)}{6} \\
&= \frac{(n+1)(2n^2 + 2n + 5n + 6)}{6} \\
&= \frac{(n+1)((n+2)(2n+n) - n^2 - 4n + 5n + 6)}{6} \\
&= \frac{(n+1)((n+2)(2(n+1) + 1 + n - 3) - n^2 + n + 6)}{6} \\
&= \frac{(n+1)((n+2)(2(n+1)+1) + (n+2)(n-3) - n^2 + n + 6)}{6} \\
&= \frac{(n+1)((n+2)(2(n+1)+1) + n^2 - 3n + 2n - 6 - n^2 + n + 6)}{6} \\
&= \frac{(n+1)(n+2)(2(n+1)+1)}{6}.
\end{aligned}$$

Solution 1.7 (Solution to Exercise 1.7)**.**

(a) We have that $x \leq 1$ for any $x \in A$. Thus, 1 is an upper bound for A. Further, $\sup A = 1$, i.e., 1 is the smallest upper bound of A. This is because $1 \in A$ and hence $s' < 1$ can not be a smaller upper bound of A. In particular, as $1 \in A$, the maximum of A is 1. It is true that $x \geq 0$ for any $x \in A$. Thus, 0 is a lower bound for A. Further, 0 is the largest lower bound of A. This is because for any $s' > 0$ there exists $n \in \mathbb{N}$ s.t. $1/n < s'$ (Proposition 1.7). Thus, it can not be the case that there exists a larger lower bound than 0. This shows that $\inf A = 0$. Notice that $0 \notin A$, hence A has no minimum.

(b) First, $0 = \inf B$ and in particular, $0 \in B$. Hence 0 is the minimum of B. Further, for any $x \in B$, $x \leq 2$. That is, 2 is an upper bound for B. By Proposition 1.6, 2 must be the smallest upper bound of B. Hence $\sup B = 2$. We note that since $2 \notin B$, B does not have a maximum.

(c) Clearly, there does not exists $s \in \mathbb{R}$ s.t. $x \geq s$ for any $x \in C$. This shows that $\inf B = -\infty$. In particular, C does not have a minimum. On the other hand, $x \leq 0$ for any $x \in C$. Thus, because $0 \in C$, $\sup C = 0$ and in particular, 0 is the maximum of C.

1.5 Additional exercises

Exercise 1.8. *Let I be an arbitrary set and $\{A_i : i \in I\}$ be a collection of subsets of a set Ω. Let $A \subset \Omega$. Show that*

$$A \subset A_i \;\forall i \in I \Leftrightarrow A \subset \bigcap_{i \in I} A_i.$$

Exercise 1.9. *Prove the following items of Proposition 1.4:*

- $A \setminus B = A \cap B^c$;
- $(A \subset B) \Rightarrow (B^c \subset A^c)$;
- $(A \cap B)^c = A^c \cup B^c$;
- $(A \cup B)^c = A^c \cap B^c$.

Exercise 1.10. *Let $n \in \mathbb{N}$ and A_i, $i = 1, \ldots, n$, be a collection of subsets of a set Ω. Show that*

$$\left(\bigcup_{i=1}^{n} A_i \right)^c = \bigcap_{i=1}^{n} A_i^c.$$

Exercise 1.11. *Let A be a set with n elements. Show that*

- *the number of permutations of the elements from A is $n!$;*
- *for any $0 \leq k \leq n$, the number of subsets of A having k elements is given by*

$$\frac{n!}{(n-k)!k!}.$$

Note 1: *If $A = \{\omega_1, \ldots, \omega_n\}$, then the vector $(\omega_{n_1}, \ldots, \omega_{n_n})$, $n_1, \ldots, n_n \in \{1, \ldots, n\}$, $n_i \neq n_j$, $i \neq j$, represents a permutation of the elements from A.*
Note 2: *The number $n!/((n-k)!k!)$ is denoted with $\binom{n}{k}$.*

Exercise 1.12. *Verify that*

$$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}.$$

2 Introduction: Part II

2.1 Functions

Definition 2.1. *Let A and B be two sets. A function $f\colon A \to B$ is a rule which assigns to each element $a \in A$ exactly one element $f(a) \in B$.*

Definition 2.2. *Let $f\colon A \to B$ be a function.*

Image: *The image of f under $C \subset A$ is the set*

$$f(C) = \{f(a)\colon a \in C\}. \tag{3}$$

Preimage: *If $D \subset B$, the preimage of f under D is the set*

$$f^{-1}(D) = \{a \in A\colon f(a) \in D\}.$$

Example 2.1. *Let $f(x) = x^2$, $x \in \mathbb{R}$. We have that*

$$f(\mathbb{R}) = \{y\colon y \in [0, \infty)\}.$$

To see it, it is clear that $f(\mathbb{R}) \subset \{y\colon y \in [0, \infty)\}$. For the other inclusion, take $y \in \{y\colon y \in [0, \infty)\}$. If we set $x = \sqrt{y}$, then $x^2 = y$. Thus, $y \in f(\mathbb{R})$. Thus also $\{y\colon y \in [0, \infty)\} \subset f(\mathbb{R})$.

Example 2.2. *A function $f\colon A \to \mathbb{R}^k$ assigns to each element $a \in A$, a vector $f(a) \in \mathbb{R}^k$. We use the notation $f(a) = (f_1(a), \ldots, f_k(a))$ for the value of f at a, where $f_i(a)$, $i = 1, \ldots, k$, are referred to as the coordinate functions of f.*

To indicate that an element a is assigned to an element $f(a)$ we often use the notation $a \mapsto f(a)$ to define the assignment, i.e., the function.

Definition 2.3. *Let $f\colon A \to B$ be a function.*

Surjective: *f is called surjective, if $f(A) = B$.*

Injective: *f is called injective, if $a_1 \neq a_2 \Rightarrow f(a_1) \neq f(a_2)$.*

Bijective: *f is called bijective if it is surjective and injective.*

Example 2.3. *Let $f\colon \mathbb{R} \to [0, \infty)$, $f(x) = x^2$. We already know that f is surjective. However, it is not injective, since for $x_1 = -1$ and $x_2 = 1$, $f(x_1) = f(x_2) = 1$. This shows that $x \mapsto x^2$ is not bijective.*

Definition 2.4. *Let $f\colon A \to B$ be a function and $E \subset A$. The restriction of f to E is the function $f|_E\colon E \to B$ given by $f|_E(a) = f(a)$ for any $a \in E$.*

Example 2.4. *The function $f\colon \mathbb{R} \to [0, \infty)$, $f(x) = x^2$ is not bijective. However, its restriction to $[0, \infty)$ is.*

Definition 2.5. *Let $f\colon A \to B$ and $g\colon B \to C$ be two functions. The composition $g \circ f$ or $g(f)$ is defined pointwise, i.e.,*

$$(g \circ f)(a) = g(f)(a) = g(f(a)).$$

Thus, $g \circ f\colon A \to C$.

Definition 2.6. *Let $f\colon A \to B$ be a function. A function $g\colon B \to A$ is called an inverse of f if*

$$g(f(a)) = a \;\forall\, a \in A \quad \text{and} \quad f(g(b)) = b \;\forall\, b \in B.$$

If f has an inverse it is called invertible.

Proposition 2.1. *Let $f\colon A \to B$ be a function. If f is bijective, then it is invertible.*

Proof. Since f is surjective, $f(A) = B$. In particular, for any $b \in B$, there exists $a \in A$, s.t. $f(a) = b$. Thus, we can assign to each $b \in B$, a respective element $a \in A$ via the assignment $b \mapsto g(b) = a$, $b = f(a)$. Clearly, g assigns elements from B to elements from A. In order to show that the latter assignment is a function, the assignment rule must be unique in the sense that if $b = f(a) \in f(A)$ then g assigns b to exactly one element $g(b)$ of A. Suppose that this assignment rule is not unique, i.e., there exists $a^* \in A$, $a^* \neq a$ s.t. $g(b) = a^*$, $b = f(a)$ (two outputs for the same input). This can only happen if $f(a) = f(a^*)$, since if $f(a) \neq f(a^*)$, $g(f(a^*)) = a^* \neq a = g(f(a)) = g(b)$, i.e., $g(b) \neq a^*$. But the case $f(a) = f(a^*)$ for $a^* \neq a$ is not possible since f is surjective. Hence, it can not happen that there exists $a^* \in A$, $a^* \neq a$ s.t. $g(b) = a^*$. □

Example 2.5. *Let $f(x) = \mathrm{e}^x$, $x \in \mathbb{R}$, i.e., f is the (natural) exponential function* ($\mathrm{e} = \mathrm{e}^1$ *is Euler's number). One can show that $f\colon \mathbb{R} \to (0, \infty)$ is bijective where the inverse is given by the natural logarithm* $\log(y)\colon (0, \infty) \to \mathbb{R}$.

Exercise 2.1. *Let $f(x) = 1 - e^{-x}$, $x \in [0, \infty)$. Is $f\colon [0, \infty) \to [0, 1)$ invertible? If yes, what is its inverse?*

Example 2.6. *The trigonometric functions* $\sin, \cos\colon \mathbb{R} \to \mathbb{R}$ *are clearly not bijective. However,* $\sin|_{[-\pi/2,\pi/2]}\colon [-\pi/2, \pi/2] \to [0, 1]$ *and* $\cos|_{[0,\pi]}\colon [0, \pi] \to [-1, 1]$ *are bijective with inverse* $\arcsin\colon [-1, 1] \to [-\pi/2, \pi/2]$ *and* $\arccos\colon [-1, 1] \to [0, \pi]$*. In general,* $\sin$ *and* $\cos$ *satisfy the following addition formulas: Given any $\theta_1, \theta_2 \in \mathbb{R}$,*

$$\begin{aligned}
\sin(\theta_1 + \theta_2) &= \sin(\theta_1)\cos(\theta_2) + \sin(\theta_2)\cos(\theta_1) \\
\sin(\theta_1 - \theta_2) &= \sin(\theta_1)\cos(\theta_2) - \sin(\theta_2)\cos(\theta_1) \\
\cos(\theta_1 + \theta_2) &= \cos(\theta_1)\cos(\theta_2) - \sin(\theta_2)\sin(\theta_1) \\
\cos(\theta_1 - \theta_2) &= \cos(\theta_1)\cos(\theta_2) + \sin(\theta_2)\sin(\theta_1).
\end{aligned}$$

As a further example, the cotangent $\cot(\theta) = \cos(\theta)/\sin(\theta)$*, $\theta \in (0, \pi)$ is s.t.* $\cot\colon (0, \pi) \to \mathbb{R}$ *is bijective with inverse* $\operatorname{arccot}\colon \mathbb{R} \to (0, \pi)$.

Example 2.7. *Let*

$$U = \{(\rho, \theta) \in \mathbb{R}^2 \colon \rho > 0,\ 0 < \theta < 2\pi\} = (0, \infty) \times (0, 2\pi).$$

Define the function

$$T(\rho, \theta) = (\rho\cos(\theta), \rho\sin(\theta)), \quad (\rho, \theta) \in U.$$

Set $V = \mathbb{R}^2 \setminus ([0, \infty) \times \{0\})$, i.e., V is $\mathbb{R}^2$ with the ray $[0, \infty) \times \{0\}$ removed. We notice that $T(U) \subset V$, since if $(x, y) \in T(U)$, then $x = \rho\cos(\theta)$ and $y = \rho\sin(\theta)$ for some $\rho > 0$ and $\theta \in (0, 2\pi)$. Thus, if $(x, y) \in V^c$, then, by definition of V, $\sin(\theta) = 0$ and hence, $\theta = \pi$. This is not possible as it implies that $x = -\rho$ and then $(x, y) = (-\rho, 0) \in V$. We verify that $T\colon U \to V$ is bijective, i.e., injective and surjective. To see that T is injective we show that for any two elements $x_1 = (\rho_1, \theta_1)$ and $x_2 = (\rho_2, \theta_2)$ in U, $T(x_1) = T(x_2)$ implies that $x_1 = x_2$. We note that $T(x_1) = T(x_2)$ implies that

$$\rho_1\cos(\theta_1) - \rho_2\cos(\theta_2) = 0 \tag{4}$$
$$\rho_1\sin(\theta_1) - \rho_2\sin(\theta_2) = 0. \tag{5}$$

Assume by contradiction that $x_1 \neq x_2$. If $\rho_1 \neq \rho_2$ but $\theta_1 = \theta_2$, we use (5) and conclude that $\sin(\theta_1) = \sin(\theta_2) = 0$. Since $\theta_1, \theta_2 \in (0, 2\pi)$, this implies that $\theta_1 = \theta_2 = \pi$. Upon (4), we get that $-(\rho_1 - \rho_2) = 0$, which is a contradiction. For the remaining case assume that $\theta_1 \neq \theta_2$ (and either $\rho_1 = \rho_2$ or $\rho_1 \neq \rho_2$). Notice first that it is not possible that

$\sin(\theta_1) = \sin(\theta_2) = 0$, *since this would imply that* $\theta_1 = \theta_2 = \pi$. *Further, it is not possible that* $\cos(\theta_1) = \cos(\theta_2) = 0$, *since then either* $\theta_1 = \pi/2$ *and* $\theta_2 = (3\pi)/2$ *or vice versa. This implies that either* $\rho_1 + \rho_2 = 0$ *or* $-(\rho_1 + \rho_2) = 0$. *Thus, as a first case, we assume that* $\sin(\theta_1) \neq 0$ *and* $\cos(\theta_2) \neq 0$. *Then, by* (4) *and* (5),

$$\rho_2 = \rho_1 \frac{\cos(\theta_1)}{\cos(\theta_2)} \quad \text{and} \quad \rho_1 = \rho_2 \frac{\sin(\theta_2)}{\sin(\theta_1)}.$$

The latter display is equivalent to

$$\sin(\theta_1)\cos(\theta_2) = \sin(\theta_2)\cos(\theta_1).$$

Upon the identity $\sin(\theta_1 - \theta_2) = \sin(\theta_1)\cos(\theta_2) - \sin(\theta_2)\cos(\theta_1)$, *we conclude that* $\sin(\theta_1 - \theta_2) = 0$. *Since* $\theta_1 \neq \theta_2$, *this implies that* $\theta_1 = \theta_2 + \pi$. *Since* $\sin(\theta_2 + \pi) = -\sin(\theta_2)$, *we deduce from* (5) *that*

$$-\rho_1 \sin(\theta_2) - \rho_2 \sin(\theta_2) = -\sin(\theta_2)(\rho_1 + \rho_2) = 0.$$

Thus, $\theta_2 = \pi$. *This is not possible as it implies that* $\theta_1 = 2\pi$ *and we have assumed that* $\theta_1 < 2\pi$. *The remaining cases are similar and we thus conclude that* T *is indeed injective. In order to verify that* T *is also surjective we need to show that* $T(U) = V$. *We already know that* $T(U) \subset V$, *hence it remains to show the opposite inclusion. Let* $(x, y) \in V$. *First, it is not possible that* $x = y = 0$, *since* $y = 0$ *implies that* $x < 0$. *Assume first that* $y \neq 0$. *We define* $\rho = \sqrt{x^2 + y^2}$ *and* $\theta = \operatorname{arccot}(x/y)$ *(cf. Example 2.6). With this choice,* $(\rho, \theta) \in U$ *and* $\rho\cos(\theta) = x$ *and* $\rho\sin(\theta) = y$. *If* $y = 0$, *then again we define* $\rho = \sqrt{x^2 + y^2} = |x| > 0$ *and* $\theta = \pi$. *Again,* $(\rho, \theta) \in U$ *and* $\rho\cos(\theta) = x$ *and* $\rho\sin(\theta) = y$.

Definition 2.7. *Let* $f\colon I \to \mathbb{R}$, $I \subset \mathbb{R}$ *be a function.* f *is called increasing (resp. decreasing) if* $x_1 \leq x_2 \Rightarrow f(x_1) \leq f(x_2)$ *(resp.* $x_1 \leq x_2 \Rightarrow f(x_2) \leq f(x_1)$*).* f *is strictly increasing (resp. strictly decreasing) if* $x_1 < x_2 \Rightarrow f(x_1) < f(x_2)$ *(resp.* $x_1 < x_2 \Rightarrow f(x_2) < f(x_1)$*).* f *is called monotonic (resp. strictly monotonic) if it is either increasing or decreasing (resp. strictly increasing or decreasing).*

Proposition 2.2. *Let* $f\colon I \to \mathbb{R}$, $I \subset \mathbb{R}$ *be a strictly monotonic function. Then,* $f\colon I \to f(I)$ *is a bijection. Further, if* f *is strictly increasing (resp. strictly decreasing) on* I, *then an inverse of* f *is strictly increasing (resp. strictly decreasing) on* $f(I)$.

Example 2.8. *Let* $f\colon [0, \infty) \to [0, \infty)$ *be the function* $f(x) = x^2$. *Then, using the same argument as in Example 2.1, we readily see that* $f([0, \infty)) = [0, \infty)$. *Further, if* $x < y$, *then* $f(x) = x^2 < y^2 = f(y)$. *Thus,* f *is strictly increasing. By Proposition 2.2, we know that* f *has inverse* $f^{-1}(y)$, $y \in [0, \infty)$, *which is also strictly increasing. Actually, in this case, we know that* $f^{-1}(y) = \sqrt{y}$, $y \in [0, \infty)$, *is the unique inverse of* f, *since* $y = x^2$ *has only one non-negative solution.*

Exercise 2.2. *Let* $x, y \geq 0$. *Show that* $x < y \Rightarrow x^2 < y^2$.

In order to verify whether a function f is monotone, the following proposition is helpful.

Proposition 2.3. *Let* $a, b \in \mathbb{R}$ *and* $f\colon (a, b) \to \mathbb{R}$ *be a function that is differentiable on* (a, b). *That is, the derivative* f' *of* f *exists on* (a, b). *Then:*

- $f'(x) \geq 0\ \forall x \in (a, b) \Rightarrow f$ *is increasing on* (a, b);
- $f'(x) > 0\ \forall x \in (a, b) \Rightarrow f$ *is strictly increasing on* (a, b);
- $f'(x) \leq 0\ \forall x \in (a, b) \Rightarrow f$ *is decreasing on* (a, b);
- $f'(x) < 0\ \forall x \in (a, b) \Rightarrow f$ *is strictly decreasing on* (a, b).

Remark 2.1. *Upon the latter proposition, the claim of Exercise 2.2 follows readily.*

Exercise 2.3. *Are the following functions monotone. If yes, are they increasing or decreasing (resp. strictly increasing or decreasing).*

(a) $f\colon (0,\infty) \to (0,\infty)$, $f(x) = 1/x$;

(b) $g\colon (0,\infty) \to \mathbb{R}$, $g(x) = \log(x)$;

(c) $h\colon \mathbb{R} \to [0,\infty)$, $h(x) = x^4$.

Definition 2.8. *Let $f, g\colon A \to \mathbb{R}$ be two real-valued functions.*

Sum: $f + g$ *is the function* $a \mapsto (f+g)(a) = f(a) + g(a)$;

Product: fg *is the function* $a \mapsto (fg)(a) = f(a)g(a)$;

Quotient: *If* $g(a) \neq 0\ \forall a \in A$, *then* f/g *is the function* $a \mapsto (f/g)(a) = f(a)/g(a)$.

Proposition 2.4. *Let $f\colon A \to B$ be a function. Let $B_* \subset B$. Then,*

(a) $f^{-1}(B_*^c) = f^{-1}(B_*)^c$.

Let I and J be some sets and $A_i \subset A$, $i \in I$, and $B_j \subset B$, $j \in J$, be a collection of sets from A and B, respectively. Then,

(b) $f(\cup_{i\in I} A_i) = \cup_{i\in I} f(A_i)$;

(c) $f^{-1}(\cup_{j\in J} B_j) = \cup_{j\in J} f^{-1}(B_j)$;

(d) $f^{-1}(\cap_{j\in J} B_j) = \cap_{j\in J} f^{-1}(B_j)$.

Definition 2.9. *Let $f\colon A \to \overline{\mathbb{R}}$ be a function and $E \subset A$. We use the notation*

$$\sup f(E) = \sup\{f(e)\colon e \in E\} = \sup_{e\in E} f(e),$$

for the supremum of $f(E)$. Similarly, we write

$$\inf f(E) = \inf\{f(e)\colon e \in E\} = \inf_{e\in E} f(e).$$

for the infimum of $f(E)$.

2.2 Cardinality of Sets

Definition 2.10. *Let A be a set. A is said to be finite if it contains a finite number of elements. Otherwise, if the number of elements in A is not finite, A is referred to as infinite.*

Example 2.9. *The sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ are all infinite. There does not exist a number which counts the number of elements within the latter sets.*

Definition 2.11. *Let A and B be two arbitrary sets. A and B are said to have the same cardinality ($\#A = \#B$) if and only if there exists a bijection $f\colon A \to B$.*

In terms of a visual interpretation, if $\#A = \#B$, then we can draw a line from each element of A to an element of B (surjectivity) and it is not possible that a line emerges from two different elements of A to the same element of B.

Example 2.10. *If A and B are finite, then $\#A = \#B$ precisely means that A and B have the same number of elements ($\#A$ is given by the number of connecting lines between A and B). In particular, if we let $A = B$, $\#A$ gives the number of elements in A. This shows that $\#\emptyset = 0$, since $\emptyset$ is clearly finite.*

Definition 2.12. *Let A and B be two sets and $C \subset B$.*

- *If there exists a bijection $g: A \to C$ we write $\#A \leq \#B$;*
- *If there exists a bijection $g: A \to C$ but there exists no bijection $f: A \to B$, we write $\#A < \#B$.*

Example 2.11. *Let A and B be two sets. According to Definition 2.12, if $A \subset B$, $\#A \leq \#B$. This is because the identity map $g: A \to A$, $g(a) = a$, $a \in A$, is a bijection.*

Proposition 2.5. *We have that*

- *$\#\mathbb{N} = \#\mathbb{Z} = \#\mathbb{Q}$;*
- *$\#\mathbb{R} > \#\mathbb{N}$;*
- *any interval $I \subset \mathbb{R}$ is s.t. $\#I > \#\mathbb{N}$.*

Definition 2.13. *Let A be a set. If*

- *$\#A \leq \#\mathbb{N}$, then A is countable;*
- *$\#A > \#\mathbb{N}$, then A is uncountable.*

Proposition 2.6. *Let A be a set. If A is countable but not finite, then $\#A = \#\mathbb{N}$ and A is said to be countably infinite.*

Proposition 2.7. *Let $\{A_i : i \in \mathbb{N}\}$ be a collection of sets s.t. for any $i \in \mathbb{N}$, A_i is countable. Then the union $\cup_{i\in\mathbb{N}} A_i$ is countable as well.*

Proposition 2.8. *Let A_i, $i = 1, \dots, n$, be countable sets. Then, their Cartesian product, $\prod_{i=1}^{n} A_i$ is countable.*

We remark that if $\{A_i : i \in I\}$ is some collection of sets, where I is some set, then if I is countable we might always set $I = \mathbb{N}$ or $I = \{1, \dots, n\}$, $n \in \mathbb{N}$.

To conclude this section we state the following result:

Proposition 2.9. *Let $[a, b]$, $a < b \in \mathbb{R}$, be any closed interval. Let I be some countable set ($\#I \leq N$) and assume that there exists a family of open intervals (a_i, b_i), a_i, b_i, $i \in I$, s.t., $[a, b] \subset \cup_{i\in I}(a_i, b_i)$. Then, there exists $N \in \mathbb{N}$ and $i_1, \dots, i_N \in I$, s.t. $[a, b] \subset \cup_{j=1}^{N}(a_{i_j}, b_{i_j})$.*

The latter result is known as the Heine-Borel theorem for intervals, it shows that any closed interval on the real line which is covered by a countable collection of open intervals can be covered by a finite sub-collection.

2.3 Euclidean distance

Definition 2.14. *Given two points $x, y \in \mathbb{R}^k$, the Euclidean distance between x and y is given by*

$$\|x - y\| = \sqrt{(x_1 - y_1)^2 + \cdots (x_k - y_k)^2}, \quad x = (x_1, \dots, x_k),\ y = (y_1, \dots, y_k).$$

Example 2.12. *If $k = 1$, then we write $\|x - y\| = |x - y|$, where the function $x \mapsto |x|$ maps $x \in \mathbb{R}$ to its absolute value*

$$|x| = \begin{cases} x, & \text{if } x \geq 0, \\ -x, & \text{if } x < 0. \end{cases}$$

Proposition 2.10. *For any $x, y, z \in \mathbb{R}^k$, The Euclidean distance satisfies:*

(i) $\|x - y\| \geq 0$ *and* $\|x - y\| = 0 \Leftrightarrow x = y$;

(ii) $\|x - y\| = \|y - x\|$ *(symmetry);*

(iii) $\|x + y\| \le \|x\| + \|y\|$ *(triangular inequality);*

(iv) $\|x - y\| \ge \big|\|x\| - \|y\|\big|$ *(reverse triangular inequality).*

Notice that we use the notation 0 for the vector in $\mathbb{R}^k$, $k \in \mathbb{N}$, with all of its coordinates equal to zero.

Example 2.13. *In Proposition 1.6 of Example 1.4 we saw that there exists a rational number in between any two distinct real numbers. Let $\varepsilon > 0$ and $x \in \mathbb{R}$. Chose $n \in \mathbb{N}$ s.t. $1/n < \varepsilon$. Then, $y = x + 1/n$ is s.t. $|x - y| = 1/n < \varepsilon$. Still, $x \neq y$ and hence by Proposition 1.6 we find $q \in \mathbb{Q}$ s.t. $x < q < y$. In particular, $0 < q - x < y - x$ and hence $|q - x| < \varepsilon$. Since $\varepsilon > 0$ and $x \in \mathbb{R}$ were arbitrary, we can conclude that for any $\varepsilon > 0$ and $x \in \mathbb{R}$, there exists $q \in \mathbb{Q}$ s.t. $|x - q| < \varepsilon$.*

Definition 2.15. *An open (resp. closed) ball of radius $r > 0$ with center $y \in \mathbb{R}^k$ is denoted with*

$$B_r(y) = \{x \in \mathbb{R}^k \colon \|y - x\| < r\} \text{ (resp. } B_r[y] = \{x \in \mathbb{R}^k \colon \|y - x\| \le r\})$$

Definition 2.16. *A set $U \subset \mathbb{R}^k$ is called open if for any point $x \in U$, there exists $\varepsilon > 0$, s.t. $B_\varepsilon(x) \subset U$. That is any point in U is the center of an open Ball contained in U.*

Example 2.14. *If U_1 and U_2 are two open subsets of $\mathbb{R}^k$, then $U_1 \cap U_2$ is open. By definition, if $x \in U_1 \cap U_2$, there exist $B_{\varepsilon_1}(x)$ and $B_{\varepsilon_2}(x)$ s.t. $B_{\varepsilon_1}(x) \subset U_1$ and $B_{\varepsilon_2}(x) \subset U_2$. Thus, set $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$ and we obtain that $B_\varepsilon(x) \subset U_1 \cap U_2$.*

Example 2.15. *Let $a, b \in \mathbb{R}$. Then, the open interval (a, b) is an open set of $\mathbb{R}$. Let $x \in (a, b)$, i.e., $a < x < b$. Let $\varepsilon < \min\{|x - a|, |x - b|\}$. Then, take any $y \in (x - \varepsilon, x + \varepsilon) = B_\varepsilon(x)$. It follows that $x - y \le |x - y| < |x - a| = x - a$ and hence, $y > a$. Also, $y - x \le |x - y| < |x - b| = b - x$, i.e., $y < b$. Therefore, $y \in (a, b)$.*

Example 2.16. *An open ball $B_r(y) \subset \mathbb{R}^k$ is an open set. This is a consequence of the triangular inequality. Let $x \in B_r(y)$ be an arbitrary point and set $\delta = \|x - y\|$. Then, by definition of $B_r(y)$, $\delta < r$. Let $\varepsilon = r - \delta$. Then, $B_\varepsilon(x) \subset B_r(y)$, since for any $z \in B_\varepsilon(x)$, $\|z - y\| \le \|z - x\| + \|x - y\| = \varepsilon + \delta = r$. Since $x \in B_r(y)$ was arbitrary, the result follows.*

Exercise 2.4. *Show that the open rectangle $\prod_{i=1}^k (a_i, b_i)$, a_i, b_i, $i = 1, \dots, k$, is an open set of $\mathbb{R}^k$.*

We remark that continuous functions on Euclidean spaces are characterized in terms of open sets.

Definition 2.17. *A function $f \colon \mathbb{R}^m \to \mathbb{R}^k$ is continuous if for any open set $U \subset \mathbb{R}^k$, the set $f^{-1}(U)$ is open in $\mathbb{R}^m$.*

Example 2.17. *Let $f(x) = x$, $x \in \mathbb{R}$. Then, $f \colon \mathbb{R} \to \mathbb{R}$ is continuous. Take any $U \subset \mathbb{R}$ open. Then, $f^{-1}(U) = U$. Hence, $f^{-1}(U)$ is an open subset of $\mathbb{R}$.*

A motivation for Definition 2.17 is given in the appendix (Section A.4). In terms of functions defined on subsets of $\mathbb{R}^m$, continuity is characterized as follows (a proof is given in Section A.4):

Proposition 2.11. *Let $E \subset \mathbb{R}^m$, $E \neq \emptyset$. A function $f \colon E \to \mathbb{R}^k$ is continuous if and only if for any open set $U \subset \mathbb{R}^k$,*

$$f^{-1}(U) \in \{G \cap E \colon G \text{ open in } \mathbb{R}^m\}.$$

We recall some classical examples of continuous functions.

Proposition 2.12. *The following functions are continuous:*

- $f\colon \mathbb{R}^k \to \mathbb{R}$, $f(x) = \sum_{i=1}^k x_k$, $x = (x_1, \ldots, x_k)$;
- $g\colon \mathbb{R}^k \to \mathbb{R}$, $g(x) = \prod_{i=1}^k x_k$, $x = (x_1, \ldots, x_k)$;
- $h\colon \mathbb{R} \setminus \{0\} \to \mathbb{R} \setminus \{0\}$, $h(x) = 1/x$.

For more details on the notion of continuity we refer to Section A.4.

Definition 2.18. *A set $V \subset \mathbb{R}^k$ is said to be closed if V^c is open.*

Example 2.18. *Given $r > 0$ and $y \in \mathbb{R}^k$, a closed ball $B_r[y] \subset \mathbb{R}^k$ is closed. To see it, we show that $B_r[y]^c = \mathbb{R}^k \setminus B_r[y]$ is open in $\mathbb{R}^k$. Let $x \in B_r[y]^c$, i.e., $\|x - y\| > r$. Set $\delta = \|x-y\| - r$. Then, $\delta > 0$. Consider the open ball $B_\delta(x)$. If we show that $B_\delta(x) \subset B_r[y]^c$, we are done. Hence, let $z \in B_\delta(x)$. Using the reverse triangular inequality (item (iv) of Proposition 2.10),*

$$\|z - y\| \geq \|x - y\| - \|z - x\| > \|x - y\| - \delta = r,$$

i.e., $z \in B_r[y]^c$ and we are done.

Example 2.19. *The set $V = \{(x, y) \in \mathbb{R}^2 \colon x \geq 0, y = 0\}$ is a closed set of $\mathbb{R}^2$. We show that V^c is open. Let $v = (v_1, v_2) \in V^c$. If $v_2 = 0$, then $v_1 < 0$ otherwise $v \in V$. Hence, if $v_2 = 0$, $v \in V^c$ implies that $B_r(v) \subset V^c$ with $r = |v_1|$. To see it, it is sufficient to verify that $z = (a, b) \in B_r(v)$ implies that $a < 0$ (this implies that $(a, b) \in V^c$). We notice that $z \in B_r(v)$ implies that*

$$(a - v_1)^2 + b^2 < v_1^2 \Leftrightarrow a^2 + b^2 < 2av_1.$$

Hence, by the previous inequality, since $v_1 < 0$ it can not be the case that $a \geq 0$. The remaining case is that $v_2 \neq 0$. Set $r = |v_2|$. We verify that $B_r(v) \subset V^c$. Let $z = (a, b) \in B_r(v)$. We verify that $b \neq 0$ (this shows that $(a, b) \in V^c$). We note that $z \in B_r(v)$ implies that $(a - v_1)^2 + (b - v_2)^2 < v_2^2$. But then, $b = 0$ is not possible, as it would imply that $(a - v_1)^2 < 0$. Hence, the set V^c is open and thus V is closed.

Definition 2.19. *A set $A \subset \mathbb{R}^k$ is said to be bounded if there exists $r > 0$ and $y \in A$ s.t. $A \subset B_r[y]$, i.e., A is contained in a closed ball.*

Definition 2.20. *Let $f\colon A \to \overline{\mathbb{R}}$ be a function and $E \subset A$, where A is some set. The function f is said to be bounded on E if there exists $0 \leq M < \infty$, s.t. $|f(a)| \leq M$ for any $a \in E$.*

2.4 Solution to exercises

Solution 2.1 (Solution to Exercise 2.1)**.** *Yes, f is invertible. The inverse of f is the function $g\colon [0, 1) \to [0, \infty)$, $g(y) = -\log(1 - y)$.*

Solution 2.2 (Solution to Exercise 2.2)**.** *We show first that if $a \in \mathbb{R}$ s.t. $a > 1$, then, for any $z \in (0, \infty)$, $az > z$. Write $a = 1 + \varepsilon$ with $\varepsilon = a - 1 > 0$. Then, $az = z + \varepsilon z$. Since $\varepsilon, z > 0$, $az > z$. Since $x < y$, $0 < (x - y)^2 = x^2 + y^2 - 2xy$. Then, since $x < y$, $2xy \geq 2x^2$. Notice that this is because $x \geq 0$. If $x = 0$, then $2xy = 0 = 2x^2$. Otherwise, if $x > 0$, then, since $x < y$, $y/x > 1$. Hence, $2xy = 2x^2x^{-1}y > 2x^2$. In particular, $0 < x^2 + y^2 - 2xy < x^2 + y^2 - 2x^2 = y^2 - x^2$. This solves the exercises.*

Solution 2.3 (Solution to Exercise 2.3)**.**

(a) *Given any* $0 < a < b < \infty$, $f'(x) = -1/x^2$, $x \in (a, b)$. *Thus,* $f'(x) < 0$ *for any* $x \in (a, b) \subset (0, \infty)$. *Since* a *and* b *were arbitrary, we easily verify that* f *must be strictly decreasing on* $(0, \infty)$. *If not, there exists* $x_1, x_2 \in (0, \infty)$ *with* $x_1 < x_2$ *and* $f(x_2) > f(x_1)$. *This is not possible, since we always find* $a < x_1 < x_2 < b$, *and upon Proposition 2.3, since* f *has strictly negative derivative on* (a, b), f *is strictly decreasing on* (a, b), *in particular* $f(x_2) < f(x_1)$.

(b) *Given* $a, b \in \mathbb{R}$, *we calculate* $g'(x) = 1/x$, $x \in (0, \infty)$. *Then, using the same reasoning as in (a), we conclude that* g *is strictly increasing.*

(c) h *is not monotone. To see it, we notice that* $h(-1) \geq h(x)$ *for any* $x \in [-1, 1]$ *but* $h(x) > h(-1)$ *for any* $x \in (1, 2)$.

Solution 2.4 (Solution to Exercise 2.4)**.** *We need to show that for any* $x \in \prod_{i=1}^k (a_i, b_i)$, *there exists* $\varepsilon > 0$, *s.t.* $B_\varepsilon(x) \subset \prod_{i=1}^k (a_i, b_i)$ *(cf. Definition 2.16). Let* $x = (x_1, \ldots, x_k) \in \prod_{i=1}^k (a_i, b_i)$ *and set* $m_i = \min\{|x_i - a_i|, |x_i - b_i|\}$, $i = 1, \ldots, k$. *Then, since for any* $i = 1, \ldots, k$, $m_i > 0$, *choose* $\varepsilon < \min\{m_i : i = 1, \ldots, k\}$. *Then, let* $y = (y_1, \ldots, y_k) \in B_\varepsilon(x)$. *For any* $i = 1, \ldots, k$, *we have that*

$$|y_i - x_i| \leq \|y - x\| < \varepsilon < m_i.$$

In particular, $x_i - y_i \leq |y_i - x_i| < |x_i - a_i| = x_i - a_i$ *and therefore,* $y_i > a_i$. *Also,* $y_i - x_i \leq |y_i - x_i| < |x_i - b_i| = b_i - x_i$ *and hence* $y_i < b_i$. *This shows that* $y_i \in (a_i, b_i)$ *for any* $i = 1, \ldots, k$.

2.5 Additional exercises

Exercise 2.5. *Show that for any* $x, y \in \mathbb{R}$ *and* $n \in \mathbb{N}$, $(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$.

Exercise 2.6. *Prove Proposition 2.4.*

Exercise 2.7. *Let* $f : \mathbb{N} \to \mathbb{Z}$ *be defined as follows:*

$$f(n) = \begin{cases} \frac{n}{2}, & \text{if } n \text{ is even,} \\ \frac{1-n}{2}, & \text{if } n \text{ is odd.} \end{cases}$$

Show that f *is bijective.*
Note: *This shows that* $\#\mathbb{N} = \#\mathbb{Z}$.

Exercise 2.8. *Let* A *and* B *be two sets. Show that if* A *and* B *are countable, then,* $A \cup B$ *is countable.*

Exercise 2.9. *Show that,*

(a) *if* $f : A \to B$ *and* $g : B \to C$ *are two functions s.t.* f *and* g *are bijective, then, the composition* $g \circ f : A \to C$ *is bijective as well;*

(b) $\#\mathbb{R} = \#(0, 1)$.

3 Introduction: Part III

3.1 Real valued sequences

Definition 3.1. *A real-valued sequence is a function* $f\colon \mathbb{N} \to \mathbb{R}$*, i.e.,* $f(n) \in \mathbb{R}$ *for any* $n \in \mathbb{N}$*. We use the notation* $f = (a_n)_{n\in\mathbb{N}}$ *for a real-valued sequence and* $f(n) = a_n$ *for the values of* f *at* n*.*

This section only treats real-valued sequences. Thus, for now, a sequence is a real-valued sequence.

Definition 3.2. *Let* $(a_n)_{n\in\mathbb{N}}$ *be a sequence.* $(a_n)_{n\in\mathbb{N}}$ *is said to be convergent if there exists a number* $a \in \mathbb{R}$ *s.t. for any* $\varepsilon > 0$ *there exists* $N \in \mathbb{N}$ *s.t.* $|a_n - a| < \varepsilon$ *for any* $n \geq N$*.*

The number a in Definition 3.2 is called the limit of $(a_n)_{n\in\mathbb{N}}$.

Example 3.1. *Let* $a_n = 1/n$*,* $n \in \mathbb{N}$*. We show that* $(a_n)_{n\in\mathbb{N}}$ *is convergent with limit* 0*. Let* $\varepsilon > 0$ *be an arbitrary strictly positive real number. According to Proposition 1.7, we pick* $N \in \mathbb{N}$ *s.t.* $N > 1/\varepsilon$*. Then, for any* $n \geq N$*,*

$$|a_n - 0| = \frac{1}{n} \leq \frac{1}{N} < \varepsilon.$$

Exercise 3.1. *If* $a_n = c$ *for any* $n \in \mathbb{N}$ *then* $(a_n)_{n\in\mathbb{N}}$ *is convergent with limit* c

Exercise 3.2. *Let* $a_n = (-1)^n$*,* $n \in \mathbb{N}$*. Is* $(a_n)_{n\in\mathbb{N}}$ *convergent? Try to only use Definition 3.2.*

Proposition 3.1. *If* $(a_n)_{n\in\mathbb{N}}$ *is convergent, then its limit* a *is unique and we write*

$$\lim_{n\to\infty} a_n = a \qquad \text{or} \qquad a_n \xrightarrow{n\to\infty} a.$$

In the following, we list some important results on real valued sequences.

Definition 3.3. *A sequence* $(a_n)_{n\in\mathbb{N}}$ *is said to be bounded if there exists* $M > 0$ *s.t. for any* $n \in \mathbb{N}$*,* $|a_n| \leq M$*.* $(a_n)_{n\in\mathbb{N}}$ *is said to be bounded from below (resp. above) if there exists* $M \in \mathbb{R}$ *s.t.* $a_n \geq M$ *(resp.* $a_n \leq M$*) for any* $n \in \mathbb{N}$*.*

Proposition 3.2. *If* $(a_n)_{n\in\mathbb{N}}$ *is convergent, then it is bounded.*

Definition 3.4. *Let* $(a_n)_{n\in\mathbb{N}}$ *be a sequence.* $(a_n)_{n\in\mathbb{N}}$ *is increasing (resp. decreasing) if* $a_n \leq a_{n+1}\ \forall\, n \in \mathbb{N}$ *(resp.* $a_n \geq a_{n+1}\ \forall\, n \in \mathbb{N}$*).* $(a_n)_{n\in\mathbb{N}}$ *is said to be monotonic if it is either increasing or decreasing.*

Definition 3.5. *If* $(a_n)_{n\in\mathbb{N}}$ *is increasing (resp. decreasing) with limit* a*, we write* $a_n \uparrow a$ *(resp.* $a_n \downarrow a$*).*

Proposition 3.3. *A bounded and monotonic sequence* $(a_n)_{n\in\mathbb{N}}$ *is convergent.*

Example 3.2. *Let* $|r| < 1$*, and consider* $a_n = |r|^n$*,* $n \in \mathbb{N}$*. For any* $n \in \mathbb{N}$*,* $a_n < 1$ *and*

$$|r|^{n+1} = |r|^n|r| \leq |r|^n.$$

Thus, $(a_n)_{n\in\mathbb{N}}$ *is bounded and decreasing. By Proposition 3.3, there exists* L *s.t.,*

$$\lim_{n\to\infty} a_n = L.$$

Proposition 3.4. *Let* $(a_n)_{n\in\mathbb{N}}$ *and* $(b_n)_{n\in\mathbb{N}}$ *be two convergent sequences s.t.* $a_n \xrightarrow{n\to\infty} a$ *and* $b_n \xrightarrow{n\to\infty} b$*. Then,* *(i)* $a_n + b_n \xrightarrow{n\to\infty} a + b$;

(ii) $a_n b_n \xrightarrow{n\to\infty} ab$;

(iii) $a_n/b_n \xrightarrow{n\to\infty} a/b$, *if* $b \neq 0$.

Proposition 3.5. *Let* $(a_n)_{n\in\mathbb{N}}$ *and* $(b_n)_{n\in\mathbb{N}}$ *be two convergent sequences s.t.* $a_n \xrightarrow{n\to\infty} a$ *and* $b_n \xrightarrow{n\to\infty} b$. *Assume that* $a_n \leq b_n$ *(resp.* $a_n \geq b_n$*) for any* $n \in \mathbb{N}$*, then* $a \leq b$ *(resp.* $a \geq b$*).*

Proposition 3.6. *Let* $(a_n)_{n\in\mathbb{N}}$ *and* $(b_n)_{n\in\mathbb{N}}$ *be two convergent sequences that converge to the same limit, i.e.,* $a_n \xrightarrow{n\to\infty} a$ *and* $b_n \xrightarrow{n\to\infty} a$. *Let* $(c_n)_{n\in\mathbb{N}}$ *be another sequence which is s.t. for any* $n \in \mathbb{N}$, $a_n \leq c_n \leq b_n$. *Then,* $c_n \xrightarrow{n\to\infty} a$.

Example 3.3. *Let* $|r| < 1$*, and consider* $a_n = r^n$, $n \in \mathbb{N}$. *We show that* $(a_n)_{n\in\mathbb{N}}$ *is convergent with limit* 0. *To do so, we consider* $(b_n)_{n\in\mathbb{N}}$*, with* $b_n = |r|^n$, $n \in \mathbb{N}$ *as in Example 3.2. We have that*

$$r^n = \begin{cases} r^n, & r \in [0,1), \\ (-1)^n |r|^n, & r \in (-1,0). \end{cases}$$

In particular, for any $n \in \mathbb{N}$,

$$-|r|^n \leq r^n \leq |r|^n.$$

Thus, by Propositions 3.4 and 3.6, it is sufficient to show that $(b_n)_{n\in\mathbb{N}}$ *converges to zero. We already know (see Example 3.2) that there exists* L *s.t.,*

$$\lim_{n\to\infty} b_n = L.$$

Now we have that

$$L = \lim_{n\to\infty} b_n = \lim_{n\to\infty} b_{n+1} = |r| \lim_{n\to\infty} b_n = |r|L.$$

Thus, since $r \notin \{-1,1\}$*, it must be the case that* $L = 0$. *In conclusion,* $\lim_{n\to\infty} a_n = 0$.

Example 3.4. *In Example 2.13 we have seen that for any real number* x*, the Euclidean distance between* x *and elements from* $\mathbb{Q}$ *can be made arbitrary small. Let us show that for any* $x \in \mathbb{R}$*, there exists a sequence of rational numbers* $(q_n)_{n\in\mathbb{N}}$, $q_n \in \mathbb{Q}$, $n \in \mathbb{N}$*, s.t.* $q_n \uparrow x$. *By Proposition 1.6, for any* $\varepsilon > 0$*, there exists* $q \in \mathbb{Q}$*, s.t.* $x - \varepsilon < q \leq x$. *For* $n = 1$*, choose* $q_1 \in (x-1, x] \setminus (x - 1/2, x]$. *For* $n = 2$*, choose* $q_2 \in (x - 1/2, x] \setminus (x - 1/3, x]$ *and so on until we choose* q_n *in*

$$\left(x - \frac{1}{n}, x\right] \setminus \left(x - \frac{1}{n+1}, x\right].$$

Then, $q_n < q_{n+1}$ *and for any* $n \in \mathbb{N}$ *and*

$$x - \frac{1}{n} < q_n \leq x.$$

Using Proposition 3.6, this shows that $q_n \uparrow x$. *A similar argument shows there exists a sequence* $(q_n^*)_{n\in\mathbb{N}}$, $q_n^* \in \mathbb{Q}$, $n \in \mathbb{N}$*, s.t.* $q_n^* \downarrow x$.

Exercise 3.3. *Let*

$$a_n = \frac{n^2 + 3n^3 + n}{1 + n^4}, \quad n \in \mathbb{N}.$$

Is $(a_n)_{n\in\mathbb{N}}$ *convergent? If yes, what is its limit?*

Exercise 3.4. *Let $a_n = n!/n^n$, $n \in \mathbb{N}$. Is $(a_n)_{n\in\mathbb{N}}$ convergent? If yes, what is its limit?*

Definition 3.6. *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence. We write:*

(i) $a_n \xrightarrow{n\to\infty} \infty$ *(or $\lim_{n\to\infty} a_n = \infty$) if $\forall\, M \in \mathbb{R}\ \exists\, N \in \mathbb{N}$ s.t. $a_n \geq M\ \forall\, n \geq N$.*

(ii) $a_n \xrightarrow{n\to\infty} -\infty$ *(or $\lim_{n\to\infty} a_n = -\infty$) if $\forall\, M \in \mathbb{R}\ \exists\, N \in \mathbb{N}$ s.t. $a_n \leq M\ \forall\, n \geq N$.*

If $\lim_{n\to\infty} a_n = \infty$ (resp. $\lim_{n\to\infty} a_n = -\infty$) we say that $(a_n)_{n\in\mathbb{N}}$ diverges to ∞ (resp. $-\infty$). We say that $(a_n)_{n\in\mathbb{N}}$ diverges if it either diverges to ∞ or $-\infty$.

Remark 3.1. *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence. For now, if $\lim_{n\to\infty} a_n$ is well defined, i.e., $\lim_{n\to\infty} a_n \in \mathbb{R}$ (i.e., $(a_n)_{n\in\mathbb{N}}$ converges) or $(a_n)_{n\in\mathbb{N}}$ diverges, we understand $\lim_{n\to\infty} a_n$ as an element of the extended real line $\overline{\mathbb{R}}$ (cf. Remark 1.1). We write that $\lim_{n\to\infty} a_n$ exists, if $\lim_{n\to\infty} a_n \in \overline{\mathbb{R}}$. Further, if $\lim_{n\to\infty} a_n$ exists, then it is unique.*

Proposition 3.7. *Let $(a_n)_{n\in\mathbb{N}}$ be a monotone sequence. Then, $\lim_{n\to\infty} a_n$ exists. If $(a_n)_{n\in\mathbb{N}}$ is increasing and $(a_n)_{n\in\mathbb{N}}$ diverges, then it diverges to ∞. If $(a_n)_{n\in\mathbb{N}}$ is decreasing and $(a_n)_{n\in\mathbb{N}}$ diverges, then it diverges to $-\infty$.*

We notice that for increasing (resp. decreasing) sequences, Proposition 3.5 remains true even if the sequences do not converge.

Proposition 3.8. *Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be two increasing (resp. decreasing) sequences. Then, if for any $n \in \mathbb{N}$, $a_n \leq b_n$, $\lim_{n\to\infty} a_n \leq \lim_{n\to\infty} b_n$.*

Exercise 3.5. *Prove Proposition 3.8.*

Definition 3.7. *Let $(a_i)_{i\in\mathbb{N}}$ be a sequence. The series*

$$\sum_{i\in\mathbb{N}} a_i = \sum_{i=1}^{\infty} a_i,$$

is understood as the sequence $(s_n)_{n\in\mathbb{N}}$, where $s_n = \sum_{i=1}^{n} a_i$, $n \in \mathbb{N}$. If $\lim_{n\to\infty} s_n$ exists we write $\lim_{n\to\infty} s_n = \sum_{i=1}^{\infty} a_i$ for the limit.

Proposition 3.9. *Let $\sum_{i\in\mathbb{N}} a_i$ be a series where $a_i \geq 0$ for any $i \in \mathbb{N}$. Then, either $\sum_{i\in\mathbb{N}} a_i < \infty$ or $\sum_{i\in\mathbb{N}} a_i = \infty$.*

Proof. This follows from Proposition 3.7. Notice that since $a_i \geq 0$ for any $i \in \mathbb{N}$, the sequence $(s_n)_{n\in\mathbb{N}}$, $s_n = \sum_{i=1}^{n} a_i$, is increasing. $\square$

Example 3.5. *We have that*

$$\sum_{i\in\mathbb{N}} \frac{1}{i} = \infty,$$

i.e., the sequence $(s_n)_{n\in\mathbb{N}}$, $s_n = \sum_{i=1}^{n}(1/i)$, diverges to ∞. We will give an argument in the next section.

Example 3.6. *We have that*

$$\sum_{i\in\mathbb{N}} \frac{1}{i(i+1)} = 1.$$

We notice that

$$\frac{1}{i} - \frac{1}{i+1} = \frac{1}{i(i+1)}.$$

Then, for any $n \in \mathbb{N}$,

$$\begin{aligned} s_n &= \sum_{i=1}^{n} \frac{1}{i(i+1)} = \sum_{i=1}^{n} \left(\frac{1}{i} - \frac{1}{i+1} \right) \\ &= 1 - \frac{1}{1+1} + \frac{1}{2} - \frac{1}{2+1} + \frac{1}{3} - \frac{1}{3+1} + \cdots - \frac{1}{n-1+1} + \frac{1}{n} - \frac{1}{n+1} \\ &= 1 - \frac{1}{n+1}. \end{aligned}$$

Therefore, since $\lim_{n\to\infty} s_n = 1$, we have that $\sum_{i\in\mathbb{N}} 1/(i(i+1)) = 1$.

Example 3.7. *The functions $x \mapsto \mathrm{e}^x$, $x \mapsto \sin(x)$ and $x \mapsto \cos(x)$ are all defined in terms of a series:*

- $\mathrm{e}^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$, $x \in \mathbb{R}$;
- $\sin(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1}$, $x \in \mathbb{R}$;
- $\cos(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} x^{2k}$, $x \in \mathbb{R}$.

To conclude this section, we list two useful results for series.

Proposition 3.10. *Let $\sum_{i\in\mathbb{N}} a_i$ be a series and $\sum_{i\in\mathbb{N}} b_i$ be a series s.t. $b_i \geq 0$ for any $i \in \mathbb{N}$ and $\sum_{i\in\mathbb{N}} b_i < \infty$. Suppose that $|a_i| \leq b_i$ for any $i \in \mathbb{N}$. Then $\sum_{i\in\mathbb{N}} a_i < \infty$.*

Proposition 3.11. *Let $I, J \subset \mathbb{N}$ and $f \colon I \times J \to \mathbb{R}$. For any $i, j \in \mathbb{N}$, set $a_{ij} = f(i,j)$. Thus, we obtain a doubly indexed sequence of real numbers $(a_{ij})_{(i,j)\in I\times J}$. Suppose that either $a_{ij} \geq 0$ for any $(i,j) \in I \times J$ or $\sum_{(i,j)\in I\times J} |a_{ij}| < \infty$. Then, $\sum_{(i,j)\in I\times J} a_{ij}$ is well defined and*

$$\sum_{(i,j)\in I\times J} a_{ij} = \sum_{i\in I} \left(\sum_{j\in J} a_{ij} \right) = \sum_{j\in J} \left(\sum_{i\in I} a_{ij} \right). \tag{6}$$

That is, we are allowed to change the order of summation. If $I = J$, we use the notation $\sum_{(i,j)\in I^2} a_{ij} = \sum_{i,j\in I} a_{ij}$ for the sum over all the pairs $(i,j) \in I^2$.

3.2 Subsequences: Limit inferior and limit superior

We remain in the setting of the previous Section, i.e., any sequence $(a_n)_{n\in\mathbb{N}}$ is a real-valued sequence according to Definition 3.1.

Definition 3.8. *Let $f = (a_n)_{n\in\mathbb{N}}$ be a sequence (cf. Definition 3.1). A subsequence of $(a_n)_{n\in\mathbb{N}}$ is a new sequence $g = (b_n)_{n\in\mathbb{N}}$, where $g = f \circ s$, with $s \colon \mathbb{N} \to \mathbb{N}$ s.t. $s(n) < s(n+1)$, i.e., for any $k \in \mathbb{N}$, $b_n = g(n) = f(s(n)) = a_{s(n)}$.*

Example 3.8. *Let $a_n = 1/n$, $n \in \mathbb{N}$. Then, $(a_{2n})_{n\in\mathbb{N}}$, is a subsequence of $(a_n)_{n\in\mathbb{N}}$.*

The following result is known as the Bolzano–Weierstrass theorem.

Proposition 3.12. *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence. If $(a_n)_{n\in\mathbb{N}}$ is bounded, then there exists a subsequence of $(a_n)_{n\in\mathbb{N}}$ which is convergent.*

Example 3.9. *Let $a_n = (-1)^n$, $n \in \mathbb{N}$. We have seen that $(a_n)_{n\in\mathbb{N}}$ is not convergent. However, $(a_{2n})_{n\in\mathbb{N}}$ is a subsequence of $(a_n)_{n\in\mathbb{N}}$ with limit 1.*

An application of Proposition 3.12 is the following result (a proof is given in Section A.4).

Proposition 3.13. *Let $f \colon [a,b] \to \mathbb{R}$ be continuous, then there exists $x_M, x_m \in [a,b]$ s.t. $f(x_M) = \sup_{x\in[a,b]} f(x) = \max_{x\in[a,b]} f(x)$ and $f(x_m) = \inf_{x\in[a,b]} f(x) = \min_{x\in[a,b]} f(x)$, i.e., f attains its maximum and minimum in $[a,b]$. In particular, f is bounded.*

Definition 3.9. *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence and $(a_{s(n)})_{n\in\mathbb{N}}$ be a subsequence of $(a_n)_{n\in\mathbb{N}}$ s.t. $\lim_{n\to\infty} a_{s(n)} = a$. Then, a is said to be an accumulation point of $(a_n)_{n\in\mathbb{N}}$.*

Example 3.10. *Let $a_n = (-1)^n$, $n \in \mathbb{N}$. Then, $(a_n)_{n\in\mathbb{N}}$ has two accumulation points, -1 and 1.*

Proposition 3.14. *Let a be an accumulation point of $(a_n)_{n\in\mathbb{N}}$. Then, for any $\varepsilon > 0$, there are infinitely a_n s.t. $a_n \in (a-\varepsilon, a+\varepsilon)$.*

Proof. Since $a = \lim_{n\to\infty} a_{s(n)}$, it follows that for any $\varepsilon > 0$, there exists $N \in \mathbb{N}$, s.t. for any $n \geq N$, $|a_{s(n)} - a| < \varepsilon \Rightarrow a_{s(n)} \in (a-\varepsilon, a+\varepsilon)$. □

Proposition 3.15. *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence. If $(a_n)_{n\in\mathbb{N}}$ is convergent with limit a, then every subsequence of $(a_n)_{n\in\mathbb{N}}$ converges to a. That is, a convergent sequence has only one accumulation point.*

Proof. Clearly a is an accumulation point of $(a_n)_{n\in\mathbb{N}}$. Suppose by contradiction that $b \neq a$ is another accumulation point of $(a_n)_{n\in\mathbb{N}}$. Set $\delta = (a-b)/2$. Then $\delta > 0$ and hence, there exists $N \in \mathbb{N}$ s.t. $|a_n - a| < \delta$ for any $n \geq N$. Then,

$$|a_n - b| = |a_n - a + a - b| \geq \big||a_n - a| - |a-b|\big|,$$

by the reverse triangular inequality (cf. Proposition 2.10). Therefore, for any $n \geq N$,

$$|a_n - b| \geq \big||a_n - a| - 2\delta\big| = 2\delta - |a_n - a| > \delta.$$

This shows that there exists $\varepsilon > 0$ ($\varepsilon = \delta$) s.t. $|a_n - b| > \varepsilon$ for any $n \geq N$. Thus, for that particular ε, only at most finitely many a_n, are s.t. $a_n \in (b-\varepsilon, b+\varepsilon)$. This contradicts Proposition 3.14. Hence, b is not an accumulation point of $(a_n)_{n\in\mathbb{N}}$. □

Proposition 3.16. *Let $(a_n)_{n\in\mathbb{N}}$ be an increasing (resp. decreasing) sequence. Suppose that there exists a subsequence $(a_{s(n)})_{n\in\mathbb{N}}$ which is s.t. $\lim_{n\to\infty} a_{s(n)} = \infty$ (resp. $\lim_{n\to\infty} a_{s(n)} = -\infty$). Then, $\lim_{n\to\infty} a_n = \infty$ (resp. $\lim_{n\to\infty} a_n = -\infty$).*

Proof. Using Proposition 3.7, this follows directly from Proposition 3.15. If $\lim_{n\to\infty} a_n \neq \infty$ (resp. $\lim_{n\to\infty} a_n \neq -\infty$) it means that $(a_n)_{n\in\mathbb{N}}$ converges (cf. Proposition 3.7). In particular, there exists $a < \infty$ s.t. any accumulation point of $(a_n)_{n\in\mathbb{N}}$ is equal to a (cf. Proposition 3.15). This gives a contradiction with the assumption that $\lim_{n\to\infty} a_{s(n)} = \infty$ (resp. $\lim_{n\to\infty} a_{s(n)} = -\infty$). □

Example 3.11. *We show that $\sum_{i\in\mathbb{N}}(1/i) = \infty$ (cf. Example 3.5). Set, $s_n = \sum_{i=1}^{n}(1/i)$, $n \in \mathbb{N}$. Then, $(s_n)_{n\in\mathbb{N}}$ is increasing. Define the subsequence $(s_{(2^n-1)})_{n\in\mathbb{N}}$. Let $n \in \mathbb{N}$ and $k \in \{2,\ldots,n\}$. We have that*

$$s_{(2^n-1)} = 1 + \sum_{k=2}^{n}\left(\sum_{i=2^{k-1}}^{2^k-1}\frac{1}{i}\right).$$

This is because, $\{2,\ldots,2^n-1\} = \cup_{k=2}^{n}\{2^{k-1},\ldots,2^k-1\}$. We can use induction to see it. If $n=2$, then $\{2,3\} = \{2^{2-1}, 2^2-1\}$. Suppose that $\{2,\ldots,2^n-1\} = \cup_{k=2}^{n}\{2^{k-1},\ldots,2^k-1\}$. Then,

$$\begin{aligned}\{2,\ldots,2^{n+1}-1\} &= \{2,\ldots,2^n-1\} \cup \{2^n,\ldots,2^{n+1}-1\}\\ &= \cup_{k=2}^{n}\{2^{k-1},\ldots,2^k-1\} \cup \{2^n,\ldots,2^{n+1}-1\}\\ &= \cup_{k=2}^{n+1}\{2^{k-1},\ldots,2^k-1\}.\end{aligned}$$

For any $n \in \mathbb{N}$, the cardinality of $\{2^{k-1},\ldots,2^k-1\}$ is $2^k - 1 - (2^{k-1}-1) = 2^k - 2^{k-1} = 2^{k-1}(2-1) = 2^{k-1}$. Hence, for any $n \in \mathbb{N}$, $\sum_{i=2^{k-1}}^{2^k-1}(1/i) \geq 2^{k-1}(1/2^k - 1) \geq 1/2$. Thus, for any $n \in \mathbb{N}$, $s_{(2^n-1)} \geq 1 + (n-1)/2$. Therefore, $(s_{(2^n-1)})_{n\in\mathbb{N}}$ can not be convergent and we conclude that $\lim_{n\to\infty} s_{(2^n-1)} = \infty$. Using Proposition 3.16, this shows that $\lim_{n\to\infty} s_n = \infty$ as well.

Proposition 3.17. *Suppose that* $(a_n)_{n\in\mathbb{N}}$ *is not bounded from below (resp. bounded from above). Then,* $\inf_{n\in\mathbb{N}} a_n = -\infty$ *(resp.* $\sup_{n\in\mathbb{N}} a_n = \infty$*).*

Exercise 3.6. *Prove Proposition 3.17.*

Exercise 3.7. *Let* $(a_n)_{n\in\mathbb{N}}$ *be a sequence bounded from below. Define the sequence*

$$m_n = \inf\{a_k \colon k \geq n\} = \inf_{k\geq n} a_k, \quad n \in \mathbb{N}.$$

Show that $(m_n)_{n\in\mathbb{N}}$ *is increasing.*

Exercise 3.8. *Let* $(a_n)_{n\in\mathbb{N}}$ *be a sequence bounded from above. Define the sequence*

$$M_n = \sup\{a_k \colon k \geq n\} = \sup_{k\geq n} a_k, \quad n \in \mathbb{N}.$$

Show that $(M_n)_{n\in\mathbb{N}}$ *is decreasing.*

Proposition 3.18. *Let* $(a_n)_{n\in\mathbb{N}}$ *and* $(m_n)_{n\in\mathbb{N}}$ *be as in Exercise 3.7. Then,*

$$\lim_{n\longrightarrow\infty} m_n = \sup_{n\in\mathbb{N}} m_n = \sup_{n\in\mathbb{N}} \inf_{k\geq n} a_k.$$

Exercise 3.9. *Prove Proposition 3.18.*

Similarly, we have the following result.

Proposition 3.19. *Let* $(a_n)_{n\in\mathbb{N}}$ *and* $(M_n)_{n\in\mathbb{N}}$ *be as in Exercise 3.8. Then,*

$$\lim_{n\longrightarrow\infty} M_n = \inf_{n\in\mathbb{N}} M_n = \inf_{n\in\mathbb{N}} \sup_{k\geq n} a_k.$$

Example 3.12. *Let* $a_n = (-1)^n$*,* $n \in \mathbb{N}$*, then* $m_n = -1$ *and* $M_n = 1$ *for any* $n \in \mathbb{N}$*. In particular,* $\lim_{n\to\infty} m_n = -1$ *and* $\lim_{n\to\infty} M_n = 1$*.*

We are in place to make the following definition:

Definition 3.10. *Let* $(a_n)_{n\in\mathbb{N}}$ *be a sequence. We define:*

$$\liminf_{n\longrightarrow\infty} a_n = \begin{cases} \sup_{n\in\mathbb{N}} \inf_{k\geq n} a_k, & \text{if } (a_n)_{n\in\mathbb{N}} \text{ is bounded or bounded from below,} \\ -\infty, & \text{otherwise,} \end{cases}$$

and

$$\limsup_{n\longrightarrow\infty} a_n = \begin{cases} \inf_{n\in\mathbb{N}} \sup_{k\geq n} a_k, & \text{if } (a_n)_{n\in\mathbb{N}} \text{ is bounded or bounded from above,} \\ \infty, & \text{otherwise.} \end{cases}$$

$\liminf_{n\longrightarrow\infty} a_n$ *and* $\limsup_{n\longrightarrow\infty} a_n$ *are referred to as limit inferior and limit superior of* $(a_n)_{n\in\mathbb{N}}$*.*

Exercise 3.10. *Suppose that* $\lim_{n\to\infty} a_n = -\infty$ *(resp.* $\lim_{n\to\infty} a_n = \infty$*), then,*

$$\liminf_{n\longrightarrow\infty} a_n = -\infty = \limsup_{n\longrightarrow\infty} a_n \ \text{(resp. } \liminf_{n\longrightarrow\infty} a_n = \infty = \limsup_{n\longrightarrow\infty} a_n\text{)}.$$

Proposition 3.20. *Let* $(a_n)_{n\in\mathbb{N}}$ *be a sequence. We have that*

$$\liminf_{n\longrightarrow\infty} a_n \leq \limsup_{n\longrightarrow\infty} a_n.$$

Proof. Given any $n \in \mathbb{N}$, $\inf_{k\geq n} a_k \leq \sup_{k\geq n} a_k$. Assume that $(a_n)_{n\in\mathbb{N}}$ is bounded. Then, $(\inf_{k\geq n} a_k)_{n\in\mathbb{N}}$ and $(\sup_{k\geq n} a_k)_{n\in\mathbb{N}}$ converge (cf. Proposition 3.3). We use Propositions 3.18 and 3.19 and obtain,

$$\liminf_{n\to\infty} a_n = \lim_{n\to\infty} \inf_{k\geq n} a_k \leq \lim_{n\to\infty} \sup_{k\geq n} a_k = \inf_{n\in\mathbb{N}} \sup_{k\geq n} a_k = \limsup_{n\to\infty} a_n.$$

Clearly, by Definition 3.10, $\liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n$ in all the other cases. □

The following result gives another characterization of convergence:

Proposition 3.21. *Let $(a_n)_{n\in\mathbb{N}}$ be a bounded sequence. Then, $(a_n)_{n\in\mathbb{N}}$ is convergent with limit a if and only if*

$$\liminf_{n\to\infty} a_n = a = \limsup_{n\to\infty} a_n.$$

We use the following result to proof the latter proposition.

Proposition 3.22. *Assume that $(a_n)_{n\in\mathbb{N}}$ is a bounded sequence and define the set*

$$A = \{a\colon a \text{ is an accumulation point of } (a_n)_{n\in\mathbb{N}}\}.$$

Then, $\min A = \liminf_{n\to\infty} a_n$ *and* $\max A = \limsup_{n\to\infty} a_n$.

Proof of Proposition 3.21. Assume that $(a_n)_{n\in\mathbb{N}}$ is convergent with limit a. It follows by Proposition 3.15 that the set of accumulation points of $(a_n)_{n\in\mathbb{N}}$ is given by $\{a\}$. Using Proposition 3.22, this shows that $\liminf_{n\to\infty} a_n = a = \limsup_{n\to\infty} a_n$. For the other direction, assume that $(a_n)_{n\in\mathbb{N}}$ is s.t. $\liminf_{n\to\infty} a_n = a = \limsup_{n\to\infty} a_n$. Then, we have that for any $n \in \mathbb{N}$,

$$m_n = \inf_{k\geq n} a_k \leq a_n \leq \sup_{k\geq n} a_k = M_n$$

Therefore, using Propositions 3.6, 3.18 and 3.19 we conclude that $(a_n)_{n\in\mathbb{N}}$ has limit a. □

A more general statement is the following.

Proposition 3.23. $\lim_{n\to\infty} a_n$ *exists if and only if*

$$\liminf_{n\to\infty} a_n = \lim_{n\to\infty} a_n = \limsup_{n\to\infty} a_n.$$

Proof. Using Proposition 3.21 and Exercise 3.10 it remains to show that if

$$\liminf_{n\to\infty} a_n = -\infty = \limsup_{n\to\infty} a_n \ (\text{or } \liminf_{n\to\infty} a_n = \infty = \limsup_{n\to\infty} a_n),$$

then $\lim_{n\to\infty} a_n$ exists. Actually, if

$$\liminf_{n\to\infty} a_n = -\infty = \limsup_{n\to\infty} a_n,$$

then $\lim_{n\to\infty} a_n = -\infty$ as well. To see it we first notice that $(a_n)_{n\in\mathbb{N}}$ must be bounded from above, otherwise, $\limsup_{n\to\infty} a_n = \infty$, by definition. Hence, $(M_n)_{n\in\mathbb{N}}$ with $M_n = \sup_{k\geq n} a_k$ is decreasing (cf. Exercise 3.8). Also since $\liminf_{n\to\infty} a_n = -\infty$, $(a_n)_{n\in\mathbb{N}}$ is not bounded from below. Therefore, for any $M \in \mathbb{R}$, there exists $N \in \mathbb{N}$, s.t. for any $n \geq N$, $M_n \leq M$. Further, for any $n \in \mathbb{N}$, $a_n \leq \sup_{k\geq n} a_k$. In conclusion, we have shown that for any $M \in \mathbb{R}$, there exists $N \in \mathbb{N}$, s.t. $a_n \leq M$ for any $n \geq N$. This shows that $(a_n)_{n\in\mathbb{N}}$ diverges to $-\infty$ (cf. Definition 3.6). A similar argument shows that if

$$\liminf_{n\to\infty} a_n = \infty = \limsup_{n\to\infty} a_n,$$

then $\lim_{n\to\infty} a_n = \infty$. □

Example 3.13. *Let $a_n = (-1)^n$, $n \in \mathbb{N}$, then, $\liminf_{n\to\infty} a_n = -1$ and $\limsup_{n\to\infty} a_n = 1$ (cf. Example 3.12). Hence, using Proposition 3.21, $(a_n)_{n\in\mathbb{N}}$ can not be convergent (cf. Exercise 3.2).*

Exercise 3.11. *Let $a_n = \cos(n\pi)$, $n \in \mathbb{N}$. Find $\liminf_{n\to\infty} a_n$ and $\limsup_{n\to\infty} a_n$.*

We note that the following proposition gives another useful characterization of convergence in terms of subsequences — it states that if for an arbitrary subsequence of a real-valued sequence one can extract a subsequence that converges to some real number, then the original sequence converges with limit given by the aforementioned number:

Proposition 3.24. *Let $(a_n)_{n\in\mathbb{N}}$ be a real-valued sequence. Consider the follwong assumption:*

(A) for any subsequence $(a_{s(n)})_{n\in\mathbb{N}}$ of $(a_n)_{n\in\mathbb{N}}$ there exists a subsequence $(a_{t(s(n))})_{n\in\mathbb{N}}$ of $(a_{s(n)})_{n\in\mathbb{N}}$ s.t. $a_{t(s(n))} \xrightarrow{n\to\infty} a$.

Then, if (A) holds, $a_n \xrightarrow{n\to\infty} a$.

The latter is known as the subsequence criterion for convergent sequences – for a proof we refer to Section A.3 of the appendix.

3.3 Vector-valued sequences

The previous section on real-valued sequences can easily be extended to the notion of vector-valued sequences.

Definition 3.11. *An $\mathbb{R}^k$-valued sequence is a function $f\colon \mathbb{N} \to \mathbb{R}^k$, where we write*

$$f(n) = (f_1(n), \dots, f_k(n)) = (a_1^n, \dots, a_k^n), \quad n \in \mathbb{N},$$

We use the notation $f = (a_n)_{n\in\mathbb{N}}$ for a $\mathbb{R}^k$-valued sequence.

We notice that the coordinate functions of an $\mathbb{R}^k$-valued sequence $(a_n)_{n\in\mathbb{N}}$ are real-valued sequences (cf. Definition 3.1). The space of all $\mathbb{R}^k$-valued sequences is denoted with $(\mathbb{R}^k)^{\mathbb{N}}$, $k \in \mathbb{N}$. If $k = 1$, then $(a_n)_{n\in\mathbb{N}}$ is a real-valued sequence. In particular, Definition 3.11 contains Definition 3.1. Upon the Euclidean metric $\|x - y\| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_k - y_k)^2}$, $x = (x_1, \dots, x_k)$, $y = (y_1, \dots, y_k)$, we can introduce the notion of convergence for $\mathbb{R}^k$-valued sequence.

Definition 3.12. *Let $(a_n)_{n\in\mathbb{N}} \in (\mathbb{R}^k)^{\mathbb{N}}$. $(a_n)_{n\in\mathbb{N}}$ is said to be convergent if there exists a number $a = (a_1, \dots, a_k) \in \mathbb{R}^k$ s.t. for any $\varepsilon > 0$ there exists $N \in \mathbb{N}$ s.t. $\|a_n - a\| < \varepsilon$ for any $n \geq N$.*

If $(a_n)_{n\in\mathbb{N}} \in (\mathbb{R}^k)^{\mathbb{N}}$, we write $a_n \xrightarrow{n\to\infty} a$ (resp. $\lim_{n\to\infty} a_n = a$) to indicate that $(a_n)_{n\in\mathbb{N}}$ converges to a. The following result shows that in order to proof that an $\mathbb{R}^k$-valued sequence converges, it is enough to study the convergence of the individual coordinates.

Proposition 3.25. *Let $(a_n)_{n\in\mathbb{N}} \in (\mathbb{R}^k)^{\mathbb{N}}$. Then,*

$$(a_1^n, \dots, a_k^n) = a_n \xrightarrow{n\to\infty} a = (a_1, \dots, a_k) \ \Leftrightarrow\ a_i^n \xrightarrow{n\to\infty} a_i \ \forall\, i = 1, \dots, k.$$

The following result is called the sequence criterion for continuous functions (a proof is given in the appendix, Section A.4).

Proposition 3.26. *Let $f\colon E \to \mathbb{R}^k$, $E \subset \mathbb{R}^m$ and $x \in E$. Then, the following two statements are equivalent:*

(i) f is continuous at x;

(ii) $\forall\,(x_n)_{n\in\mathbb{N}} \subset E$ with $\lim_{n\to\infty} x_n = x$ it follows that $\lim_{n\to\infty} f(x_n) = f(x)$.

In the following, we will use the term sequence for $(a_n)_{n\in\mathbb{N}} \in (\mathbb{R}^k)^{\mathbb{N}}$, $k \in \mathbb{N}$, and it will be clear from the context whether $k = 1$ or $k > 1$.

3.4 Sequences of Functions

Definition 3.13. *In general, a sequence of functions, taking values in the extended real numbers, defined on some common set A, is a collection of functions $g_n\colon A \to \overline{\mathbb{R}}$, $n \in \mathbb{N}$. Then, given $E \subset \mathbb{N}$, the quantities $\inf_{n\in E} g_n$ and $\sup_{n\in E} g_n$ are defined pointwise on A, i.e.,*

$$(\inf_{n\in E} g_n)(x) = \inf_{n\in E} g_n(x) \text{ and } (\sup_{n\in E} g_n)(x) = \sup_{n\in E} g_n(x), \quad x \in A.$$

Hence, for any $x \in A$, $(\inf_{n\in E} g_n)(x), (\sup_{n\in E} g_n)(x) \in \overline{\mathbb{R}}$. Additionally, if for any $x \in A$, $\lim_{n\to\infty} g_n(x) \in \overline{\mathbb{R}}$, then also $\lim_{n\to\infty} g_n$ is defined pointwise, i.e.,

$$(\lim_{n\to\infty} g_n)(x) = \lim_{n\to\infty} g_n(x), \quad x \in A.$$

We say that the sequence of functions g_n, $n \in \mathbb{N}$, converges pointwise to a function $g\colon A \to \overline{\mathbb{R}}$, if for any $x \in A$, $\lim_{n\to\infty} g_n(x) = g(x)$.

Example 3.14. *Given $x \in [0, \pi]$, let $g_n(x) = \cos(nx)$, $n \in \mathbb{N}$. Then, g_n, $n \in \mathbb{N}$, is a sequence of functions. We have seen that $(\liminf_{n\to\infty} g_n)(\pi) = -1$ and $(\limsup_{n\to\infty} g_n)(\pi) = 1$. Therefore, using Proposition 3.21 it is not true that $x \mapsto \cos(nx)$, $n \in \mathbb{N}$, converges pointwise on $[0, \pi]$. Notice that for any $x \in [0, 2\pi]$, $(\liminf_{n\to\infty} g_n)(x) = \sup_{n\in\mathbb{N}}(\inf_{k\geq n} g_k)(x)$.*

3.5 Solution to exercises

Solution 3.1 (Solution to Exercise 3.1)**.** *Let $\varepsilon > 0$ and $a = c$. Then, for any $n \in \mathbb{N}$, $|a_n - c| = 0 < \varepsilon$. Thus, $(a_n)_{n\in\mathbb{N}}$ is convergent with limit c.*

Solution 3.2 (Solution to Exercise 3.2)**.** *The sequence $(a_n)_{n\in\mathbb{N}}$ is not convergent. To see it, assume by contradiction that $(a_n)_{n\in\mathbb{N}}$ is convergent with limit a. Suppose that $a \notin \{-1, 1\}$. Then, since $a \neq 1$ and $a \neq -1$, it follows that $|1-a| > 0$. Thus, if $\varepsilon < |1-a|$, for any $N \in \mathbb{N}$, $|a_n - a| = |1 - a| > \varepsilon$ for infinitely many n. Hence, it must be the case that $a \in \{-1, 1\}$. Again, this not possible, since if for example $a = 1$, we have that $|a_n - a| = 2$ for infinitely many n. A similar argument shows that $a = -1$ is also not possible. Hence, there is no chance that $(a_n)_{n\in\mathbb{N}}$ can be convergent.*

Solution 3.3 (Solution to Exercise 3.3)**.** *Yes, $(a_n)_{n\in\mathbb{N}}$ is convergent with limit 0. To see it, we write*

$$\frac{n^2 + 3n^3 + n}{1 + n^4} = \frac{\frac{n^2+3n^3+n}{n^4}}{\frac{1+n^4}{n^4}} = \frac{\frac{1}{n^2} + \frac{3}{n} + \frac{1}{n^3}}{\frac{1}{n^4} + 1}.$$

We then remark that for any $k \in \mathbb{N}$,

$$0 \leq \frac{1}{n^k} \leq \frac{1}{n}.$$

Thus, using Propositions 3.6 and 3.4,

$$\lim_{n\to\infty} a_n = \frac{0}{1} = 0.$$

Solution 3.4 (Solution to Exercise 3.4)**.** *Since*

$$0 \leq \frac{n!}{n^n} \leq \frac{1}{n},$$

it follows from Proposition 3.6 that $\lim_{n\to\infty} a_n = 0$.

Solution 3.5 (Solution to Exercise 3.5)**.** *Let us first consider the case where $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ are both increasing. If both sequences converge, this is Proposition 3.5. The case where $(b_n)_{n\in\mathbb{N}}$ converges but $(a_n)_{n\in\mathbb{N}}$ does not is not possible by Proposition 3.7 and Proposition 3.2. If $(b_n)_{n\in\mathbb{N}}$ does not converge, it diverges to ∞ and we are left with two cases, either $(a_n)_{n\in\mathbb{N}}$ converges or it does not. In both cases it is clearly true that $\lim_{n\to\infty} a_n \leq \lim_{n\to\infty} b_n$, where we have equality when both series diverge (cf. Remark 3.1). The other case, i.e., when $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ are both decreasing is proved similarly. Here, it is not possible that $(a_n)_{n\in\mathbb{N}}$ converges but $(b_n)_{n\in\mathbb{N}}$ does not (cf. Propositions 3.7 and 3.2). If both sequences diverge, we have equality in the limit and if $(b_n)_{n\in\mathbb{N}}$ converges but $(a_n)_{n\in\mathbb{N}}$ diverges, we have strict inequality in the limit, i.e., $a < b$.*

Solution 3.6 (Solution to Exercise 3.6)**.** *Suppose by contradiction that $\inf\{a_n : n \in \mathbb{N}\} > -\infty$, i.e., there exists a real number $M > 0$ s.t. $\inf\{a_n : n \in \mathbb{N}\} > -M$. Since $(a_n)_{n\in\mathbb{N}}$ is not bounded from below, then, for any integer $N \in \mathbb{N}$, there exists $n(N) \in \mathbb{N}$ s.t. $a_{n(N)} < -N$ (cf. Definition 3.3). Define the subsequence $a_{n(N)}$, $N \in \mathbb{N}$, where for any $N \in \mathbb{N}$, $a_{n(N)} < -N$. Clearly, $\{a_{n(N)} : N \in \mathbb{N}\} \subset \{a_n : n \in \mathbb{N}\}$, hence, by Proposition 1.9, for any integer $k \in \mathbb{N}$,*

$$-M < \inf\{a_n : n \in \mathbb{N}\} \leq \inf\{a_{n(N)} : n \in \mathbb{N}\} \leq a_{n(k)} < -k.$$

Hence, we have deduced that there exists a real number $M > 0$ s.t. for any integer $k \in \mathbb{N}$, $-M < -k$. This gives a contradiction with Proposition 1.7. Hence, $\inf\{a_n : n \in \mathbb{N}\} = -\infty$. A similar argument shows that if $(a_n)_{n\in\mathbb{N}}$ is not bounded from above, then $\sup\{a_n : n \in \mathbb{N}\} = \infty$.

Solution 3.7 (Solution to Exercise 3.7)**.** *Let $n \in \mathbb{N}$. Write, $\{a_k : k \geq n\} = \{a_n\} \cup \{a_k : k \geq n+1\}$. Hence, $\{a_k : k \geq n+1\} \subset \{a_k : k \geq n\}$. It follows that $m_n \leq m_{n+1}$ (cf. Proposition 1.9).*

Solution 3.8 (Solution to Exercise 3.8)**.** *Since $\{a_k : k \geq n+1\} \subset \{a_k : k \geq n\}$, it follows that $M_{n+1} \leq M_n$ (cf. Proposition 1.9).*

Solution 3.9 (Solution to Exercise 3.9)**.** *If $S = \sup_{n\in\mathbb{N}} m_n < \infty$, define $b_n = S$ for any $n \in \mathbb{N}$. Then, $m_n \leq b_n$ for any $n \in \mathbb{N}$ and by Proposition 3.8, $\lim_{n\to\infty} m_n \leq S$. If $S = \infty$, clearly, $\lim_{n\to\infty} m_n \leq S$. For the other inequality, if $S < \infty$, by definition of $\sup_{n\in\mathbb{N}} m_n$ (cf. Proposition 1.10), there exists $n \in \mathbb{N}$, s.t. $m_n \geq S - \varepsilon$ for any $\varepsilon > 0$. Thus we can again use Proposition 3.8 to conclude that $\lim_{n\to\infty} m_n \geq S$ (take for example $b_n = S - (1/n)$). For the final case, suppose by contradiction that $S = \infty$ and $\lim_{n\to\infty} m_n < \infty$. By Proposition 3.2, this means that $(m_n)_{n\in\mathbb{N}}$ is bounded, i.e., there exists an upper bound for $\{m_n : n \in \mathbb{N}\}$. This is not the case, as $S = \sup\{m_n : n \in \mathbb{N}\} = \infty$ (cf. Definition 1.10). Hence, if $S = \infty$, then $\lim_{n\to\infty} m_n = \infty$.*

Solution 3.10 (Solution to Exercise 3.10)**.** *If $\lim_{n\to\infty} a_n = -\infty$, then $(a_n)_{n\in\mathbb{N}}$ can not be bounded from below. Therefore, by Definition 3.10, $\liminf_{n\to\infty} a_n = -\infty$. Further, $(a_n)_{n\in\mathbb{N}}$ must be bounded from above, since there exists $N \in \mathbb{N}$, s.t. $a_n \leq 0$ for any $n \geq N$ (cf. Definition 3.6) and for any $n < N$, we have that $a_n \leq \max\{a_i : i = 1, \ldots, N-1\} = M$. In particular, $a_n \leq \max\{0, M\}$. Therefore, by Definition 3.10, $\limsup_{n\to\infty} a_n = \inf_{n\in\mathbb{N}} \sup_{k\geq n} a_k$. We know that $(M_n)_{n\in\mathbb{N}} = (\sup_{k\geq n} a_k)_{n\in\mathbb{N}}$ is decreasing and $\lim_{n\to\infty} M_n = \limsup_{n\to\infty} a_n$ (cf. Proposition 3.19). Suppose by contradiction that $(M_n)_{n\in\mathbb{N}}$ converges. Then, $(M_n)_{n\in\mathbb{N}}$ is bounded. That is, there exists $L \in \mathbb{R}$ s.t. $M_n > L$ for any $n \in \mathbb{N}$. Since $\lim_{n\to\infty} a_n = -\infty$, let $N \in \mathbb{N}$ s.t. $a_n \leq L$ for any $n \geq N$ (cf. Definition 3.6). Then, $\sup_{k\geq N} a_k = M_N \leq L$. This gives a contradiction, hence, $(M_n)_{n\in\mathbb{N}}$ must diverge and hence by Proposition 3.9, $\lim_{n\to\infty} M_n = \limsup_{n\to\infty} a_n = -\infty$. Similarly, one can show that if $\lim_{n\to\infty} a_n = \infty$, then $\liminf_{n\to\infty} a_n = \infty = \limsup_{n\to\infty} a_n$.*

Solution 3.11 (Solution to Exercise 3.11)**.** *We have that $|a_n| \leq 1$ for any $n \in \mathbb{N}$. Thus, $(a_n)_{n\in\mathbb{N}}$ is bounded. Given any $n \in \mathbb{N}$, $\inf_{k\geq n} a_k = -1$ and hence $\liminf_{n\to\infty} a_n = -1$. Further, for any $n \in \mathbb{N}$, $\sup_{k\geq n} a_k = 1$. Therefore, $\limsup_{n\to\infty} a_n = 1$*

3.6 Additional exercises

Exercise 3.12. *Let* $(a_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ *be a convergent sequence. Show that* $\lim_{n\to\infty} a_n$ *is unique, i.e., if* $\lim_{n\to\infty} a_n = a$ *and* $\lim_{n\to\infty} a_n = b$*, then* $a = b$.

Exercise 3.13. *Let* $(a_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ *be a sequence. Show that:*

(i) If $a_n > 0$ *for any* $n \in \mathbb{N}$*, then if* $a_n \xrightarrow{n\to\infty} 0$ *it follows that* $1/a_n \xrightarrow{n\to\infty} \infty$*;*

(ii) If $a_n < 0$ *for any* $n \in \mathbb{N}$*, then if* $a_n \xrightarrow{n\to\infty} 0$ *it follows that* $1/a_n \xrightarrow{n\to\infty} -\infty$.

Exercise 3.14. *Let* $a \in \mathbb{R}$ *and* $|r| < 1$*. We consider the sequence*

$$s_n = \sum_{k=0}^{n} ar^k = a + ar + ar^2 + \cdots + ar^n, \quad n \in \mathbb{N}.$$

Show that $(s_n)_{n\in\mathbb{N}}$ *is convergent with limit* $a/(1-r)$.
Hint: *Compare* s_n *and* rs_n.

Exercise 3.15. *Let* $(a_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ *be a bounded sequence. Show that*

$$\min\{a\colon a \textit{ is an accumulation point of } (a_n)_{n\in\mathbb{N}}\} = \liminf_{n\to\infty} a_n.$$

Exercise 3.16. *Prove Proposition 3.25.*

4 Measurable sets: Part I

4.1 Measurable spaces

Definition 4.1 (σ-field)**.** *Let Ω be a nonempty set. A family of subsets $\mathcal{F}$ of Ω is called a σ-field on Ω if the following three items are satisfied:*

(i) $\Omega \in \mathcal{F}$;

(ii) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$;

(iii) if $\{A_i \colon i \in \mathbb{N}\}$ is a collection of sets s.t. $A_i \in \mathcal{F}$ for any $i \in \mathbb{N}$, then $\cup_{i\in\mathbb{N}} A_i \in \mathcal{F}$.

Example 4.1. *Let $\Omega \neq \emptyset$ be an arbitrary set. Let*

$$\mathcal{F} = \{\emptyset, \Omega\}.$$

Then, $\mathcal{F}$ is a σ-field on Ω. Clearly, $\Omega \in \mathcal{F}$. Further, let $A \in \mathcal{F}$. Then, there are only two cases, either $A = \emptyset$ or $A = \Omega$. In each case, $A^c \in \mathcal{F}$. Consider a countable collection $\{A_i \colon i \in \mathbb{N}\} \subset \mathcal{F}$. This collection is composed only of the sets $A_i = \emptyset$ or $A_i = \Omega$, $i \in \mathbb{N}$. Thus (recall Exercise 1.3),

$$\bigcup_{i\in\mathbb{N}} A_i = \begin{cases} \Omega, & \text{if } \exists\, i \text{ s.t. } A_i = \Omega, \\ \emptyset, & \text{otherwise.} \end{cases}$$

This, items (i), (ii) and (iii) of Definition 4.1 are satisfied and hence $\mathcal{F}$ is a σ-field. We remark that $\mathcal{F}$ is referred to as the trivial σ-field. It is the smallest possible σ-field on Ω.

Example 4.2. *Let $\Omega \neq \emptyset$ be an arbitrary set. Let $\mathcal{F}$ be the family which consists of all possible subsets of Ω, i.e.,*

$$\mathcal{F} = \{A \colon A \subset \Omega\}.$$

Then, $\mathcal{F}$ is a σ-field on Ω. Since $\Omega \subset \Omega$, $\Omega \in \mathcal{F}$. Let $A \in \mathcal{F}$, then by definition, $A^c = \Omega \setminus A \subset \Omega$ and hence $A^c \in \mathcal{F}$. Let $\{A_i \colon i \in \mathbb{N}\} \subset \mathcal{F}$. Then, by definition,

$$\bigcup_{i\in\mathbb{N}} A_i = \{\omega \in \Omega \colon \exists\, i \text{ s.t. } \omega \in A_i\} \subset \Omega.$$

Hence, $\mathcal{F}$ is a σ-field. The given σ-field $\mathcal{F}$ is referred to as the power set of Ω and denoted with $\mathcal{P}(\Omega)$ (or 2^Ω). It is the largest possible σ-field on Ω.

Example 4.3. *Let Ω be an uncountable set. We consider the family*

$$\mathcal{F} = \{A \colon A \subset \Omega \text{ s.t. } A \text{ is countable or } A^c \text{ is countable}\}.$$

Then, $\mathcal{F}$ is a σ-field on Ω. We have that $\Omega^c = \emptyset$. We know that $\#\emptyset = 0$, in particular $\emptyset$ is countable. Thus, $\Omega \in \mathcal{F}$. Let $A \in \mathcal{F}$. Thus, A is countable or A^c is countable. Since $(A^c)^c = A$, $A^c \in \mathcal{F}$. Let $\{A_i \colon i \in \mathbb{N}\} \subset \mathcal{F}$. If there exists $j \in \mathbb{N}$ s.t. A_j is not countable, we have that

$$\left(\bigcup_{i\in\mathbb{N}} A_i\right)^c = \bigcap_{i\in\mathbb{N}} A_i^c = \bigcap_{\substack{i\in\mathbb{N}\\ i\neq j}} A_i^c \cap A_j^c \subset A_j^c.$$

Thus,

$$\#\left(\bigcup_{i\in\mathbb{N}} A_i\right)^c \le \#A_j^c \le \#\mathbb{N},$$

since A_j^c is countable. Hence $(\cup_{i\in\mathbb{N}}A_i)^c$ is countable as well and therefore an element of $\mathcal{F}$. If for any $i \in \mathbb{N}$, A_i is countable we rely on Proposition 2.7 and conclude that $\cup_{i\in\mathbb{N}}A_i$ must be countable as well and hence, $\cup_{i\in\mathbb{N}}A_i \in \mathcal{F}$. Notice that since Ω is uncountable it is not true that $\mathcal{F} = \mathcal{P}(\Omega)$ (cf. Example 4.2). As a simple example consider $\Omega = [0, 1)$, then $A = [0, 1/2)$ and $A^c = [1/2, 0)$ are not countable.

Exercise 4.1. *Let $\Omega \neq \emptyset$ be countable and define $\mathcal{F}$ as in Example 4.3. Is it true that $\mathcal{F} = \mathcal{P}(\Omega)$?*

Exercise 4.2. *Let Ω be a non empty set an $\mathcal{F}$ be a σ-field on Ω. Show that*

(a) if $\{A_i : i \in \mathbb{N}\} \subset \mathcal{F}$, then $\cap_{i\in\mathbb{N}}A_i \in \mathcal{F}$.

(b) if $A \in \mathcal{F}$ and $B \in \mathcal{F}$ then $A \setminus B \in \mathcal{F}$;

Exercise 4.3. *Let $A \subset \Omega$, $\Omega \neq \emptyset$. Show that*

$$\{\emptyset, A, A^c, \Omega\},$$

is a σ-field on Ω.

Example 4.4. *Let Ω be a non empty set an $\mathcal{F}$ be a σ-field on Ω. Let $\Omega_0 \subset \Omega$ s.t. $\Omega_0 \neq \emptyset$. Then, the collection $\mathcal{F} \cap \Omega_0$ defined by*

$$\mathcal{F} \cap \Omega_0 = \{A \cap \Omega_0 : A \in \mathcal{F}\},$$

is a σ-field on Ω_0. Clearly, $\Omega_0 = \Omega \cap \Omega_0 \in \mathcal{F} \cap \Omega_0$. If $B \in \mathcal{F} \cap \Omega_0$, then $B = A \cap \Omega_0$ for some $A \in \mathcal{F}$. Then,

$$\Omega_0 \setminus B = (\Omega_0 \setminus A) \cup (\Omega_0 \setminus \Omega_0) = \Omega_0 \setminus A = A^c \cap \Omega_0 \in \mathcal{F} \cap \Omega_0.$$

Suppose that $\{B_i : i \in \mathbb{N}\} \subset \mathcal{F} \cap \Omega_0$. Therefore, for any $i \in \mathbb{N}$, $B_i = A_i \cap \Omega_0$ for some $A_i \in \mathcal{F}$. Then, since $\cup_{i\in\mathbb{N}}B_i = (\cup_{i\in\mathbb{N}}A_i) \cap \Omega_0$, it follows that $\cup_{i\in\mathbb{N}}B_i \in \mathcal{F} \cap \Omega_0$.

Example 4.5. *Let Ω be an infinite set. Define the family*

$$\mathcal{G} = \{A : A \subset \Omega \text{ s.t. } A \text{ is finite or } A^c \text{ is finite}\}.$$

Then, $\mathcal{G}$ is not a σ-field on Ω. To see this, let $\{\omega_i : i \in \mathbb{N}\}$ be a countably infinite sequence of distinct points of Ω. This is possible, since Ω is not finite. Define the set $A = \{\omega_{2i} : i \in \mathbb{N}\}$. Let $A_i = \{\omega_{2i}\}$, $i \in \mathbb{N}$, be the singleton sets of A. Thus, $A = \cup_{i\in\mathbb{N}}A_i$. It is clear that $A_i \in \mathcal{G}$ for any $i \in \mathbb{N}$. It is also true that $A \notin \mathcal{G}$ since A and A^c are both not finite ($\{\omega_{2i+1} : i \in \mathbb{N}\} \subset A^c$). Therefore, $\mathcal{G}$ is not a σ-field on Ω.

Exercise 4.4. *Let $\Omega \neq \emptyset$ be finite and define $\mathcal{G}$ as in Example 4.5. Is it true that $\mathcal{G}$ is a σ-field on Ω?*

Example 4.6. *Let $\Omega = \mathbb{R}$ and consider the family*

$$\mathcal{R} = \{A : A = (a, b],\ a, b \in \mathbb{R}\} \cup \{\emptyset\}.$$

That is, the members of $\mathcal{R}$ are either the empty set or a left-open interval. The family $\mathcal{R}$ is not a σ-field on $\mathbb{R}$. To see it, if $A = (a, b] \in \mathcal{R}$, then $A^c = (-\infty, a] \cup (b, \infty) \notin \mathcal{R}$. For another way to see it, let $x \in \mathbb{R}$. Then (cf. Exercise 4.1) we must have

$$\bigcap_{n\in\mathbb{N}} (x - n^{-1}, x] \in \mathcal{R}.$$

But, we readily see that

$$\bigcap_{n\in\mathbb{N}} (x - n^{-1}, x] = \{x\} \notin \mathcal{R}. \tag{7}$$

In order to verify (7), we notice first that $\{x\} \subset \cap_{n\in\mathbb{N}}(x-n^{-1},x]$*. For the other inclusion, let* $y \in \cap_{n\in\mathbb{N}}(x-n^{-1},x]$*. Then* $y \in (x-n^{-1},x]$ *for any* $n \in \mathbb{N}$*. That is, for any* $n \in \mathbb{N}$*,*

$$x - \frac{1}{n} < y \leq x.$$

If we let $a_n = x - n^{-1}$*,* $n \in \mathbb{N}$*, and* $b_n = x$*,* $n \in \mathbb{N}$*, using Proposition 3.5, we have that*

$$\lim_{n\to\infty} a_n = x \leq y \leq x.$$

Hence, $y = x$ *and therefore* $\cap_{n\in\mathbb{N}}(x-n^{-1},x] \subset \{x\}$*.*

The next result is of general importance as it shows that even though a family of subsets $\mathcal{G}$ might not be σ-field, one can always find a σ-filed which is the smallest possible σ-filed that contains $\mathcal{G}$.

Proposition 4.1. *Let* $\Omega \neq \emptyset$ *and* $\mathcal{G}$ *be a family of subsets of* Ω*. Then, there exists a* σ*-field* $\sigma(\mathcal{G})$ *which satisfies:*

(i) $\mathcal{G} \subset \sigma(\mathcal{G})$*;*

(ii) If $\mathcal{G} \subset \mathcal{U}$ *and* $\mathcal{U}$ *is a* σ*-field, then* $\sigma(\mathcal{G}) \subset \mathcal{U}$*.*

To prove Proposition 4.1 we rely on the following result:

Proposition 4.2. *Let* $\Omega \neq \emptyset$ *be a set. Let* $\mathcal{F}_i$*,* $i \in I$*, be a collection of* σ*-fields on* Ω *over an arbitrary set* I*. Then,*

$$\mathcal{F} = \bigcap_{i\in I} \mathcal{F}_i,$$

is a σ*-field on* Ω*.*

Exercise 4.5. *Prove Proposition 4.2.*

Proof of Proposition 4.1. We know that there exists at least one σ-field which contains $\mathcal{G}$, namely $\mathcal{P}(\Omega)$. Thus, we define

$$\sigma(\mathcal{G}) = \bigcap_{i\in I} \mathcal{F}_i,$$

where

$$\{\mathcal{F}_i : i \in I\} = \{\mathcal{U} : \ \mathcal{U} \text{ is a } \sigma\text{-field on } \Omega \text{ s.t. } \mathcal{G} \subset \mathcal{U}\}$$

Notice that the set I might not be countable. By Proposition 4.2, $\sigma(\mathcal{G})$ is a σ-field. It remains to show (i) and (ii) of Proposition 4.1. Since for each $i \in I$, $\mathcal{G} \subset \mathcal{F}_i$, it is not possible that $\mathcal{G}$ is not a subset of $\sigma(\mathcal{G})$ (see Exercise 1.8). With regard to (ii), let $g \in \sigma(\mathcal{G})$ and $\mathcal{U}$ be any σ-field which is s.t. $\mathcal{G} \subset \mathcal{U}$. Then, there exists $j \in I$ s.t. $\mathcal{U} = \mathcal{F}_j$. Since $g \in \sigma(\mathcal{G})$, $g \in \mathcal{F}_i$ for any $i \in I$. In particular, $g \in \mathcal{F}_j$. Thus, $\sigma(\mathcal{G}) \subset \mathcal{U}$. □

The σ-field $\sigma(\mathcal{G})$ of Proposition 4.1 is referred to as the σ-field generated by $\mathcal{G}$.

Proposition 4.3. *Let* $\sigma(\mathcal{G})$ *be the* σ*-field generated by a family of subsets* $\mathcal{G}$ *of* Ω*. Let* $\mathcal{A}$ *be another family of subsets of* Ω*. Then,*

(a) if $\mathcal{A}$ *is a* σ*-field s.t.* $\mathcal{G} \subset \mathcal{A}$ *and* $\mathcal{A} \subset \sigma(\mathcal{G})$*, then* $\mathcal{A} = \sigma(\mathcal{G})$*.*

(b) $\mathcal{A} \subset \mathcal{G} \Rightarrow \sigma(\mathcal{A}) \subset \sigma(\mathcal{G})$*;*

(c) $\mathcal{A} \subset \mathcal{G} \subset \sigma(\mathcal{A}) \Rightarrow \sigma(\mathcal{A}) = \sigma(\mathcal{G})$*;*

Example 4.7. *Let $\Omega \neq \emptyset$ and let $\mathcal{G} = \{\emptyset\}$. Then, $\sigma(\mathcal{G}) = \{\emptyset, \Omega\}$, the trivial σ-field on Ω (cf. Example 4.1). By (a) of Proposition 4.3 it is enough to show that $\{\emptyset, \Omega\} \subset \sigma(\mathcal{G})$ since $\{\emptyset, \Omega\}$ is a σ-field that contains $\mathcal{G}$. It is clear that $\{\emptyset, \Omega\} \subset \sigma(\mathcal{G})$ since $\sigma(\mathcal{G})$ is a σ-field, hence it must contain Ω.*

Example 4.8. *Let $\Omega = \{1, 2, 3\}$ and define $\mathcal{G} = \{\{1\}\}$. Clearly, $\sigma(\mathcal{G})$ must contain $\{1\}$. Since $\sigma(\mathcal{G})$ is a σ-field, it must contain Ω and its complement $\emptyset$. Further, it must contain the complement of $\{1\}$, i.e., it contains $\{1\}^c = \{2, 3\}$. Thus, we claim that*

$$\sigma(\mathcal{G}) = \big\{\emptyset, \{1\}, \{2, 3\}, \Omega\big\}.$$

By (a) of Proposition 4.3, if $\big\{\emptyset, \{1\}, \{2, 3\}, \Omega\big\}$ is a σ-field on Ω, the claim is true. This is the case (cf. Exercise 4.3). Generally, upon Exercise 4.3, if $A \subset \Omega$, then,

$$\sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\}.$$

We often omit the braces and use the notation $\sigma(A) = \sigma(\{A\})$.

Example 4.9. *Let $\Omega \neq \emptyset$ and*

$$\mathcal{F} = \{A \colon A \subset \Omega \text{ s.t. } A \text{ is countable or } A^c \text{ is countable}\},$$

i.e., $\mathcal{F}$ is the σ-field introduced in Example 4.3. We show that

$$\mathcal{F} = \sigma(\mathcal{G}),$$

where

$$\mathcal{G} = \big\{\{\omega\},\ \omega \in \Omega\big\}.$$

Clearly $\mathcal{G} \subset \mathcal{F}$ since each set $\{\omega\}$, $\omega \in \Omega$, has cardinality one. Thus, it remains to show that $\mathcal{F} \subset \sigma(\mathcal{G})$. Let $A \in \mathcal{F}$. Then either A or A^c is countable. Suppose that A is countable, then

$$A = \{\omega_i \colon i \in \mathbb{N}\},$$

for some collection of singletons $\omega_i \in \Omega$, $i \in \mathbb{N}$. Therefore, $A = \cup_{i \in \mathbb{N}}\{\omega_i\}$ and thus $A \in \sigma(\mathcal{G})$ since $\{\omega_i\} \in \sigma(\mathcal{G})$ for any $i \in \mathbb{N}$ and $\sigma(\mathcal{G})$ is a σ-field. If A^c is countable, by the latter argument, $A^c \in \sigma(\mathcal{G})$. But then, $(A^c)^c = A \in \sigma(\mathcal{G})$. The set $\mathcal{G}$ is referred to as the point-partition on Ω.

Exercise 4.6. *Let $\Omega \neq \emptyset$ and*

$$\mathcal{G} = \{A \colon A \subset \Omega \text{ s.t. } A \text{ is finite or } A^c \text{ is finite}\},$$

i.e., $\mathcal{G}$ is the family introduced in Example 4.5. Let $\mathcal{A} = \big\{\{\omega\} \colon \omega \in \Omega\big\}$. Show that

(a) $\mathcal{G} \subset \sigma(\mathcal{A})$;

(b) $\sigma(\mathcal{A}) = \mathcal{G}$ if Ω is finite.

Proposition 4.4. *Let Ω be a non empty set an $\mathcal{F}$ be a σ-field on Ω. Let $\Omega_0 \subset \Omega$ s.t. $\Omega_0 \neq \emptyset$. Define $\mathcal{F} \cap \Omega_0$ as in Example 4.4 and assume $\mathcal{F} = \sigma(\mathcal{G})$ for some family $\mathcal{G}$ of subsets of Ω. Then,*

$$\mathcal{F} \cap \Omega_0 = \sigma(\{G \cap \Omega_0 \colon G \in \mathcal{G}\}).$$

Proof. Given any $G \in \mathcal{G}$, since $\mathcal{F} = \sigma(\mathcal{G})$, it follows that $G \cap \Omega_0 \in \mathcal{F} \cap \Omega_0$. That is $\mathcal{F} \cap \Omega_0$ is a σ-field on Ω_0 that contains $\{G \cap \Omega_0 \colon G \in \mathcal{G}\}$. Since $\sigma(\{G \cap \Omega_0 \colon G \in \mathcal{G}\})$ is the smallest such σ-field, it follows that $\sigma(\{G \cap \Omega_0 \colon G \in \mathcal{G}\}) \subset \mathcal{F} \cap \Omega_0$. To make the notation easier, we write $\sigma(\{G \cap \Omega_0 \colon G \in \mathcal{G}\}) = \mathcal{F}_0$. Hence, it remains to show that $\mathcal{F} \cap \Omega_0 \subset \mathcal{F}_0$. Suppose that $A \in \mathcal{F}$ implies that $A \in \{A \subset \Omega \colon A \cap \Omega_0 \in \mathcal{F}_0\}$. Then, given any $B \in \mathcal{F} \cap \Omega_0$, i.e., $B = A \cap \Omega_0$ for $A \in \mathcal{F}$, it follows that $B \in \mathcal{F}_0$. Therefore, it is sufficient to show that $\mathcal{F} \subset \{A \subset \Omega \colon A \cap \Omega_0 \in \mathcal{F}_0\}$. We set $\mathcal{A} = \{A \subset \Omega \colon A \cap \Omega_0 \in \mathcal{F}_0\}$. We notice first that $\mathcal{G} \subset \mathcal{A}$, since any set $G \in \mathcal{G}$ is s.t. $G \cap \Omega_0 \in \mathcal{F}_0$ by definition. If we show that $\mathcal{A}$ is a σ-field on Ω, we are done, since then $\mathcal{F} = \sigma(\mathcal{G}) \subset \mathcal{A}$. We notice that by definition $\mathcal{F}_0$ is a σ-field on Ω_0, i.e., it contains Ω_0. Thus, $\Omega \in \mathcal{A}$ since $\Omega \cap \Omega_0 = \Omega_0 \in \mathcal{F}_0$. If $A \in \mathcal{A}$, then $A \cap \Omega_0 \in \mathcal{F}_0$ and hence $\Omega_0 \setminus A \cap \Omega_0 = \Omega_0 \setminus A \in \mathcal{F}_0$. Hence, $\Omega \setminus A$ is s.t. $(\Omega \setminus A) \cap \Omega_0 = \Omega_0 \setminus A \in \mathcal{F}_0$. Therefore, $\Omega \setminus A \in \mathcal{A}$. Suppose that $\{A_i \colon i \in \mathbb{N}\} \subset \mathcal{A}$. Then, $A_i \cap \Omega_0 \in \mathcal{F}_0$ for any $i \in \mathbb{N}$. Therefore, $\cup_{i \in \mathbb{N}}(A_i \cap \Omega_0) = (\cup_{i \in \mathbb{N}} A_i) \cap \Omega_0 \in \mathcal{F}_0$. That is, $\cup_{i \in \mathbb{N}} A_i \in \mathcal{A}$. □

Example 4.10. *Let $\Omega = \mathbb{R}$ and $\mathcal{R}$ be the family of left-open intervals with the empty set adjoined (cf. Example 4.6). The σ-field $\mathfrak{B}(\mathbb{R}) = \sigma(\mathcal{R})$ is referred to as the Borel σ-field on $\mathbb{R}$.*

Exercise 4.7. *Let $\mathcal{R}' = \{(-\infty, x] \colon x \in \mathbb{R}\}$. Show that $\mathfrak{B}(\mathbb{R}) = \sigma(\mathcal{R}')$.*

Example 4.11. *Let $\Omega = \mathbb{R}^k$, $k \in \mathbb{N}$. Given $a_i, b_i \in \mathbb{R}$, $i = 1, \dots, k$, a set*

$$\prod_{i=1}^{k} (a_i, b_i],$$

is referred to as a rectangle on $\mathbb{R}^k$. Define

$$\mathcal{R}_k = \big\{A \colon A = \prod_{i=1}^{k} (a_i, b_i],\ a_i, b_i \in \mathbb{R},\ i = 1, \dots, k\big\} \cup \{\emptyset\},$$

i.e., the family of rectangles in $\mathbb{R}^k$. Then, the σ-field $\mathfrak{B}(\mathbb{R}^k) = \sigma(\mathcal{R}_k)$ is referred to as the Borel σ-field on $\mathbb{R}^k$.

Exercise 4.8. *Let*

$$\mathcal{R}'_k = \{(-\infty, x_1] \times \cdots (-\infty, x_k] \colon x = (x_1, \dots, x_k) \in \mathbb{R}^k\} \cup \{\emptyset\}.$$

Show that $\mathfrak{B}(\mathbb{R}^k) = \sigma(\mathcal{R}'_k)$.

Having in mind Example 4.4, we define the Borel σ-field restricted to a subset of $\mathbb{R}^k$ as follows:

Definition 4.2. *Let $E \subset \mathbb{R}^k$, $E \neq \emptyset$. The Borel σ-field on E is defined by*

$$\mathfrak{B}(E) = \mathfrak{B}(\mathbb{R}^k) \cap E = \{A \cap E \colon A \in \mathfrak{B}(\mathbb{R}^k)\}.$$

Then, using Proposition 4.4, we obtain that

Proposition 4.5. *Given any $E \subset \mathbb{R}^k$ s.t. $E \neq \emptyset$,*

$$\mathfrak{B}(E) = \sigma(\{G \cap E \colon G \in \mathcal{R}_k\}).$$

Definition 4.3. *Let $\Omega \neq \emptyset$ and $\mathcal{F}$ be a σ-field on Ω. The pair $(\Omega, \mathcal{F})$ is referred to as a measurable space. If $A \in \mathcal{F}$, then A is said to be measurable. If $\mathcal{A} \subset \mathcal{F}$ and $\mathcal{A}$ is a σ-field on Ω, $\mathcal{A}$ is referred to as a sub-σ-field on Ω.*

4.2 Solution to exercises

Solution 4.1 (Solution to Exercise 4.1). *Yes, if Ω is countable, then $\mathcal{F} = \mathcal{P}(\Omega)$. This follows from the fact that in this case, any set from $\mathcal{P}(\Omega)$ is countable and hence a member of $\mathcal{F}$.*

Solution 4.2 (Solution to Exercise 4.2). *We have seen that (cf. Exercise 1.10) for any $n \in \mathbb{N}$,*

$$\bigcap_{i=1}^{n} A_i = \left(\bigcup_{i=1}^{n} A_i^c \right)^c .$$

Hence, (a) follows from (ii) and (iii) of Definition 4.1. With regard to (b), using Proposition 1.4, $A \setminus B = A \cap B^c$ and thus (b) follows from (a) and (ii) of Definition 4.1.

Solution 4.3 (Solution to Exercise 4.3). *Let $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$. We need to verify items (i), (ii) and (iii) of Definition 4.1. Items (i) and (ii) are clearly satisfied. Since $\mathcal{F}$ only contains 4 sets, we can argue case by case. Let $k \in \{2, 3\}$ and*

$$\bigcap_{i_k=1}^{k} A_{i_k},$$

be any intersection of subsets $A_{i_k} \in \mathcal{F}$. Clearly, for any k, if $A_{i_k} = \emptyset$ for some i_k, $\cap_{i_k=1}^{k} A_{i_k} = \emptyset \in \mathcal{F}$. Thus, we have that for any $k = 2, 3$,

$$\bigcap_{i_k=1}^{k} A_{i_k} = \begin{cases} \emptyset, & \text{if } \exists\, i_k \text{ s.t. } A_{i_k} = \emptyset, \\ \bigcap_{i_k=1}^{k} A_{i_k}^1, & \text{otherwise,} \end{cases}$$

where $A_{i_k}^1$ are updated in the sense that $A_{i_k}^1 \in \{A, A^c, \Omega\}$. If $k = 2$, i.e., we intersect two subsets of $\{A, A^c, \Omega\}$,

$$\bigcap_{i_k=1}^{k} A_{i_k}^1 = \begin{cases} A, & \text{if } A_{i_k}^1 \in \{A, \Omega\},\ i_k = 1, 2 \\ A^c, & \text{if } A_{i_k}^1 \in \{A^c, \Omega\},\ i_k = 1, 2 \\ \emptyset, & \text{if } A_{i_k}^1 \in \{A^c, A\},\ i_k = 1, 2. \end{cases}$$

Otherwise, if $k = 3$, $\cap_{i_k=1}^{k} A_{i_k} = \emptyset$, since $A \cap A^c = \emptyset$. This shows that any possible intersection of sets from $\mathcal{F}$ is again a member of $\mathcal{F}$. Therefore, since for any $k \in \{2, 3\}$,

$$\bigcup_{i_k=1}^{k} A_{i_k} = \left(\bigcap_{i_k=1}^{k} A_{i_k}^c \right)^c,$$

item (iii) is satisfied.

Solution 4.4 (Solution to Exercise 4.4). *Yes, if Ω is finite, then $\mathcal{G} = \mathcal{P}(\Omega)$. This follows from the fact that in this case, any set from $\mathcal{P}(\Omega)$ is finite and hence a member of $\mathcal{G}$.*

Solution 4.5 (Solution to Exercise 4.5). *We need to verify items (i), (ii) and (iii) of Definition 4.1. Since $\Omega \in \mathcal{F}_i$ for any $i \in I$, $\Omega \in \mathcal{F} = \cap_{i \in I} \mathcal{F}_i$ by definition. Also, if $A \in \mathcal{F}$, then $A \in \mathcal{F}_i$ for any $i \in I$. In particular, $A^c \in \mathcal{F}_i$ for any $i \in I$. Thus, $A^c \in \mathcal{F}$. If $\{A_n \colon n \in \mathbb{N}\} \subset \mathcal{F}$ we have that $\{A_n \colon n \in \mathbb{N}\} \subset \mathcal{F}_i$ for any $i \in I$ (cf. Exercise 1.8). Thus, $\cup_{i \in I} A_i \in \mathcal{F}_i$ for any $i \in I$. Hence, $\cup_{i \in I} A_i \in \mathcal{F}$.*

Solution 4.6 (Solution to Exercise 4.6). *Item (a) follows from the fact that $\sigma(\mathcal{A})$ is a σ-field: If $A \in \mathcal{G}$, then since either A or A^c is finite, either of the two is a countable union of elements from $\mathcal{A}$, i.e., $A \in \sigma(\mathcal{A})$. With regard to (b), it remains to show that $\sigma(\mathcal{A}) \subset \mathcal{G}$ if Ω is finite. This is certainly true since in this case $\mathcal{G} = \mathcal{P}(\Omega)$.*

Solution 4.7 (Solution to Exercise 4.7). *We show that $\sigma(\mathcal{R})$ contains $\mathcal{R}'$ ($\Rightarrow \sigma(\mathcal{R}') \subset \sigma(\mathcal{R})$) and $\sigma(\mathcal{R}')$ contains $\mathcal{R}$ ($\Rightarrow \sigma(\mathcal{R}) \subset \sigma(\mathcal{R}')$). Let $x \in \mathbb{R}$, then, $(-\infty, x] = \cup_{n\in I}(-n, x] \in \sigma(\mathcal{R})$, where $I = \{n \in \mathbb{N}\colon -n < x\}$. Let $(a, b] \in \mathcal{R}$, then $(a, b]^c = (-\infty, a] \cup (b, \infty) = (-\infty, a] \cup (-\infty, b]^c \in \sigma(\mathcal{R}')$.*

Solution 4.8 (Solution to Exercise 4.8). *The inequality $\sigma(\mathcal{R}'_k) \subset \sigma(\mathcal{R}_k)$ follows from the fact that for any $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$,*

$$(-\infty, x_1] \times \ldots \times (-\infty, x_n] = \bigcup_{n\in I}\Big((-n, x_1] \times \ldots \times (-n, x_n]\Big),$$

where $I = \{n \in \mathbb{N}\colon -n < \min\{x_i\colon i = 1, \ldots, k\}\}$. For the other inequality, let $A = \prod_{i=1}^{k}(a_i, b_i] \in \mathcal{R}_k$. Consider the sets $(-\infty, b_i]$, $i = 1, \ldots, k$. Define the sets

$$S_i = (-\infty, b_1] \times \cdots \times \underbrace{(-\infty, a_i]}_{\text{Position } i} \times \cdots \times (-\infty, b_k], \quad i = 1, \ldots, k.$$

Then,

$$A = \Big(\prod_{i=1}^{k}(-\infty, b_i]\Big) \setminus \Big(\bigcup_{i=1}^{k} S_i\Big).$$

To see it, we notice that by Proposition 1.3,

$$\Big(\prod_{i=1}^{k}(-\infty, b_i]\Big) \setminus \Big(\bigcup_{i=1}^{k} S_i\Big) = \bigcap_{i=1}^{k}\Big(\Big(\prod_{i=1}^{k}(-\infty, b_i]\Big) \setminus S_i\Big)$$

Then, we have that for any $i = 1, \ldots, k$,

$$\Big(\prod_{i=1}^{k}(-\infty, b_i]\Big) \setminus S_i = \Big(\prod_{i=1}^{k}(-\infty, b_i]\Big) \cap S_i^c = (-\infty, b_1] \times \cdots \times \underbrace{(a_i, b_i]}_{\text{Position } i} \times \cdots \times (-\infty, b_k].$$

Thus,

$$\bigcap_{i=1}^{k}\Big((-\infty, b_1] \times \cdots \times \underbrace{(a_i, b_i]}_{\text{Position } i} \times \cdots \times (-\infty, b_k]\Big) = A,$$

as desired. Now,

$$\Big(\prod_{i=1}^{k}(-\infty, b_i]\Big) \setminus \Big(\bigcup_{i=1}^{k} S_i\Big) \in \sigma(\mathcal{R}'_k).$$

Hence, $\sigma(\mathcal{R}'_k)$ is another σ-field which contains sets of the form A. We know that $\sigma(\mathcal{R}_k)$ is the smallest such σ-field. Hence, $\sigma(\mathcal{R}_k) \subset \sigma(\mathcal{R}'_k)$.

4.3 Additional exercises

Exercise 4.9. *Let $f\colon X \to Y$ be a function and $\mathfrak{B}$ be a σ-field on Y. Show that*

$$\sigma(f) = \{f^{-1}(B)\colon B \in \mathfrak{B}\},$$

is a σ-field on X. $\sigma(f)$ is referred to as the σ-field generated by f.

Exercise 4.10. *Find an example of a set $\Omega \neq \emptyset$ and two σ-fields $\mathcal{F}_1$ and $\mathcal{F}_2$ on Ω s.t. the union $\mathcal{F}_1 \cup \mathcal{F}_2$ is not a σ-field on Ω. This shows that in contrast to the intersection of σ-fields (cf. Proposition 4.2), the union of σ-fields is not necessarily a σ-field.*
Hint: *Example 4.8.*

Exercise 4.11. *Let* $\Omega = [0, 1]$ *and* $\mathcal{G} = \{[0, 1/2], [1/2, 1]\}$*. Find* $\sigma(\mathcal{G})$*.*

Exercise 4.12. *Let* $\Omega = \mathbb{R}$*, equipped with the Borel* σ*-field* $\mathfrak{B}(\mathbb{R})$ *(cf. Example 4.10). Show that*

$$\sigma(\mathcal{R}_*) = \mathfrak{B}(\mathbb{R}),$$

where

$$\mathcal{R}_* = \{A\colon A = [a, b),\ a, b \in \mathbb{R}\} \cup \{\emptyset\},$$

i.e., the right-open intervals with the empty set adjoined.

Exercise 4.13. *We remain in the setting of Exercise 4.12. Show that*

$$\mathfrak{B}(\mathbb{R}) = \sigma(\mathcal{R}_{**}),$$

where

$$\mathcal{R}_{**} = \{A\colon A = (q_1, q_2],\ q_1, q_2 \in \mathbb{Q}\} \cup \{\emptyset\},$$

i.e., the left-open intervals with rational end-points and the empty set adjoined.
Hint: *Recall Example 3.4.*

5 Measurable sets: Part II

5.1 Measure spaces

Definition 5.1. *Let $(\Omega, \mathcal{F})$ be a measurable space. A function $\mu\colon \mathcal{F} \to \overline{\mathbb{R}}_+$ is said to be a measure on $\mathcal{F}$ if the following two items are satisfied:*

(i) $\mu(\emptyset) = 0$;

(ii) Given any disjoint family $\{A_i\colon i \in \mathbb{N}\}$ of measurable sets (i.e., $\{A_i\colon i \in \mathbb{N}\} \subset \mathcal{F}$),

$$\mu\Big(\bigcup_{i\in\mathbb{N}} A_i\Big) = \sum_{i\in\mathbb{N}} \mu(A_i).$$

Example 5.1. *Let $(\Omega, \mathcal{F})$ be a measurable space and $x \in \Omega$ be a given point of Ω. Define the function*

$$\delta_x(A) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

Then, $A \mapsto \delta_x(A)$ is a measure on $\mathcal{F}$. From the definition of δ_x, we immediately see that $\delta_x(\emptyset) = 0$. Let $\{A_i\colon i \in \mathbb{N}\} \subset \mathcal{F}$ be disjoint and set $A = \cup_{i\in\mathbb{N}} A_i$. If $x \notin A$ (i.e., $x \in \cap_{i\in\mathbb{N}} A_i^c$), there does not exists i s.t. $x \in A_i$, hence $\delta_x(A) = 0 = \sum_{i\in\mathbb{N}} \delta_x(A_i)$. Otherwise, if $x \in A$, since $\{A_i\colon i \in \mathbb{N}\}$ is disjoint, there exists a unique $j \in \mathbb{N}$, s.t. $x \in A_j$. Hence also in this case

$$\delta_x(A) = 1 = \delta_x(A_j) = \sum_{i\in\mathbb{N}} \delta_x(A_i).$$

Example 5.2. *We consider the measurable space $(\Omega, \mathcal{P}(\Omega))$, where Ω is a finite set and $\mathcal{P}(\Omega)$ is the power set of Ω (cf. Example 4.2). Define $\mu(A) = \#A$, $A \in \mathcal{P}(\Omega)$. Then, μ is a measure on $\mathcal{F}$. Clearly, for any $A \in \mathcal{P}(\Omega)$, $\mu(A) \geq 0$. We need to verify (i) and (ii) of Definition 5.1. Since $\#\emptyset = 0$, (i) is satisfied. Let $\{A_i\colon i \in \mathbb{N}\} \subset \mathcal{P}(\Omega)$ be disjoint. Naturally, since $\{A_i\colon i \in \mathbb{N}\}$ is a disjoint family of sets, the cardinality of its union is the sum of the individual set cardinalities. Let us show it. Clearly, since Ω is finite, the sets A_i, $i \in \mathbb{N}$, and $A = \cup_{i\in\mathbb{N}} A_i \subset \Omega$ are all finite sets. Let I_1 be s.t. $A_i = \emptyset$ for any $i \in I_1$ and $I_2 = \mathbb{N} \setminus I_1$, i.e., $A_i \neq \emptyset$ for any $i \in I_2$. Then, since A is finite, I_2 must be finite as well. That is, $A_i \neq \emptyset$ only for finitely many i. Then, for any $i \in I_2$, let $A_i = \{x_{i_1}, \ldots, x_{i_{n_i}}\}$, $n_i \in \mathbb{N}$. Since $\{A_i\colon i \in I_2\}$ is disjoint,*

$$\mu\Big(\bigcup_{i\in I_2} A_i\Big) = \#\bigcup_{i\in I_2} A_i = \sum_{i\in I_2} \#\{x_{i_k}\colon k = 1, \ldots, n_i\} = \sum_{i\in I_2} \mu(A_i).$$

In conclusion, we obtain that

$$\begin{aligned} \mu(A) &= \mu\Big(\bigcup_{i\in I_1} A_i \cup \bigcup_{i\in I_2} A_i\Big) \\ &= \mu\Big(\bigcup_{i\in I_2} A_i\Big) \\ &= \sum_{i\in I_2} \mu(A_i) \\ &= \sum_{i\in I_1} \mu(A_i) + \sum_{i\in I_2} \mu(A_i) = \sum_{i\in\mathbb{N}} \mu(A_i). \end{aligned}$$

We can generalize the previous example.

Example 5.3. *Consider the measurable space $(\Omega, \mathcal{P}(\Omega))$, where $\mathcal{P}(\Omega)$ is the power set of Ω and Ω is not necessarily finite. Define*

$$\mu(A) = \begin{cases} \#A, & \text{if } A \in \mathcal{P}(\Omega) \text{ s.t. } A \text{ is finite,} \\ \infty & \text{otherwise.} \end{cases}$$

Then, μ is a measure on $\mathcal{P}(\Omega)$. We have already seen in the previous example that (i) of Definition 5.1 is satisfied. Thus, let $\{A_i : i \in \mathbb{N}\} \subset \mathcal{P}(\Omega)$ be disjoint. Suppose that there exists $j \in \mathbb{N}$ s.t. A_j is not finite. Then,

$$\mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \infty = \sum_{i\in\mathbb{N}} \mu(A_i).$$

Thus, it remains to show (ii) for the case where A_i is finite for any $i \in \mathbb{N}$. Given $i \in \mathbb{N}$, let $A_i = \{x_{i_1}, \dots, x_{i_{n_i}}\}$, $n_i \in \mathbb{N}$. Set $U_n = \cup_{i=1}^n A_i$ and $A = \cup_{i\in\mathbb{N}} A_i$. As in the previous example, since the family $\{A_i : i \in \mathbb{N}\}$ is disjoint, we have that for any $n \in \mathbb{N}$,

$$\mu(U_n) = \#\{x_{i_k} : k = 1, \dots, n_i,\ 1 \le i \le n\} = \sum_{i=1}^n \#\{x_{i_k} : k = 1, \dots, n_i\} = \sum_{i=1}^n \mu(A_i).$$

We show that

$$\lim_{n\to\infty} \mu(U_n) = \mu(A), \tag{8}$$

then, (ii) of Definition 5.1 is verified. There are two cases, either A is finite or not. If A is finite, we repeat the arguments from Example 5.1 and obtain

$$\mu(A) = \sum_{i\in\mathbb{N}} \mu(A_i).$$

Thus, if we show (8) for the case wehen A is not finite, we are done. If A is not finite, then, for any $N \in \mathbb{N}$, there exists an integer $n \ge N$, s.t. A_n is not the empty set. In particular, for any $N \in \{n \in \mathbb{N} : n \ge 2\}$ there exists an integer $n \ge N$ s.t. $\#U_n - \#U_{n-1} = \#A_n = m_n$ for some positive integer m_n. Hence, we let $n_1 \in \mathbb{N}$ be s.t.

$$\#U_{n_1} = \#U_{n_1-1} + m_{n_1}, \quad n_1 \ge 1 + 1.$$

Then, we let n_2 be s.t.

$$\#U_{n_2} = \#U_{n_2-1} + m_{n_2}, \quad n_2 \ge n_1 + 1.$$

We continue like that and let n_k, $k \in \mathbb{N}$, be s.t.

$$\#U_{n_k} = \#U_{n_k-1} + m_{n_k}, \quad n_k \ge n_{k-1} + 1.$$

Now we note that for any $n \in \mathbb{N}$, $\#U_n \le \#U_{n+1}$, i.e., $(\#U_n)_{n\in\mathbb{N}}$ is increasing. Therefore,

$$\begin{aligned} \#U_{n_k} = \#U_{n_k-1} + m_{n_k} \ge \#U_{n_{k-1}+1-1} + m_{n_k} &= \#U_{n_{k-1}} + m_{n_k} \\ &= \#U_{n_{k-1}-1} + m_{n_{k-1}} + m_{n_k}. \end{aligned}$$

This shows that for any $k \in \mathbb{N}$,

$$\#U_{n_k} \ge \#U_{n_1-1} + \sum_{i=0}^{k-1} m_{n_{k-i}}.$$

Hence, given any $M \in \mathbb{R}$, if we let $k \in \mathbb{N}$ be large enough s.t. $\#U_{n_1-1} + \sum_{i=0}^{k-1} m_{n_{k-i}} \geq M$, we have that $\#U_n \geq M$ for any $n \geq n_k$. Finding such an integer k is certainly possible, since $\#U_{n_1-1}$ is a finite number and

$$\sum_{i=0}^{k-1} m_{n_{k-i}} \geq \underbrace{1 + \cdots + 1}_{k\text{-times}} = k.$$

In Particular, for any $M \in \mathbb{R}$ there exists $N = n_k \in \mathbb{N}$ s.t. $\#U_n \geq M$ for any $n \geq N$. Hence, by Definition 3.6, $\lim_{n\to\infty} \#U_n = \lim_{n\to\infty} \mu(U_n) = \infty$. Clearly, by definition of μ, since A is not finite, $\mu(A) = \infty$. This shows (8) and completes the argument. The measure μ is called the counting measure.

Let us also consider a more general version of Example 5.1.

Example 5.4. *Let $(\Omega, \mathcal{F})$ be a measurable space, I be a countable set and $E = \{x_i \colon i \in I\} \subset \Omega$ be a collection of points in Ω. Assume that α_i, $i \in I$, are s.t. $\alpha_i \in [0, \infty)$ for any $i \in I$. Define*

$$\mu(A) = \sum_{i \in I} \alpha_i \delta_{x_i}(A) = \sum_{x \in E} \alpha_x \delta_x(A), \quad A \in \mathcal{F},$$

where for any $i \in I$,

$$\delta_{x_i}(A) = \begin{cases} 1, & \text{if } x_i \in A, \\ 0, & \text{if } x_i \notin A. \end{cases}$$

Then, μ is a measure on $\mathcal{F}$. It is clear that (i) of Definition 5.1 is satisfied. Let $\{A_k \colon k \in \mathbb{N}\} \subset \mathcal{F}$ be disjoint and set $A = \cup_{k\in\mathbb{N}} A_k$. We need to verify that

$$\mu(A) = \sum_{k \in \mathbb{N}} \mu(A_k).$$

Certainly, this is true if for any $i \in I$, $x_i \notin A$. Assume that there exists $i \in I$ s.t. $x_i \in A$. Let us label these i's, i.e., i_1 is s.t. $x_{i_1} \in A$, i_2 is s.t. $x_{i_2} \in A$ and so on to obtain a family $\{x_{i_j} \colon j \in J\}$, where the set J is s.t. $J \subset I$ and for any $j \in J$, $x_{i_j} \in A$. We assume that any $x \in \{x_i \colon i \in I\} \setminus \{x_{i_j} \colon j \in J\}$ is s.t. $x \notin A$. Otherwise, we can always enlarge the set J. Then, since $\{A_k \colon k \in \mathbb{N}\}$ is disjoint, for each $j \in J$, there exists a unique $k_j \in \mathbb{N}$ s.t. $x_{i_j} \in A_{k_j}$. Therefore,

$$\begin{aligned} \mu(A) &= \sum_{i \in I} \alpha_i \delta_{x_i}(A) \\ &= \sum_{x \in \{x_{i_j} \colon j \in J\}} \alpha_x \delta_x(A) + \sum_{x \notin \{x_{i_j} \colon j \in J\}} \alpha_x \delta_x(A) \\ &= \sum_{j \in J} \alpha_{i_j} \delta_{x_{i_j}}(A) \\ &= \sum_{j \in J} \alpha_{i_j} \delta_{x_{i_j}}(A_{k_j}) = \sum_{j \in J} \alpha_{i_j}. \end{aligned}$$

Then, we notice that if $k \in \mathbb{N}$,

$$\begin{aligned} \sum_{i \in I} \alpha_i \delta_{x_i}(A_k) &= \sum_{x \in \{x_{i_j} \colon j \in J\}} \alpha_x \delta_x(A_k) + \sum_{x \notin \{x_{i_j} \colon j \in J\}} \alpha_x \delta_x(A_k) \\ &= \sum_{x \in \{x_{i_j} \colon j \in J\}} \alpha_x \delta_x(A_k) \\ &= \sum_{j \in J} \alpha_{i_j} \delta_{x_{i_j}}(A_k). \end{aligned}$$

Hence, since for any $j \in J$, $\sum_{k\in\mathbb{N}} \delta_{x_{i_j}}(A_k) = \delta_{x_{i_j}}(A_{k_j}) = 1$,

$$\begin{aligned}\sum_{k\in\mathbb{N}} \mu(A_k) &= \sum_{k\in\mathbb{N}} \Big(\sum_{i\in I} \alpha_i \delta_{x_i}(A_k) \Big) \\ &= \sum_{k\in\mathbb{N}} \Big(\sum_{j\in J} \alpha_{i_j} \delta_{x_{i_j}}(A_k) \Big) \\ &= \sum_{j\in J} \alpha_{i_j} \Big(\sum_{k\in\mathbb{N}} \delta_{x_{i_j}}(A_k) \Big) \\ &= \sum_{j\in J} \alpha_{i_j} \delta_{x_{i_j}}(A_{k_j}) = \sum_{j\in J} \alpha_{i_j}.\end{aligned}$$

It is important to remark that since $\alpha_{i_j}\delta_{x_{i_j}}(A_k) = a_{k,j} \geq 0$, the double sum $\sum_{(k,j)\in\mathbb{N}\times J} a_{k,j}$ is well defined and we are allowed to interchange the order of summation (cf. Proposition 3.11). Clearly, if I is finite, $\mu(A) < \infty$ for any $A \in \mathcal{F}$. If I is countably infinite, we have that $\#I = \#\mathbb{N}$ and hence, for any $A \in \mathcal{F}$, $\mu(A) = \sum_{n\in\mathbb{N}} \alpha_n \delta_{x_n}(A)$. Therefore, since $\alpha_i\delta_{x_i}(A) \leq \alpha_i$ and $\alpha_i \geq 0$, if follows from Proposition 3.10 that $\mu(A) < \infty$ for any $A \in \mathcal{F}$ if $\sum_{i\in I} \alpha_i < \infty$.

Exercise 5.1. *Consider the measurable space $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, where $\mathfrak{B}(\mathbb{R})$ is the Borel σ-field on $\mathbb{R}$. Define $\mu(B) = \sum_{n\in B\cap\mathbb{N}} 2^{-n}$, $B \in \mathfrak{B}(\mathbb{R})$. Is μ a measure on $\mathfrak{B}(\mathbb{R})$?*

Exercise 5.2. *Let $E = \{0, 1\}$, $p \in (0, 1)$ and set $p_0 = 1 - p$ and $p_1 = p$. Define the function*

$$P(B) = \sum_{x\in E\cap B} p_x, \quad B \in \mathfrak{B}(\mathbb{R}).$$

That is,

$$P(B) = \begin{cases} 0, & \text{if } 0 \notin B \text{ and } 1 \notin B, \\ 1-p, & \text{if } 0 \in B \text{ and } 1 \notin B, \\ p, & \text{if } 0 \notin B \text{ and } 1 \in B, \\ 1, & \text{if } 0 \in B \text{ and } 1 \in B. \end{cases}$$

Is P is a measure on $\mathfrak{B}(\mathbb{R})$?

Example 5.5. *Consider the measurable space $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, where $\mathfrak{B}(\mathbb{R}))$ is the Borel σ-field on $\mathbb{R}$ (cf. Example 4.6 and 4.10). Then, there exists a unique measure λ on $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ s.t. for any left-open interval $(a, b]$, $a, b \in \mathbb{R}$, $\lambda((a, b])$ returns the length of $(a, b]$, i.e., $\lambda((a, b]) = b - a$. The measure λ is referred to as the Lebesgue measure on $\mathfrak{B}(\mathbb{R})$. An explicit construction of λ is given in the next section.*

In the following we list some general properties of measures.

Proposition 5.1. *Let $(\Omega, \mathcal{F})$ be a measurable space and μ be a measure on $\mathcal{F}$. Then,*

(i) given $n \in \mathbb{N}$ and $\{A_i : 1 \leq i \leq n\} \subset \mathcal{F}$ disjoint, it follows that

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

(ii) if $A, B \in \mathcal{F}$ s.t. $A \subset B$ it follows that $\mu(A) \leq \mu(B)$;

(iii) if $A, B \in \mathcal{F}$ s.t. $A \subset B$ and $\mu(A) < \infty$ it follows that $\mu(B \setminus A) = \mu(B) - \mu(A)$;

(iv) given $A, B \in \mathcal{F}$, $\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B)$;

(v) if $\{A_i\colon i\in\mathbb{N}\}\subset\mathcal{F}$ *is s.t.* $A_i\subset A_{i+1}$*,*

$$\mu\Big(\bigcup_{i=1}^{n}A_i\Big)=\mu(A_n)\uparrow\mu\Big(\bigcup_{i\in\mathbb{N}}A_i\Big);$$

(vi) if $\{A_i\colon i\in\mathbb{N}\}\subset\mathcal{F}$ *is s.t.* $\mu(A_1)<\infty$ *and* $A_{i+1}\subset A_i$*,*

$$\mu\Big(\bigcap_{i=1}^{n}A_i\Big)=\mu(A_n)\downarrow\mu\Big(\bigcap_{i\in\mathbb{N}}A_i\Big).$$

(vii) if $\{A_i\colon n\in\mathbb{N}\}\subset\mathcal{F}$*,*

$$\mu\Big(\bigcup_{i\in\mathbb{N}}A_i\Big)\le\sum_{i\in\mathbb{N}}\mu(A_i);$$

Proof of item (v) of Proposition 5.1. Let $B_1=A_1$, $B_2=A_2\setminus A_1$ and so on until we set for any $i\in\mathbb{N}$, $B_i=A_i\setminus A_{i-1}$. It is true that for any $i\neq j$, $B_i\cap B_j=\emptyset$. Either, $i<j$, and then

$$B_i=A_i\setminus A_{i-1}\subset A_i\subset A_{j-1}\subset(A_j^c\cup A_{j-1})=(A_j\cap A_{j-1}^c)^c=(A_j\setminus A_{j-1})^c=B_j^c,$$

or $j>i$ and then $B_j\subset B_i^c$. Thus, The family $\{B_i\colon i\in\mathbb{N}\}$ is disjoint. By item (ii) of Definition 5.1,

$$\mu\Big(\bigcup_{i\in\mathbb{N}}B_i\Big)=\sum_{i\in\mathbb{N}}\mu(B_i),$$

and hence,

$$\sum_{i=1}^{n}\mu(B_i)\uparrow\mu\Big(\bigcup_{i\in\mathbb{N}}B_i\Big).$$

Further, for any $n\in\mathbb{N}$, $\cup_{i=1}^{n}B_i=\cup_{i=1}^{n}A_i$. To see it, we use an argument by induction. By definition, $B_1=A_1$. Then,

$$\begin{aligned}\bigcup_{i=1}^{n+1}B_i=\bigcup_{i=1}^{n}A_i\cup B_{n+1}&=\bigcup_{i=1}^{n}A_i\cup A_{n+1}\setminus A_n\\&=\bigcup_{i=1}^{n-1}A_i\cup A_n\cup A_{n+1}\setminus A_n=\bigcup_{i=1}^{n-1}A_i\cup A_n\cup A_{n+1}.\end{aligned}$$

Therefore, since for any $i\in\mathbb{N}$, $A_i\subset A_{i+1}$, it follows that $\cup_{i=1}^{n}B_i=A_n$. In conclusion, since

$$\bigcup_{i\in\mathbb{N}}A_i=\bigcup_{n\in\mathbb{N}}A_n=\bigcup_{n\in\mathbb{N}}\Big(\bigcup_{i=1}^{n}B_i\Big)=\bigcup_{i\in\mathbb{N}}B_i,$$

we have under application of item (i) of Proposition 5.1,

$$\mu(A_n)=\mu\Big(\bigcup_{i=1}^{n}B_i\Big)=\sum_{i=1}^{n}\mu(B_i)\uparrow\mu\Big(\bigcup_{i\in\mathbb{N}}B_i\Big)=\mu\Big(\bigcup_{i\in\mathbb{N}}A_i\Big).$$

□

Proof of item (vii) of Proposition 5.1. Let $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, $B_3 = A_3 \setminus A_1 \cup A_2$, and so on s.t. for any $i \in \mathbb{N}$,

$$B_i = A_i \setminus \bigcup_{k=1}^{i-1} A_k.$$

We notice that for $i \neq j$, $B_i \cap B_j = \emptyset$ since either $i < j$ and hence

$$B_i = A_i \setminus \cup_{k=1}^{i-1} A_k \subset A_i \subset \cup_{k=1}^{j-1} A_k \cup A_j^c = B_j^c,$$

or $j < i$ and then $B_j \subset B_i^c$. Hence, the family $\{B_i : i \in \mathbb{N}\}$ is disjoint. Further, we note that for any $n \in \mathbb{N}$,

$$\bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A_i.$$

To see it, we can use an argument by induction. We have that $A_1 = B_1$. Then, assume that $\cup_{i=1}^n B_i = \cup_{i=1}^n A_i$. It follows that

$$\bigcup_{i=1}^{n+1} B_i = \left(\bigcup_{i=1}^{n} B_i\right) \cup B_{n+1} = \left(\bigcup_{i=1}^{n} A_i\right) \cup B_{n+1} = \bigcup_{i=1}^{n} A_i \cup \left(A_{n+1} \setminus \bigcup_{i=1}^{n} A_i\right) = \bigcup_{i=1}^{n+1} A_i.$$

Therefore, by item (i) of Proposition 5.1, we have that for any $n \in \mathbb{N}$,

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) = \mu\left(\bigcup_{i=1}^{n} B_i\right) = \sum_{i=1}^{n} \mu(B_i) \leq \sum_{i=1}^{n} \mu(A_i), \tag{9}$$

where we used that for any $1 \leq i \leq n$, $B_i \subset A_i$ and hence $\mu(B_i) \leq \mu(A_i)$ by (ii) of Proposition 5.1. To conclude, we define $C_n = \cup_{i=1}^n A_i$ and obtain $C_n \subset C_{n+1}$ and hence

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) \uparrow \mu\left(\bigcup_{i \in \mathbb{N}} A_i\right),$$

by (v) of Proposition 5.1. This concludes the argument (cf. Proposition 3.8). □

Exercise 5.3. *Prove items (i), (ii), (iii) and (iv) of Proposition 5.1.*

Remark 5.1. *We remark that item (iv) of Proposition 5.1 can be stated more generally for a finite collection of sets with finite measure (the details are given in Section B.1).*

Definition 5.2. *Let $(\Omega, \mathcal{F})$ be a measurable space and μ be a measure on $\mathcal{F}$. The triple $(\Omega, \mathcal{F}, \mu)$ is referred to as a measure space.*

5.2 Semirings

Definition 5.3. *Let Ω be a nonempty set. A family of subsets $\mathcal{A}$ of Ω is said to be a semiring on Ω if*

(i) $\emptyset \in \mathcal{A}$;

(ii) $A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$;

(iii) if $A, B \in \mathcal{A}$ and $A \subset B$, then there exists a disjoint collection of sets $\{C_1, \dots, C_n\} \subset \mathcal{A}$ s.t. $B \setminus A = \cup_{k=1}^n C_k$.

Example 5.6. *We have seen in Example 4.6 that the family of left-open intervals*

$$\mathcal{R} = \{A\colon A = (a,b],\ a,b \in \mathbb{R}\} \cup \{\emptyset\},$$

is not a σ-field on $\mathbb{R}$. Still, it is a semiring on $\mathbb{R}$. By definition, $\emptyset \in \mathcal{R}$. Let $A, B \in \mathcal{R}$, i.e., $A = (a_1, a_2]$ and $B = (b_1, b_2]$. If $A \cap B = \emptyset$, then $A \cap B \in \mathcal{R}$. Otherwise,

$$A \cap B = (\max\{a_1, b_1\}, \min\{a_2, b_2\}].$$

Thus, item (ii) of Definition 5.3 is satisfied. With regard to (iii), let $A, B \in \mathcal{R}$ s.t. $A \subset B$. If $A = B = \emptyset$, then $A \setminus B = \emptyset$ and with $C = \emptyset$, the result follows. Similarly, if $A = B$, the result follows with $C = \emptyset$. If $B \neq \emptyset$ but $A = \emptyset$, then $B \setminus A = B$ and with $C = B$, the result follows. Thus, suppose that $A, B \neq \emptyset$, $A \subset B$ and $A \neq B$, i.e., $b_1 < a_1$ and $a_2 < b_2$. We have that $B \setminus A = (b_1, a_1] \cup (a_2, b_2]$. Hence, with $C_1 = (b_1, a_1]$ and $C_2 = (a_2, b_2]$, the result follows.

Example 5.7. *Let Ω be an infinite set and consider the family*

$$\mathcal{G} = \{A\colon A \subset \Omega \text{ s.t. } A \text{ is finite or } A^c \text{ is finite}\}.$$

Then, $\mathcal{G}$ is not a σ-field on Ω (cf. Example 4.5). However, $\mathcal{G}$ is a semiring on Ω.

5.3 Solution to exercises

Solution 5.1 (Solution to Exercise 5.1)**.** *Notice that we can write for any $B \in \mathfrak{B}(\mathbb{R})$,*

$$\mu(B) = \sum_{n \in B \cap \mathbb{N}} 2^{-n}\delta_n(B) = \sum_{n \in B \cap \mathbb{N}} 2^{-n}\delta_n(B) + \sum_{n \in B^c \cap \mathbb{N}} 2^{-n}\delta_n(B) = \sum_{n \in \mathbb{N}} 2^{-n}\delta_n(B).$$

Thus, μ is a measure on $\mathfrak{B}(\mathbb{R})$ (cf. Example 5.4). We remark that upon Exercise 3.14, we have that $\sum_{n\in\mathbb{N}} 2^{-n} = 1$. That is, $\mu(A) < \infty$ for any $A \subset \mathfrak{B}(\mathbb{R})$.

Solution 5.2 (Solution to Exercise 5.2)**.** *We set $\alpha_0 = 1 - p$ and $\alpha_1 = p$ and obtain that for any $B \in \mathfrak{B}(\mathbb{R})$,*

$$\sum_{x \in E} \alpha_x \delta_x(B) = \sum_{x \in E \cap B} p_x.$$

Thus, P is a measure on $\mathfrak{B}(\mathbb{R})$ (cf. Example 5.4). We remark that $\sum_{x\in E} p_x = 1$.

Solution 5.3 (Solution to Exercise 5.3)**.**

(i) We define $A_i = \emptyset$ for any $i > n$ and the statement is verified as a consequence of (i) and (ii) in Definition 5.1.

(ii) In general, $A \cup B = A \cup (B \setminus A)$. Sine $A \subset B$, it follows that $B = A \cup (B \setminus A)$. Hence by (i), $\mu(B) = \mu(A) + \mu(B \setminus A)$. Since $\mu(B \setminus A) \geq 0$, the result follows. Notice that $\mu(A) = \infty$ is possible.

(iii) As with the previous solution, $\mu(B) = \mu(A) + \mu(B \setminus A)$. Now, since $\mu(A) < \infty$, we obtain $\mu(B) - \mu(A) = \mu(B \setminus A)$.

(iv) We write $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$. Then, since $\{A \setminus B, B \setminus A, A \cap B\}$ is disjoint, it follows that

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

Similarly, since $A \cap B$ is a subset of both, A and B, we write

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

and

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

Hence,

$$\mu(A) + \mu(B) = \mu(A \setminus B) + \mu(B \setminus A) + 2\mu(A \cap B) = \mu(A \cup B) + \mu(A \cap B).$$

5.4 Additional exercises

Exercise 5.4. *Suppose that $E \subset \mathbb{R}$ is a countable set. Let $f\colon E \to [0, \infty)$ be a function and define*

$$\mu(B) = \sum_{x \in E \cap B} f(x), \quad B \in \mathfrak{B}(\mathbb{R}).$$

Verify that μ is a measure on $\mathfrak{B}(\mathbb{R})$.

Exercise 5.5. *Let $E = \{0, 1, \ldots, n\}$, $n \in \mathbb{N}$, and $p \in (0, 1)$. For any $k \in E$, let*

$$p_k = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Define the function

$$P(B) = \sum_{k \in E \cap B} p_k, \quad B \in \mathfrak{B}(\mathbb{R}).$$

Is P a measure on $\mathfrak{B}(\mathbb{R})$?

Exercise 5.6. *Let $(\Omega, \mathcal{F})$ be a measurable space and $\mu\colon \mathcal{F} \to \overline{\mathbb{R}}_+$ be a function s.t. $\mu(\emptyset) = 0$. Suppose that*

- *For any finite disjoint collection $\{A_i\colon i = 1, \ldots, n\} \subset \mathcal{F}$, $\mu(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i)$ (μ is finitely additive on $\mathcal{F}$);*
- *for any collection $\{A_i\colon i \in \mathbb{N}\} \subset \mathcal{F}$, $\mu(\cup_{i=1}^\infty A_i) \le \sum_{i=1}^\infty \mu(A_i)$ (μ is countable subadditive on $\mathcal{F}$).*

Show that μ is a measure on $\mathcal{F}$.

Exercise 5.7. *Show that the family*

$$\mathcal{G} = \{A\colon A \subset \Omega \text{ s.t. } A \text{ is finite or } A^c \text{ is finite}\},$$

of Example 5.7 is a semiring on Ω.

Exercise 5.8. *Prove item (vi) of Proposition 5.1.*

6 Measurable sets: Part III

A selection of omitted proofs of this chapter are found in Section B.2 of the appendix.

6.1 The Lebesgue measure

The main result of this chapter is the following proposition:

Proposition 6.1. *There exists a measure λ on $\mathfrak{B}(\mathbb{R})$ (referred to as the Lebesgue measure on $\mathfrak{B}(\mathbb{R})$) which is s.t.*

$$\text{for any left-open interval } (a,b],\ \lambda \text{ returns its length, i.e., } \lambda((a,b]) = b - a. \tag{10}$$

Further, λ is the unique measure on $\mathfrak{B}(\mathbb{R})$ which satisfies (10).

The intention of this chapter is to explore the tools that provide arguments for the latter proposition.

6.2 Measure extensions

Proposition 6.2. *Let $(a,b]$, $a < b \in \mathbb{R}$, be any left-open interval. Let I be countable and $(a_i, b_i]$, $i \in I$, be s.t., $(a,b] \subset \cup_{i\in I}(a_i, b_i]$, then*

$$b - a \leq \sum_{i\in I}(b_i - a_i). \tag{11}$$

If the collection $\{(a_i, b_i]\colon i \in I\}$ is disjoint we also have the following result (Exercise 6.10).

Proposition 6.3. *Let $(a,b]$, $a < b \in \mathbb{R}$, be any left-open interval. Let I be countable and $\{(a_i, b_i]\colon i \in I\}$ be a disjoint collection of left-open intervals s.t. $\cup_{i\in I}(a_i, b_i] \subset (a,b]$. Then*

$$\sum_{i\in I}(b_i - a_i) \leq b - a.$$

Definition 6.1. *Let $\Omega \neq \emptyset$ be a set and $\mathcal{A}$ be a collection of subsets from Ω. Let $A \in \mathcal{P}(\Omega)$ be any subset of Ω. A collection $\{U_i\colon i \in I\}$ is said to be a covering of A by sets from $\mathcal{A}$ if $\{U_i\colon i \in I\} \subset \mathcal{A}$ and $A \subset \cup_{i\in I}U_i$. A covering $\{U_i\colon i \in I\}$ of A by sets from $\mathcal{A}$ is referred to as countable (resp. finite) if I is countable (resp. finite). We write $C_{\mathcal{A}}(A)$ for the set which contains all the countable coverings of A by sets from $\mathcal{A}$, i.e.,*

$$C_{\mathcal{A}}(A) = \{\xi\colon \xi \text{ is a countable covering of } A \text{ by sets from } \mathcal{A}\}.$$

Example 6.1. *Consider the setting of Example 4.6 and let $\Omega = \mathbb{R}$ and $\mathcal{R}$ be the family of left-open intervals with the empty set adjoined:*

$$\mathcal{R} = \{A\colon A = (a,b],\ a,b \in \mathbb{R}\} \cup \{\emptyset\}.$$

Let $B_r(x)$ be any open ball with center $x \in \mathbb{R}$ and radius $r > 0$. That is, $B_r(x) = (x - r, x + r)$ is an open interval with endpoints $a = x - r$ and $b = x + r$. Consider the set $\xi_1 = \{(a,x], (x,b]\}$. Then, $\xi_1 \in C_{\mathcal{R}}((a,b))$. As another example, let for $n \in \mathbb{N}$,

$$U_i^n = \left(a + \frac{2ri}{2^n}, a + \frac{2r(i+1)}{2^n}\right], \quad i = 0, \dots 2^n - 1.$$

Then, $\xi_2 = \{U_i^n\colon i = 0, \dots 2^n - 1\} \in C_{\mathcal{R}}((a,b))$ for any $n \in \mathbb{N}$. As a final example, define

$$U_k = \left(\frac{a}{2^k}, \frac{b}{2^k}\right], \quad k \in \mathbb{N} \cup \{0\}.$$

Then, $\xi_3 = \{U_k \colon k \in \mathbb{N} \cup \{0\}\} \in C_{\mathcal{R}}((a,b))$. Each of the coverings ξ_1, ξ_2 and ξ_3 of (a,b) by sets from $\mathcal{R}$ offers an approach to quantify the length of (a,b) by summing up the respective lengths of the sets from $\mathcal{R}$. Given $A \in \mathcal{P}(\mathbb{R})$, we define the function $v_\ell(\xi) = \sum_{U \in \xi} \ell(U)$, $\xi \in C_{\mathcal{R}}(A)$ where $\ell \colon \mathcal{R} \to [0,\infty)$ is s.t.

$$\ell(U) = \begin{cases} b-a, & \text{if } U = (a,b], \\ 0, & \text{if } U = \emptyset. \end{cases}$$

As an example, we have that $v_\ell(\xi_1) = x - a + b - x = b - a$. Notice also, that

$$\begin{aligned} v_\ell(\xi_2) &= \sum_{i=0}^{2^n-1} \frac{2r(i+1)-i}{2^n} \\ &= \frac{2r}{2^n} + \frac{4r}{2^n} - \frac{2r}{2^n} + \frac{6r}{2^n} - \frac{4r}{2^n} + \cdots + \frac{2r(2^n-1)}{2^n} + 2r - \frac{2r(2^n-1)}{2^n} \\ &= 2r = b - a. \end{aligned}$$

Exercise 6.1. *Verify that $v_\ell(\xi_3) = 2(b-a)$.*

In the following we show that

$$\inf\{v_\ell(\xi) \colon \xi \in C_{\mathcal{R}}((a,b])\} = \inf_{\xi \in C_{\mathcal{R}}((a,b])} v_\ell(\xi) = b - a, \tag{12}$$

i.e., $b-a$ is a lower bound for the values of $v_\ell(\xi)$, $\xi \in C_{\mathcal{R}}((a,b])$.

Exercise 6.2. *Verify that $\inf_{\xi \in C_{\mathcal{R}}((a,b])} v_\ell(\xi) \leq b - a$.*

Upon the later exercise, it remains to show that $b - a \leq \inf_{\xi \in C_{\mathcal{R}}((a,b])} v_\ell(\xi)$. Let ξ be any countable covering of $(a,b]$ by sets from $\mathcal{R}$. That is, $\xi = \{U_i \colon i \in I\}$, with $U_i = (a_i, b_i]$ or $U_i = \emptyset$, $i \in I$, where I is countable. Since $\ell(\emptyset) = 0$, we assume without loss of generality that $U_i = (a_i, b_i]$ for any $i \in I$. Therefore, we have that $(a,b] \subset \cup_{i \in I}(a_i, b_i]$ and $v_\ell(\xi) = \sum_{i \in I}(b_i - a_i)$. Since I is countable, either I is finite or $\#I = \#\mathbb{N}$. Using Proposition 6.2 we obtain,

$$b - a \leq \sum_{i \in I}(b_i - a_i) = v_\ell(\xi).$$

It follows that $b - a \leq \inf_{\xi \in C_{\mathcal{R}}((a,b])} v_\ell(\xi)$. We also saw that there exists $\xi \in C_{\mathcal{R}}((a,b])$ s.t. $b - a = v_\ell(\xi)$. Hence, the latter infimum is actually a minimum. In conclusion we have proven the following result.

Proposition 6.4. *Given any left-open interval $(a,b]$, $\min_{\xi \in C_{\mathcal{R}}((a,b])} v_\ell(\xi) = b - a$.*

We build on the latter result and define the function

$$\ell^*(A) = \inf_{\xi \in C_{\mathcal{R}}(A)} v_\ell(\xi), \quad A \in \mathcal{P}(\mathbb{R}).$$

Exercise 6.3. *Verify that $\ell^*(\{a\}) = 0$ for any point $a \in \mathbb{R}$.*

The given examples show that the function $\mathcal{P}(\mathbb{R}) \ni A \mapsto \ell^*(A)$ is in alignment with our intuitive understanding of the length of an interval. However, we will see in the following that ℓ^* is not a measure on $\mathcal{P}(\mathbb{R})$. Nevertheless, we will show that it is always possible to restrict ℓ^* to $\sigma(\mathcal{R}) \subset \mathcal{P}(\mathbb{R})$ to obtain a measure λ on $\sigma(\mathcal{R}) = \mathfrak{B}(\mathbb{R})$ which agrees with ℓ on $\mathcal{R}$. This measure λ will be referred to as the Lebesgue measure on $\mathfrak{B}(\mathbb{R})$ (cf. Example 5.5). The function ℓ^* will be termed an outer measure, in accordance with the property that restricted to a smaller σ-field it becomes a measure.

Leaving the framework of the latter example, we introduce the following general definition.

Definition 6.2. *Let Ω be a nonempty set. An outer measure μ^* is a function $\mu^*\colon \mathcal{P}(\Omega) \to \overline{\mathbb{R}}_+$ that satisfies the following properties:*

(i) $\mu^*(\emptyset) = 0$;

(ii) $A, B \in \mathcal{P}(\Omega)$, *s.t.* $A \subset B \Rightarrow \mu^*(A) \leq \mu^*(B)$;

(iii) μ^* *is countable subadditive on* $\mathcal{P}(\Omega)$, *i.e., for any collection* $\{A_n\colon n \in \mathbb{N}\} \subset \mathcal{P}(\Omega)$, $\mu^*(\cup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$.

Then, the following result offers a general classification of the function ℓ^* covered in Example 6.1:

Proposition 6.5. *Let $\mathcal{A} \subset \mathcal{P}(\Omega)$ where Ω is some nonempty set. Suppose that $\emptyset \in \mathcal{A}$ and $\rho\colon \mathcal{A} \to \overline{\mathbb{R}}_+$ is a function s.t. $\rho(\emptyset) = 0$. Let $v_\rho(\xi) = \sum_{U\in\xi} \rho(U)$, $\xi \in C_{\mathcal{A}}(A)$. Then, the function,*

$$\mathcal{P}(\Omega) \ni A \mapsto \rho^*(A) = \inf_{\xi \in C_{\mathcal{A}}(A)} v_\rho(\xi),$$

is an outer measure.

Exercise 6.4. *Let ℓ be as in Example 6.1. Verify that $\ell^*(\mathbb{R}) = \infty$.*

Definition 6.3. *Let $\mu^*\colon \mathcal{P}(\Omega) \to \overline{\mathbb{R}}_+$ be a function. Define the set*

$$\mathcal{M}(\mu^*) = \{A \in \mathcal{P}(\Omega)\colon \mu^*(A \cap E) + \mu^*(A^c \cap E) = \mu^*(E)\ \forall E \in \mathcal{P}(\Omega)\}.$$

We say that A is μ^*-measurable if $A \in \mathcal{M}(\mu^*)$.

We obtain the following result.

Proposition 6.6. *Let μ^* be an outer measure on $\mathcal{P}(\Omega)$. Then,*

(I) $\mathcal{M}(\mu^*)$ *is a σ-field;*

(II) The restriction $\mu^*|_{\mathcal{M}(\mu^*)}$ *is a measure.*

The main result of this section is the following:

Proposition 6.7. *Let $\mathcal{A} \subset \mathcal{P}(\Omega)$ be a semiring and $\rho\colon \mathcal{A} \to \overline{\mathbb{R}}_+$ be a function which is s.t.,*

- $\rho(\emptyset) = 0$;
- *ρ is finitely additive on $\mathcal{A}$, i.e., for any finite disjoint collection $\{A_i\colon i = 1, \dots, n\} \subset \mathcal{A}$, $n \in \mathbb{N}$, with $\cup_{i=1}^n A_i \in \mathcal{A}$, $\rho(\cup_{i=1}^n A_i) = \sum_{i=1}^n \rho(A_i)$;*
- *ρ is countable subadditive on $\mathcal{A}$, i.e., for any collection $\{A_i\colon i \in \mathbb{N}\} \subset \mathcal{A}$ with $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$, $\rho(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \rho(A_i)$.*

Then, ρ extends to a measure on $\sigma(\mathcal{A})$. That is, there exists a measure $\rho_\uparrow\colon \sigma(\mathcal{A}) \to \overline{\mathbb{R}}_+$ which is s.t. $\rho_\uparrow(A) = \rho(A)$ for any $A \in \mathcal{A}$.

Exercise 6.5. *Let ρ be as in Proposition 6.7. Show that ρ is monotone on $\mathcal{A}$, i.e., $A, B \in \mathcal{A}$ s.t. $A \subset B \Rightarrow \rho(A) \leq \rho(B)$.*

We remark that in the statement of Proposition 6.7, since $\mathcal{A}$ is not necessarily a σ-field, to make sense of the condition that ρ is finitely additive (resp. countable subadditive) on $\mathcal{A}$, it is required to demand that $\cup_{i=1}^n A_i \in \mathcal{A}$ (resp. $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$).

Example 6.2. *We remain in the setting of Example 6.1, where $\ell\colon \mathcal{R} \to [0,\infty)$ is s.t.*

$$\ell(U) = \begin{cases} b-a, & \text{if } U = (a,b], \\ 0, & \text{if } U = \emptyset, \end{cases}$$

and for $A \in \mathcal{P}(\mathbb{R})$,

$$\ell^*(A) = \inf_{\xi \in C_{\mathcal{R}}(A)} v_\ell(\xi), \quad v_\ell(\xi) = \sum_{U \in \xi} \ell(U), \quad \xi \in C_{\mathcal{R}}(A),$$

with

$$C_{\mathcal{R}}(A) = \{\xi\colon \xi \text{ is a countable covering of } A \text{ by sets from } \mathcal{R}\}.$$

Proposition 6.5 shows that ℓ^* is an outer measure on $\mathcal{P}(\mathbb{R})$. By Proposition 6.6, $\mathcal{M}(\ell^*)$ is in fact a σ-field on $\mathbb{R}$ and the restriction of the outer measure ℓ^* to $\mathcal{M}(\ell^*)$ written as $\ell^*|_{\mathcal{M}(\ell^*)}$ is a measure on $\mathcal{M}(\ell^*)$. In addition, a set $A \subset \mathbb{R}$ is called Lebesgue measurable, if $A \in \mathcal{M}(\ell^*)$, i.e., if A is ℓ^*-measurable. We also know that the family of left-open intervals $\mathcal{R}$ is a semiring (cf. Example 5.6). Then, let $\{A_i, i \in \mathbb{N}\} \subset \mathcal{R}$, $A_i = (a_i, b_i]$, be disjoint and s.t., $\cup_{i\in\mathbb{N}}(a_i, b_i] = A \in \mathcal{R}$. Using Propositions 6.2 and 6.3, we conclude that $\ell(A) = \sum_{i\in\mathbb{N}} \ell(A_i)$. Hence, ℓ is additive on $\mathcal{R}$ and hence clearly finitely additive on $\mathcal{R}$. If $\{A_i, i \in \mathbb{N}\} \subset \mathcal{R}$ is not disjoint, then still, Propositions 6.2 shows that $\ell(A) \leq \sum_{i\in\mathbb{N}} \ell(A_i)$. Hence, ℓ satisfies all the conditions of Proposition 6.7. Therefore, there exists a measure $\ell_\uparrow$ on $\sigma(\mathcal{R})$ which is s.t. for any $A \in \mathcal{R}$, $\ell_\uparrow(A) = \ell(A)$. The measure $\ell_\uparrow$ is called the Lebesgue measure on $\sigma(\mathcal{R})$ and we use the notation $\ell_\uparrow = \lambda$ (cf. Example 5.5). In fact, we have that $\lambda = \ell^*|_{\sigma(\mathcal{R})}$, i.e., the Lebesgue measure on $\mathfrak{B}(\mathbb{R})$ equals the restriction of the outer measure ℓ^* to $\mathfrak{B}(\mathbb{R})$ (cf. the proof of Proposition 6.7). We also have that $\mathfrak{B}(\mathbb{R}) \subset \mathcal{M}(\ell^*)$ (cf. the proof of Proposition 6.7). Thus, every Borel set is also Lebesgue measurable. One can show that $\mathcal{M}(\ell^*) \setminus \mathfrak{B}(\mathbb{R}) \neq \emptyset$, i.e., there are sets that are Lebesgue measurable but not Borel measurable. Even further, the following result shows that $\mathcal{P}(\mathbb{R}) \setminus \mathcal{M}(\ell^*) \neq \emptyset$.

Proposition 6.8. *There are sets $V \subset \mathbb{R}$ which are not Lebesgue measurable.*

We omit a proof and refer to [1] or [2].

Definition 6.4. *Let $(\Omega, \mathcal{F})$ be a measurable space and μ be a measure on $\mathcal{F}$. Let $\mathcal{A} \subset \mathcal{F}$. The measure μ is called σ-finite on $\mathcal{A}$ if $\Omega = \cup_{i\in I} A_i$ for some countable collection of sets $\{A_i\colon i \in I\} \subset \mathcal{A}$ which are s.t. $\mu(A_i) < \infty$ for any $i \in I$.*

A proof of the following result is given in Section B.3 of the appendix.

Proposition 6.9. *Let $\Omega \neq \emptyset$ be a set and assume that $\mathcal{A} \subset \mathcal{P}(\Omega)$ is s.t. $A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$. Let μ_1 and μ_2 be two measures on $\sigma(\mathcal{A})$ where at least one of the measures μ_1 and μ_2 is a σ-finite measure on $\mathcal{A}$. If $\mu_1(A) = \mu_2(A)$ for any $A \in \mathcal{A}$, then $\mu_1(A) = \mu_2(A)$ for any $A \in \sigma(\mathcal{A})$, i.e., the two measures agree on $\sigma(\mathcal{A})$.*

Proposition 6.10. *The Lebesgue measure λ on $\mathfrak{B}(\mathbb{R})$ is the unique measure on $\mathfrak{B}(\mathbb{R})$ which is s.t. for any left-open interval $(a,b]$, $\lambda((a,b]) = b-a$.*

Remark 6.1. *We remark that upon Example 6.2 and the latter proposition, we have proven Proposition 6.1.*

6.3 The Lebesgue measure on real coordinate spaces

Let $\Omega = \mathbb{R}^k$, $k \in \mathbb{N}$, and consider the function

$$\ell_k(A) = \begin{cases} \prod_{i=1}^k (b_i - a_i) & \text{if } A = \prod_{i=1}^k (a_i, b_i] \\ 0, & \text{otherwise}, \end{cases}$$

defined on the set

$$\mathcal{R}_k = \{A\colon A = \prod_{i=1}^{k}(a_i, b_i],\ a_i, b_i \in \mathbb{R},\ i = 1, \dots, k\} \cup \{\emptyset\},$$

i.e., the family of rectangles in $\mathbb{R}^k$ (cf. Example 4.11). That is, ℓ_k $k \in \mathbb{N}$, returns length ($k = 1$), area ($k = 2$), volume ($k = 3$) and hypervolume ($k \geq 4$). Then, Proposition 6.1 is generalized as follows:

Proposition 6.11. *There exists a measure λ_k on $\mathfrak{B}(\mathbb{R}^k) = \sigma(\mathcal{R}_k)$ (referred to as the (k-dimensional) Lebesgue measure on $\mathfrak{B}(\mathbb{R}^k)$) which is s.t.*

$$\text{for any rectangle } A = \prod_{i=1}^{k}(a_i, b_i],\ \lambda_k \text{ satisfies } \lambda_k(A) = \ell_k(A). \tag{13}$$

Further, λ_k is the unique measure on $\mathfrak{B}(\mathbb{R}^k)$ which satisfies (13).

6.4 Solution to exercises

Solution 6.1 (Solution to Exercise 6.1)**.** *We have that*

$$v_\ell(\xi_3) = \sum_{k=0}^{\infty} \frac{(b-a)}{2^k} = (b-a)\sum_{k=0}^{\infty}\left(\frac{1}{2}\right)^k = 2(b-a).$$

Solution 6.2 (Solution to Exercise 6.2)**.** *This is clearly the case, as an example, take $\xi_* = \{(a, b]\}$, then $\xi_* \in C_{\mathcal{R}}((a, b])$ and $v_\ell(\xi_*) = b - a$. Hence,*

$$\inf_{\xi \in C_{\mathcal{R}}((a,b])} v_\ell(\xi) \leq b - a.$$

Solution 6.3 (Solution to Exercise 6.3)**.** *Let $a \in \mathbb{R}$. We know that $\ell^*(A) \geq 0$ for any $A \in \mathcal{P}(\mathbb{R})$. In particular, $\ell^*(\{a\}) \geq 0$. Define for any $\varepsilon > 0$,*

$$\xi_a^\varepsilon = \left\{\left(a - \frac{\varepsilon}{2^n}, a\right] : n \in \mathbb{N}\right\} \in C_{\mathcal{R}}(\{a\}).$$

We have that $v_\ell(\xi_a^\varepsilon) = \varepsilon$. Hence, since ε was arbitrary, $v_\ell(\xi_a^\varepsilon) = 0$. This shows that $\ell^*(\{a\}) \leq 0$ and hence $\ell^*(\{a\}) = 0$.

Solution 6.4 (Solution to Exercise 6.4)**.** *By Proposition 6.5, ℓ^* is an outer measure on $\mathcal{P}(\mathbb{R})$. In particular, it is monotone. That is, $A \subset B \Rightarrow \ell^*(A) \leq \ell^*(B)$. Also, because of Proposition 6.4, for any $n \in \mathbb{N}$, $\ell^*((-n/2, n/2]) = n$. This shows that for any $n \in \mathbb{N}$, $n = \ell^*((-n/2, n/2]) \leq \ell^*(\mathbb{R})$ and hence $\ell^*(\mathbb{R})$ can not be finite.*

Solution 6.5 (Solution to Exercise 6.5)**.** *Let $A, B \in \mathcal{A}$ s.t. $A \subset B$. Then, since $\mathcal{A}$ is a semiring, there exists a disjoint collection $\{C_i : i = 1, \dots, n\} \subset \mathcal{A}$, $n \in \mathbb{N}$, s.t. $B \setminus A = \cup_{i=1}^n C_i$. In particular, $B = A \cup (B \setminus A)$ and hence, since ρ is assumed to be finitely additive on $\mathcal{A}$,*

$$\rho(B) = \rho\left(A \cup \left(\bigcup_{i=1}^{n} C_k\right)\right) = \rho(A) + \rho\left(\bigcup_{i=1}^{n} C_k\right) \geq \rho(A).$$

6.5 Additional exercises

Exercise 6.6. *Argue that each of the following sets is a Borel set, i.e., a member of* $\mathfrak{B}(\mathbb{R})$ *and deduce its Lebesgue measure:*

- $\mathbb{R}$*;*
- $\{a\}$*,* $a \in \mathbb{R}$ *(singleton sets);*
- (a, b)*,* $[a, b)$ *and* $[a, b]$*, where* $a, b \in \mathbb{R}$*.*

Exercise 6.7. *Prove Proposition 6.10.*

Exercise 6.8. *Show that* ℓ^* *is not finitely additive on* $\mathcal{P}(\mathbb{R})$*, i.e., there exists a disjoint collection* $\{A_i \colon i = 1, \ldots, n\} \subset \mathcal{P}(\mathbb{R})$ *s.t.* $\ell^*(\cup_{i=1}^n A_i) \neq \sum_{i=1}^n \ell^*(A_i)$*. Conclude that* ℓ^* *is not a measure on* $\mathcal{P}(\mathbb{R})$*.*
Hint: *Proposition 6.8.*

Exercise 6.9. *Prove Proposition 6.3.*
Hint: *Induction.*

Exercise 6.10. *Let* $\mathcal{U} = \{U \colon U \subset \mathbb{R}^k \text{ open}\}$*, i.e.,* $\mathcal{U}$ *contains all the open sets of* $\mathbb{R}^k$ *(cf. Definition 2.16). Show that* $\mathfrak{B}(\mathbb{R}^k) = \sigma(\mathcal{U})$*.*

$\mathfrak{B}(\mathbb{R}^k)$ **contains** $\mathcal{U}$**:**

Step1: *Define*

$$\mathcal{R}_k(\mathbb{Q}) = \Big\{A \colon A = \prod_{i=1}^{k} (a_i, b_i],\ a_i, b_i \in \mathbb{Q},\ i = 1, \ldots, k\Big\}.$$

Verify that $\mathcal{R}_k(\mathbb{Q})$ *is countable.*

Step2: *Let* $U \in \mathcal{U}$ *be nonempty. Argue that for any* $x \in U$*, there exists* $R_x \in \mathcal{R}_k(\mathbb{Q})$ *s.t.* $x \in R_x$ *and* $R_x \subset U$*.*
Hint: *Recall that since* U *is open, it follows that for any* $x \in U$*, there exists an open ball* $B_{\varepsilon_x}(x)$ *s.t.* $B_{\varepsilon_x}(x) \subset U$*. Hence, it is sufficient to find*

$$R_x = (a_1, b_1] \times \cdots \times (a_k, b_k] \in \mathcal{R}_k(\mathbb{Q})$$

s.t. $x \in R_x$ *and for* $i = 1, \ldots, k$*,* $b_i - a_i < r(\varepsilon_x, k)$*, where* $r(\varepsilon_x, k)$ *is chosen s.t.* $\|x - y\| < \varepsilon_x$ *for any* $y \in R_x$*.*

Step3: *Write* U *as a countable union of rectangles from* $\mathcal{R}_k(\mathbb{Q})$*.*

$\sigma(\mathcal{U})$ **contains** $\mathcal{R}_k$**:** *Show that any element of* $\mathcal{R}_k$ *can be written as a countable intersection of open rectangles (cf. Exercise 2.4).*

7 Measurable functions

The omitted proofs of this chapter are found in Section B.4 of the appendix.

7.1 The concept of measurable functions

Definition 7.1. *Let $(\Omega, \mathcal{F})$ and $(\Omega^*, \mathcal{F}^*)$ be two measurable spaces (cf. Definition 4.3). A function $f\colon \Omega \to \Omega^*$ is said to be measurable $\mathcal{F}/\mathcal{F}^*$ if for any $A^* \in \mathcal{F}^*$, $f^{-1}(A^*) \in \mathcal{F}$.*

Proposition 7.1. *Let $(\Omega, \mathcal{F})$ and $(\Omega^*, \mathcal{F}^*)$ be two measurable spaces and $f\colon \Omega \to \Omega^*$ be a function. Suppose that $\mathcal{F}^* = \sigma(\mathcal{G})$ and for any $G \in \mathcal{G}$, $f^{-1}(G) \in \mathcal{F}$. Then, f is $\mathcal{F}/\mathcal{F}^*$ measurable.*

Proof. It is enough to show that $\Sigma^* = \{A^*\colon f^{-1}(A^*) \in \mathcal{F}\}$ is a σ-field on Ω^*. We have that $\Omega^* \in \Sigma^*$, since $f^{-1}(\Omega^*) = \Omega$ and $\mathcal{F}$ is a σ-field, i.e., it contains Ω. Let $A^* \in \Sigma^*$. Then, $f^{-1}((A^*)^c) = f^{-1}(A^*)^c$ (cf. Proposition 2.4) and hence $(A^*)^c \in \Sigma^*$. Let $\{A_i^*\colon i \in \mathbb{N}\} \subset \Sigma^*$. Then, since $f^{-1}(\cup_{i=1}^\infty A_i^*) = \cup_{i=1}^\infty f^{-1}(A_i^*)$ (cf. Proposition 2.4), $\cup_{i=1}^\infty A_i^* \in \Sigma^*$. □

Example 7.1. *Consider the measurable space $(\mathbb{R}^m, \mathfrak{B}(\mathbb{R}^m))$ and $(\mathbb{R}^k, \mathfrak{B}(\mathbb{R}^k))$. Let $f\colon \mathbb{R}^m \to \mathbb{R}^k$ be continuous (cf. Definition 2.17). Then, f is measurable $\mathfrak{B}(\mathbb{R}^m)/\mathfrak{B}(\mathbb{R}^k)$. To see it, let $\mathcal{U} = \{U\colon U \text{ open in } \mathbb{R}^k\}$. Then, since f is continuous, for any $U \in \mathcal{U}$, $f^{-1}(U)$ is an open set of $\mathbb{R}^m$. Since $\mathfrak{B}(\mathbb{R}^m)$ contains the open sets of $\mathbb{R}^m$, it follows that for any $U \in \mathcal{U}$, $f^{-1}(U) \in \mathfrak{B}(\mathbb{R}^m)$. We know that $\mathfrak{B}(\mathbb{R}^k) = \sigma(\mathcal{U})$ (cf. Exercise 6.10). Hence, by Proposition 7.1, f is measurable.*

Definition 7.2. *A function $f\colon \mathbb{R}^m \to \mathbb{R}^k$ is called Borel function if it is measurable $\mathfrak{B}(\mathbb{R}^m)/\mathfrak{B}(\mathbb{R}^k)$.*

Upon Example 7.2 we have proven the following result.

Proposition 7.2. *Any continuous function $f\colon \mathbb{R}^m \to \mathbb{R}^k$ is a Borel function.*

Remark 7.1. *Suppose that $f\colon E \to \mathbb{R}^k$, where $E \subset \mathbb{R}^m$, $E \neq \emptyset$. Using Proposition 4.4 and Exercise 6.10, we know that*

$$\mathfrak{B}(E) = \sigma(\{G \cap E\colon G \text{ open in } \mathbb{R}^m\}).$$

Further, by Proposition 2.11, if $f\colon E \to \mathbb{R}^k$ is continuous, then, for any $U \subset \mathbb{R}^k$ open, $f^{-1}(U) \in \{G \cap E\colon G \text{ open in } \mathbb{R}^m\}$. Hence, by Proposition 7.1, f is $\mathfrak{B}(E)/\mathfrak{B}(\mathbb{R}^k)$ measurable.

Using the structure of the Borel σ-field, the next result is helpful to verify that a function $f\colon \Omega \to \mathbb{R}$ is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable.

Proposition 7.3. *Let $(\Omega, \mathcal{F})$ be a measurable space and $f\colon \Omega \to \mathbb{R}$ be a real-valued function. Suppose that $\{\omega \in \Omega\colon f(\omega) \leq x\} \in \mathcal{F}$ for any $x \in \mathbb{R}$, then f is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable.*

Proof. Notice that for any $x \in \mathbb{R}$, $\{\omega \in \Omega\colon f(\omega) \leq x\} = f^{-1}((-\infty, x])$. Since $\mathfrak{B}(\mathbb{R}) = \sigma(\{(-\infty, x]\colon x \in \mathbb{R}\})$ (cf. Exercise 4.7), f is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable (cf. Proposition 7.1). □

Exercise 7.1. *Let $(\Omega, \mathcal{F})$ be a measurable space and $f(\omega) = \alpha$ for any $\omega \in \Omega$, where $\alpha \in \mathbb{R}$. Show that f is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable.*

Example 7.2. *In Example 5.1, we have considered the measure $A \mapsto \delta_\omega(A)$ for fixed $\omega \in \Omega$. If now $A \subset \Omega$ is fixed and ω is variable, then, the function*

$$\omega \mapsto \mathbb{1}_A(\omega) = \delta_\omega(A) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{otherwise.} \end{cases},$$

is referred to as the indicator function of the set A. Then, if $A \in \mathcal{F}$, $\omega \mapsto \mathbb{1}_A(\omega)$ is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable. To see it, we notice that for any $x \in \mathbb{R}$,

$$\{\omega \in \Omega \colon \mathbb{1}_A(\omega) > x\} = \begin{cases} \emptyset, & \text{if } x \geq 1, \\ A, & \text{if } 0 \leq x < 1, \\ \Omega & \text{if } x < 0. \end{cases}$$

Since $A \in \mathcal{F}$, $\{\omega \in \Omega \colon \mathbb{1}_A(\omega) > x\} \in \mathcal{F}$ for any $x \in \mathbb{R}$. Thus, $\{\omega \in \Omega \colon \mathbb{1}_A(\omega) \leq x\} \in \mathcal{F}$ for any $x \in \mathbb{R}$. By Proposition 7.3, $\omega \mapsto \mathbb{1}_A(\omega)$ is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable.

Example 7.3. *Let $\Omega = \{h, t\}$ and $\mathcal{F} = \mathcal{P}(\{h, t\}) = \{\emptyset, \{h\}, \{t\}, \{h, t\}\}$. Then, $\{h\} \in \mathcal{P}(\{h, t\})$. Thus,*

$$f(\omega) = \begin{cases} 1, & \text{if } \omega = h, \\ 0, & \text{if } \omega = t, \end{cases}$$

is $\mathcal{P}(\{h, t\})/\mathfrak{B}(\mathbb{R})$ measurable.

Example 7.4. *Let $(\Omega, \mathcal{F})$ be a measurable space and $f_i \colon \Omega \to \mathbb{R}$, $i = 1, \ldots, k$, be $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable. Then, $f = \min\{f_i \colon i = 1, \ldots, k\}$ is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable. To see it, we notice that for any $x \in \mathbb{R}$ and $i = 1, \ldots, k$, since f_i is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable, $f_i^{-1}((x, \infty)) \in \mathcal{F}$. Therefore, for any $x \in \mathbb{R}$,*

$$f^{-1}((x, \infty)) = \{\omega \in \Omega \colon f(\omega) > x\} = \cap_{i=1}^k \{\omega \in \Omega \colon f_i(\omega) > x\} = \cap_{i=1}^k f_i^{-1}((x, \infty)) \in \mathcal{F}.$$

Therefore, for any $x \in \mathbb{R}$, $(f^{-1}((x, \infty)))^c = f^{-1}((-\infty, x]) \in \mathcal{F}$. Using Proposition 7.3, f is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable.

The following result shows that the measurability of a vector valued function is characterized in terms of the measurability of the respective coordinate functions.

Proposition 7.4. *Let $(\Omega, \mathcal{F})$ be a measurable space and $f \colon \Omega \to \mathbb{R}^k$, i.e.,*

$$f(\omega) = (f_1(\omega), \ldots, f_k(\omega)).$$

Then, f is $\mathcal{F}/\mathfrak{B}(\mathbb{R}^k)$ measurable if and only if for any $i = 1, \ldots, k$, $f_i \colon \Omega \to \mathbb{R}$ is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable.

The latter result offers a helpful tool to show that if $f_1, \ldots, f_k \colon \Omega \to \mathbb{R}$ are $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable functions, then, the functions $\sum_{i=1}^k f_i$ and $\prod_{i=1}^k f_i$ are $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable as well. In order to arrive there, the conclusion of the following exercise is valuable.

Exercise 7.2. *Let $(\Omega, \mathcal{F})$ and $(\Omega^*, \mathcal{F}^*)$ be two measurable spaces and $f \colon \Omega \to \Omega^*$ be a function. Let $(\Omega^{**}, \mathcal{F}^{**})$ be a third measurable space and $f^* \colon \Omega^* \to \Omega^{**}$ be another function. Show that if f is $\mathcal{F}/\mathcal{F}^*$ measurable and f^* is $\mathcal{F}^*/\mathcal{F}^{**}$ measurable, then the composition $f^*(f) \colon \Omega \to \Omega^{**}$ is $\mathcal{F}/\mathcal{F}^{**}$ measurable.*

Proposition 7.5. *Let $(\Omega, \mathcal{F})$ be a measurable space and $f_i \colon \Omega \to \mathbb{R}$, $i = 1, \ldots, k$, be $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable. Suppose that $g \colon \mathbb{R}^k \to \mathbb{R}$ is $\mathfrak{B}(\mathbb{R}^k)/\mathfrak{B}(\mathbb{R})$ measurable. Then,*

$$\omega \mapsto g((f_1(\omega), \ldots, f_k(\omega))),$$

is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable.

Notice that we usually avoid the double bracket and simply write

$$\omega \mapsto g((f_1(\omega), \ldots, f_k(\omega))) = g(f_1(\omega), \ldots, f_k(\omega)).$$

Proof of Proposition 7.5. Since $f_i\colon \Omega \to \mathbb{R}$, $i = 1, \dots, k$, are $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable, we use Proposition 7.4 and deduce that the map

$$\omega \mapsto (f_1(\omega), \dots, f_k(\omega)),$$

is $\mathcal{F}/\mathfrak{B}(\mathbb{R}^k)$ measurable. Then, since g is $\mathfrak{B}(\mathbb{R}^k)/\mathfrak{B}(\mathbb{R})$ measurable, we rely on Exercise 7.2 and verify that

$$\omega \mapsto g(f_1(\omega), \dots, f_k(\omega)),$$

is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable. □

A direct consequence of the latter result is the following (cf. Example 7.1).

Proposition 7.6. *Let $(\Omega, \mathcal{F})$ be a measurable space and $f_i\colon \Omega \to \mathbb{R}$, $i = 1, \dots, k$, be $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable. Then, if $g\colon \mathbb{R}^k \to \mathbb{R}$ is continuous,*

$$\omega \mapsto g(f_1(\omega), \dots, f_k(\omega)),$$

is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable.

Example 7.5. *Let $(\Omega, \mathcal{F})$ be a measurable space and $f_i\colon \Omega \to \mathbb{R}$, $i = 1, \dots, k$, be $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable. Then, $\sum_{i=1}^k f_i$ is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable (cf. Proposition 2.12).*

Example 7.6. *Let $(\Omega, \mathcal{F})$ be a measurable space and $f_i\colon \Omega \to \mathbb{R}$, $i = 1, \dots, k$, be $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable. Then, $\prod_{i=1}^k f_i$ is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable (cf. Proposition 2.12).*

Exercise 7.3. *Let $(\Omega, \mathcal{F})$ be a measurable space and $f_i\colon \Omega \to \mathbb{R}$, $i = 1, \dots, k$, be $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable. Let $c_1, \dots, c_k \in \mathbb{R}$. Then, the function*

$$\omega \mapsto f(\omega) = \sum_{i=1}^k c_i f_i(\omega),$$

is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable.

Remark 7.2. *Let $(\Omega, \mathcal{F})$ be a measurable space and $f\colon \Omega \to \mathbb{R}$ be $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable s.t. $f(\Omega) \subset E$, where $E \in \mathfrak{B}(\mathbb{R})$. Suppose that $g\colon E \to \mathbb{R}$ is continuous. Then, the composition $g(f)\colon \Omega \to \mathbb{R}$ is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable. To see it, we notice that for any $B \in \mathfrak{B}(\mathbb{R})$,*

$$g(f)^{-1}(B) = \{\omega\colon g(f(\omega)) \in B\} = \{\omega\colon f(\omega) \in g^{-1}(B)\} = f^{-1}(g^{-1}(B)).$$

Therefore, since by assumption f is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable, it remains to check that $g^{-1}(B) \in \mathfrak{B}(\mathbb{R})$. Upon Remark 7.1 we know that since g is continuous, it is $\mathfrak{B}(E)/\mathfrak{B}(\mathbb{R})$ measurable. In particular, $g^{-1}(B) \in \mathfrak{B}(E)$. Further, since $E \in \mathfrak{B}(\mathbb{R})$, it follows that $\mathfrak{B}(E) \subset \mathfrak{B}(\mathbb{R})$ since for any $A \in \mathfrak{B}(\mathbb{R})$, $A \cap E \in \mathfrak{B}(\mathbb{R})$ (recall Definition 4.2). In conclusion, $g^{-1}(B) \in \mathfrak{B}(\mathbb{R})$.

Definition 7.3. *A function $f\colon \Omega \to \mathbb{R}$ is called simple if there exists $n \in \mathbb{N}$, $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ and sets $A_1, \dots, A_n \subset \Omega$ s.t.*

$$f(\omega) = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}(\omega), \quad \omega \in \Omega.$$

That is, a simple function is a finite linear combination of indicator functions.

Example 7.7. *Let $(\Omega, \mathcal{F})$ be a measurable space and f be a simple function on Ω, i.e., $f(\omega) = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}(\omega)$. Then, if $A_i \in \mathcal{F}$ for any $i = 1, \dots, n$, f is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable.*

Example 7.8. *Let $n \in \mathbb{N}$ and $\Omega = \{\omega \colon \omega = (\omega_1, \ldots, \omega_n) \colon \omega_i \in \{0,1\},\ i = 1, \ldots, n\}$. We write $\Omega = A_0 \cup A_1 \cup \cdots A_n$, where*

$$A_k = \{\omega \in \Omega \colon \sum_{i=1}^{n} \omega_i = k\}, \quad k = 0, \ldots, n.$$

Define

$$f(\omega) = \sum_{k=0}^{n} k \mathbb{1}_{A_k}(\omega).$$

Then, since $A_k \in \mathcal{P}(\Omega)$ for any $k = 0, \ldots, n$, f is $\mathcal{P}(\Omega)/\mathfrak{B}(\mathbb{R})$ measurable.

Definition 7.4. *Let $(\Omega, \mathcal{F})$ be a measurable space and $f \colon \Omega \to \mathbb{R}$ be a simple function on Ω, i.e., $f = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}$. f is called standard if $\cup_{i=1}^{n} A_i = \Omega$ and $\{A_1, \ldots, A_n\} \subset \mathcal{F}$ is disjoint. If f is standard, we say that f is a simple function in standard form.*

The following result shows that simple functions can be written in standard form.

Proposition 7.7. *Let $(\Omega, \mathcal{F})$ be a measurable space and $f(\omega) = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}(\omega)$, $\omega \in \Omega$, be a simple function on Ω s.t. $A_i \in \mathcal{F}$ for any $i = 1, \ldots, n$. Then, there exists a standard simple function $g \colon \Omega \to \mathbb{R}$ s.t. $f(\omega) = g(\omega)$ for any $\omega \in \Omega$.*

7.2 Functions taking values in the extended real numbers

In the context of measurable functions, it is helpful to work with the extended real numbers $\overline{\mathbb{R}}$. Thus, we extend the definition of a measurable function to allow for function $f \colon \Omega \to \overline{\mathbb{R}}$. This will be useful in the context of limits of measurable functions.

Definition 7.5. *Let $(\Omega, \mathcal{F})$ be a measurable space and $f \colon \Omega \to \overline{\mathbb{R}}$. We say that f is $\mathcal{F}$ measurable if for any $A \in \mathfrak{B}(\mathbb{R})$, $\{\omega \in \Omega \colon f(\omega) \in A\} \in \mathcal{F}$ and $\{\omega \in \Omega \colon f(\omega) = -\infty\} \in \mathcal{F}$ and $\{\omega \in \Omega \colon f(\omega) = \infty\} \in \mathcal{F}$.*

Remark 7.3. *To combine terminology, if $f \colon \Omega \to \mathbb{R}$, then f is said to be $\mathcal{F}$ measurable if it is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable, i.e., in this case there is no need to bother about the sets $\{\omega \in \Omega \colon f(\omega) = -\infty\}$ and $\{\omega \in \Omega \colon f(\omega) = -\infty\}$. Notice that if $f(\omega) \in \mathbb{R}$ for any $\omega \in \Omega$, we anyway have that $\{\omega \in \Omega \colon f(\omega) = -\infty\} = \{\omega \in \Omega \colon f(\omega) = -\infty\} = \emptyset \in \mathcal{F}$. Hence, any results on $\mathcal{F}$ measurable functions $f \colon \Omega \to \overline{\mathbb{R}}$ apply directly to $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable functions $f \colon \Omega \to \mathbb{R}$.*

Remark 7.4. *If $(\Omega, \mathcal{F})$ be a measurable space and $f \colon \Omega \to \mathbb{R}^k$, $k \geq 1$, then the statement f is $\mathcal{F}$ measurable always means that f is $\mathcal{F}/\mathfrak{B}(\mathbb{R}^k)$ measurable. As an example, if the measurable space $(\Omega, \mathcal{F})$ is given by $(\mathbb{R}^m, \mathfrak{B}(\mathbb{R}^m))$, i.e., $\mathcal{F} = \mathfrak{B}(\mathbb{R}^m)$ and $f \colon \mathbb{R}^m \to \mathbb{R}^k$, then f is $\mathfrak{B}(\mathbb{R}^m)$ measurable if it is $\mathfrak{B}(\mathbb{R}^m)/\mathfrak{B}(\mathbb{R}^k)$ measurable.*

Proposition 7.8. *Let $(\Omega, \mathcal{F})$ be a measurable space and $f, g \colon \Omega \to \overline{\mathbb{R}}$ be two $\mathcal{F}$ measurable functions. Then, $\{\omega \in \Omega \colon f(\omega) = g(\omega)\} \in \mathcal{F}$.*

Exercise 7.4. *Let $(\Omega, \mathcal{F})$ be a measurable space and $f \colon \Omega \to \overline{\mathbb{R}}$ be $\mathcal{F}$ measurable. Show that the function $\omega \mapsto cf(\omega)$, $c \in \mathbb{R}$, is $\mathcal{F}$ measurable.*

7.3 Sequences of measurable functions

Proposition 7.9. *Let $(\Omega, \mathcal{F})$ be a measurable space and $f_n \colon \Omega \to \overline{\mathbb{R}}$, $n \in \mathbb{N}$, be a sequence of functions s.t. f_n is $\mathcal{F}$ measurable for any $n \in \mathbb{N}$. Then,*

(i) given $E \subset \mathbb{N}$, the functions $\sup_{n \in E} f_n$ and $\inf_{n \in E} f_n$ are $\mathcal{F}$ measurable;

(ii) *The functions* $\liminf_{n\to\infty} f_n$ *and* $\limsup_{n\to\infty} f_n$ *are* $\mathcal{F}$ *measurable;*

(iii) *If for any* $\omega \in \Omega$, $\lim_{n\to\infty} f_n(\omega)$ *exists, then* $\omega \mapsto (\lim_{n\to\infty} f_n)(\omega)$ *is* $\mathcal{F}$ *measurable;*

(iv) *We have that* $\{\omega \in \Omega\colon (f_n(\omega))_{n\in\mathbb{N}} \text{ converges}\} \in \mathcal{F}$*;*

(v) *Let* $f\colon \Omega \to \overline{\mathbb{R}}$ *be* $\mathcal{F}$ *measurable, then,* $\{\omega \in \Omega\colon f_n(\omega) \xrightarrow{n\to\infty} f(\omega)\} \in \mathcal{F}$.

The following result shows that any non-negative $\mathcal{F}$ measurable function can be understood as a monotone limit of non-negative standard simple functions.

Proposition 7.10. *Let* $(\Omega, \mathcal{F})$ *be a measurable space* $f\colon \Omega \to \overline{\mathbb{R}}$ *be a* $\mathcal{F}$ *measurable function s.t.* $f(\omega) \geq 0$ *for any* $\omega \in \Omega$*. Then, there exists a sequence of standard simple functions* $f_n\colon \Omega \to [0,\infty)$, $n \in \mathbb{N}$, *s.t. for any* $\omega \in \Omega$, $f_n(\omega) \uparrow f(\omega)$.

Definition 7.6. *Let* $f\colon \Omega \to \overline{\mathbb{R}}$ *be a function. We define the positive part of* f *as the function*

$$f^+ = \max\{f, 0\},$$

and the negative part of f *as*

$$f^- = \max\{-f, 0\},$$

Exercise 7.5. *Show that*

(a) *for any* $\omega \in \Omega$, $f^+(\omega) \geq 0$ *and* $f^-(\omega) \geq 0$*;*

(b) $f(\omega) = f^+(\omega) - f^-(\omega)$*;*

(c) $|f(\omega)| = f^+(\omega) + f^-(\omega)$.

Verify further that if $(\Omega, \mathcal{F})$ *is a measurable space and* $f\colon \Omega \to \overline{\mathbb{R}}$ *is a* $\mathcal{F}$ *measurable function, then its positive and negative parts are* $\mathcal{F}$ *measurable.*

Upon the latter exercise, we have the following result.

Proposition 7.11. *Let* $(\Omega, \mathcal{F})$ *be a measurable space and* $f\colon \Omega \to \overline{\mathbb{R}}$ *be a* $\mathcal{F}$ *measurable function. Then, there exists a sequence of standard simple functions* $(f_n)_{n\in\mathbb{N}}$ *s.t. for any* $\omega \in \Omega$, $\lim_{n\to\infty} f_n(\omega) = f(\omega)$.

Proof. We write $f = f^+ - f^-$. Since both, f^+ and f^- are non-negative and $\mathcal{F}$ measurable, we use Proposition 7.10 to find simple functions $(f_n^+)_{n\in\mathbb{N}}$ and $(f_n^-)_{n\in\mathbb{N}}$ s.t. $f_n^+(\omega) \uparrow f^+(\omega)$ and $f_n^-(\omega) \uparrow f^-(\omega)$. Therefore, for any $\omega \in \Omega$, $f_n^+(\omega) - f_n^-(\omega) \xrightarrow{n\to\infty} f^+(\omega) - f^-(\omega) = f(\omega)$. □

We note that the statement of Exercise 7.3 can be generalized.

Proposition 7.12. *Let* $(\Omega, \mathcal{F})$ *be a measurable space and* $f_i\colon \Omega \to \overline{\mathbb{R}}$, $i = 1, \dots, k$, *be* $\mathcal{F}$ *measurable. Then, the functions*

- $\omega \mapsto \prod_{i=1}^k f_i(\omega)$*;*
- $\omega \mapsto \sum_{i=1}^k c_i f_i(\omega)$, $c_1, \dots, c_k \in \mathbb{R}$*;*

are $\mathcal{F}$ *measurable.*

Proof. We may see this result as a consequence of Propositions 7.11 and 7.9. Since for any $i = 1, \dots, k$, f_i is $\mathcal{F}$ measurable, we write $f_i(\omega) = \lim_{n\to\infty} f_i^n(\omega)$, $\omega \in \Omega$, where $(f_i^n)_{n\in\mathbb{N}}$ is a sequence of $\mathcal{F}$ measurable functions (cf. Propositions 7.11). Therefore, by item (iii) of Propositions 7.9,

$$\omega \mapsto \lim_{n\to\infty} \left(\prod_{i=1}^k f_i^n \right)(\omega) = \prod_{i=1}^k f_i(\omega),$$

and

$$\omega \mapsto \lim_{n\to\infty} \left(\sum_{i=}^{k} c_i f_i^n\right)(\omega) = \sum_{i=1}^{k} c_i f_i(\omega),$$

are $\mathcal{F}$ measurable. $\square$

7.4 Minimal measurability

Let $f(\omega) = (f_1(\omega), \dots, f_k(\omega))$, $\omega \in \Omega$, be a function defined on some set Ω, taking values in $\mathbb{R}^k$. We recall (cf. Exercise 4.9) that the σ-field generated by f is given by

$$\sigma(f) = \{f^{-1}(B) \colon B \in \mathfrak{B}(\mathbb{R}^k)\}.$$

Let

$$I = \{\mathcal{F} \colon \mathcal{F} \text{ is a } \sigma\text{-field on } \Omega \text{ s.t. } \sigma(f) \subset \mathcal{F}\}.$$

If $\mathcal{F} \in I$, then f is $\mathcal{F}$ measurable, since for any $B \in \mathfrak{B}(\mathbb{R}^k)$, $f^{-1}(B) \in \sigma(f) \subset \mathcal{F}$. Let

$$\Sigma = \bigcap_{\mathcal{F}\in I} \mathcal{F}.$$

We know that Σ is a σ-field (cf. Proposition 4.2) and by definition, we know that

$$\Sigma = \sigma(\sigma(f)) = \sigma(f),$$

since $\sigma(f)$ is a σ-field. Further, since $\sigma(f)$ is a σ-field that contains $\sigma(f)$, $\sigma(f) \in I$. Consider the set

$$J = \{\mathcal{F} \colon \mathcal{F} \text{ is a } \sigma\text{-field on } \Omega \text{ and } f \text{ is } \mathcal{F} \text{ measurable}\}.$$

We have that $I \subset J$. Take $\mathcal{F} \in J$, hence f is $\mathcal{F}$ measurable. Then, $\sigma(f) \subset \mathcal{F}$. Hence, $I = J$. In conclusion, $\sigma(f)$ is the smallest σ-field on Ω s.t. f is $\sigma(f)$ measurable since any other $\mathcal{A} \in J = I$, is s.t. $\sigma(f) = \Sigma = \cap_{\mathcal{F}\in I}\mathcal{F} \subset \mathcal{A}$. We notice that any function $h \colon \Omega \to \mathbb{R}$ that is $\sigma(f)$ measurable is s.t. for any $A \in \mathfrak{B}(\mathbb{R})$ there exists $B \in \mathfrak{B}(\mathbb{R}^k)$ s.t.

$$\{\omega \in \Omega \colon g(\omega) \in A\} = \{\omega \in \Omega \colon f(\omega) \in B\}.$$

The next result shows that the condition that $h \colon \Omega \to \mathbb{R}$ is $\sigma(f)$ measurable is equivalent to the condition that $h = g(f)$, for some function $g \colon \mathbb{R}^k \to \mathbb{R}$.

Proposition 7.13. *Let $f(\omega) = (f_1(\omega), \dots, f_k(\omega))$, $\omega \in \Omega$, be a function defined on some set Ω, taking values in $\mathbb{R}^k$. Let $\sigma(f)$ be the σ-field generated by f and $h \colon \Omega \to \mathbb{R}$ be a function. Then, h is $\sigma(f)$ measurable if and only if there exists a function $g \colon \mathbb{R}^k \to \mathbb{R}$ which is $\mathfrak{B}(\mathbb{R}^k)$ measurable and s.t. $h(\omega) = g(f(\omega))$.*

Example 7.9. *Let $\Omega = \{h, t\}$, $\mathcal{F} = \mathcal{P}(\{h, t\})$ and f_1 be the $\mathcal{P}(\{h, t\})/\mathfrak{B}(\mathbb{R})$ measurable map*

$$f_1(\omega) = \begin{cases} 1, & \text{if } \omega = h, \\ 0, & \text{if } \omega = t, \end{cases}$$

given in Example 7.3. Further, let

$$f_2(\omega) = \begin{cases} 1, & \text{if } \omega = t, \\ 0, & \text{if } \omega = h. \end{cases}$$

That is, $f_1 = \mathbb{1}_H$ and $f_2 = \mathbb{1}_T$, where $H = \{\omega \in \Omega \colon \omega = h\}$ and $T = \{\omega \in \Omega \colon \omega = t\}$. Define the map $h = \mathbb{1}_H + \mathbb{1}_T$. Then, h is $\sigma(f)$ measurable, where $f = (f_1, f_2)$. This follows from Proposition 7.13, with $g(x) = x_1 + x_2$, $x = (x_1, x_2) \in \mathbb{R}^2$, which is $\mathfrak{B}(\mathbb{R}^k)$ measurable since it is continuous (cf. Propositions 2.12 and 7.2).

7.5 Solution to exercises

Solution 7.1 (Solution to Exercise 7.1). *This follows from the fact that for any $x \in \mathbb{R}$,*

$$\{\omega \in \Omega \colon f(\omega) > x\} \in \{\emptyset, \Omega\} \subset \mathcal{F}.$$

Solution 7.2 (Solution to Exercise 7.2). *We need to show that for any $A^{**} \in \mathcal{F}^{**}$,*

$$f^*(f)^{-1}(A^{**}) \in \mathcal{F}.$$

Therefore, let $A^{**} \in \mathcal{F}^{**}$. We have that

$$\begin{aligned} f^*(f)^{-1}(A^{**}) &= \{\omega \in \Omega \colon f^*(f)(\omega) \in A^{**}\} \\ &= \{\omega \in \Omega \colon f^*(f(\omega)) \in A^{**}\} \\ &= \{\omega \in \Omega \colon f(\omega) \in (f^*)^{-1}(A^{**})\} \\ &= f^{-1}\big((f^*)^{-1}(A^{**})\big) \end{aligned}$$

Then, since by assumption, f^* is $\mathcal{F}^*/\mathcal{F}^{**}$ measurable, it follows that $(f^*)^{-1}(A^{**}) \in \mathcal{F}^*$. Further, since f is $\mathcal{F}/\mathcal{F}^*$ measurable, we conclude that $f^{-1}((f^*)^{-1}(A^{**})) \in \mathcal{F}$.

Solution 7.3 (Solution to Exercise 7.3). *Since for any $i = 1, \ldots, k$, $\omega \mapsto c_i$ and $\omega \mapsto f_i(\omega)$ are $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable, $\omega \mapsto c_i f_i(\omega)$ is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable (cf. Example 7.6). Therefore, f is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable (cf. Example 7.5).*

Solution 7.4 (Solution to Exercise 7.4). *If $c = 0$, then $\{\omega \in \Omega \colon cf(\omega) = 0\} = \Omega$. Thus, for any set $A \subset \overline{\mathbb{R}}$, the set $\{\omega \in \Omega \colon cf(\omega) \in A\}$ is either the empty set or Ω. Suppose that $c \neq 0$. Then,*

$$\{\omega \in \Omega \colon cf(\omega) = -\infty\} = \{\omega \in \Omega \colon f(\omega) = -\infty\} \in \mathcal{F},$$

and

$$\{\omega \in \Omega \colon cf(\omega) = \infty\} = \{\omega \in \Omega \colon f(\omega) = \infty\} \in \mathcal{F},$$

since f is $\mathcal{F}$ measurable. Consider the function

$$f^*(\omega) = cf(\omega)\mathbb{1}_F(\omega), \quad \omega \in \Omega,$$

where $F = \{\omega \in \Omega \colon cf(\omega) \in \mathbb{R}\}$. We notice that $F \in \mathcal{F}$, since

$$F^c = \{\omega \in \Omega \colon cf(\omega) = -\infty\} \cup \{\omega \in \Omega \colon cf(\omega) = \infty\}.$$

Let $x \in \mathbb{R}$ and define $z(\omega) = 0$ for any $\omega \in \Omega$, i.e., the function that is constant and equal to zero on Ω. We have that

$$\begin{aligned} \{\omega \in \Omega \colon f^*(\omega) \le x\} &= \big(\{\omega \in \Omega \colon f^*(\omega) \le x\} \cap F\big) \cup (\{\omega \in \Omega \colon f^*(\omega) \le x\} \cap F^c) \\ &= \big(\{\omega \in \Omega \colon cf(\omega) \le x\} \cap F\big) \cup (\{\omega \in \Omega \colon z(\omega) \le x\} \cap F^c) \\ &= \big(\{\omega \in \Omega \colon f(\omega) \le x/c\} \cap F\big) \cup (\{\omega \in \Omega \colon z(\omega) \le x\} \cap F^c) \in \mathcal{F}, \end{aligned}$$

since f is $\mathcal{F}$ measurable, $F \in \mathcal{F}$ and z is $\mathcal{F}$ measurable. This shows that f^* is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable (cf. Proposition 7.3). Therefore, given $A \in \mathfrak{B}(\mathbb{R})$,

$$\begin{aligned} \{\omega \in \Omega \colon cf(\omega) \in A\} &= \big(\{\omega \in \Omega \colon f^*(\omega) \in A\} \cap F\big) \cup (\{\omega \in \Omega \colon cf(\omega) \in A\} \cap F^c) \\ &= \{\omega \in \Omega \colon f^*(\omega) \in A\} \cap F \in \mathcal{F}. \end{aligned}$$

This shows that $\omega \mapsto cf(\omega)$ is $\mathcal{F}$ measurable.

Solution 7.5 (Solution to Exercise 7.5). *By definition,* $f^+(\omega) = \max\{f(\omega), 0\} \geq 0$ *and* $f^-(\omega) = \max\{-f(\omega), 0\} \geq 0$. *Regarding (b), let* $\omega \in \Omega$, *then, if* $f(\omega) < 0$, $-f(\omega) > 0$, *and hence* $f^+(\omega) - f^-(\omega) = f(\omega)$. *Similarly, if* $f(\omega) > 0$, *then* $f^+(\omega) - f^-(\omega) = f(\omega)$. *If* $f(\omega) = 0$, *then* $f^+(\omega) = f^-(\omega) = 0$. *To see (c), notice that for any* $\omega \in \Omega$,

$$f^+(\omega) + f^-(\omega) = \max\{f(\omega), 0\} + \max\{-f(\omega), 0\} = \begin{cases} f(\omega), & \text{if } f(\omega) \geq 0, \\ -f(\omega), & \text{if } f(\omega) < 0. \end{cases}$$

Finally, by item (i) of Proposition 7.9, if f *is* $\mathcal{F}$ *measurable,* f^+ *and* f^- *are* $\mathcal{F}$ *measurable as well (cf. Exercise 7.4).*

7.6 Additional exercises

Exercise 7.6. *Let* $(\Omega, \mathcal{F})$ *be a measurable space and* $f\colon \Omega \to \mathbb{R}$ *be* $\mathcal{F}$ *measurable. Verify that the following functions are* $\mathcal{F}$ *measurable:*

- $\omega \mapsto \mathrm{e}^{f(\omega)}$;
- $\omega \mapsto 1/f(\omega)$ *(provided that for any* $\omega \in \Omega$, $f(\omega) \neq 0$*);*
- $\omega \mapsto \log(f(\omega))$ *(provided that for any* $\omega \in \Omega$, $f(\omega) > 0$*).*

Exercise 7.7. *Let* $(\Omega, \mathcal{F})$ *be a measurable space and* $f_1\colon \Omega \to \mathbb{R}$ *and* $f_2\colon \Omega \to \mathbb{R}$ *be two* $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ *measurable functions s.t.* $f_1(\Omega) \subset E_1$ *and* $f_2(\Omega) \subset E_2$, *where* $E_1, E_2 \in \mathfrak{B}(\mathbb{R})$. *Let* $g_1\colon E_1 \to \mathbb{R}$ *and* $g_2\colon E_2 \to \mathbb{R}$ *be continuous. Verify that the function*

$$f(\omega) = (g_1(f_1)(\omega), g_2(f_2)(\omega)), \quad \omega \in \Omega,$$

is $\mathcal{F}/\mathfrak{B}(\mathbb{R}^2)$ *measurable.*

Exercise 7.8. *Let* $(\Omega, \mathcal{F})$ *and* $(\Omega^*, \mathcal{F}^*)$ *be two measurable spaces and* $g\colon \Omega \to \Omega^*$ *be* $\mathcal{F}/\mathcal{F}^*$ *measurable. Let* μ *be a measure on* $\mathcal{F}$. *Define the function*

$$\mu g^{-1}(A^*) = \mu(g^{-1}(A^*)) = \mu(\{\omega \in \Omega\colon g(\omega) \in A^*\}), \quad A^* \in \mathcal{F}^*.$$

Show that μg^{-1} *is a measure on* $\mathcal{F}^*$.

Exercise 7.9. *Let* f *and* h *be as in Proposition 7.13. Suppose that there exists* $g\colon \mathbb{R}^k \to \mathbb{R}$ *which is* $\mathfrak{B}(\mathbb{R}^k)$ *measurable and s.t.* $h(\omega) = g(f(\omega))$. *Then,* h *is* $\sigma(f)$ *measurable.*
Hint: f *is* $\sigma(f)$ *measurable.*

Exercise 7.10. *Let* f *and* h *be as in Proposition 7.13. Suppose that* h *is a simple function in standard form (cf. Definition 7.4 with* $\mathcal{F} = \sigma(f)$*). Show that there exists* $g\colon \mathbb{R}^k \to \mathbb{R}$ *which is* $\mathfrak{B}(\mathbb{R}^k)$ *measurable and s.t.* $h(\omega) = g(f(\omega))$.
Hint: *Consider the sets* $B_i = \{\omega \in \Omega\colon h(\omega) = \alpha_i\}$, $i = 1\ldots, n$, *where* $\alpha_i \in \mathbb{R}$, $i = 1\ldots, n$, *are the* n *values that* h *can take.*

8 Integration: Part I

We rely on the following conventions regarding infinity:

$$\begin{aligned}
&x+\infty=\infty+x=\infty, \quad x-\infty=-\infty+x=-\infty, \quad x\in\mathbb{R},\\
&x\cdot\infty=\infty\cdot x=\infty, \quad x\cdot(-\infty)=(-\infty)\cdot x=-\infty, \quad x>0,\\
&0\cdot\infty=\infty\cdot 0=0,\\
&\infty\cdot\infty=\infty.
\end{aligned}$$

The omitted proofs of this chapter are found in Section B.5 of the appendix.

8.1 The integral for nonnegative functions

If $f\colon \Omega\to\overline{\mathbb{R}}$ is s.t. $f(\omega)\geq 0$ for any $\omega\in\Omega$, f is said to be nonnegative.

Definition 8.1. *Let Ω be a set. A partition of Ω is a disjoint collection $\{A\colon A\in P\}$, $P\subset\mathcal{P}(\Omega)$, s.t. $\cup_{A\in P}A=\Omega$. That is, a partition of Ω is a disjoint collection of subsets of Ω whose union is Ω. If ξ is a partition of Ω, a set $A\in\xi$ is referred to as an atom of ξ. A partition ξ of Ω is said to be finite, if it contains a finite number of atoms.*

Example 8.1. *Let $\Omega=\{0,1,\ldots,N\}$, $N\in\mathbb{N}$. Then, $\xi=\{\{\omega\}\colon\omega\in\Omega\}$ is a partition of Ω.*

Definition 8.2. *Let $(\Omega,\mathcal{F})$ be a measurable space. We use the notation $Z_0^{\mathcal{F}}(\Omega)=Z_0^{\mathcal{F}}$ for the set which contains all the finite partitions of Ω with atoms form $\mathcal{F}$. That is,*

$$Z_0^{\mathcal{F}}=\{\xi\colon \xi \text{ is a finite partition of } \Omega \text{ s.t. for any } A\in\xi,\ A\in\mathcal{F}\}.$$

Definition 8.3. *Let $(\Omega,\mathcal{F},\mu)$ be a measure space and $f\colon\Omega\to\overline{\mathbb{R}}$ be nonnegative and $\mathcal{F}$ measurable. Then, we define*

$$S_\mu^f(\xi)=\sum_{A\in\xi}\Big(\inf_{\omega\in A}f(\omega)\Big)\mu(A), \quad \xi\in Z_0^{\mathcal{F}},$$

and

$$\int_\Omega f(\omega)\mu(d\omega)=\sup_{\xi\in Z_0^{\mathcal{F}}}S_\mu^f(\xi).$$

Upon the latter definition, we deduce the integral for a (nonnegative) standard simple function (cf. Definition 7.4).

Proposition 8.1. *Let $(\Omega,\mathcal{F},\mu)$ be a measure space and $f(\omega)=\sum_{i=1}^N\alpha_i\mathbb{1}_{A_i}$, where $N\in\mathbb{N}$, $\alpha_i\in[0,\infty)$, $i=1,\ldots,N$, and $\{A_i\colon i=1,\ldots,N\}\in Z_0^{\mathcal{F}}$. That is, f is a simple function in standard form, with nonnegative coefficients α_i. We have that*

$$\int_\Omega f(\omega)\mu(d\omega)=\sum_{i=1}^N\alpha_i\mu(A_i).$$

Example 8.2. *Consider the measure space $(\mathbb{R},\mathfrak{B}(\mathbb{R}),\lambda)$, i.e., the real numbers equipped with the Borel σ-field and the Lebesgue measure (cf. Example 6.2). Let $-\infty<a=a_0<a_1<\cdots<a_N=b<\infty$ and consider the partition*

$$\xi=\big\{(-\infty,a_0]\big\}\cup\big\{(a_{i-1},a_i]\colon i=1,\ldots,N\big\}\cup\big\{(a_N,\infty)\big\}.$$

Define

$$f(x)=\begin{cases}\sum_{i=1}^N\alpha_i\mathbb{1}_{(a_{i-1},a_i]}(x), & \text{if } x\in(a,b],\\ 0, & \text{otherwise.}\end{cases}$$

Using the convention that $0 \cdot \infty = 0$ (see also Exercise 6.5), we obtain that

$$\int_{\mathbb{R}} f(x)\lambda(dx) = \sum_{i=1}^{N} \alpha_i(a_i - a_{i-1}),$$

i.e., $\int_{\mathbb{R}} f(x)\lambda(dx)$ gives the "area under the curve" of f.

Exercise 8.1. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f, g\colon \Omega \to \overline{\mathbb{R}}$ be two nonnegative and $\mathcal{F}$ measurable functions s.t. for any $\omega \in \Omega$, $f(\omega) \leq g(\omega)$. Show that $\int_\Omega f(\omega)\mu(d\omega) \leq \int_\Omega g(\omega)\mu(d\omega)$.*

The following two results are important tools in integration theory. The first is known as the monotone convergence theorem and the second shows that the integral of nonnegative functions is linear.

Proposition 8.2. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f_n\colon \Omega \to \overline{\mathbb{R}}$, $n \in \mathbb{N}$, be a sequence of nonnegative $\mathcal{F}$ measurable functions s.t. for any $\omega \in \Omega$ $f_n(\omega) \uparrow f(\omega)$ for some $f\colon \Omega \to \overline{\mathbb{R}}$. Then,*

$$\int_\Omega f_n(\omega)\mu(d\omega) \uparrow \int_\Omega f(\omega)\mu(d\omega).$$

Proposition 8.3. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $f, g\colon \Omega \to \overline{\mathbb{R}}$ be two nonnegative and $\mathcal{F}$ measurable functions. Given $\alpha, \beta \in [0, \infty)$ we have that*

$$\int_\Omega (\alpha f + \beta g)(\omega)\mu(d\omega) = \alpha \int_\Omega f(\omega)\mu(d\omega) + \beta \int_\Omega g(\omega)\mu(d\omega).$$

As a consequence of the latter two propositions we have the following result:

Proposition 8.4. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f_i\colon \Omega \to \overline{\mathbb{R}}$, $i \in \mathbb{N}$, be a sequence of nonnegative $\mathcal{F}$ measurable functions, then*

$$\int_\Omega \Big(\sum_{i\in\mathbb{N}} f_i\Big)(\omega)\mu(d\omega) = \sum_{i\in\mathbb{N}} \Big(\int_\Omega f_i(\omega)\mu(d\omega)\Big).$$

Proof. Since for any $i \in \mathbb{N}$, $f_i(\omega) \geq 0$, we have that $\sum_{i=1}^n f_i(\omega) \uparrow \sum_{i\in\mathbb{N}} f_i(\omega)$. Notice that $(\sum_{i=1}^n f_i(\omega))_{n\in\mathbb{N}}$ is a sequence of nonnegative $\mathcal{F}$ measurable functions. By Proposition 8.2,

$$\int_\Omega \Big(\sum_{i=1}^{n} f_i\Big)(\omega)\mu(d\omega) \uparrow \int_\Omega \Big(\sum_{i\in\mathbb{N}} f_i\Big)(\omega)\mu(d\omega).$$

Further, by Proposition 8.3,

$$\int_\Omega \Big(\sum_{i=1}^{n} f_i\Big)(\omega)\mu(d\omega) = \sum_{i=1}^{n} \Big(\int_\Omega f_i(\omega)\mu(d\omega)\Big) \uparrow \sum_{i\in\mathbb{N}} \Big(\int_\Omega f_i(\omega)\mu(d\omega)\Big).$$

Since a limit of a real valued sequence is unique, the result follows. □

Remark 8.1. *We remark that the monotone convergence theorem (cf. Proposition 8.2), allows for another interpretation of the integral for nonnegative functions. If $(\Omega, \mathcal{F}, \mu)$ is a measure space and $f\colon \Omega \to \overline{\mathbb{R}}$ is a nonnegative and $\mathcal{F}$ measurable function, then, we have seen in Proposition 7.10 that it is always possible to approximate f by a sequence of (nonnegative) standard simple functions, i.e., for any $\omega \in \Omega$, $f_n(\omega) \uparrow f(\omega)$ where for any $n \in \mathbb{N}$, f_n is a simple function in standard form. Hence, upon the monotone convergence theorem, we obtain,*

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

where the latter convergence is monotone. Hence, the integral for a nonnegative and $\mathcal{F}$ measurable function can be understood as the monotone limit of the integral of simple functions.

Definition 8.4. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Suppose that for any $\omega \in \Omega$, $S(\omega)$ is a statement on Ω. We say S is true μ almost everywhere (a.e.) if $\mu(\{\omega \colon S(\omega) \text{ is false}\}) = 0$.*

Example 8.3. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f \colon \Omega \to \overline{\mathbb{R}}$ be $\mathcal{F}$ measurable and nonnegative. If $f = 0$ μ a.e., then $\int_\Omega f(\omega)\mu(d\omega) = 0$. To see it take any $\xi \in Z_0^{\mathcal{F}}$. Let $A \in \xi$. Suppose that $A \cap \{\omega \colon f(\omega) = 0\} \neq \emptyset$, then $\inf_{\omega \in A} f(\omega) = 0$. Otherwise, if $A \cap \{\omega \colon f(\omega) = 0\} = \emptyset$,*

$$\mu(A) = \mu(A \cap \{\omega \colon f(\omega) \neq 0\}) \leq \mu(\{\omega \colon f(\omega) \neq 0\}) = 0.$$

Hence, $S_\mu^f(\xi) = 0$.

We can derive further properties of the integral for nonnegative functions.

Proposition 8.5. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Assume that $f, g \colon \Omega \to \overline{\mathbb{R}}$ be two nonnegative and $\mathcal{F}$ measurable functions.*

(i) If $\mu(\{\omega \colon f(\omega) > 0\}) > 0$, then $\int_\Omega f(\omega)\mu(d\omega) > 0$;

(ii) If $\int_\Omega f(\omega)\mu(d\omega) < \infty$, then $f < \infty$ μ a.e.;

(iii) If $f \leq g$ μ a.e., then $\int_\Omega f(\omega)\mu(d\omega) \leq \int_\Omega g(\omega)\mu(d\omega)$;

(iv) If $f = g$ μ a.e., then $\int_\Omega f(\omega)\mu(d\omega) = \int_\Omega g(\omega)\mu(d\omega)$.

Exercise 8.2. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f_n \colon \Omega \to \overline{\mathbb{R}}$, $n \in \mathbb{N}$, be a sequence of nonnegative $\mathcal{F}$ measurable functions s.t. $f_n \uparrow f$ μ a.e. for some $f \colon \Omega \to \overline{\mathbb{R}}$. Show that $\int_\Omega f_n(\omega)\mu(d\omega) \uparrow \int_\Omega f(\omega)\mu(d\omega)$.*

8.2 Integrable functions

We recall the definition of the positive (f^+) and negative (f^-) parts of a function (cf. Definition 7.6, recall also Exercise 7.5).

Definition 8.5. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f \colon \Omega \to \overline{\mathbb{R}}$ be a $\mathcal{F}$ measurable function. The integral of f is defined by*

$$\int_\Omega f(\omega)\mu(d\omega) = \int_\Omega f^+(\omega)\mu(d\omega) - \int_\Omega f^-(\omega)\mu(d\omega),$$

unless $\int_\Omega f^+(\omega)\mu(d\omega) = \int_\Omega f^-(\omega)\mu(d\omega) = \infty$, in which case $\int_\Omega f(\omega)\mu(d\omega)$ is not defined. If both, $\int_\Omega f^+(\omega)\mu(d\omega) < \infty$ and $\int_\Omega f^-(\omega)\mu(d\omega) < \infty$, f is said to be integrable.

Remark 8.2. *If $(\Omega, \mathcal{F}, \mu)$ is a measure space and f is as in the latter definition, then the assumption that f is integrable is defined upon the measure μ, i.e., if one wants to further refer to the measure of integration one specifies that f is integrable with respect to μ.*

Exercise 8.3. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f, g \colon \Omega \to \overline{\mathbb{R}}$ be $\mathcal{F}$ measurable. Suppose that f is integrable and $f = g$ μ a.e. Show that g is integrable and $\int_\Omega f(\omega)\mu(d\omega) = \int_\Omega g(\omega)\mu(d\omega)$.*

Another characterization of integrability is the following:

Proposition 8.6. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f \colon \Omega \to \overline{\mathbb{R}}$ be $\mathcal{F}$ measurable. Then, f is integrable if and only if $\int_\Omega |f(\omega)|\mu(d\omega) < \infty$.*

Proof. By definition, f is integrable if $\int_\Omega f^+(\omega)\mu(d\omega) < \infty$ and $\int_\Omega f^-(\omega)\mu(d\omega) < \infty$. This is equivalent to

$$\int_\Omega f^+(\omega)\mu(d\omega) + \int_\Omega f^-(\omega)\mu(d\omega) = \int_\Omega f^+(\omega) + f^-(\omega)\mu(d\omega) = \int_\Omega |f(\omega)(\omega)|\mu(d\omega) < \infty.$$

Recall also Exercise 7.5. Notice also that f integrable and $\int_\Omega |f(\omega)(\omega)|\mu(d\omega) = \infty$ gives a contradiction. Hence, if f is integrable, it must follow that $\int_\Omega |f(\omega)(\omega)|\mu(d\omega) < \infty$. $\square$

Exercise 8.4. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f, g \colon \Omega \to \overline{\mathbb{R}}$ be $\mathcal{F}$ measurable. Show that*

(a) if $|f| \leq |g|$ a.e., and g is integrable, then f is integrable as well;

(b) if $\mu(\Omega) < \infty$ and f is bounded on Ω, then f is integrable.

The following is a general version of (iii) in Proposition 8.5.

Proposition 8.7. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f, g \colon \Omega \to \overline{\mathbb{R}}$ be $\mathcal{F}$ measurable. If f and g are integrable and $f \leq g$ a.e., then, $\int_\Omega f(\omega)\mu(d\omega) \leq \int_\Omega g(\omega)\mu(d\omega)$.*

The extension of Proposition 8.3 reads as follows:

Proposition 8.8. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f, g \colon \Omega \to \overline{\mathbb{R}}$ be $\mathcal{F}$ measurable. If f and g are integrable, then for any $\alpha, \beta \in \mathbb{R}$, $\alpha f + \beta g$ is integrable and*

$$\int_\Omega (\alpha f + \beta g)(\omega)\mu(d\omega) = \alpha \int_\Omega f(\omega)\mu(d\omega) + \beta \int_\Omega g(\omega)\mu(d\omega).$$

Exercise 8.5. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $A \in \mathcal{F}$ be s.t. $\mu(A) < \infty$. Let $f = \sum_{i=1}^N \alpha_i \mathbb{1}_{A_i}$, $N \in \mathbb{N}$, $\alpha_i \in \mathbb{R}$, $i = 1, \ldots, N$, be a simple function where $\{A_i \colon i = 1, \ldots, N\} \subset \mathcal{F}$ is disjoint and $\cup_{i=1}^N A_i = A$. Show that*

$$\int_A f(\omega)\mu(d\omega) = \sum_{i=1}^N \alpha_i \mu(A_i).$$

Exercise 8.6. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f, g \colon \Omega \to \overline{\mathbb{R}}$ be $\mathcal{F}$ measurable and integrable. Show that*

$$\left| \int_\Omega f(\omega)\mu(d\omega) - \int_\Omega g(\omega)\mu(d\omega) \right| \leq \int_\Omega |f(\omega) - g(\omega)|\mu(d\omega).$$

8.3 Fatou's lemma and Lebesgue's dominated convergence theorem

The following is known as Fatou's lemma.

Proposition 8.9. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f_n \colon \Omega \to \overline{\mathbb{R}}$, $n \in \mathbb{N}$, be a sequence of nonnegative and $\mathcal{F}$ measurable functions. Then,*

$$\int_\Omega \liminf_{n\to\infty} f_n(\omega)\mu(d\omega) \leq \liminf_{n\to\infty} \int_\Omega f_n(\omega)\mu(d\omega).$$

Fatou's lemma is used to prove Lebesgue's dominated convergence (cf. Section B.5 of the appendix):

Proposition 8.10. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f_n \colon \Omega \to \overline{\mathbb{R}}$, $n \in \mathbb{N}$, be a sequence of $\mathcal{F}$ measurable functions. Suppose that there exist $f, g \colon \Omega \to \overline{\mathbb{R}}$ s.t. $f_n \to f$ a.e. and for any $n \in \mathbb{N}$, $|f_n| \leq g$ a.e. where g is integrable. Then f is integrable and*

$$\int_\Omega f_n(\omega)\mu(d\omega) \xrightarrow{n\to\infty} \int_\Omega f(\omega)\mu(d\omega).$$

In the following we discuss some applications of the latter proposition. As a first consequence, we can further extend Proposition 8.4.

Proposition 8.11. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f_i \colon \Omega \to \overline{\mathbb{R}}$, $i \in \mathbb{N}$, be a sequence of $\mathcal{F}$ measurable functions. If $\lim_{n\to\infty} \sum_{i=1}^n f_i$ exists a.e. and there exists integrable g s.t. $|\sum_{i=1}^n f_i| \leq g$ a.e., then $\sum_{i\in\mathbb{N}} f_i$ and f_i, $i \in \mathbb{N}$, are integrable and*

$$\sum_{i\in\mathbb{N}} \left(\int_\Omega f_i(\omega)\mu(d\omega) \right) = \int_\Omega \left(\sum_{i\in\mathbb{N}} f_i \right)(\omega)\mu(d\omega).$$

Proof. Since $\lim_{n\to\infty}\sum_{i=1}^n f_i$ exists a.e., we have that $\sum_{i=1}^n f_i(\omega) \xrightarrow{n\to\infty} \sum_{i\in\mathbb{N}} f_i(\omega)$ a.e. By Proposition 8.10, we can interchange limit and integration, i.e.,

$$\int_\Omega \left(\sum_{i=1}^n f_i\right)(\omega)\mu(d\omega) \xrightarrow{n\to\infty} \int_\Omega \left(\sum_{i\in\mathbb{N}} f_i\right)(\omega)\mu(d\omega).$$

In particular, $\lim_{n\to\infty}\int_\Omega(\sum_{i=1}^n f_i)(\omega)\mu(d\omega)$ exists. Hence, since

$$\lim_{n\to\infty}\int_\Omega \left(\sum_{i=1}^n f_i\right)(\omega)\mu(d\omega) = \lim_{n\to\infty}\sum_{i=1}^n \left(\int_\Omega f_i(\omega)\mu(d\omega)\right) = \sum_{i\in\mathbb{N}}\left(\int_\Omega f_i(\omega)\mu(d\omega)\right),$$

the result follows. □

Another consequence is the following result:

Proposition 8.12. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f\colon U\times\Omega \to \mathbb{R}$ be a function where $U \subset \mathbb{R}^k$. Assume that*

(i) for any $u \in U$ $\omega \mapsto f(u,\omega)$ is $\mathcal{F}$ measurable;

(ii) for any $u_0 \in U$, $u \mapsto f(u,\omega)$ is continuous in u_0 μ a.e.;

(iii) there exists a nonnegative and μ integrable $g\colon \Omega \to \mathbb{R}$ s.t. for any $u \in U$, $|f(u,\omega)| \le g(\omega)$, μ a.e.

Then, $F(u) = \int_\Omega f(u,\omega)\mu(d\omega)$, $u \in U$, is s.t. $F\colon U \to \mathbb{R}$ and F is continuous on U.

Proof. Using Exercise 8.4, (iii) implies that for any $u \in U$, $\omega \mapsto f(u,\omega)$ is μ integrable. Hence, $F\colon U \to \mathbb{R}$. Let u_n, $n \in \mathbb{N}$, be s.t. $u_n \xrightarrow{n\to\infty} u_0$, $u_0 \in U$. Then, (ii) implies that $f(u_n,\omega) \xrightarrow{n\to\infty} f(u_0,\omega)$, μ a.e. on Ω (cf. Proposition 3.26). Therefore, we apply Proposition 8.10, and conclude that $F(u_n) \xrightarrow{n\to\infty} F(u_0)$, as well. Since $u_0 \in U$ was arbitrary, F is continuous on U. □

Example 8.4. *Consider the measure space $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), \lambda)$, where λ is the Lebesgue measure on $\mathfrak{B}(\mathbb{R})$. Let $\varphi\colon \mathbb{R} \to \mathbb{R}$ be $\mathfrak{B}(\mathbb{R})$ measurable and integrable, i.e., $\int_\mathbb{R}|\varphi(x)|\lambda(dx) < \infty$. The Fourier transform $\hat\varphi$ of φ is defined as*

$$\hat\varphi(u) = \int_\mathbb{R} \mathrm{e}^{iux}\,\varphi(x)\lambda(dx),$$

where $i^2 = -1$, the imaginary number. Let $f(u,x) = \mathrm{e}^{iux}\,\varphi(x)$, $u, x \in \mathbb{R}$. Then, given any $u \in \mathbb{R}$, $x \mapsto \mathrm{e}^{iux}$ is $\mathfrak{B}(\mathbb{R})$ measurable since it is continuous on $\mathbb{R}$. Therefore, $x \mapsto \mathrm{e}^{iux}\,\varphi(x)$ is $\mathfrak{B}(\mathbb{R})$ measurable (cf. Example 7.6). Then, by the continuity of $u \mapsto \mathrm{e}^{iux}$, for any $x \in \mathbb{R}$, the function $u \mapsto f(u,x)$ is continuous on $\mathbb{R}$. Further, it follows with $|\mathrm{e}^{iux}| = 1$, that for any $u, x \in \mathbb{R}$, $|f(u,x)| \le |\varphi(x)|$, where by assumption $\int_\mathbb{R}|\varphi(x)|\lambda(dx) < \infty$. Therefore, we apply Proposition 8.12, and conclude that the Fourier transform of φ is s.t. $\hat\varphi\colon \mathbb{R} \to \mathbb{R}$ and is continuous on $\mathbb{R}$.

Example 8.5. *Consider $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), \lambda)$ as in the previous example. Let $\varphi\colon \mathbb{R} \to \mathbb{R}$ be $\mathfrak{B}(\mathbb{R})$ measurable and integrable. Suppose that $h\colon \mathbb{R} \to \mathbb{R}$ is bounded and continuous. Let*

$$h * \varphi(u) = \int_\mathbb{R} h(u-x)\varphi(x)\lambda(dx), \quad u \in \mathbb{R}.$$

Similar to the previous example, we apply Proposition 8.12 and readily see that $h * \varphi$ is continuous and bounded on $\mathbb{R}$.

8.4 Integration over measurable sets

Definition 8.6. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f\colon \Omega \to \overline{\mathbb{R}}$ be a $\mathcal{F}$ measurable function. The integral of f over a set $A \in \mathcal{F}$ is defined as*

$$\int_A f(\omega)\mu(d\omega) = \int_\Omega (\mathbb{1}_A f)(\omega)\mu(d\omega).$$

Exercise 8.7. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f\colon \Omega \to \overline{\mathbb{R}}$ be $\mathcal{F}$ measurable and either nonnegative or integrable. Show that if $\mu(A) = 0$, $A \in \mathcal{F}$, then $\int_A f(\omega)\mu(d\omega) = 0$.*

Exercise 8.8. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $F \in \mathcal{F}$. Show that*

$$\mathcal{F} \ni A \mapsto \mu_F(F \cap A),$$

is a measure on $\mathcal{F}$ and for any nonnegative and $\mathcal{F}$ measurable function $f\colon \Omega \to \overline{\mathbb{R}}$,

$$\int_\Omega f(\omega)\mu_F(d\omega) = \int_F f(\omega)\mu(d\omega).$$

Exercise 8.9. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f\colon \Omega \to \overline{\mathbb{R}}$ be a $\mathcal{F}$ measurable function. Suppose that either f is nonnegative or integrable and let $\{A_i \colon i \in I\} \subset \mathcal{F}$ be disjoint, where $I \subset \mathbb{N}$. Show that*

$$\int_{\cup_{i\in I} A_i} f(\omega)\mu(d\omega) = \sum_{i\in I} \left(\int_{A_i} f(\omega)\mu(d\omega) \right).$$

Example 8.6. *Consider the measure space $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$, where μ is the counting measure and $\mathcal{P}(\mathbb{N})$ is the power set of $\mathbb{N}$ (cf. Example 5.3). We have that for any nonnegative $\mathcal{P}(\mathbb{N})$ measurable function $f\colon \mathbb{N} \to \mathbb{R}$,*

$$\int_\mathbb{N} f(k)\mu(dk) = \sum_{k\in\mathbb{N}} f(k).$$

To see it, we define the sequence of functions

$$f_n(k) = (f\mathbb{1}_{\{0,\dots,n\}})(k) = \begin{cases} f(k), & \text{if } 0 \le k \le n, \\ 0, & \text{otherwise.} \end{cases}$$

Then, $(f_n)_{n\in\mathbb{N}}$ is a sequence of nonnegative $\mathcal{P}(\mathbb{N})$ measurable functions. Further, given any $k \in \mathbb{N}$, $f_n(k) \uparrow f(k)$. By Proposition 8.2,

$$\int_\mathbb{N} f_n(k)\mu(dk) \uparrow \int_\mathbb{N} f(k)\mu(dk).$$

We write

$$\mathbb{N} = \{1\} \cup \ldots \cup \{n\} \cup (\mathbb{N} \setminus \{1, \dots, n\}).$$

Therefore, using Exercise 8.9, we obtain,

$$\begin{aligned}
\int_\mathbb{N} f_n(k)\mu(dk) &= \int_{\{1\}} f_n(k)\mu(dk) + \cdots + \int_{\{n\}} f_n(k)\mu(dk) + \int_{\mathbb{N}\setminus\{1,\dots,n\}} f_n(k)\mu(dk) \\
&= \int_\mathbb{N} f_n(k)\mathbb{1}_{\{1\}}(k)\mu(dk) + \cdots + \int_\mathbb{N} f_n(k)\mathbb{1}_{\{n\}}(k)\mu(dk) \\
&= f_n(1)\int_\mathbb{N} \mathbb{1}_{\{1\}}\mu(dk) + \cdots + f_n(n)\int_\mathbb{N} \mathbb{1}_{\{1\}}\mu(dk) \\
&= f_n(1)\mu(\{1\}) + \cdots f_n(n)\mu(\{n\}) \\
&= \sum_{k=1}^n f_n(k) = \sum_{k=1}^n f(k).
\end{aligned}$$

Hence, $\sum_{k\in\mathbb{N}} f(k) = \int_\mathbb{N} f(k)\mu(dk)$.

Example 8.7. *Let $f\colon \mathbb{N}\times\mathbb{N}\to\mathbb{R}$, $f(i,j)=a_{ij}$ with $a_{ij}\geq 0$ for any $(i,j)\in\mathbb{N}\times\mathbb{N}$. We consider the measure space $(\mathbb{N},\mathcal{P}(\mathbb{N}),\mu)$, where μ is the counting measure and $\mathcal{P}(\mathbb{N})$ is the power set of $\mathbb{N}$ (cf. Example 5.3). Fix any $j\in\mathbb{N}$, and define the function $f_j(i)=f(i,j)$, $i\in\mathbb{N}$, i.e., $f_j\colon\mathbb{N}\to\mathbb{R}$ is a sequence of real numbers. Clearly, given any $j\in\mathbb{N}$, $f_j^{-1}(B)\in\mathcal{P}(\mathbb{N})$, $B\in\mathfrak{B}(\mathbb{R})$. That is, f_j, $j\in\mathbb{N}$, is a sequence of nonnegative $\mathcal{P}(\mathbb{N})$ measurable functions. Let $S(i)=\sum_{j\in\mathbb{N}}f_j(i)$, $i\in\mathbb{N}$. Notice that as the limit of a sequence of measurable functions $(\sum_{j=1}^n f_j(i))$, S is $\mathcal{P}(\mathbb{N})$ measurable (cf. Proposition 7.9). Using Example 8.6, we know that*

$$\int_{\mathbb{N}} S(i)\mu(di)=\sum_{i\in\mathbb{N}}S(i)=\sum_{i\in\mathbb{N}}\Big(\sum_{j\in\mathbb{N}}f_j(i)\Big)=\sum_{i\in\mathbb{N}}\Big(\sum_{j\in\mathbb{N}}a_{ij}\Big).$$

Also, by Proposition 8.4, using Example 8.6 again,

$$\int_{\mathbb{N}} S(i)\mu(di)=\sum_{j\in\mathbb{N}}\Big(\int_{\mathbb{N}}f_j(i)\mu(di)\Big)=\sum_{j\in\mathbb{N}}\Big(\sum_{i\in\mathbb{N}}f_j(i)\Big)=\sum_{j\in\mathbb{N}}\Big(\sum_{i\in\mathbb{N}}a_{ij}\Big).$$

Which shows that

$$\sum_{i\in\mathbb{N}}\Big(\sum_{j\in\mathbb{N}}a_{ij}\Big)=\sum_{j\in\mathbb{N}}\Big(\sum_{i\in\mathbb{N}}a_{ij}\Big).$$

Example 8.8. *Let $(\Omega,\mathcal{F},\mu)$ be a measure space and $f\colon\Omega\to\overline{\mathbb{R}}$ be $\mathcal{F}$ measurable and integrable. Suppose that I is a countable set and $\mu=\sum_{i\in I}\mu_i$, where μ_i is a collection of measures on $\mathcal{F}$. Then,*

$$\int_\Omega f(\omega)\mu(d\omega)=\sum_{i\in I}\Big(\int_\Omega f(\omega)\mu_i(d\omega)\Big).$$

To see it, suppose first that f is nonnegative. If $f(\omega)=\mathbb{1}_A(\omega)$, $A\in\mathcal{F}$, then

$$\int_\Omega f(\omega)\mu(d\omega)=\mu(A)=\sum_{i\in I}\mu_i(A)=\sum_{i\in I}\Big(\int_\Omega f(\omega)\mu_i(d\omega)\Big).$$

Now suppose that $f=\sum_{k=1}^N\alpha_k\mathbb{1}_{A_k}$ is a nonnegative simple function (f is assumed to be in standard from, cf. Proposition 7.7). By Proposition 8.3, we obtain:

$$\begin{aligned}\int_\Omega f(\omega)\mu(d\omega)&=\int_\Omega\Big(\sum_{k=1}^N\alpha_k\mathbb{1}_{A_k}(\omega)\Big)\mu(d\omega)=\sum_{k=1}^N\alpha_k\Big(\int_\Omega\mathbb{1}_{A_k}\mu(d\omega)\Big)\\&=\sum_{k=1}^N\alpha_k\Big(\sum_{i\in I}\mu_i(A_k)\Big)=\sum_{i\in I}\Big(\sum_{k=1}^N\alpha_k\mu_i(A_k)\Big)=\sum_{i\in I}\Big(\int_\Omega f(\omega)\mu_i(d\omega)\Big).\end{aligned}$$

Notice that we are allowed to interchange the order of summation by Proposition 3.11. If f is $\mathcal{F}$ measurable and nonnegative function, we rely on Proposition 7.10 and find a sequence of nonnegative standard simple functions $(f_n)_{n\in\mathbb{N}}$ s.t. for any $\omega\in\Omega$, $f_n(\omega)\uparrow f(\omega)$. Using the previous case, we deduce that

$$\int_\Omega f(\omega)\mu(d\omega)=\lim_{n\to\infty}\Big(\int_\Omega f_n(\omega)\mu(d\omega)\Big)=\lim_{n\to\infty}\Big(\sum_{i\in I}\Big(\int_\Omega f_n(\omega)\mu_i(d\omega)\Big)\Big).$$

Given any $n\in\mathbb{N}$, write $f_n=\sum_{k=1}^{N_n}\alpha_{k_n}\mathbb{1}_{A_{k_n}}$. Then, on the right-hand side of the latter display it reads

$$\lim_{n\to\infty}\Big(\sum_{i\in I}\Big(\sum_{k=1}^{N_n}\alpha_{k_n}\mu_i(A_{k_n})\Big)\Big)$$

Set $g_n^i = \sum_{k=1}^{N_n} \alpha_{k_n}\mu_i(A_{k_n})$, $i \in I$, $n \in \mathbb{N}$. *By Proposition 8.2, we know that for any* $i \in I$, $g_n^i \uparrow \int_\Omega f(\omega)\mu_i(d\omega)$. *Suppose that* I *is finite and for any* $i \in I$, $\int_\Omega f(\omega)\mu_i(d\omega) < \infty$. *Then, it follows from Proposition 3.4, that*

$$\lim_{n\to\infty}\Big(\sum_{i\in I} g_n^i\Big) = \sum_{i\in I}\Big(\int_\Omega f(\omega)\mu_i(d\omega)\Big), \tag{14}$$

and the result follows. If there exists $i \in I$, *s.t.* $g_n^i \uparrow \int_\Omega f(\omega)\mu_i(d\omega) = \infty$, *then, both sides of* (14) *are equal to* ∞ *and the result remains true. Thus, if* I *is finite, the statement is verified. If* I *is countably infinite,* $\#I = \#\mathbb{N}$, *and upon Example 8.7, we write* $\sum_{m=1}^\infty g_n^m = \int_\mathbb{N} g_n^m c(dm)$, *where* c *is the counting measure on* $\mathcal{P}(\mathbb{N})$. *Therefore, as a consequence of Proposition 8.2, we obtain*

$$\begin{aligned}\lim_{n\to\infty}\Big(\sum_{m=1}^\infty g_n^m\Big) &= \lim_{n\to\infty}\Big(\int_\mathbb{N} g_n^m c(dm)\Big) = \int_\mathbb{N}\Big(\int_\Omega f(\omega)\mu_m(d\omega)\Big)c(dm)\\ &= \sum_{m=1}^\infty\Big(\int_\Omega f(\omega)\mu_m(d\omega)\Big) = \sum_{i\in I}\Big(\int_\Omega f(\omega)\mu_i(d\omega)\Big).\end{aligned}$$

For the remaining case, we write $f = f^+ - f^-$ *and apply the result to* f^+ *and* f^-.

8.5 Solution to exercises

Solution 8.1 (Solution to Exercise 8.1)**.** *This follows from the fact that* $S_\mu^f(\xi) \le S_\mu^g(\xi)$ *for any* $\xi \in Z_0^\mathcal{F}$ *(cf. Definition 1.11).*

Solution 8.2 (Solution to Exercise 8.2)**.** *Let* $A = \{\omega\colon f_n(\omega) \uparrow f(\omega)\}$. *By assumption,* $\mu(A^c) = 0$. *Define* $f_n^*(\omega) = f_n(\omega)\mathbb{1}_A(\omega)$, $f^*(\omega) = f(\omega)\mathbb{1}_A(\omega)$, $\omega \in \Omega$. *Then, for any* $\omega \in \Omega$, $f_n^*(\omega) \uparrow f^*(\omega)$. *Hence, using Proposition 8.2, it follows that* $\int_\Omega f_n^*(\omega)\mu(d\omega) \uparrow \int_\Omega f^*(\omega)\mu(d\omega)$. *Further, for any* $n \in \mathbb{N}$, $f_n = f_n^*$ μ *a.e. since* $\{\omega\colon f_n \neq f_n^*\} \subset A^c$. *Similarly,* $f = f^*$ μ *a.e. Therefore, using item (iv) of Proposition 8.5,*

$$\int_\Omega f_n(\omega)\mu(d\omega) = \int_\Omega f_n^*(\omega)\mu(d\omega) \uparrow \int_\Omega f^*(\omega)\mu(d\omega) = \int_\Omega f(\omega)\mu(d\omega).$$

Solution 8.3 (Solution to Exercise 8.3)**.** *Since* f *is integrable,* $\int_\Omega f^+(\omega)\mu(d\omega) < \infty$ *and* $\int_\Omega f^-(\omega)\mu(d\omega) < \infty$. *Notice that by definition of* f^+ *and* f^-, $\{\omega\colon f^+(\omega) = g^+(\omega)\} \cap \{\omega\colon f^-(\omega) = g^-(\omega)\} = \{\omega\colon f(\omega) = g(\omega)\}$. *So that* $\{\omega\colon f(\omega) \neq g(\omega)\} = \{\omega\colon f^+(\omega) \neq g^+(\omega)\} \cup \{\omega\colon f^-(\omega) \neq g^-(\omega)\}$. *Hence,* $f^+ = g^+$ *and* $f^- = g^-$ μ *a.e. Thus, the result follows from item (iv) of Proposition 8.5.*

Solution 8.4 (Solution to Exercise 8.4)**.**

(a) First of all, $\omega \mapsto |f(\omega)|$ *and* $\omega \mapsto |g(\omega)|$ *are nonnegative and* $\mathcal{F}$ *measurable. Using item (iii) of Proposition 8.5, it follows that*

$$\int_\Omega |f(\omega)(\omega)|\mu(d\omega) \le \int_\Omega |g(\omega)(\omega)|\mu(d\omega),$$

where the latter integral is finite by Proposition 8.6. Therefore, using Proposition 8.6 again, f *is integrable.*

(b) We recall that f *bounded on* Ω *means that there exists* $0 \le M < \infty$ *s.t.* $|f(\omega)| \le M$ *for any* $\omega \in \Omega$ *(cf. Definition 2.20). Then, using the result of Exercise 8.1, we obtain*

$$\int_\Omega |f(\omega)(\omega)|\mu(d\omega) \le M\int_\Omega \mu(d\omega) = M\mu(\Omega) < \infty.$$

Solution 8.5 (Solution to Exercise 8.5). *Given any* $\omega \in \Omega$,

$$|f(\omega)| \leq |\max\{\alpha_i : i = 1, \dots, N\}|.$$

Hence, the function $\mathbb{1}_A f$ *is integrable since* $\mu(A) < \infty$. *Then, we use Proposition 8.8, and obtain*

$$\begin{aligned}\int_A f(\omega)\mu(d\omega) &= \sum_{i=1}^{n} \alpha_i \Big(\int_\Omega \mathbb{1}_A \mathbb{1}_{A_i}(\omega)\mu(d\omega) \Big) \\ &= \sum_{i=1}^{n} \alpha_i \Big(\int_\Omega \mathbb{1}_{A_i}(\omega)\mu(d\omega) \Big) \\ &= \sum_{i=1}^{n} \alpha_i \mu(A_i),\end{aligned}$$

where we used that $A_i \subset A$ *and* $\mathbb{1}_{A_i}$ *is nonnegative for any* $i = 1, \dots, N$.

Solution 8.6 (Solution to Exercise 8.6). *Let* $h \colon \Omega \to \overline{\mathbb{R}}$ *be any* $\mathcal{F}$ *measurable and integrable function. Given any* $\omega \in \Omega$, $-|h(\omega)| \leq h(\omega) \leq |h(\omega)|$. *Hence, by Proposition 8.7, we obtain* $\int_\Omega h(\omega)\mu(d\omega) \leq \int_\Omega |h(\omega)|\mu(d\omega)$ *and* $\int_\Omega -h(\omega)\mu(d\omega) \leq \int_\Omega |h(\omega)|\mu(d\omega)$. *There are two cases, either* $\int_\Omega h(\omega)\mu(d\omega) \geq 0$, *then,*

$$\Big| \int_\Omega h(\omega)\mu(d\omega) \Big| = \int_\Omega h(\omega)\mu(d\omega) \leq \int_\Omega |h(\omega)|\mu(d\omega),$$

or $\int_\Omega h(\omega)\mu(d\omega) < 0$, *then (cf. Proposition 8.3),*

$$\Big| \int_\Omega h(\omega)\mu(d\omega) \Big| = -\int_\Omega h(\omega)\mu(d\omega) = \int_\Omega -h(\omega)\mu(d\omega) \leq \int_\Omega |h(\omega)|\mu(d\omega).$$

To conclude, we set $h = f - g$.

Solution 8.7 (Solution to Exercise 8.7). *Since* $\{\omega \colon f\mathbb{1}_A \neq 0\} \subset A$, *it follows that* $f\mathbb{1}_A$ *is zero a.e. Thus, if* f *is nonnegative we rely on item (iv) of Proposition 8.5 and conclude. If* f *is integrable, we use Exercise 8.3 and arrive at the same conclusion.*

Solution 8.8 (Solution to Exercise 8.8). *It is clear that* μ_F *is a measure on* $\mathcal{F}$. *Let* $f = \mathbb{1}_A$, $A \in \mathcal{F}$. *We have that*

$$\begin{aligned}\int_\Omega f(x)\mu_F(d\omega) &= \int_A \mu_F(d\omega) \\ &= \mu_F(A) = \mu(F \cap A) = \int_{F\cap A} \mu(d\omega) = \int_\Omega \mathbb{1}_{F\cap A}(x)\mu(d\omega) \\ &= \int_\Omega \mathbb{1}_F(\omega)\mathbb{1}_A(\omega)\mu(d\omega) = \int_F f(x)\mu(d\omega).\end{aligned}$$

If $f = \sum_{i=1}^N \alpha_i \mathbb{1}_{A_i}$ *is a nonnegative simple function (*f *is assumed to be in standard from, cf. Proposition 7.7), then by Proposition 8.3,*

$$\int_\Omega f(x)\mu_F(d\omega) = \sum_{i=1}^{N} \alpha_i \Big(\int_\Omega \mathbb{1}_{A_i}(\omega)\mu_F(d\omega) \Big) = \sum_{i=1}^{N} \alpha_i \Big(\int_F \mathbb{1}_{A_i}(\omega)\mu(d\omega) \Big) = \int_F f(x)\mu(d\omega).$$

Let f *be any* $\mathcal{F}$ *measurable and nonnegative function. We rely on Proposition 7.10 and find a sequence of nonnegative standard simple functions* $(f_n)_{n\in\mathbb{N}}$ *s.t. for any* $\omega \in \Omega$, $f_n(\omega) \uparrow f(\omega)$. *By Proposition 8.2, it follows from the previous case that*

$$\int_\Omega f(x)\mu_F(d\omega) = \lim_{n\to\infty} \Big(\int_\Omega f_n(x)\mu_F(d\omega) \Big) = \lim_{n\to\infty} \int_F f_n(x)\mu(d\omega) = \int_F f(x)\mu(d\omega).$$

Solution 8.9 (Solution to Exercise 8.9). *We notice first that if I is finite, then the result follows from Proposition 8.8. Hence we assume that $I = \mathbb{N}$. Suppose that f is nonnegative. Using Definition 8.6, we have that*

$$\int_{\cup_{i\in\mathbb{N}} A_i} f(\omega)\mu(d\omega) = \int_\Omega \mathbb{1}_{\cup_{i\in\mathbb{N}} A_i}(\omega) f(\omega)\mu(d\omega) = \int_\Omega \Big(\sum_{i\in\mathbb{N}} \mathbb{1}_{A_i}(\omega) f(\omega)\Big)\mu(d\omega),$$

since $\{A_i : i \in \mathbb{N}\}$ is disjoint. Then, we use Proposition 8.4 and obtain

$$\int_{\cup_{i\in\mathbb{N}} A_i} f(\omega)\mu(d\omega) = \sum_{i\in\mathbb{N}} \Big(\int_\Omega \mathbb{1}_{A_i}(\omega) f(\omega)\mu(d\omega)\Big) = \sum_{i\in\mathbb{N}} \Big(\int_{A_i} f(\omega)\mu(d\omega)\Big).$$

Suppose now that f is integrable. Define $f_i = f\mathbb{1}_{A_i}$, $i \in \mathbb{N}$. We have that for any $\omega \in \Omega$, $\lim_{n\to\infty} \sum_{i=1}^n f_i(\omega) = f\mathbb{1}_{\cup_{i\in\mathbb{N}} A_i}(\omega)$. In particular,

$$\Big|\sum_{i=1}^n f_i(\omega)\Big| \le |f|\mathbb{1}_{\cup_{i\in\mathbb{N}} A_i}(\omega) \le \big|f(\omega)\big|.$$

Therefore, by Proposition 8.11,

$$\int_{\cup_{i\in\mathbb{N}} A_i} f(\omega)\mu(d\omega) = \sum_{i\in\mathbb{N}} \Big(\int_{A_i} f(\omega)\mu(d\omega)\Big).$$

8.6 Additional exercises

Exercise 8.10. *Let $(\Omega, \mathcal{F})$ be a measurable space and $x \in \Omega$ be given. Define the measure*

$$A \mapsto \delta_x(A) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

on $\mathcal{F}$ (cf. Example 5.1). Let $f : \Omega \to \overline{\mathbb{R}}$ be nonnegative and $\mathcal{F}$ measurable. Show that

$$\int_\Omega f(\omega)\delta_x(d\omega) = f(x).$$

Hint: *Calculate $\int_\Omega f(\omega)\delta_x(d\omega)$ if f is a nonnegative simple function in standard form and rely on Propositions 7.10 and 8.2.*
Extension: *Using positive and negative parts of f, it can be shown that it is enough to demand that $f : \Omega \to \overline{\mathbb{R}}$ is $\mathcal{F}$ measurable to obtain $\int_\Omega f(\omega)\delta_x(d\omega) = f(x)$.*

Exercise 8.11. *Calculate the integral in each of the following cases:*

(a) $\int_\mathbb{R} \big(\sum_{n\in\mathbb{N}} \mathbb{1}_{(0,1/2^n]}(x)\big)\lambda(dx)$, *where λ is the Lebesgue measure on $\mathfrak{B}(\mathbb{R})$;*

(b) $\int_\mathbb{N} x^2 \lambda(dx)$, *where λ is the Lebesgue measure on $\mathfrak{B}(\mathbb{R})$;*

(c) $\int_\mathbb{N} 1/n\, \mu(dn)$, *where μ is the counting measure on the power set of $\mathbb{N}$;*

(d) $\int_\mathbb{R} y^2 \mu(dy)$, *where μ is the measure on $\mathfrak{B}(\mathbb{R})$ given by (cf. Example 5.4),*

$$\mu(B) = \sum_{x=0}^N p_x \delta_x(B), \quad \delta_x(B) = \begin{cases} 1, & \text{if } x \in B, \\ 0, & \text{if } x \notin B, \end{cases}$$

with $N \in \mathbb{N}$, and $0 \le p_x \le 1$ for any $x \in \{1, \dots, N\}$.

Exercise 8.12. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f_n\colon \Omega \to \mathbb{R}$, $n \in \mathbb{N}$, be a sequence of $\mathcal{F}$ measurable and integrable functions s.t. $\sup_{n\in\mathbb{N}} \int_\Omega f_n(\omega)\mu(d\omega) < \infty$. Show that if for any $\omega \in \Omega$, $f_n(\omega) \uparrow f(\omega)$, then f is integrable and*

$$\int_\Omega f_n(\omega)\mu(d\omega) \uparrow \int_\Omega f(\omega)\mu(d\omega).$$

Hint: *Consider the sequence of nonnegative and $\mathcal{F}$ measurable functions $f_n - f_1$.*

Exercise 8.13. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Assume that $\mathcal{M}$ is another σ-field on Ω s.t. $\mathcal{M} \subset \mathcal{F}$ and $\mu|_{\mathcal{M}}$ is the restriction of μ to $\mathcal{M}$. Suppose that $f\colon \Omega \to \overline{\mathbb{R}}$ is $\mathcal{M}$ measurable. Show that if either f is nonnegative or integrable with respect to μ, then,*

$$\int_\Omega f(\omega)\mu(d\omega) = \int_\Omega f(\omega)\mu|_{\mathcal{M}}(d\omega).$$

Exercise 8.14. *Consider $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), \lambda)$, where λ is the Lebesgue measure on $\mathfrak{B}(\mathbb{R})$. Let $f(x) = 0$ for any $x \in \mathbb{R}$. Find a sequence of $\mathfrak{B}(\mathbb{R})$ measurable functions s.t. $f_n(x) \xrightarrow{n\to\infty} f(x)$ for any $x \in \mathbb{R}$ and $\int_\mathbb{R} f_n(x)\lambda(dx) \xrightarrow{n\to\infty} \infty$.*

9 Integration: Part II

9.1 Pushforward measure

Definition 9.1. *Let $(\Omega, \mathcal{F})$ and $(\Omega^*, \mathcal{F}^*)$ be two measurable spaces and $g\colon \Omega \to \Omega^*$ be $\mathcal{F}/\mathcal{F}^*$ measurable. Suppose that μ is a measure on $\mathcal{F}$. The measure μg^{-1} as given in Exercise 7.8 is referred to as the pushforward measure of μ.*

Proposition 9.1. *Let $(\Omega, \mathcal{F})$ and $(\Omega^*, \mathcal{F}^*)$ be two measurable spaces, $g\colon \Omega \to \Omega^*$ be $\mathcal{F}/\mathcal{F}^*$ measurable, μ be a measure on $\mathcal{F}$ and μg^{-1} be the pushforward measure of μ. Let $f\colon \Omega^* \to \overline{\mathbb{R}}$ be $\mathcal{F}^*$ measurable.*

(i) If f is nonnegative, then for any $A^* \in \mathcal{F}^*$,

$$\int_{g^{-1}(A^*)} f(g(\omega))\mu(d\omega) = \int_{A^*} f(\omega^*)\mu g^{-1}(d\omega^*); \tag{15}$$

(ii) f is integrable with respect to μg^{-1} if and only if $f(g)$ is integrable with respect to μ;

(iii) if $f(g)$ is integrable with respect to μ, then (15) holds.

Proof. Notice first that since f is $\mathcal{F}^*$ measurable and g is $\mathcal{F}/\mathcal{F}^*$, the composition $f(g)$ is $\mathcal{F}$ measurable (cf. Exercise 7.2). In order to show (i), let f be nonnegative. If $f = \mathbb{1}_{B^*}$, $B^* \in \mathcal{F}^*$, then given any $A^* \in \mathcal{F}^*$,

$$\begin{aligned}
\int_{g^{-1}(A^*)} f(g(\omega))\mu(d\omega) &= \int_{\Omega} \mathbb{1}_{g^{-1}(A^*)}(\omega)\mathbb{1}_{B^*}(g(\omega))\mu(d\omega) \\
&= \int_{\Omega} \mathbb{1}_{g^{-1}(A^*)}(\omega)\mathbb{1}_{g^{-1}(B^*)}(\omega)\mu(d\omega) \\
&= \int_{\Omega} \mathbb{1}_{g^{-1}(A^*\cap B^*)}(\omega)\mu(d\omega) \\
&= \mu(g^{-1}(A^* \cap B^*)) = \mu g^{-1}(A^* \cap B^*) = \int_{\Omega^*} \mathbb{1}_{A^*\cap B^*}(\omega^*)\mu g^{-1}(d\omega^*),
\end{aligned}$$

and thus (15) is satisfied. If $f = \sum_{i=1}^N \alpha_i \mathbb{1}_{B_i^*}$ is a nonnegative simple function (f is assumed to be in standard from, cf. Proposition 7.7), then by Proposition 8.3,

$$\begin{aligned}
\int_{g^{-1}(A^*)} f(g(\omega))\mu(d\omega) &= \int_{g^{-1}(A^*)} \Big(\sum_{i=1}^N \alpha_i \mathbb{1}_{B_i^*}\Big)(g(\omega))\mu(d\omega) \\
&= \sum_{i=1}^N \alpha_1 \Big(\int_{g^{-1}(A^*)} \mathbb{1}_{B_i^*}(g(\omega))\mu(d\omega)\Big) \\
&= \int_{A^*} f(\omega^*)\mu g^{-1}(d\omega^*).
\end{aligned}$$

where the last equality follows by the previous case. Let now f be any $\mathcal{F}^*$ measurable and nonnegative function. We rely on Proposition 7.10 and find a sequence of nonnegative standard simple functions $(f_n)_{n\in\mathbb{N}}$ s.t. for any $\omega^* \in \Omega^*$, $f_n(\omega^*) \uparrow f(\omega^*)$. By Proposition 8.2, it follows from the previous case that

$$\int_{A^*} f(\omega^*)\mu g^{-1}(d\omega^*) = \lim_{n\to\infty} \Big(\int_{g^{-1}(A^*)} f_n(g(\omega))\mu(d\omega)\Big) = \int_{g^{-1}(A^*)} f(g(\omega))\mu(d\omega).$$

This completes the proof of (i). By Proposition 8.6, f is integrable with respect to μg^{-1} if and only if

$$\int_{\Omega^*} |f(\omega^*)|\mu g^{-1}(d\omega^*) < \infty.$$

By (i), the latter integral is equla to

$$\int_{g^{-1}(\Omega^*)} |f(g(\omega))|\mu(d\omega) = \int_{\Omega} |f(g(\omega))|\mu(d\omega),$$

thus, (ii) follows. It remains to show (iii). Suppose that $f(g)$ is integrable with respect to μ, i.e., f is integrable with respect to μg^{-1}. We write $f = f^+ - f^-$ and obtain

$$\int_{A^*} f(\omega^*)\mu g^{-1}(d\omega^*) = \int_{A^*} f^+(\omega^*)\mu g^{-1}(d\omega^*) - \int_{A^*} f^-(\omega^*)\mu g^{-1}(d\omega^*).$$

Then, by (i),

$$\begin{aligned}
&\int_{A^*} f^+(\omega^*)\mu g^{-1}(d\omega^*) - \int_{A^*} f^-(\omega^*)\mu g^{-1}(d\omega^*) \\
&= \int_{g^{-1}(A^*)} f^+(g(\omega))\mu(d\omega) - \int_{g^{-1}(A^*)} f^-(g(\omega))\mu(d\omega) \\
&= \int_{g^{-1}(A^*)} f(g(\omega))\mu(d\omega).
\end{aligned}$$

□

9.2 Densities

Proposition 9.2. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $\phi\colon \Omega \to \overline{\mathbb{R}}$ be a nonnegative and $\mathcal{F}$ measurable function. Then, ν defined by*

$$\nu(A) = \int_A \phi(\omega)\mu(d\omega), \quad A \in \mathcal{F},$$

is a measure on $\mathcal{F}$.

Definition 9.2. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and ν be a measure on $\mathcal{F}$. A nonnegative and $\mathcal{F}$ measurable function $\phi\colon \Omega \to \overline{\mathbb{R}}$ is said to be a density of ν with respect to μ if for any $A \in \mathcal{F}$, $\nu(A) = \int_A \phi(\omega)\mu(d\omega)$.*

Proposition 9.3. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Suppose that ν is a measure on $\mathcal{F}$ with density ϕ with respect to μ. Then,*

(i) for any nonnegative and $\mathcal{F}$ measurable function f,

$$\int_A f(\omega)\nu(d\omega) = \int_A f(\omega)\phi(\omega)\mu(d\omega), \quad A \in \mathcal{F}; \tag{16}$$

(ii) f is integrable with respect to ν if and only if $f\phi$ is integrable with respect to μ;

(iii) if $f\phi$ is integrable with respect to μ, then (16) holds.

Proof. The proof is similar to the proof of Proposition 9.1, we leave it as an exercise (Exercise 9.6). □

Example 9.1. *If λ is the Lebesgue measure on $\mathfrak{B}(\mathbb{R})$ and f is the identity, i.e., $f(x) = x$ for any $x \in \mathbb{R}$ and s.t. $x \mapsto x\phi(x)$ is integrable with respect to λ, then if ν has density ϕ with respect to λ, we obtain*

$$\int_A x\nu(dx) = \int_A x\phi(x)\lambda(dx), \quad A \in \mathfrak{B}(\mathbb{R}).$$

9.3 Integration with respect to the Lebesgue measure on the real line

In the following we cover integration methods that relate to the Lebesgue measure. In particular, we compare the Lebesgue integral with the Riemann integral. A selection of omitted proofs are given in Section B.6 of the appendix.

Definition 9.3. *Consider the measure space $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), \lambda)$, where λ is the Lebesgue measure on the Borel σ-field $\mathfrak{B}(\mathbb{R})$. In accordance with Definition 8.5, a $\mathfrak{B}(\mathbb{R})$ measurable function $f\colon \mathbb{R} \to \overline{\mathbb{R}}$ is Lebesgue integrable if $\int_{\mathbb{R}} |f(x)| \lambda(dx) < \infty$. The integral of f with respect to λ is denoted with $\int_{\mathbb{R}} f(x)dx$, i.e., $\int_{\mathbb{R}} f(x)dx = \int_{\mathbb{R}} f(x)\lambda(dx)$. If $E \subset \mathbb{R}$ and $\lambda|_E$ is the restriction of λ to $\mathfrak{B}(E)$ (cf. Definition 4.2), then a $\mathfrak{B}(E)$ measurable function $f\colon E \to \overline{\mathbb{R}}$ is referred to as Lebesgue integrable if $\int_E |f(x)| \lambda|_E(dx) < \infty$. Also in this case we write $\int_E f(x)\lambda|_E(dx) = \int_E f(x)dx$.*

In accordance with the fact that the Lebesgue measure of a single point is zero, we adapt the following definition.

Definition 9.4. *If $f\colon E \to \overline{\mathbb{R}}$ is $\mathfrak{B}(E)$ measurable and Lebesgue integrable we adapt the following notation:*

$$\int_A f(x)dx = \int_a^b f(x)dx, \quad \text{if } A \in \mathcal{P}(E) \text{ s.t. } A \in \{[a,b],(a,b],[a,b),(a,b)\};$$
$$\int_A f(x)dx = \int_a^\infty f(x)dx \quad \text{if } A \in \mathcal{P}(E) \text{ s.t. } A \in \{(a,\infty),[a,\infty)\}, \quad a \in \mathbb{R};$$
$$\int_A f(x)dx = \int_{-\infty}^b f(x)dx \quad \text{if } A \in \mathcal{P}(E) \text{ s.t. } A \in \{(-\infty,b),(-\infty,b]\}, \quad b \in \mathbb{R}.$$

We review the definition of a Riemann integrable function:

Definition 9.5. *Let $[a,b] \subset \mathbb{R}$ be a closed interval. A function $f\colon [a,b] \to \mathbb{R}$ is called Riemann integrable with integral $I_f(a,b)$ if for any $\varepsilon > 0$ there exists $\delta > 0$ s.t. for any partition*

$$a = a_1 < a_2 < \cdots < a_{N+1} = b,$$

of $[a,b]$ with $\lambda([a_i,a_{i+1}]) < \delta$ and $x_i \in [a_i,a_{i+1}]$, $i = 1,\ldots,N$, it follows that

$$\left| I_f(a,b) - \sum_{i=1}^N f(x_i)\lambda([a_i,a_{i+1}]) \right| < \varepsilon.$$

We relate the Riemann integral with the Lebesgue integral:

Proposition 9.4. *Consider the measure space $([a,b], \mathfrak{B}([a,b]), \lambda|_{[a,b]})$, where $[a,b] \subset \mathbb{R}$ is a closed interval and $\lambda|_{[a,b]}$ is the Lebesgue measure restricted to $\mathfrak{B}([a,b])$. Assume that $f\colon [a,b] \to \mathbb{R}$ is $\mathfrak{B}([a,b])$ measurable and Riemann integrable. Then, f is Lebesgue integrable and $I_f(a,b) = \int_a^b f(x)dx$, i.e., the Riemann integral equals the Lebesgue integral of f on $[a,b]$.*

Remark 9.1. *It can be shown that if $f\colon [a,b] \to \mathbb{R}$ is continuous, then it is Riemann integrable. Notice that $f\colon [a,b] \to \mathbb{R}$ continuous, implies that f is $\mathfrak{B}([a,b])$ measurable (cf. Remark 7.1). Hence, upon the latter proposition, if $f\colon [a,b] \to \mathbb{R}$ is continuous, then it is Lebesgue integrable and $I(a,b) = \int_a^b f(x)dx$.*

With respect to improper integrals, we have the following result:

Proposition 9.5. *Let $u \in (a,\infty) \cup \{\infty\}$. Suppose that for any $a < b < u$, f restricted to $[a,b]$ is $\mathfrak{B}([a,b])$ measurable, nonnegative and Riemann integrable. Then, if the improper integral $\lim_{b\uparrow u} I_f(a,b) = I_f(a,u)$ is s.t. $I_f(a,u) \in \mathbb{R}$, $\int_a^u f(x)dx = I_f(a,u)$.*

Actually, a more general version of Proposition 9.5, with f not necessarily nonnegative, can be proven. It follows from the fact that if $f\colon [a,b] \to \mathbb{R}$ is Riemann integrable, its positive and negative parts are Riemann integrable as well.

Proposition 9.6. *Let $a, b \in \mathbb{R}$. We consider two cases,*

- *(I) Let $u \in (a, \infty) \cup \{\infty\}$. Assume that for any $a < b < u$, f restricted to $[a,b]$ is $\mathfrak{B}([a,b])$ measurable and Riemann integrable. Suppose that the improper integral $\lim_{b \uparrow u} I_f(a,b) = I_f(a,u)$ is s.t. $I_f(a,u) \in \mathbb{R}$.*
- *(II) let $u \in (-\infty, b) \cup \{-\infty\}$. Assume that for any $u < a < b$, f restricted to $[a,b]$ is $\mathfrak{B}([a,b])$ measurable and Riemann integrable. Suppose that the improper integral $\lim_{a \downarrow u} I_f(a,b) = I_f(u,b)$ is s.t. $I_f(u,b) \in \mathbb{R}$.*

We have that $\int_a^u f(x)dx = I_f(a,u)$ and $\int_u^b f(x)dx = I_f(u,b)$ in case (I) and (II), respectively.

The following result is often helpful in practice.

Proposition 9.7. *Suppose that $f\colon [a,b] \to \mathbb{R}$ is continuous with antiderivative $F\colon [a,b] \to \mathbb{R}$, i.e., $F'(x) = f(x)$ for any $x \in [a,b]$, then*

$$\int_a^b f(x)dx = [F(x)]_a^b = F(b) - F(a).$$

Example 9.2. *Let $E \subset \mathbb{R}$ s.t. $[a,b] \subset E$. Define $f = \mathbb{1}_{[a,b]}$. We can consider the measure space $(E, \mathfrak{B}(E), \lambda|_E)$ and by Proposition 8.1, $\int_E f(x)dx = \int_a^b f(x)dx = \lambda([a,b]) = b - a$. Clearly, $f|_{[a,b]}$ is continuous with antiderivative $F(x) = x$, $x \in [a,b]$. Thus,*

$$\int_a^b f|_{[a,b]}(x)dx = [x]_a^b = b - a.$$

Exercise 9.1. *Let $f(x) = \mathrm{e}^{-x}$, $x \in [0, \infty)$. Show that $\int_0^\infty \mathrm{e}^{-x}\, dx = 1$.*

Definition 9.6. *Let $f\colon [a,b] \to \mathbb{R}$ be $\mathfrak{B}([a,b])$ measurable and Lebesgue integrable. The integral of f when the limits of integration are reverted is defined as follows*

$$\int_a^b f(x)dx = -\int_b^a f(x)dx.$$

Definition 9.7. *Let $E \subset \mathbb{R}$ and $f\colon \mathbb{R} \to \mathbb{R}$ be a function. f is even if $f(-x) = f(x)$ and f is odd if $f(-x) = -f(x)$.*

Example 9.3. *The functions $x \mapsto \cos(x)$ and $x \mapsto \sin(x)$ are even and odd, respectively (cf. Example 3.7).*

Proposition 9.8. *Let $a \in (0, \infty)$ and $f\colon [-a,a] \to \mathbb{R}$ be a $\mathfrak{B}([-a,a])$ measurable and Lebesgue integrable function. Then, if f is odd,*

$$\int_{-a}^a f(x)dx = 0.$$

Proof. We have that

$$\begin{aligned}
\int_{-a}^a f(x)dx &= \int_0^a f(x)dx + \int_{-a}^0 f(x)dx \\
&= \int_0^a f(x)dx - \int_0^{-a} f(x)dx \\
&= \int_0^a f(x)dx + \int_0^a f(-x)dx \\
&= 0.
\end{aligned}$$

Notice that $\int_{-a}^{0} f(x)dx = -\int_{0}^{-a} f(x)dx$ by Definition 9.6 and $-\int_0^{-a} f(u)du = \int_0^a f(-x)dx$ upon the substitution $u = -x$. $\square$

Example 9.4. *For any $a \in (0, \infty)$, $\int_{-a}^{a} \sin(x)dx = 0$. In particular,*

$$\int_{-a}^{a} \mathrm{e}^{-\frac{x^2}{2}} \sin(x)dx = 0,$$

since $x \mapsto \mathrm{e}^{-x^2/2} \sin(x)$ is odd.

Exercise 9.2. *Let $a \in (0, \infty)$ and $f\colon [-a, a] \to \mathbb{R}$ be a $\mathfrak{B}([-a, a])$ measurable and Lebesgue integrable function. Show that if f is even,*

$$\int_{-a}^{a} f(x)dx = 2\int_0^a f(x)dx.$$

Exercise 9.3. *Show that for any $a \in (0, \infty)$, $\int_{-a}^{a} |x|dx = a^2$.*

9.4 Change of variable

Definition 9.8. *Let $U \subset \mathbb{R}^m$ be an open set and $T\colon U \to \mathbb{R}^k$ be a function. T is said to be continuously differentiable on U if for any $x_0 \in U$ the partial derivatives $\partial_{x_j} T_i(x_0)$, $j = 1, \dots, m$, $i = 1, \dots, k$, exist and are continuous on U.*

Remark 9.2. *For further details on the notion of differentiability of a function $T\colon U \to \mathbb{R}^k$, we refer to Section A.7. In particular, we note that the differential of a differentiable map $T\colon U \to \mathbb{R}^k$ in $x_0 \in U$ is represented in terms of the Jacobian matrix $J_T(x_0)$ of f in x_0 (cf. Definition A.15 and Proposition A.24).*

The following result is known as the change of variable theorem. For now, we omit a proof and rely on common references such as [1] or [2].

Proposition 9.9. *Let U and V be two open sets of $\mathbb{R}^k$. Suppose that $T\colon U \to V$ is bijective, continuously differentiable on U and s.t. $\det J_T(x_0) \neq 0$ for any $x_0 \in U$. Then, if $f\colon V \to \mathbb{R}$ is nonnegative and $\mathfrak{B}(V)/\mathfrak{B}(\mathbb{R})$ measurable we have that*

$$\int_U f(T(x))|\det J_T(x)|dx = \int_V f(y)dy. \tag{17}$$

Remark 9.3. *If f in Proposition 9.9 is $\mathfrak{B}(V)/\mathfrak{B}(\mathbb{R})$ measurable but not necessarily non-negative but s.t. both $\int_U |f(T(x)) \det J_T(x)|dx$ and $\int_V |f(y)|dy$ are finite, (17) still holds.*

As a main application, Proposition 9.9 allows to integrate in polar coordinates.

Example 9.5. *Let $T\colon U \to V$ be as in Example 2.7, i.e.,*

$$T(\rho, \theta) = (\rho\cos(\theta), \rho\sin(\theta)), \quad (\rho, \theta) \in U,$$

where $U = (0, \infty) \times (0, 2\pi)$ and $V = \mathbb{R}^2 \setminus ([0, \infty) \times \{0\})$. Clearly, U is an open subset of $\mathbb{R}^2$ (cf. Example 2.4). If we write $V = \mathbb{R}^2 \cap ([0, \infty) \times \{0\})^c$ we readily see that also V is an open subset of $\mathbb{R}^2$ (cf. Examples 2.14 and 2.19). Given $(\rho, \theta) \in U$, we have that

$$\partial_{x_1} T_1(\rho, \theta) = \cos(\theta),\ \partial_{x_2} T_1(\rho, \theta) = -\rho\sin(\theta),\ \partial_{x_1} T_2(\rho, \theta) = \sin(\theta),\ \partial_{x_2} T_2(\rho, \theta) = \rho\cos(\theta).$$

Thus, the partial derivatives exist and are all continuous on U (cf. Proposition 3.26). Hence, T is continuously differentiable on U. Also, given any $(\rho, \theta) \in U$, $\det J_T(\rho, \theta) = \rho > 0$. We have already seen in Example 2.7 that $T\colon U \to V$ is bijective. Hence, T satisfies the *assumptions of Proposition 9.9 and we deduce that for any nonnegative and $\mathfrak{B}(V)/\mathfrak{B}(\mathbb{R})$ measurable function $f\colon V \to \mathbb{R}$,*

$$\int_V f(y)dy = \int_{(0,\infty)\times(0,2\pi)} f(\rho\cos(\theta), \rho\sin(\theta))\rho d(\rho,\theta).$$

If we recall Definition 4.2, we notice that if $f\colon \mathbb{R}^2 \to \mathbb{R}$ is nonnegative and $\mathfrak{B}(\mathbb{R}^2)/\mathfrak{B}(\mathbb{R})$ measurable (i.e., a Borel function), then, the restriction $f|_V$ is $\mathfrak{B}(V)/\mathfrak{B}(\mathbb{R})$ measurable. This is because for any $A \in \mathfrak{B}(\mathbb{R})$,

$$\begin{aligned}\{y \in V\colon f|_V(y) \in A\} &= \{y \in V\colon f(y) \in A\}\\ &= \{y \in \mathbb{R}^2\colon f(y) \in A\} \cap V \in \{B \cap V\colon B \in \mathfrak{B}(\mathbb{R}^2)\},\end{aligned}$$

since f is $\mathfrak{B}(\mathbb{R}^2)/\mathfrak{B}(\mathbb{R})$ measurable. We conclude that for any nonnegative Borel function,

$$\int_V f(y)dy = \int_{(0,\infty)\times(0,2\pi)} f(\rho\cos(\theta), \rho\sin(\theta))\rho d(\rho,\theta). \tag{18}$$

In order to discover the full potential of (18), the next section allows for a further refinement of (18), i.e., we introduce tools to integrate on product spaces.

9.5 Integration on product spaces

The following covers useful results concerning the integral on product spaces. The proofs are given in the appendix (cf. Section B.7)

Definition 9.9. *Let $(X, \mathscr{X})$ and $(Y, \mathscr{Y})$ be two measurable spaces. The product σ-field on the cartesian product $X \times Y$ is defined by*

$$\mathscr{X} \otimes \mathscr{Y} = \sigma\big(\{A \times B\colon A \in \mathscr{X}, B \in \mathscr{Y}\}\big).$$

Remark 9.4. *The latter definition extends to products of higher order. Consider a collection of measure spaces $(X_1, \mathscr{X}_1), \ldots, (X_n, \mathscr{X}_n)$. We define*

$$\otimes_{i=1}^n \mathscr{X}_i = \mathscr{X}_1 \otimes \cdots \otimes \mathscr{X}_n = \sigma\big(\{A_1 \times \cdots \times A_n\colon A_i \in \mathscr{X}_i,\ i = 1, \ldots, n\}\big).$$

One can then show that the latter product is associative. If $n = 3$, that means

$$(\mathscr{X}_1 \otimes \mathscr{X}_2) \otimes \mathscr{X}_3 = \mathscr{X}_1 \otimes (\mathscr{X}_2 \otimes \mathscr{X}_3) = \mathscr{X}_1 \otimes \mathscr{X}_2 \otimes \mathscr{X}_3.$$

Example 9.6. *Assume that X and Y are countable and consider the measurable spaces $(X, \mathcal{P}(X))$ and $(Y, \mathcal{P}(Y))$, where $\mathcal{P}(X)$ and $\mathcal{P}(Y)$ are the power sets on X and Y, respectively. Then,*

$$\mathcal{P}(X \times Y) = \mathcal{P}(X) \otimes \mathcal{P}(Y).$$

It is clear that $\mathcal{P}(X) \otimes \mathcal{P}(Y) \subset \mathcal{P}(X \times Y)$ (recall that $\mathcal{P}(X \times Y)$ is the largest possible σ-field on $X \times Y$). With regard to the other inequality. Let $S \subset X \times Y$. Then, since X and Y are countable, $X \times Y$ is countable (cf. Proposition 2.8) and in particular S is countable. We write $S = \cup_{(x,y)\in S}\{(x, y)\} \in \mathcal{P}(X) \otimes \mathcal{P}(Y)$, since for any $(x, y) \in S$, $\{(x, y)\} \in \mathcal{P}(X) \otimes \mathcal{P}(Y)$. Therefore, $\mathcal{P}(X \times Y) \subset \mathcal{P}(X) \otimes \mathcal{P}(Y)$.

One of the most central results of measure theory is the following:

Proposition 9.10. *Let $(X, \mathscr{X}, \mu)$ and $(Y, \mathscr{Y}, \nu)$ be two measure spaces where μ and ν are σ-finite on $\mathscr{X}$ and $\mathscr{Y}$, respectively. Then there exists a unique σ-finite measure $\mu \otimes \nu$ on $\mathscr{X} \otimes \mathscr{Y}$ which is s.t. for any $A \in \mathscr{X}$ and $B \in \mathscr{Y}$,*

$$\mu \otimes \nu(A \times B) = \mu(A)\nu(B).$$

The measure $\mu \otimes \nu$ of the latter proposition is called the product measure on $\mathscr{X} \otimes \mathscr{Y}$.

Remark 9.5. *Suppose that $(X_1, \mathscr{X}_1), (X_2, \mathscr{X}_2)$ and $(X_3, \mathscr{X}_3)$ are three measure spaces with σ-finite measures μ_1, μ_2 and μ_3. Then, upon the previous proposition, we obtain a unique and σ-finite measure $\mu_1 \otimes \mu_2$ on $\mathscr{X}_1 \otimes \mathscr{X}_2$ s.t. for any $A_i \in \mathscr{X}_i$, $i = 1, 2$,*

$$\mu_1 \otimes \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2).$$

Then, we consider the measurable spaces $(X_1 \times X_2, \mathscr{X}_1 \otimes \mathscr{X}_2)$ and $(X_3, \mathscr{X}_3)$ and upon Proposition 9.10 again, we obtain a unique measure $(\mu_1 \otimes \mu_2) \otimes \mu_3$ on $(\mathscr{X}_1 \otimes \mathscr{X}_2) \otimes \mathscr{X}_3$ which is s.t. for any $A_i \in \mathscr{X}_i$, $i = 1, 2, 3$,

$$(\mu_1 \otimes \mu_2) \otimes \mu_3(A_1 \times A_2 \times A_3) = \mu_1 \otimes \mu_2(A_1 \times A_2)\mu_3(A_3) = \mu_1(A_1)\mu_2(A_2)\mu_3(A_3).$$

Then, since $(\mathscr{X}_1 \otimes \mathscr{X}_2) \otimes \mathscr{X}_3 = \mathscr{X}_1 \otimes \mathscr{X}_2 \otimes \mathscr{X}_3$, $(\mu_1 \otimes \mu_2) \otimes \mu_3$ is a measure on $\mathscr{X}_1 \otimes \mathscr{X}_2 \otimes \mathscr{X}_3$. We define $\mu_1 \otimes \mu_2 \otimes \mu_3 = (\mu_1 \otimes \mu_2) \otimes \mu_3$ as the product measure on $\mathscr{X}_1 \otimes \mathscr{X}_2 \otimes \mathscr{X}_3$. We remark that the latter strategy can be iterated for products of higher order. If $(X_1, \mathscr{X}_1), \ldots, (X_n, \mathscr{X}_n)$ is a collection of measurable spaces where for any $i = 1, \ldots, n$, μ_i is a σ-finite measure on $\mathscr{X}_i$, then we obtain a unique and σ-finite measure $\otimes_{i=1}^n \mu_i$ on $\otimes_{i=1}^n \mathscr{X}_i$ which is s.t. for any $A_i \in \mathscr{X}_i$, $i = 1, \ldots, n$.

$$\otimes_{i=1}^n \mu_i \left(\prod_{i=1}^n A_i \right) = \prod_{i=1}^n \mu_i(A_i).$$

Example 9.7. *Assume that X and Y are countable and consider the measurable spaces $(X, \mathcal{P}(X))$ and $(Y, \mathcal{P}(Y))$, where $\mathcal{P}(X)$ and $\mathcal{P}(Y)$ are the power sets on X and Y, respectively. If μ is the counting measure on $\mathcal{P}(X)$ and ν is the counting measure on $\mathcal{P}(Y)$, then $\mu \otimes \nu$ is the counting measure on $\mathcal{P}(X \times Y)$. To see it, let $S \in \mathcal{P}(X) \otimes \mathcal{P}(Y)$, i.e., $S \subset X \times Y$. Write m for the counting measure on $\mathcal{P}(X \times Y)$. Since S is countable, $S = \cup_{(x,y) \in S} \{(x, y)\}$. Then, we obtain*

$$\mu \otimes \nu(S) = \bigcup_{(x,y)\in S} \mu \otimes \nu(\{(x, y)\}) = \bigcup_{(x,y)\in S} \mu(\{x\})\nu(\{y\}) = \bigcup_{(x,y)\in S} m(\{(x, y)\}) = m(S).$$

Example 9.8. *Let $E \times F \subset \mathbb{R} \times \mathbb{R}$. One can show that $\mathfrak{B}(E) \otimes \mathfrak{B}(F) = \mathfrak{B}(E \times F)$, i.e., the product σ field of $\mathfrak{B}(E)$ and $\mathfrak{B}(F)$ on the cartesian product $E \times F$ is equal to the Borel σ-field on $E \times F$. This can be generalized to products of higher orders: If $E_i \subset \mathbb{R}$, $i = 1, \ldots, n$, we have that*

$$\mathfrak{B}(E_1 \times \cdots \times E_n) = \mathfrak{B}(E_1) \otimes \cdots \otimes \mathfrak{B}(E_n). \tag{19}$$

In particular, we have that

$$\underbrace{\mathfrak{B}(\mathbb{R}) \otimes \cdots \otimes \mathfrak{B}(\mathbb{R})}_{n\text{-times}} = \mathfrak{B}(\mathbb{R}^n).$$

Further, if λ is the Lebesgue measures on $\mathfrak{B}(\mathbb{R})$, then one can prove that $\lambda_2 = \lambda \otimes \lambda$, i.e., the Lebesgue measure on $\mathfrak{B}(\mathbb{R}^2)$ is equal to the product measure on $\mathfrak{B}(\mathbb{R}) \otimes \mathfrak{B}(\mathbb{R})$ (cf. Exercise 9.9). We recall that λ is σ-finite on $\mathfrak{B}(\mathbb{R})$ (cf. Definition 6.4). In general, we have that $\lambda \otimes \cdots \otimes \lambda = \lambda_n$, i.e., the n-fold product of the Lebesgue measure on $\mathfrak{B}(\mathbb{R})$ is the Lebesgue measure on $\mathfrak{B}(\mathbb{R}^n)$.

The next result is known as the Fubini-Tonnelli theorem.

Proposition 9.11. *Let μ and ν be two σ-finite measures on the measurable spaces $(X, \mathscr{X})$ and $(Y, \mathscr{Y})$, respectively. Let $f : X \times Y \to \overline{\mathbb{R}}$ be nonnegative and $\mathscr{X} \otimes \mathscr{Y}$ measurable. Then,* *(i)* *the functions*

$$X \ni x \mapsto \int_Y f(x,y)\nu(dy) \text{ and } Y \ni y \mapsto \int_X f(x,y)\mu(dx),$$

are $\mathscr{X}$ *and* $\mathscr{Y}$ *measurable, respectively;*

(ii) *we have that*

$$\begin{aligned}&\int_{X\times Y} f(x,y)\mu\otimes\nu(d(x,y))\\ &= \int_X\Big(\int_X f(x,y)\nu(dy)\Big)\mu(dx) = \int_Y\Big(\int_X f(x,y)\mu(dx)\Big)\nu(dy).\end{aligned}$$

Regarding non necessarily nonnegative functions, we have the Fubini-Lebesgue theorem.

Proposition 9.12. *Let* μ *and* ν *be two* σ*-finite measures on the measurable spaces* $(X,\mathscr{X})$ *and* $(Y,\mathscr{Y})$*, respectively. Let* $f\colon X\times Y\to\overline{\mathbb{R}}$ *be* $\mathscr{X}\otimes\mathscr{Y}$ *measurable and integrable with respect to* $\mu\otimes\nu$*, i.e.,* $\int_{X\times X}|f(x,y)|\mu\otimes\nu(d(x,y))<\infty$*. Then,*

(i) $\int_Y|f(x,y)|\nu(dy)<\infty$ μ *a.e. on* $(X,\mathscr{X})$ *and* $\int_X|f(x,y)|\mu(dx)<\infty$ ν *a.e. on* $(Y,\mathscr{Y})$*;*

(ii) *there exist sets* $F_X\in\mathscr{X}$ *and* $F_Y\in\mathscr{Y}$ *s.t.* $\mu(F_X^c)=0$ *and* $\nu(F_Y^c)=0$*, respectively, where for any* $x\in F_X$*,* $x\mapsto\int_Y f(x,y)\nu(dy)$ *is* $\mathscr{X}$ *measurable and for any* $y\in F_Y$*,* $y\mapsto\int_X f(x,y)\mu(dx)$ *is* $\mathscr{Y}$ *measurable;*

(iii) *we have that*

$$\int_X\Big|\int_Y f(x,y)\nu(dy)\Big|\mu(dx)<\infty \text{ and } \int_Y\Big|\int_X f(x,y)\mu(dx)\Big|\nu(dy)<\infty,$$

and

$$\begin{aligned}&\int_{X\times Y} f(x,y)\mu\otimes\nu(d(x,y))\\ &= \int_X\Big(\int_Y f(x,y)\nu(dy)\Big)\mu(dx) = \int_Y\Big(\int_X f(x,y)\mu(dx)\Big)\nu(dy).\end{aligned}$$

Example 9.9. *We consider the measure space* $([0,\infty),\mathfrak{B}([0,\infty)),\lambda|_{[0,\infty)})$*. Define the function*

$$f(x,y)=\mathrm{e}^{-(x+y)},\quad (x,y)\in[0,\infty)^2.$$

Clearly, f *is continuous on* $[0,\infty)^2$ *and hence* $\mathfrak{B}([0,\infty)^2)$ *measurable. Since* $\mathfrak{B}([0,\infty)^2)=\mathfrak{B}([0,\infty))\otimes\mathfrak{B}([0,\infty))$*,* f *is measurable with respect to the product* σ*-field on* $[0,\infty)^2$*. Further, for* $A_i\in\mathfrak{B}([0,\infty))$*,* $i=1,2$*,*

$$\lambda|_{[0,\infty)}\otimes\lambda|_{[0,\infty)}(A_1\times A_2)=\lambda|_{[0,\infty)}(A_1)\lambda|_{[0,\infty)}(A_2)=\lambda(A_1)\lambda(A_2).$$

Also, since $\lambda\otimes\lambda=\lambda_2$*, it follows in particular that* $(\lambda\otimes\lambda)|_{[0,\infty)^2}=\lambda_2|_{[0,\infty)^2}$ *and we obtain*

$$(\lambda\otimes\lambda)|_{[0,\infty)^2}(A_1\times A_2)=\lambda_2|_{[0,\infty)^2}(A_1\times A_2)=\lambda(A_1)\lambda(A_2).$$

Since a product measure is uniquely defined upon sets $A_i\in\mathfrak{B}([0,\infty))$*, we deduce that*

$$\lambda_2|_{[0,\infty)^2}=\lambda|_{[0,\infty)}\otimes\lambda|_{[0,\infty)}.$$

Clearly, for any $(x, y) \in [0, \infty)^2$, $f(x, y) > 0$. Hence, we are in place to apply Proposition 9.11, and conclude that

$$\begin{aligned}\int_{[0,\infty)^2} \mathrm{e}^{-(x+y)}\, d(x,y) &= \int_{[0,\infty)^2} \mathrm{e}^{-(x+y)}\, \lambda_2|_{[0,\infty)^2}(d(x,y)) \\ &= \int_{[0,\infty)\times[0,\infty)} \mathrm{e}^{-(x+y)}\, \lambda|_{[0,\infty)} \otimes \lambda|_{[0,\infty)}(d(x,y)) \\ &= \int_0^\infty \left(\int_0^\infty \mathrm{e}^{-(x+y)}\, dy \right) dx = \left(\int_0^\infty \mathrm{e}^{-x}\, dx \right)\left(\int_0^\infty \mathrm{e}^{-y}\, dy \right) = 1.\end{aligned}$$

Remark 9.6. *Let $E \times F \subset \mathbb{R}^k \times \mathbb{R}^m$. We equip the product space $E \times F$ with the product σ-field $\mathfrak{B}(E) \otimes \mathfrak{B}(F)$ and the product measure $\lambda_k|_E \otimes \lambda_m|_F$, where $\lambda_k|_E$ and $\lambda_m|_F$ represent the Lebesgue measures on $\mathbb{R}^k$ and $\mathbb{R}^m$ restricted to E and F, respectively. To further simplify the notation, if $f\colon E \times F \to \overline{\mathbb{R}}$ is $\mathfrak{B}(E) \otimes \mathfrak{B}(F)$ measurable (either nonnegative or integrable w.r.t. $\lambda_k|_E \otimes \lambda_m|_F$), we adapt the notation*

$$\int_{E\times F} f(x,y)\lambda_k|_E \otimes \lambda_m|_F(d(x,y)) = \int_{E\times F} f(x,y)d(x,y). \tag{20}$$

Notice that upon latter example (cf. Example 9.9 where $k = m = 1$), we always have

$$\lambda|_E \otimes \lambda|_F = \lambda_2|_{E\times F}.$$

Hence, in this case, the integral (20) is the usual integral of f w.r.t. the Lebesgue measure on $\mathfrak{B}(E \times F)$ (recall also Example 9.8). Of course, this can be generalized to products of higher orders: If $E_i \subset \mathbb{R}$, $i = 1, \ldots, n$, we always have that

$$\otimes_{i=1}^n \lambda|_{E_i} = \lambda_n|_{\prod_{i=1}^n E_i},$$

and (20) is the usual integral of f w.r.t. the Lebesgue measure on $\mathfrak{B}(\prod_{i=1}^n E_i)$. Finally, we keep in mind (19), i.e., $\mathfrak{B}(\prod_{i=1}^n E_i) = \otimes_{i=1}^n \mathfrak{B}(E_i)$. In particular, f is $\otimes_{i=1}^n \mathfrak{B}(E_i)$ measurable if it is $\mathfrak{B}(\prod_{i=1}^n E_i)$ measurable (which is in particular the case if it is continuous on $\prod_{i=1}^n E_i$, cf. Remark 7.1).

Upon the introduced notation given in (20), the Fubini-Tonnelli theorem (cf. Proposition 9.11) leads to the following result:

Proposition 9.13. *Let $E \times F \subset \mathbb{R}^k \times \mathbb{R}^m$. Given a nonnegative and $\mathfrak{B}(E) \otimes \mathfrak{B}(F)$ measurable $f\colon E \times F \to \overline{\mathbb{R}}$, we always have*

$$\begin{aligned}&\int_{E\times F} f(x,y)d(x,y) \\ &= \int_E \left(\int_F f(x,y)dy \right) dx = \int_F \left(\int_E f(x,y)dx \right) dy.\end{aligned}$$

Similarly, upon Proposition 9.12, we obtain:

Proposition 9.14. *Let $E \times F \subset \mathbb{R}^k \times \mathbb{R}^m$. Given a $\mathfrak{B}(E) \otimes \mathfrak{B}(F)$ measurable $f\colon E \times F \to \overline{\mathbb{R}}$ which is integrable, i.e., $\int_{E\times F} |f(x,y)| d(x,y) < \infty$, we always have*

$$\begin{aligned}&\int_{E\times F} f(x,y)d(x,y) \\ &= \int_E \left(\int_F f(x,y)dy \right) dx = \int_F \left(\int_E f(x,y)dx \right) dy.\end{aligned}$$

With respect to practical applications, the latter two proposition are primary tools for the integration (w.r.t. the Lebesgue measure) of functions defined on real coordinate spaces.

Example 9.10. *Let $f\colon [0,1]\times[0,1] \to \mathbb{R}$ be defined by $f(x,y) = x^2y^2$. By Proposition 9.13, we readily calculate:*

$$\int_{[0,1]^2} f(x,y)d(x,y) = \int_0^1 x^2\Big(\int_0^1 y^2dy\Big)dx = \Big(\int_0^1 x^2dx\Big)^2 = \Big(\frac{1}{3}\Big)^2 = \frac{1}{9}.$$

Keep in mind that f is continuous on $[0,1]^2$ and hence $\mathfrak{B}([0,1]^2) = \mathfrak{B}([0,1]) \otimes \mathfrak{B}([0,1])$ measurable.

Example 9.11. *The convolution of two $\mathfrak{B}(\mathbb{R}^k)$ measurable functions $\phi_1, \phi_2\colon \mathbb{R}^k \to \mathbb{R}$ can be defined as*

$$\phi_1 * \phi_2(z) = \mathbb{1}_A(z) \int_{\mathbb{R}^k} \phi_1(z-x)\phi_2(x)dx, \quad z \in \mathbb{R}^k,$$

where $A = \{z \in \mathbb{R}^k\colon \int_{\mathbb{R}^k}|\phi_1(z-x)\phi_2(x)|dx < \infty\}$ (cf. Example 8.5, where $A = \mathbb{R}$). We remark that if $\int_{\mathbb{R}^k}|\phi_1(x)|dx < \infty$ and $\int_{\mathbb{R}^k}|\phi_2(x)|dx < \infty$ (i.e., ϕ_1 and ϕ_2 are integrable), then $\lambda_k(A^c) = 0$. To see it, we apply Fubini-Tonnelli (cf. Proposition 9.13) and write

$$\int_{\mathbb{R}^k}\Big(\int_{\mathbb{R}^k}|\phi_1(z-x)\phi_2(x)|dx\Big)dz = \Big(\int_{\mathbb{R}^k}|\phi_1(x)|dx\Big)\Big(\int_{\mathbb{R}^k}|\phi_2(x)|dx\Big) < \infty. \tag{21}$$

We remark that it can be shown that the map $(x,z) \mapsto |\phi_1(z-x)\phi_2(x)|$ is $\mathfrak{B}(\mathbb{R}^k)\otimes\mathfrak{B}(\mathbb{R}^k)$ measurable (cf. Proposition B.7). In conclusion, by item (ii) of Proposition 8.5, $\int_{\mathbb{R}^k}|\phi_1(z-x)\phi_2(x)|dx < \infty$ λ_k a.e., i.e., $\lambda_k(A^c) = 0$. This shows that for integrable ϕ_1 and ϕ_2, we have that $\phi_1 * \phi_2(z) = \int_{\mathbb{R}^k}\phi_1(z-x)\phi_2(x)dx$ almost everywhere with respect to λ_k. Notice further that (21) and Exercise 8.6 imply that that $\int_{\mathbb{R}^k}|\phi_1 * \phi_2(z)|dz < \infty$, i.e., $\phi_1 * \phi_2$ is integrable. Hence, the convolution of two integrable functions is again integrable. For integrable functions ϕ_1, ϕ_2 and ϕ_3, we define $\phi_1 * \phi_2 * \phi_3(z) = (\phi_1 * \phi_2) * \phi_3(z)$. Again, $\phi_1 * \phi_2 * \phi_3(z) = \int_{\mathbb{R}^k}\phi_1 * \phi_2(z-x)\phi_3(x)dx$ λ_k a.e. In general, for n integrable functions $\phi_1,\dots,\phi_n\colon \mathbb{R}^k \to \mathbb{R}$, we define

$$\phi_1 * \cdots * \phi_n(z) = \underbrace{(\cdots((\phi_1 * \phi_2) * \phi_3) * \cdots * \phi_{n-2}) * \phi_{n-1})}_{=\phi_1 * \cdots * \phi_{n-1}} * \phi_n(z), \quad z \in \mathbb{R}^k, \tag{22}$$

where $\phi_1 * \cdots * \phi_n(z) = \int_{\mathbb{R}^k}\phi_1 * \cdots * \phi_{n-1}(z-x)\phi_n(x)dx$ λ_k a.e. The map $\phi_1 * \cdots * \phi_n$ is referred to as the n-fold convolution of $\phi_1,\dots,\phi_n$.

9.6 Integration in polar coordinates

We reconsider Example 9.5. In particular, we further develop equation (18). The general result reads as follows:

Proposition 9.15. *Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be a nonnegative Borel function (i.e., $\mathfrak{B}(\mathbb{R}^2)$ measurable). Then,*

$$\begin{aligned}\int_{\mathbb{R}^2} f(y)dy &= \int_0^{2\pi}\Big(\int_0^\infty f(\rho\cos(\theta),\rho\sin(\theta))\rho d\rho\Big)d\theta \\ &= \int_0^\infty\Big(\int_0^{2\pi} f(\rho\cos(\theta),\rho\sin(\theta))d\theta\Big)\rho d\rho.\end{aligned}$$

Proof. Let $T\colon U \to \mathbb{R}^2$ be defined as in Example 9.5, i.e.,

$$T(\rho,\theta) = (\rho\cos(\theta),\rho\sin(\theta)), \quad (\rho,\theta) \in U.$$

Since T is continuous on U, it is $\mathfrak{B}(U)/\mathfrak{B}(\mathbb{R}^2)$ measurable (cf. Remark 7.1). Therefore, by Exercise 7.2, since f is $\mathfrak{B}(\mathbb{R}^2)/\mathfrak{B}(\mathbb{R})$ measurable, $f \circ T\colon U \to \mathbb{R}$ is nonnegative and

$\mathfrak{B}(U)/\mathfrak{B}(\mathbb{R})$ measurable, i.e., $\mathfrak{B}((0,\infty)\times(0,2\pi))$ measurable. We apply Proposition 9.13 to the right hand side of (18) and obtain

$$\int_V f(y)dy = \int_0^{2\pi}\left(\int_0^{\infty} f(\rho\cos(\theta),\rho\sin(\theta))\rho d\rho\right)d\theta$$
$$= \int_0^{\infty}\left(\int_0^{2\pi} f(\rho\cos(\theta),\rho\sin(\theta))d\theta\right)\rho d\rho,$$

where $V = \mathbb{R}^2\setminus([0,\infty)\times\{0\})$. Since $V^c = [0,\infty)\times\{0\} = \cup_{n\in\mathbb{N}}([0,n)\times\{0\})$, we obtain (cf. Example 9.8),

$$\lambda_2(V^c) = \lim_{n\to\infty}\lambda_2([0,n)\times\{0\}) = \lim_{n\to\infty}\lambda([0,n))\lambda(\{0\}) = 0.$$

Then, by Exercise 8.7, $\int_V f(y)dy = \int_{\mathbb{R}^2} f(y)dy$ and the proposition is proven. □

Example 9.12. *Let $A = \int_{-\infty}^{\infty} \mathrm{e}^{-x^2}\,dx$. By Proposition 9.13 and Proposition 9.15,*

$$A^2 = \int_{\mathbb{R}^2} \mathrm{e}^{-(x^2+y^2)}\,d(x,y) = \int_0^{\infty}\mathrm{e}^{-\rho^2}\left(\int_0^{2\pi} d\theta\right)\rho d\rho = 2\pi\int_0^{\infty}\mathrm{e}^{-\rho^2}\rho d\rho.$$

Then, upon the substitution $u=\rho^2$,

$$\int_0^{\infty}\mathrm{e}^{-\rho^2}\rho d\rho = 1/2.$$

Hence, $A=\sqrt{\pi}$.

Exercise 9.4. *Let $D_r(0) = B_r[0]\subset\mathbb{R}^2$ be the closed ball (i.e. disk) with radius r and center at the origin. Show that $\lambda_2(D_r(0)) = \pi r^2$.*

9.7 Solution to exercises

Solution 9.1 (Solution to Exercise 9.1)**.** *We have that f is continuous with antiderivative $F(x) = -\mathrm{e}^{-x}$. Therefore, we obtain,*

$$\int_0^{\infty}\mathrm{e}^{-x}\,dx = \lim_{b\uparrow\infty}\int_0^b \mathrm{e}^{-x}\,dx = \lim_{b\uparrow\infty}[-\mathrm{e}^{-x}]_0^b = 0-(-1) = 1,$$

since $\lim_{b\uparrow\infty}(-\mathrm{e}^{-b}) = 0$.

Solution 9.2 (Solution to Exercise 9.2)**.** *We have that*

$$\int_{-a}^{a} f(x)dx = \int_{-a}^{0} f(x)dx + \int_0^a f(x)dx.$$

Hence, it is sufficient to show that $\int_{-a}^0 f(x)dx = \int_0^a f(x)dx$. Using the substitution $u(x) = -x$, we obtain that

$$\int_{-a}^{0} f(x)dx = -\int_{u(-a)}^{u(0)} f(-u)du = -\int_a^0 f(-u)du = \int_0^a f(-u)du = \int_0^a f(u)du.$$

Solution 9.3 (Solution to Exercise 9.3)**.** *Since $x\mapsto|x|$ is even, we rely on Exercise 9.2 and obtain $\int_{-a}^a |x|dx = 2\int_0^a x dx = a^2$.*

Solution 9.4 (Solution to Exercise 9.4)**.** *By Proposition 9.15, we have that*

$$\lambda_2(D_r(0)) = \int_{\mathbb{R}^2}\mathbb{1}_{D_r(0)}(y)dy = \int_0^r\rho\Big(\int_0^{2\pi}d\theta\Big)d\rho = \pi r^2.$$

9.8 Additional exercises

Exercise 9.5. *Prove Proposition 9.2.*

Exercise 9.6. *Prove Proposition 9.3.*

Exercise 9.7. *Calculate the following integrals:*

(a) $\int_0^\infty x\,\mathrm{e}^{-x}\,dx$*;*

(b) $\frac{1}{2\pi}\int_{\mathbb{R}^2}\mathrm{e}^{-(\frac{x^2}{2}+\frac{y^2}{2})}\,d(x,y)$*;*

(c) $\int_{[0,1]^2} xy\mathbb{1}_{g^{-1}((-\infty,0])}(x,y)d(x,y)$*, with* $g(a,b) = b - a$*,* $a, b \in \mathbb{R}$*.*

Exercise 9.8. *Let* $(X_1, \mathscr{X}_1)$*,* $(X_2, \mathscr{X}_2)$*, and* $(X_3, \mathscr{X}_3)$ *be measurable spaces. Show that*

$$(\mathscr{X}_1 \otimes \mathscr{X}_2) \otimes \mathscr{X}_3 = \mathscr{X}_1 \otimes \mathscr{X}_2 \otimes \mathscr{X}_3.$$

Exercise 9.9. *Show that* $\lambda_2 = \lambda \otimes \lambda$*, i.e., the product of the Lebesgue measure on* $\mathfrak{B}(\mathbb{R})$ *equals the Lebesgue measure on* $\mathfrak{B}(\mathbb{R}^2)$*.*

10 General notions in Probability

A selection of omitted proofs of this chapter is found in Appendix C.

10.1 Probability spaces

Definition 10.1. *Let $(\Omega, \mathcal{F})$ be a measurable space. A probability $\mathbb{P}$ on $\mathcal{F}$ is a measure on $\mathcal{F}$ s.t. $\mathbb{P}(\Omega) = 1$. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is a referred to as a probability space.*

Example 10.1. *Let Ω be a finite and nonempty set. Define*

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega}, \quad A \in \mathcal{P}(\Omega), \tag{23}$$

where $\mathcal{P}(\Omega)$ is the power set on Ω. Then, $\mathbb{P}$ is a probability on $\mathcal{P}(\Omega)$ (cf. Example 5.2).

Example 10.2. *Let C be a set s.t. $\#C = 52$. Suppose that*

$$C = S_1 \cup S_2 \cup S_3 \cup S_4,$$

with $\{S_1, S_2, S_3, S_4\}$ disjoint and s.t. $\#S_i = 13$ for any $i = 1, 2, 3, 4$. We remain in the setting of the previous example with

$$\Omega = \{A \subset C \colon \#A = 5\},$$

and $\mathbb{P}$ on $\mathcal{P}(\Omega)$ defined as in (23). Upon Exercise 1.11, we already know that $\#\Omega = \binom{52}{5}$. Let

$$A_i = \{A \subset S_i \colon \#A = 5\}, \quad i = 1, 2, 3, 4,$$

and define $A = A_1 \cup A_2 \cup A_3 \cup A_4$. Since $\{A_1, A_2, A_3, A_4\} \subset \mathcal{P}(\Omega)$ is disjoint, it follows that

$$\mathbb{P}(A) = \frac{4\binom{13}{5}}{\binom{52}{5}}.$$

If we interpret C as the collection of 52 poker cards with suits defined upon S_1, S_2, S_3 and S_4, $\mathbb{P}(A)$ is the probability of a flush, i.e., obtaining a poker hand which consists only of cards of the same suit.

Exercise 10.1. *Let $\Omega = \{(\omega_1, \omega_2) \colon \omega_1, \omega_2 \in \{1, 2, 3, 4, 5, 6\}\}$ and define $\mathbb{P}$ on $\mathcal{P}(\Omega)$ as in (23). Let $A = \{(\omega_1, \omega_2) \in \Omega \colon \omega_2 > \omega_1\}$. Calculate $\mathbb{P}(A)$.*

Example 10.3. *Consider the measurable space $([0, 1], \mathfrak{B}([0, 1]))$, where $\mathfrak{B}([0, 1])$ is the Borel σ-field on $[0, 1]$. Then, $\lambda|_{[0,1]}$, the restriction of the Lebesgue measure to $\mathfrak{B}([0, 1])$ is a probability on $\mathfrak{B}([0, 1])$. Similarly, the measure*

$$\mathbb{P}(A) = \frac{\lambda|_{[a,b]}(A)}{\lambda|_{[a,b]}([a, b])}, \quad A \in \mathfrak{B}([a, b]),$$

is a probability on $\mathfrak{B}([a, b])$, $a < b$, $a, b \in \mathbb{R}$.

10.2 Random variables and random vectors

Definition 10.2. *Let $(\Omega, \mathcal{F})$ be a measurable space. A map $X \colon \Omega \to \mathbb{R}$ is referred to as a random variable on $(\Omega, \mathcal{F})$ if it is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable.*

Example 10.4. *Any continuous map $f \colon [a, b] \to \mathbb{R}$ is a random variable on $([a, b], \mathfrak{B}([a, b]))$ (cf. Remark 7.1). In particular, the identity map $g(x) = x$, $x \in [a, b]$, is a random variable on $([a, b], \mathfrak{B}([a, b]))$.*

Example 10.5. *Let* $\Omega = \{(i,j): i,j \in \{1,2,3,4,5,6\}\}$*, then the map* $X((i,j)) = \min\{i,j\}$*,* $(i,j) \in \Omega$*, is a random variable on* $(\Omega, \mathcal{P}(\Omega))$*.*

Example 10.6. *If* X_i*,* $i = 1,\dots,n$*, are s.t.* X_i *is a random variable on* $(\Omega, \mathcal{F})$ *for any* $i = 1,\dots,n$*, then* $\sum_{i=1}^n X_i$ *is a random variable on* $(\Omega, \mathcal{F})$ *(cf. Proposition 7.12).*

Definition 10.3. *Let* $(\Omega, \mathcal{F})$ *be a measurable space. A map* $X\colon \Omega \to \mathbb{R}^k$ *is referred to as a random vector on* $(\Omega, \mathcal{F})$ *if it is* $\mathcal{F}/\mathfrak{B}(\mathbb{R}^k)$ *measurable.*

Remark 10.1. *By Proposition 7.4,* $X = (X_1,\dots,X_k)$ *is a random vector on* $(\Omega, \mathcal{F})$ *if and only if* X_i *is a random variables on* $(\Omega, \mathcal{F})$ *for any* $i = 1,\dots,k$*. In particular, for* $k = 1$*,* X *is a random variable. In what follows, if no distinction is needed, our results are primed for random vectors and the respective result for random variables follows by considering the case where* $k = 1$*.*

Example 10.7. *Let* $\Omega = \{(i,j): i,j \in \{1,2,3,4,5,6\}\}$*. Given* $(i,j) \in \Omega$*, define* $X_1((i,j)) = \min\{i,j\}$ *and* $X_2((i,j)) = \max\{i,j\}$*. Then,* $X = (X_1, X_2)$ *is a random vector on* $(\Omega, \mathcal{P}(\Omega))$*.*

Example 10.8. *Let* $X = (X_1,\dots,X_k)$ *be a random vector on* $(\Omega, \mathcal{F})$ *and* $g\colon \mathbb{R}^k \to \mathbb{R}$ *be* $\mathfrak{B}(\mathbb{R}^k)/\mathfrak{B}(\mathbb{R})$ *measurable. Then,* $g(X)$ *is a random variable on* $(\Omega, \mathcal{F})$*. In particular, if* g *is continuous,* $g(X)$ *is a random variable on* $(\Omega, \mathcal{F})$*.*

We note that a direct consequence of Proposition 7.13 is the following result.

Proposition 10.1. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and* X *be a random vector on* $(\Omega, \mathcal{F})$*. A random variable* Y *on* $(\Omega, \mathcal{F})$ *is* $\sigma(X)$ *measurable if and only if there exists a function* $f\colon \mathbb{R}^k \to \mathbb{R}$ *which is* $\mathfrak{B}(\mathbb{R}^k)$ *measurable and s.t.* $Y = f(X)$*.*

Definition 10.4. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. The distribution or law of a random vector* X *on* $(\Omega, \mathcal{F})$ *is the pushforward measure* $P_X = \mathbb{P}X^{-1}$ *on* $\mathfrak{B}(\mathbb{R}^k)$ *(cf. Definition 9.1). In particular, for any* $B \in \mathfrak{B}(\mathbb{R}^k)$ *we use the simplified notation*

$$\{\omega \in \Omega\colon X(\omega) \in B\} = \{X \in B\},$$

and hence

$$P_X(B) = \mathbb{P}(X \in B).$$

For now, unless mentioned otherwise, if $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space any random vector X is a random vector on $(\Omega, \mathcal{F})$, i.e., a $\mathcal{F}$ measurable functions with values in $\mathbb{R}^k$.

10.3 Discrete laws

Definition 10.5. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. A random vector is referred to as discrete if there exists a countable set* $E = E_1 \times \cdots \times E_k \subset \mathbb{R}^k$ *s.t.* $P_X(E) = 1$*. That is to say that the law of* X *has a countable support.*

Proposition 10.2. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. A random vector* X *is discrete if and only if*

$$P_X = \sum_{x \in E} p_x \delta_x, \quad p_x = \mathbb{P}(X = x), \tag{24}$$

for some countable set $E = E_1 \times \cdots \times E_k \subset \mathbb{R}^k$*. In particular, for any* $B \in \mathfrak{B}(\mathbb{R}^k)$*,* $P_X(B) = \sum_{x \in B \cap E} p_x$*.*

Remark 10.2. *The latter proposition shows that the law of a discrete random vector* X *with support* E *is determined by* $\mathbb{P}(X = x)$*,* $x \in E$*. Notice that the measure* P_X *given in (24) was already introduced in the more general setting of Example 5.4.*

Proof of Proposition 10.2. Suppose that X is discrete. Let $B \in \mathfrak{B}(\mathbb{R}^k)$. We have that

$$P_X(B) = P_X(B \cap E) = \mathbb{P}(X \in B \cap E) = \mathbb{P}\Big(\bigcup_{x \in B \cap E} \{X = x\}\Big) = \sum_{x \in B \cap E} p_x = \sum_{x \in E} p_x \delta_x(B).$$

With respect to the other direction, if P_X is given by (24), then

$$1 = P_X(\mathbb{R}^k) = \sum_{x \in E} p_x \delta_x(\mathbb{R}^k) = \sum_{x \in E} p_x = \mathbb{P}(X \in E) = P_X(E),$$

i.e., X is a discrete random vector according to Definition 10.5. $\square$

Example 10.9. *Let $\Omega = \{t, h\}$ and*

$$X(\omega) = \begin{cases} 0, & \text{if } \omega = t, \\ 1, & \text{if } \omega = h. \end{cases}$$

Then, X is a random variable on $(\Omega, \mathcal{P}(\Omega))$. Suppose that $\mathbb{P}$ is a probability on $\mathcal{P}(\Omega)$ s.t. $\mathbb{P}(X = 0) = 1 - p$ and $\mathbb{P}(X = 1) = p$. Clearly, $P_X(\{0, 1\}) = 1$. By Proposition 10.2, we deduce that the law of X is given by

$$P_X = (1 - p)\delta_0 + p\delta_1.$$

That is, for any $B \in \mathfrak{B}(\mathbb{R})$,

$$P_X(B) = \begin{cases} 0, & \text{if } 0 \notin B \text{ and } 1 \notin B, \\ 1 - p, & \text{if } 0 \in B \text{ and } 1 \notin B, \\ p, & \text{if } 0 \notin B \text{ and } 1 \in B, \\ 1, & \text{if } 0 \in B \text{ and } 1 \in B. \end{cases}$$

Notice that P_X was already introduced in Exercise 5.2.

Example 10.10. *Let $\emptyset \neq E \subset \mathbb{R}^k$ be a countable set. Generally, a measure*

$$P = \sum_{x \in E} p_x \delta_x, \quad p_x \geq 0,$$

defined on $\mathfrak{B}(\mathbb{R}^k)$ which satisfies $\sum_{x \in E} p_x = 1$, is referred to as a discrete probability distribution. In the following we list some classical examples.

Discrete uniform: *$E \subset \mathbb{R}$ is a finite set s.t. $\#E = n$ and $p_x = 1/n$ for any $x \in E$;*

Bernoulli: *$E = \{0, 1\}$ and $p_0 = 1 - p$ and $p_1 = p$, $p \in [0, 1]$;*

Binomial: *$E = \{0, 1, \ldots, n\}$, $n \in \mathbb{N}$ and $p_x = \binom{n}{x} p^x (1 - p)^{n - x}$, $p \in [0, 1]$;*

Geometric: *$E = \mathbb{N}$ and $p_x = (1 - p)^{x - 1} p$, $p \in (0, 1)$;*

Poisson: *$E = \mathbb{N} \cup \{0\}$ and $p_x = (\lambda^x / x!)\, \mathrm{e}^{-\lambda}$, $\lambda > 0$;*

Multinomial *$E = \{(x_1, \ldots, x_k) \in \{0, \ldots, N\}^k \colon \sum_{i=1}^k x_i = N\}$, $N \in \mathbb{N}$, and*

$$p_{(x_1, \ldots, x_k)} = \left(\frac{N!}{\prod_{i=1}^k (x_i!)}\right) p_1^{x_1} \cdot \ldots \cdot p_k^{x_k}, \quad \sum_{i=1}^k p_k = 1,$$

with $p_i \in [0, 1]$ for any $i = 1, \ldots, k$

Exercise 10.2. *Verify that $\sum_{x=0}^n p_x = 1$, $p_x = \binom{n}{x} p^x (1 - p)^{n - x}$, $p \in [0, 1]$.*

Exercise 10.3. *Verify that* $\sum_{x=0}^{\infty} p_x = 1$*,* $p_x = (\lambda^x/x!)\,\mathrm{e}^{-\lambda}$*,* $\lambda > 0$*.*

Remark 10.3. *If* $X = (X_1, \ldots, X_k) \colon \Omega \to \mathbb{R}^k$ *is discrete with support* $E = E_1 \times \cdots \times E_k$*, we apply Proposition 10.2 and deduce that for any* $i = 1, \ldots, k$*,*

$$\begin{aligned}
\mathbb{P}(X_i = x) &= \mathbb{P}(X_1 \in \mathbb{R}, \ldots, X_{i-1} \in \mathbb{R}, X_i = x, X_{i+1} \in \mathbb{R}, \ldots, X_k \in \mathbb{R}) \\
&= P_X(\mathbb{R} \times \cdots \times \mathbb{R} \times \{x\} \times \mathbb{R} \times \cdots \times \mathbb{R}) \\
&= \sum_{\substack{(x_1,\ldots,x_k)\in E \\ x_i = x}} p_{x_1,\ldots,x_k}.
\end{aligned}$$

Given $i = 1, \ldots, k$*, we apply the notation,*

$$x_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k),$$

and

$$E_{-i} = E_1 \times \cdots \times E_{i-1} \times E_{i+1} \cdots \times E_k.$$

Then, we obtain

$$\mathbb{P}(X_i = x) = \sum_{x_{-i} \in E_{i-1}} p_{x_1,\ldots,x_{i-1},x,x_{i+1},\ldots,x_k}. \tag{25}$$

Notice that the sum in (25) is zero if $x \notin E_i$*. Thus, for any* $i = 1, \ldots, k$*,* X_i *has support* E_i*. Further, since the law of* X_i *is determined by* E_i*,* $i = 1, \ldots, k$ *(cf. Remark 10.2), it is sufficient to compute (25) for* $x \in E_i$*. If* $n = 2$*, (25) gives,*

$$\begin{aligned}
\mathbb{P}(X_1 = x) &= \sum_{x_2 \in E_2} \mathbb{P}(\{X_1 = x\} \cap \{X_2 = x_2\}), \quad x \in E_1; \\
\mathbb{P}(X_2 = y) &= \sum_{x_1 \in E_1} \mathbb{P}(\{X_1 = x_1\} \cap \{X_2 = y\}), \quad y \in E_2.
\end{aligned}$$

A formal treatment on how to compute the sum (25) for general n *is given in Section B.8 of the appendix (cf. Proposition B.11).*

10.4 Continuous laws

Definition 10.6. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. A random vector is referred to as continuous if the law of* X *has density* $\phi \colon \mathbb{R}^k \to [0, \infty)$ *with respect to the Lebesgue measure on* $\mathfrak{B}(\mathbb{R}^k)$ *(cf. Definition 9.2). That is, for any* $B \in \mathfrak{B}(\mathbb{R}^k)$*,*

$$P_X(B) = \int_B \phi(x)dx.$$

The density ϕ *of* P_X *is referred to as a probability density function.*

Remark 10.4. *Recall that by definition of a density, a probability density function is* $\mathfrak{B}(\mathbb{R}^k)$ *measurable. Notice that if* X *is a continuous random variable with probability density function* ϕ*, then, for any* $a < b$*,* $\mathbb{P}(a \leq X \leq b) = \int_a^b \phi(x)dx$*.*

Example 10.11. *Suppose that* $X = (X_1, X_2)$ *is a random vector with probability density function*

$$\phi(x_1, x_2) = \begin{cases} \mathrm{e}^{-(x_1+x_2)}, & \text{if } (x_1, x_2) \in [0, \infty)^2, \\ 0, & \text{otherwise.} \end{cases}$$

By Proposition 9.13 (take $E = F = \mathbb{R}$, recall also Example 9.9 and that $\mathfrak{B}(\mathbb{R}^2) = \mathfrak{B}(\mathbb{R}) \otimes \mathfrak{B}(\mathbb{R})$), we obtain that for any $A \times B$, $A \in \mathfrak{B}(\mathbb{R})$ and $B \in \mathfrak{B}(\mathbb{R})$,

$$P_X(A \times B) = \mathbb{P}(X_1, X_2) \in A \times B) = \Big(\int_A \mathrm{e}^{-x_1}\, dx_1\Big)\Big(\int_B \mathrm{e}^{-x_2}\, dx_2\Big). \tag{26}$$

Therefore, if we set $P_{X_1}(C) = P_{X_2}(C) = \int_C \mathrm{e}^{-x}\, dx$, $C \in \mathfrak{B}(\mathbb{R})$, we deduce that

$$P_X(A \times B) = P_{X_1}(A)P_{X_2}(B) = P_{X_1}(A)P_{X_1}(B).$$

Example 10.12. *Generally, the map $B \mapsto P(B) = \int_B \phi(x)dx$, $\phi\colon \mathbb{R}^k \to [0, \infty)$ is a measure on $\mathfrak{B}(\mathbb{R}^k)$ (cf. Proposition 9.2) and if in addition $\int_{\mathbb{R}} \phi(x)dx = 1$, P is referred to as a probability distribution with probability density function ϕ. We use the notation $P(dx) = \phi(x)dx$ to indicate that for any $B \in \mathfrak{B}(\mathbb{R}^k)$, $P(B) = \int_B \phi(x)dx$. In the following we give some classical examples of probability distributions with probability density function ϕ.*

Continuous uniform: *Given $a, b \in \mathbb{R}$, $a < b$,*

$$\phi(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x), \quad x \in \mathbb{R};$$

Exponential: *Given $\lambda > 0$,*

$$\phi(x) = \lambda\, \mathrm{e}^{-\lambda x}\, \mathbb{1}_{[0,\infty)}(x), \quad x \in \mathbb{R};$$

Normal: *Given $\mu \in \mathbb{R}$ and $\sigma \in (0, \infty)$,*

$$\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R};$$

Multivariate Normal: *Given $\Sigma \in \mathbb{R}^{k\times k}$, positive definite and symmetric, and $\mu \in \mathbb{R}^k$,*

$$\phi(x) = \frac{1}{\sqrt{(2\pi)^k \det \Sigma}}\, \mathrm{e}^{-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)}, \quad x \in \mathbb{R}^k. \tag{27}$$

Definition 10.7. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Suppose that for any $\omega \in \Omega$, $S(\omega)$ is a statement on Ω. We say S is true $\mathbb{P}$ almost surely (a.s.) if $\mathbb{P}(\{\omega\colon S(\omega)$ is true$\}) = 1$ (cf. Definition 8.4).*

Definition 10.8. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X be a random variable. Suppose that $\mu \in \mathbb{R}$ and $\sigma \geq 0$. Then, $X \sim \mathcal{N}(\mu, \sigma^2)$ is to say that X is Gaussian (or normal) with mean μ and variance σ^2. If $\sigma > 0$, $X \sim \mathcal{N}(\mu, \sigma^2)$ indicates that the law of X has probability density function*

$$\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}. \tag{28}$$

If $\sigma = 0$, $X \sim \mathcal{N}(\mu, 0)$ indicates that $\mathbb{P}(X = \mu) = 1$, i.e., X is constant and equal to μ $\mathbb{P}$ a.s. If $X \sim \mathcal{N}(0, 1)$, X is said to be standard Gaussian (or normal).

Remark 10.5. *Let $\mu \in \mathbb{R}$ and $\sigma > 0$. If we substitute $u = (x - \mu)/(\sqrt{2}\sigma)$, we obtain with Example 9.12,*

$$\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \sqrt{2}\sigma \int_{-\infty}^{\infty} \mathrm{e}^{-u^2}\, du = \sqrt{2}\sigma\sqrt{\pi} = \sqrt{2\pi\sigma^2}.$$

10.5 Expectation

Definition 10.9. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X be a random variable. In either case*

- *X is nonnegative;*
- *X is integrable with respect to $\mathbb{P}$;*

the expectation of X is defined by

$$\mathbb{E}[X] = \int_{\Omega} X(\omega)\mathbb{P}(d\omega).$$

If $X = (X_1, \ldots, X_k)$ is a random vector, $\mathbb{E}[X] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_k])$ is defined if $\mathbb{E}[X_i]$ is defined for any $i = 1, \ldots, k$.

Proposition 10.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X be a random vector. Then, for any nonnegative and $\mathfrak{B}(\mathbb{R}^k)$ measurable map $f\colon \mathbb{R}^k \to \overline{\mathbb{R}}$,*

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}^k} f(x)P_X(dx). \tag{29}$$

In addition, if f is not necessarily nonnegative, (29) is also satisfied if $\mathbb{E}[|f(X)|] < \infty$.

Proof. Since $f(X)\colon \Omega \to \overline{\mathbb{R}}$ is nonnegative and $\mathcal{F}$ measurable, it follows that $\mathbb{E}[f(X)]$ is well defined. Using Proposition 9.1 with $g = X$, we obtain that

$$\mathbb{E}[f(X)] = \int_{\Omega} f(X(\omega))\mathbb{P}(d\omega) = \int_{\mathbb{R}^k} f(x)P_X(dx).$$

For the case where $\mathbb{E}[|f(X)|] < \infty$, the result also follows from Proposition 9.1. □

Example 10.13. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Suppose that X is a discrete random variable with support E and s.t. either*

(i) for any $x \in E$, $x \geq 0$, i.e., $\mathbb{P}(X \geq 0) = 1$, that is $X \geq 0$ $\mathbb{P}$ a.s.;

or

(ii) $\mathbb{E}[|X|] < \infty$.

Then, in case (i) $\mathbb{E}[X]$ is well defined and we have that $X = X\mathbb{1}_{[0,\infty)}$ $\mathbb{P}$ a.s. We apply Proposition 10.3 with $f(x) = x\mathbb{1}_{[0,\infty)}(x)$ and deduce that

$$\mathbb{E}[X] = \mathbb{E}[f(X)] = \int_0^{\infty} xP_X(dx) = \sum_{y\in E}\left(\int_0^{\infty} xp_y\delta_y(dx)\right) = \sum_{y\in E} yp_y = \sum_{x\in E} x\mathbb{P}(X = x).$$

Keep in mind Example 8.8, Exercise 8.10 and Proposition 10.2. In case (ii), $\mathbb{E}[X]$ is well defined as well and we apply Proposition 10.3 with $f\colon \mathbb{R} \to \mathbb{R}$ be the identity map, i.e., $f(x) = x$, $x \in \mathbb{R}$ and obtain

$$\mathbb{E}[X] = \int_{\mathbb{R}} xP_X(dx) = \sum_{y\in E}\left(\int_{\mathbb{R}} xp_y\delta_y(dx)\right) = \sum_{y\in E} yp_y = \sum_{x\in E} x\mathbb{P}(X = x).$$

Example 10.14. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Suppose that X is a continuous random variable with probability density function ϕ. Suppose that either*

(i) $X \geq 0$ $\mathbb{P}$ a.s.;

or

(ii) $\mathbb{E}[|X|] < \infty$.

In case (i), $\mathbb{E}[X]$ *is well defined,* $X = X\mathbb{1}_{[0,\infty)}$ $\mathbb{P}$ *a.s. and we apply Proposition 10.3 with* $f(x) = x\mathbb{1}_{[0,\infty)}(x)$ *to deduce that*

$$\mathbb{E}[X] = \mathbb{E}[f(X)] = \int_0^\infty x P_X(dx) = \int_0^\infty x\phi(x)dx.$$

Keep in mind Proposition 9.3. In case (ii), $\mathbb{E}[X]$ *is well defined and we let* $f\colon \mathbb{R} \to \mathbb{R}$ *be the identity map and again apply Propositions 10.3 and 9.3 to deduce that*

$$\mathbb{E}[X] = \int_{\mathbb{R}} x P_X(dx) = \int_{\mathbb{R}} x\phi(x)dx.$$

Example 10.15. *Let* X *be a discrete random variable with Poisson distribution, i.e.,* $\mathbb{P}(X = x) = (\lambda^x/x!)\,\mathrm{e}^{-\lambda}$, $x \in \mathbb{N} \cup \{0\}$, $\lambda > 0$. *We notice that* $1 = P_X(\mathbb{N} \cup \{0\}) = \mathbb{P}(X \in \mathbb{N} \cup \{0\})$. *That is,* $X \geq 0$ $\mathbb{P}$ *a.s. Hence, upon Example 10.13,* $\mathbb{E}[X]$ *is well defined and we calculate*

$$\mathbb{E}[X] = \mathrm{e}^{-\lambda}\sum_{k=0}^{\infty} k\frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!} = \lambda\,\mathrm{e}^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda\,\mathrm{e}^{-\lambda}\,\mathrm{e}^{\lambda} = \lambda < \infty.$$

Example 10.16. *Let* X *be a continuous random variable with uniform distribution, i.e.,* P_X *has probability density function* $\phi(x) = 1/(b-a)\mathbb{1}_{[a,b]}(x)$, $x \in \mathbb{R}$. *We apply Proposition 10.3 with* $f(x) = |x|$, $x \in \mathbb{R}$ *and readily notice that* $\mathbb{E}[|X|] < \infty$. *Recall that* $x \mapsto |x|$ *is continuous on* $[a,b]$ *and hence Riemann integrable, i.e.,* $\mathbb{E}[|X|] < \infty$ *follows directly from Proposition 9.4. Thus, by Example 10.15, we have that* $\mathbb{E}[X]$ *is well defined and*

$$\mathbb{E}[X] = \frac{1}{b-a}\int_a^b x dx = \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{(a+b)}{2}.$$

Example 10.17. *Let* $\Omega = \{(i,j)\colon i,j \in \{1,2,3,4,5,6\}\}$ *as given in Example 10.7. Define the random variable* $X((i,j)) = i+j$, $(i,j) \in \{1,2,3,4,5,6\}^2$. X *is nonnegative and hence* $\mathbb{E}[X]$ *is well defined and we calculate*

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{(i,j)\in\Omega}(i+j)\mathbb{P}(X = (i,j)) = \sum_{(i,j)\in\Omega}(i+j)\frac{1}{36} = \frac{1}{36}\sum_{i=1}^{6}\left(\sum_{j=1}^{6}(i+j)\right)\\
&= \frac{1}{36}\left(\sum_{i=1}^{6}\left(6i + \sum_{j=1}^{6} j\right)\right) = \frac{1}{36}\left(\sum_{i=1}^{6}\left(6i + \frac{6(6+1)}{2}\right)\right)\\
&= \frac{1}{36}\left(\frac{36(6+1)}{2} + \frac{36(6+1)}{2}\right) = 7.
\end{aligned}$$

Keep in mind Example 1.7.

Example 10.18. *Suppose that* $X \sim \mathcal{N}(\mu,\sigma^2)$. *We verify that* $\mathbb{E}[X] = \mu$. *If* $\sigma = 0$, *then, by Definition 10.8,* $\mathbb{P}(\mu = X) = 1$, *i.e., by Exercise 8.3,*

$$\mu = \mu\mathbb{P}(\Omega) = \int_\Omega \mu\mathbb{P}(d\omega) = \int_\Omega X(\omega)\mathbb{P}(d\omega) = \mathbb{E}[X].$$

Hence, suppose that $\sigma > 0$. *Consider the case where* $\mu = 0$ *and* $\sigma = 1$ *(i.e.,* X *is standard normal). Upon a substitution* $u = x^2/2$ *we obtain*

$$\int_0^b \frac{1}{\sqrt{2\pi}} x\,\mathrm{e}^{-x^2/2}\,dx = \int_0^{b^2}\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-u}\,du = \frac{1}{\sqrt{2\pi}}\big[-\mathrm{e}^{-u}\big]_0^{b^2} = \frac{1-\mathrm{e}^{-b^2}}{\sqrt{2\pi}}.$$

Hence, $\lim_{b\uparrow\infty}\int_0^b (1/\sqrt{2\pi})x\,\mathrm{e}^{-x^2/2}\,dx = 1/(\sqrt{2\pi})$. *Similarly, we verify that*

$$\lim_{a\downarrow-\infty}\int_a^0 \frac{1}{\sqrt{2\pi}}x\,\mathrm{e}^{-x^2/2}\,dx = \frac{-1}{\sqrt{2\pi}}.$$

Therefore $\mathbb{E}[X]=0$. *Now, if* $\mu\in\mathbb{R}$ *and* $\sigma>0$, *we obtain*

$$\begin{aligned}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}x\,\mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}\,dx &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}(x\sigma+\mu)\,\mathrm{e}^{-x^2/2}\,dx\\ &= \mu\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2}\,dx = \mu,\end{aligned}$$

since $\int_{-\infty}^{\infty}(1/\sqrt{2\pi})\,\mathrm{e}^{-x^2/2}\,dx = 1$. *Notice that* $x\mapsto |x|\,\mathrm{e}^{-x^2/2}$ *is continuous on* $[a,b]$ *for any* $a\neq b$, $a,b\in\mathbb{R}$. *Therefore, the latter map is Riemann integrable on* $[a,b]$ *and in particular, Lebesgue integrable and both integrals agree (cf. Proposition 9.4). Further, since* $x\mapsto |x|\,\mathrm{e}^{-x^2/2}$ *is even we recall Exercise 9.2 and obtain*

$$\int_{-\infty}^{\infty}|x|\,\mathrm{e}^{-x^2/2}\,dx = \lim_{b\uparrow\infty}2\int_0^b |x|\,\mathrm{e}^{-x^2/2}\,dx = \lim_{b\uparrow\infty}2\int_0^b x\,\mathrm{e}^{-x^2/2}\,dx = 2.$$

Hence, by Proposition 9.6, $x\mapsto x\,\mathrm{e}^{-x^2/2}$ *is Lebesgue integrable.*

Remark 10.6. *Let* $(\Omega,\mathcal{F},\mathbb{P})$ *be a probability space and* X *be a random vector where the law of* X *is unknown. We notice that by Proposition 10.3, for any* $A\in\mathfrak{B}(\mathbb{R}^k)$,

$$\mathbb{E}[\mathbb{1}_A(X)] = \int_\Omega \mathbb{1}_A(X(\omega))\mathbb{P}(d\omega) = \int_{\mathbb{R}^k}\mathbb{1}_A(x)P_X(dx) = P_X(A).$$

Hence, Proposition 10.3 allows us to identify the law of X.

Example 10.19. *Let* $(\Omega,\mathcal{F},\mathbb{P})$ *be a probability space and* $X=(X_1,\dots,X_k)$ *be a random vector where the distribution of* X *has probability density function* $\phi(x_1,\dots,x_k)$, $(x_1,\dots,x_k)\in\mathbb{R}^k$. *Then, for any* $i=1,\dots,k$, *the law of* X_i *has probability density function*

$$p_i(x) = \int_{\mathbb{R}^{k-1}}\phi(x_1,\dots,x_{i-1},x,x_{i+1},\dots,x_k)d(x_1,\dots,x_{i-1},x_{i+1},\dots,x_k),\quad x\in\mathbb{R}. \tag{30}$$

To see it, let $i\in\{1,\dots,k\}$. *Given* $(x_1,\dots,x_k)\in\mathbb{R}^k$, *define the map* $\pi_i(x_1,\dots,x_k)=x_i$. *Then, for any nonnegative* $\mathfrak{B}(\mathbb{R})$ *measurable function* $f\colon\mathbb{R}\to\mathbb{R}$, *we apply Proposition 10.3 with* $f(\pi_i)$ *and obtain*

$$\begin{aligned}\mathbb{E}[f(X_i)] &= \mathbb{E}[f(\pi_i(X)]\\ &= \int_{\mathbb{R}^k} f(x_i)\phi(x_1,\dots,x_k)d(x_1,\dots,x_k)\\ &= \int_{\mathbb{R}} f(x_i)\Big(\int_{\mathbb{R}^{k-1}}\phi(x_1,\dots,x_k)d(x_1,\dots,x_{i-1},x_{i+1},\dots,x_k)\Big)dx_i,\end{aligned}$$

where we made use of Proposition 9.13 (recall also Example 9.8).

Example 10.20. *Let* $(\Omega,\mathcal{F},\mathbb{P})$ *be a probability space and* U *be a random variable with uniform law on* $[0,1]$, *i.e.,* $P_U(dx)=\mathbb{1}_{[0,1]}(x)dx$. *Define the random variable* $X=-2\log(U)$. *By Proposition 10.3, for any* $f\colon\mathbb{R}\to\mathbb{R}$, *nonnegative and* $\mathfrak{B}(\mathbb{R})$ *measurable,*

$$\mathbb{E}[f(X)] = \mathbb{E}[f(-2\log(U))] = \int_{\mathbb{R}} f(-2\log(u))\mathbb{1}_{[0,1]}(u)du = \int_0^1 f(-2\log(u))du.$$

We substitute $x(u) = -2\log(\mathrm{u})$ $(dx/du = -2/u \Leftrightarrow du = -u/2)$ *and obtain for any* $\varepsilon > 0$

$$\int_{\varepsilon}^{1} f(-2\log(u))du = \int_{x(\varepsilon)}^{x(1)} f(x)\Big(-\frac{\mathrm{e}^{-x/2}}{2}\Big)dx = \int_{0}^{x(\varepsilon)} f(x)\frac{\mathrm{e}^{-x/2}}{2}dx,$$

where we recall Definition 9.6. Hence,

$$\lim_{\varepsilon\to 0}\int_{0}^{x(\varepsilon)} f(x)\frac{\mathrm{e}^{-x/2}}{2}dx = \int_{0}^{\infty} f(x)\frac{\mathrm{e}^{-x/2}}{2}dx = \lim_{\varepsilon\to 0}\int_{\varepsilon}^{1} f(-2\log(u))du = \mathbb{E}[f(X)].$$

By Remark 10.6, $P_X(dx) = 2^{-1}\,\mathrm{e}^{-x/2}\,dx$*, i.e., the law of* X *is exponential with* $\lambda = 1/2$.

Exercise 10.4. *Let* $X \sim \mathcal{N}(\mu, \sigma^2)$, $\sigma > 0$*. Show that* $X - \mu \sim \mathcal{N}(0, \sigma^2)$.

The following is known as Markov's inequality:

Proposition 10.4. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and* X *be a random variable s.t.* $\mathbb{E}[|X|^p] < \infty$ *where* $p \in \mathbb{N}$*. Then, for any* $a > 0$, $\mathbb{P}(|X| \geq a) \leq a^{-p}\mathbb{E}[|X|^p]$.

Proof. Let $a > 0$. We have that

$$\mathbb{P}(|X| \geq a) = \int_{\Omega} \mathbb{1}_{\{|X|\geq a\}}(\omega)\mathbb{P}(d\omega) \leq \int_{\Omega} \frac{|X|^p(\omega)}{a^p}\mathbb{1}_{\{|X|\geq a\}}(\omega)\mathbb{P}(d\omega) \leq \frac{\mathbb{E}[|X|^p]}{a^p}.$$

□

10.6 Distribution function

Definition 10.10. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and* X *be a random variable. The distribution function* F *of* X *is defined by*

$$F_X(t) = \mathbb{P}(X \leq t) = P_X((-\infty, t]), \quad t \in \mathbb{R}.$$

Exercise 10.5. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and* X *and* Y *be two random variables with equal distribution function* F*. Verify that* $P_X = P_Y$*, i.e.,* X *and* Y *have the same law.*

Remark 10.7. *Using Proposition 10.2, if* X *is discrete, we have that for any* $t \in \mathbb{R}$,

$$F_X(t) = \sum_{\substack{x\in E\\ x\leq t}} p_x.$$

If X *is continuous with law that has probability density function* ϕ*, we have upon Definition 10.6 that for any* $t \in \mathbb{R}$,

$$F_X(t) = \int_{-\infty}^{t} \phi(x)dx.$$

Proposition 10.5. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and* X *be a random variable. Then, the distribution function* $F_X\colon \mathbb{R} \to [0, 1]$ *of* X *is continuous if and only if for any* $t \in \mathbb{R}$, $\mathbb{P}(X = t) = 0$.

Example 10.21. *Suppose that the law of* X *is Binomial with parameters* n *and* p*, i.e.,*

$$P_X = \sum_{x=0}^{n} \binom{n}{x} p^x(1-p)^{n-x}\delta_x.$$

Then, $F_X(t)$ *is equal to zero for any* $t < 0$ *and* 1 *for any* $t \geq n$*. The jumps are at the points* $x = 0, \dots, n$.

Proposition 10.6. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X be a continuous random variable with distribution function F_X. Let*

$$E = \{t \in \mathbb{R} \colon 0 < F_X(t) < 1\}.$$

Then, if F_X is strictly increasing on E, the restriction $F_X|_E \colon E \to (0,1)$ is invertible with inverse $F_X|_E^{-1}$ that is strictly in increasing on $(0,1)$.

Exercise 10.6. *Let X be a random variable with uniform law on $[a,b]$, i.e., $P_X(dx) = 1/(b-a)\mathbb{1}_{[a,b]}(x)dx$. Calculate $F_X(t)$, $t \in \mathbb{R}$. Show that F_X restricted to (a,b) is invertible and calculate its inverse.*

Exercise 10.7. *Let $X \sim \mathcal{N}(0,1)$. Show that for any $t \in \mathbb{R}$, $1 - F_X(-t) = F_X(t)$.*

10.7 Variance and covariance

Definition 10.11. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X be a random variable s.t. $\mathbb{E}[|X|] < \infty$. The variance of X is defined by*

$$\mathrm{Var}(X) = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big] \in \overline{\mathbb{R}}_+.$$

Proposition 10.7. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X be a random variable s.t. $\mathbb{E}[|X|] < \infty$. Then,*

(i) $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$*;*

(ii) $X = \mathbb{E}[X]$ $\mathbb{P}$ *a.s. if and only if* $\mathrm{Var}(X) = 0$*;*

(iii) Given $a > 0$, $\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq a^{-2}\,\mathrm{Var}(X)$.

We remark that the last item is known as Chebyshev's inequality.

Proof. We readily deduce items (i) and (ii) from Definition 10.11. Item (iii) is just Markov's inequality (cf. Proposition 10.4) applied to $X - \mathbb{E}[X]$ with $p = 2$. □

Exercise 10.8. *Let X be a random variable s.t.* $\mathrm{Var(X)} < \infty$*. Show that* $\mathrm{Var(a + X)} = \mathrm{Var(X)}$*.*

Remark 10.8. *Notice that if X is a random variable s.t. $\mathbb{E}[|X|^2] < \infty$, then $\mathbb{E}[|X|] < \infty$. To see it, take the set $A = \{\omega \colon |X| \leq 1\}$. We have that*

$$\mathbb{E}[|X|] = \int_A |X|(\omega)\mathbb{P}(d\omega) + \int_{A^c} |X|(\omega)\mathbb{P}(d\omega) \leq \mathbb{P}(A) + \int_{A^c} |X|^2(\omega)\mathbb{P}(d\omega) < \infty.$$

In particular, if $\mathbb{E}[|X|^2] < \infty$, then $\mathrm{Var}(X) < \infty$. Clearly, if $\mathrm{Var}(X) < \infty$, then $\mathbb{E}[|X|^2] < \infty$. Thus, $\mathrm{Var}(X) < \infty$ if and only if $\mathbb{E}[|X|^2] < \infty$.

Example 10.22. *Let X be a random variable with law P_X that is continuous uniform with probability density $\phi(x) = 1/(b-a)\mathbb{1}_{[a,b]}(x)$, $x \in \mathbb{R}$. By Example 10.16, we know that $\mathbb{E}[X]^2 = (a+b)^2/4$. Further, by Proposition 10.3,*

$$\begin{aligned}
\mathbb{E}[X^2] &= \frac{1}{b-a}\int_a^b x^2 dx \\
&= \frac{1}{b-a}\left[\frac{1}{3}x^3\right]_a^b = \frac{b^3 - a^3}{3(b-a)} = \frac{(b-a)(b^2 + ab + a^2)}{3(b-a)} = \frac{(b^2 + ab + a^2)}{3}.
\end{aligned}$$

Hence,

$$\mathrm{Var}(X) = \frac{4b^2 + 4ab + 4a^2 - 3a^2 - 3b^2 - 6ab}{12} = \frac{b^2 + a^2 - 2ab}{12} = \frac{(a-b)^2}{12}.$$

Remark 10.9. *One can show that if $X \sim \mathcal{N}(\mu, \sigma^2)$ then* $\operatorname{Var}(X) = \sigma^2$*. Notice that if $X \sim \mathcal{N}(\mu, 0)$, then $X = \mu = \mathbb{E}[X]$ $\mathbb{P}$ a.s., i.e., by item (ii) of Proposition 10.7,* $\operatorname{Var}(X) = 0$.

The following is known as a version of the Cauchy–Schwarz inequality.

Proposition 10.8. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X and Y be two random variables s.t. $\mathbb{E}[|X|^2] < \infty$ and $\mathbb{E}[|Y|^2] < \infty$. Then,*

$$\mathbb{E}[|XY|] \leq \sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}. \tag{31}$$

We define the covariance between two random variables.

Definition 10.12. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.*

(i) If X and Y are two random variables s.t. $\operatorname{Var}(X) < \infty$ *and* $\operatorname{Var}(Y) < \infty$*, then the covaraince of X and Y is defined as*

$$\operatorname{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])];$$

(ii) if $\operatorname{Cov}(X, Y) = 0$ *for two random variables, then X and Y are referred to as uncorrelated;*

(iii) If $X = (X_1, \dots, X_k)$ is a random vector, where for any $i = 1, \dots, k$, $\operatorname{Var}(X_i) < \infty$*, then the covariance matrix $\Sigma(X)$ is defined by $\Sigma(X)_{i,j} = \operatorname{Cov}(X_i, X_j)$, $1 \leq i, j \leq k$.*

Remark 10.10. *Notice that upon Proposition 10.8, we have that*

$$|\operatorname{Cov}(X, Y)| \leq \sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}.$$

Hence, the condition $\operatorname{Var}(X) < \infty$ *and* $\operatorname{Var}(Y) < \infty$ *is sufficient to ensure that* $\operatorname{Cov}(X, Y)$ *is well defined.*

Proposition 10.9. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X = (X_1, \dots, X_k)$ be a random vector, where for any $i = 1, \dots, k$,* $\operatorname{Var}(X_i) < \infty$*. Then, for any $v = (v_1, \dots, v_k) \in \mathbb{R}^k$, the random variable $v^t X$ has variance $v^t \Sigma(X) v$. If $k = 1$, that is* $\operatorname{Var}(vX) = v^2 \operatorname{Var}(X)$*, $v \in \mathbb{R}$.*

Proposition 10.10. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X = (X_1, \dots, X_k)$ be a random vector s.t. for any $i = 1, \dots, k$,* $\operatorname{Var}(X_i) < \infty$*. Then, the covariance matrix $\Sigma(X)$ is symmetric and positive semidefinite.*

Proof. By definition, $\Sigma(X)$ is symmetric. Further, given any $v_1, \dots, v_k \in \mathbb{R}$, we have with Proposition 10.9 that $0 \leq \operatorname{Var}(v^t X) = v^t \Sigma(X) v$. □

Remark 10.11. *Let $X_1, \dots, X_k$ be k random variables s.t. for any $i = 1, \dots, k$,* $\operatorname{Var}(X_i) < \infty$*. Then, if we set $v = (1, \dots, 1)$ in Proposition 10.9, we obtain*

$$\operatorname{Var}\Big(\sum_{i=1}^{k} X_i\Big) = \sum_{i=1}^{k} \operatorname{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \operatorname{Cov}(X_i, X_j).$$

Hence, if $X_1, \dots, X_k$ are pairwise uncorrelated, then, $\operatorname{Var}(\sum_{i=1}^{k} X_i) = \sum_{i=1}^{k} \operatorname{Var}(X_i)$.

10.8 Characteristic function

Definition 10.13. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f \colon \Omega \to \mathbb{C}$, i.e., for any $\omega \in \Omega$, $f(\omega) = g(\omega) + i h(\omega)$, where $g \colon \Omega \to \mathbb{R}$ and $h \colon \Omega \to \mathbb{R}$ and $i^2 = -1$, the imaginary unit. Then, f is said to be $\mathcal{F}$ measurable if g and h are $\mathcal{F}$ measurable. Further, if g and h are μ integrable, f is μ integrable and we write*

$$\int_\Omega f(\omega)\mu(d\omega) = \int_\Omega g(\omega)\mu(d\omega) + i \int_\Omega h(\omega)\mu(d\omega).$$

Definition 10.14. *Let P be a probability measure on $\mathfrak{B}(\mathbb{R}^k)$. The Fourier transform of P is defined by*

$$\widehat{P}(v) = \int_{\mathbb{R}^k} \mathrm{e}^{iv^t x}\, P(dx), \quad v \in \mathbb{R}^k.$$

If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and X a random vector with law P_X on $\mathbb{B}(\mathbb{R}^k)$, the characteristic function of X is defined by $\Phi_X(v) = \widehat{P}_X(v)$, $v \in \mathbb{R}^k$. That is, $\Phi_X(v) = \mathbb{E}[\mathrm{e}^{iv^t X}]$.

Remark 10.12. *Notice that for any $v \in \mathbb{R}^k$, $|\Phi_X(v)| \leq 1$ (cf. Exercise 8.6) and $\Phi_X(0) = 1$. Further, by Proposition 8.12, $v \mapsto \Phi_X(v)$ is continuous on $\mathbb{R}^k$.*

As a well known example, the following proposition gives the characteristic function for Gaussian random variables.

Proposition 10.11. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X \sim \mathcal{N}(\mu, \sigma^2)$. Then,*

$$\Phi_X(v) = \mathrm{e}^{i\mu v - \frac{\sigma^2 v^2}{2}}, \quad v \in \mathbb{R}. \tag{32}$$

In general, the law of a random vector is determined by its characteristic function — this is known as the uniqueness theorem of the characteristic function:

Proposition 10.12. *Let P and P' be two probability measures on $\mathfrak{B}(\mathbb{R}^k)$. If $\widehat{P}(v) = \widehat{P}'(v)$ for any $v \in \mathbb{R}^k$, then $P = P'$.*

Combining the latter two propositions, (32) characterizes the law of a Gaussian random variable — a summary is given in the following remark.

Remark 10.13. *Let $\mu \in \mathbb{R}$ and $\sigma \geq 0$. Then, Propositions 10.11 and 10.12 show that a random variable X is Gaussian with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \geq 0$ if and only if its characteristic function is given by (32). Notice that if X has characteristic function given by (32) with $\sigma = 0$, then if we define $X'(\omega) = \mu$ for any $\omega \in \Omega$, we apply Proposition 10.12 and deduce that $P_{X'} = P_X$. Hence, $\mathbb{P}(X = \mu) = \mathbb{P}(\mu = \mu) = 1$, i.e., X is again Gaussian according to Definition 10.8.*

Exercise 10.9. *Show that if $X \sim \mathcal{N}(\mu, \sigma^2)$, then for any $a, b \in \mathbb{R}$, $a + bX \sim \mathcal{N}(a + b\mu, b^2\sigma^2)$. Conclude that if $\sigma > 0$, $(X - \mu)/\sigma$ is standard normal.*

10.9 Solution to exercises

Solution 10.1 (Solution to Exercise 10.1)**.** *Clearly, $\#\Omega = 6^2 = 36$. We split*

$$\Omega = A \cup A^c = A \cup \underbrace{\{(\omega_1, \omega_2) \in \Omega \colon \omega_2 < \omega_1\}}_{=A'} \cup \{(\omega_1, \omega_2) \in \Omega \colon \omega_1 = \omega_2\}.$$

If $\omega = (\omega_1, \omega_2) \in A$, then $(\omega_2, \omega_1) \in A'$ (and vice versa). This shows that $\#A = \#A'$. Thus,

$$\#\Omega = 2\#A + 6 = 36 \Rightarrow \#A = 15.$$

Hence, $\mathbb{P}(A) = 15/36$.

Solution 10.2 (Solution to Exercise 10.2)**.** *This is Exercise 2.5, we have that $\sum_{x=0}^{n} p_x = (p + 1 - p)^n = 1$.*

Solution 10.3 (Solution to Exercise 10.3)**.** *This follows from the definition of e^x, $x \in \mathbb{R}$ (cf. Example 3.7). We have that*

$$\sum_{x=0}^{\infty} \frac{\lambda^x \mathrm{e}^{-\lambda}}{x!} = \mathrm{e}^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = \mathrm{e}^{-\lambda}\, \mathrm{e}^{\lambda} = 1.$$

Solution 10.4 (Solution to Exercise 10.4)**.** *Let* $g(x) = x - \mu$, $x \in \mathbb{R}$. *Then, for any nonnegative and* $\mathfrak{B}(\mathbb{R})$ *measurable function* $f\colon \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[f(X-\mu)] = \mathbb{E}[f(g(X))] = \int_{\mathbb{R}} f(x-\mu)\frac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx = \int_{\mathbb{R}} f(u)\frac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{e}^{-\frac{u^2}{2\sigma^2}}\,du.$$

Thus, $X - \mu \sim \mathcal{N}(0, \sigma^2)$.

Solution 10.5 (Solution to Exercise 10.5)**.** *This is a simple consequence of Proposition 6.9. We know that* $\mathfrak{B}(\mathbb{R}) = \sigma(\mathcal{R})$, *where* $\mathcal{R}$ *is the semiring of left-open intervals with the empty set adjoined. By assumption, for any* $a, b \in \mathbb{R}$, $a < b$,

$$F_X(b) - F_X(a) = P_X((-\infty, b]) - P_X((-\infty, a]) = P_X((a, b]) = P_Y((a, b]).$$

Thus, since P_X *and* P_Y *agree on* $\mathcal{R}$, *they also agree on* $\mathfrak{B}(\mathbb{R})$.

Solution 10.6 (Solution to Exercise 10.6)**.** *We have that*

$$F_X(t) = P_X((-\infty, t]) = \frac{1}{b-a}\lambda\big((-\infty, t] \cap [a, b]\big) = \begin{cases} 0, & \text{if } t < a, \\ \frac{1}{b-a}\int_a^t dx = \frac{t-a}{b-a}, & \text{if } a \le t < b, \\ 1, & \text{if } t \ge b. \end{cases} \quad (33)$$

Recall that λ *is the Lebesgue measure on* $\mathfrak{B}(\mathbb{R})$. *By* (33), $\{t \in \mathbb{R}\colon 0 < F_X(t) < 1\} = (a, b)$. *Further, the restriction* $F_X|_{(a,b)}$ *is strictly increasing since for any* $t \in (a, b)$, $F'_X|_{(a,b)}(t) > 0$ *(cf. Proposition 2.3). Therefore, by Proposition 10.6,* $F_X|_{(a,b)}\colon (a, b) \to (0, 1)$ *is a bijection, i.e., for any* $p \in (0, 1)$, *there exists a unique* $t \in (a, b)$ *s.t.* $F_X|_{(a,b)}(t) = p$. *By* (33), $F_X|_{(a,b)}(t) = p$ *is equivalent to* $(t-a)/(b-a) = p \Leftrightarrow t = a+p(b-a)$. *Therefore,* $F_X^{-1}|_{(a,b)}(p) = a + p(b - a)$, $p \in (0, 1)$ *(cf. Definition 2.6).*

Solution 10.7 (Solution to Exercise 10.7)**.** *Let* $t \in \mathbb{R}$. *We recall Definition 9.6 and obtain*

$$\sqrt{2\pi}F_X(t) = \int_{-\infty}^{t} \mathrm{e}^{-\frac{x^2}{2}}\,dx = \int_{-\infty}^{t} \mathrm{e}^{-\frac{(-x)^2}{2}}\,dx = -\int_{\infty}^{-t} \mathrm{e}^{-\frac{u^2}{2}}\,du = \int_{-t}^{\infty} \mathrm{e}^{-\frac{u^2}{2}}\,du.$$

Therefore,

$$F_X(t) = \frac{1}{\sqrt{2\pi}}\int_{-t}^{\infty} \mathrm{e}^{-\frac{x^2}{2}}\,dx = \mathbb{P}(X > -t) = 1 - \mathbb{P}(X \le -t) = 1 - F_X(-t).$$

Solution 10.8 (Solution to Exercise 10.8)**.** *By definition,* $\mathrm{Var}(a+X) = \mathbb{E}[(a+X)^2] - \mathbb{E}[a+X]^2$. *Developing the terms on the right of the latter equation gives the result.*

Solution 10.9 (Solution to Exercise 10.9)**.** *We calculate the characteristic function of the random variable* $a + bX$*: Given* $v \in \mathbb{R}$, *we apply Proposition 10.11 and deduce that*

$$\begin{aligned} \Phi_{a+bX}(v) &= \mathbb{E}[\mathrm{e}^{iv(a+bX)}] \\ &= \mathbb{E}[\mathrm{e}^{iva}\,\mathrm{e}^{i(vb)X)}] \\ &= \mathrm{e}^{iva}\,\mathbb{E}[\mathrm{e}^{i(vb)X)}] = \mathrm{e}^{iva}\,\Phi_X(vb) = \mathrm{e}^{iva}\,\mathrm{e}^{i(b\mu)v - \frac{(\sigma^2 b^2)v^2}{2}} = \mathrm{e}^{i(a+b\mu)v - \frac{(\sigma^2 b^2)v^2}{2}}. \end{aligned}$$

Thus, by Proposition 10.12 (cf. Remark 10.13), $a + bX \sim \mathcal{N}(a + b\mu, b^2\sigma^2)$. *It follows that,* $X - \mu \sim \mathcal{N}(0, \sigma^2)$ *and hence* $(X - \mu)/\sigma \sim \mathcal{N}(0, 1)$.

10.10 Additional exercises

Exercise 10.10. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and* X *be a random variable s.t.* $X \sim \mathcal{N}(\mu, \sigma^2)$, $\sigma > 0$. *Show that* $\mathbb{P}(X \le \mu) = 1/2$.
Hint: *Consider first the case where* $X \sim \mathcal{N}(0, 1)$.

Exercise 10.11. *Define*

$$\phi(x) = \begin{cases} 0 & x < -2 \\ \frac{1}{2} + \frac{1}{4}x & -2 \leq x < 0 \\ \frac{1}{2} - \frac{1}{4}x & 0 \leq x < 2 \\ 0 & x \geq 2. \end{cases}$$

(a) Show that $\int_{\mathbb{R}} \phi(x)dx = 1$.

Suppose that X is a random variable with law $P_X(dx) = \phi(x)dx$.

(b) Find the distribution function F_X of X.

(c) Calculate the expected value and the variance of X.

Exercise 10.12. *Suppose that X is a random variable with law $P_X(dx) = \mathrm{e}^{-x}\,\mathbb{1}_{[0,\infty)}(x)dx$, $x \in \mathbb{R}$. That is P_X is exponential with parameter $\lambda = 1$. Find the distribution function F_X of X. Verify that $F_X|_{[0,\infty)}$ is invertible and find its inverse.*

Exercise 10.13. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X be a random variable s.t. $X(\Omega) \subset \mathbb{N}$. Verify that*

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} \mathbb{P}(X \geq n).$$

Hint: *Proposition 3.11.*

Exercise 10.14. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X be a random variable with continuous distribution function F_X that is strictly increasing on $\mathbb{R}$. Find the law of $\omega \mapsto F_X(X)(\omega)$.*

11 Collections of random vectors

11.1 Independence

Regarding omitted proofs of this chapter, we refer to Appendix C.

Definition 11.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{A}_1, \ldots, \mathcal{A}_n$ be n sub-σ-fields on Ω. $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are said to be independent if for any $A_1 \in \mathcal{A}_1, \ldots, A_n \in \mathcal{A}_n$,*

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdot \ldots \cdot \mathbb{P}(A_n).$$

Remark 11.1. *If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $A_1, \ldots, A_n$ are n sets (events) s.t. $A_i \in \mathcal{F}$, $i = 1, \ldots, n$. Then, $A_1, \ldots, A_n$ are said to be independent if the sub-σ-fields $\sigma(A_1), \ldots, \sigma(A_n)$ are independent. Recall that $\sigma(A_i) = \{A_i, A_i^c, \Omega, \emptyset\}$, $i = 1, \ldots, n$ (cf. Example 4.8).*

Remark 11.2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are independent sub-σ-fields on Ω, then*

$$\mathbb{P}(A_{k_1} \cap \cdots \cap A_{k_j}) = \mathbb{P}(A_{k_1}) \cdot \ldots \cdot \mathbb{P}(A_{k_j}),$$

for any choice $A_{k_1} \in \mathcal{A}_{k_1}, \ldots, A_{k_j} \in \mathcal{A}_{k_j}$, $1 \leq k_1 < k_2 < \cdots < k_j \leq n$, $2 \leq j \leq n$. This is because for any $i = 1, \ldots, n$, $\Omega \in \mathcal{A}_i$. Therefore, by Definition 11.1,

$$\mathbb{P}(A_{k_1} \cap \cdots \cap A_{k_j}) = \mathbb{P}(A_{k_1} \cap \cdots \cap A_{k_j} \cap \underbrace{\Omega \cap \cdots \cap \Omega}_{n-j \text{ times}}) = \mathbb{P}(A_{k_1}) \cdot \ldots \cdot \mathbb{P}(A_{k_j}).$$

Definition 11.2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X_i \colon \Omega \to \mathbb{R}^{k_i}$, $i = 1, \ldots, n$, be n random vectors on $(\Omega, \mathcal{F})$. $X_1, \ldots, X_n$ are said to be independent if the sub-σ-fields $\sigma(X_1), \ldots, \sigma(X_n)$ are independent. By definition of $\sigma(X_i) = \{X_i^{-1}(B) \colon B \in \mathfrak{B}(\mathbb{R}^{k_i})\}$, $i = 1, \ldots, n$, that is equivalent to assume that for any $B_1 \in \mathfrak{B}(\mathbb{R}^{k_1}), \ldots, B_n \in \mathfrak{B}(\mathbb{R}^{k_n})$,*

$$\mathbb{P}(\{X_1 \in B_1\} \cap \cdots \cap \{X_n \in B_n\}) = \mathbb{P}(X_1 \in B_1) \cdot \ldots \cdot \mathbb{P}(X_n \in B_n). \tag{34}$$

Remark 11.3. *We remark that the left hand side of (34) is written as*

$$\mathbb{P}(\{X_1 \in B_1\} \cap \cdots \cap \{X_n \in B_n\}) = \mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n).$$

Notice also that by Definition 11.2 (cf. Remark 11.2), if $X_1, \ldots, X_n$ are n independent random vectors, then they are pairwise independent, i.e., for any $i, j = 1, \ldots, n$, $i \neq j$, X_i is independent of X_j.

In what follows we avoid to explicitly mention the underlying probability space and if nothing else is mentioned, any random vector shall be defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Example 11.1. *Let $E = \{-5, -4, \ldots, -1, 0, 1, \ldots, 5\}$. Suppose that X_1 and X_2 are two random variables with common law defined upon $\mathbb{P}(X_1 = i) = \mathbb{P}(X_2 = i) = 1/11$ for any $i \in E$. That is, X_1 and X_2 have law that is discrete uniform on E (recall also Remark 10.2). If we make the assumption that X_1 and X_2 are independent, then, for any $i, j \in E$, $\mathbb{P}(\{X_1 = i\} \cap \{X_2 = j\}) = 1/11^2$. If we want to interpret the events $\{X_1 = i\}$ and $\{X_2 = j\}$, $i, j \in E$, as the outcome of two consecutive independent draws from an urn with balls labeled with digits $-5, \ldots, 5$, where each ball is equally likely to be drawn, then the number $1/11^2$ gives the chance to view label i in the first draw and label j in the second draw.*

Exercise 11.1. *Let $X_1 = X$ be as in Example 11.1. Let G be a discrete random variable with law defined by*

$$\mathbb{P}(G = 1) = \mathbb{P}(G = -1) = 1/2.$$

Suppose that G is independent of X. What is the law of $Y = GX$?

Exercise 11.2. *Let X_1 and X_2 be as in Example 11.1 with the assumption that X_1 and X_2 are independent. What is the law of $M = \max\{X_1, X_2\}$?*

Remark 11.4. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Notice that if $X_1, \dots, X_n$ are n random vectors s.t. $X_i = (X_i^1, \dots, X_i^{k_i})\colon \Omega \to \mathbb{R}^{k_i}$, $i = 1, \dots, n$, then, the n-tuple of random vectors $X = (X_1, \dots, X_n)$ is regarded as a function $X\colon \Omega \to \mathbb{R}^{k_1} \times \dots \times \mathbb{R}^{k_n}$, where $\mathbb{R}^{k_1} \times \dots \times \mathbb{R}^{k_n}$ is equipped with the product σ-field $\mathfrak{B}(\mathbb{R}^{k_1}) \otimes \dots \otimes \mathfrak{B}(\mathbb{R}^{k_n})$. Then, since X_i are by assumption $\mathcal{F}/\mathfrak{B}(\mathbb{R}^{k_i})$ measurable for any $i = 1, \dots, n$, the map X is $\mathcal{F}/(\mathfrak{B}(\mathbb{R}^{k_1}) \otimes \dots \otimes \mathfrak{B}(\mathbb{R}^{k_n}))$ measurable. To see it, we recall that by definition of the product σ-field (cf. Definition 9.9),*

$$\mathfrak{B}(\mathbb{R}^{k_1}) \otimes \dots \otimes \mathfrak{B}(\mathbb{R}^{k_n}) = \sigma\left(\left\{\prod_{i=1}^{n} B_i \colon B_i \in \mathfrak{B}(\mathbb{R}^{k_i}),\ i = 1, \dots, n\right\}\right).$$

Thus, if $B = B_1 \times \cdots B_n$, $B_1 \in \mathfrak{B}(\mathbb{R}^{k_1}), \dots, B_n \in \mathfrak{B}(\mathbb{R}^{k_n})$,

$$X^{-1}(B) = \{\omega \in \Omega \colon X(\omega) \in B\} = X_1^{-1}(B_1) \cap \dots \cap X_n^{-1}(B_n) \in \mathcal{F},$$

since X_i, $i = 1, \dots, n$, are random vectors. Then, we apply Proposition 7.1 and deduce that for any $B \in \mathfrak{B}(\mathbb{R}^{k_1}) \otimes \dots \otimes \mathfrak{B}(\mathbb{R}^{k_n})$, $X^{-1}(B) \in \mathcal{F}$, i.e., X is $\mathcal{F}/(\mathfrak{B}(\mathbb{R}^{k_1}) \otimes \dots \otimes \mathfrak{B}(\mathbb{R}^{k_n}))$ measurable. The law of X is defined as

$$P_X(B) = \mathbb{P}X^{-1}(B) = \mathbb{P}(X \in B), \quad B \in \mathfrak{B}(\mathbb{R}^{k_1} \times \dots \times \mathbb{R}^{k_n}) = \mathfrak{B}(\mathbb{R}^{k_1}) \otimes \dots \otimes \mathfrak{B}(\mathbb{R}^{k_n}).$$

Therefore, by Proposition 9.1 for any $\mathfrak{B}(\mathbb{R}^{k_1}) \otimes \dots \otimes \mathfrak{B}(\mathbb{R}^{k_n})$ measurable map $f\colon \mathbb{R}^{k_1} \times \dots \times \mathbb{R}^{k_n} \to \overline{\mathbb{R}}$ s.t. f is either nonnegative or $\mathbb{E}[|f(X)|] < \infty$, it follows that

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}^{k_1} \times \dots \times \mathbb{R}^{k_n}} f(x) P_X(dx). \tag{35}$$

We show that the law of an independent collection of random vectors is precisely the product measure of the individual laws of the random vectors (cf. Proposition 9.10).

Proposition 11.1. *Let $X = (X_1, \dots, X_n)$ be an n-tuple of random vectors $X_i \colon \Omega \to \mathbb{R}^{k_i}$, $i = 1, \dots, n$. Then, $X_1, \dots, X_n$ are independent if and only if*

$$P_X = P_{X_1} \otimes \dots \otimes P_{X_1},$$

where $P_{X_1} \otimes \dots \otimes P_{X_1}$ is the product measure on $\mathfrak{B}(\mathbb{R}^{k_1}) \otimes \dots \otimes \mathfrak{B}(\mathbb{R}^{k_n})$.

Proof. Suppose that $X_1, \dots, X_n$ are independent. Then, for any $B = B_1 \times \dots \times B_n$, $B_i \in \mathfrak{B}(\mathbb{R}^{k_i})$, $i = 1, \dots, n$,

$$P_X(B) = P_{X_1}(B_1) \cdot \ldots \cdot P_{X_1}(B_1) = P_{X_1} \otimes \dots \otimes P_{X_1}(B). \tag{36}$$

By Proposition 9.10, $P_{X_1} \otimes \dots \otimes P_{X_1}$ is the unique measure on $\mathfrak{B}(\mathbb{R}^{k_1}) \otimes \dots \otimes \mathfrak{B}(\mathbb{R}^{k_n})$ that satisfies (36). Thus, $P_X = P_{X_1} \otimes \dots \otimes P_{X_1}$. The other direction follows readily, it is a consequence of the definition of $P_{X_1} \otimes \dots \otimes P_{X_1}$ on $\mathfrak{B}(\mathbb{R}^{k_1}) \otimes \dots \otimes \mathfrak{B}(\mathbb{R}^{k_n})$. □

Proposition 11.2. *Let $X_1, \dots, X_n$ be n random vectors s.t. $X_i \colon \Omega \to \mathbb{R}^{k_i}$, $i = 1, \dots, n$. Suppose that $X_1, \dots, X_n$ are independent. Then, if for any $i = 1, \dots, n$, $f_i \colon \mathbb{R}^{k_i} \to \overline{\mathbb{R}}$ are nonnegative and $\mathfrak{B}(\mathbb{R}^{k_i})$ measurable,*

$$\mathbb{E}\left[\prod_{i=1}^{n} f_i(X_i)\right] = \prod_{i=1}^{n} \mathbb{E}\big[f_i(X_i)\big]. \tag{37}$$

If $f_i \colon \mathbb{R}^{k_i} \to \overline{\mathbb{R}}$ are not necessarily nonnegative but s.t. $\mathbb{E}[|f_i(X_i)|] < \infty$ for any $i = 1, \dots, n$, (37) remains valid.

Example 11.2. *If $X_1, \dots, X_n$ are independent random variables s.t. $\mathbb{E}[|X_i|] < \infty$ for any $i = 1, \dots, n$, then*

$$\mathbb{E}\Big[\prod_{i=1}^n X_i\Big] = \prod_{i=1}^n \mathbb{E}[X_i].$$

Another consequence of the Proposition 11.2 is the following:

Proposition 11.3. *Let X and Y be two independent random variables s.t. $\mathrm{Var}(X) < \infty$ and $\mathrm{Var}(Y) < \infty$. Then, X and Y are uncorrelated, i.e., $\mathrm{Cov}(X, Y) = 0$.*

Remark 11.5. *We note that the converse of the latter proposition is generally not true. As an example, let $X \sim \mathcal{N}(0, 1)$. Define G as in Exercise 11.1 and assume that G is independent of X. Let $Y = GX$. One can show that $Y \sim \mathcal{N}(0, 1)$ as well (cf. Exercise 11.3). Further, $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[GX^2]$. Since $\mathbb{E}[|GX^2|] < \infty$ and X and G are independent we apply Proposition 11.2 and obtain that $\mathbb{E}[GX^2] = \mathbb{E}[G]\mathbb{E}[X^2]$. Hence, since $\mathbb{E}[G] = 0$, $\mathrm{Cov}(X, Y) = 0$. Thus X and Y are uncorrelated. Assume by contradiction that X and Y are independent. Then it must be the case that*

$$\mathbb{P}(\{0 < X < 1/2\} \cap \{Y > 1\}) = \mathbb{P}(0 < X < 1/2)\mathbb{P}(Y > 1).$$

where the expression on the right is not equal to zero since $X, Y \sim \mathcal{N}(0, 1)$. On the other hand, suppose that $\omega \in \{0 < X < 1/2\} \cap \{Y > 1\}$. Then, $Y(\omega) = G(\omega)X(\omega) > 1$ and $X(\omega) > 0$. Hence, $G(\omega) = 1$ (otherwise we would have $Y(\omega) = -X(\omega) < 0$, which is not possible). Therefore, $\omega \in \{0 < X < 1/2\} \cap \{Y > 1\}$ implies that $X(\omega) > 1$. Hence,

$$\{0 < X < 1/2\} \cap \{Y > 1\} \subset \{0 < X < 1/2\} \cap \{X > 1\} = \emptyset.$$

Thus $\mathbb{P}(\{0 < X < 1/2\} \cap \{Y > 1\}) = 0$, which gives a contradiction. In conclusion, X and Y are uncorrelated but not independent.

Exercise 11.3. *Let X, G and Y be as in Remark 11.5. Verify that $Y \sim \mathcal{N}(0, 1)$.*

Proposition 11.4. *Let $X_1, \dots, X_n$ be n random variables.*

(i) Suppose that for any $i = 1, \dots, n$, $P_{X_i}(dx) = \phi_i(x)dx$, i.e., P_{X_i} has probability density function ϕ_i. Then, if $X_1, \dots, X_n$ are independent, the law of the random vector $X = (X_1, \dots, X_n)$ has probability density function

$$\phi(x) = \prod_{i=1}^n \phi_i(x_i), \quad x = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

(ii) Suppose that the random vector $X = (X_1, \dots, X_n)$ is s.t. $P_X(dx) = \phi(x)dx$, with $\phi(x) = \prod_{i=1}^n \phi_i(x_i)$, $= (x_1, \dots, x_n) \in \mathbb{R}^n$, where for any $i = 1, \dots, n$, ϕ_i are non-negative and $\mathfrak{B}(\mathbb{R})$ measurable. Then, $X_1, \dots, X_n$ are independent where for any $i = 1, \dots, n$, P_{X_i} has probability density function $f_i = K_i\phi_i$, with $K_i \in \mathbb{R}$, $K_i > 0$.

Proof. By Proposition 11.1, if $X_1, \dots, X_n$ are independent, the law of the random vector X is given by $P_X = P_{X_1} \otimes \cdots \otimes P_{X_n}$ on $\mathfrak{B}(\mathbb{R}^n)$. Thus, by Proposition 10.3, for any $B \in \mathfrak{B}(\mathbb{R}^n)$,

$$P_X(B) = \int_{\mathbb{R}^n} \mathbb{1}_B(x) P_{X_1} \otimes \cdots \otimes P_{X_n}(dx) = \int_{\mathbb{R}^n} \mathbb{1}_B(x)\phi(x)dx,$$

where $\phi(x) = \prod_{i=1}^n \phi_i(x_i)$. This shows (i). With regard to (ii), we first notice that

$$1 = \int_{\mathbb{R}^n} \phi(x)dx = \int_{\mathbb{R}} \phi_i(x)dx \Big(\int_{\mathbb{R}^{n-1}} \Big(\prod_{j \neq i} \phi_i(x_i) \Big) d(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \Big).$$

Define $C_i = \int_{\mathbb{R}} \phi_i(x)dx$, $i = 1, \dots, n$. Upon Example 10.19 we deduce that for any $i = 1, \dots, n$, the probability density function of P_{X_i} is

$$f_i(x_i) = \int_{\mathbb{R}^{n-1}} \left(\prod_{i=1}^{n} \phi_i(x_i) \right) d(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = C_i^{-1} \phi_i(x_i).$$

Hence, we obtain that

$$\phi(x) = \prod_{i=1}^{n} \phi_i(x_i) = \prod_{i=1}^{n} f_i(x_i), \quad x = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Then, for any $B \in \mathfrak{B}(\mathbb{R}^n)$,

$$P_X(B) = \int_{\mathbb{R}^n} P_X(dx) = \int_{\mathbb{R}^n} \phi(x)dx = \int_{\mathbb{R}^n} P_{X_1} \otimes \cdots \otimes P_{X_n}(dx) = P_{X_1} \otimes \cdots \otimes P_{X_n}(B),$$

and by Proposition 11.1, $X_1, \dots, X_n$ are independent. $\square$

Example 11.3. *The random vector $X = (X_1, X_2)$ of Example 10.11 is s.t. X_1 and X_2 are independent. This is either seen by (26) under application of Propositions 9.10 and 11.1 or by applying (ii) of Proposition 11.4.*

Example 11.4. *Let U have law that is continuous uniform on $[0, 1]$, i.e., $P_U(du) = \mathbb{1}_{[0,1]}(u)du$. Suppose that R is independent of U with law that is exponential with parameter $\lambda = 1/2$, i.e., $P_R(dr) = (1/2)\,\mathrm{e}^{-r/2}\,\mathbb{1}_{[0,\infty)}(r)dr$. Define $X = \sqrt{R}\cos(2\pi U)$ and $Y = \sqrt{R}\sin(2\pi U)$. Then, X and Y are independent and identically distributed with law $\mathcal{N}(0, 1)$. To see that X and Y are independent, let $h\colon \mathbb{R}^2 \to \mathbb{R}$ be any nonnegative and $\mathfrak{B}(\mathbb{R}^2)$ measurable function. In particular,*

$$h(X, Y) = h(\sqrt{R}\cos(2\pi U), \sqrt{R}\sin(2\pi U)) = h(g(U, R)),$$

with $g(u, r) = (\sqrt{r}\cos(2\pi u), \sqrt{r}\sin(2\pi u))\mathbb{1}_{[0,1]\times[0,\infty)}(u, r)$, $(u, r) \in \mathbb{R}^2$. Thus, by Proposition 10.3,

$$\mathbb{E}[h(X, Y)] = \int_0^\infty \left(\int_0^1 h(\sqrt{r}\cos(2\pi u), \sqrt{r}\sin(2\pi u))2^{-1}\,\mathrm{e}^{-r/2}\,du \right) dr,$$

where we used the assumption that U and R are independent. We substitute $\theta = 2\pi u$ and $\sqrt{r} = \rho$ obtain:

$$\mathbb{E}[h(X, Y)] = \frac{1}{2\pi} \int_0^\infty \left(\int_0^{2\pi} h(\rho\cos(\theta), \rho\sin(\theta))\,\mathrm{e}^{-\rho^2/2}\,\rho d\theta \right) d\rho.$$

Thus, we apply Proposition 9.15 with $f(x, y) = h(x, y)\,\mathrm{e}^{-(x^2+y^2)/2}$, $(x, y) \in \mathbb{R}^2$ and obtain:

$$\mathbb{E}[h(X, Y)] = \frac{1}{2\pi} \int_{\mathbb{R}^2} h(x, y)\,\mathrm{e}^{-\frac{(x^2+y^2)}{2}}\,d(x, y)$$

Therefore (cf. Remark 10.6), the law of (X, Y) has probability density function

$$\phi(x, y) = \frac{1}{2\pi}\,\mathrm{e}^{-\frac{(x^2+y^2)}{2}} = \left(\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{x^2}{2}} \right) \left(\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{y^2}{2}} \right), \quad (x, y) \in \mathbb{R}^2.$$

Upon (ii) of Proposition 11.4, X and Y are independent and s.t. $X \sim \mathcal{N}(0, 1)$ and $Y \sim \mathcal{N}(0, 1)$.

Remark 11.6. *Notice that if $X_1, \dots, X_n$ are n discrete random variables with support $E_1, \dots, E_n$, respectively, then it follows from Proposition 11.1 that $X_1, \dots, X_n$ are independent if and only if for any $(x_1, \dots, x_n) \in \mathbb{R}^n$, the law of $X = (X_1, \dots, X_n)$ satisfies*

$$P_X(\{x_1\} \times \cdots \times \{x_n\}) = P_{X_1}(\{x_1\}) \cdot \ldots \cdot P_{X_n}(\{x_1\}).$$

The following result shows how independence is determined by the characteristic function:

Proposition 11.5. *Let $X_1, \dots, X_n$ be n random variables. Then, $X_1, \dots, X_n$ are independent if and only if the characteristic function Φ_X of the random vector $X = (X_1, \dots, X_n)$ satisfies:*

$$\Phi_X(v) = \prod_{i=1}^{n} \Phi_{X_i}(v_i), \quad v = (v_1, \dots, v_n) \in \mathbb{R}^n,$$

where Φ_{X_i} is the characteristic function of X_i, $i = 1, \dots, n$.

A more general definition for independence, allowing for a notion of independence of tuples of random vectors reads as follows:

Definition 11.3. *Let $Y_1, \dots, Y_N$ be a collection of N tuples of random vectors, i.e., for any $i = 1, \dots, N$, Y_i is the n_i-tuple,*

$$Y_i = (X_1^i, \dots, X_{n_i}^i)$$

where for any $j = 1, \dots, n_i$, $X_j^i \colon \Omega \to \mathbb{R}^{k_{ij}}$ is a random vector. Then, $Y_1, \dots, Y_N$ are said to be independent if the σ-fields $\sigma(Y_1), \dots, \sigma(Y_N)$ are independent.

Remark 11.7. *Recall that by Remark 11.4, for any $i = 1, \dots, N$, the tuple Y_i in the latter definition is a $\mathcal{F}/\mathfrak{B}(\mathbb{R}^{k_{i1}}) \otimes \cdots \otimes \mathfrak{B}(\mathbb{R}^{k_{in_i}})$ measurable mapping from Ω to $\mathbb{R}^{k_{i1}} \times \cdots \times \mathbb{R}^{k_{in_i}}$.*

The following result supports our intuitive understanding of independence (a proof is given in Section C.5 of the appendix).

Proposition 11.6. *Let $X_1, \dots, X_n$ be independent random vectors s.t. $X_i \colon \Omega \to \mathbb{R}^{k_i}$, $i = 1, \dots, n$. Then, for any $n_0 = 0 < n_1 < n_2 < \cdots < n_p = n$, the tuples*

$$Y_1 = (X_1, \dots, X_{n_1}),\ Y_2 = (X_{n_1+1}, \dots, X_{n_2}),\ \dots,\ Y_p = (X_{n_{p-1}+1}, \dots, X_n)$$

are independent. In particular, for any collection of functions

$$f_i \colon \mathbb{R}^{k_{n_i+1}} \times \cdots \times \mathbb{R}^{k_{n_{i+1}}} \to \mathbb{R}, \quad i = 0, \dots, p-1,$$

each of which is $\mathfrak{B}(\mathbb{R}^{k_{n_i+1}}) \otimes \cdots \otimes \mathfrak{B}(\mathbb{R}^{k_{n_{i+1}}})$ measurable, the random variables

$$T_1 = f_1(Y_1),\ \dots,\ T_p = f_p(Y_p),$$

are independent.

Example 11.5. *Suppose that X_1, X_2, X_3 are independent random variables. Then, X_1^2 and $X_2 + X_3$ are independent. In particular, if $X_1, \dots, X_n$ are n independent random vectors s.t. $X_i \colon \Omega \to \mathbb{R}^k$ for any $i = 1, \dots, n$, $X_1 + \cdots + X_{n-1}$ is independent of X_n. Another application of the latter proposition is that if two random vectors $X = (X_1, \dots, X_k)$ and $Y = (Y_1, \dots, Y_k)$ are independent, then, for any $i, j \in \{1, \dots, k\}$, the random variables X_i and Y_j are independent. To see it, we define the coordinate map $\pi_i \colon \mathbb{R}^k \to \mathbb{R}$ ($i = 1, \dots, k$),*

$$\pi_i(x) = x_i, \quad x = (x_1, \dots, x_k),$$

and notice that upon

$$\mathbb{R} \times \cdots \times \mathbb{R} \times B \times \mathbb{R} \cdots \times \mathbb{R} = \pi_i^{-1}(B), \quad B \in \mathfrak{B}(\mathbb{R}),$$

that for any $i = 1, \dots, k$, π_i is $\mathfrak{B}(\mathbb{R}^k)/\mathfrak{B}(\mathbb{R})$ measurable. Then, the result follows by Proposition 11.6, with $Y_1 = X$ and $Y_2 = Y$ and $T_1 = \pi_i(X)$ and $T_2 = \pi_j(Y)$.

Finally we note that the definition of independence can be extended to arbitrary collections of sub-σ-fields and random vectors (or tuples of random vectors).

Definition 11.4. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and I be a non empty set.*

(i) A collection $\mathcal{A}_i$, $i \in I$, of sub-σ-fields on Ω is referred to as independent if for any $i_1, \dots, i_k$, $\mathcal{A}_{i_1}, \dots, \mathcal{A}_{i_k}$ are independent;

(ii) a collection of events $\{A_i : i \in I\} \subset \mathcal{F}$ is said to be independent if $\sigma(A_i)$, $i \in I$, is independent;

(iii) a collection of random vectors $(X_i)_{i\in I}$ is said to be independent if $\sigma(X_i)$, $i \in I$, is independent;

(iv) a collection of random vectors $(X_i)_{i\in I}$ is said to be independent and identically distributed (i.i.d.) if $(X_i)_{i\in I}$ is independent and for any $i \in I$, X_i has law P_{X_1};

(v) a collection of tuples of random vectors $(X_i)_{i\in I}$ is said to be independent if $\sigma(X_i)$, $i \in I$, is independent;

(vi) a collection of tuples of random vectors $(X_i)_{i\in I}$ is said to be independent and identically distributed (i.i.d.) if $(X_i)_{i\in I}$ is independent and for any $i \in I$, X_i has law P_{X_1};

Remark 11.8. *Notice that items (iii) and (iv) of the latter definition are special cases of items (v) and (vi).*

11.2 Sums of independent random vectors

Definition 11.5. *Let $X_1, \dots, X_n$ be n random vectors s.t. $X_i : \Omega \to \mathbb{R}^k$, $i = 1, \dots, n$. We define the measure*

$$P_{X_1} * \cdots * P_{X_n}(B) = (P_{X_1} \otimes \cdots \otimes P_{X_n})s^{-1}(B), \quad B \in \mathfrak{B}(\mathbb{R}^k),\ s(x_1, \dots, x_n) = \sum_{i=1}^{n} x_i.$$

Remark 11.9. *We note that the map $s : \mathbb{R}^k \times \cdots \times \mathbb{R}^k \to \mathbb{R}^k$ given in Definition 11.2 is $\mathfrak{B}(\mathbb{R}^k) \otimes \cdots \otimes \mathfrak{B}(\mathbb{R}^k)$ measurable (cf. Proposition B.7). Further, $P_{X_1} \otimes \cdots \otimes P_{X_n}$ is a measure on $\mathfrak{B}(\mathbb{R}^k) \otimes \cdots \otimes \mathfrak{B}(\mathbb{R}^k)$. Thus, $P_{X_1} * \cdots * P_{X_n}$ is well defined. In particular, $P_{X_1} * \cdots * P_{X_n}$ is the pushforward measure of $P_{X_1} \otimes \cdots \otimes P_{X_n}$ (cf. Definition 9.1).*

Proposition 11.7. *Assume that $X_1, \dots, X_n$ are n independent random vectors s.t. $X_i : \Omega \to \mathbb{R}^k$, $i = 1, \dots, n$. Let $Z = \sum_{i=1}^{n} X_i$.*

(i) The law of Z is $P_{X_1} * \cdots * P_{X_n}$;

(ii) if P_{X_i} has probability density function ϕ_i, $i = 1, \dots, n$, then, the law of Z has probability density function $\phi_Z(z) = \phi_1 * \cdots * \phi_n(z)$, $z \in \mathbb{R}^k$ (cf. Example 9.11). That is, the probability density function of the law of Z is the n-fold convolution of the probability density functions $\phi_1, \dots, \phi_n$ of $P_{X_1}, \dots, P_{X_n}$, respectively.

Proof. By Proposition 11.1, the law of the n-tuple $X = (X_1, \dots, X_n)$ is given by the product measure $P_{X_1} \otimes \cdots \otimes P_{X_n}$. Hence, for any nonnegative $\mathfrak{B}(\mathbb{R}^k)$ measurable function $f : \mathbb{R}^k \to \mathbb{R}$,

$$\mathbb{E}[f(Z)] = \mathbb{E}[f(s(X)] = \int_{\mathbb{R}^k \times \cdots \times \mathbb{R}^k} f(s(x_1, \dots, x_n)) P_{X_1} \otimes \cdots \otimes P_{X_n}(d(x_1, \dots, x_n)),$$

with $s(x_1, \dots, x_n) = \sum_{i=1}^{n} x_i$. Therefore, by Proposition 9.1,

$$\mathbb{E}[f(Z)] = \int_{\mathbb{R}^k} f(z) P_{X_1} \otimes \cdots \otimes P_{X_n} s^{-1}(dz) = \int_{\mathbb{R}^k} f(z) P_{X_1} * \cdots * P_{X_n}(dz).$$

This shows (i). In order to show (ii), we notice that it is enough to consider the case where $n = 2$, since for a general n, the argument follows by induction under application of Proposition 11.6 (cf. Example 11.5) and (22) of Example 9.11. By (i), the law of $X_1 + X_2$ is given by $P_{X_1} * P_{X_2}$. Thus, again by Proposition 9.1, with $s(x_1, x_2) = x_1 + x_2$, for any $B \in \mathfrak{B}(\mathbb{R}^k)$,

$$\begin{aligned}
P_{X_1+X_2}(B) &= \int_{\mathbb{R}^k} \mathbb{1}_B(z) P_{X_1} * P_{X_2}(dz) = \int_{\mathbb{R}^k \times \mathbb{R}^k} \mathbb{1}_B(s(x_1, x_2)) P_{X_1} \otimes P_{X_2}(d(x_1, x_2)) \\
&= \int_{\mathbb{R}^k} \left(\int_{\mathbb{R}^k} \mathbb{1}_B(x_1 + x_2) P_{X_1}(dx_1) \right) P_{X_2}(dx_2) \\
&= \int_{\mathbb{R}^k} \left(\int_{\mathbb{R}^k} \mathbb{1}_B(x_1 + x_2) \phi_1(x_1) dx_1 \right) \phi_2(x_2) dx_2 \\
&= \int_{\mathbb{R}^k} \left(\int_{\mathbb{R}^k} \mathbb{1}_B(z) \phi_1(z - x_2) \phi_2(x_2) dz \right) dx_2 \\
&= \int_{\mathbb{R}^k} \mathbb{1}_B(z) \left(\int_{\mathbb{R}^k} \phi_1(z - x_2) \phi_2(x_2) dx_2 \right) dz.
\end{aligned}$$

Finally, we recall that $\int_{\mathbb{R}^k} \phi_1(z - x_2)\phi_2(x_2) dx_2 = \phi_1 * \phi_2(z)$ λ_k a.e. (cf. Example 9.11). Hence, for any $B \in \mathfrak{B}(\mathbb{R}^k)$, $P_{X_1+X_2}(B) = \int_B \phi_1 * \phi_2(z) dz$, i.e., $P_{X_1+X_2}(dz) = \phi_1 * \phi_2(z) dz$. $\square$

Remark 11.10. *Suppose that X_1 and X_2 are two independent discrete random vectors s.t. $X_1, X_2 \colon \Omega \to \mathbb{R}^k$. Let E_1 and E_2 be the support of X_1 and X_2, respectively. We remark that by Proposition 11.1 (i.e., $P_{(X_1,X_2)} = P_{X_1} \otimes P_{X_2}$), the support of (X_1, X_2) is $E_1 \times E_2$. In particular, the support of $X_1 + X_2$ is $E_1 + E_2 = \{x_1 + x_2 \colon (x_1, x_2) \in E_1 \times E_2\}$. This is because by item (i) of Proposition 11.7,*

$$\begin{aligned}
P_{X_1+X_2}(E_1 + E_2) &= P_{X_1} \otimes P_{X_2}(\{(x_1, x_2) \in \mathbb{R}^k \times \mathbb{R}^k \colon x_1 + x_2 \in E_1 + E_2\}) \\
&= P_{X_1} \otimes P_{X_2}(E_1 \times E_2) = 1.
\end{aligned}$$

We further deduce that

$$P_{X_1+X_2}(\{z\}) = P_{(X_1,X_2)}(\{(x_1, x_2) \in \mathbb{R}^k \times \mathbb{R}^k \colon x_1 + x_2 = z\}), \quad z \in E_1 + E_2.$$

Then, $\{(x_1, x_2) \in \mathbb{R}^k \times \mathbb{R}^k \colon x_1 + x_2 = z\} = \{z - x_2\} \times \mathbb{R}^k$. Hence, by Proposition 10.2,

$$\begin{aligned}
P_{X_1+X_2}(\{z\}) &= \sum_{(x_1,x_2) \in (\{z - x_2\} \cap E_1) \times E_2} P_{X_1}(\{x_1\}) P_{X_2}(\{x_2\}) \\
&= \sum_{x_2 \in E_2} P_{X_1}(\{z - x_2\}) P_{X_2}(\{x_2\}), \quad z \in E_1 + E_2. \qquad (38)
\end{aligned}$$

As with the continuous case, for 3 independent and discrete random vectors X_1, X_2 and X_3 with support E_1, E_2 and E_3, respectively, we obtain for $z \in E_1 + E_2 + E_3$,

$$P_{X_1+X_2+X_2}(\{z\}) = \sum_{x_3 \in E_3} \left(\sum_{x_2 \in E_2} P_{X_1}(\{z - x_3 - x_2\}) P_{X_2}(\{x_2\}) \right) P_{X_3}(\{x_3\}).$$

Generally, if $X_1, \dots, X_n$ are n independent and discrete random vectors with $X_i \colon \Omega \to \mathbb{R}^k$ and X_i has support E_i, $i = 1, \dots, n$, we obtain for $z \in \sum_{i=1}^n E_i$, with $Z = \sum_{i=1}^n X_i$,

$$P_Z(\{z\}) = \sum_{x_n \in E_n} \left(\cdots \left(\sum_{x_3 \in E_3} \left(\sum_{x_2 \in E_2} p_1(z - (x_2 + \cdots + x_n)) p_2(x_2) \right) p_3(x_3) \right) \cdots \right) p_n(x_n),$$

where $p_i(y) = P_{X_i}(\{y\})$, $y \in \mathbb{R}^k$.

Example 11.6. *Suppose that X_1 and X_2 are independent where the law of X_1 is Binomial with parameters $n-1$ and p and X_2 has Bernoulli law with parameter p. Then, X_1+X_2 is Binomial with parameters n and p. To see it, we first notice that X_1+X_2 has support $\{0,\ldots,n\}$ (cf. Remark 11.10). Thus, let $z\in\{0,\ldots,n\}$. We apply (38) and obtain*

$$\begin{aligned}
P_{X_1+X_2}(\{z\}) &= P_{X_1}(\{z-0\})P_{X_2}(\{0\}) + P_{X_1}(\{z-1\})P_{X_2}(\{1\})\\
&= \binom{n-1}{z}p^z(1-p)^{n-1-z}(1-p) + \binom{n-1}{z-1}p^{z-1}(1-p)^{n-1-(z-1)}p\\
&= \left(\binom{n-1}{z}+\binom{n-1}{z-1}\right)p^z(1-p)^{n-z}.
\end{aligned}$$

Then, we rely on Exercise 1.12 and obtain $\binom{n-1}{z}+\binom{n-1}{z-1}=\binom{n}{z}$. In conclusion, for any $z\in\{0,\ldots,n\}$, $\mathbb{P}(X_1+X_2=z)=\binom{n}{z}p^z(1-p)^{n-z}$. This shows that the law of X_1+X_2 is Binomial with parameters n and p. We remark that upon an argument by induction, we have shown that if $X_1,\ldots,X_n$ are n independent random variables with common law that is Bernoulli with parameter p, then the law of $\sum_{i=1}^n X_i$ is Binomial with parameters n and p.

Example 11.7. *Suppose that X_1 and X_2 are two independent random variables where for $i=1,2$, the law of X_i is exponential with parameter $\lambda>0$, i.e.,*

$$P_{X_i}(dx) = \underbrace{\mathbb{1}_{[0,\infty)}(x)\lambda\,\mathrm{e}^{-\lambda x}}_{=\phi(x)}\,dx, \quad i=1,2.$$

Apart from a set of Lebesgue measure zero, we have that

$$\begin{aligned}
\phi*\phi(z) &= \int_{\mathbb{R}}\phi(z-x)\phi(x)dx\\
&= \int_{\mathbb{R}}\mathbb{1}_{[0,\infty)}(z-x)\lambda\,\mathrm{e}^{-\lambda(z-x)}\,\mathbb{1}_{[0,\infty)}(x)\lambda\,\mathrm{e}^{-\lambda x}\,dx\\
&= \mathbb{1}_{[0,\infty)}(z)\lambda^2\,\mathrm{e}^{-\lambda z}\left(\int_0^z dx\right)\\
&= \mathbb{1}_{[0,\infty)}(z)\lambda^2 z\,\mathrm{e}^{-\lambda z}.
\end{aligned}$$

We apply item (ii) of Proposition 11.7 and deduce that the probability density function of the law of $Z=X_1+X_2$ is given by $\phi*\phi(z)$, $z\in\mathbb{R}$. We remark that a random variable with law $P_Z(dz)=\phi_Z(z)dz$ is referred to as Gamma distributed with parameters 2 and λ.

Exercise 11.4. *Suppose that X_1 and X_2 are two independent random variables with law that is geometric with parameter p, i.e., X_1 and X_2 have support $\mathbb{N}$ and $\mathbb{P}(X_i=k)=(1-p)^{k-1}p$, $k\in\mathbb{N}$, $i=1,2$. Find the law of X_1+X_2.*

The following result is helpful in characterizing the law of the sum of independent random vectors:

Proposition 11.8. *Assume that $X_1,\ldots,X_n$ are n independent random vectors s.t. $X_i\colon\Omega\to\mathbb{R}^k$, $i=1,\ldots,n$. Let $Z=\sum_{i=1}^n X_i$. Then,*

$$\Phi_Z(v)=\prod_{i=1}^n\Phi_{X_i}(v), \quad v\in\mathbb{R}^k,$$

i.e., the characteristic function of Z equals the product of the characteristic functions of X_i, $i=1,\ldots,k$.

Example 11.8. *Upon the latter result we readily check that if $X_1, \dots, X_n$ are n independent random variables s.t. $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, \dots, n$, then $\sum_{i=1}^n X_i \sim \mathcal{N}(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$. To see it, we apply Proposition 11.8 and obtain with Proposition 10.11 that for any $v \in \mathbb{R}$,*

$$\Phi_{\sum_{i=1}^n X_i}(v) = \mathrm{e}^{i(\sum_{i=1}^n \mu_i)v - \frac{(\sum_{i=1}^n \sigma_i^2)v^2}{2}}.$$

Then, by Proposition 10.12, the sum $\sum_{i=1}^n X_i$ is Gaussian with mean $\sum_{i=1}^n \mu_i$ and variance $\sum_{i=1}^n \sigma_i^2$.

11.3 Gauss vectors

Definition 11.6. *A random vector $X = (X_1, \dots, X_k)$ is said to be a Gauss vector if and only if for any $v \in \mathbb{R}^k$, the random variable*

$$v^t X = v_1 X_1 + \cdots v_k X_k,$$

is Gaussian.

Remark 11.11. *Regarding the previous definition, several notes are helpful.*

(i) If $X = (X_1, \dots, X_k)$ is a Gauss vector, then for any $i = 1, \dots, k$, X_i is Gaussian. We take $v = (0, \dots, 0, 1, 0, \dots, 0) = e_i$, $i = 1, \dots, k$, and obtain that $e_i^t X = X_i$ is Gaussian.

(ii) If for any $i = 1, \dots, k$, X_i is Gaussian, then $X = (X_1, \dots, X_k)$ is not necessarily a Gauss vector. For a counter example, let $X_1 \sim \mathcal{N}(0, 1)$ and $X_2 = GX_1$, with G as in Example 11.1. That is, X_1 and X_2 are as in Remark 11.5 and we already know that X_2 is Gaussian as well. We have that

$$\begin{aligned}
\mathbb{P}(X_1 + X_2 = 0) &= \mathbb{P}(X_1(1+G) = 0) \\
&= \mathbb{P}(X_1(1+G) = 0, X_1 = 0) + \mathbb{P}(X_1(1+G) = 0, X_1 \neq 0) \\
&= \mathbb{P}(X_1(1+G) = 0, X_1 \neq 0) = \mathbb{P}(G = -1) = 1/2.
\end{aligned}$$

Thus, $X_1 + X_2$ is not Gaussian and hence (X_1, X_2) is not Gaussian.

(iii) If $X_1, \dots, X_k$ are independent Gaussian random variables, then $X = (X_1, \dots, X_k)$ is a Gauss vector. This is Example 11.8 (see also Exercise 10.9). It is not sufficient that $X_1, \dots, X_k$ are pairwise uncorrelated (cf. Remark 11.5). However, we will see later that if X is a Gauss vector, then $X_1, \dots, X_k$ are independent if and only if $X_1, \dots, X_k$ are pairwise uncorrelated (cf. Proposition 11.10).

Definition 11.7. *We use the notation $X \sim \mathcal{N}(\mu, \Sigma(X))$ to indicate that X is a Gauss vector with mean vector μ and covariance matrix $\Sigma(X)$. If $\mu = 0$ and $\Sigma(X) = I$, where*

$$I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix},$$

X is referred to as a standard Gauss vector.

Proposition 11.9. *Let $X = (X_1, \dots, X_k)$ be a Gauss vector with mean $\mu = (\mu_1, \dots, \mu_k)$ and covariance matrix $\Sigma(X)$. Then, the characteristic function of X is given by*

$$\Phi_X(v) = \mathrm{e}^{i\mu^t v - \frac{v^t \Sigma(X) v}{2}}, \quad v \in \mathbb{R}^k.$$

Proof. Let $v \in \mathbb{R}^k$. Since X is a Gauss vector, $v^t X$ is a Gaussian random variable with expectation $\mu^t v$ and variance $v^t \Sigma(X) v$ (cf. Proposition 10.9). Therefore, by Proposition 10.11,

$$\Phi_X(v) = \mathbb{E}[\mathrm{e}^{i v^t X}] = \Phi_{v^t X}(1) = \mathrm{e}^{i(\mu^t v) - \frac{(v^t \Sigma(X) v)}{2}} .$$

□

Exercise 11.5. *Suppose that $X \sim \mathcal{N}(\mu, \Sigma(X))$, $\mu = (\mu_1, \dots, \mu_k)$ and $\Sigma(X)_{i,i} = \sigma_i^2$, $i = 1, \dots, k$. Show that $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, \dots, k$.*

Proposition 11.10. *Suppose that $X = (X_1, \dots, X_k) \sim \mathcal{N}(\mu, \Sigma(X))$. Then, $X_1, \dots, X_k$ are independent if and only if*

$$\Sigma(X) = \begin{pmatrix} \sigma_{1,1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{2,2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{k,k}^2 \end{pmatrix},$$

i.e., $\Sigma(X)$ is diagonal.

Proof. Suppose that $X_1, \dots, X_k$ are independent. Then, for any $i, j = 1, \dots, k$, $i \neq j$, $\mathrm{Cov}(X_i, X_j) = 0$ (cf. Remark 11.3 and Proposition 11.3). In particular, $\Sigma(X)$ is diagonal. For the other direction, assume that $\Sigma(X)$ is diagonal with $\Sigma(X)_{i,i} = \sigma_i^2$, $i = 1, \dots, k$. Then, for any $v = (v_1, \dots, v_k) \in \mathbb{R}^k$, by Proposition 11.9,

$$\Phi_X(v) = \mathrm{e}^{i\mu^t v - \frac{v^t \Sigma(X) v}{2}} = \mathrm{e}^{\sum_{i=1}^k \left(i\mu_i v_i - \frac{\sigma_i^2 v_i^2}{2}\right)} = \prod_{i=1}^k \mathrm{e}^{i\mu_i v_i - \frac{\sigma_i^2 v_i^2}{2}} .$$

Then, since $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ (cf. Exercise 11.5), it follows that $\Phi_X(v) = \prod_{i=1}^k \Phi_{X_i}(v_i)$. Hence, we apply Proposition 11.5 and the proposition is proven. □

Remark 11.12. *Upon Remark 11.5, it is possible that $X_1 \sim \mathcal{N}(0,1)$ and $X_2 \sim \mathcal{N}(0,1)$ s.t. $\Sigma(X) = I$, $X = (X_1, X_2)$, but X_1 and X_2 are not independent. Thus, we already know that Proposition 11.10 is not true in general (i.e. it holds for Gauss vectors).*

Proposition 11.11. *Assume that $X \sim \mathcal{N}(\mu, \Sigma(X))$, where $\Sigma(X)$ is positive definite. Then, the law of X has probability density function $\phi(x)$, $x \in \mathbb{R}^k$, given by (27), i.e., P_X is multivariate normal with probability density function*

$$\phi(x) = \frac{1}{\sqrt{(2\pi)^k \det \Sigma(X)}} \mathrm{e}^{-\frac{1}{2}(x-\mu)^t \Sigma(X)^{-1} (x-\mu)}, \quad x \in \mathbb{R}^k.$$

Remark 11.13. *Assuming that $\Sigma(X)$ is positive definite implies that the k eigenvalues of $\Sigma(X)$ are strictly positive (recall that $\Sigma(X)$ is symmetric). In particular, if $\lambda_1, \dots, \lambda_k$ are the k eigenvalues of $\Sigma(X)$, $\det \Sigma(X) = \prod_{i=1}^k \lambda_i \neq 0$. That is, $\Sigma(X)$ is invertible.*

A proof of Proposition 11.11 is given in Section C.6 of the appendix.

11.4 A note on conditional probabilities

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $B \in \mathcal{F}$ s.t. $\mathbb{P}(B) > 0$. Then, define the function

$$\mathbb{P}_B(A) = \mathbb{P}(A \,|\, B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad A \in \mathcal{F}.$$

We readily verify that $A \mapsto \mathbb{P}_B(A)$ defines a new measure on $\mathcal{F}$ according to Definition 5.1. In particular, $\mathbb{P}_B(\Omega) = 1$, i.e., $A \mapsto \mathbb{P}_B(A)$ is a probability on $\mathcal{F}$. The latter measure is referred to as the conditional probability given the event B. If X is a random variable, either nonnegative or integrable with respect to $\mathbb{P}$, then the conditional expectation of X given the event B is defined as

$$\mathbb{E}[X \mid B] = \frac{\mathbb{E}[X\mathbb{1}_B]}{\mathbb{P}(B)}.$$

We remark that by Exercise 8.8, if we set $\mu_B(A) = \mathbb{P}(A \cap B)$, $A \in \mathcal{F}$, we deduce that

$$\int_\Omega X(\omega)\mathbb{P}_B(d\omega) = \frac{\int_\Omega X(\omega)\mu_B(d\omega)}{\mathbb{P}(B)} = \mathbb{E}[X \mid B].$$

Thus, the expectation of X with respect to the conditional measure $\mathbb{P}_B$ is the conditional expectation of X given the event B. For further reading on conditional probabilities we refer to [1] and [2].

11.5 Solution to exercises

Solution 11.1 (Solution to Exercise 11.1)**.** *Suppose that $i \notin E$, then $\mathbb{P}(Y = i) = 0$. Hence, Y has support E. Let $i \in E$. Using the assumption that X and G are independent, we have that*

$$\begin{aligned}
\mathbb{P}(Y = i) &= \mathbb{P}(\{Y = i\} \cap \{G = 1\}) + \mathbb{P}(\{Y = i\} \cap \{G = -1\}) \\
&= \mathbb{P}(\{X = i\} \cap \{G = 1\}) + \mathbb{P}(\{X = -i\} \cap \{G = -1\}) \\
&= \mathbb{P}(X = i)\mathbb{P}(G = 1) + \mathbb{P}(X = -i)\mathbb{P}(G = -1) = 1/11.
\end{aligned}$$

Thus, Y has the same law as X.

Solution 11.2 (Solution to Exercise 11.2)**.** *Let $m \notin E$. We have that $\mathbb{P}(M = m) = 0$. Thus, M has support E. Let $m \in E$, $m \geq -4$. We recall Remark 10.7 and notice that*

$$\mathbb{P}(M = m) = \mathbb{P}(M \leq m) - \mathbb{P}(M \leq m - 1).$$

Then, since X_1 and X_2 are assumed to be independent with same law,

$$\mathbb{P}(M \leq m) = \mathbb{P}(\{X_1 \leq m\} \cap \{X_2 \leq m\}) = \mathbb{P}(X_1 \leq m)^2.$$

We calculate

$$\begin{aligned}
\mathbb{P}(X_1 \leq m) &= \sum_{\{i \in E\colon\, i \leq m\}} \mathbb{P}(X_1 = i) \\
&= \frac{\#\{i \in E\colon i \leq m\}}{11} \\
&= \frac{\#\{i\colon\, -5 \leq i \leq m\}}{11} = \frac{\#\{i\colon\, -5+5+1 \leq i \leq m+5+1\}}{11} = \frac{m+5+1}{11}.
\end{aligned}$$

The same argument shows that $\mathbb{P}(X_1 \leq m - 1) = (m + 5)/11$. Therefore, for $m \in E$, $m \geq -4$,

$$\mathbb{P}(M = m) = \frac{1}{11^2}\Big((m+5+1)^2 - (m+5)^2\Big) = \frac{2m+11}{121}. \tag{39}$$

If $m = -5$, $\mathbb{P}(M = m) = \mathbb{P}(\{X_1 = -5\} \cap \{X_2 = -5\}) = 1/121 = (2m + 11)/121$. In conclusion, for any $m \in E$, the law of M is given by (39).

Solution 11.3 (Solution to Exercise 11.3)**.** *Given any $t \in \mathbb{R}$, using the assumption that G and X are independent, we have that*

$$\begin{aligned}
\mathbb{P}(Y \leq t) &= \mathbb{P}(Y \leq t, G = -1) + \mathbb{P}(Y \leq t, G = 1) \\
&= \mathbb{P}(X \geq -t)\mathbb{P}(G = -1) + \mathbb{P}(X \leq t)\mathbb{P}(G = 1) \\
&= \big(1 - F_X(-t)\big)\frac{1}{2} + F_X(t)\frac{1}{2}.
\end{aligned}$$

Thus, by Exercise 10.7, we deduce that for any $t \in \mathbb{R}$, $F_Y(t) = F_X(t)$. Hence, upon Exercise 10.5, this shows that $Y \sim \mathcal{N}(0,1)$.

Solution 11.4 (Solution to Exercise 11.4)**.** *The support of $X_1 + X_2$ is $\mathbb{N} + \mathbb{N} = \mathbb{N} \setminus \{1\}$. Given $z \in X_1 + X_2$, we have*

$$\begin{aligned}\mathbb{P}(X_1 + X_2 = z) &= \sum_{k \in \mathbb{N}} \mathbb{P}(X_1 = z-k)\mathbb{P}(X_2 = k) \\ &= \sum_{k=1}^{z-1} \mathbb{P}(X_1 = z-k)\mathbb{P}(X_2 = k) = (z-1)p^2(1-p)^{z-2}.\end{aligned}$$

Solution 11.5 (Solution to Exercise 11.5)**.** *Let $i = 1, \ldots, k$. We know that X_i is Gaussian (cf. item (i) of Remark 11.11). Define the random vector*

$$Y_i = (0, \ldots, 0, X_i, 0, \ldots, 0).$$

We have that $\mathbb{E}[Y_i] = (0, \ldots, 0, \mu_i, 0, \ldots, 0)$ and

$$\Sigma(Y_i)_{j,k} = \begin{cases} \sigma_i^2, & \text{if } j = k = i, \\ 0, & \text{otherwise.} \end{cases}$$

We deduce that for any $v = (v_1, \ldots, v_k) \in \mathbb{R}^k$,

$$\Phi_{Y_i}(v) = \mathbb{E}[\mathrm{e}^{iv_iX_i}] = \mathbb{E}[\mathrm{e}^{iw^tX}] = \Phi_X(w),$$

with $w = (0, \ldots, 0, v_i, 0, \ldots, 0)$. Thus, by Proposition 11.9, $\Phi_{Y_i}(v) = \mathrm{e}^{i\mu_i v_i - \frac{v_i^2\sigma_i^2}{2}}$. Therefore, for any $a \in \mathbb{R}$, with $b = (b_1, \ldots, b_{i-1}, a, b_{i+1}, \ldots, b_k)$,

$$\Phi_{X_i}(a) = \mathbb{E}[\mathrm{e}^{iaX_i}] = \mathbb{E}[\mathrm{e}^{ib^tY_i}] = \Phi_{Y_i}(b) = \mathrm{e}^{i\mu_i a - \frac{a^2\sigma_i^2}{2}}.$$

Thus, by Proposition 10.12, $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$.

11.6 Additional exercises

Exercise 11.6. *Let $X_1, \ldots, X_n$ be n random variables. Verify that if $X_1, \ldots, X_n$ are independent, then for any $t_1, \ldots, t_n \in \mathbb{R}$,*

$$\mathbb{P}(X_1 \leq t_1, \ldots, X_n \leq t_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq t_i). \tag{40}$$

Hint: *Proposition 11.1.*
Note: *One can also verify the converse of the latter statement, i.e., if (40) holds, then $X_1, \ldots, X_n$ are independent.*

Exercise 11.7. *Let $X_1, \ldots, X_n$ be a n independent discrete random variables where for any $i = 1, \ldots, n$, the law of X_i is discrete uniform on $\{1, \ldots, p\}$, $p \in \mathbb{N}$. Find the law of $M = \max\{X_1, \ldots, X_n\}$.*

Exercise 11.8. *Let X_1 and X_2 be two independent random variables s.t. X_1 and X_2 have Poisson law with parameters λ and μ, respectively. Show that the law of $X_1 + X_2$ is Poisson with parameter $\lambda + \mu$.*

Exercise 11.9. *Let $X = (X_1, \ldots, X_k)$ and $Y = (Y_1, \ldots, Y_k)$ be two independent random vectors s.t. for any $i = 1, \ldots, k$, $\mathrm{Var}(X_i)$ and $\mathrm{Var}(Y_i)$ are finite. Let $\Sigma(X)$ and $\Sigma(Y)$ be the covariance matrices of X and Y, respectively. Show that $\Sigma(X+Y) = \Sigma(X) + \Sigma(Y)$, where $\Sigma(X+Y)$ is the covariance matrix of $X+Y$.*

Exercise 11.10. *Let X and Y be two random variables.*

(a) *Show that the characteristic function Φ_X of X is real valued (i.e., $\Phi_X(v) \in \mathbb{R}$ for any $v \in \mathbb{R}$) if and only if $P_X = P_{-X}$, i.e., the law of X is symmetric around the origin.*

Hint: *Recall that the maps $x \mapsto \cos(x)$ and $x \mapsto \sin(x)$ are even and odd, respectively (cf. Example 9.3).*

(b) *Show that if X and Y are independent with equal law, then the random variable $Z = X - Y$ has law that is symmetric around the origin.*

12 Convergence of random vectors

12.1 General notions of Convergence

We recall that if nothing else is mentioned, any collection of random vectors is defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Further, if not specified otherwise, any sequence of random vectors $(X_n)_{n\in\mathbb{N}}$ is s.t. $X_n \colon \Omega \to \mathbb{R}^k$ for any $n \in \mathbb{N}$. If $k = 1$, $(X_n)_{n\in\mathbb{N}}$ is a sequence of random variables. We also keep in mind Sections 3.4 and 7.3.

Definition 12.1. *Let $(X_n)_{n\in\mathbb{N}}$ be s sequence of random vectors.*

Almost sure convergence: *$(X_n)_{n\in\mathbb{N}}$ converges almost surely to a random vector X if*

$$\mathbb{P}(\{\omega \in \Omega \colon X_n(\omega) \xrightarrow{n\to\infty} X(\omega)\}) = \mathbb{P}(X_n \xrightarrow{n\to\infty} X) = \mathbb{P}(\lim_{n\to\infty} X_n = X) = 1.$$

To indicate that $(X_n)_{n\in\mathbb{N}}$ converges almost surely to X, we use the notation $X_n \to_{a.s.} X$.

L^1 and L^2 convergence: *Let $p = 1, 2$. Assume that $\mathbb{E}[\|X_n\|^p] < \infty$ for any $n \in \mathbb{N}$ and suppose that X is a random vector s.t. $\mathbb{E}[\|X\|^p] < \infty$. Then, $(X_n)_{n\in\mathbb{N}}$ converges in L^p to X if*

$$\lim_{n\to\infty} \mathbb{E}[\|X_n - X\|^p] = 0. \tag{41}$$

To indicate that $(X_n)_{n\in\mathbb{N}}$ converges in L^p to X, we use the notation $X_n \to_{L^p} X$.

Convergence in probability: *$(X_n)_{n\in\mathbb{N}}$ converges in probability to a random vector X if for any $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}(\|X_n - X\| > \varepsilon) = 0.$$

To indicate that $(X_n)_{n\in\mathbb{N}}$ converges in probability to X, we use the notation $X_n \to_{\mathbb{P}} X$.

Remark 12.1. *In either case, $X_n \to_{a.s.} X$, $X_n \to_{L^p} X$ or $X_n \to_{\mathbb{P}} X$, the limit X is assumed to be a random vector, i.e., X is $\mathcal{F}/\mathfrak{B}(\mathbb{R}^k)$ measurable. In particular, by Proposition 7.9, the set $\{X_n \xrightarrow{n\to\infty} X\}$ is an element of $\mathcal{F}$. Notice that almost sure convergence $X_n \to_{a.s.} X$ is a special case of a.e. convergence of a sequence of measurable function X_n, $n \in \mathbb{N}$, to X where the measure is a probability and the limit is a measurable function as well. If $p = 1$, then upon the triangular inequality (cf. Proposition 2.10), $\mathbb{E}[\|X\|] < \infty$ and $\mathbb{E}[\|X\|] < \infty$ imply that $\mathbb{E}[\|X_n - X\|] < \infty$. If $p = 2$, then,*

$$\mathbb{E}[\|X_n - X\|^2] = \mathbb{E}\bigg[\sum_{i=1}^{k} (X_i^n - X_i)^2\bigg] = \sum_{i=1}^{k} \mathbb{E}[(X_i^n - X_i)^2],$$

where $X_n = (X_1^n, \dots, X_k^n)$ and $X = (X_1, \dots, X_k)$. Hence, by Proposition 10.8, $\mathbb{E}[\|X_n\|^2] < \infty$ and $\mathbb{E}[\|X\|^2] < \infty$ imply that $\mathbb{E}[\|X_n - X\|^2] < \infty$. Further, if $X_n \to_{L^2} X$, then $X_n \to_{L^1} X$. This is because $\mathbb{E}[\|X_n\|^2] < \infty$ and $\mathbb{E}[\|X\|^2] < \infty$ imply that $\mathbb{E}[\|X_n\|] < \infty$ and $\mathbb{E}[\|X\|] < \infty$ (cf. Proposition 8.5) and

$$\mathrm{Var}(\|X_n - X\|) = \mathbb{E}[\|X_n - X\|^2] - (\mathbb{E}[\|X_n - X\|])^2 \geq 0, \tag{42}$$

i.e., $\mathbb{E}[\|X_n - X\|] \leq (\mathbb{E}[\|X_n - X\|^2])^{1/2}$. Finally we remark that L^p convergence is not restricted to $p = 1, 2$. Generally, $(X_n)_{n\in\mathbb{N}}$ converges in L^p $(p \geq 1)$ to X if (41) is satisfied.

When studying the convergence of random vectors, the following statement is helpful. The given result is known as the Borel-Cantelli lemma.

Proposition 12.1. *Let $\{A_n \colon n \in \mathbb{N}\} \subset \mathcal{F}$ be a sequence of events. Define*

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigg(\bigcup_{k=n}^{\infty} A_k \bigg).$$

(i) *If* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$*, then* $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$*.*

(ii) *If* $\{A_n : n \in \mathbb{N}\}$ *is independent (cf. Definition 11.4) and* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$*, then,* $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$*.*

Exercise 12.1. *Show that* $\limsup_{n\to\infty} A_n = \{\omega \in \Omega : \omega \in A_n$ *for infinitely many* $n\}$*.*

Remark 12.2. *Upon Exercise 12.1, if* $\{A_n : n \in \mathbb{N}\} \subset \mathcal{F}$ *is a sequence of events s.t.* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$*, (i) of Proposition 12.1 shows that*

$$\mathbb{P}(\{\omega \in \Omega : \omega \in A_n \text{ for finitely many } n\}) = 1.$$

If $\{A_n : n \in \mathbb{N}\}$ *is independent and* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$*, then, by (ii) of Proposition 12.1,*

$$\mathbb{P}(\{\omega \in \Omega : \omega \in A_n \text{ for infinitely many } n\}) = 1.$$

Proof of Proposition 12.1. We notice that by Proposition 8.4,

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} \mathbb{E}[\mathbb{1}_{A_n}] = \mathbb{E}\bigg[\sum_{i=1}^{\infty} \mathbb{1}_{A_n}\bigg].$$

Thus, if $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then, by Proposition 8.5, $\sum_{i=1}^{\infty} \mathbb{1}_{A_n} < \infty$ $\mathbb{P}$ a.s. This implies that $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$ (cf. Exercise 12.1). This shows (i). For the remaining, assume that $\{A_n : n \in \mathbb{N}\}$ is independent and $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$. Notice first that by Example A.5, we have that for any $x \in \mathbb{R}$, $1 - x \le \mathrm{e}^{-x}$. Hence, for any $N \in \mathbb{N}$,

$$\mathbb{P}\bigg(\bigcap_{k=N}^{n} A_k^c\bigg) = \prod_{k=N}^{n} \mathbb{P}(A_k^c) = \prod_{k=N}^{n} (1 - \mathbb{P}(A_k)) \le \mathrm{e}^{-\sum_{k=N}^{n} \mathbb{P}(A_k)}.$$

Then, since by assumption, $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, it follows that $\lim_{n\to\infty} \mathbb{P}(\cap_{k=N}^{n} A_k^c) = 0$. Therefore, by Proposition 5.1, for any $N \in \mathbb{N}$, $\mathbb{P}(\cap_{k=N}^{\infty} A_k^c) = 0$. Hence (cf. Proposition 5.1),

$$\mathbb{P}\bigg(\bigcup_{N=1}^{\infty}\bigg(\bigcap_{k=N}^{\infty} A_k^c\bigg)\bigg) \le \sum_{N=1}^{\infty} \mathbb{P}\bigg(\bigcap_{k=N}^{\infty} A_k^c\bigg) = 0.$$

This shows that $\mathbb{P}((\limsup_{n\to\infty} A_n)^c) = 0$ and completes the proof of (ii). ☐

The following two propositions show that convergence in probability is the weakest form of the three notions of convergence given in Definition 12.1.

Proposition 12.2. *Let* $p = 1, 2$*. Suppose that* $X_n \to_{L^p} X$*. Then,* $X_n \to_{\mathbb{P}} X$*.*

Proof. Since $X_n \to_{L^2} X$ implies that $X_n \to_{L^1} X$ (cf. Remark 12.1) it is enough to assume that $p = 1$. Let $\varepsilon > 0$. The proposition is a consequence of Proposition 10.4, we have that

$$\mathbb{P}(\|X_n - X\| > \varepsilon) \le \frac{\mathbb{E}[\|X_n - X\|]}{\varepsilon}.$$

☐

Proposition 12.3. *Suppose that* $X_n \to_{a.s.} X$*. Then,* $X_n \to_{\mathbb{P}} X$*.*

Proof. Consider the real valued sequence $d_n = \mathbb{E}[\min\{\|X_n - X\|, 1\}]$, $n \in \mathbb{N}$. We notice that if $\lim_{n\to\infty} d_n = 0$, then $X_n \to_{\mathbb{P}} X$. To see it, we rely on Proposition 10.4 and deduce that for any $\varepsilon \in (0, 1)$,

$$\mathbb{P}(\|X_n - X\| > \varepsilon) \le \mathbb{P}(\min\{\|X_n - X\|, 1\} > \varepsilon) \le \frac{d_n}{\varepsilon}.$$

Therefore, it is sufficient to show that $X_n \to_{a.s.} X$ implies that $\lim_{n\to\infty} d_n = 0$. Upon $X_n \to_{a.s.} X$ it follows that $\min\{\|X_n - X\|, 1\} \to_{a.s.} 0$. Hence, by the dominated convergence theorem (cf. Proposition 8.10), $\lim_{n\to\infty} d_n = 0$. ☐

Proposition 12.4. *If $X_n \to_{\mathbb{P}} X$ and $X_n \to_{\mathbb{P}} Y$, then $X = Y$ $\mathbb{P}$ a.s. The same is true if $X_n \to_{a.s.} X$ and $X_n \to_{a.s.} Y$ or $X_n \to_{L^p} X$ and $X_n \to_{L^p} Y$.*

Proof. By Propositions 12.2 and 12.3 it is sufficient to show that if $X_n \to_{\mathbb{P}} X$ and $X_n \to_{\mathbb{P}} Y$, then $X = Y$ $\mathbb{P}$ a.s. Given $\varepsilon > 0$, we have that

$$\mathbb{P}(\{\|X_n - X\| > \varepsilon\} \cup \{\|X_n - Y\| > \varepsilon\}) \leq \mathbb{P}(\|X_n - X\| > \varepsilon) + \mathbb{P}(\|X_n - Y\| > \varepsilon).$$

Thus, if $X_n \to_{\mathbb{P}} X$ and $X_n \to_{\mathbb{P}} Y$, we deduce that

$$\lim_{n\to\infty} \mathbb{P}(\{\|X_n - X\| > \varepsilon\} \cup \{\|X_n - Y\| > \varepsilon\}) = 0.$$

In particular,

$$\lim_{n\to\infty} \mathbb{P}(\{\|X_n - X\| \leq \varepsilon\} \cap \{\|X_n - Y\| \leq \varepsilon\}) = 1. \tag{43}$$

Therefore, if $\omega \in \{\|X_n - X\| \leq \varepsilon/2\} \cap \{\|X_n - Y\| \leq \varepsilon/2\}$, it follows that

$$\|X(\omega) - Y(\omega)\| = \|X(\omega) - X_n(\omega) + X_n(\omega) - Y(\omega)\| \leq \|X_n - X\| + \|X_n - Y\| \leq \varepsilon.$$

Since $\mathbb{P}(\|X(\omega) - Y(\omega)\| \leq \varepsilon)$ is constant in n, it follows from (43) that

$$\mathbb{P}(\|X(\omega) - Y(\omega)\| \leq \varepsilon) = 1.$$

Therefore, $\sum_{n\in\mathbb{N}} \mathbb{P}(\|X(\omega) - Y(\omega)\| > 1/n) = 0$. Hence, by item (i) of Proposition 12.1,

$$\mathbb{P}(X = Y) = \mathbb{P}\left(\bigcup_{n=1}^{\infty}\left(\bigcap_{k=n}^{\infty}\{\|X(\omega) - Y(\omega)\| \leq 1/k\}\right)\right) = 1.$$

□

Example 12.1. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables s.t. for any $n \in \mathbb{N}$, X_n takes the values $1 - (1/n)$ and $1 + (1/n)$ with probability $1/2$, i.e., for any $n \in \mathbb{N}$, the law of X_n is s.t. $P_{X_n}(\{1 - (1/n)\}) = P_{X_n}(\{1 + (1/n)\}) = 1/2$. We have that*

$$\mathbb{E}[|X_n - 1|^2] = \frac{1}{n^2} \xrightarrow{n\to\infty} 0.$$

Hence, $X_n \to_{L^2} 1$. In particular, $X_n \to_{L^1} 1$ (cf. Remark 12.1). Upon Proposition 12.2, $X_n \to_{\mathbb{P}} 1$. Let $A_n = \{|X_n - 1| > 1/n\}$, $n \in \mathbb{N}$. We have that $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = 0$. Hence, by item (i) of Proposition 12.1,

$$1 = \mathbb{P}\left(\bigcup_{n=1}^{\infty}\left(\bigcap_{k=n}^{\infty}\{|X_k - 1| \leq 1/k\}\right)\right) = \mathbb{P}(X_n \xrightarrow{n\to\infty} 1).$$

Example 12.2. *Assume that $(X_n)_{n\in\mathbb{N}}$ is an independent sequence of random variables (cf. Definition 11.4) s.t. for any $n \in \mathbb{N}$,*

$$\mathbb{P}(X_n = 0) = 1 - \frac{1}{n} \quad \textit{and} \quad \mathbb{P}(X_n = 1) = \frac{1}{n}.$$

We have that

$$\mathbb{E}[|X_n - 0|^2] = \left(1 - \frac{1}{n}\right)0^2 + \frac{1}{n}1^2 \xrightarrow{n\to\infty} 0,$$

i.e., $X_n \to_{L^p} 0$, $p = 1, 2$, and in particular $X_n \to_{\mathbb{P}} 0$. Set $A_n = \{X_n = 1\}$. We have that $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} 1/n = \infty$ (cf. Example 3.5). Therefore, by item (ii) of Proposition 12.1,

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty}\left(\bigcup_{k=n}^{\infty}\{X_k = 1\}\right)\right) = 1.$$

Thus, $\mathbb{P}(X_n = 1$ *for infinitely many* $n) = 1$*. In particular, since*

$$\{X_n = 1 \text{ for infinitely many } n\} \subset \{\limsup_{n\to\infty} X_n \geq 1\} \subset \{\lim_{n\to\infty} X_n \neq 0\},$$

we obtain $\mathbb{P}(X_n \xrightarrow{n\to\infty} 0) = 0$*. That is,* $(X_n)_{n\in\mathbb{N}}$ *does not converges almost surely to* 0*.*

Example 12.3. *Assume that* $(X_n)_{n\in\mathbb{N}}$ *is a sequence of random variables s.t. for any* $n \in \mathbb{N}$*,*

$$\mathbb{P}(X_n = 0) = 1 - \frac{1}{2^n} \quad \text{and} \quad \mathbb{P}(X_n = 2^{n+1}) = \frac{1}{2^n}.$$

Then, for any $n \in \mathbb{N}$*,* $\mathbb{E}[|X_n - 0|] = 2$*, i.e., it is not true that* $X_n \to_{L^p} 0$*,* $p = 1, 2$*. However, we have that* $\sum_{n=1}^{\infty} \mathbb{P}(X_n \neq 0) = 1 < \infty$*, i.e., by item (i) of Proposition 12.1,*

$$\mathbb{P}(X_n \neq 0 \text{ for infinitely many } n) = 0.$$

We observe that

$$\{X_n \neq 0 \text{ for infinitely many } n\}^c \subset \{\exists\, N \text{ s.t. } X_n = 0 \;\forall\, n \geq N\}.$$

This is because $\{\exists\, N \text{ s.t. } X_n = 0 \;\forall\, n \geq N\}^c \subset \{X_n \neq 0 \text{ for infinitely many } n\}$*. In conclusion,* $\mathbb{P}(\{\exists\, N \text{ s.t. } X_n = 0 \;\forall\, n \geq N\}) = 1$ *and hence* $X_n \to_{a.s.} 0$*.*

Exercise 12.2. *Let* $(X_n)_{n\in\mathbb{N}}$ *be a sequence of random variables s.t. for any* $n \in \mathbb{N}$*,*

$$\mathbb{P}(X_n = n) = 1 - \frac{1}{2^n} \quad \text{and} \quad \mathbb{P}(X_n = 1/n) = \frac{1}{2^n}.$$

Show that there exists no random variable X *s.t.* $(X_n)_{n\in\mathbb{N}}$ *converges to* X *in probability.*
Note: *This implies that* $(X_n)_{n\in\mathbb{N}}$ *can not converge in* L^1 *or almost surely (cf. Propositions 12.2 and 12.3).*

Exercise 12.3. *Let* $(X_n)_{n\in\mathbb{N}}$ *be a sequence of independent random variables s.t. for any* $n \in \mathbb{N}$*,*

$$\mathbb{P}(X_n = 0) = 1 - \frac{1}{n} \quad \text{and} \quad \mathbb{P}(X_n = n) = \frac{1}{n}.$$

Show that $X_n \to_{\mathbb{P}} 0$ *but* $(X_n)_{n\in\mathbb{N}}$ *neither converges in* L^1 *nor almost surely to zero.*

Remark 12.3. *By Exercise 12.2, it is possible that* $(X_n)_{n\in\mathbb{N}}$ *converges in* L^2 *to a random variable* X *but it does not hold that* $(X_n)_{n\in\mathbb{N}}$ *converges almost surely to* X*. In addition, by Example 12.3, it is possible that* $X_n \to_{a.s.} X$ *where the convergence is not in* L^1*. By Exercise 12.3, there exists* $(X_n)_{n\in\mathbb{N}}$ *that converge in probability to* X *but the convergence is neither in* L^1 *nor almost surely.*

The following proposition shows that given a sequence of random vectors that converges in probability, one can always extract a subsequence that converges almost surely.

Proposition 12.5. *Let* $(X_n)_{n\in\mathbb{N}}$ *be a sequence of random vectors s.t.* $X_n \to_{\mathbb{P}} X$*. Then, there exists a subsequence* $(X_{s(n)})_{n\in\mathbb{N}}$ *s.t.* $X_{s(n)} \to_{a.s.} X$*.*

Proof. Define $\varepsilon_m = 1/m$ and $a_m = 2^{-m}$, $m \in \mathbb{N}$. Then, given any $m \in \mathbb{N}$, since $X_n \to_{\mathbb{P}} X$,

$$\lim_{n\to\infty} \mathbb{P}(\|X_n - X\| > \varepsilon_m) = 0.$$

By the latter convergence, we set $m = 1$ and find $s(1) \in \mathbb{N}$ s.t. $\mathbb{P}(\|X_{s(1)} - X\| > \varepsilon_1) \leq a_1$. Then, for $m = 2$, we find $s(2) > s(1)$ s.t. $\mathbb{P}(\|X_{s(2)} - X\| > \varepsilon_2) \leq a_2$. If we continue like this, we obtain a subsequence $(X_{s(n)})_{n\in\mathbb{N}}$ s.t. for any $n \in \mathbb{N}$, $\mathbb{P}(\|X_{s(n)} - X\| > \varepsilon_n) \leq a_n$.

We notice that $\sum_{n\in\mathbb{N}} \mathbb{P}(\|X_{s(n)} - X\| > \varepsilon_n) \le \sum_{n=1}^{\infty} 2^{-n} = 1$. Thus, by the first item of the Borel-Cantelli lemma (item (i) of Proposition 12.1),

$$1 = \mathbb{P}\bigg(\bigcup_{n\in\mathbb{N}} \bigcap_{k=n}^{\infty} \{\|X_{s(k)} - X\| \le \varepsilon_k\}\bigg) = \mathbb{P}(\exists\, N \in \mathbb{N} \text{ s.t. } \forall\, n \ge N\, \|X_{s(n)} - X\| \le \varepsilon_n).$$

Thus, $X_{s(n)} \to_{a.s.} X$. □

A consequence for sequences of random variables is the following result:

Proposition 12.6. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables s.t. $X_n \to_{\mathbb{P}} X$. Suppose that there exists a random variable Y s.t. $\mathbb{E}[|Y|] < \infty$ and for any $n \in \mathbb{N}$, $|X_n| \le Y$ $\mathbb{P}$ a.s. Then, $X_n \to_{L^1} X$.*

Proof. Let $(X_{s(n)})_{n\in\mathbb{N}}$ be any subsequence of $(X_{(n)})_{n\in\mathbb{N}}$. Since $X_n \to_{\mathbb{P}} X$, we readily check that $X_{s(n)} \to_{\mathbb{P}} X$ as well. Notice that given any $\varepsilon > 0$, $(\mathbb{P}(|X_{s(n)} - X| > \varepsilon))_{n\in\mathbb{N}}$ is a subsequence of $(\mathbb{P}(|X_n - X| > \varepsilon))_{n\in\mathbb{N}}$. Since $X_{s(n)} \to_{\mathbb{P}} X$, we rely on Proposition 12.5 and find a subsequence $(X_{t(s(n))})_{n\in\mathbb{N}}$ s.t. $X_{t(s(n))} \to_{a.s.} X$. Then, by assumption, for any $n \in \mathbb{N}$, $|X_{t(s(n))}| \le Y$ $\mathbb{P}$ a.s. Hence, upon Lebesgues dominated convergence theorem (cf. Proposition 8.10), X is integrable with respect to $\mathbb{P}$, i.e., $\mathbb{E}[|X|] < \infty$. Similarly, since $|X_{t(s(n))} - X| \le 2|Y|$ $\mathbb{P}$ a.s. and $|X_{t(s(n))} - X| \xrightarrow{n\to\infty} 0$ $\mathbb{P}$ a.s., we apply Lebesgues dominated convergence theorem to the sequence $(|X_{t(s(n))} - X|)_{n\in\mathbb{N}}$ and deduce that

$$\lim_{n\to\infty} \mathbb{E}[|X_{t(s(n))} - X|] = 0. \tag{44}$$

This shows that $X_n \to_{L^1} X$ since we started with an arbitrary subsequence $(\mathbb{E}[|X_{s(n)} - X|])_{n\in\mathbb{N}}$ of $(\mathbb{E}[|X_n - X|])_{n\in\mathbb{N}}$ and found a subsequnece $(\mathbb{E}[|X_{t(s(n))} - X|])_{n\in\mathbb{N}}$ s.t. (44) holds (cf. Proposition 3.24). □

Proposition 12.7. *Let $X_n = (X_1^n, \dots, X_k^n)$, $n \in \mathbb{N}$, be a sequence of random vectors and $X = (X_1, \dots, X_k)$ be a random vector.*

(i) $X_n \to_{a.s.} X$ if and only if for any $i = 1, \dots, k$, $X_i^n \to_{a.s.} X_i$;

(ii) if $X_n \to_{a.s.} X$, then for any continuous function $g\colon \mathbb{R}^k \to \mathbb{R}^m$, $g(X_n) \to_{a.s.} g(X)$;

(iii) $X_n \to_{\mathbb{P}} X$ if and only if for any $i = 1, \dots, k$, $X_i^n \to_{\mathbb{P}} X_i$;

(iv) if $X_n \to_{\mathbb{P}} X$, then for any continuous function $g\colon \mathbb{R}^k \to \mathbb{R}^m$, $g(X_n) \to_{\mathbb{P}} g(X)$.

Proof. We leave items (i) and (ii) as an exercise (cf. Exercise 12.6). We show item (iii). Suppose that $X_n \to_{\mathbb{P}} X$. Let $\varepsilon > 0$. Given any $i = 1, \dots, k$, we have that

$$\mathbb{P}(|X_i^n - X_i^n| > \varepsilon) \le \mathbb{P}(\|X_n - X\| > \varepsilon) \xrightarrow{n\to\infty} 0.$$

For the other direction, suppose that for any $i = 1, \dots, k$, $X_i^n \to_{\mathbb{P}} X_i$. Let $\varepsilon > 0$. We have that

$$\begin{aligned} \mathbb{P}(\|X_n - X\| > \varepsilon) &\le \mathbb{P}\bigg(\bigcup_{i=1}^{k} \Big\{|X_i^n - X_i| > \frac{\varepsilon}{\sqrt{k}}\Big\}\bigg) \\ &\le \sum_{i=1}^{k} \mathbb{P}\big(|X_i^n - X_i| > \varepsilon/\sqrt{k}\big) \xrightarrow{n\to\infty} 0. \end{aligned}$$

Therefore, $X_n \to_{\mathbb{P}} X$. With regard to (iv), suppose that $X_n \to_{\mathbb{P}} X$ and let $g\colon \mathbb{R}^k \to \mathbb{R}^m$ be continuous. Let $(X_{s(n)})_{n\in\mathbb{N}}$ be an arbitrary subsequence of $(X_n)_{n\in\mathbb{N}}$. As in the proof of Proposition 12.6, $X_{s(n)} \to_{\mathbb{P}} X$. By Proposition 12.5, we find a subsequence $(X_{t(s(n))})_{n\in\mathbb{N}}$

s.t. $X_{t(s(n))} \to_{a.s.} X$. By (i), $g(X_{t(s(n))}) \to_{a.s.} g(X)$ and in particular, $g(X_{t(s(n))}) \to_{\mathbb{P}} g(X)$. Let $\varepsilon > 0$. We use the notation $\|\cdot\|_m$ for the Eucildean norm on $\mathbb{R}^m$. We have shown that for any arbitrary subsequnece $\mathbb{P}(\|g(X_{s(n)}) - g(X)\|_m > \varepsilon)$, $n \in \mathbb{N}$, there exists a subsequence $\mathbb{P}(\|g(X_{t(s(n))}) - g(X)\|_m > \varepsilon)$, $n \in \mathbb{N}$, s.t. $\lim_{n\to\infty} \mathbb{P}(\|g(X_{t(s(n))}) - g(X)\|_m > \varepsilon) = 0$. This shows that $\lim_{n\to\infty} \mathbb{P}(\|g(X_n) - g(X)\|_m > \varepsilon) = 0$, i.e., $g(X_n) \to_{\mathbb{P}} g(X)$ (cf. Proposition 3.24). □

Exercise 12.4. *Let $X_n = (X_1^n, \dots, X_k^n)$, $n \in \mathbb{N}$, and $X = (X_1, \dots, X_k)$ be random vectors. Given $p = 1, 2$, show that $X_n \to_{L^p} X$ if and only if for any $i = 1, \dots, k$, $X_i^n \to_{L^p} X_i$.*

Exercise 12.5. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables s.t. $X_n \to_{L^1} X$. Show that for any $g\colon \mathbb{R} \to \mathbb{R}$ continuous and bounded, $g(X_n) \to_{L^1} g(X)$.*

Remark 12.4. *One can show that the result of the previous exercise is not true in general if g is only assumed to be continuous but not necessarily bounded.*

The following is known as Scheffé's lemma:

Proposition 12.8. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables. Suppose that for any $n \in \mathbb{N}$, $\mathbb{E}[|X_n|] < \infty$ and $X_n \to_{\mathbb{P}} X$ where $\mathbb{E}[|X|] < \infty$. Then, $\mathbb{E}[|X_n|] \xrightarrow{n\to\infty} \mathbb{E}[|X|]$ implies that $X_n \to_{L^1} X$.*

Proof. Let $(X_{s(n)})_{n\in\mathbb{N}}$ be an arbitrary subsequence of $(X_n)_{n\in\mathbb{N}}$. Since $X_{s(n)} \to_{\mathbb{P}} X$, we apply Proposition 12.5 and find a subsequnece $(X_{t(s(n))})_{n\in\mathbb{N}}$ s.t. $X_{t(s(n))} \to_{a.s.} X$. Define

$$Y_n = |X_{t(s(n))}| + |X| - |X_{t(s(n))} - X|, \quad n \in \mathbb{N}.$$

We have that $Y_n(\omega) \geq 0$ for any $\omega \in \Omega$ and $Y_n \to_{a.s.} 2|X|$. By Fatou's lemma (cf. Proposition 8.9),

$$2\mathbb{E}[|X|] = \mathbb{E}[\liminf_{n\to\infty} Y_n] \leq \liminf_{n\to\infty} \mathbb{E}[Y_n].$$

Since by assumption, $\mathbb{E}[|X_n|] \xrightarrow{n\to\infty} \mathbb{E}[|X|]$, it follows that (cf. Proposition A.4)

$$\liminf_{n\to\infty} \mathbb{E}[Y_n] = 2\mathbb{E}[|X|] - \limsup_{n\to\infty} \mathbb{E}[|X_{t(s(n))} - X|].$$

Hence,

$$2\mathbb{E}[|X|] \leq 2\mathbb{E}[|X|] - \limsup_{n\to\infty} \mathbb{E}[|X_{t(s(n))} - X|],$$

and hence, $\limsup_{n\to\infty} \mathbb{E}[|X_{t(s(n))} - X|] = \limsup_{n\to\infty} \mathbb{E}[|X_{t(s(n))} - X|] = 0$. This shows that $X_{t(s(n))} \to_{L^1} X$. In particular, by Proposition 3.24, $X_n \to_{L^1} X$. □

Remark 12.5. *Notice that if $(X_n)_{n\in\mathbb{N}}$ is a sequence of random variables s.t. $X_n \to_{L^1} X$, then, upon Exercise 8.6 and the reverse triangular inequality,*

$$\big|\mathbb{E}[|X_n|] - \mathbb{E}[|X|]\big| \leq \mathbb{E}\big[\big||X_n| - |X|\big|\big] \leq \mathbb{E}[|X_n - X|],$$

and hence $\mathbb{E}[|X_n|] \xrightarrow{n\to\infty} \mathbb{E}[|X|]$. However, in general, $\mathbb{E}[|X_n|] \xrightarrow{n\to\infty} \mathbb{E}[|X|]$ does not imply that $X_n \to_{L^1} X$. As an example, let X_n, $n \in \mathbb{N}$, be s.t. $\mathbb{P}(X_n = 1) = 1/n$ and $\mathbb{P}(X_n = -1) = 1 - (1/n)$. Then, $\mathbb{E}[|X_n|] \xrightarrow{n\to\infty} 1$. However, $\lim_{n\to\infty} \mathbb{E}[|X_n - 1|] = 2 \neq 0$.

The following is known as the (strong) law of large numbers (a proof is given in Section C.7 of the appendix).

Proposition 12.9. *Let $(X_n)_{n\in\mathbb{N}}$ be an i.i.d. (cf. Definition 11.4) sequence of random variables s.t. $\mathbb{E}[|X_1|] < \infty$. Then,*

$$\frac{1}{n}(X_1 + \cdots + X_n) \to_{a.s.} \mathbb{E}[X_1].$$

Example 12.4. *Let* $(X_n)_{n\in\mathbb{N}}$ *be an i.i.d. sequence of random variables s.t.* $\mathbb{E}[|X_1|] < \infty$. *Assume that* $\mathbb{E}[X_1] = \mu$ *and* $\mathrm{Var}(X_1) = \sigma^2$, $\mu, \sigma^2 \in \mathbb{R}$, $\sigma^2 > 0$. *Define,* $\overline{X}_n = n^{-1}(X_1 + \cdots + X_n)$, $n \in \mathbb{N}$. *Then,*

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 \to_{a.s.} \sigma^2. \tag{45}$$

To see it, we write

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu + \mu - \overline{X}_n\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\Big((X_i-\mu)^2 + 2(X_i-\mu)(\mu-\overline{X}_n) - (\mu-\overline{X}_n)^2\Big) \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 + 2(\mu-\overline{X}_n)\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu) + (\mu-\overline{X}_n)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - (\mu-\overline{X}_n)^2.
\end{aligned}$$

By Proposition 11.6, the sequence of random variables $((X_n-\mu)^2)_{n\in\mathbb{N}}$ *is an independent sequence of random variables. Clearly, given any* $n \in \mathbb{N}$, *the law of* $(X_n-\mu)^2$ *is the same as the law of* $(X_1-\mu)^2$ *(cf. Proposition 10.3). Thus,* $((X_n-\mu)^2)_{n\in\mathbb{N}}$ *is an i.i.d. sequence of random variables s.t.* $\mathbb{E}[|(X_1-\mu)^2|] = \mathbb{E}[(X_1-\mu)^2] = \sigma^2$. *Upon the law of large numbers,*

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 \to_{a.s.} \sigma^2.$$

By the law of large number again, it also follows that $\mu - \overline{X}_n \to_{a.s.} 0$ *and in particular, by Proposition 12.7,* $(\mu-\overline{X}_n)^2 \to_{a.s.} 0$. *In conclusion, we apply Proposition 12.7 once more and (45) follows.*

Remark 12.6. *If* $X_n = (X_1^n, \ldots, X_k^n)$, $n \in \mathbb{N}$, *is a sequence of i.i.d. random vectors s.t.* $\mathbb{E}[|X_i^1|] < \infty$, $i = 1, \ldots, k$. *Then, for any* $i = 1, \ldots, k$, $(X_i^n)_{n\in\mathbb{N}}$ *is a sequence of i.i.d. random variables (independence follows from Proposition 11.6 with* $f_i^1(X_1) = X_i^1, \ldots, f_i^n(X_n) = X_i^n$, $n \in \mathbb{N}$, *and* X_i^n, $n \in \mathbb{N}$, *are identically distributed because of Proposition 10.3 upon the identification of* $X_i^1, \ldots, X_i^n$ *with the coordinate functions* $f_i^1, \ldots, f_i^n$, *as above). In particular, for any* $i = 1, \ldots, k$, *we apply the law of large numbers (cf. Proposition 12.9) to* $(X_i^n)_{n\in\mathbb{N}}$ *and obtain that* $n^{-1}(X_1 + \cdots + X_n) \to_{a.s.} \mathbb{E}[X_1]$.

12.2 A note on convergence in distribution

Definition 12.2. *A sequence of random vectors* $X_n = (X_1^n, \ldots, X_k^n)$, $n \in \mathbb{N}$, *converges in distribution to a random vector* X *(we write* $X_n \to_d X$*) if for any continuous and bounded function* $\varphi\colon \mathbb{R}^k \to \mathbb{R}$,

$$\mathbb{E}[\varphi(X_n)] \xrightarrow{n\to\infty} \mathbb{E}[X].$$

That is to say that the sequence of measures $(P_{X_n})_{n\in\mathbb{N}}$ *converges weakly to the measure* P_X.

The following proposition shows that convergence in distribution is the weakest form of the types of convergence we have seen so far.

Proposition 12.10. *Let* $(X_n)_{n\in\mathbb{N}}$ *be a sequence of random vectors s.t.* $X_n \to_{\mathbb{P}} X$. *Then,* $X_n \to_d X$.

To conclude, we list some results regarding convergence in distribution (for further reading we refer to [1] and [2]).

Remark 12.7. *Let $X_n = (X_1^n, \dots, X_k^n)$, $n \in \mathbb{N}$, be a sequence of random vectors and $X = (X_1, \dots, X_k)$ be a random vector.*

(i) *$X_n \to_d X$ does not imply that $P_{X_n}(B) \xrightarrow{n\to\infty} P_X(B)$ for any $B \in \mathfrak{B}(\mathbb{R}^k)$.*

(ii) *If $X_n \to_d X$, then for any rectangle $R_k = \prod_{i=1}^k [a_i, b_i]$, $a_i, b_i \in \mathbb{R}$, $P_{X_n}(R_k) \xrightarrow{n\to\infty} P_X(R_k)$.*

(iii) *Let X be a random vector and $A = \{t \in \mathbb{R}^k \colon F_X \text{ is continuous at } t\}$. Then, $X_n \to_d X$ if and only if $F_{X_n}(t) \xrightarrow{n\to\infty} F_X(t)$ for any $t \in A$.*

(iv) *Let $c \in \mathbb{R}^k$ and set $X(\omega) = c$ for any $\omega \in \Omega$. If $X_n \to_d c$, then $X_n \to_{\mathbb{P}} c$. That is, convergence in distribution to a constant implies convergence in probability to the given constant.*

(v) *$X_n \to_d X$ if and only if $\Phi_{X_n}(v) \xrightarrow{n\to\infty} \Phi_X(v)$ for any $v \in \mathbb{R}^k$. This is known as Lévy's theorem. It shows that convergence in distribution of a sequence of random vectors $(X_n)_{n\in\mathbb{N}}$ to X is equivalent to the pointwise convergence of the respective sequence of characteristic functions $(\Phi_{X_n})_{n\in\mathbb{N}}$ to the characteristic function of X.*

(vi) *Suppose that $(X_n)_{n\in\mathbb{N}}$ are i.i.d. and s.t. $\mathrm{Var}(X_i^n) < \infty$ for any $i = 1, \dots, k$. Then,*

$$\frac{X_1 + \cdots + X_n - n\mathbb{E}[X_1]}{\sqrt{n}} \to_d \mathcal{N}(0, \Sigma(X_1)). \tag{46}$$

For $k = 1$, (46) reads as $n^{-1/2}(X_1 + \cdots + X_n - n\mathbb{E}[X_1]) \to_d \mathcal{N}(0, \mathrm{Var}(X_1))$. The result given in (46) is known as the central limit theorem.

12.3 Solution to exercises

Solution 12.1 (Solution to Exercise 12.1)**.** *Let $\omega \in \limsup_{n\to\infty} A_n$. Then, for any $n \in \mathbb{N}$, there exists $k_n \geq n$ s.t. $\omega \in A_{k_n}$, i.e., $\omega \in A_{k_n}$, $n \in \mathbb{N}$. For the other inclusion, suppose that $\omega \in \Omega$ is s.t. $\omega \in A_k$, $k \in \mathbb{N}$. Then, for any $n \in \mathbb{N}$, there exists $k \geq N$, s.t. $\omega \in A_k$. Thus, $\omega \in \limsup_{n\to\infty} A_n$.*

Solution 12.2 (Solution to Exercise 12.2)**.** *Define $A_n = \{X_n \neq n\}$. We deduce that $\sum_{n=1}^\infty \mathbb{P}(A_n) = \sum_{n=1}^\infty (1/2)^n = 1$ (cf. Exercise 3.14). Therefore, by item (i) of Proposition 12.1,*

$$\mathbb{P}(X_n \neq n \textit{ for infinitely many } n) = 0.$$

Then, we notice that

$$\{X_n \neq n \textit{ for infinitely many } n\}^c \subset \{\exists\, N \textit{ s.t. } X_n = n \;\forall\, n \geq N\}.$$

Hence, $\mathbb{P}(\{\exists\, N \text{ s.t. } X_n = n \;\forall\, n \geq N\}) = 1$. Then, assume by contradiction that X is a random variable s.t. $X_n \to_{\mathbb{P}} X$. Let $\omega \in \{\exists\, N \text{ s.t. } X_n = n \;\forall\, n \geq N\}$. Find $N^* \in \mathbb{N}$ s.t. $N^* \geq N$ and $|X_n(\omega) - X(\omega)| > 1$ for any $n \geq N^*$. This shows that

$$\{\exists\, N \textit{ s.t. } X_n = n \;\forall\, n \geq N\} \subset \{\exists\, N^* \textit{ s.t. } |X_n - X| > 1 \;\forall\, n \geq N^*\}.$$

Write $A = \{\exists\, N^* \text{ s.t. } |X_n - X(\omega)| > 1 \;\forall\, n \geq N^*\}$. Given any $k \geq N^*$, we obtain

$$1 = \mathbb{P}(A) = \mathbb{P}(\{|X_k - X| > 1\} \cap A) \leq \mathbb{P}(|X_k - X| > 1).$$

Therefore, for any $k \geq N^*$, $\mathbb{P}(|X_k - X| > 1) = 1$. This contradicts $X_n \to_{\mathbb{P}} X$.

Solution 12.3 (Solution to Exercise 12.3). *As in Example 12.2, we apply item (ii) of Proposition 12.1 and deduce that*

$$\mathbb{P}(X_n = n \textit{ for infinitely many } n) = 1.$$

Hence, it is not possible that X_n converges almost surely to zero. Additionally, $\mathbb{E}[|X_n|] = 1$ for any $n \in \mathbb{N}$, i.e., it is also not true that X_n converges in L^1 to zero. Let $\varepsilon > 0$, we have that $\mathbb{P}(|X_n| > \varepsilon) = 1/n \xrightarrow{n\to\infty} 0$. Thus, $X_n \to_{\mathbb{P}} 0$.

Solution 12.4 (Solution to Exercise 12.4). *Let $p = 2$. Suppose that for any $i = 1, \dots, k$, $X_i^n \to_{L^2} X_i$. We have that*

$$\mathbb{E}[\|X_n - X\|^2] = \sum_{i=1}^{k} \mathbb{E}[|X_i^n - X_i|^2] \xrightarrow{n\to\infty} 0.$$

Therefore $X_n \to_{L^2} X$. For the other direction, if $X_n \to_{L^2} X$, then by the previous display, $\mathbb{E}[|X_i^n - X_i|^2] \le \mathbb{E}[\|X_n - X\|^2]$, $i = 1, \dots, k$, i.e., $X_i^n \to_{L^2} X_i$ for any $i = 1, \dots, k$. If $p = 1$ we readily check that $X_n \to_{L^1} X$ implies that $X_i^n \to_{L^1} X_i$ for any $i = 1, \dots, k$ (keep in mind that $\mathbb{E}[\|X_n - X\|] \ge \mathbb{E}[|X_i^n - X_i|]$, $i = 1, \dots, k$). We also notice that

$$\mathbb{E}[\|X_n - X\|] \le k^{1/2}\mathbb{E}\big[\big(\max_{i=1,\dots,k} |X_i^n - X_i|^2\big)^{1/2}\big] = k^{1/2}\mathbb{E}\big[\max_{i=1,\dots,k} |X_i^n - X_i|\big].$$

Hence, if $X_i^n \to_{L^1} X_i$ for any $i = 1, \dots, k$, $X_n \to_{L^1} X$ as well.

Solution 12.5 (Solution to Exercise 12.5). *Since $X_n \to_{L^1} X$ it follows that $X_n \to_{\mathbb{P}} X$ (cf. Proposition 12.2). By item (iv) of Proposition 12.7, $g(X_n) \to_{\mathbb{P}} g(X)$. Also, since g is bounded, $|g(X_n(\omega))| \le C$, $\omega \in \Omega$, where $C \in \mathbb{R}$. Therefore, by Proposition 12.6, $g(X_n) \to_{L^1} g(X)$.*

12.4 Additional exercises

Exercise 12.6. *Verify items (i) and (ii) of Proposition 12.7.*

Exercise 12.7. *Let $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ be two sequences of random variables. Verify the following:*

(a) If $X_n \to_{a.s.} X$ and $Y_n \to_{a.s.} Y$, then $X_n + Y_n \to_{a.s.} X + Y$ and $X_nY_n \to_{a.s.} XY$;

(b) If $X_n \to_{\mathbb{P}} X$ and $Y_n \to_{\mathbb{P}} Y$, then $X_n + Y_n \to_{\mathbb{P}} X + Y$ and $X_nY_n \to_{\mathbb{P}} XY$.

Exercise 12.8. *Let $\{A_n : n \in \mathbb{N}\}$ be a collection of independent events (cf. Definition 11.4) s.t. $\mathbb{P}(A_n) = \mathbb{P}(A_1)$ for any $n \in \mathbb{N}$. Show that*

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{A_i} \to_{a.s.} \mathbb{P}(A_1)$$

Exercise 12.9. *Let $p \in [0, 1]$. Show that for any continuous function $f : [0, 1] \to \mathbb{R}$,*

$$\lim_{n\to\infty} \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} f(k/n) = f(p).$$

Hint: *You might want to apply the law of large numbers (cf. Proposition 12.9).*

Exercise 12.10. *Prove Proposition 12.10.*

A Results from analysis

A.1 On infima and suprema: arithmetic set operations

We derive further results on the infimum and supremum of subsets of the real numbers. With regard to arithmetics on sets, we make the following definition:

Definition A.1. *Let $\emptyset \neq A \subset \mathbb{R}$. Then,*

(1.) $-A = \{-x \colon x \in A\}$;

(2.) $cA = \{ca \colon a \in A\}$, $c \in \mathbb{R}$;

(3.) $A + B = \{a + b \colon a \in A, b \in B\}$.

Proposition A.1. *Let $\emptyset \neq A \subset \mathbb{R}$. be bounded from above.*

(i) If A is bounded from above, then $-A$ is bounded from below and $\inf -A = -\sup A$;

(ii) if A is bounded from below, then $-A$ is bounded from above and $\sup -A = -\inf A$.

Proof. We only verify item (i) (the argument for item (ii) is similar). We notice first that since A is bounded from above, we rely on Proposition 1.5 and know that $\sup A \in \mathbb{R}$. Hence, since $\sup A$ is an upper bound for A, it follows that for any $x \in A$, $x \leq \sup A$. In particular, for any $x \in A$. $-x \geq -\sup A$. Thus, $-\sup A$ is a lower bound of $-A$. By Proposition 1.5, the infimum of $-A$ exists as an element of the real numbers. We thus write $L = \inf -A$. Then, we show that $L = -\sup A$. Clearly, since L is the infimum of $-A$ and $-\sup A$ is a lower bound of $-A$ it follows that $L \geq -\sup A$ (L is the greatest lower bound). On the other hand, given any $x \in A$, $-x \geq L$, it follows that $x \leq -L$ and hence $-L$ is an upper bound for A. This leads to $\sup A \leq -L$ and hence $-\sup A \geq L$. □

Proposition A.2. *Let $\emptyset \neq A \subset \mathbb{R}$ and $c > 0$.*

(i) If $\sup(cA) < \infty$, then $\sup A < \infty$ and $\sup(cA) = c \sup A$;

(ii) if $\inf(cA) > -\infty$, then $\inf A > -\infty$ and $\inf(cA) = c \inf(A)$.

Proof. Given any $a \in A$, $ca \leq \sup(cA)$. Hence, since $c > 0$, $a \leq \sup(cA)/c$. Thus, since by assumption $\sup(cA) < \infty$ and $a \in A$ was arbitrary, it follows that $\sup A \leq \sup(cA)/c < \infty$. Notice that the latter inequality shows that $c \sup A \leq \sup(cA)$. Also, since $c > 0$, $ca \leq c \sup A$ for any $a \in A$, hence, $\sup(cA) \leq c \sup A$ and we conclude that $\sup(cA) = c \sup A$. This shows (i). The argument for (ii) is similar. □

Proposition A.3. *Let $A, B \subset \mathbb{R}$, A and B not empty.*

(i) If $\sup(A + B) < \infty$, then $\sup A < \infty$, $\sup B < \infty$ and $\sup(A + B) = \sup A + \sup B$;

(ii) If $\inf(A + B) > -\infty$, then $\inf A > -\infty$, $\inf B > -\infty$ and $\inf(A + B) = \inf A + \inf B$.

Proof. We show item (i). Notice that since $\sup(A + B) < \infty$, it follows that for any $a \in A$, $a + b \leq \sup(A + B)$ and hence $a \leq \sup(A + B) - b$. Therefore, since $a \in A$ was arbitrary, $\sup A \leq \sup(A + B) - b < \infty$ and we conclude that $\sup A < \infty$. A similar argument shows that $\sup B < \infty$. We readily see that $\sup(A + B) \leq \sup A + \sup B$. Now either $\sup(A + B) < \sup A + \sup B$ or $\sup(A + B) = \sup A + \sup B$. We show that if we assume that $\sup(A + B) < \sup A + \sup B$, then we arrive at a contradiction. Hence, suppose that $\sup(A + B) < \sup A + \sup B$. Let $\delta = \sup A + \sup B - \sup(A + B) > 0$. By Proposition 1.10, there exists $a \in A$ s.t. $a > \sup A - (\delta/2)$. Similarly, there exists $b \in B$ s.t. $b > \sup B - (\delta/2)$. Hence,

$$a + b > \sup A + \sup B - \delta = \sup(A + B),$$

which is a contradiction. Therefore, $\sup(A + B) = \sup A + \sup B$. The argument for (ii) is similar. □

A.2 On limit inferior and limit superior

We derive further properties of the limit inferior and limit superior.

Proposition A.4. *Let $(b_n)_{n\in\mathbb{N}}$ be a real valued sequence s.t.* $\lim_{n\to\infty} b_n = b$. *Then,*

(i) $\liminf_{n\to\infty}(b_n + a_n) = b + \liminf_{n\to\infty} a_n$ *and* $\liminf_{n\to\infty}(-a_n) = -\limsup_{n\to\infty} a_n$*;*

(ii) $\limsup_{n\to\infty}(b_n + a_n) = b + \limsup_{n\to\infty} a_n$ *and* $\limsup_{n\to\infty}(-a_n) = -\liminf_{n\to\infty} a_n$*.*

Proof. We only show (i), since the arguments for (ii) are similar. Assume first that $(a_n)_{n\in\mathbb{N}}$ is bounded from below. We have that for any $n \in \mathbb{N}$ and any given $j \geq n$, $\inf_{k\geq n} b_k + \inf_{k\geq n} a_k \leq b_j + a_j$. Therefore, $\inf_{k\geq n} b_k + \inf_{k\geq n} a_k \leq \inf_{k\geq n}(b_k + a_k)$. Hence,

$$b + \liminf_{n\to\infty} a_n = \lim_{n\to\infty}\big(\inf_{k\geq n} b_k + \inf_{k\geq n} a_k\big) \leq \lim_{n\to\infty} \inf_{k\geq n}(b_k + a_k) = \liminf_{n\to\infty}(b_n + a_n).$$

If $(a_n)_{n\in\mathbb{N}}$ is not bounded from below, then by definition, $-\infty = b + \liminf_{n\to\infty} a_n = \liminf_{n\to\infty}(b_n + a_n)$ (where we used the convention that $-\infty + x = -\infty$ for $x \in \mathbb{R}$). For the other inequality we notice that by the previous inequality,

$$\liminf_{n\to\infty} a_n = \liminf_{n\to\infty}(b_n + a_n - b_n) \geq \liminf_{n\to\infty}(b_n + a_n) + \liminf_{n\to\infty}(-b_n) = \liminf_{n\to\infty}(b_n + a_n) - b.$$

Thus also $\liminf_{n\to\infty}(b_n + a_n) \leq b + \liminf_{n\to\infty} a_n$. Let us show that $\liminf_{n\to\infty}(-a_n) = -\limsup_{n\to\infty} a_n$. We assume first that $(a_n)_{n\in\mathbb{N}}$ is bounded from above. By Proposition A.1, we know that $\{-a_n : n \in \mathbb{N}\}$ is bounded from below and $\inf_{k\geq n}(-a_k) = -\sup_{k\geq n} a_k$. Hence,

$$\liminf_{n\to\infty}(-a_n) = \lim_{n\to\infty} \inf_{k\geq n}(-a_k) = -\lim_{n\to\infty} \sup_{k\geq n} a_k = -\limsup_{n\to\infty} a_n.$$

If $(a_n)_{n\in\mathbb{N}}$ is not bounded from above, then $\{-a_n : n \in \mathbb{N}\}$ is not bounded from below and by definition of the limit inferior and limit superior $\liminf_{n\to\infty}(-a_n) = -\limsup_{n\to\infty} a_n = -\infty$. □

Proposition A.5. *Let $f_n : A \to \overline{\mathbb{R}}$, $n \in \mathbb{N}$, be a sequence of functions. Define the sequences of functions $g_n = \inf_{k\geq n} f_k$ and $h_n = \sup_{k\geq n} f_k$, $n \in \mathbb{N}$. Then,*

(i) for any $x \in A$, $(g_n(x))_{n\in\mathbb{N}}$ is increasing and $g_n(x) \uparrow \liminf_{n\to\infty} f_n(x)$;

(ii) for any $x \in A$, $(h_n(x))_{n\in\mathbb{N}}$ is decreasing and $h_n(x) \downarrow \limsup_{n\to\infty} f_n(x)$.

Proof.

(i) Let $x \in A$. Suppose that $(g_n(x))$ is bounded from below. Then, using Exercise 3.7 and Proposition 3.18, $(g_n(x))$ is increasing with limit $\liminf_{n\to\infty} f_n(x)$. If $(g_n(x))$ is not bounded from below, still, for any $n \in \mathbb{N}$, $g_{n+1}(x) = \inf_{k\geq n+1} f_k(x) \geq \inf_{k\geq n} f_k(x) = g_n(x)$, i.e., $(g_n(x))$ is increasing. Suppose that there exists $N \in \mathbb{N}$ s.t. $(g_N(x)) > -\infty$. Then, we rely on the sequence $g_n^*(x) = g_{n-1+N}(x)$, $n \in \mathbb{N}$, and repeat the arguments given in the proof of Proposition 3.18 to show that $\lim_{n\to\infty} g_n^*(x) = \sup_{n\in\mathbb{N}} g_n^*(x)$. Therefore, $\lim_{n\to\infty} g_n(x) = \sup_{n\in\mathbb{N}} g_n^*(x)$. Since $\sup_{n\in\mathbb{N}} g_n(x) \leq \sup_{n\in\mathbb{N}} g_n^*(x)$, this shows that $\lim_{n\to\infty} g_n(x) \geq \sup_{n\in\mathbb{N}} g_n(x)$. The other inequality, $\lim_{n\to\infty} g_n(x) \leq \sup_{n\in\mathbb{N}} g_n(x)$, is obtained from the arguments already given in the proof of Proposition 3.18. Finally, if for any $n \in \mathbb{N}$, $g_n(x) = -\infty$, $\lim_{n\to\infty} g_n(x) = -\infty = \liminf_{n\to\infty} f_n(x)$ (cf. Definition 3.10).

(ii) We may repeat a similar argument as in (i).

□

A.3 On convergent sequences

Proposition A.6. *Let $(a_n)_{n\in\mathbb{N}}$ be a real-valued sequence s.t. for any $n \in \mathbb{N}$, $|a_n| > n$. Then, $(a_n)_{n\in\mathbb{N}}$ has no convergent subsequence.*

Proof. This follows from the fact that any subsequence $(a_{s(n)})_{n\in\mathbb{N}}$ of $(a_n)_{n\in\mathbb{N}}$ is not bounded. Notice that if $(a_{s(n)})_{n\in\mathbb{N}}$ is a subsequence of $(a_n)_{n\in\mathbb{N}}$, then for any $n \in \mathbb{N}$, $|a_{s(n)}| > s(n)$. Since $s(n) < s(n+1)$ for any $n \in \mathbb{N}$ the set $\{s(n): n \in \mathbb{N}\}$ is countably infinite and hence not bounded. Thus, $\sup\{s(n): n \in \mathbb{N}\} = \infty$. Therefore, for any $M \in \mathbb{R}$, there exists $n \in \mathbb{N}$ s.t. $s(n) > M$. Since $|a_{s(n)}| > s(n)$, it follows in particular that $|a_{s(n)}| > M$. Therefore, $(a_{s(n)})_{n\in\mathbb{N}}$ is not bounded (cf. Definition 3.3). $\square$

We prove the subsequence criterion for convergent sequences stated at the end of Section 3.2.

Proof of Proposition 3.24. First, we prove that (A) implies that $(a_n)_{n\in\mathbb{N}}$ is bounded. To see it, suppose by contradiction that $(a_n)_{n\in\mathbb{N}}$ is not bounded. Then, find $n_1 \in \mathbb{N}$ s.t. $1 < |a_{n_1}| \leq 1 + N_1$ for some $N_1 \in \mathbb{N}\setminus\{1\}$. Since $(a_n)_{n\in\mathbb{N}}$ is not bounded, there exists $n_2 \in \mathbb{N}$ s.t. $2 < 1 + N_1 < |a_{n_2}| < 1 + N_1 + N_2$ for some $N_2 \in \mathbb{N}\setminus\{1\}$. We continue like this and obtain a subsequence $(b_k)_{k\in\mathbb{N}} = (a_{s(k)})_{k\in\mathbb{N}}$, $s(k) = n_k$, $k \in \mathbb{N}$ where for any $k \in \mathbb{N}$, $|b_k| > k$. By Proposition A.6, $(b_k)_{k\in\mathbb{N}}$ can not have any convergent subsequence. This contradicts (A). Hence, (A) implies that $(a_n)_{n\in\mathbb{N}}$ is bounded. We verify that (A) implies that $\lim_{n\to\infty} a_n = a$. Again, we consider an argument by contradiction and suppose that (A) is true but $\lim_{n\to\infty} a_n \neq a$. If $\lim_{n\to\infty} a_n \neq a$, (A) implies that $(a_n)_{n\in\mathbb{N}}$ can not be convergent, since if $(a_n)_{n\in\mathbb{N}}$ was convergent with limit L, then any subsequence of $(a_n)_{n\in\mathbb{N}}$ must converge to the same limit L and then (A) implies that $L = \lim_{n\to\infty} a_n = a$. In summary, under the assumption that (A) holds and $\lim_{n\to\infty} a_n \neq a$, $(a_n)_{n\in\mathbb{N}}$ is bounded but does not converge. Using Propositions 3.20 and 3.21, this implies that $m = \liminf_{n\to\infty} a_n < \limsup_{n\to\infty} a_n = M$, where $m = \lim_{n\to\infty} m_n$ and $M = \lim_{n\to\infty} M_n$ with $m_n = \inf\{a_k: k \geq n\}$ and $M_n = \sup\{a_k: k \geq n\}$, respectively (cf. Propositions 3.18 and 3.19). Thus, if we write $a_{s(n)} = m_n$, $n \in \mathbb{N}$, and $a_{\tilde{s}(n)} = M_n$, $n \in \mathbb{N}$, we identify two subsequences $(a_{s(n)})_{n\in\mathbb{N}}$ and $(a_{\tilde{s}(n)})_{n\in\mathbb{N}}$ of $(a_n)_{n\in\mathbb{N}}$ with limit m and M, respectively. But then, for any subsequence $(a_{t(s(n))})_{n\in\mathbb{N}}$ and $(a_{\tilde{t}(\tilde{s}(n))})_{n\in\mathbb{N}}$ of $(a_{s(n)})_{n\in\mathbb{N}}$ and $(a_{\tilde{s}(n)})_{n\in\mathbb{N}}$, respectively, it follows that $\lim_{n\to\infty} a_{t(s(n))} = m$ and $\lim_{n\to\infty} a_{\tilde{t}(\tilde{s}(n))} = M$. Since $m \neq M$, (A) is violated and hence we arrive at a contradiction. We conclude that $\lim_{n\to\infty} a_n = a$. $\square$

A.4 Continuity

For this section, we use the notation $\|\cdot\|_m$ and $\|\cdot\|_k$ for the Euclidean distance on $\mathbb{R}^m$ and $\mathbb{R}^k$, respectively (cf. Definition 2.14).

Definition A.2. *Let $f\colon E \to \mathbb{R}^k$, $E \subset \mathbb{R}^m$. Then, f is continuous at a point $x \in E$, if for any $\varepsilon > 0$, there exists $\delta > 0$ s.t. for any $y \in E$, if $\|x - y\|_m < \delta$, then $\|f(x) - f(y)\|_k < \varepsilon$.*

Definition A.3. *A function $f\colon E \to \mathbb{R}^k$, $E \subset \mathbb{R}^m$ is referred to as continuous (or continuous on E) if it is continuous at any point $x \in E$.*

Example A.1. *Let $m = k = 1$ and consider $f(x) = x$, $x \in \mathbb{R}$. Then, clearly $f\colon \mathbb{R} \to \mathbb{R}$ is continuous. For any $\varepsilon > 0$, let $\delta = \varepsilon$, then $|x - y| < \delta$ implies that $|f(x) - f(y)| < \varepsilon$.*

Example A.2. *Let $c \in \mathbb{R}$ and $f(x) = c$ for any $x \in \mathbb{R}$. Then, clearly $f\colon \mathbb{R} \to \mathbb{R}$ is continuous since for any $\varepsilon > 0$, $|f(x) - f(y)| = 0 < \varepsilon$.*

The following is known as the intermediate value theorem, it supports our natural understanding of a continuous function defined on a closed interval.

Proposition A.7. *Let $f\colon [a,b] \to \mathbb{R}$ be continuous s.t. $f(a) \neq f(b)$. Suppose that γ is a value s.t.*

$$\gamma \in \big[\min\{f(a), f(b)\}, \max\{f(a), f(b)\}\big],$$

i.e., γ is between $f(a)$ and $f(b)$, then there exists $s \in [a,b]$ s.t. $f(s) = \gamma$. That is, any value γ between $f(a)$ and $f(b)$ is attained by f.

Proof. Let $\gamma \in \big[\min\{f(a), f(b)\}, \max\{f(a), f(b)\}\big]$. Assume that that $f(a) < f(b)$. Suppose that $\gamma \in (f(a), f(b))$, since if $\gamma = f(a)$ or $\gamma = f(b)$, the result follows. Define

$$A = \{x \in [a,b] \colon f(x) < \gamma\}.$$

A is not empty, since $f(a) < \gamma$. Clearly, since A is bounded, $s = \sup A < \infty$. Further, $s \in [a,b]$. There are three cases, $f(s) < \gamma$, $f(s) > \gamma$ or $f(s) = \gamma$. We show that only $f(s) = \gamma$ is possible. Suppose first that that $f(s) < \gamma$, i.e., $s \in A$. In particular, $s < b$ since $s = b$ is not possible ($s = b$ implies with $s \in A$ that $\gamma > f(b)$). Then, since f is continuous on $[a,b]$ and $s \in [a,b]$, f is continuous at s, i.e., there exists $\delta > 0$ s.t. for any $x \in [a,b]$, if $|x - s| < \delta$, then $|f(x) - f(s)| < \gamma - f(s)$ (recall that $\gamma - f(s) > 0$). In particular, $f(x) - f(s) \leq |f(x) - f(s)| < \gamma - f(s)$ and hence, if $x \in [a,b]$ s.t. $|x - s| < \delta$, we obtain $f(x) < \gamma$. Since $s < b$, we let $\varepsilon > 0$ s.t. $s + \varepsilon \leq b$. Set $x = s + \min\{(\delta/2), \varepsilon\}$. Then, $x \in [a,b]$ and $|x - s| < \delta$ and hence $f(x) < \gamma$. This is a contradiction with $s = \sup A$. Hence, the case $f(s) < \gamma$ is not possible. Assume now $f(s) > \gamma$. Using the same reasoning as before, there exists $\delta > 0$ s.t. for any $x \in [a,b]$, if $|x - s| < \delta$ then $f(x) > \gamma$. Then, since $s < \infty$, by Proposition 1.10, there exists $x \in A$ s.t. $x > s - \delta$. Since $s \geq x$, that particular x is s.t. $f(x) > \gamma$ and $f(x) < \gamma$, which is again a contradiction. Hence, we obtain $f(s) = \gamma$ and the proposition is proven. □

We prove the sequence criterion for continuous functions stated in Section 3.3.

Proof of Proposition 3.26. We first show that item (i) implies item (ii). Hence, let f be continuous at x. Take any $(x_n)_{n\in\mathbb{N}} \subset E$ which is s.t. $\lim_{n\to\infty} x_n = x$. Since f is continuous at x, for any $\varepsilon > 0$, there exists $\delta > 0$ s.t. $\|f(x) - f(y)\|_k < \varepsilon$ if $\|x - y\|_m < \delta$. Since $\lim_{n\to\infty} x_n = x$, there exists $N \in \mathbb{N}$, s.t. $\|x - x_n\|_m < \delta$ for any $n \geq N$. Therefore, $\|f(x) - f(x_n)\|_k < \varepsilon$ for any $n \geq N$. Since $\varepsilon > 0$ was arbitrary, the result follows. For the other direction, assume that item (ii) is true. Assume by contradiction that f is not continuous at x. Hence, there exists $\varepsilon > 0$ s.t. for any $\delta > 0$, there exists $y \in E$ with $\|x - y\|_m < \delta$ but $\|f(x) - f(y)\|_k \geq \varepsilon$. Let $\delta = 1/n$, $n \in \mathbb{N}$. For any $n \in \mathbb{N}$, let y_n be s.t. $\|x - y_n\|_m < 1/n$ and $\|f(x) - f(y_n)\|_k \geq \varepsilon$. Then $\lim_{n\to\infty} y_n = y$ but $\lim_{n\to\infty} f(y_n) \neq f(x)$. This contradicts item (ii), and hence the proof is complete. □

Proposition A.8. *Let $f = (f_1, \ldots, f_k)\colon E \to \mathbb{R}^k$, $E \subset \mathbb{R}^m$. Then, f is continuous at $x \in E$ if and only if f_i is continuous at x for any $i = 1, \ldots, k$.*

Proof. This follows from Propositions 3.26 and 3.25. □

Proof of Proposition 2.12. Let $f\colon \mathbb{R}^k \to \mathbb{R}$, $f(x) = \sum_{i=1}^k x_k$, $x = (x_1, \ldots, x_k)$. Assume that $(a_n)_{n\in\mathbb{N}} \in (\mathbb{R}^k)^{\mathbb{N}}$, $a_n = (a_1^n, \ldots, a_k^n)$, s.t. $a_n \xrightarrow{n\to\infty} x$ where $x \in \mathbb{R}^k$. Then, by Proposition 3.25, $\lim_{n\to\infty} a_i^n = x_i$ for any $i = 1, \ldots, k$. Hence, using Proposition 3.4,

$$f(a_n) = \sum_{i=1}^k a_i^n \xrightarrow{n\to\infty} \sum_{i=1}^k x_k = f(x).$$

Therefore, by Proposition 3.26, f is continuous at x. Since x was arbitrary, $f\colon \mathbb{R}^k \to \mathbb{R}$ is continuous. A similar argument shows that g und h are continuous. □

Proposition A.9. *Let $f\colon A \to B$, $A \subset \mathbb{R}^m$ and $B \subset \mathbb{R}^k$. Further, let $g\colon B \to C$, $C \subset \mathbb{R}^l$. If f is continuous at $x \in A$ and g is continuous at $f(x)$, then $g(f)\colon A \to C$ is continuous at x.*

Proof. Let $(x_n)_{n\in\mathbb{N}} \in A^{\mathbb{N}}$ s.t. $\lim_{n\to\infty} x_n = x$. Since f is continuous at x, it follows that $\lim_{n\to\infty} f(x_n) = f(x)$. Then, since g is continuous at $f(x)$, $\lim_{n\to\infty} g(f)(x_n) = g(f)(x)$. $\square$

Example A.3. *Let $f_i\colon E \to \mathbb{R}$, $E \subset \mathbb{R}^m$, $i = 1, \dots, k$, be continuous. Then, $f = \sum_{i=1}^k f_i\colon E \to \mathbb{R}$ and $g = \prod_{i=1}^k f_i\colon E \to \mathbb{R}$ are continuous as well. To see it, let $x \in E$. We first use Proposition A.8 and conclude that $x \mapsto \psi(x) = (f_1(x), \dots, f_k(x))$ is continuous at x. Further, by Proposition 2.12, given any $y = (y_1, \dots, y_k) \in \mathbb{R}^k$, $y \mapsto s(y) = \sum_{i=1}^k y_i$ and $y \mapsto p(y) = \prod_{i=1}^k y_i$ are continuous at y. Therefore, using Proposition A.9, it follows that $f(x) = s(\psi)(x)$ and $g(x) = p(\psi)(x)$ are continuous at x. Since $x \in E$ was arbitrary, the result follows.*

Proposition A.10. *Let $f\colon \mathbb{R}^m \to \mathbb{R}^k$. The following are equivalent:*

(i) f is continuous;

(ii) for any open set $U \subset \mathbb{R}^k$, $f^{-1}(U)$ is open in $\mathbb{R}^m$.

Proof. To avoid confusion, we write $B_r^m(a)$ and $B_r^k(b)$ for an open ball with center $a \in \mathbb{R}^m$ and $b \in \mathbb{R}^k$, respectively. We first show that (i) implies (ii). Therefore, let $U \subset \mathbb{R}^k$ be an open set. Let $x \in f^{-1}(U)$, i.e., $f(x) \in U$. Since U is open, there exists $\varepsilon > 0$, s.t. $B_\varepsilon^k(f(x)) \subset U$. Since f is continuous at x, there exists $\delta > 0$ s.t. if $y \in B_\delta^m(x)$, then $f(y) \in B_\varepsilon^k(f(x))$. Hence, $f(B_\delta^m(x)) \subset B_\varepsilon^k(f(x))$. Hence, $B_\delta^m(x) \subset f^{-1}(B_\varepsilon^k(f(x))) \subset f^{-1}(U)$. Therefore, $f^{-1}(U)$ is open in $\mathbb{R}^m$. We show that (ii) implies (i). Let $x \in \mathbb{R}^m$, $\varepsilon > 0$ and consider the open set $B_\varepsilon^k(f(x))$. Then, by assumption, $f^{-1}(B_\varepsilon^k(f(x)))$ is open in $\mathbb{R}^m$. Clearly, $x \in f^{-1}(B_\varepsilon^k(f(x)))$. Hence, there exists $\delta > 0$, s.t. $B_\delta^m(x) \subset f^{-1}(B_\varepsilon^k(f(x)))$. That is, $f(B_\delta^m(x)) \subset B_\varepsilon^k(f(x))$. This implies that f is continuous at x. Since $x \in \mathbb{R}^m$ was arbitrary, the result follows. $\square$

In accordance with Definition 2.16, we can define open sets in E, where $E \subset \mathbb{R}^m$.

Definition A.4. *Let $E \subset \mathbb{R}^m$, $E \neq \emptyset$. The Euclidean distance $\|\cdot\|_m$ restricted to E is denoted with $\|\cdot\|_E$. Given $x \in E$, we write $B_r^E(x) = \{y \in E\colon \|y - x\|_E < r\}$ for an open ball in E of radius $r > 0$ with center x. A set $V \subset E$ is open if for any $x \in V$, there exists $\varepsilon > 0$ s.t. $B_\varepsilon^E(x) \subset V$.*

The latter definition leads to the following result.

Proposition A.11. *Let $E \subset \mathbb{R}^m$, $E \neq \emptyset$. A function $f\colon E \to \mathbb{R}^k$ is continuous if and only if for any $U \subset \mathbb{R}^k$ open, $f^{-1}(U)$ is an open set in E.*

Proof. We repeat the arguments given in Proposition A.10 (replace open balls in $\mathbb{R}^m$ with open balls in E) and the result follows. $\square$

We prove Propositions 2.11 and 3.13.

Proof of Proposition 2.11. Using Proposition A.11, we show that a set $V \subset E$ is open if and only if there exists $G \subset \mathbb{R}^m$ which is open in $\mathbb{R}^m$ and s.t. $V = G \cap E$. Let $V \subset E$. Suppose that there exists $G \subset \mathbb{R}^m$, open in $\mathbb{R}^m$, s.t. $V = G \cap E$. Let $x \in V$, then $x \in G \cap E$, and since G is open and $x \in G$, there exists $B_\varepsilon^m(x) \subset G$. Thus, $B_\varepsilon^m(x) \cap E = B_\varepsilon^E(x) \subset G \cap E = V$. This shows that V is open in E. For the other direction, let $V \subset E$ be open in E. Then, given any $x \in V$, there exists $B_{\varepsilon_x}^E(x)$ s.t. $B_{\varepsilon_x}^E(x) \subset V$. We know that for any $x \in V$, $B_{\varepsilon_x}^m(x)$ is open in $\mathbb{R}^m$ (cf. Example 2.16). Define

$$G = \bigcup_{x \in V} B_{\varepsilon_x}^m(x).$$

It is clear that G is open in $\mathbb{R}^m$, since given any $y \in G$, there exists $x \in V$, s.t. $y \in B^m_{\varepsilon_x}(x) \subset G$. Certainly $V \subset G$ and hence $V = V \cap E \subset G \cap E$. On the other hand, $G \cap E = \cup_{x \in V} B^E_{\varepsilon_x}(x) \subset V$, since for any $x \in V$, $B^E_{\varepsilon_x}(x) \subset V$. Therefore, $V = G \cap E$. $\square$

Proof of Proposition 3.13. We only show the existence of x_M since the arguments for the minimum are similar. Let $S = \sup_{x \in \mathbb{R}} f(x)$. There are two cases, either $S < \infty$ or $S = \infty$. If $S = \infty$, let $(x_n)_{n \in \mathbb{N}}$ be a sequence s.t. for any $n \in \mathbb{N}$ $x_n \in [a, b]$ and $f(x_n) \geq n$ (for example, given any $n \in \mathbb{N}$, set $x_n = \min\{x \in [a, b] \colon f(x) \geq n\}$). Since $x_n \in [a, b]$ for any $n \in \mathbb{N}$, it follows that $(x_n)_{n \in \mathbb{N}}$ is bounded and hence by Proposition 3.12, there exists a subsequence $(x_{s(n)})_{n \in \mathbb{N}}$ s.t. $\lim_{n \to \infty} x_{s(n)} = \xi \in [a, b]$. Notice that $\xi \in [a, b]$, since $a \leq x_n \leq b$ for any $n \in \mathbb{N}$. In particular, since f is continuous on $[a, b]$, it follows that $\lim_{n \to \infty} f(x_{s(n)}) = f(\xi)$. On the other hand, since for any $n \in \mathbb{N}$, $f(x_n) \geq n$, it follows that $\lim_{n \to \infty} f(x_{s(n)}) = \infty$. Which gives a contradiction. Hence it must be the case that $S < \infty$. In this case, we set $a_n = S - 1/n$, $n \in \mathbb{N}$, and by Proposition 1.10, there exists $(x_n)_{n \in \mathbb{N}}$ s.t. for any $n \in \mathbb{N}$ $x_n \in [a, b]$ and $f(x_n) > S_n$. Again, we apply Proposition 3.12, and find a subsequence $(x_{s(n)})_{n \in \mathbb{N}}$ s.t. $\lim_{n \to \infty} x_{s(n)} = x_M \in [a, b]$. Since f is continuous on $[a, b]$, $\lim_{n \to \infty} f(x_{s(n)}) = f(x_M)$. Then, we have that for any $n \in \mathbb{N}$,

$$f(x_{s(n)}) \geq S - \frac{1}{s(n)},$$

and hence, $\lim_{n \to \infty} f(x_{s(n)}) \geq S$. Clearly, $\lim_{n \to \infty} f(x_{s(n)}) \leq S$ as well and hence $S = f(x_M) = \max_{x \in [a,b]} f(x)$. To see that f is bounded (cf. Definition 2.20), we apply Proposition A.9 and note that $x \mapsto |f(x)|$ is continuous on $[a, b]$. Therefore, for any $x \in [a, b]$, $|f(x)| \leq |f(x_M)|$. $\square$

Example A.4. *We list some examples of continuous functions:*

- $x \mapsto \mathrm{e}^x$ *as a function from* $\mathbb{R}$ *to* $\mathbb{R}$*;*
- $x \mapsto \log(x)$ *as a function from* $(0, \infty)$ *to* $\mathbb{R}$*;*
- $x \mapsto \sin(x)$ *as a function from* $\mathbb{R}$ *to* $\mathbb{R}$*;*
- $x \mapsto \cos(x)$ *as a function from* $\mathbb{R}$ *to* $\mathbb{R}$*.*

A.5 Limit points

Definition A.5. *Let $E \subset \mathbb{R}^m$ be a nonempty set. A point $a \in \mathbb{R}^m$ is said to be a limit or accumulation point of E if there exists a vector-valued sequence $(x_n)_{n \in \mathbb{N}}$ which satisfies $x_n \in E \setminus \{a\}$ for any $n \in \mathbb{N}$ and $x_n \xrightarrow{n \to \infty} a$.*

For this section, unless stated otherwise, any sequence is a real-valued sequence according to Definition 3.1.

Definition A.6. *Let $E \subset \mathbb{R}$ be a nonempty set and $f \colon E \to \mathbb{R}$ be a function. If $a \in \mathbb{R}$ is a limit point of E, we write $\lim_{x \to a} f(x) = L$ if there exists $L \in \mathbb{R}$ s.t. for any sequence $(x_n)_{n \in \mathbb{N}}$ which satisfies $x_n \in E \setminus \{a\}$ for any $n \in \mathbb{N}$ and $x_n \xrightarrow{n \to \infty} a$, it follows that $f(x_n) \xrightarrow{n \to \infty} L$.*

Proposition A.12. *Let $E \subset \mathbb{R}$ be a nonempty set and $f \colon E \to \mathbb{R}$ be a function. Then, $\lim_{x \to a} f(x) = L$ if and only if for any $\varepsilon > 0$, there exists $\delta > 0$ s.t. for any $x \in E \setminus \{a\}$, if $|x - a| < \delta$, $|f(x) - L| < \varepsilon$.*

Proof. This is essentially the proof of the sequence criterion of continuous function where the limit $f(x)$ is replaced with L (cf. the proof of Proposition 3.26). $\square$

Definition A.7. *Let $E \subset \mathbb{R}$ be a nonempty set and $f\colon E \to \mathbb{R}$ be a function. If $a \in \mathbb{R}$ is a limit point of $E \cap (-\infty, a)$, we write $\lim_{x\uparrow a} f(x) = L_l$ if there exists $L_l \in \mathbb{R}$ s.t. for any sequence $(x_n)_{n\in\mathbb{N}}$ which is s.t.*

$$\{x_n\colon n \in \mathbb{N}\} \subset E \cap (-\infty, a) \text{ and } x_n \xrightarrow{n\to\infty} a,$$

it follows that $f(x_n) \xrightarrow{n\to\infty} L_l$. Similarly, if $a \in \mathbb{R}$ is a limit point of $E \cap (a, \infty)$, we use the notation $\lim_{x\downarrow a} f(x) = L_r$ if there exists $L_r \in \mathbb{R}$ s.t. for any sequence $(x_n)_{n\in\mathbb{N}}$ which is s.t.

$$\{x_n\colon n \in \mathbb{N}\} \subset E \cap (a, \infty) \text{ and } x_n \xrightarrow{n\to\infty} a,$$

it follows that $f(x_n) \xrightarrow{n\to\infty} L_r$. If they exists, L_l and L_r are referred to as the left-hand and right-hand limit of f as x approaches a.

If we adapt the proof of Proposition 3.26 accordingly, we readily verify the following result:

Proposition A.13. *Let $E \subset \mathbb{R}$ be a nonempty set and $f\colon E \to \mathbb{R}$ be a function. Then, $\lim_{x\uparrow a} f(x) = L_l$ if and only if for any $\varepsilon > 0$, there exists $\delta_l > 0$ s.t. for any $x \in E \setminus \{a\}$, $x \in (a-\delta_l, a)$ implies that $|f(x) - L_l| < \varepsilon$. Similarly, $\lim_{x\downarrow a} f(x) = L_r$ if and only if for any $\varepsilon > 0$, there exists $\delta_r > 0$ s.t. for any $x \in E\setminus\{a\}$, $x \in (a, a+\delta_l)$ implies that $|f(x) - L_r| < \varepsilon$.*

Proposition A.14. *Let $E \subset \mathbb{R}$ be a nonempty set and $f\colon E \to \mathbb{R}$ be a function. Then, $\lim_{x\to a} f(x) = L$ if and only if*

$$\lim_{x\uparrow a} f(x) = L = \lim_{x\downarrow a} f(x).$$

Proof. Let δ_l and δ_r be as in Proposition A.13 and apply Proposition A.12 with $\delta = \min\{\delta_l, \delta_r\}$. □

Upon the definition of continuity (cf. Definition A.2), Proposition A.14 shows the following:

Proposition A.15. *Let $E \subset \mathbb{R}$ be a nonempty set and $f\colon E \to \mathbb{R}$ be a function. Then, f is continuous at a point $a \in E$, if and only if $\lim_{x\uparrow a} f(x) = f(a) = \lim_{x\downarrow a} f(x)$.*

Proposition A.16. *Let $E \subset \mathbb{R}$ be a nonempty set and $f\colon E \to \mathbb{R}$ be a function. Then, $\lim_{x\uparrow a} f(x) = L_l$ if and only if*

$$\forall\, (x_n)_{n\in\mathbb{N}} \text{ s.t. } x_n \in E \setminus \{a\}\ \forall n \in \mathbb{N} \text{ and } x_n \uparrow a \text{ it follows that } f(x_n) \xrightarrow{n\to\infty} L_l. \tag{47}$$

Similarly, $\lim_{x\downarrow a} f(x) = L_r$ if and only if

$$\forall\, (x_n)_{n\in\mathbb{N}} \text{ s.t. } x_n \in E \setminus \{a\}\ \forall n \in \mathbb{N} \text{ and } x_n \downarrow a \text{ it follows that } f(x_n) \xrightarrow{n\to\infty} L_r. \tag{48}$$

Proof. We only show that (47) implies $\lim_{x\uparrow a} f(x) = L_l$ and vice versa. The argument for the right-hand limit is the same. By Definition A.7, if $\lim_{x\uparrow a} f(x) = L_l$, then, (47) holds. Thus, suppose that (47) holds and let $(x_n)_{n\in\mathbb{N}}$ be a sequence s.t.

$$\{x_n\colon n \in \mathbb{N}\} \subset E \cap (-\infty, a) \text{ and } x_n \xrightarrow{n\to\infty} a.$$

Let $(x_{s(n)})_{n\in\mathbb{N}}$ be any subsequence of $(x_n)_{n\in\mathbb{N}}$. Since $x_n \xrightarrow{n\to\infty} a$, it follows that $x_{s(n)} \xrightarrow{n\to\infty} a$. In particular, $(x_{s(n)})_{n\in\mathbb{N}}$ is bounded. Define $x_{t(s(n))} = \inf\{x_{s(k)}\colon k \geq n\}$, $n \in \mathbb{N}$. Then, $(x_{t(s(n))})_{n\in\mathbb{N}}$ is increasing and s.t. $x_{t(s(n))} \uparrow a$ (cf. Exercise 3.7). Thus, by (47), $\lim_{n\to\infty} f(x_{t(s(n))}) = L_l$. Therefore, we have shown that for any subsequence $(f(x_{s(n)}))_{n\in\mathbb{N}}$ of $(f(x_n))_{n\in\mathbb{N}}$ there exists a subsequence $(f(x_{t(s(n))}))_{n\in\mathbb{N}}$ s.t. $\lim_{n\to\infty} f(x_{t(s(n))}) = L_l$. By Proposition 3.24, $\lim_{n\to\infty} f(x_n) = L_l$ and hence (47) implies that $\lim_{x\uparrow a} f(x) = L_l$. □

Definition A.8. *Let $E \subset \mathbb{R}$ be a nonempty set and $f\colon E \to \mathbb{R}$ be a function. If E is not bounded from above, we write $\lim_{x\to\infty} f(x) = L$ if there exists $L \in \mathbb{R}$ s.t. for any sequence $(x_n)_{n\in\mathbb{N}}$ which is s.t. $\{x_n : n \in \mathbb{N}\} \subset E$ and $x_n \xrightarrow{n\to\infty} \infty$, it follows that $f(x_n) \xrightarrow{n\to\infty} L$. Similarly, if E is not bounded from below, we write $\lim_{x\to-\infty} f(x) = L$ if there exists $L \in \mathbb{R}$ s.t. for any sequence $(x_n)_{n\in\mathbb{N}}$ which is s.t. $\{x_n : n \in \mathbb{N}\} \subset E$ and $x_n \xrightarrow{n\to\infty} -\infty$, it follows that $f(x_n) \xrightarrow{n\to\infty} L$.*

Proposition A.17. *Let $E \subset \mathbb{R}$ be a nonempty set and $f\colon E \to \mathbb{R}$ be a function. If E is not bounded from above, then $\lim_{x\to\infty} f(x) = L$ if and only if for any $\varepsilon > 0$ there exists a real number $M > 0$ s.t. for any $x \in E$ with $x > M$, $|f(x) - L| < \varepsilon$. Similarly, if E is not bounded from below, $\lim_{x\to-\infty} f(x) = L$ if and only if for any $\varepsilon > 0$ there exists a real number $M > 0$ s.t. for any $x \in E$ with $x < -M$, $|f(x) - L| < \varepsilon$.*

Proof. Suppose that $\lim_{x\to\infty} f(x) = L$ and there exists $\varepsilon > 0$ s.t. for any $M > 0$ there exists $x \in E$ s.t. $x > M$ with $|f(x) - L| \geq \varepsilon$. Let $n_1 = 1$ and obtain $n_1 < x_1 \leq n_2$ which is s.t. $|f(x_1) - L| \geq \varepsilon$. Then, find $n_2 < x_2 \leq n_3$ which also satisfies $|f(x_2) - L| \geq \varepsilon$. If we continue like this, we obtain a sequence $(x_n)_{n\in\mathbb{N}}$ which is s.t. for any $n \in \mathbb{N}$, $|f(x_n) - L| \geq \varepsilon$. Since $x_n \xrightarrow{n\to\infty} \infty$ this contradicts $\lim_{x\to\infty} f(x) = L$. Therefore, $\lim_{x\to\infty} f(x) = L$ implies that for any $\varepsilon > 0$ there exists a real number $M > 0$ s.t. for any $x \in E$ s.t. $x > M$, $|f(x) - L| < \varepsilon$. For the other direction, suppose that for any $\varepsilon > 0$ there exists a real number $M > 0$ s.t. for any $x \in E$ with $x > M$, $|f(x) - L| < \varepsilon$. Let $\varepsilon > 0$ and consider $\{x_n : n \in \mathbb{N}\} \subset E$ s.t. $x_n \xrightarrow{n\to\infty} \infty$. Since $x_n \xrightarrow{n\to\infty} \infty$, there exists $N \in \mathbb{N}$, s.t. for any $n \geq N$, $x_n > M$. Thus, by assumption, $|f(x_n) - L| < \varepsilon$, i.e., $f(x_n) \xrightarrow{n\to\infty} L$. The argument for $\lim_{x\to-\infty} f(x) = L$ is similar and we consider the proposition as proven. ☐

Similar to Proposition A.16, we have the following result:

Proposition A.18. *Suppose that $f\colon E \to \mathbb{R}$ is a function where $E \subset \mathbb{R}$ is a nonempty set. Suppose that E is not bounded from above. Then, $\lim_{x\to\infty} f(x) = L$ if and only if*

$$\forall\, (x_n)_{n\in\mathbb{N}} \text{ s.t. } x_n \in E\ \forall n \in \mathbb{N} \text{ and } x_n \uparrow \infty \text{ it follows that } f(x_n) \xrightarrow{n\to\infty} L. \tag{49}$$

Similarly, if E is not bounded from below, $\lim_{x\to-\infty} f(x) = L$ if and only if

$$\forall\, (x_n)_{n\in\mathbb{N}} \text{ s.t. } x_n \in E\ \forall n \in \mathbb{N} \text{ and } x_n \downarrow -\infty \text{ it follows that } f(x_n) \xrightarrow{n\to\infty} L. \tag{50}$$

Proof. We only show that (50) implies that $\lim_{x\to-\infty} f(x) = L$ and vice versa (the remaining argument is similar). Suppose that E is not bounded from below. Clearly, if $\lim_{x\to-\infty} f(x) = L$, then by Definition A.8, (50) is satisfied. Hence, suppose that (50) is satisfied. Assume by contradiction that there exists $\varepsilon > 0$ s.t. for any $M > 0$ there exists $x \in E$ s.t. $x < -M$ with $|f(x) - L| \geq \varepsilon$. Let $n_1 = -1$ and obtain $n_2 \leq x_1 < n_1$ which is s.t. $|f(x_1) - L| \geq \varepsilon$. Then, find $n_3 \leq x_2 < n_2$ which also satisfies $|f(x_2) - L| \geq \varepsilon$. If we continue like this, we obtain a sequence $(x_n)_{n\in\mathbb{N}}$ which is s.t. $x_n \downarrow -\infty$ but $\lim_{n\to\infty} f(x_n) \neq L$. This contradicts (50). Hence, for any $\varepsilon > 0$ there exists a real number $M > 0$ s.t. for any $x \in E$ with $x < -M$, $|f(x) - L| < \varepsilon$. By Proposition A.17 this implies that $\lim_{x\to-\infty} f(x) = L$. ☐

Remark A.1. *Notice that if $E \subset \mathbb{R}^m$, $a \in \mathbb{R}^m$ is a limit point of E and $f\colon E \to \mathbb{R}^k$ is a function, $\lim_{x\to a} f(x) = L$, $L \in \mathbb{R}^k$, is defined as in Definition A.6 with $(x_n)_{n\in\mathbb{N}}$ vector-valued.*

A.6 Differentiability in one variable and the mean value theorem

Definition A.9. *Let $f\colon [a,b] \to \mathbb{R}$ be a function. The derivative of f at $x_0 \in (a,b)$ is defined as the limit*

$$f'(x_0) = \lim_{x\to x_0} \frac{f(x) - f(x_0)}{x - x_0} \quad \left(= \lim_{h\to 0} \frac{f(x_0+h) - f(x_0)}{h} \right).$$

f is referred to as differentiable if $f'(x_0) \in \mathbb{R}$ for any $x_0 \in (a,b)$, i.e., $f'\colon (a,b) \to \mathbb{R}$.

Remark A.2. *If $f\colon [a,b] \to \mathbb{R}$ is differentiable, then f is continuous on (a,b). In particular, if f is continuous in a and b, f is continuous on the entire $[a,b]$.*

Proposition A.19. *Let $f\colon [a,b] \to \mathbb{R}$ be a differentiable function. Suppose that $x_0 \in (a,b)$ is a maximum point (resp. minimum point) of f, i.e., $f(x) \le f(x_0)$ for any $x \in [a,b]$ (resp. $f(x) \ge f(x_0)$ for any $x \in [a,b]$). Then, $f'(x_0) = 0$.*

Proof. Suppose that $x_0 \in (a,b)$ is a maximum point of f. By Proposition A.14, we know that

$$\lim_{x \uparrow x_0} \frac{\overbrace{f(x) - f(x_0)}^{\le 0}}{\underbrace{x - x_0}_{\le 0}} = f'(x_0) = \lim_{x \downarrow x_0} \frac{\overbrace{f(x) - f(x_0)}^{\le 0}}{\underbrace{x - x_0}_{\ge 0}}.$$

Hence, $f'(x_0) = 0$. The same argument works if $x_0 \in (a,b)$ is a minimum point of f. $\square$

The next result is known as Rolle's theorem.

Proposition A.20. *Let $f\colon [a,b] \to \mathbb{R}$ be continuous and differentiable. Suppose that $f(a) = f(b)$. Then, there exists $x_0 \in (a,b)$ s.t. $f'(x_0) = 0$.*

Proof. Since $f(a) = f(b)$, there are three cases:

(i) for any $x \in (a,b)$ $f(x) = f(b)$;

(ii) there exists $x \in (a,b)$ s.t. $f(x) > f(b)$;

(iii) there exists $x \in (a,b)$ s.t. $f(x) < f(b)$.

In case (i), f is constant on (a,b), i.e., $f'(x) = 0$ for any $x \in (a,b)$. If case (ii) is true, then, since f is continuous on $[a,b]$, there exists $x_0 \in [a,b]$ s.t. $f(x_0) \ge f(y)$ for any $y \in [a,b]$ (cf. Proposition 3.13). We notice that $x_0 = a$ or $x_0 = b$ is not possible since this would imply that $f(x_0) = f(a) = f(b) < x$. Thus, the maximum point is s.t. $x_0 \in (a,b)$. By Proposition A.19, $f'(x_0) = 0$. Finally suppose that case (iii) holds. By Proposition 3.13 again, let x_0 be a minimum point of $[a,b]$. Again, $x_0 \in (a,b)$ and hence, $f'(x_0) = 0$. $\square$

The mean value theorem (or Lagrange theorem) reads as follows:

Proposition A.21. *Let $f\colon [a,b] \to \mathbb{R}$ be continuous and differentiable. Then, there exists $m \in (a,b)$ s.t.*

$$f(b) - f(a) = f'(m)(b-a).$$

Proof. Define

$$g(x) = f(x) - \frac{x-a}{b-a}(f(b) - f(a)).$$

We notice that g is s.t. $g(a) = g(b)$, g is continuous on $[a,b]$ and the derivative of g exists for any point $x_0 \in (a,b)$. Thus, by Proposition A.20, there exists $m \in (a,b)$ s.t. $g'(m) = 0$. That is,

$$0 = g'(m) = f'(m) - \frac{f(b) - f(a)}{b-a}.$$

$\square$

Example A.5. *Given any $x \in \mathbb{R}$, $1 + x \leq \mathrm{e}^x$. Clearly, the inequality becomes an equality if $x = 0$. Let $x > 0$ and consider the interval $[0, x]$. By the mean value theorem, there exists $m \in (0, x)$ s.t.*

$$\mathrm{e}^x - \mathrm{e}^0 = \mathrm{e}^x - 1 = \mathrm{e}^m (x - 0) = \mathrm{e}^m x.$$

Since $m > 0$, $\mathrm{e}^m > 1$. Therefore, the previous display shows that $\mathrm{e}^x - 1 > x \Leftrightarrow 1 + x < \mathrm{e}^x$. If $x < 0$, then, again by the mean value theorem, there exists $m \in (x, 0)$ s.t. $1 - \mathrm{e}^x = -\mathrm{e}^m x$. Then, since $m \in (x, 0)$ and $-x > 0$, we obtain with $0 < \mathrm{e}^m < 1$ that $1 - \mathrm{e}^x < -x \Leftrightarrow 1 + x < \mathrm{e}^x$.

A.7 Differentiability in several variables

Definition A.10. *A function $L \colon \mathbb{R}^m \to \mathbb{R}^k$ is linear if for any $v_1, v_2 \in \mathbb{R}^m$ and for any $\lambda_1, \lambda_2 \in \mathbb{R}$,*

$$L(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 L(v_1) + \lambda_2 L(v_2).$$

Suppose that $f \colon [a, b] \to \mathbb{R}$ is differentiable in $x_0 \in (a, b)$, i.e., the limit $f'(x_0)$ exists according to Definition A.9. Define the map: $L_{x_0}(h) = f'(x_0)h$, $h \in \mathbb{R}$. Then, $L_{x_0} \colon \mathbb{R} \to \mathbb{R}$ is linear and

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0) - L_{x_0}(h)}{h} = 0. \tag{51}$$

On the other hand if there exists a linear map $L_{x_0} \colon \mathbb{R} \to \mathbb{R}$ s.t. f satisfies (51) for some $x_0 \in (a, b)$, then,

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0) - L_{x_0}(h)}{h} + L_{x_0}(1) = L_{x_0}(1),$$

that is f is differentiable in x_0 according to Definition A.9. This shows that the following definition is equivalent to Definition A.9.

Definition A.11. *A function $f \colon [a, b] \to \mathbb{R}$ is differentiable in $x_0 \in (a, b)$ if there exists a linear map $L_{x_0} \colon \mathbb{R} \to \mathbb{R}$ s.t. (51) is satisfied.*

In general, differentiability is defined as follows (again $\|\cdot\|_m$ and $\|\cdot\|_k$ denote the Euclidean distance on $\mathbb{R}^m$ and $\mathbb{R}^k$, respectively):

Definition A.12. *Let $U \subset \mathbb{R}^m$ be an open set. A function $f \colon U \to \mathbb{R}^k$ is differentiable in $x_0 \in U$ if there exists a linear map $L_{x_0} \colon \mathbb{R}^m \to \mathbb{R}^k$ s.t.*

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0) - L_{x_0}(h)}{\|h\|_m} = 0. \tag{52}$$

The map f is referred to as differentiable on U if f is differentiable for any $x_0 \in U$. Further, the linear map L_{x_0} is said to be differential of f in x_0.

Remark A.3. *Let $U \subset \mathbb{R}^m$ be an open set and $f \colon U \to \mathbb{R}^k$ be differentiable in $x_0 \in U$. Then the linear map L_{x_0} in (52) is unique. That is, if L_{x_0} and $\tilde{L}_{x_0}$ are two linear maps that satisfy (52), then $L_{x_0} = \tilde{L}_{x_0}$. To see it, if L_{x_0} and $\tilde{L}_{x_0}$ satisfy (52), then for any $v \in \mathbb{R}^m$, $v \neq 0$, by linearity and (52),*

$$L_{x_0}(v/\|v\|_m) - \tilde{L}_{x_0}(v/\|v\|_m) = 0.$$

Thus, $L_{x_0} = \tilde{L}_{x_0}$.

Example A.6. *Let $f(x) = Ax + b$, $A \in \mathbb{R}^{k\times m}$, $b \in \mathbb{R}^k$. Let $L(x) = Ax$, $x \in \mathbb{R}^m$. Given any $x_0 \in \mathbb{R}^m$, we have that*

$$\lim_{h\to 0} \frac{f(x_0 + h) - f(x_0) - L(h)}{\|h\|_m} = 0.$$

Therefore, $f\colon \mathbb{R}^m \to \mathbb{R}^k$ is differentiable with differential L (cf. Remark A.3).

The following proposition is of central importance:

Proposition A.22. *Let $U \subset \mathbb{R}^m$ be an open set and $f = (f_1, \dots, f_k)\colon U \to \mathbb{R}^k$ be a map. Assume that $x_0 \in U$. Then, f is differentiable in x_0 if and only if any function $f_i\colon U \to \mathbb{R}$, $i = 1, \dots, k$, is differentiable in x_0. Further, if f is differentiable in x_0, the differential of f in x_0 is given by the map*

$$L_{x_0} = (L^1_{x_0}, \dots, L^k_{x_0}),$$

where $L^i_{x_0}\colon \mathbb{R}^m \to \mathbb{R}$ is the differential of f_i in x_0, $i = 1, \dots, k$.

Proof. Suppose first that f is differentiable in x_0, i.e., there exists a linear map $L_{x_0}\colon \mathbb{R}^m \to \mathbb{R}^k$ s.t. (52) is satisfied. Define the linear map

$$L^i_{x_0}(v) = \sum_{j=1}^m a_{ij} v_j, \quad i = 1, \dots, k,$$

where $A = (a_{ij})_{1\le i\le k, 1\le j\le m}$ is the matrix representation of L_{x_0}. Notice that this implies that

$$Av = (L^1_{x_0}(v), \dots, L^k_{x_0}(v)), \quad v \in \mathbb{R}^m.$$

Further, for any $i = 1, \dots, k$,

$$|f_i(x_0 + v) - f_i(x_0) - L^i_{x_0}(v)| \le \|f(x_0 + v) - f(x_0) - L_{x_0}(v)\|_k.$$

Thus, by (52),

$$\lim_{v\to 0} \frac{|f_i(x_0 + v) - f_i(x_0) - L^i_{x_0}(v)|}{\|v\|_m} = 0,$$

and in particular (recall that $t \mapsto |t| = \sqrt{t^2}$ is continuous),

$$\lim_{v\to 0} \frac{f_i(x_0 + v) - f_i(x_0) - L^i_{x_0}(v)}{\|v\|_m} = 0.$$

For the other direction, suppose that for any $i = 1, \dots, k$, f_i is differentiable in x_0 with differential $L^i_{x_0}$ in x_0. Then, we define the map

$$L_{x_0}(v) = (L^1_{x_0}(v), \dots, L^k_{x_0}(v)), \quad v \in \mathbb{R}^m,$$

and notice that $L_{x_0}\colon \mathbb{R}^m \to \mathbb{R}^k$ is linear since $L^i_{x_0}$, $i = 1, \dots, k$, are linear. Then, since for any $i = 1, \dots, k$,

$$\lim_{v\to 0} \frac{f_i(x_0 + v) - f_i(x_0) - L^i_{x_0}(v)}{\|v\|_m} = 0,$$

it follows with

$$\frac{\|f(x_0 + v) - f(x_0) - L_{x_0}(v)\|_k}{\|v\|_m} = \sqrt{\sum_{i=1}^k \left(\frac{f_i(x_0 + v) - f_i(x_0) - L^i_{x_0}(v)}{\|v\|_m}\right)^2},$$

that

$$\lim_{v\to 0} \frac{\|f(x_0+v)-f(x_0)-L_{x_0}(v)\|_k}{\|v\|_m} = 0.$$

In particular,

$$\lim_{v\to 0} \frac{f(x_0+v)-f(x_0)-L_{x_0}(v)}{\|v\|_m} = 0,$$

and hence f is differentiable in x_0 with differential L_{x_0}. □

In the following, we aim to find an explicit description of the matrix representation of the differential of a differentiable map. This matrix will be referred to as the Jacobian matrix.

Definition A.13. *Let $U \subset \mathbb{R}^m$ be open, $x_0 \in U$ and $f\colon U \to \mathbb{R}$ be a function. The directional derivative of f in direction $v \neq 0$ is defined as the limit (if it exists)*

$$\partial_v f(x_0) = \frac{\partial f}{\partial v}(x_0) = \lim_{t\to 0} \frac{f(x_0+tv)-f(x_0)}{t}.$$

Proposition A.23. *Let $U \subset \mathbb{R}^m$ be an open set and $f\colon U \to \mathbb{R}$ be differentiable in $x_0 \in U$. Then, for any $v \neq 0$, the directional derivative $\partial_v f(x_0)$ exists and is given by*

$$\frac{\partial f}{\partial v}(x_0) = L_{x_0}(v).$$

Proof. By assumption, f is differentiable in x_0, hence there exists a linear map $L_{x_0} : \mathbb{R}^m \to \mathbb{R}$ s.t. (52) is satisfied. Then, given any $v \in \mathbb{R}^m$, $v \neq 0$, we obtain

$$\begin{aligned}\frac{f(x_0+tv)-f(x_0)}{\|tv\|_m} &= \frac{f(x_0+tv)-f(x_0)-L_{x_0}(tv)}{\|tv\|_m} + \frac{L_{x_0}(tv)}{\|tv\|_m}\\ &= \frac{f(x_0+tv)-f(x_0)-L_{x_0}(tv)}{\|tv\|_m} + \frac{L_{x_0}(tv)}{\|tv\|_m}.\end{aligned}$$

By (52), it follows that

$$\lim_{t\downarrow 0} \frac{f(x_0+tv)-f(x_0)}{\|tv\|_m} = \frac{L_{x_0}(v)}{\|v\|_m}.$$

Thus,

$$\lim_{t\downarrow 0} \frac{f(x_0+tv)-f(x_0)}{t} = \|v\|_m \lim_{t\downarrow 0} \frac{f(x_0+tv)-f(x_0)}{\|tv\|_m} = L_{x_0}(v).$$

Similarly,

$$\lim_{t\uparrow 0} \frac{f(x_0+tv)-f(x_0)}{t} = L_{x_0}(v),$$

and the result follows. □

Definition A.14. *Let $U \subset \mathbb{R}^m$ be open, $x_0 \in U$ and $f\colon U \to \mathbb{R}$ be a function. Let $e_1, \ldots, e_m$, be the standard basis of $\mathbb{R}^m$, i.e.,*

$$e_j = (0, \ldots, 0, \underbrace{1}_{position\ j}, 0, \ldots, 0), \quad j = 1, \ldots, m.$$

The directional derivatives of f in direction e_j, $j = 1, \ldots, m$, are referred to as the partial derivatives of f in x_0. We use the notation

$$\frac{\partial f}{\partial e_j}(x_0) = \frac{\partial f}{\partial x_j}(x_0) = \partial_{x_j} f(x_0), \quad j = 1, \ldots, m.$$

Remark A.4. *Let $U \subset \mathbb{R}^m$ be an open set and $x_0 \in U$. Suppose that $f\colon U \to \mathbb{R}$ is differentiable in x_0. Then, by Proposition A.23, the partial derivatives $\partial_{x_j} f(x_0)$ exists for any $j = 1, \dots, m$. In particular, for any $j = 1, \dots, m$,*

$$\frac{\partial f}{\partial x_j}(x_0) = L_{x_0}(e_j).$$

Write $v = \sum_{j=1}^m v_j e_j$, where $v_j \in \mathbb{R}$ and $e_1, \dots, e_m$, is the standard basis of $\mathbb{R}^m$. By linearity, we obtain that

$$L_{x_0}(v) = \sum_{j=1}^m v_j L_{x_0}(e_j) = \sum_{j=1}^m v_j \frac{\partial f}{\partial x_j}(x_0). \tag{53}$$

The identity (53) is the representation of the differential of f in x_0 in terms of the partial derivatives.

Definition A.15. *Let $f\colon U \to \mathbb{R}^k$, $U \subset \mathbb{R}^m$ open. Suppose that the partial derivatives $\partial_{x_j} f_i(x_0)$ exist for any $j = 1, \dots, m$ and $i = 1, \dots, k$. Then, the matrix*

$$J_f(x_0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \frac{\partial f_1}{\partial x_2}(x_0) & \dots & \frac{\partial f_1}{\partial x_m}(x_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial x_1}(x_0) & \frac{\partial f_k}{\partial x_2}(x_0) & \dots & \frac{\partial f_k}{\partial x_m}(x_0) \end{pmatrix}$$

is referred to as the Jacobian matrix of f in x_0.

We have all the tools to easily verify the following:

Proposition A.24. *Let $f\colon U \to \mathbb{R}^k$, $U \subset \mathbb{R}^m$ open. Suppose that f is differentiable in $x_0 \in U$ with differential L_{x_0}. Then, the matrix representation of the linear map L_{x_0} is given by the Jacobian matrix $J_f(x_0)$.*

Proof. Suppose that $v \neq 0$ (for $v = 0$, the result is trivial). By Proposition A.22 and (53), we have that

$$L_{x_0}(v) = (L^1_{x_0}(v), \dots, L^k_{x_0}(v)) = \left(\sum_{j=1}^m v_j \frac{\partial f_1}{\partial x_j}(x_0), \dots, \sum_{j=1}^m v_j \frac{\partial f_k}{\partial x_j}(x_0) \right) = J_f(x_0)v.$$

□

Remark A.5. *One can show that $f\colon U \to \mathbb{R}^k$, $U \subset \mathbb{R}^m$ open, is differentiable in $x_0 \in U$ if for any $v \in U$, the partial derivatives $\partial_{x_j} f_i(v)$, $j = 1, \dots, m$, $i = 1, \dots, k$, exist and are continuous in x_0. This provides a useful criterion to verify the differentiability of f in x_0.*

B Measure and integration

B.1 Inclusion-exclusion principle

The following is known as the inclusion-exclusion formula:

Proposition B.1. *Let $(\Omega, \mathcal{F})$ be a measurable space and μ be a measure on $\mathcal{F}$. Assume that $A_1, \dots, A_n \in \mathcal{F}$ are s.t. $\mu(A_i) < \infty$ for any $i = 1, \dots, n$. We have that*

$$\mu\left(\bigcup_{i=1}^n A_i \right) = \sum_{k=1}^n \left((-1)^{k-1} \sum_{I \in A_k^n} \mu(A_I) \right), \tag{54}$$

where

$$A_k^n = \{I \subset \{1, \dots, n\} \colon \#I = k\}, \quad k = 1, \dots, n, \quad n \in \mathbb{N},$$

and for any $I \in A_k^n$, $\mu(A_I) = \mu(\cap_{i \in I} A_i)$.

Proof of Proposition B.1. If $n = 1$, (54) is trivial. If $n = 2$, (54) states that

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

which is true since by item (iv) of Proposition 5.1, $\mu(A_1 \cup A_2) + \mu(A_1 \cap A_2) = \mu(A_1) + \mu(A_2)$ and since $\mu(A_1)$ and $\mu(A_2)$ are assumed to be finite, we can subtract $\mu(A_1 \cap A_2)$ on both sides of the latter equation. By induction, assume that (54) holds for $n \in \mathbb{N}$. We have that

$$\begin{aligned} \mu\left(\bigcup_{i=1}^{n+1} A_i\right) &= \mu\left(\bigcup_{i=1}^{n} A_i\right) + \mu(A_{n+1}) - \mu\left(\bigcup_{i=1}^{n}(A_i \cap A_{n+1})\right) \\ &= \sum_{k=1}^{n}\left((-1)^{k-1} \sum_{I \in A_k^n} \mu(A_I)\right) + \mu(A_{n+1}) \\ &\quad - \left(\sum_{k=1}^{n}\left((-1)^{k-1} \sum_{I \in A_k^n} \mu(A_I \cap A_{n+1})\right)\right). \end{aligned} \tag{55}$$

Then, for any $k = 1, \dots, n$, $A_k^n \subset A_k^{n+1}$, and therefore,

$$\begin{aligned} &\sum_{k=1}^{n}\left((-1)^{k-1} \sum_{I \in A_k^n} \mu(A_I))\right) - \sum_{k=1}^{n+1}\left((-1)^{k-1} \sum_{I \in A_k^{n+1}} \mu(A_I))\right) \\ &= -\left(\sum_{k=1}^{n}\left((-1)^{k-1} \sum_{I \in (A_k^{n+1} \setminus A_k^n)} \mu(A_I))\right)\right). \end{aligned}$$

We notice that the latter sum is equal to

$$\begin{aligned} &-\mu(A_{n+1}) - \left(\sum_{k=1}^{n}\left((-1)^{k+1-1} \sum_{I \in (A_k^{n+1} \setminus A_k^n)} \mu(A_I))\right)\right) \\ &= -\mu(A_{n+1}) + \left(\sum_{k=1}^{n}\left((-1)^{k-1} \sum_{I \in (A_k^{n+1} \setminus A_k^n)} \mu(A_I))\right)\right). \end{aligned}$$

Further, for any $I \in A_k^{n+1} \setminus A_k^n$, $A_I = A_J \cap A_{n+1}$ for $J \in A_k^n$. Hence, the latter sum reads as

$$-\mu(A_{n+1}) + \left(\sum_{k=1}^{n}\left((-1)^{k-1} \sum_{I \in A_k^n} \mu(A_I \cap A_{n+1}))\right)\right).$$

If we add and subtract $\sum_{k=1}^{n+1}((-1)^{k-1} \sum_{I \in A_k^{n+1}} \mu(A_I)))$ from (55), we obtain that

$$\mu\left(\bigcup_{i=1}^{n+1} A_i\right) = \sum_{k=1}^{n+1}\left((-1)^{k-1} \sum_{I \in A_k^{n+1}} \mu(A_I))\right).$$

□

B.2 On measure extensions

Proof of Proposition 6.2. The finite case, i.e., there exists $N \in \mathbb{N}$ s.t.

$$\cup_{i \in I}(a_i, b_i] = \cup_{i=1}^{N}(a_i, b_i],$$

follows by induction. The base step of the induction is clear, if $(a,b] \subset (c,d]$, then $c \le a < b \le d$ and hence $b-a \le d-c$. For the induction step assume that (11) holds for $N-1$ intervals. Let $(a,b] \subset \cup_{i=1}^N (a_i,b_i]$. We want to show that $b-a \le \sum_{i=1}^N (b_i - a_i)$. Notice first that we can always assume that $b_1 \le b_2 \le \cdots \le b_N$. If not, we can just consider a relabeling and the union would remain unchanged. Assume first that $b \notin (a_N, b_N]$. Then, $b \le a_N$ since $b > b_N$ is not possible. To see this, assume by contradiction that $b > b_N$. Then, since $b_1 \le b_2 \le \cdots \le b_N$, $b \notin (a_i,b_i]$ for any $i = 1,\ldots,N$. Since $b \in (a,b] \subset \cup_{i=1}^N (a_i,b_i]$, this is not possible. Hence, $b \notin (a_N,b_N] \Rightarrow b \le a_N$. Hence, $(a,b] \subset \cup_{i=1}^{N-1}(a_i,b_i]$ since if $y \in (a,b]$, $y \le b \le a_N$ and hence $y \notin (a_N,b_N]$. By the induction hypothesis, the result follows. Thus, in the remaining we assume that $b \in (a_N,b_N]$. If $a_N \le a$, then $a_N \le a < b \le b_N$ and the result follows. Hence, assume that $a < a_N$. Then, $(a,a_N] \subset \cup_{i=1}^{N-1}(a_i,b_i]$. This is because $y \in (a,a_N]$ implies that $y \notin (a_N,b_N]$. Further $y \in (a,a_N]$ implies that $a < y \le a_N < b$ ($b \in (a_N,b_N]$) and hence $(a,a_N] \subset (a,b]$. Since $(a,b]$ is a subset of $\cup_{i=1}^N (a_i,b_i]$ it follows that $y \in (a_i,b_i]$ for some $i \ne N$, i.e., $(a,a_N] \subset \cup_{i=1}^{N-1}(a_i,b_i]$. By the induction hypothesis, $\sum_{i=1}^{N-1}(b_i - a_i) \ge a_N - a$. Therefore, $\sum_{i=1}^N (b_i - a_i) \ge a_N - a + b_N - a_N \ge a_N - a + b - a_N = b - a$. We use the Heine-Borel theorem for intervals (cf. Proposition 2.9) to prove the infinite case, i.e., $\cup_{i\in I}(a_i,b_i] = \cup_{i=1}^\infty (a_i,b_i]$. Suppose that $(a,b] \subset \cup_{i=1}^\infty (a_i,b_i]$. Let $\varepsilon > 0$ be s.t. $b - a > \varepsilon$. This is possible since $b \ne a$. Clearly, the family of intervals $(a_i, b_i + \varepsilon 2^{-i})$, $i \in \mathbb{N}$, are s.t.

$$[a+\varepsilon, b] \subset \bigcup_{i\in\mathbb{N}} (a_i, b_i + \varepsilon 2^{-i}).$$

By Proposition 2.9 it follows that there exists $i_1,\ldots,i_N$, s.t.

$$[a+\varepsilon, b] \subset \bigcup_{k=1}^N (a_{i_k}, b_{i_k} + \varepsilon 2^{-i_k}).$$

Hence, by the finite case,

$$\begin{aligned} b - a + \varepsilon &\le \sum_{k=1}^N (b_{i_k} - a_{i_k} + \varepsilon 2^{-i_k}) = \sum_{k=1}^N (b_{i_k} - a_{i_k}) + \varepsilon \sum_{k=1}^N 2^{-i_k} \\ &\le \sum_{i=1}^\infty (b_i - a_i) + \varepsilon \sum_{i=1}^\infty 2^{-i} = \sum_{i=1}^\infty (b_i - a_i) + \frac{\varepsilon}{2}\sum_{i=0}^\infty 2^{-i}. \end{aligned}$$

By Exercise 3.14, we obtain that $b - a + \varepsilon \le \sum_{i=1}^\infty (b_i - a_i) + \varepsilon$. This completes the argument. $\square$

Proof of Proposition 6.5. Given any $\xi \in C_{\mathcal{A}}(A)$, $A \in \mathcal{P}(\Omega)$, $v_\rho(\xi) \ge 0$. It follows that $\rho^*(A) \ge 0$ for any $A \in \mathcal{P}(\Omega)$. By assumption, $\emptyset \in \mathcal{A}$. Hence we can take $\xi = \{\emptyset\}$ and have that ξ is a covering of $\emptyset$ by sets from $\mathcal{A}$. Further, since $\rho(\emptyset) = 0$, it follows that $v_\rho(\xi) = 0$. This shows that $\rho^*(\emptyset) \le 0$. Since $\rho^*(A) \ge 0$ for any $A \in \mathcal{P}(\Omega)$, it follows that $\rho^*(\emptyset) = 0$. Let $A, B \in \mathcal{P}(\Omega)$ s.t. $A \subset B$. Since $A \subset B$, we have that

$$\{v_\rho(\xi) \colon \xi \in C_{\mathcal{A}}(B)\} \subset \{v_\rho(\xi) \colon \xi \in C_{\mathcal{A}}(A)\}.$$

This shows that $\inf_{\xi\in C_{\mathcal{A}}(A)} v_\rho(\xi) \le \inf_{\xi \in C_{\mathcal{A}}(B)} v_\rho(\xi)$ (cf. Proposition 1.9). To complete the argument, it remains to show that ρ^* is countable subadditive on $\mathcal{P}(\Omega)$. Let $\{A_n \colon n \in \mathbb{N}\} \subset \mathcal{P}(\Omega)$. We need to show that

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

Clearly, if there exists $n \in \mathbb{N}$ s.t. $\rho^*(A_n) = \infty$, $\rho^*(\cup_{n=1}^\infty A_n) \le \sum_{n=1}^\infty \rho^*(A_n)$. Thus suppose that for any $n \in \mathbb{N}$, $\rho^*(A_n) < \infty$. Given $n \in \mathbb{N}$, let $\xi_\varepsilon^n = \{U_{n_k} \colon k \in \mathbb{N}\} \in C_{\mathcal{A}}(A_n)$ be a covering of A_n by sets from $\mathcal{A}$ s.t.

$$v_\rho(\xi_\varepsilon^n) < \rho^*(A_n) + \frac{\varepsilon}{2^n}.$$

This is possible since $\inf_{\xi \in C_{\mathcal{A}}(A_n)} v_\rho(\xi)$ exists as an element of the real numbers (cf. Proposition 1.10). Since, $\{U_{n_k}: k \in \mathbb{N}, n \in \mathbb{N}\}$ is a covering of $\cup_{n\in\mathbb{N}} A_n$ by sets from $\mathcal{A}$, it follows that

$$\rho^*\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n=1}^{\infty} v_\rho(\xi_\varepsilon^n) < \sum_{n=1}^{\infty} \rho^*(A_n) + \varepsilon.$$

Since ε was arbitrary, the result follows (cf. Example 1.13). □

Proposition B.2. *Let μ^* be an outer measure on $\mathcal{P}(\Omega)$. Suppose that $\{A_i: i \in I\} \subset \mathcal{M}(\mu^*)$ is disjoint, where I is either finite or countably infinite. Then, for any $E \in \mathcal{P}(\Omega)$,*

$$\mu^*\left(\left(\bigcup_{i\in I} A_i\right) \cap E\right) = \sum_{i\in I} \mu^*(A_i \cap E).$$

Proof. Suppose first that I is finite, i.e., $\{A_i: i = 1, \ldots, n\} \subset \mathcal{M}(\mu^*)$, $n \in \mathbb{N}$. We prove by induction that for any $n \in \mathbb{N}$, $\mu^*((\cup_{i=1}^n A_i) \cap E) = \sum_{i=1}^n \mu^*(A_i \cap E)$. If $n = 1$, then the result follows immediately. If $n = 2$, and $A_1 \cup A_2 = \Omega$, we have that $A_2 = A_1^c$ and since $A_1 \in \mathcal{M}(\mu^*)$, it follows that

$$\begin{aligned}\mu^*(A_1 \cap E) + \mu^*(A_2 \cap E) &= \mu^*(A_1 \cap E) + \mu^*(A_1^c \cap E) \\ &= \mu^*(\Omega \cap E) = \mu^*((A_1 \cup A_2) \cap E).\end{aligned}$$

Suppose that $A_1 \cup A_2$ is a proper subset of Ω, i.e., $\Omega \setminus (A_1 \cup A_2) \neq \emptyset$. Since A_1 and A_2 are disjoint and $A_1^c \subset A_2$, we have that $((E \cap (A_1 \cup A_2)) \cap A_1) = E \cup A_1$ and $((E \cap (A_1 \cup A_2)) \cap A_1^c) = E \cup A_2$. Hence, since $A_1 \in \mathcal{M}(\mu^*)$, it follows that

$$\underbrace{\mu^*\big((E \cap (A_1 \cup A_2)) \cap A_1\big)}_{=\mu^*(E \cup A_1)} + \underbrace{\mu^*\big((E \cap (A_1 \cup A_2)) \cap A_1^c\big)}_{=\mu^*(E \cup A_2)} = \mu^*(E \cap (A_1 \cup A_2)). \tag{56}$$

Assume that $\mu^*((\cup_{i=1}^{n-1} A_i) \cap E) = \sum_{i=1}^{n-1} \mu^*(A_i \cap E)$. Using (56), we have that

$$\begin{aligned}\mu^*\left(\left(\bigcup_{i=1}^{n} A_i\right) \cap E\right) &= \mu^*\left(\left(\bigcup_{i=1}^{n-1} A_i \cup A_n\right) \cap E\right) \\ &= \mu^*\left(\left(\bigcup_{i=1}^{n-1} A_i\right) \cap E\right) + \mu^*(A_n \cap E).\end{aligned}$$

Then, by the induction hypothesis, the result follows. Assume now that I is countably infinite, i.e., $\cup_{i\in I} A_i = \cup_{i\in\mathbb{N}} A_i$. Since μ^* is an outer measure, it satisfies (ii) of Definition 6.2. It follows that

$$\mu^*\left(\left(\bigcup_{i\in\mathbb{N}} A_i\right) \cap E\right) \geq \mu^*\left(\left(\bigcup_{i=1}^{n} A_i\right) \cap E\right) = \sum_{i=1}^{n} \mu^*(A_i \cap E).$$

If we let $n \to \infty$, we obtain that

$$\mu^*\left(\left(\bigcup_{i\in\mathbb{N}} A_i\right) \cap E\right) \geq \sum_{i\in\mathbb{N}} \mu^*(A_i \cap E).$$

This completes the argument since the other inequality follows from the fact that μ^* is countable subadditive on $\mathcal{P}(\Omega)$. □

Proof of Proposition 6.6. We first notice that under the assumption that (I) is true, i.e., $\mathcal{M}(\mu^*)$ is a σ-field, item (II) is to show that for any disjoint collection $\{A_i : i \in \mathbb{N}\} \subset \mathcal{M}(\mu^*)$, $\mu^*(\cup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} \mu^*(A_i)$. This follows immediately from Proposition B.2 if we take $E = \Omega$. Hence, it remains to show that $\mathcal{M}(\mu^*)$ is a σ-field. We need to verify that $\mathcal{M}(\mu^*)$ satisfies the items of Definition 4.1. Let $E \in \mathcal{P}(\Omega)$. We have that

$$\mu^*(\Omega \cap E) + \mu^*(\Omega^c \cap E) = \mu^*(E).$$

Thus, $\Omega \in \mathcal{M}(\mu^*)$. Let $A \in \mathcal{M}(\mu^*)$, then

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

since $A \in \mathcal{M}(\mu^*)$. Thus, items (i) and (ii) of Definition 4.1 are clearly satisfied. We notice that for any given $A \in \mathcal{P}(\Omega)$,

$$E = \Omega \cap E = (A \cup A^c) \cap E = (A \cap E) \cup (A^c \cap E).$$

Hence, since μ^* is an outer measure and therefore countable subadditive on $\mathcal{P}(\Omega)$, it follows that for any $A \in \mathcal{P}(\Omega)$, $\mu^*(E) \leq \mu^*(A \cap E) + \mu^*(A^c \cap E)$. Thus, A is μ^* measurable if A is s.t. for any $E \in \mathcal{P}(\Omega)$, $\mu^*(E) \geq \mu^*(A \cap E) + \mu^*(A^c \cap E)$. Hence, if μ^* is an outer measure,

$$\mathcal{M}(\mu^*) = \{A \in \mathcal{P}(\Omega) \colon \mu^*(A \cap E) + \mu^*(A^c \cap E) \leq \mu^*(E) \; \forall E \in \mathcal{P}(\Omega)\}.$$

We show that if $A, B \in \mathcal{M}(\mu^*)$, then $A \cup B \in \mathcal{M}(\mu^*)$. Since $A \in \mathcal{M}(\mu^*)$,

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

It follows that

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

Since $B \in \mathcal{M}(\mu^*)$, $\mu^*(A^c \cap E) = \mu^*(B^c \cap A^c \cap E) + \mu^*(B \cap A^c \cap E)$. Hence,

$$\mu^*\big((A \cup B) \cap E\big) + \mu^*\big((A \cup B)^c \cap E\big) \leq \mu^*\big(A \cap E\big) + \mu^*(A^c \cap E) = \mu^*(E),$$

which shows that $A \cup B \in \mathcal{M}(\mu^*)$. By induction, if $A_1, \dots, A_n \in \mathcal{M}(\mu^*)$, we have that $\cup_{i=1}^n A_i \in \mathcal{M}(\mu^*)$. To show (iii) of Definition 4.1, assume first that $\{A_i : i \in \mathbb{N}\} \subset \mathcal{M}(\mu^*)$ is disjoint. Write $A = \cup_{i\in\mathbb{N}} A_i$. We show that $A \in \mathcal{M}(\mu^*)$. We know that for any $n \in \mathbb{N}$, $F_n = \cup_{i=1}^n A_i \in \mathcal{M}(\mu^*)$, i.e., $\mu^*(E) = \mu^*(F_n \cap E) + \mu^*(F_n^c \cap E)$. Using Proposition B.2, we know that $\mu^*(F_n \cap E) = \sum_{i=1}^n \mu^*(A_i \cap E)$. Also, $A^c \subset F_n^c$ ($F_n \subset A$). Hence, $\mu^*(F_n^c \cap E) \geq \mu^*(A^c \cap E)$. Therefore, we obtain, $\mu^*(E) \geq \sum_{i=1}^n \mu^*(A_i \cap E) + \mu^*(A^c \cap E)$. If we let $n \to \infty$, we obtain $\mu^*(E) \geq \sum_{i=1}^\infty \mu^*(A_i \cap E) + \mu^*(A^c \cap E)$. Thus, using Proposition B.2 again, we have that $\mu^*(E) \geq \mu^*(A \cap E) + \mu^*(A^c \cap E)$ and thus $A \in \mathcal{M}(\mu^*)$. Let now $\{B_i : i \in \mathbb{N}\} \subset \mathcal{M}(\mu^*)$, not necessarily disjoint. Write $B = \cup_{i\in\mathbb{N}} B_i$. We want to show that $B \in \mathcal{M}(\mu^*)$. Let $A_1 = B_1$, $A_2 = B_2 \setminus B_1 = B_2 \cap B_1^c$ and so on until we define

$$A_i = B_i \setminus \left(\bigcup_{k=1}^{i-1} B_k\right) = B_i \cap B_1^c \cap \ldots \cap B_{i-1}^c.$$

Then, see the proof of item (vii) in Proposition 5.1, $\{A_i : i \in \mathbb{N}\}$ is disjoint and s.t. for any $n \in \mathbb{N}$, $\cup_{i=1}^n A_i = \cup_{i=1}^n B_i$. In particular, $\cup_{i\in\mathbb{N}} A_i = \cup_{i\in\mathbb{N}} B_i$. Clearly, $\{A_i : i \in \mathbb{N}\} \subset \mathcal{M}(\mu^*)$ and hence, since $\{A_i : i \in \mathbb{N}\}$ is disjoint, $\cup_{i\in\mathbb{N}} A_i \in \mathcal{M}(\mu^*)$ and therefore $\cup_{i\in\mathbb{N}} B_i \in \mathcal{M}(\mu^*)$ as well. This completes the proof of item (I). $\square$

Proof of Proposition 6.7. Using Proposition 6.5, we know that the function

$$\rho^*(A) = \inf_{\xi \in C_{\mathcal{A}}(A)} v_\rho(\xi), \quad A \in \mathcal{P}(\Omega),$$

is an outer measure on $\mathcal{P}(\Omega)$. We also know that $\mathcal{M}(\rho^*)$ is a σ-field and ρ^* restricted to $\mathcal{M}(\rho^*)$ is a measure (cf. Proposition 6.6). As a first step, we show that $\mathcal{M}(\rho^*)$ contains $\mathcal{A}$, i.e., $\mathcal{A} \subset \mathcal{M}(\rho^*)$. Let $A \in \mathcal{A}$. We need to show that for any $E \in \mathcal{P}(\Omega)$,

$$\rho^*(E) \geq \rho^*(A \cap E) + \rho^*(A^c \cap E). \tag{57}$$

It is clear that if $\rho^*(E) = \infty$, (57) is true. Thus consider the case where $\rho^*(E) < \infty$. We apply the same strategy as in the proof of Proposition 6.5 and choose for any $\varepsilon > 0$, a covering $\xi_\varepsilon = \{U_n : n \in \mathbb{N}\} \in C_{\mathcal{A}}(E)$ s.t. $v_\rho(\xi) = \sum_{n\in\mathbb{N}} \rho(U_n) < \rho^*(E) + \varepsilon$. Since $\xi_\varepsilon \in C_{\mathcal{A}}(E)$, we know that for any $n \in \mathbb{N}$, $U_n \in \mathcal{A}$. In particular, since $\mathcal{A}$ is a semiring, $B_n = A \cap U_n \in \mathcal{A}$ for any $n \in \mathbb{N}$. Hence, since $B_n \subset U_n$, we have that for any $n \in \mathbb{N}$, the set $U_n \setminus B_n$ has the form $\cup_{k=1}^{m_n} C_{n_k}$ where $\{C_{n_k} : k = 1, \ldots, m_n\} \subset \mathcal{A}$ is disjoint. Now we notice the following,

1. $U_n = (U_n \setminus B_n) \cup B_n = (\cup_{k=1}^{m_n} C_{n_k}) \cup B_n$, where $\{C_{n_1}, \ldots, C_{n_{m_n}}, B_n\}$ is disjoint;
2. $A \cap E \subset A \cap (\cup_{n\in\mathbb{N}} U_n) = \cup_{n\in\mathbb{N}}(A \cap U_n) = \cup_{n\in\mathbb{N}} B_n$;
3. $U_n \setminus B_n = U_n \cap B_n^c = U_n \cap A^c$;
4. $A^c \cap E \subset A^c \cap (\cup_{n\in\mathbb{N}} U_n) = \cup_{n\in\mathbb{N}}(A^c \cap U_n) = \cup_{n\in\mathbb{N}}(\cup_{k=1}^{m_n} C_{n_k})$.

Therefore, by 1, since ρ is finitely additive on $\mathcal{A}$, for any $n \in \mathbb{N}$, $\sum_{k=1}^{m_k} \rho(C_{n_k}) + \rho(B_n) = \rho((\cup_{k=1}^{m_n} C_{n_k}) \cup B_n) = \rho(U_n)$. Then, using 2 and 4 and the fact that ρ^* is an outer measure it follows that

$$\begin{aligned}
\rho^*(A \cap E) + \rho^*(A^c \cap E) &\leq \rho^*\left(\bigcup_{n\in\mathbb{N}} B_n\right) + \rho^*\left(\bigcup_{n\in\mathbb{N}}\left(\bigcup_{k=1}^{m_n} C_{n_k}\right)\right) \\
&\leq \sum_{n\in\mathbb{N}} \rho^*(B_n) + \sum_{n\in\mathbb{N}}\left(\sum_{k=1}^{m_k} \rho^*(C_{n_k})\right) \\
&\leq \sum_{n\in\mathbb{N}} \rho(B_n) + \sum_{n\in\mathbb{N}}\left(\sum_{k=1}^{m_k} \rho(C_{n_k})\right) \\
&= \sum_{n\in\mathbb{N}}\left(\rho(B_n) + \sum_{k=1}^{m_k} \rho(C_{n_k})\right) \\
&= \sum_{n\in\mathbb{N}} \rho\left(B_n \cup \left(\bigcup_{k=1}^{m_n} C_{n_k}\right)\right) = \sum_{n\in\mathbb{N}} \rho(U_n) < \rho^*(E) + \varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, (57) is satisfied. Therefore, $\mathcal{A} \subset \mathcal{M}(\rho^*)$. We show that $\rho^*(A) = \rho(A)$ for any $A \in \mathcal{A}$. Notice that we have already shown this for the special case where $\rho = \ell$ (cf. Proposition 6.4 in Example 6.1). To see the general case, let $A \in \mathcal{A}$ and consider any covering $\xi = \{U_n : n \in \mathbb{N}\} \in C_{\mathcal{A}}(A)$. In particular, $A \subset \cup_{n\in\mathbb{N}} U_n$, and therefore, $A = (\cup_{n\in\mathbb{N}} U_n) \cap A = \cup_{n\in\mathbb{N}}(U_n \cap A)$. Since ρ is countable subadditive on $\mathcal{A}$, it follows that $\rho(A) = \rho(\cup_{n\in\mathbb{N}}(U_n \cap A)) \leq \sum_{n\in\mathbb{N}} \rho(U_n \cap A)$. By Exercise 6.5, ρ is monotone and hence $\rho(A) \leq \sum_{n\in\mathbb{N}} \rho(U_n) = v_\rho(\xi)$. Therefore, it follows that $\rho(A) \leq \inf_{\xi\in C_{\mathcal{A}}(A)} v_\rho(\xi) = \rho^*(A)$, since $\xi \in C_{\mathcal{A}}(A)$ was arbitrary. The other inequality follows immediately, since if $A \in \mathcal{A}$, $\{A\} \in C_{\mathcal{A}}(A)$ and $\rho(A) = v_\rho(\{A\}) \geq \rho^*(A)$. Hence, ρ^* and ρ agree on $\mathcal{A}$. By Proposition 6.6, $\mathcal{M}(\rho^*)$ is a σ-filed. Further, $\mathcal{M}(\rho^*)$ contains $\mathcal{A}$. By definition, $\sigma(\mathcal{A})$ is the smallest σ-field that contains $\mathcal{A}$. Hence, $\sigma(\mathcal{A}) \subset \mathcal{M}(\rho^*)$. Using Proposition 6.6 again, we know that $\rho^*|_{\mathcal{M}(\rho^*)}$ is a measure. Since $\sigma(\mathcal{A})$ is a σ-field and $\sigma(\mathcal{A}) \subset \mathcal{M}(\rho^*)$, it follows that $\rho^*|_{\sigma(\mathcal{A})}$ is a measure as well. Hence, we set $\rho_\uparrow = \rho^*|_{\sigma(\mathcal{A})}$ and obtain a measure on $\sigma(\mathcal{A})$ which is s.t. for any $A \in \mathcal{A}$, $\rho_\uparrow(A) = \rho^*|_{\sigma(\mathcal{A})}(A) = \rho^*(A) = \rho(A)$. $\square$

B.3 π-λ theorem

Definition B.1. *Let $\Omega \neq \emptyset$. A collection $\mathscr{P} \subset \mathcal{P}(\Omega)$ is called a π-system if $A, B \in \mathscr{P}$ implies that $A \cap B \in \mathscr{P}$.*

Definition B.2. *Let $\Omega \neq \emptyset$. A collection $\mathscr{L} \subset \mathcal{P}(\Omega)$ is called a λ-system if*

(i) $\Omega \in \mathscr{L}$*;*

(ii) $A \in \mathscr{L}$ *implies that* $A^c \in \mathscr{L}$*;*

(iii) if $\{A_i \colon i \in \mathbb{N}\} \subset \mathscr{L}$ *s.t.* $\{A_i \colon i \in \mathbb{N}\}$ *is disjoint, then* $\cup_{i\in\mathbb{N}} A_i \in \mathscr{L}$*.*

Proposition B.3. *Let $\Omega \neq \emptyset$. If $\mathcal{F}$ is a π-system and a λ-system, then it is a σ-field on Ω.*

Proof. Since $\mathcal{F}$ is a λ-system, it follows that $\Omega \in \mathcal{F}$ and $\mathcal{F}$ is closed under the formation of complements. Hence it remains to show item (iii) of Definition 4.1. Let $\{A_i \colon i \in \mathbb{N}\} \subset \mathcal{F}$. Given $i \in \mathbb{N}$, we write

$$B_i = A_i \setminus \left(\bigcup_{j=1}^{i-1} A_j \right) = A_i \cap A_1^c \cap \ldots \cap A_{i-1}^c.$$

We already know that $\{B_i \colon i \in \mathbb{N}\}$ is disjoint (recall the proof of item (vii) of Proposition 5.1). Since $\mathcal{F}$ is a π-system, $B_i \in \mathcal{F}$ for any $i \in \mathbb{N}$ and hence, by item (iii) of Definition B.2, $\cup_{i\in\mathbb{N}} B_i \in \mathcal{F}$. Then, $\mathcal{F}$ is a σ-field, since for any $n \in \mathbb{N}$, $\cup_{i=1}^n B_i = \cup_{i=1}^n A_i$ and hence $\cup_{i\in\mathbb{N}} B_i = \cup_{i\in\mathbb{N}} A_i$. $\square$

The following is known as Dynkin's π-λ theorem.

Proposition B.4. *Let $\Omega \neq \emptyset$. If $\mathscr{P}$ is a π-system and $\mathscr{L}$ is a λ system then, if $\mathscr{P} \subset \mathscr{L}$, it follows that $\sigma(\mathscr{P}) \subset \mathscr{L}$.*

Proof. Suppose that $\mathscr{P} \subset \mathscr{L}$. Define the set

$$\mathscr{C} = \{\mathcal{L} \colon \mathcal{L} \text{ is a } \lambda\text{-system s.t. } \mathscr{P} \subset \mathcal{L}\}.$$

Then, similar to the definition of the σ-field generated by a family of subsets of Ω, we set

$$\mathscr{L}_0 = \bigcap_{\mathcal{L} \in \mathscr{C}} \mathcal{L}.$$

Again, $\mathscr{L}_0$ is not empty, since $\mathcal{P}(\Omega) \in \mathscr{C}$. Further, $\mathscr{L}_0$ is a λ-system (cf. Exercise 4.5). Upon the assumption that $\mathscr{P} \subset \mathscr{L}$, we obtain $\mathscr{P} \subset \mathscr{L}_0 \subset \mathscr{L}$. If $\mathscr{L}_0$ is a π-system, then $\mathscr{L}_0$ is a σ-field on Ω (Proposition B.3). Hence, since $\sigma(\mathscr{P})$ is the smallest σ-field that contains $\mathscr{P}$, it follows that $\sigma(\mathscr{P}) \subset \mathscr{L}_0 \subset \mathscr{L}$ and we are done. Therefore, we prove that $\mathscr{L}_0$ is a π-system. Given any $A \in \mathcal{P}(\Omega)$ we define

$$\mathscr{L}_A = \{B \in \mathcal{P}(\Omega) \colon A \cap B \in \mathscr{L}_0\}.$$

Suppose that $A \in \mathscr{P}$. Then $\mathscr{L}_A$ is a λ-system. To see it, notice that $\Omega \in \mathscr{L}_A$, since $A \cap \Omega = A \in \mathscr{L}_0$ (recall that by definition of $\mathscr{L}_0$, $A \in \mathscr{P}$ implies that $A \in \mathcal{L}$ for any $\mathcal{L} \in \mathscr{C}$). Suppose that $B \in \mathscr{L}_A$. Then, $A \cap B \in \mathscr{L}_0$. We also know that $A \cap \Omega \in \mathscr{L}_0$. Hence, since $\mathscr{L}_0$ is a λ-system, it contains $(A\cap\Omega)\cap(A\cap B)^c = (A\cap\Omega)\setminus(A\cap B)$ (notice that $(A\cap\Omega)^c \cap A \cap B = \emptyset$). Then, since $(A \cap \Omega) \setminus (A \cap B) = A \cap (\Omega \setminus B)$, it follows that $\Omega \setminus B = B^c \in \mathscr{L}_A$. Suppose that $\{B_i \colon i \in \mathbb{N}\} \subset \mathscr{L}_A$ disjoint, then $\{A \cap B_i \colon i \in \mathbb{N}\} \subset \mathscr{L}_0$, where also $\{A \cap B_i \colon i \in \mathbb{N}\}$ is disjoint. Thus, since $\mathscr{L}_0$ is a λ-system, $\cup_{i\in\mathbb{N}}(A\cap B_i) = A\cap(\cup_{i\in\mathbb{N}} B_i) \in \mathscr{L}_0$, i.e., $\cup_{i\in\mathbb{N}} B_i \in \mathscr{L}_A$. Therefore, we have shown that if $A \in \mathscr{P}$, then $\mathscr{L}_A$ is a λ-system. Further, if $A \in \mathscr{P}$, then $\mathscr{P} \subset \mathscr{L}_A$, since in this case, for any $B \in \mathscr{P}$, $A \cap B \in \mathscr{P}$ ($\mathscr{P}$ is a π-system) and since $\mathscr{P} \subset \mathscr{L}_0$, it follows that $A \cap B \in \mathscr{L}_0$, i.e., $B \in \mathscr{L}_A$. Then, since $\mathscr{L}_0$ is the smallest λ-system that contains $\mathscr{P}$, $\mathscr{L}_0 \subset \mathscr{L}_A$ if $A \in \mathscr{P}$. This means that if $A \in \mathscr{P}$ and $B \in \mathscr{L}_0$, then $B \in \mathscr{L}_A$ and hence $A \in \mathscr{L}_B$. Notice that $B \in \mathscr{L}_A$ if and only if $A \in \mathscr{L}_B$. Thus, $B \in \mathscr{L}_0$ implies that $\mathscr{P} \subset \mathscr{L}_B$. Then, since $\mathscr{L}_B$ is a λ-system, it follows that $B \in \mathscr{L}_0$ implies that $\mathscr{L}_0 \subset \mathscr{L}_B$. In conclusion, $B \in \mathscr{L}_0$ and $C \in \mathscr{L}_0$ imply that $C \in \mathscr{L}_B$ and hence $B \cap C \in \mathscr{L}_0$. This shows that $\mathscr{L}_0$ is a π-system. □

Proof of Proposition 6.9. Assume that μ_1 is σ-finite on $\mathcal{A}$ (otherwise μ_2 is). With the terminology of Definition B.1, $\mathcal{A}$ is a π-system. Let $B \in \mathcal{A}$ s.t. $\mu_1(B) < \infty$ (this is possible since μ_1 is σ-finite on $\mathcal{A}$). Hence, since by assumption, μ_1 and μ_2 agree on $\mathcal{A}$, $\mu_1(B) = \mu_2(B)$. Define

$$\mathscr{L}_B = \{A \in \sigma(\mathcal{A}) \colon \mu_1(B \cap A) = \mu_2(B \cap A)\}.$$

Notice that $\mathcal{A} \subset \mathscr{L}_B$ since if $A \in \mathcal{A}$, $A \cap B \in \mathcal{A}$ ($\mathcal{A}$ is a π-system) and as μ_1 and μ_2 agree on $\mathcal{A}$, $\mu_1(B \cap A) = \mu_2(B \cap A)$, i.e., $A \in \mathscr{L}_B$. We show that $\mathscr{L}_B$ is a λ-system. Clearly, $\Omega \in \mathscr{L}_B$, since $\Omega \in \sigma(\mathcal{A})$ s.t. $\mu_1(B \cap \Omega) = \mu_1(B) = \mu_2(B) = \mu_2(B \cap \Omega)$. If $A \in \mathscr{L}_B$, then $A^c \in \sigma(\mathcal{A})$. Then, since $B \cap A^c = B \setminus (B \cap A)$ and $\mu_1(A \cap B) = \mu_2(A \cap B) < \infty$, we obtain (cf. item (iii) of Proposition 5.1),

$$\mu_1(B \cap A^c) = \mu_1(B \setminus (B \cap A)) = \mu_1(B) - \mu_1(B \cap A) = \mu_2(B) - \mu_2(B \cap A) = \mu_2(B \cap A^c),$$

it follows that $A^c \in \mathscr{L}_B$. If $\{A_i \colon i \in \mathbb{N}\} \subset \mathscr{L}_B$, disjoint, then $\cup_{i\in\mathbb{N}} A_i \in \sigma(\mathcal{A})$ and $\{B \cap A_i \colon i \in \mathbb{N}\}$ is disjoint as well. Then,

$$\begin{aligned}\mu_1(B \cap (\cup_{i\in\mathbb{N}} A_i)) &= \mu_1(\cup_{i\in\mathbb{N}}(B \cap A_i)) \\ &= \sum_{i\in\mathbb{N}} \mu_1(B \cap A_i) = \sum_{i\in\mathbb{N}} \mu_2(B \cap A_i) = \mu_2(B \cap (\cup_{i\in\mathbb{N}} A_i)),\end{aligned}$$

and hence $\cup_{i\in\mathbb{N}} A_i \in \mathscr{L}_B$. Thus, $\mathscr{L}_B$ is a λ-system that contains the π-system $\mathcal{A}$ and with Proposition B.4, we obtain that $\sigma(\mathcal{A}) \subset \mathscr{L}_B$. Therefore, we have shown that $\sigma(\mathcal{A}) \subset \mathscr{L}_B$ for any set $B \in \mathcal{A}$ for which $\mu_1(B) < \infty$. We consider a collection $\{B_i \colon i \in \mathbb{N}\} \subset \mathcal{A}$ s.t. $\cup_{i\in\mathbb{N}} B_i = \Omega$ and for any $i \in \mathbb{N}$, $\mu_1(B_i) < \infty$. In particular, given any $n \in \mathbb{N}$ and $I \subset \{1, \ldots, n\}$, since $\mathcal{A}$ is a π-system, $\cap_{i\in I} B_i \in \mathcal{A}$ and $\mu_1(\cap_{i\in I} B_i) < \infty$. Therefore, $\sigma(\mathcal{A}) \subset \mathscr{L}_{\cap_{i\in I} B_i}$. This shows that for any $A \in \sigma(\mathcal{A})$,

$$\mu_1\left(\left(\bigcap_{i\in I} B_i\right) \cap A\right) = \mu_2\left(\left(\bigcap_{i\in I} B_i\right) \cap A\right).$$

Then, upon (Proposition B.1) we obtain that for any $n \in \mathbb{N}$,

$$\mu_1\left(\bigcup_{i=1}^n (B_i \cap A)\right) = \mu_2\left(\bigcup_{i=1}^n (B_i \cap A)\right).$$

Then, we apply item (v) of Proposition 5.1, and conclude that for any $A \in \sigma(\mathcal{A})$,

$$\mu_1(A) = \mu_1\left(\bigcup_{i\in\mathbb{N}} (B_i \cap A)\right) = \mu_2\left(\bigcup_{i\in\mathbb{N}} (B_i \cap A))\right) = \mu_2(A).$$

□

B.4 On measurable functions

Proof of Proposition 7.4. We first show that if for any $i = 1, \ldots, k$, $f_i \colon \Omega \to \mathbb{R}$ is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable, then, f is $\mathcal{F}/\mathfrak{B}(\mathbb{R}^k)$ measurable. By Exercise 4.8, we have that $\mathfrak{B}(\mathbb{R}^k) = \sigma(\mathcal{R}'_k)$, where

$$\mathcal{R}'_k = \{(-\infty, x_1] \times \cdots (-\infty, x_k] \colon x = (x_1, \ldots, x_k) \in \mathbb{R}^k\}.$$

Hence, by Proposition 7.1, it is sufficient to show that for any $R \in \mathcal{R}'_k$, $f^{-1}(R) \in \mathcal{F}$. Therefore, let $R \in \mathcal{R}'_k$, i.e.,

$$R = (-\infty, x_1] \times \cdots (-\infty, x_k], \quad x = (x_1, \ldots, x_k) \in \mathbb{R}^k.$$

Then,

$$f^{-1}(R) = \{\omega \in \Omega \colon f(\omega) \in R\} = \bigcap_{i=1}^{k} \{\omega \in \Omega \colon f_i(\omega) \in (-\infty, x_1]\} = \bigcap_{i=1}^{k} f_i^{-1}((-\infty, x_1]).$$

Then, since for any $i = 1, \ldots, k$, $(-\infty, x_1] \in \mathfrak{B}(\mathbb{R})$ and f_i is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable, it follows that $f_i^{-1}((-\infty, x_1]) \in \mathcal{F}$. Therefore, $f^{-1}(R) \in \mathcal{F}$ as well. For the other direction, assume that f is $\mathcal{F}/\mathfrak{B}(\mathbb{R}^k)$ measurable. Given any $i = 1, \ldots, k$, let

$$A_i^n = (-\infty, n] \times \cdots \times \underbrace{(-\infty, x]}_{\text{position } i} \times \cdots \times (-\infty, n],$$

where $x \in \mathbb{R}$ is arbitrary. Then,

$$f^{-1}(A_i^n) = f_1^{-1}((-\infty, n]) \cap \cdots \cap \underbrace{f_i^{-1}((-\infty, x])}_{\text{position } i} \cap \cdots \cap f_k^{-1}((-\infty, n]).$$

Then,

$$\bigcup_{n \in \mathbb{N}} A_i^n = \mathbb{R} \times \cdots \times \underbrace{(-\infty, x]}_{\text{position } i} \times \cdots \times \mathbb{R}.$$

Therefore,

$$\begin{aligned}\bigcup_{n \in \mathbb{N}} f^{-1}(A_i^n) &= f^{-1}\left(\bigcup_{n \in \mathbb{N}} A_i^n\right) \\ &= f_1^{-1}(\mathbb{R}) \cap \cdots \cap \underbrace{f_i^{-1}((-\infty, x])}_{\text{position } i} \cap \cdots \cap f_k^{-1}(\mathbb{R}) = f_i^{-1}((-\infty, x]).\end{aligned}$$

Since f is $\mathcal{F}/\mathfrak{B}(\mathbb{R}^k)$ measurable, for any $i = 1, \ldots, k$, $\cup_{n \in \mathbb{N}} f^{-1}(A_i^n) \in \mathcal{F}$. Thus, for any $i = 1, \ldots, k$ and any $x \in \mathbb{R}$, $f_i^{-1}((-\infty, x]) \in \mathcal{F}$. Then, $\mathfrak{B}(\mathbb{R}) = \sigma(\{(-\infty, x] \colon x \in \mathbb{R}\})$ and hence, by Proposition 7.1, f_i is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable. This completes the proof. □

Proof of Proposition 7.7. We show by induction (over $n \geq 2$) that there exists $m \in \mathbb{N}$, s.t. f satisfies

$$f(\omega) = \sum_{i=1}^{m} \alpha_i^* \mathbb{1}_{A_i^*}(\omega), \quad \omega \in \Omega,$$

where $\{A_1^*, \ldots, A_m^*\} \subset \mathcal{F}$ is disjoint. Let $n = 2$, i.e., $f(\omega) = \alpha_1 \mathbb{1}_{A_1}(\omega) + \alpha_2 \mathbb{1}_{A_2}(\omega)$, $A_1, A_2 \in \mathcal{F}$. We define the sets $A_1^* = A_1 \setminus A_2$, $A_2^* = A_2 \setminus A_1$ and $A_3^* = A_1 \cap A_2$. Further, we set $\alpha_1^* = \alpha_1$, $\alpha_2^* = \alpha_2$ and $\alpha_3^* = \alpha_1 + \alpha_2$. Then, $\{A_1^*, A_2^*, A_3^*\} \subset \mathcal{F}$ is disjoint. Also, for any $\omega \in \Omega$,

$$f(\omega) = \alpha_1^* \mathbb{1}_{A_1^*}(\omega) + \alpha_2^* \mathbb{1}_{A_2^*}(\omega) + \alpha_3^* \mathbb{1}_{A_3^*}(\omega).$$

With regard to the induction step, assume that $f(\omega) = \sum_{i=1}^{n-1} \alpha_i \mathbb{1}_{A_i}(\omega) + \alpha_n \mathbb{1}_{A_n}(\omega)$, $A_i \in \mathcal{F}$, $i = 1, \ldots, n$, and the induction hypothesis is that $\{A_1, \ldots, A_{n-1}\} \subset \mathcal{F}$ is disjoint. Then, we set

$$A_i^* = A_i \setminus A_n, \quad i = 1, \ldots, n-1, \quad A_n^* = A_n \setminus (\cup_{i=1}^{n-1} A_i),$$

and

$$A^*_{n+i} = A_n \cap A_i, \quad i = 1, \ldots, n-1.$$

It follows that

$$\{A^*_1, \ldots, A^*_n, A^*_{n+1}, \ldots, A^*_{2n-1}\} \subset \mathcal{F}$$

is disjoint. We set $\alpha^*_i = \alpha_i$, $i = 1, \ldots, n$, and $\alpha^*_{n+i} = \alpha_n + \alpha_i$, $i = 1, \ldots, n-1$. It follows that for any $\omega \in \Omega$,

$$f(\omega) = \sum_{i=1}^{2n-1} \alpha^*_i \mathbb{1}_{A^*_i}(\omega).$$

This completes the induction step. Finally, we notice that

$$f(\omega) = \sum_{i=1}^{2n-1} \alpha^*_i \mathbb{1}_{A^*_i}(\omega) + \alpha^*_{2n} \mathbb{1}_{A^*_{2n}}(\omega),$$

with $\alpha^*_{2n} = 0$ and $A^*_{2n} = \Omega \setminus (\cup_{i=1}^{2n-1} A^*_i)$. Thus, $\cup_{i=1}^{2n} A^*_i = \Omega$ and $\{A^*_i : i = 1, \ldots, 2n\} \subset \mathcal{F}$ is disjoint. □

Proof of Proposition 7.8. We have that

$$\{\omega \in \Omega \colon f(\omega) < g(\omega)\} = \bigcup_{q \in \mathbb{Q}} \{\omega \in \Omega \colon f(\omega) < q < g(\omega)\}.$$

To see it, take any ω in the left set of the above equation. Then, by Proposition 1.6, there exists $q \in \mathbb{Q}$ s.t. $f(\omega) < q < g(\omega)$. Notice that is is also true if $g(\omega) = \infty$. Clearly, if there exists $q \in \mathbb{Q}$ s.t. $f(\omega) < q < g(\omega)$, then $f(\omega) < g(\omega)$. Then, we have that for any $q \in \mathbb{Q}$,

$$\{\omega \in \Omega \colon f(\omega) < q < g(\omega)\} = \{\omega \in \Omega \colon f(\omega) < q\} \cap \{\omega \in \Omega \colon g(\omega) > q\}.$$

Since f and g are both $\mathcal{F}$ measurable, $\{\omega \in \Omega \colon f(\omega) < q\}$ and $\{\omega \in \Omega \colon g(\omega) > q\}$ are members of $\mathcal{F}$. This shows that $\{\omega \in \Omega \colon f(\omega) < g(\omega)\} \in \mathcal{F}$ since $\mathbb{Q}$ is countable. A similar argument can be used to show that $\{\omega \in \Omega \colon f(\omega) > g(\omega)\} \in \mathcal{F}$. Therefore,

$$\mathcal{F} \ni \{\omega \in \Omega \colon f(\omega) = g(\omega)\} = \big(\{\omega \in \Omega \colon f(\omega) < g(\omega)\} \cup \{\omega \in \Omega \colon f(\omega) > g(\omega)\}\big)^c.$$

□

Proof of Proposition 7.9. Let $x \in \overline{\mathbb{R}}$. We have that

$$\{\omega \in \Omega \colon \sup_{n \in E} f_n(\omega) \leq x\} = \cap_{n \in E} \{\omega \in \Omega \colon f_n(\omega) \leq x\}.$$

Hence, for any $x \in \overline{\mathbb{R}}$, $\{\omega \in \Omega \colon \sup_{n \in E} f_n \leq x\} \in \mathcal{F}$ since we have assumed that f_n is $\mathcal{F}$ measurable for any $n \in E$. We notice that

$$\{\omega \in \Omega \colon \sup_{n \in E} f_n(\omega) < \infty\} = \{\omega \in \Omega \colon \sup_{n \in E} f_n(\omega) \leq C\},$$

for some $C \in \mathbb{R}$. Therefore, $\{\omega \in \Omega \colon \sup_{n \in E} f_n(\omega) < \infty\} \in \mathcal{F}$. Hence,

$$\{\omega \in \Omega \colon \sup_{n \in E} f_n(\omega) = \infty\} = \{\omega \in \Omega \colon \sup_{n \in E} f_n(\omega) < \infty\}^c \in \mathcal{F}.$$

Also,

$$\{\omega \in \Omega \colon \sup_{n\in E} f_n(\omega) = -\infty\} = \{\omega \in \Omega \colon \sup_{n\in E} f_n(\omega) \le -\infty\}$$
$$= \cap_{n\in E}\{\omega \in \Omega \colon f_n(\omega) = -\infty\}.$$

Thus, since for any $n \in E$, f_n is $\mathcal{F}$ measurable, we have that$\{\omega \in \Omega \colon f_n(\omega) = -\infty\} \in \mathcal{F}$. Therefore $\{\omega \in \Omega \colon \sup_{n\in E} f_n(\omega) = -\infty\} \in \mathcal{F}$. As in the solution of Exercise 7.4, we define

$$F = \{\omega \in \Omega \colon \sup_{n\in E} f_n(\omega) \in \mathbb{R}\},$$

and set

$$f^*(\omega) = \sup_{n\in E} f_n(\omega)\mathbb{1}_F(\omega), \quad \omega \in \Omega.$$

We notice that $F \in \mathcal{F}$. Also, since for any $x \in \overline{\mathbb{R}}$, $\{\omega \in \Omega \colon \sup_{n\in E} f_n(\omega) \le x\} \in \mathcal{F}$, we obtain that

$$\{\omega \in \Omega \colon f^*(\omega) \le x\} \in \mathcal{F}.$$

Hence, f^* is $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable (cf. Proposition 7.3). Let $A \in \mathfrak{B}(\mathbb{R})$. We obtain

$$\{\omega \in \Omega \colon \sup_{n\in E} f_n(\omega) \in A\} = \big(\{\omega \in \Omega \colon f^*(\omega) \in A\} \cap F\big) \cup (\{\omega \in \Omega \colon \sup_{n\in E} f_n(\omega) \in A\} \cap F^c)$$
$$= \{\omega \in \Omega \colon f^*(\omega) \in A\} \cap F \in \mathcal{F}.$$

This shows that $\omega \mapsto \sup_{n\in E} f_n(\omega)$ is $\mathcal{F}$ measurable. A similar argument shows that $\omega \mapsto \inf_{n\in E} f_n(\omega)$ is $\mathcal{F}$ measurable. With respect to (ii) of Proposition 7.9, we have that

$$\omega \mapsto (\liminf_{n\longrightarrow\infty} f_n)(\omega) = \sup_{n\in\mathbb{N}}(\inf_{k\ge n} f_n)(\omega),$$

and

$$\omega \mapsto (\limsup_{n\longrightarrow\infty} f_n)(\omega) = \inf_{n\in\mathbb{N}}(\sup_{k\ge n} f_n)(\omega),$$

are $\mathcal{F}$ measurable as a consequence of item (i) (cf. Exercise 7.2). Item (iii) is a consequence of Proposition 3.23, we have that

$$\{\omega \in \Omega \colon \lim_{n\to\infty} f_n(\omega) = -\infty\}$$
$$= \{\omega \in \Omega \colon (\liminf_{n\longrightarrow\infty} f_n)(\omega) = -\infty\} \cap \{\omega \in \Omega \colon (\limsup_{n\longrightarrow\infty} f_n)(\omega) = -\infty\}.$$

Therefore, since $\liminf_{n\longrightarrow\infty} f_n$ and $\limsup_{n\longrightarrow\infty} f_n$ are $\mathcal{F}$ measurable, we conclude that

$$\{\omega \in \Omega \colon \lim_{n\to\infty} f_n(\omega) = -\infty\} \in \mathcal{F}.$$

Similarly, we obtain

$$\{\omega \in \Omega \colon \lim_{n\to\infty} f_n(\omega) = \infty\}$$
$$= \{\omega \in \Omega \colon (\liminf_{n\longrightarrow\infty} f_n)(\omega) = \infty\} \cap \{\omega \in \Omega \colon (\limsup_{n\longrightarrow\infty} f_n)(\omega) = \infty\} \in \mathcal{F}.$$

Let $A \in \mathfrak{B}(\mathbb{R})$, we have that

$$\{\omega \in \Omega \colon \lim_{n\to\infty} f_n(\omega) \in A\}$$
$$= \{\omega \in \Omega \colon (\liminf_{n\longrightarrow\infty} f_n)(\omega) = (\limsup_{n\longrightarrow\infty} f_n)(\omega)\} \cap \{\omega \in \Omega \colon (\liminf_{n\longrightarrow\infty} f_n)(\omega) \in A\}.$$

Then, $\{\omega \in \Omega\colon \lim_{n\to\infty} f_n(\omega) \in A\} \in \mathcal{F}$ (cf. Proposition 7.8). Using Propositions 3.21 and 7.8, we obtain

$$\{\omega \in \Omega\colon (f_n(\omega))_{n\in\mathbb{N}} \text{ converges}\} = \{\omega \in \Omega\colon (\liminf_{n\to\infty} f_n)(\omega) = (\limsup_{n\to\infty} f_n)(\omega)\} \in \mathcal{F}.$$

This shows item (iv). Finally, to see item (v), we use again Proposition 3.23 and conclude that

$$\begin{aligned}&\{\omega \in \Omega\colon f_n(\omega) \xrightarrow{n\to\infty} f(\omega)\}\\ &= \{\omega \in \Omega\colon (\liminf_{n\to\infty} f_n)(\omega) = f(\omega)\} \cap \{\omega \in \Omega\colon (\limsup_{n\to\infty} f_n)(\omega) = f(\omega)\}.\end{aligned}$$

Then, since by assumption f is $\mathcal{F}$ measurable, the sets $\{\omega \in \Omega\colon (\liminf_{n\to\infty} f_n)(\omega) = f(\omega)\}$ and $\{\omega \in \Omega\colon (\limsup_{n\to\infty} f_n)(\omega) = f(\omega)\}$ are both elements of $\mathcal{F}$ (cf. Proposition 7.8). This completes the proof of the proposition. □

Proof of Proposition 7.10. The proof is to find such an approximating sequence of simple functions. Given any $n \in \mathbb{N}$, we partition $[0, n)$ as follows:

$$\Big[0, n\Big) = \Big[0, \frac{1}{2^n}\Big) \cup \Big[\frac{1}{2^n}, \frac{2}{2^n}\Big) \cup \cdots \cup \Big[\frac{n2^n - 1}{2^n}, \frac{n2^n}{2^n}\Big) = \bigcup_{i=1}^{n2^n} \Big[\frac{i-1}{2^n}, \frac{i}{2^n}\Big).$$

Then, we label,

$$I_{n,i} = \Big[\frac{i-1}{2^n}, \frac{i}{2^n}\Big), \quad i = 1, \ldots, n2^n,$$

and $I_n = [n, \infty) \cup \{\infty\}$. Accordingly, we define the sets

$$A_{n,i} = \{\omega \in \Omega\colon f(\omega) \in I_{n,i}\}, \quad i = 1, \ldots, n2^n,$$

and $A_n = \{\omega \in \Omega\colon f(\omega) \in I_n\}$. We let

$$f_n(\omega) = \sum_{i=1}^{n2^n} \frac{i-1}{2^n} \mathbb{1}_{A_{n,i}}(\omega) + n\mathbb{1}_{A_n}(\omega).$$

Clearly, for any $\omega \in \Omega$, $f_n(\omega) \le f_{n+1}(\omega)$ and for any $n \in \mathbb{N}$, f_n is $\mathcal{F}$ measurable. Suppose first that $f(\omega) = \infty$. Then, for any $n \in \mathbb{N}$, $f_n(\omega) = n$, i.e., $f_n(\omega) \uparrow f(\omega)$. Assume that $f(\omega) \in [0, \infty)$. Then, there exists $N \in \mathbb{N}$ s.t. $f(\omega) < N$. Hence, there exists precisely one $i \in \{1, \ldots, N2^N\}$, s.t. $f(\omega) \in I_{N,i}$, i.e., $(i-1)2^{-N} \le f(\omega) < i2^{-N}$. In particular, $f(\omega) \ge (i-1)2^{-N} = f_N(\omega)$. Therefore,

$$|f(\omega) - f_N(\omega)| = f(\omega) - f_N(\omega) < \frac{i}{2^N} - \frac{i-1}{2^N} = \frac{1}{2^N}.$$

Then, for any $n \ge N$, $f(\omega) < N \le n$. Hence, we repeat the above argument and conclude that for any $n \ge N$, there is precisely one $k \in \{1, \ldots, n2^n\}$ s.t. $f(\omega) \in I_{n,k}$ and $|f(\omega) - f_n(\omega)| < 2^{-n}$. Therefore, $\lim_{n\to\infty} f_n(\omega) = f(\omega)$. The proof is complete upon application of Proposition 7.7, i.e., for any $n \in \mathbb{N}$ there exists a standard simple function g_n s.t. for any $\omega \in \Omega$, $f_n(\omega) = g_n(\omega)$. □

Proof of Proposition 7.13. Let $g\colon \mathbb{R}^k \to \mathbb{R}$ be $\mathfrak{B}(\mathbb{R}^k)$ measurable and s.t. $h(\omega) = g(f(\omega))$. Then, h is $\sigma(f)$ measurable (Exercise 7.9). For the other direction, we prove the claim in two steps, first we show it for the case where h is a standard simple function and then we approximate a general h with simple functions. Thus, suppose that h is a simple function in standard form (cf. Definition 7.4 with $\mathcal{F} = \sigma(f)$). Then, there exists $g\colon \mathbb{R}^k \to \mathbb{R}$ which is $\mathfrak{B}(\mathbb{R}^k)$ measurable and s.t. $h(\omega) = g(f(\omega))$ (Exercise 7.10). Let $h\colon \Omega \to \mathbb{R}$ be a general $\sigma(f)$

measurable function. By Proposition 7.11, we find a sequence of $\sigma(f)$ measurable standard simple functions $(h_n)_{n\in\mathbb{N}}$ s.t. $h_n \xrightarrow{n\to\infty} h$. By the previous case, for each $n \in \mathbb{N}$, there exists a $\mathfrak{B}(\mathbb{R}^k)$ measurable function g_n s.t. for any $\omega \in \Omega$, $h_n(\omega) = g_n(f(\omega))$. Consider the set $A = \{x \in \mathbb{R}^k : (g_n(x))_{n\in\mathbb{N}} \text{ converges}\}$. Since for any $n \in \mathbb{N}$, g_n is $\mathfrak{B}(\mathbb{R}^k)$ measurable, it follows that $A \in \mathfrak{B}(\mathbb{R}^k)$ (cf. item (iv) of Proposition 7.9). Define

$$g(x) = \begin{cases} \lim_{n\to\infty} g_n(x), & \text{if } x \in A, \\ 0, & \text{otherwise.} \end{cases}$$

We have that $g(x) = \lim_{n\to\infty} g_n(x)\mathbb{1}_A(x)$ and hence, g is $\mathfrak{B}(\mathbb{R}^k)$ measurable by item (iii) of Proposition 7.9. Let $\omega \in \Omega$, then $f(\omega)$ is s.t.

$$\lim_{n\to\infty} g_n(f(\omega)) = \lim_{n\to\infty} h_n(\omega) = h(\omega).$$

Thus, $f(\omega) \in A$ for any $\omega \in \Omega$. Hence, for any $\omega \in \Omega$,

$$h(\omega) = \lim_{n\to\infty} h_n(\omega) = \lim_{n\to\infty} g_n(f(\omega)) = \lim_{n\to\infty} g_n(f(\omega))\mathbb{1}_A(f(\omega)) = g(f(\omega)).$$

□

B.5 The integral

Proof of Proposition 8.1. Notice first that we have seen in Example 7.7 that f is $\mathcal{F}$ measurable. Let $\eta = \{B_j : j = 1, \ldots, M\} \in Z_0^{\mathcal{F}}$. Since $\eta \in Z_0^{\mathcal{F}}$, we can represent the atoms of ξ using the atoms of η and vice versa, i.e.,

$$A_i = A_i \cap \Omega = \bigcup_{j=1}^{M} (A_i \cap B_j), \quad i = 1, \ldots, N,$$

and

$$B_j = \bigcup_{i=1}^{N} (A_i \cap B_j), \quad j = 1, \ldots, M.$$

Further, since ξ is disjoint, it follows that for any $j = 1, \ldots, M$, $\{A_i \cap B_j : i = 1, \ldots, N\}$ is disjoint. Therefore,

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

Further, given any $(j, i) \in \{1, \ldots, M\} \times \{1, \ldots, N\}$, either $A_i \cap B_j \neq \emptyset$, then

$$\Big(\inf_{\omega\in B_j} f(\omega)\Big)\mu(A_i \cap B_j) \leq \Big(\inf_{\omega\in B_j\cap A_i} f(\omega)\Big)\mu(A_i \cap B_j) = \alpha_i\mu(A_i \cap B_j),$$

or $A_i \cap B_j = \emptyset$, then, $(\inf_{\omega\in B_j} f(\omega))\mu(A_i \cap B_j) = 0 = \alpha_i\mu(A_i \cap B_j)$. Hence,

$$\begin{aligned} S_\mu^f(\eta) &= \sum_{j=1}^{M} \Big(\inf_{\omega\in B_j} f(\omega)\Big)\Big(\sum_{i=1}^{N} \mu(A_i \cap B_j)\Big) = \sum_{j=1}^{M} \Big(\sum_{i=1}^{N} \Big(\inf_{\omega\in B_j} f(\omega)\Big)\mu(A_i \cap B_j)\Big) \\ &\leq \sum_{j=1}^{M} \Big(\sum_{i=1}^{N} \alpha_i\mu(A_i \cap B_j)\Big) = \sum_{i=1}^{N} \alpha_i\Big(\sum_{j=1}^{M} \mu(A_i \cap B_j)\Big) = \sum_{i=1}^{N} \alpha_i\mu(A_i). \end{aligned}$$

This shows that

$$\int_\Omega f(\omega)\mu(d\omega) = \sup_{\eta\in Z_0^{\mathcal{F}}} S_\mu^f(\eta) \le \sum_{i=1}^N \alpha_i \mu(A_i).$$

The reverse inequality follows from the fact that ξ is s.t. $\xi \in Z_0^{\mathcal{F}}$ and $S_\mu^f(\xi) = \sum_{i=1}^N \alpha_i\mu(A_i)$. □

We prove the monotone convergence theorem for nonnegative functions.

Proof of Proposition 8.2. First of all, using the result of the latter exercise, we have that for any $n \in \mathbb{N}$,

$$\int_\Omega f_n(\omega)\mu(d\omega) \le \int_\Omega f_{n+1}(\omega)\mu(d\omega)$$

and

$$\sup_{n\in\mathbb{N}} \left(\int_\Omega f_n(\omega)\mu(d\omega) \right) \le \int_\Omega f(\omega)\mu(d\omega).$$

Hence, using Proposition 3.8,

$$\lim_{n\to\infty} \left(\int_\Omega f_n(\omega)\mu(d\omega) \right) \le \int_\Omega f(\omega)\mu(d\omega).$$

Thus, it remains to show that

$$\lim_{n\to\infty} \left(\int_\Omega f_n(\omega)\mu(d\omega) \right) \ge \int_\Omega f(\omega)\mu(d\omega).$$

That is, we need to show that for any $\xi = \{A_i : i = 1, \dots, N\} \in Z_0^{\mathcal{F}}$,

$$\lim_{n\to\infty} \left(\int_\Omega f_n(\omega)\mu(d\omega) \right) \ge S_\mu^f(\xi). \tag{58}$$

Hence, let $\xi = \{A_i : i = 1, \dots, N\} \in Z_0^{\mathcal{F}}$. We use the notation

$$S_\mu^f(\xi) = \sum_{i=1}^N a_i\mu(A_i), \quad a_i = \inf_{\omega\in A_i} f(\omega),\ i = 1, \dots, N.$$

First we consider the case where ξ is s.t. $S_\mu^f(\xi) < \infty$ and for any $i = 1, \dots, N$, $0 < a_i < \infty$ and $0 < \mu(A_i) < \infty$. Let $\varepsilon > 0$, s.t. $\varepsilon < \min\{a_i : i = 1, \dots, N\}$ and

$$\varepsilon < \frac{\delta}{\sum_{i=1}^N \mu(A_i)}, \tag{59}$$

where $\delta > 0$ is arbitrary but nonnegative. Define the sets

$$A_{i,n} = \{\omega \in A_i : f_n(\omega) > a_i - \varepsilon\}, \quad i = 1, \dots, N,\ n \in \mathbb{N}.$$

Since $(f_n(\omega))_{n\in\mathbb{N}}$ is increasing, we observe that for any $i = 1, \dots, N$, $A_{i,n} \subset A_{i,n+1}$. Further, for any $i = 1, \dots, N$,

$$\bigcup_{n\in\mathbb{N}} A_{i,n} = A_i.$$

Clearly, if $\omega' \in \cup_{n\in\mathbb{N}} A_{i,n}$, $\omega' \in A_i$. For the other direction, suppose that $\omega' \in A_i$, then, since $f_n(\omega') \uparrow f(\omega')$, there exists $K \in \mathbb{N}$ s.t. for any $n \geq K$,

$$f(\omega') - f_n(\omega') = |f(\omega') - f_n(\omega')| < \varepsilon.$$

Therefore,

$$f_n(\omega') > f(\omega') - \varepsilon \geq \inf_{\omega \in A_i} f(\omega) - \varepsilon.$$

Given any $n \in \mathbb{N}$, consider the partition

$$\xi_n = \big\{A_{i,n} \colon i = 1, \dots, N\big\} \cup \big\{\Omega \setminus (\cup_{i=1}^N A_{i,n})\big\}.$$

We notice that since f_n is $\mathcal{F}$ measurable, $\xi_n \in Z_0^{\mathcal{F}}$. Hence,

$$\begin{aligned}\int_\Omega f_n(\omega)\mu(d\omega) &\geq S_\mu^{f_n}(\xi_n) = \sum_{A\in\xi_n} \Big(\inf_{\omega\in A} f_n(\omega)\Big)\mu(A) \geq \sum_{i=1}^N \Big(\inf_{\omega\in A_{i,n}} f_n(\omega)\Big)\mu(A_{i,n}) \\ &\geq \sum_{i=1}^N \big(a_i - \varepsilon\big)\mu(A_{i,n}).\end{aligned}$$

Then, using item (v) of Proposition 5.1, we obtain that

$$\sum_{i=1}^N \big(a_i - \varepsilon\big)\mu(A_{i,n}) \uparrow \sum_{i=1}^N \big(a_i - \varepsilon\big)\mu(A_i) = S_\mu^f(\xi) - \varepsilon \sum_{i=1}^N \mu(A_i).$$

Therefore,

$$\lim_{n\to\infty} \left(\int_\Omega f_n(\omega)\mu(d\omega)\right) \geq S_\mu^f(\xi) - \varepsilon \sum_{i=1}^N \mu(A_i).$$

Hence, by (59),

$$\lim_{n\to\infty} \left(\int_\Omega f_n(\omega)\mu(d\omega)\right) \geq S_\mu^f(\xi) - \delta.$$

Since $\delta > 0$ was arbitrary, (58) is shown. Consider now the case, where ξ is s.t. $S_\mu^f(\xi) < \infty$. This implies that for any $i = 1, \dots, N$, $a_i\mu(A_i) < \infty$. Clearly, if for any $i = 1, \dots, N$, $a_i\mu(A_i) = 0$, then (58) is true. Thus, suppose that there exists $i_1, \dots, i_{N_0} \in \{1, \dots, N\}$, $N_0 \leq N$, s.t. $a_{i_j}\mu(A_{i_j}) > 0$ for any $j = 1, \dots, N_0$, and $a_i\mu(A_i) = 0$ for any $i \in \{1, \dots, N\} \setminus \{i_1, \dots, i_{N_0}\}$. Hence, since μ is a measure, $a_{i_j} > 0$ and $\mu(A_{i_j}) > 0$ for any $j = 1, \dots, N_0$. Therefore, we obtain

$$S_\mu^f(\xi) = \sum_{j=1}^{N_0} a_{i_j}\mu(A_{i_j}),$$

and proceed towards (58) as in the previous case. Finally, we consider the case where ξ is s.t. $S_\mu^f(\xi) = \infty$. We need to show that

$$\lim_{n\to\infty} \left(\int_\Omega f_n(\omega)\mu(d\omega)\right) = \infty. \tag{60}$$

Since $S_\mu^f(\xi) = \infty$, there exists $j \in \{1, \dots, N\}$ s.t. $a_j\mu(A_j) = \infty$. Hence, $a_j > 0$ and $\mu(A_j) > 0$ and either $a_j = \infty$ or $\mu(A_j) = \infty$. Let $p, q > 0$ s.t. $0 < p < a_j \leq \infty$ and $0 < q < \mu(A_j) \leq \infty$. Set

$$A_{j,n} = \{\omega \in A_j \colon f_n(\omega) > p\}.$$

We have that $A_{j,n} \subset A_{j,n+1}$ and $A_j = \cup_{n\in\mathbb{N}} A_{j,n}$. We notice that $A_j \subset \cup_{n\in\mathbb{N}} A_{j,n}$ follows from the fact that if $\omega \in A_j$, then since $f_n(\omega) \uparrow f(\omega)$, it follows that there exists $K \in \mathbb{N}$ s.t. for any $n \geq K$, $f(\omega) - f_n(\omega) < a_j - p$, i.e., $f_n(\omega) > f(\omega) - a_j + p \geq p$. Therefore, using item (v) of Proposition 5.1 there exists $K_0 \in \mathbb{N}$ s.t. for any $n \geq K_0$, $\mu(A_j) - \mu(A_{j,n}) < \mu(A_j) - q$. That is, $\mu(A_{j,n}) > q$ for any $n \geq K_0$. Then, consider the partition ξ_n composed of the sets $A_{j,n}$ and $A^c_{j,n}$. We have that for any $n \geq K_0$,

$$\int_\Omega f_n(\omega)\mu(d\omega) \geq S^{f_n}_\mu(\xi_n) \geq p\mu(A_{j,n}) > pq.$$

Therefore,

$$\lim_{n\to\infty} \left(\int_\Omega f_n(\omega)\mu(d\omega) \right) \geq pq.$$

Since either $a_j = \infty$ or $\mu(A_j) = \infty$, either p or q can be made arbitrary large and (60) follows. $\square$

We show the linearity propoerty of the integral for nonnegative functions.

Proof of Proposition 8.3. Suppose that $f = \sum_{i=1}^N \alpha_i \mathbb{1}_{A_i}$, $\alpha_i \in [0,\infty)$, $\{A_i : i = 1,\dots,N\} \subset \mathcal{F}$, $i = 1,\dots,N$, and $g = \sum_{j=1}^M \beta_j \mathbb{1}_{B_j}$, $\beta_j \in [0,\infty)$, $\{B_j : j = 1,\dots,M\} \subset \mathcal{F}$, $j = 1,\dots,M$, i.e., f and g are nonnegative, $\mathcal{F}$ measurable simple functions (cf. Definition 7.3). According to Proposition 7.7, we assume that f and g are already in standard form, i.e., $\cup_{i=1}^N A_i = \cup_{j=1}^M B_j = \Omega$. We write

$$\alpha f = \sum_{i=1}^N \alpha\alpha_i \mathbb{1}_{A_i} = \sum_{i=1}^N \alpha\alpha_i \left(\sum_{j=1}^M \mathbb{1}_{A_i\cap B_j} \right),$$

and

$$\beta g = \sum_{j=1}^M \beta\beta_i \mathbb{1}_{B_j} = \sum_{j=1}^M \beta\beta_j \left(\sum_{i=1}^N \mathbb{1}_{B_j\cap A_i} \right).$$

Hence,

$$\begin{aligned}
\alpha f + \beta g &= \sum_{i=1}^N \alpha\alpha_i \left(\sum_{j=1}^M \mathbb{1}_{A_i\cap B_j} \right) + \sum_{j=1}^M \beta\beta_j \left(\sum_{i=1}^N \mathbb{1}_{B_j\cap A_i} \right) \\
&= \sum_{i=1}^N \alpha\alpha_i \left(\sum_{j=1}^M \mathbb{1}_{A_i\cap B_j} \right) + \sum_{i=1}^N \beta\beta_j \left(\sum_{j=1}^M \mathbb{1}_{A_i\cap B_j} \right) \\
&= \sum_{i=1}^N \left(\sum_{j=1}^M \alpha\alpha_i \mathbb{1}_{A_i\cap B_j} \right) + \sum_{i=1}^N \left(\sum_{j=1}^M \beta\beta_j \mathbb{1}_{A_i\cap B_j} \right) \\
&= \sum_{i=1}^N \left(\sum_{j=1}^M (\alpha\alpha_i + \beta\beta_j) \mathbb{1}_{A_i\cap B_j} \right).
\end{aligned}$$

Assume that $N \geq M$ and set $B_j = \emptyset$ for $j = M+1,\dots,N$ (if $M \geq N$, we set $A_i = \emptyset$ for $i = N+1,\dots,M$). Then, we write

$$\alpha f + \beta g = \sum_{i,j=1}^N \gamma_{i,j} \mathbb{1}_{C_{i,j}},$$

where $\gamma_{i,j} = \alpha\alpha_i + \beta\beta_j$ and $C_{i,j} = A_i \cap B_j$, $i, j = 1, \ldots, N$. Therefore, $\alpha f + \beta g$ is a standard simple function. Using Proposition 8.1, we obtain

$$\begin{aligned}\int_\Omega (\alpha f + \beta g)(\omega)\mu(d\omega) &= \sum_{i,j=1}^{N} \gamma_{i,j}\mu(C_{i,j}) \\ &= \alpha \sum_{i=1}^{N} \alpha_i \mu(A_i) + \beta \sum_{j=1}^{M} \beta_j \mu(B_j) \\ &= \alpha \int_\Omega f(\omega)\mu(d\omega) + \beta \int_\Omega g(\omega)\mu(d\omega).\end{aligned}$$

At this point we remark that α or β equal to ∞, would not change the latter result. For the general case, assume that f and g are nonnegative and $\mathcal{F}$ measurable. Using Proposition 7.10, there exists sequences of nonnegative standard simple functions (f_n) and (g_n) s.t. $f_n(\omega) \uparrow f(\omega)$ and $g_n(\omega) \uparrow g(\omega)$, $\omega \in \Omega$. That is, $(\alpha f_n + \beta g_n)(\omega) \uparrow \alpha f + \beta g$. Then, we rely on Proposition 8.2 and conclude that

$$\begin{aligned}\int_\Omega (\alpha f + \beta g)(\omega)\mu(d\omega) &= \lim_{n\to\infty} \left(\int_\Omega (\alpha f_n + \beta g_n)(\omega)\mu(d\omega) \right) \\ &= \alpha \lim_{n\to\infty} \left(\int_\Omega f_n(\omega)\mu(d\omega) \right) + \beta \lim_{n\to\infty} \left(\int_\Omega g_n(\omega)\mu(d\omega) \right) \\ &= \alpha \int_\Omega f(\omega)\mu(d\omega) + \beta \int_\Omega g(\omega)\mu(d\omega).\end{aligned}$$

□

Proof of Proposition 8.5. Let $A_i = \{\omega\colon f(\omega) \geq 1/i\}$, $i \in \mathbb{N}$. Then, clearly $\cup_{i\in\mathbb{N}} A_i = \{\omega\colon f(\omega) > 0\}$. Using item (v) of Proposition 5.1, $\mu(A_n) \uparrow \mu(\{\omega\colon f(\omega) > 0\})$. Thus, since $\mu(\{\omega\colon f(\omega) > 0\}) > 0$, let $\varepsilon > 0$ s.t. $\varepsilon < \mu(\{\omega\colon f(\omega) > 0\})$. We write $\delta = \mu(\{\omega\colon f(\omega) > 0\}) - \varepsilon$. Then, there exists $N \in \mathbb{N}$ s.t. $\mu(\{\omega\colon f(\omega) > 0\}) - \mu(A_N) \leq \varepsilon$, i.e., $\mu(A_N) \geq \mu(\{\omega\colon f(\omega) > 0\}) - \varepsilon = \delta > 0$. Now consider $\xi = \{A_N, A_N^c\}$. Then, $\xi \in Z_0^{\mathcal{F}}$ and $\int_\Omega f(\omega)\mu(d\omega) \geq S_\mu^f(\xi) \geq \frac{1}{N}\mu(A_N) \geq \delta/N > 0$. This completes the proof of (i). Regarding (ii), suppose by contradiction that $\int_\Omega f(\omega)\mu(d\omega) < \infty$ but $\mu(\{\omega\colon f(\omega) = \infty\}) > 0$, i.e., the negation of $f < \infty$ μ a.e. Then, we consider $\xi = \{A, A^c\}$, where $A = \{\omega\colon f(\omega) = \infty\}$. We have that $S_\mu^f(\xi) \geq \infty \cdot \mu(\{\omega\colon f(\omega) = \infty\}) = \infty$, which gives a contradiction. In order to verify (iii), let $G = \{\omega\colon f(\omega) \leq g(\omega)\}$. Let $\xi = \{A_1, \ldots, A_N\} \in Z_0^{\mathcal{F}}$. Then, we consider $\xi^* = \{A_1 \cap G, \ldots, A_N \cap G\} \cup \{G^c\}$ and have that $\xi^* \in Z_0^{\mathcal{F}}$. In particular,

$$S_\mu^f(\xi) \leq S_\mu^f(\xi^*) \leq S_\mu^g(\xi^*) \leq \int_\Omega g(\omega)\mu(d\omega),$$

where the second inequality follows since by assumption on G, for any $A \in \xi$, $\mu(A) = \mu(A \cap G) + \mu(A \cap G^c) = \mu(A \cap G)$. This shows (iii) and in particular (iv) (then also $\int_\Omega g(\omega)\mu(d\omega) \leq \int_\Omega f(\omega)\mu(d\omega)$). □

Proof of Proposition 8.7. We notice that

$$\{\omega\colon f^+(\omega) \leq g^+(\omega)\} \cap \{\omega\colon f^-(\omega) \geq g^-(\omega)\} = \{\omega\colon f(\omega) \leq g(\omega)\}.$$

This follows from the definition of f^+ and f^{-1}, since if ω is s.t. $f(\omega) \leq g(\omega)$, then, $\max\{f(\omega), 0\} \leq \max\{g(\omega), 0\}$ and $\max\{-f(\omega), 0\} \geq \max\{-g(\omega), 0\}$. We obtain

$$\{\omega\colon f^+(\omega) \leq g^+(\omega)\}^c \cup \{\omega\colon f^-(\omega) \geq g^-(\omega)\}^c = \{\omega\colon f(\omega) \leq g(\omega)\}^c$$

Hence, if $f \leq g$ a.e., then, $f^+ \leq g^+$ a.e. and $f^- \geq g^-$ a.e. Therefore, using (iii) in Proposition 8.5, $\int_\Omega f^+(\omega)\mu(d\omega) \leq \int_\Omega g^+(\omega)\mu(d\omega)$ and $\int_\Omega f^-(\omega)\mu(d\omega) \geq \int_\Omega g^-(\omega)\mu(d\omega)$, which makes

$$\begin{aligned}\int_\Omega f(\omega)\mu(d\omega) &= \int_\Omega f^+(\omega)\mu(d\omega) - \int_\Omega f^-(\omega)\mu(d\omega)\\ &\leq \int_\Omega g^+(\omega)\mu(d\omega) - \int_\Omega g^-(\omega)\mu(d\omega) = \int_\Omega g(\omega)\mu(d\omega).\end{aligned}$$

□

We show that the integral for integrable functions is linear.

Proof of Proposition 8.8. Notice first that for any $\omega \in \Omega$,

$$|(\alpha f + \beta g)(\omega)| \leq |\alpha||f(\omega)| + |\beta||g(\omega)|,$$

by the triangular inequality. Thus, since $|\alpha|\int_\Omega |f(\omega)|\mu(d\omega)$ and $|\beta|\int_\Omega |g(\omega)|\mu(d\omega)$ are finite upon the integrability of f and g (cf. Proposition 8.6), $\alpha f + \beta g$ is integrable as well (cf. Exercise 8.1 and Proposition 8.3). Notice first that $\int_\Omega \alpha f(\omega)\mu(d\omega) = \alpha\int_\Omega f(\omega)\mu(d\omega)$. We can see it by considering the cases: $\alpha > 0$, $\alpha < 0$ and $\alpha = 0$. If $\alpha = 0$, then, for any $\omega \in \Omega$, $(\alpha f)^+(\omega) = (\alpha f)^-(\omega) = 0 = \alpha f^-(\omega) = \alpha f^+(\omega)$. If $\alpha > 0$, then,

$$(\alpha f)^+(\omega) = \begin{cases}\alpha f(\omega), & \text{if } \alpha f(\omega) \geq 0,\\ 0, & \text{if } \alpha f(\omega) < 0,\end{cases} = \alpha\begin{cases} f(\omega), & \text{if } f(\omega) \geq 0,\\ 0, & \text{if } f(\omega) < 0,\end{cases} = \alpha f^+(\omega),$$

and

$$(\alpha f)^-(\omega) = \begin{cases}-\alpha f(\omega), & \text{if } -\alpha f(\omega) \geq 0,\\ 0, & \text{if } -\alpha f(\omega) < 0,\end{cases} = \alpha\begin{cases} -f(\omega), & \text{if } -f(\omega) \geq 0,\\ 0, & \text{if } -f(\omega) < 0,\end{cases} = \alpha f^-(\omega),$$

Similarly, if $\alpha < 0$, then $(\alpha f)^+(\omega) = -\alpha f^-(\omega)$ and $(\alpha f)^-(\omega) = -\alpha f^+(\omega)$. Hence, in each case, we use Proposition 8.3 and obtain

$$\begin{aligned}\int_\Omega \alpha f(\omega)\mu(d\omega) &= \int_\Omega (\alpha f)^+(\omega)\mu(d\omega) - \int_\Omega (\alpha f)^-(\omega)\mu(d\omega)\\ &= \alpha\int_\Omega f^+(\omega)\mu(d\omega) - \alpha\int_\Omega f^-(\omega)\mu(d\omega) = \alpha\int_\Omega f(\omega)\mu(d\omega).\end{aligned}$$

Of course the same argument applies for $\int_\Omega \beta g(\omega)\mu(d\omega)$. Hence, if we show that

$$\int_\Omega (\alpha f + \beta g)(\omega)\mu(d\omega) = \int_\Omega \alpha f(\omega)\mu(d\omega) + \int_\Omega \beta g(\omega)\mu(d\omega), \tag{61}$$

we are done. Write $f_* = \alpha f$ and $g_* = \beta g$. We know that $(f_* + g_*)^+ - (f_* + g_*)^- = f_* + g_*$ (cf. Exercise 7.5). Hence, $(f_* + g_*)^+ - (f_* + g_*)^- = f_*^+ - f_*^- + g_*^+ - g_*^-$. This shows that $(f_* + g_*)^+ + f_*^- + g_*^- = (f_* + g_*)^- + f_*^+ + g_*^+$. Hence,

$$\int_\Omega \big((f_* + g_*)^+ + f_*^- + g_*^-\big)(\omega)\mu(d\omega) = \int_\Omega \big((f_* + g_*)^- + f_*^+ + g_*^+\big)(\omega)\mu(d\omega),$$

which gives (cf. Proposition 8.3)

$$\begin{aligned}&\int_\Omega (f_* + g_*)^+(\omega)\mu(d\omega) - \int_\Omega (f_* + g_*)^-(\omega)\mu(d\omega)\\ &= \int_\Omega f_*^+(\omega)\mu(d\omega) - \int_\Omega f_*^-(\omega)\mu(d\omega) + \int_\Omega g_*^+(\omega)\mu(d\omega) - \int_\Omega g_*^-(\omega)\mu(d\omega)\\ &= \int_\Omega f(\omega)\mu(d\omega) + \int_\Omega g(\omega)\mu(d\omega).\end{aligned}$$

This shows (61) and the proof of the proposition is complete. □

We prove Fatou's lemma and Lebesgue's dominated convergence theorem.

Proof of Proposition 8.9. Define $g_n = \inf_{k\geq n} f_k$, then $(g_n)_{n\in\mathbb{N}}$ is a sequence of $\mathcal{F}$ measurable and nonnegative functions s.t. for any $n \in \mathbb{N}$, $\int_\Omega g_n(\omega)\mu(d\omega) \leq \int_\Omega f_n(\omega)\mu(d\omega)$ (cf. Exercise 8.1), Therefore,

$$\liminf_{n\to\infty} \int_\Omega g_n(\omega)\mu(d\omega) \leq \liminf_{n\to\infty} \int_\Omega f_n(\omega)\mu(d\omega).$$

Further, for any $\omega \in \Omega$, $g_n(\omega) \uparrow \liminf_{n\to\infty} f_n(\omega)$ (cf. Proposition A.5). Therefore, using Proposition 8.2, $\int_\Omega g_n(\omega)\mu(d\omega) \uparrow \int_\Omega \liminf_{n\to\infty} f_n(\omega)\mu(d\omega)$. In conclusion, using Proposition 3.23, we obtain

$$\int_\Omega \liminf_{n\to\infty} f_n(\omega)\mu(d\omega) = \liminf_{n\to\infty} \int_\Omega g_n(\omega)\mu(d\omega) \leq \liminf_{n\to\infty} \int_\Omega f_n(\omega)\mu(d\omega).$$

□

Proof of Proposition 8.10. Notice that f is integrable, i.e., $\int_\Omega |f(\omega)|\mu(d\omega) < \infty$. To see it, we notice that (compare also to Proposition 3.26),

$$A = \underbrace{\{\omega\colon \lim_{n\to\infty} f_n(\omega) = f(\omega)\}}_{=A_1} \cap \left(\bigcap_{n\in\mathbb{N}} \underbrace{\{\omega\colon |f_n(\omega)| \leq g(\omega)\}}_{=A_{2,n}} \right) \subset \{\omega\colon f(\omega) \leq g(\omega)\}.$$

Then, upon Proposition 5.1, since $\mu(A_1^c) = 0$ and $\mu(A_{2,n}^c) = 0$ for any $n \in \mathbb{N}$,

$$\mu(A^c) \leq \sum_{n\in\mathbb{N}} \big(\mu(A_1^c) + \mu(A_{2,n}^c)\big) = 0.$$

Thus, $f \leq g$ a.e. and hence f is integrable by Exercise 8.4. Similarly $f^* = \limsup_{n\to\infty} f_n$ and $f_* = \liminf_{n\to\infty} f_n$ are integrable. Then, we notice that $(g + f_n)_{n\in\mathbb{N}}$ and $(g - f_n)_{n\in\mathbb{N}}$ are nonnegative and $\mathcal{F}$ measurable sequences of functions. Given any $\omega \in A$, we rely on Proposition A.4 and obtain

$$(g + f_*)(\omega) = g(\omega) + \liminf_{n\to\infty} f_n(\omega) = \liminf_{n\to\infty} (g + f_n)(\omega),$$

and

$$(g - f^*)(\omega) = \liminf_{n\to\infty} (g - f_n)(\omega).$$

Hence, $g + f_* = \liminf_{n\to\infty}(g + f_n)$ a.e. and $g - f^* = \liminf_{n\to\infty}(g - f_n)$ a.e. Therefore, we use Proposition 8.9 and obtain

$$\begin{aligned}
\int_\Omega g(\omega)\mu(d\omega) + \int_\Omega f_*(\omega)\mu(d\omega) &= \int_\Omega (g + f_*)(\omega)\mu(d\omega) \\
&= \int_\Omega \liminf_{n\to\infty} (g + f_n)(\omega)\mu(d\omega) \\
&\leq \liminf_{n\to\infty} \int_\Omega (g + f_n)(\omega)\mu(d\omega) \\
&= \int_\Omega g(\omega)\mu(d\omega) + \liminf_{n\to\infty} \int_\Omega f_n(\omega)\mu(d\omega),
\end{aligned}$$

where the last equality again follows from Proposition A.4. Similarly, we obtain

$$\begin{aligned}
\int_\Omega g(\omega)\mu(d\omega) - \int_\Omega f^*(\omega)\mu(d\omega) &= \int_\Omega \liminf_{n\to\infty} (g - f_n)(\omega)\mu(d\omega) \\
&\leq \liminf_{n\to\infty} \int_\Omega (g - f_n)(\omega)\mu(d\omega) \\
&= \int_\Omega g(\omega)\mu(d\omega) - \limsup_{n\to\infty} \int_\Omega f_n(\omega)\mu(d\omega).
\end{aligned}$$

This makes (cf. Proposition 3.20)

$$\int_\Omega \liminf_{n\to\infty} f_n(\omega)\mu(d\omega) \le \liminf_{n\to\infty} \int_\Omega f_n(\omega)\mu(d\omega)$$
$$\le \limsup_{n\to\infty} \int_\Omega f_n(\omega)\mu(d\omega) \le \int_\Omega \limsup_{n\to\infty} f_n(\omega)\mu(d\omega).$$

Since $f_n \to f$ a.e. it follows that $\liminf_{n\to\infty} f_n = \limsup_{n\to\infty} f_n = f$ a.e. and hence upon the previous display,

$$\liminf_{n\to\infty} \int_\Omega f_n(\omega)\mu(d\omega) = \limsup_{n\to\infty} \int_\Omega f_n(\omega)\mu(d\omega) = \int_\Omega f(\omega)\mu(d\omega),$$

and therefore (cf. Proposition 3.23),

$$\lim_{n\to\infty} \int_\Omega f_n(\omega)\mu(d\omega) = \int_\Omega f(\omega)\mu(d\omega).$$

□

In order to verify that integration can be interchanged with differentiation, Lebesgue's dominated convergence theorem is a key tool.

Proposition B.5. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f\colon (a,b)\times\Omega \to \mathbb{R}$ be a function where $(a,b) \subset \mathbb{R}$ is an open interval. Let $u_0 \in (a,b)$. Suppose that*

(i) for any $u \in (a,b)$ $\omega \mapsto f(u,\omega)$ is $\mathcal{F}$ measurable and integrable with respect to μ;

(ii) There exists a set $F \in \mathcal{F}$ s.t. $\mu(F^c) = 0$ and for any $\omega \in F$, the map $u \mapsto f(u,\omega)$ is differentiable in u_0 with derivative $(\partial f/\partial u)(u_0,\omega)$ (that is, $u \mapsto f(u,\omega)$ is differentiable in u_0 μ a.e.);

(iii) there exists a $\mathcal{F}$ measurable and nonnegative function $g\colon \Omega \to \mathbb{R}$ s.t. $\int_\Omega g(\omega)\mu(d\omega) < \infty$ and for any $u \in (a,b)$ $(u \neq u_0)$ we have that μ a.e.,

$$|f(u,\omega) - f(u_0,\omega)| \le g(\omega)|u - u_0|.$$

Then, the function $F(u) = \int_\Omega f(u,\omega)\mu(d\omega)$ is differentiable in u_0 with derivative

$$F'(u_0) = \int_\Omega (\partial f/\partial u)(u_0,\omega)\mu(d\omega),$$

i.e., we are allowed to interchange integration with differentiation.

Proof. Let $(u_n)_{n\in\mathbb{N}}$ be a sequence in $(a,b)\setminus\{u_0\}$ s.t. $u_n \xrightarrow{n\to\infty} u_0$ and define

$$\varphi_n(u_0,\omega) = \frac{f(u_n,\omega) - f(u_0,\omega)}{u_n - u_0}, \quad \omega \in \Omega.$$

Let f_{n,u_0} and f_{u_0} be the map $\omega \mapsto \varphi_n(u_0,\omega)$ and $\omega \mapsto (\partial f/\partial u)(u_0,\omega)\mathbb{1}_F(\omega)$, respectively. By item (ii), we have that μ a.e., $f_{n,u_0} \xrightarrow{n\to\infty} f_{u_0}$. Further, by item (iii), for any $n \in \mathbb{N}$,

$$|f_{n,u_0}| \le g(\omega), \quad \mu \text{ a.e.},$$

and $\int_\Omega |g(\omega)|\mu(d\omega) < \infty$. Hence, we are in position to apply Proposition 8.10, and deduce that

$$\lim_{n\to\infty} \frac{F(u_n) - F(u_0)}{u_n - u_0} = \lim_{n\to\infty} \int_\Omega f_{n,u_0}(\omega)\mu(d\omega) = \int_\Omega (\partial f/\partial u)(u_0,\omega)\mu(d\omega).$$

□

We remark that stronger conditions as given in Proposition B.5 can be of interest for practical applications.

Proposition B.6. *Let $(\Omega, \mathcal{F}, \mu)$ and $f\colon (a,b) \times \Omega \to \mathbb{R}$ be as in Proposition B.5. Suppose that*

(i) for any $u \in (a,b)$ $\omega \mapsto f(u,\omega)$ is $\mathcal{F}$ measurable and integrable with respect to μ;

(ii) for any $\omega \in \Omega$, the map $f_\omega(u) = f(u,\omega)$, $u \in (a,b)$ is differentiable (cf. Definition A.9) with derivative f'_ω that satisfies

$$|f'_\omega(u_0)| \leq g(\omega), \quad u_0 \in (a,b),$$

where $g\colon \Omega \to \mathbb{R}$ is nonnegative and $\mathcal{F}$ measurable and s.t. $\int_\Omega g(\omega)\mu(d\omega) < \infty$.

Then, the function $F(u) = \int_\Omega f(u,\omega)\mu(d\omega)$, $u \in (a,b)$, is differentiable with derivative

$$F'(u_0) = \int_\Omega (\partial f/\partial u)(u_0,\omega)\mu(d\omega), \quad u_0 \in (a,b).$$

Proof. Clearly, item (ii) of Proposition B.5 is s satisfied. We show that also (iii) holds. It is a consequence of the mean value theorem (cf. Proposition A.21). Let $u_0 \in (a,b)$. Then, for any $u \in (a,b)$ $(u \neq u_0)$, either $u > u_0$ or $u < u_0$. If $u > u_0$, we apply the mean value theorem and obtain $m_1 \in (u_0, u)$ s.t. for any $\omega \in \Omega$,

$$f_\omega(u) - f_\omega(u_0) = f'_\omega(m_1)(u - u_0).$$

If $u < u_0$, we obtain $m_2 \in (u, u_0)$ s.t.

$$f_\omega(u_0) - f_\omega(u) = f'_\omega(m_2)(u_0 - u).$$

In particular, for any $\omega \in \Omega$,

$$|f_\omega(u) - f_\omega(u_0)| \leq \max\{f'_\omega(m_1), f'_\omega(m_2)\}|u - u_0| \leq g(\omega)|u - u_0|,$$

and hence (iii) of Proposition B.5 is satisfied. This completes the proof. $\square$

B.6 On the Riemann and the Lebesgue integral

Proof of Proposition 9.4. Since f is Riemann integrable, let $\varepsilon > 0$ and choose $\delta > 0$ s.t. for any partition $a = a_1 < a_2 < \cdots < a_{N+1} = b$ of $[a,b]$ with $\lambda([a_i, a_{i+1}]) < \delta$ and $x_i \in [a_i, a_{i+1}]$, $i = 1, \ldots, N$, we obtain

$$\left| I_f(a,b) - \sum_{i=1}^{N} f(x_i)\lambda([a_i, a_{i+1}]) \right| < \varepsilon. \tag{62}$$

In particular,

$$I_f(a,b) - \varepsilon < \sum_{i=1}^{N} f(x_i)\lambda([a_i, a_{i+1}]) < I_f(a,b) + \varepsilon, \tag{63}$$

where $x_1, \ldots, x_N$ is any collection of points s.t. $x_i \in [a_i, b_i]$. Let

$$A_i = \left\{ f(x)\lambda([a_i, a_{i+1}])\colon x \in [a_i, b_i] \right\}, \quad i = 1, \ldots, N.$$

By (62),

$$I_f(a,b) - \varepsilon < \sup(A_1 + \cdots + A_N) < I_f(a,b) + \varepsilon.$$

Then, we rely on Proposition A.3 and conclude that for any $i = 1, \dots, N$, $\sup A_i < \infty$ and

$$\sup(A_1 + \cdots + A_N) = \sup A_1 + \cdots + \sup A_N.$$

In particular, upon Proposition A.2, for any $i = 1, \dots, N$, $\sup_{x\in[a_i,b_i]} f(x) < \infty$. Therefore, since $a = a_1 < a_2 < \cdots < a_{N+1} = b$ is a partition of $[a, b]$, $\sup_{x\in[a,b]} f(x) < M$ for some $M \in \mathbb{R}$. Hence, we have shown that (62) implies that

$$\Big| I_f(a, b) - \sum_{i=1}^{N} M_i \lambda([a_i, a_{i+1}]) \Big| < \varepsilon, \tag{64}$$

where $M_i = \sup_{x\in[a_i,b_i]} f(x)$, $i = 1, \dots, N$. Upon the same reasoning, we find $m \in \mathbb{R}$, s.t. $\inf_{x\in[a,b]} f(x) > m$ and deduce that (62) also implies that

$$\Big| I_f(a, b) - \sum_{i=1}^{N} m_i \lambda([a_i, a_{i+1}]) \Big| < \varepsilon, \tag{65}$$

where $m_i = \inf_{x\in[a_i,b_i]} f(x)$, $i = 1, \dots, N$. In particular, for any $x \in [a, b]$, $|f(x)| < \max\{-m, M\}$. Thus, since $\lambda([a, b]) < \infty$, it follows from (b) of Exercise 8.4, that $f\colon [a, b] \to \mathbb{R}$ is Lebesgue integrable. We define the simple functions

$$g(x) = \sum_{i=1}^{N} M_i \mathbb{1}_{[a_i,b_i]}(x), \quad x \in [a, b],$$

and

$$h(x) = \sum_{i=1}^{N} m_i \mathbb{1}_{[a_i,b_i]}(x), \quad x \in [a, b].$$

We have that, $\int_a^b g(x)dx = \sum_{i=1}^{N} M_i \lambda([a_i, a_{i+1}])$ and $\int_a^b h(x)dx = \sum_{i=1}^{N} m_i \lambda([a_i, a_{i+1}])$ (cf. Exercise 8.5). Further for any $x \in [a, b]$, $h(x) \le f(x) \le g(x)$. Hence, by (64) and (65),

$$I_f(a, b) - \varepsilon \le \int_a^b h(x)dx \le \int_a^b f(x)dx \le \int_a^b g(x)dx \le I_f(a, b) + \varepsilon.$$

□

Proof of Proposition 9.5. Let $(b_n)_{n\in\mathbb{N}}$ be any sequence which is s.t. $b_n \uparrow u$ and $a < b_n < u$ for any $n \in \mathbb{N}$. We define the function $x \mapsto f_n(x) = \mathbb{1}_{[a,b_n)}(x)f(x)$, $n \in \mathbb{N}$. It follows that for any $x \in [a, u)$, $f_n(x) \uparrow f(x)$. Since f is nonnegative, we apply Proposition 8.2 and conclude that

$$\lim_{n\to\infty} \int_a^{b_n} f(x) = \int_a^u f(x)dx.$$

Then, upon Proposition 9.4, for any $a < b < u$, $I_f(a, b) = \int_a^b f(x)dx$. Hence, the assumption $\lim_{b\uparrow u} I_f(a, b) = I_f(a, u)$ becomes equivalent to the assumption that $\lim_{b\uparrow u} \int_a^b f(x)dx = I_f(a, u)$. Hence, by Proposition A.16 we obtain that $\int_a^u f(x)dx = I_f(a, u)$. □

B.7 Fubini-Tonnelli-Lebesgue

Proposition B.7. *Let $(X_i, \mathscr{X}_i)$, $i = 1, \dots, n$, be n measurable spaces and $f_i\colon X_i \to \overline{\mathbb{R}}$ be $\mathscr{X}_i$ measurable for any $i = 1, \dots, n$.*

(i) The map $f\colon X_1 \times \cdots \times X_n \to \overline{\mathbb{R}}$ *given by*

$$f(x_1, \ldots, x_n) = \prod_{i=1}^n f(x_i),$$

is $\mathscr{X}_1 \otimes \cdots \otimes \mathscr{X}_n$ *measurable;*

(ii) if for any $i = 1, \ldots, n$, $g_i = (g_i^1, \ldots, g_i^k)\colon X_i \to \mathbb{R}^k$ *is* $\mathscr{X}_i$ *measurable, then, the map*

$$s(x_1, \ldots, x_n) = \sum_{i=1}^n g_i(x_i) = \left(\sum_{i=1}^n g_i^1(x_i), \cdots, \sum_{i=1}^n g_i^k(x_i)\right),$$

is $\mathscr{X}_1 \otimes \cdots \otimes \mathscr{X}_n$ *measurable.*

Proof. Let $i = 1, \ldots, n$, and define the map $\tilde{f}_i\colon X_1 \times \cdots \times X_n \to \overline{\mathbb{R}}$ by

$$\tilde{f}_i(x_1, \ldots, x_n) = f_i(x_i), \quad (x_1, \ldots, x_n) \in X_1 \times \cdots \times X_n.$$

Then, if either $B \in \mathfrak{B}(\mathbb{R})$, $B = \{\infty\}$ or $B = \{-\infty\}$, we have that

$$\tilde{f}_i^{-1}(B) = X_1 \times \cdots \times \underbrace{f_i^{-1}(B)}_{\text{position } i} \times \cdots \times X_n \in \mathscr{X}_1 \otimes \cdots \otimes \mathscr{X}_n,$$

since by assumption $f_i\colon X_i \to \overline{\mathbb{R}}$ is $\mathscr{X}_i$ measurable. This shows that for any $i = 1, \ldots, n$, the map $\tilde{f}_i\colon X_1 \times \cdots \times X_n \to \overline{\mathbb{R}}$ is $\mathscr{X}_1 \otimes \cdots \otimes \mathscr{X}_n$ measurable. By Proposition 7.12 the map f is $\mathscr{X}_1 \otimes \cdots \otimes \mathscr{X}_n$ measurable as well. This shows item (i). The proof for item (ii) follows the same reasoning (recall also Proposition 7.4). □

Proposition B.8. *Let* $(X, \mathscr{X})$ *and* $(Y, \mathscr{Y})$ *be two measurable spaces. If* $E \in \mathscr{X} \otimes \mathscr{Y}$, *then for any* $x \in X$, $\{y \in Y\colon (x, y) \in E\} \in \mathscr{Y}$ *and for any* $y \in Y$, $\{x \in X\colon (x, y) \in E\} \in \mathscr{X}$.

Proof. Let $x \in X$, and define the map $y \mapsto g_x(y) = (x, y)$. Suppose that $E = A \times B$, $A \in \mathscr{X}$, $B \in \mathscr{Y}$. Then,

$$g_x^{-1}(E) = \begin{cases} B, & \text{if } x \in A, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Hence, in this case $g_x^{-1}(E) \in \mathscr{Y}$. Since $\mathscr{X} \otimes \mathscr{Y}$ is generated by the sets $A \times B$, $A \in \mathscr{X}$, $B \in \mathscr{Y}$, g is $\mathscr{Y}/(\mathscr{X} \otimes \mathscr{Y})$ measurable according to Proposition 7.1. Therefore, $\{y \in Y\colon (x, y) \in E\} = g_x^{-1}(E) \in \mathscr{Y}$ for any $E \in \mathscr{X} \otimes \mathscr{Y}$. We adapt the same reasoning to show that $\{x \in X\colon (x, y) \in E\} \in \mathscr{X}$ for any $E \in \mathscr{X} \otimes \mathscr{Y}$. □

Proposition B.9. *Let* $(X, \mathscr{X})$ *and* $(Y, \mathscr{Y})$ *be two measurable spaces. Let* $f\colon X \times Y \to \overline{\mathbb{R}}$ *and define for any* $x \in X$, $y \mapsto f_x(y) = f(x, y)$ *and for any* $y \in Y$, $x \mapsto f_y(x) = f(x, y)$. *Then, if* f *is* $\mathscr{X} \otimes \mathscr{Y}$ *measurable,* f_x *is* $\mathscr{Y}$ *measurable for any* $x \in X$ *and* f_y *is* $\mathscr{X}$ *measurable for any* $y \in Y$.

Proof. Define $y \mapsto g_x(y) = (x, y)$ as in the proof of the latter proposition. By assumption, $f\colon X \times Y \to \overline{\mathbb{R}}$ is $\mathscr{X} \otimes \mathscr{Y}$ measurable and by the latter proof, g_x is $\mathscr{Y}/(\mathscr{X} \otimes \mathscr{Y})$ measurable. This shows that for any $x \in X$, $f_x(y) = f(g_x(y)) = f(x, y)$ is $\mathscr{Y}$ measurable (cf. Exercise 7.2). Similarly, we show that for any $y \in Y$, f_y is $\mathscr{X}$ measurable. □

Proposition B.10. *Let* $(X, \mathscr{X})$ *and* $(Y, \mathscr{Y})$ *be two measurable spaces. Then, the collection*

$$\mathscr{P} = \{A \times B\colon A \in \mathscr{X}, B \in \mathscr{Y}\},$$

is a π*-system.*

Proof. Let $E, F \in \mathscr{P}$, i.e., $E = A \times B$, and $F = C \times D$, $A, C \in \mathscr{X}$, $B, D \in \mathscr{Y}$. We readily see that $E \cap F = A \cap C \times B \cap D$. Therefore, $E \cap F \in \mathscr{P}$. $\square$

Proof of Proposition 9.10. Suppose first that μ and ν are finite measures on $\mathscr{X}$ and $\mathscr{Y}$, respectively. That is, $\mu(X) < \infty$ and $\nu(Y) < \infty$. Define the collection

$$\mathscr{L} = \{E \in \mathscr{X} \otimes \mathscr{Y} \colon X \ni x \mapsto \nu(\{y \in Y \colon (x, y) \in E\}) \text{ is } \mathscr{X} \text{ measurable}\}.$$

We show that $\mathscr{L}$ is a λ-system on $X \times Y$. In order to simplify the notation, we write $f_E(x) = \nu(\{y \in Y \colon (x, y) \in E\})$, $E \in \mathscr{X} \otimes \mathscr{Y}$. Notice that for any $E \in \mathscr{X} \otimes \mathscr{Y}$, $f_E \colon X \to \mathbb{R}$, since ν is assumed to be finite. If $E = X \times Y$, then we observe that for any $x \in X$, $f_{X \times Y}(x) = \nu(Y)$. Thus, $x \mapsto f_{X \times Y}(x)$ is $\mathscr{X}$ measurable since it is constant. Hence, $X \times Y \in \mathscr{L}$. Suppose that $E \in \mathscr{L}$. Then, f_E is $\mathscr{X}$ measurable. We observe that for any $A \in \mathfrak{B}(\mathbb{R})$, since ν is finite,

$$\begin{aligned} f_{E^c}^{-1}(A) &= \{x \in X \colon \nu(\{y \in Y \colon (x, y) \in E^c\}) \in A\} \\ &= \{x \in X \colon \nu(\{y \in Y \colon (x, y) \in E\}^c) \in A\} \\ &= \{x \in X \colon \nu(Y) - \nu(\{y \in Y \colon (x, y) \in E\}) \in A\} \\ &= \{x \in X \colon \nu(Y) - f_E(x) \in A\} \in \mathscr{X}, \end{aligned}$$

since $\nu(Y) - f_E$ is $\mathscr{X}$ measurable. This shows that f_{E^c} is $\mathscr{X}$ measurable and hence $E^c \in \mathscr{L}$. Let $\{E_i \colon i \in \mathbb{N}\} \subset \mathscr{L}$ disjoint. We have that for any $A \in \mathfrak{B}(\mathbb{R})$,

$$\begin{aligned} f_{\cup_{i \in \mathbb{N}} E_i}^{-1}(A) &= \{x \in X \colon \nu(\{y \in Y \colon (x, y) \in \cup_{i \in \mathbb{N}} E_i\}) \in A\} \\ &= \{x \in X \colon \nu(\cup_{i \in \mathbb{N}} \{y \in Y \colon (x, y) \in E_i\}) \in A\} \\ &= \{x \in X \colon \sum_{i \in \mathbb{N}} \nu(\{y \in Y \colon (x, y) \in E_i\}) \in A\} \\ &= \{x \in X \colon \sum_{i \in \mathbb{N}} f_{E_i}(x) \in A\} \in \mathscr{X}, \end{aligned}$$

since $\sum_{i \in \mathbb{N}} f_{E_i} = \lim_{n \to \infty} \sum_{i=1}^n f_{E_i}$ is $\mathscr{X}$ measurable (cf. Proposition 7.9). Let

$$\mathscr{P} = \{A \times B \colon A \in \mathscr{X}, B \in \mathscr{Y}\}.$$

We know that $\mathscr{P}$ is a π-system (cf. Proposition B.10). Further, given $E = A \times B \in \mathscr{P}$,

$$f_{A \times B}(x) = \nu(\{y \in Y \colon (x, y) \in A \times B\}) = \mathbb{1}_A(x)\nu(B),$$

and hence $f_{A \times B}$ is $\mathscr{X}$ measurable. Thus, $\mathscr{P} \subset \mathscr{L}$. Therefore, $\mathscr{L}$ is a λ-system on $X \times Y$ which contains the π-system $\mathscr{P}$. Using Proposition B.4, $\sigma(\mathscr{P}) = \mathscr{X} \otimes \mathscr{Y} \subset \mathscr{L}$. Clearly, by definition of $\mathscr{L}$, $\mathscr{L} \subset \mathscr{X} \otimes \mathscr{Y}$ and hence, $\mathscr{L} = \mathscr{X} \otimes \mathscr{Y}$. Therefore, given any $E \in \mathscr{X} \otimes \mathscr{Y}$, f_E is $\mathscr{X}$ measurable and we define

$$\pi_1(E) = \int_X f_E(x)\mu(dx). \tag{66}$$

We readily check that π_1 is a measure on $\mathscr{X} \otimes \mathscr{Y}$. Further, $\pi_1(X \times Y) = \mu(X)\nu(Y) < \infty$, i.e., π_1 is finite. We repeat the previous arguments with the function $g_E(y) = \mu(\{x \in X \colon (x, y) \in E\})$, and obtain a finite measure π_2:

$$\pi_2(E) = \int_Y g_E(y)\nu(dy). \tag{67}$$

Then, for any $E = A \times B \in \mathscr{P}$, $\pi_1(A \times B) = \pi_2(A \times B) = \mu(A)\nu(B)$. If we now set

$$\mathscr{L}_* = \{E \in \mathscr{X} \otimes \mathscr{Y} \colon \pi_1(E) = \pi_2(E)\},$$

then, we verify, as in the proof of Proposition 6.9, that $\mathscr{L}_*$ is a λ-system. In conclusion, since $\mathscr{P} \subset \mathscr{L}_*$, we obtain with Proposition B.4, $\mathscr{X} \otimes \mathscr{Y} \subset \mathscr{L}_*$. Thus, π_1 and π_2 agree on $\mathscr{X} \otimes \mathscr{Y}$ and we define $\mu \otimes \nu = \pi_1$ on $\mathscr{X} \otimes \mathscr{Y}$. Finally, we notice that by Proposition 6.9, $\mu \otimes \nu$ is the only measure on $\mathscr{X} \otimes \mathscr{Y}$ which is s.t. $\mu \otimes \nu(A \times B) = \mu(A)\nu(B)$ for any $A \times B \in \mathscr{P}$. Clearly, $\mu \otimes \nu$ is finite since $X \times Y \in \mathscr{P}$ and μ and ν are assumed to be finite. For the remaining case, assume that μ and ν are not necessarily finite but σ-finite on $\mathscr{X} \otimes \mathscr{Y}$. Since ν is σ-finite on $\mathscr{Y}$, there exists $\{B_n \colon n \in \mathbb{N}\} \subset \mathscr{Y}$ s.t. $\nu(B_n) < \infty$ for any $n \in \mathbb{N}$ and $Y = \cup_{n\in\mathbb{N}} B_n$. As in the proof of item (vii) of Proposition 5.1, we define $C_n = B_n \setminus (\cup_{k=1}^{n-1} B_k)$, $n \in \mathbb{N}$, and obtain a disjoint collection $\{C_n \colon n \in \mathbb{N}\} \subset \mathscr{Y}$ s.t. $\nu(C_n) < \infty$ for any $n \in \mathbb{N}$ and $Y = \cup_{n\in\mathbb{N}} C_n$. Define $\nu_n(B) = \nu(B \cap C_n)$, $B \in \mathscr{Y}$. Then, for any $n \in \mathbb{N}$, ν_n is a finite measure on $\mathscr{Y}$. Thus, we rely on the finite case and deduce that for any $E \in \mathscr{X} \otimes \mathscr{Y}$, $x \mapsto f_E^n(x) = \nu_n(\{y \in Y \colon (x, y) \in E\})$ is $\mathscr{X}$ measurable for any $n \in \mathbb{N}$. We adapt the same strategy for the measure μ and find $\{A_m \colon m \in \mathbb{N}\} \subset \mathscr{X}$ disjoint s.t. $\cup_{m\in\mathbb{N}} A_m = X$ and $\mu(A_m) < \infty$ for any $m \in \mathbb{N}$. Then, via $\mu_m(A) = \mu(A \cap A_m)$, $A \in \mathscr{X}$, we obtain with $y \mapsto g_E^m(y) = \mu_m(\{x \in X \colon (x, y) \in E\})$, $E \in \mathscr{X} \otimes \mathscr{Y}$, that $(g_E^m)_{m\in\mathbb{N}}$ is a sequence of $\mathscr{Y}$ measurable functions for any $E \in \mathscr{X} \otimes \mathscr{Y}$. We notice that since for any $B \in \mathscr{Y}$, $\sum_{n\in\mathbb{N}} \nu_n(B) = \nu(B)$, it follows that for any $x \in X$ and $E \in \mathscr{X} \otimes \mathscr{Y}$,

$$\lim_{n\to\infty} \sum_{i=1}^{n} f_E^n(x) = \sum_{n\in\mathbb{N}} f_E^n(x) = \nu(\{y \in Y \colon (x, y) \in E\}),$$

which is $\mathscr{X}$ measurable by Proposition 7.9. Hence, as in the previous case,

$$\tilde{\pi}_1(E) = \int_X \nu(\{y \in Y \colon (x, y) \in E\})\mu(dx), \quad E \in \mathscr{X} \otimes \mathscr{Y},$$

is a well defined measure on $\mathscr{X} \otimes \mathscr{Y}$. Notice that in this case, $\tilde{\pi}_1$ is σ-finite on $\mathscr{X} \otimes \mathscr{Y}$ since $\cup_{m,n\in\mathbb{N}}(A_m \times C_n) = X \times Y$ and $\tilde{\pi}_1(A_m \times C_n) = \mu(A_m)\nu(C_n) < \infty$ for any $m, n \in \mathbb{N}$. Similarly, we deduce that

$$\tilde{\pi}_2(E) = \int_Y \mu(\{x \in X \colon (x, y) \in E\})\nu(dy), \quad E \in \mathscr{X} \otimes \mathscr{Y},$$

is a σ-finite measure on $\mathscr{X} \otimes \mathscr{Y}$. We show that $\tilde{\pi}_1(E) = \tilde{\pi}_2(E)$ for any $E \in \mathscr{X} \otimes \mathscr{Y}$. Define

$$\pi_1^{nm}(E) = \int_X f_E^n(x)\mu_m(dx), \quad E \in \mathscr{X} \otimes \mathscr{Y},$$

and

$$\pi_2^{mn}(E) = \int_Y g_E^m(y)\nu_n(dy), \quad E \in \mathscr{X} \otimes \mathscr{Y},$$

We observe that

$$\pi_1^{nm}(A \times B) = \int_X \mathbb{1}_{A_m}(x)\mathbb{1}_A(x)\nu_n(B)\mu(dx) = \mu_m(A)\nu_n(B) = \pi_2^{nm}(A \times B), \tag{68}$$

for any $A \times B$ be s.t. $A \in \mathscr{X}$ and $B \in \mathscr{Y}$ (cf. Exercise 8.8). Hence, as in the finite case, since $\mathscr{P}$ is a π-system, we deduce that for any $E \in \mathscr{X} \otimes \mathscr{Y}$, $\pi_1^{mn}(E) = \pi_2^{mn}(E)$, $m, n \in \mathbb{N}$. We also observe that for any $E \in \mathscr{X} \otimes \mathscr{Y}$,

$$\begin{aligned}
\sum_{m\in\mathbb{N}} \left(\sum_{n\in\mathbb{N}} \pi_1^{nm}(E) \right) &= \sum_{m\in\mathbb{N}} \left(\int_X \left(\sum_{n\in\mathbb{N}} f_E^n(x) \right) \mu_m(dx) \right) \\
&= \sum_{m\in\mathbb{N}} \left(\int_X \nu(\{y \in Y \colon (x, y) \in E\})\mu_m(dx) \right) \\
&= \sum_{m\in\mathbb{N}} \left(\int_{A_m} \nu(\{y \in Y \colon (x, y) \in E\})\mu(dx) \right) \\
&= \int_X \nu(\{y \in Y \colon (x, y) \in E\})\mu(dx) = \tilde{\pi}_1(E),
\end{aligned}$$

and similarly, $\sum_{n\in\mathbb{N}}(\sum_{m\in\mathbb{N}}\pi_2^{nm}(E)) = \sum_{m\in\mathbb{N}}(\sum_{n\in\mathbb{N}}\pi_2^{nm}(E)) = \tilde{\pi}_2(E)$ (cf. Example 8.7). Therefore, by (68), $\tilde{\pi}_1(E) = \tilde{\pi}_2(E)$ for any $E \in \mathscr{X}\otimes\mathscr{Y}$. Hence, again, we define $\mu\otimes\nu(E) = \tilde{\pi}_1(E)$, $E \in \mathscr{X}\otimes\mathscr{Y}$, and observe that for any $A\times B$ s.t. $A\in\mathscr{X}$ and $B\in\mathscr{Y}$,

$$\mu\otimes\nu(A\times B) = \sum_{m\in\mathbb{N}}\left(\sum_{n\in\mathbb{N}}\pi_1^{nm}(A\times B)\right) = \sum_{m\in\mathbb{N}}\left(\sum_{n\in\mathbb{N}}\mu_m(A)\nu_n(B)\right) = \mu(A)\nu(B). \tag{69}$$

It remains to show that $\mu\otimes\nu$ is the unique measure on $\mathscr{X}\otimes\mathscr{Y}$ that satisfies (69). This follows, as in the finite case, from the fact that $\mathscr{P}$ generates $\mathscr{X}\otimes\mathscr{Y}$, i.e, we rely on Proposition 6.9. □

Proof of Proposition 9.11. Suppose that $f = \mathbb{1}_E$, where $E\in\mathscr{X}\otimes\mathscr{Y}$. Then, for any $x\in X$, we obtain that $\mathbb{1}_E(x,y) = \mathbb{1}_{\{y\in Y\colon (x,y)\in E\}}(y)$. Similarly, for any $y\in Y$, $\mathbb{1}_E(x,y) = \mathbb{1}_{\{x\in X\colon (x,y)\in E\}}(x)$ Hence,

$$\int_Y f(x,y)\nu(dy) = \nu(\{y\in Y\colon (x,y)\in E\}).$$

and

$$\int_X f(x,y)\mu(dx) = \mu(\{x\in X\colon (x,y)\in E\}).$$

As already shown in the proof of Proposition 9.10, the functions $x\mapsto\int_Y f(x,y)\nu(dy)$ and $y\mapsto\int_X f(x,y)\mu(dx)$ are $\mathscr{X}$ and $\mathscr{Y}$ measurable, respectively, and

$$\begin{aligned}\int_{X\times Y} f(x,y)(\mu\otimes\nu)(d(x,y)) &= \mu\otimes\nu(E)\\ &= \int_X \nu(\{y\in Y\colon (x,y)\in E\})\mu(dx)\\ &= \int_Y \mu(\{x\in X\colon (x,y)\in E\})\nu(dy).\end{aligned}$$

Thus, the proposition is already verified for $f=\mathbb{1}_E$, $E\in\mathscr{X}\otimes\mathscr{Y}$. If $f=\sum_{i=1}^N\alpha_i\mathbb{1}_{E_i}$ is a nonnegative simple function, where $E_i\in\mathscr{X}\otimes\mathscr{Y}$ for any $i=1,\ldots,N$, we apply Proposition 8.3 and the result is deduced from the previous case. Finally, if f is any $\mathscr{X}\otimes\mathscr{Y}$ measurable and nonnegative function, we apply Proposition 8.2 and the proposition is proven. □

Proof of Proposition 9.12. We focus on $\int_Y f(x,y)\nu(dy)$, the arguments for $\int_X f(x,y)\mu(dx)$ are the same. Since f is $\mathscr{X}\otimes\mathscr{Y}$ measurable, $|f|$ is a nonnegative and $\mathscr{X}\otimes\mathscr{Y}$ measurable function. By Proposition 9.11,

$$\int_{X\times Y}\big|f(x,y)\big|\mu\otimes\nu(d(x,y)) = \int_X\left(\int_Y\big|f(x,y)\big|\nu(dy)\right)\mu(dx). \tag{70}$$

Thus, by (ii) of Proposition 8.5, $\int_Y\big|f(x,y)\big|\nu(dy)<\infty$ μ a.e. Thus, we define

$$F_X = \left\{x\in X\colon \int_Y\big|f(x,y)\big|\nu(dy)<\infty\right\},$$

Then, given any $x\in F_X$,

$$\int_Y f(x,y)\nu(dy) = \int_Y f^+(x,y)\nu(dy) - \int_Y f^-(x,y)\nu(dy). \tag{71}$$

Then, we again apply Proposition 8.5 and deduce that the right hand side of (71) is $\mathscr{X}$ measurable. Therefore, for any $x\in F_X$, $\int_Y f(x,y)\nu(dy)$ is $\mathscr{X}$ measurable. Further, for any

$x \in F_X$, since $\int_Y f^+(x,y)\nu(dy) \le \int_Y |f(x,y)|\nu(dy)$ and $\int_Y f^-(x,y)\nu(dy) \le \int_Y |f(x,y)|\nu(dy)$, it follows from (70), that f^+ and f^- are μ integrable on F_X. Hence, for any $x \in F_X$, $\int_Y f(x,y)\nu(dy)$ is integrable with respect to μ. We define $g(x) = \mathbb{1}_{F_X}(x)\int_Y f(x,y)\nu(dy)$, $x \in X$, and have that $g(x) = \int_Y f(x,y)\nu(dy)$ μ a.e. By Exercise 8.3,

$$\int_X \left| \int_Y f(x,y)\nu(dy) \right| \mu(dx) < \infty,$$

and

$$\int_X \left(\int_Y f(x,y)\nu(dy) \right) \mu(dx) = \int_{F_X} \left(\int_Y f(x,y)\nu(dy) \right) \mu(dx). \tag{72}$$

On F_X, we apply Proposition 9.11 to f^+ and f^- and obtain

$$\int_{F_X \times Y} f^+(x,y)\mu \otimes \nu(d(x,y)) = \int_{F_X} \left(\int_Y f^+(x,y)\nu(dy) \right) \mu(dx),$$

and

$$\int_{F_X \times Y} f^-(x,y)\mu \otimes \nu(d(x,y)) = \int_{F_X} \left(\int_Y f^-(x,y)\nu(dy) \right) \mu(dx),$$

Hence,

$$\int_{F_X \times Y} f(x,y)\mu \otimes \nu(d(x,y)) = \int_{F_X} \left(\int_Y f(x,y)\nu(dy) \right) \mu(dx). \tag{73}$$

By (72), the right-hand side of the latter display is equal to $\int_X (\int_Y f(x,y)\nu(dy))\mu(dx)$. Then, we readily see that $(F_X \times Y)^c = (F_X \times Y^c) \cup (F_X^c \times Y)$ and hence

$$\mu \otimes \nu\big((F_X \times Y)^c\big) = \mu(F_X)\nu(Y^c) + \mu(F_X^c)\nu(Y) = 0.$$

Thus the right-hand side of (73) is equal to $\int_{X \times Y} f(x,y)\mu \otimes \nu(d(x,y))$ and the proof is complete. □

B.8 Summation: Integration with respect to the counting measure

We use Propositions 9.11 and 9.12 to prove Proposition 3.11.

Proof of Proposition 3.11. Let $I \times J \subset \mathbb{N} \times \mathbb{N}$ and $f \colon I \times J \to \mathbb{R}$, $f(i,j) = a_{ij}$. Consider the measure spaces $(I, \mathcal{P}(I), \mu_I)$ and $(J, \mathcal{P}(J), \mu_J)$, where μ_I is the counting measure on $\mathcal{P}(I)$ and μ_J is the counting measure and $\mathcal{P}(J)$. Clearly, μ_I is σ-finite on I: It is finite if I is finite and if I is not finite, then since I is countable, we write

$$I = \{i_k \colon k \in \mathbb{N}\} = \cup_{n \in \mathbb{N}}\{i_1, \dots, i_n\},$$

and observe that $\mu_I(\{i_1, \dots, i_n\}) < \infty$ for any $n \in \mathbb{N}$. Thus, μ_I and μ_J are σ-finite measures on $(I, \mathcal{P}(I), \mu_I)$ and $(J, \mathcal{P}(J), \mu_J)$, respectively. Further, we observe that for any $B \in \mathfrak{B}(\mathbb{R})$, $f^{-1}(B) \in \mathcal{P}(I \times J) = \mathcal{P}(I) \otimes \mathcal{P}(J)$ (cf. Example 9.6). That is, f is $\mathcal{P}(I) \otimes \mathcal{P}(J)$ measurable. Suppose that $f(i,j) \ge 0$ for any $(i,j) \in I \times J$. By Proposition 9.11, we know that

$$\begin{aligned}&\int_{I \times J} f(i,j)\mu_I \otimes \mu_J(d(i,j)) \\ &= \int_I \left(\int_J f(i,j)\mu_J(dj) \right) \mu_I(di) = \int_J \left(\int_I f(i,j)\mu_I(di) \right) \mu_J(dj).\end{aligned} \tag{74}$$

Given any $i \in I$, let $\mathbb{N} \ni j \mapsto \tilde{f}(i,j) = f(i,j)\mathbb{1}_J(j)$. Notice that for any $i \in I$, $j \mapsto \tilde{f}(i,j)$ is $\mathcal{P}(J)$ measurable by Proposition B.9. We obtain,

$$\int_J f(i,j)\mu_J(dj) = \int_J f(i,j)\mu(dj) = \int_{\mathbb{N}} \tilde{f}(i,j)\mu(dj),$$

where μ is the counting measure on $\mathcal{P}(\mathbb{N})$. Then, by Example 8.6, we know that

$$\int_{\mathbb{N}} \tilde{f}(i,j)\mu(dj) = \sum_{j\in\mathbb{N}} \tilde{f}(i,j) = \sum_{j\in\mathbb{N}} f(i,j)\mathbb{1}_J(j) = \sum_{j\in J} f(i,j).$$

Hence, by (74), we get

$$\sum_{i\in I}\left(\sum_{j\in J} a_{ij}\right) = \sum_{j\in J}\left(\sum_{i\in I} a_{ij}\right).$$

It remains to show that

$$\int_{I\times J} f(i,j)\mu_I \otimes \mu_J(d(i,j)) = \sum_{(i,j)\in I\times J} a_{ij}$$

First, $\mu \otimes \mu = \mu_2$, where μ_2 is the counting measure on $\mathcal{P}(\mathbb{N}\times\mathbb{N})$ (cf. Example 9.7). Let $A \in \mathcal{P}(I)$ and $B \in \mathcal{P}(J)$, then,

$$\mu_I \otimes \mu_J(A\times B) = \mu_I(A)\mu_J(B) = \mu(A)\mu(B).$$

Also,

$$(\mu\otimes\mu)|_{I\times J}(A\times B) = \mu_2|_{I\times J}(A\times B) = \mu(A)\mu(B).$$

Hence, $\mu_2|_{I\times J} = \mu_I \otimes \mu_J$. We write $\overline{f}(i,j) = f(i,j)\mathbb{1}_{I\times J}(i,j)$, $(i,j)\in\mathbb{N}\times\mathbb{N}$. Then,

$$\int_{I\times J} f(i,j)\mu_I\otimes\mu_J(d(i,j)) = \int_{\mathbb{N}\times\mathbb{N}} \overline{f}(i,j)\mu_2(d(i,j)).$$

We then repeat the arguments given in Example 8.6 with

$$\overline{f}_n(i,j) = \overline{f}(i,j)\mathbb{1}_{\{1,\dots,n\}\times\{1,\dots,n\}},$$

and the decomposition

$$\mathbb{N}^2 = \left(\bigcup_{(i,j)\in\{1,\dots,n\}^2} \{i,j\}\right) \cup \mathbb{N}^2 \setminus \{1,\dots,n\}^2,$$

and obtain

$$\int_{\mathbb{N}\times\mathbb{N}} \overline{f}_n(i,j)\mu_2(d(i,j)) = \sum_{(i,j)\in\{1,\dots,n\}^2} \overline{f}_n(i,j) = \sum_{(i,j)\in\{1,\dots,n\}^2} \overline{f}(i,j),$$

where

$$\lim_{n\to\infty}\left(\int_{\mathbb{N}\times\mathbb{N}} \overline{f}_n(i,j)\mu_2(d(i,j))\right) = \int_{\mathbb{N}\times\mathbb{N}} \overline{f}(i,j)\mu_2(d(i,j)).$$

This conludes the argument since

$$\lim_{n\to\infty}\left(\sum_{(i,j)\in\{1,\dots,n\}^2} \overline{f}(i,j)\right) = \sum_{(i,j)\in\mathbb{N}^2} \overline{f}(i,j) = \sum_{(i,j)\in I\times J} f(i,j).$$

Assume now that $\sum_{(i,j)\in I\times J}|f(i,j)| < \infty$. Then, by the previous arguments, we know that this implies that $\int_{I\times J}|f(i,j)|\mu_I\otimes\mu_J(d(i,j)) < \infty$. Then, we rely on Proposition 9.12 and Proposition 3.11 is proven. □

Upon the latter proof, we deduce the following result:

Proposition B.11. *Let $E_1, \dots, E_n$ be countable sets. Suppose that $f\colon E_1 \times \cdots \times E_n \to \mathbb{R}$ is a nonnegative function. Let $E = E_1 \times \cdots \times E_n$. Then,*

$$\sum_{x \in E} f(x_1, \dots, x_n) = \sum_{x_1 \in E_1} \left(\cdots \left(\sum_{x_{n-1} \in E_{n-1}} \left(\sum_{x_n \in E_n} f(x_1, \dots, x_n) \right) \right) \cdots \right), \tag{75}$$

where the sum in the latter display can be computed in arbitrary order. If f is not necessarily nonnegative but s.t. $\sum_{x \in E} |f(x_1, \dots, x_n)| < \infty$, (75) remains valid.

Proof. If $f\colon E \to \mathbb{R}$ is nonnegative, this is a consequence of Proposition 9.11. We notice that since $E_1, \dots, E_n$ are assumed to be countable, there exists a bijection $g\colon I_1 \times \cdots \times I_n \to E_1 \times \cdots \times E_n$ s.t. for any $(x_1, \dots, x_n) \in E_1 \times \cdots \times E_n$,

$$f(x_1, \dots, x_n) = f(g(i_1, \dots, i_n)), \quad (i_1, \dots, i_n) \in I_1 \times \cdots \times I_n.$$

Thus, if we write $h = f \circ g$, we obtain

$$\sum_{x \in E} f(x_1, \dots, x_n) = \sum_{(i_1, \dots, i_n) \in I_1 \times \cdots \times I_n} h(i_1, \dots, i_n). \tag{76}$$

Then, we rely on the same reasoning as in the proof of Proposition 3.11 and verify that the right hand side of (76) equals

$$\int_{\mathbb{N} \times \cdots \times \mathbb{N}} h(i_1, \dots, i_n) \mathbb{1}_{I_1 \times \cdots \times I_n}(i_1, \dots, i_n) \mu \otimes \cdots \otimes \mu(d(i_1, \dots, i_n)),$$

where μ is the counting measure on $\mathcal{P}(\mathbb{N})$. Therefore, (75) follows from Proposition 9.11. In particular, under the assumption that $\sum_{x \in E} |f(x_1, \dots, x_n)| < \infty$, (75) follows from Proposition 9.12. □

C Probability

C.1 On the distribution function

Proof of Proposition 10.5. Upon Proposition A.15, F_X is continuous at a point $t \in \mathbb{R}$, if and only $\lim_{y \uparrow t} F_X(y) = F(t) = \lim_{y \downarrow t} F_X(y)$. We have that

$$\lim_{y \uparrow t} F_X(y) = \lim_{y \uparrow t} P_X((-\infty, y]).$$

Take any sequence of real numbers $(y_n)_{n \in \mathbb{N}}$ s.t. $y_n \uparrow t$ ($y_n \neq t$ for any $n \in \mathbb{N}$). By Proposition 5.1, we have that

$$\lim_{n \to \infty} F_X(y_n) = \lim_{n \to \infty} P_X((-\infty, y_n]) = P_X\left(\bigcup_{n \in \mathbb{N}} (-\infty, y_n] \right).$$

Then, we observe that

$$\bigcup_{n \in \mathbb{N}} (-\infty, y_n] = (-\infty, t).$$

To see it, let $y \in (-\infty, t)$, i.e., there exists $\varepsilon > 0$, s.t. $y \leq t - \varepsilon$. Then, since $y_n \uparrow t$, there exists $N \in \mathbb{N}$ s.t. $t - y_n \leq \varepsilon$ for any $n \geq N$. Thus, $y_N \geq t - \varepsilon \geq y$, i.e., $y \in \cup_{n \in \mathbb{N}} (-\infty, y_n]$. The other inclusion is obvious. Hence,

$$\lim_{n \to \infty} F_X(y_n) = P_X((-\infty, t)).$$

This shows that $\lim_{y\uparrow t} F_X(y) = P_X((-\infty, t))$ (cf. Proposition A.16). Similarly, we deduce that

$$\lim_{y\downarrow t} F_X(y) = P_X((-\infty, t]) = F_X(t).$$

Notice that in this case, the argument relies on the fact that for any sequence $(y_n)_{n\in\mathbb{N}}$ s.t. $y_n \downarrow t$, $\cap_{n\in\mathbb{N}} (-\infty, y_n] = (-\infty, t]$. Here, the inclusion $(-\infty, t] \subset \cap_{n\in\mathbb{N}} (-\infty, y_n]$ is obvious. For the other inclusion, suppose that $y \in \cap_{n\in\mathbb{N}} (-\infty, y_n]$. Then, for any $n \in \mathbb{N}$, $y \le y_n$. Since $y_n \downarrow t$, it follows that $y \le \lim_{n\to\infty} y_n = t$ (cf. Exercise 4.13). In conclusion,

$$\lim_{y\downarrow t} F_X(y) - \lim_{y\uparrow t} F_X(y) = P_X((-\infty, t]) - P_X((-\infty, t)) = \mathbb{P}(X = t).$$

Hence, F_X is continuous at t if and only if $\mathbb{P}(X = t) = 0$. $\square$

Remark C.1. *We remark that Proposition 10.5 shows that if X is continuous, the distribution function F_X is continuous on $\mathbb{R}$. Further, in the proof of Proposition 10.5 we deduced that*

$$\lim_{y\downarrow t} F_X(y) - \lim_{y\uparrow t} F_X(y) = \mathbb{P}(X = t),$$

i.e., the points of discontinuity of F_X are given by the $t \in \mathbb{R}$ for which $\mathbb{P}(X = t) \neq 0$. In addition, we have seen that for any $t \in \mathbb{R}$, the right-hand limit $\lim_{y\downarrow t} F_X(y)$ is equal to $F_X(t)$, that is to say that F_X is right-continuous. By definition, F_X is increasing. Further, $\lim_{y\to-\infty} F_X(y) = 0$ and $\lim_{y\to\infty} F_X(y) = 1$ (cf. Proposition A.18). In summary, F_X is right-continuous with jumps at the points $t \in \mathbb{R}$ for which $\mathbb{P}(X = t) \neq 0$.

Remark C.2. *Notice that the set E in Proposition 10.6 is not empty. To see it, let $p \in (0, 1)$. Choose $\varepsilon > 0$ s.t. $\varepsilon < p < 1-\varepsilon$. Since $\lim_{y\to-\infty} F_X(y) = 0$ there exists a real number $M_1 > 0$ s.t. if $y < -M_1$, $F_X(y) < \varepsilon$ (cf. Proposition A.17). Further, since $\lim_{y\to\infty} F_X(y) = 1$, there exists a real number $M_2 > 0$ s.t. if $y > M_2$, $|F_X(y) - 1| < \varepsilon \Leftrightarrow F_X(y) > 1 - \varepsilon$. Let $y_a < -M_1$ and $y_b > M_2$. We define $a = F_X(y_a)$ and $b = F_X(y_b)$ and obtain that $p \in (a, b)$. By the intermediate value theorem (cf. Proposition A.7), there exists $t \in [y_a, y_b]$ s.t. $F_X|_{[y_a,y_b]}(t) = F_X(t) = p$. In particular, there exists $t \in \mathbb{R}$ s.t. $0 < F_X(t) < 1$.*

Proof of Proposition 10.6. We aim to apply Proposition 2.2. Therefore, it remains to show that $F_X|_E(E) = (0, 1)$. By definition of E, $F_X|_E(E) \subset (0, 1)$. For the other inclusion, we have seen in Remark C.2 that for any $p \in (0, 1)$ there exists $t \in \mathbb{R}$ s.t. $F_X(t) = p$. But this implies that $t \in E$ and hence, $F_X|_E(t) = p$. $\square$

C.2 Second moments

Proof of Proposition 10.8. Notice first that if $\mathbb{E}[X^2] = \mathbb{E}[Y^2] = 0$, then $\mathbb{P}(X = 0) = \mathbb{P}(Y = 0) = 1$, i.e., (31) is an equality (cf. Proposition 8.5). Hence, we suppose that $\mathbb{E}[X^2] > 0$ (otherwise choose $\mathbb{E}[Y^2] > 0$). Let $a \in \mathbb{R}$. We have that

$$0 \le \mathbb{E}[(a|X| + |Y|)^2] = \mathbb{E}[a^2X^2 + 2a|X||Y| + Y^2] = a^2\mathbb{E}[X^2] + 2a\mathbb{E}[|XY|] + \mathbb{E}[Y^2]. \quad (77)$$

Define the quadratic function

$$q(a) = a^2\mathbb{E}[X^2] + 2a\mathbb{E}[|XY|] + \mathbb{E}[Y^2], \quad a \in \mathbb{R}. \quad (78)$$

By (77), $q(a) \ge 0$ for any $a \in \mathbb{R}$. This implies that the discriminant $d_q = 4(\mathbb{E}[|XY|])^2 - 4\mathbb{E}[X^2]\mathbb{E}[Y^2]$ defined by q has to be smaller or equal to zero. To see it, by the assumption that $\mathbb{E}[X^2] > 0$, the second derivative of q with respect to a is strictly positive and hence, q admits a unique and global minimum at zero. Therefore, $d_q > 0$ is not possible as it would imply that the equation $q(a) = 0$ has two different real solutions. Hence, the result is proven. $\square$

Proof of Proposition 10.9. We have that

$$\begin{aligned}
\mathrm{Var}(v^t X) &= \mathbb{E}\Big[\Big(\sum_{i=1}^k v_i X_i\Big)^2\Big] - \Big(\mathbb{E}\Big[\sum_{i=1}^k v_i X_i\Big]\Big)^2 \\
&= \mathbb{E}\Big[\sum_{i=1}^k\Big(\sum_{j=1}^k v_i v_j X_i X_j\Big)\Big] - \Big(\sum_{i=1}^k\Big(\sum_{j=1}^k v_i v_j \mathbb{E}[X_i]\mathbb{E}[X_j]\Big)\Big) \\
&= \sum_{i=1}^k v_i\Big(\sum_{j=1}^k v_j(\mathbb{E}[X_iX_j] - \mathbb{E}[X_i]\mathbb{E}[X_j])\Big) \\
&= \sum_{i=1}^k v_i\Big(\sum_{j=1}^k v_j \Sigma(X)_{i,j}\Big) = v^t\Sigma(X)v.
\end{aligned}$$

□

C.3 Uniqueness theorem of the characteristic function

Proof of Proposition 10.11. Let $\sigma = 0$. Since $X = \mu$ $\mathbb{P}$ a.s., it follows that for any $v \in \mathbb{R}$,

$$\Phi_X(v) = \mathbb{E}[\mathrm{e}^{ivX}] = \mathbb{E}[\mathrm{e}^{iv\mu}] = \mathrm{e}^{i\mu v}.$$

Therefore, it remains to show the proposition for the case where $\sigma > 0$. Assume that $\mu = 0$ and $\sigma = 1$. Let $v \in \mathbb{R}$. We have that

$$\Phi_X(v) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{e}^{ivx}\,dx.$$

Then, upon Euler's Formula, i.e., $\mathrm{e}^{ivx} = \cos(vx) + i\sin(vx)$, we have that

$$\Phi_X(v) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}\cos(vx)dx,$$

since the integral $\int_{\mathbb{R}} \mathrm{e}^{-x^2/2}\sin(vx)dx$ is equal to zero (cf. Example 9.4). Let $(a,b) \subset \mathbb{R}$ be an open interval. For any $x \in \mathbb{R}$, we set

$$f_x(v) = \mathrm{e}^{-\frac{x^2}{2}}\cos(vx), \quad v \in (a,b).$$

Then, for any $x \in \mathbb{R}$, f_x is differentiable with derivative

$$f_x'(v) = -x\,\mathrm{e}^{-\frac{x^2}{2}}\sin(vx), \quad v \in (a,b).$$

We notice that for any $v_0 \in (a,b)$, $|f_x'(v_0)| \le g(x)$, with $g(x) = |x|\,\mathrm{e}^{-x^2/2}$, $x \in \mathbb{R}$. By Example 10.18 we know that $\int_{\mathbb{R}} g(x)dx < \infty$. Hence, Proposition B.6 applies and we obtain that for any $v \in (a,b)$,

$$\Phi_X'(v) = -\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} x\,\mathrm{e}^{-\frac{x^2}{2}}\sin(vx)dx. \tag{79}$$

Since the interval (a,b) was arbitrary (79) holds for any $v \in \mathbb{R}$. We apply integration by parts and readily deduce that

$$\Phi_X'(v) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}\,v\cos(vx)dx = -v\Phi_X(v), \quad v \in \mathbb{R}.$$

This shows that $\Phi_X(v) = \mathrm{e}^{-v^2/2}$ for any $v \in \mathbb{R}$. Notice that a general $\sigma > 0$ does not change the argument and hence, if $X \sim \mathcal{N}(0,\sigma^2)$, $\Phi_X(v) = \mathrm{e}^{-(\sigma^2v^2)/2}$. Then, if $X \sim \mathcal{N}(\mu,\sigma^2)$, $Y = X - \mu \sim \mathcal{N}(0,\sigma^2)$ (cf. Exercise 10.4) and we obtain

$$\Phi_X(v) = \Phi_{Y+\mu}(v) = \mathbb{E}[\mathrm{e}^{iv(Y+\mu)}] = \mathrm{e}^{iv\mu}\,\mathbb{E}[\mathrm{e}^{ivY}] = \mathrm{e}^{i\mu v - \frac{\sigma^2v^2}{2}}.$$

□

In order to verify the uniqueness theorem of the characteristic function, we first consider some helpful results.

Definition C.1. *We define*

$$d(x,A)=\begin{cases}\inf_{a\in A}\|x-a\|, & \text{if } (x,A)\in\mathbb{R}\times\mathcal{P}(\mathbb{R}^k)\setminus\emptyset,\\ 1, & \text{otherwise,}\end{cases}$$

as the distance from $x\in\mathbb{R}$ to a set $A\in\mathcal{P}(\mathbb{R}^k)$.

Proposition C.1. *Given $A\in\mathcal{P}(\mathbb{R}^k)$, $x\mapsto d(x,A)$ is continuous.*

Proof. Suppose that $A=\emptyset$, then, by Definition C.1, $|d(x,A)-d(y,A)|\le\|x-y\|$. Suppose that $A\neq\emptyset$ and let $x,y\in\mathbb{R}^k$. Given any $a\in A$, we obtain

$$d(x,A)\le\|x-a\|\le\|x-y\|+\|y-a\|.$$

Since $a\in A$ was arbitrary we take the infimum on the right of the latter inequality and deduce that $d(x,A)\le\|x-y\|+d(y,A)$. That is, $d(x,A)-d(y,A)\le\|x-y\|$. Similarly, givan $a\in A$,

$$d(y,A)\le\|y-a\|\le\|y-x\|+\|x-a\|,$$

and hence $d(y,A)\le\|y-x\|+d(x,A)$, i.e., $d(y,A)-d(x,A)\le\|y-x\|$. In conclusion, for any $x,y\in\mathbb{R}^k$,

$$|d(x,A)-d(y,A)|\le\|x-y\|,\quad A\in\mathcal{P}(\mathbb{R}^k).$$

Therefore, by Definition A.2, for any $A\in\mathcal{P}(\mathbb{R}^k)$, $x\mapsto d(x,A)$ is continuous. □

Proposition C.2. *Let $U\subset\mathbb{R}^k$ be an open set. Then, there exists a sequence of continuous functions $g_n\colon\mathbb{R}^k\to\mathbb{R}$ s.t. for any $x\in\mathbb{R}^k$, $0\le g_n(x)\le 1$ and $g_n(x)\uparrow \mathbb{1}_U(x)$.*

Proof. Let $U\subset\mathbb{R}^k$, open. Set $A_n=\{y\in\mathbb{R}^k\colon d(y,U^c)\ge 1/n\}$, $n\in\mathbb{N}$. Then, for any $n\in\mathbb{N}$, $U^c\subset A_n^c$ and hence $A_n\subset U$. Further, $A_n\subset A_{n+1}$ and $\cup_{n\in\mathbb{N}}A_n=U$ (notice that $\cap_{n\in\mathbb{N}}A_n^c=U^c$). Define

$$g_n(x)=\frac{d(x,U^c)}{d(x,U^c)+d(x,A_n)},\quad x\in\mathbb{R}^k.$$

Then, for any $x\in\mathbb{R}^k$, $0\le g_n(x)\le 1$ and $g_n(x)\le g_{n+1}(x)$. It remains to show that for any $x\in\mathbb{R}^k$, $\lim_{n\to\infty}g_n(x)=\mathbb{1}_U(x)$. Let $x\in U$, then, since $\cup_{n\in\mathbb{N}}A_n=U$, there exists $N\in\mathbb{N}$, s.t. $x\in A_n$ for any $n\ge N$, i.e., $g_n(x)=1$ for any $n\ge N$. If $x\notin U$, then, $x\in A_n^c$ for any $n\in\mathbb{N}$. That is $d(x,U^c)<1/n$ for any $n\in\mathbb{N}$ and hence, $g_n(x)\xrightarrow{n\to\infty}0$. □

Proof of Proposition 10.12. Let $k=1$. Define

$$h_\sigma(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{e}^{-\frac{x^2}{2\sigma^2}},\quad x\in\mathbb{R},$$

i.e., h_σ is the probability density function of a $\mathcal{N}(0,\sigma^2)$ distribution. Let μ be a probability measure on $\mathfrak{B}(\mathbb{R})$. Define

$$f_\sigma(x)=\int_{\mathbb{R}}h_\sigma(x-y)\mu(dy)$$
$$\mu_\sigma(dx)=f_\sigma(x)dx.$$

By the Fubini-Tonnelli theorem (cf. Proposition 9.11), the map $x \mapsto f_\sigma(x)$ is $\mathfrak{B}(\mathbb{R})$ measurable. Clearly, f_σ is nonnegative and we recall that μ_σ is the measure on $\mathfrak{B}(\mathbb{R})$ defined by (cf. Proposition 9.2)

$$\mu_\sigma(A) = \int_A f_\sigma(x)dx.$$

By the Fubini-Tonnelli theorem again,

$$\mu_\sigma(\mathbb{R}) = \int_\mathbb{R} \Big(\int_\mathbb{R} h_\sigma(x-y)\mu(dy) \Big) dx = \int_\mathbb{R} \Big(\int_\mathbb{R} h_\sigma(x-y)dx \Big) \mu(dy) = \int_\mathbb{R} \mu(dy) = 1,$$

that is for any $\sigma > 0$, μ_σ is a probability measure on $\mathfrak{B}(\mathbb{R})$. Define

$$C_b(\mathbb{R}) = \{g\colon g\colon \mathbb{R} \to \mathbb{R}\ g \text{ continuous and bounded}\}.$$

Introduce the following conditions:

(1) If ν is another probability measure on $\mathfrak{B}(\mathbb{R})$ s.t. $\widehat{\nu}(v) = \widehat{\mu}(v)$ for any $v \in \mathbb{R}$, then $\nu_\sigma = \mu_\sigma$;

(2) for any $g \in C_b(\mathbb{R})$, $\lim_{\sigma\to 0} \int_\mathbb{R} g(x)\mu_\sigma(dx) = \int_\mathbb{R} g(x)\mu(dx)$.

Suppose that (1) and (2) are satisfied and let P and P' as in the statement of the proposition. Condition (1) shows that for any $\sigma > 0$, $P_\sigma = P'_\sigma$. Then, upon (2), for any $g \in C_b(\mathbb{R})$,

$$\int_\mathbb{R} g(x)P(dx) = \int_\mathbb{R} g(x)P'(dx).$$

By Proposition C.2, for any open set $U \subset \mathbb{R}$, there exists a sequence of continuous, bounded and nonnegative functions $g_n\colon \mathbb{R} \to \mathbb{R}$ s.t. for any $x \in \mathbb{R}$, $g_n(x) \uparrow \mathbb{1}_U(x)$. We conclude with Proposition 8.2, that $P(U) = P'(U)$ for any $U \subset \mathbb{R}$, open. Since $\mathfrak{B}(\mathbb{R}) = \sigma(\{U\colon U \subset \mathbb{R} \text{ open}\})$ we apply Proposition 6.9 and conclude that $P = P'$. Hence it remains to verify (1) and (2). In order to show (1), we use Proposition 10.11 and obtain that for any $x \in \mathbb{R}$,

$$\sqrt{2\pi}\sigma h_\sigma(x) = \mathrm{e}^{-\frac{x^2}{2\sigma^2}} = \int_\mathbb{R} \mathrm{e}^{ixt}\, h_{1/\sigma}(t)dt.$$

Further, we notice that $|\mathrm{e}^{i(x-y)t}\, h_{1/\sigma}(t)| = |h_{1/\sigma}(t)|$ for any $x, y, t \in \mathbb{R}$ and therefore,

$$\int_{\mathbb{R}^2} |\mathrm{e}^{i(x-y)t}\, h_{1/\sigma}(t)| \lambda \otimes \mu(d(t,y)) = \int_{\mathbb{R}^2} |h_{1/\sigma}(t)| \lambda \otimes \mu(d(t,y)) = \mu(\mathbb{R}) \int_\mathbb{R} |h_{1/\sigma}(t)|dt,$$

where we used Fubini-Tonnelli for the last equality. Therefore, since $t \mapsto h_{1/\sigma}(t)$ is integrable with respect to the Lebesgue measure, $(t,y) \mapsto \mathrm{e}^{i(x-y)t}\, h_{1/\sigma}(t)$ is integrable with respect to the product measure $\lambda \otimes \mu$ on $\mathfrak{B}(\mathbb{R}^2)$. Hence, upon Fubini-Lebesgue (cf. Proposition 9.12), the following is justified

$$\begin{aligned} f_\sigma(x) &= \int_\mathbb{R} h_\sigma(x-y)\mu(dy) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_\mathbb{R} \Big(\int_\mathbb{R} \mathrm{e}^{i(x-y)t}\, h_{1/\sigma}(t)dt \Big) \mu(dy) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_\mathbb{R} \mathrm{e}^{ixt}\, h_{1/\sigma}(t) \Big(\int_\mathbb{R} \mathrm{e}^{-ity}\, \mu(dy) \Big) dt \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_\mathbb{R} \mathrm{e}^{ixt}\, h_{1/\sigma}(t)\widehat{\mu}(-t)dt. \end{aligned}$$

This shows (1). We verify (2). Using again Fubini-Lebesgue, we write for any $g \in C_b(\mathbb{R})$,

$$\int_{\mathbb{R}} g(x)\mu_\sigma(dx) = \int_{\mathbb{R}} g(x)f_\sigma(x)dx = \int_{\mathbb{R}} g(x)\Big(\int_{\mathbb{R}} h_\sigma(x-y)\mu(dy)\Big)dx = \int_{\mathbb{R}} (h_\sigma * g)(y)\mu(dy),$$

where we recall Example 8.5. If we use the substitution $u = (x-y)/\sigma$, we deduce that for any $y \in \mathbb{R}$,

$$h_\sigma * g(y) = \int_{\mathbb{R}} h_\sigma(x-y)g(x)dx = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}\, \mathrm{e}^{-\frac{u^2}{2}}\, g(\sigma u + y)du.$$

Since g is bounded and $(1/(\sqrt{2\pi}))\int_{\mathbb{R}} \mathrm{e}^{-\frac{u^2}{2}}\, du = 1$, we obtain upon Lebesgue's dominated convergence theorem (cf. Proposition 8.10) that for any $y \in \mathbb{R}$,

$$\lim_{\sigma\to 0} h_\sigma * g(y) = g(y).$$

Keep in mind that g is assumed to be continuous. Then, we also notice that for any $y \in \mathbb{R}$ (cf. Exercise 8.6),

$$\big|h_\sigma * g(y)\big| \le \int_{\mathbb{R}} \big|h_\sigma(x-y)g(x)\big|dx \le C\int_{\mathbb{R}} h_\sigma(x-y)dx = C,$$

where $C \in \mathbb{R}$ s.t. for any $x \in \mathbb{R}$, $|g(x)| \le C$ (upon the assumption that g is bounded). Hence if we apply Lebesgue's dominated convergence again, we get that

$$\lim_{\sigma\to 0}\int_{\mathbb{R}} g(x)\mu_\sigma(dx) = \int_{\mathbb{R}} g(x)\mu(dx).$$

This completes the proof for the case where $k = 1$. If $k \in \mathbb{N}$ we replace $h_\sigma(x)$, $x \in \mathbb{R}$, with the function

$$h_\sigma^k(x) = \prod_{i=1}^k h_\sigma(x_i), \quad x = (x_1, \ldots, x_k) \in \mathbb{R}^k,$$

and condition (2) with

(2′) for any function $g\colon \mathbb{R}^k \to \mathbb{R}$ which is s.t. $g(x_1,\ldots,x_k) = \prod_{i=1}^k g_i(x_i)$, with $g_i \in C_b(\mathbb{R})$, $i = 1,\ldots,k$, we have that $\lim_{\sigma\to 0}\int_{\mathbb{R}^k} g(x)\mu_\sigma(dx) = \int_{\mathbb{R}^k} g(x)\mu(dx)$.

Notice that for any $x \in \mathbb{R}^k$,

$$(2\pi)^{k/2}\sigma^k h_\sigma^k(x) = \int_{\mathbb{R}^k} \mathrm{e}^{ix^t w}\, h_{1/\sigma}^k(w)dw.$$

Hence, we proceed as before and conclude that for any g as in (2′),

$$\int_{\mathbb{R}^k} g(x)P(dx) = \int_{\mathbb{R}^k} g(x)P'(dx). \tag{80}$$

Let $(a_1, b_1) \subset \mathbb{R}$ be an open interval of $\mathbb{R}$. Then, by Proposition C.2, there exists a sequence of functions $(g_1^n)_{n\in\mathbb{N}}$ which is s.t. $g_1^n \in C_b(\mathbb{R})$ for any $n \in \mathbb{N}$ and $g_i^n(x) \uparrow \mathbb{1}_{(a_1,b_1)}(x)$, $x \in \mathbb{R}$. Therefore, by (80), if we define $g_n(x) = g_1^n(x_1)\prod_{i=2}^k g_i(x_i)$, $g_i \in C_b(\mathbb{R})$ for any $i = 2,\ldots,k$, we obtain with Proposition 8.2,

$$\int_{\mathbb{R}^k} \mathbb{1}_{(a_1,b_1)}(x_1)\prod_{i=2}^k g_i(x_i)P(dx) = \int_{\mathbb{R}^k} \mathbb{1}_{(a_1,b_1)}(x_1)\prod_{i=2}^k g_i(x_i)P'(dx).$$

Upon k iterations of the previous argument, we deduce that $P(A) = P'(A)$ for any $A \in \mathcal{G}$, where

$$\mathcal{G} = \{A \subset \mathbb{R}^k \colon A = \prod_{i=1}^{k} (a_i, b_i),\ a_i, b_i \in \mathbb{R},\ i = 1, \ldots, k\} \cup \{\emptyset\}.$$

We know that $\mathfrak{B}(\mathbb{R}^k) = \sigma(\mathcal{R}_k)$, where

$$\mathcal{R}_k = \big\{A \colon A = \prod_{i=1}^{k} (a_i, b_i],\ a_i, b_i \in \mathbb{R},\ i = 1, \ldots, k\big\} \cup \{\emptyset\}.$$

Since for any $\mathcal{R}_k \ni A = \prod_{i=1}^{k} (a_i, b_i]$,

$$A = \bigcap_{n \in \mathbb{N}} \left(\prod_{i=1}^{k} \big(a_i, b_i + 1/n\big) \right),$$

we readily see that $\mathfrak{B}(\mathbb{R}^k) = \sigma(\mathcal{G})$. Hence, we apply Proposition 6.9 and deduce that $P(A) = P'(A)$ for any $A \in \mathfrak{B}(\mathbb{R}^k)$. □

C.4 Results on independence

Proof of Proposition 11.2. By Remark 11.4, the map $X(\omega) = (X_1(\omega), \ldots, X_n(\omega))$ is

$$\mathcal{F}/(\mathfrak{B}(\mathbb{R}^{k_1}) \otimes \cdots \otimes \mathfrak{B}(\mathbb{R}^{k_n}))$$

measurable. Suppose first that for any $i = 1, \ldots, n$, f_i is nonnegative and $\mathfrak{B}(\mathbb{R}^{k_i})$ measurable. Upon Proposition B.7, the map $f \colon \mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_n} \to \overline{\mathbb{R}}$ defined by

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_i(x_i), \quad (x_1, \ldots, x_n) \in \mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_n}, \tag{81}$$

is $\mathfrak{B}(\mathbb{R}^{k_1}) \otimes \cdots \otimes \mathfrak{B}(\mathbb{R}^{k_n})$ measurable. We thus apply (35) of Remark 11.4 and deduce that

$$\mathbb{E}\left[\prod_{i=1}^{n} f_i(X_i)\right] = \mathbb{E}[f(X)] = \int_{\mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_n}} f(x) P_X(dx). \tag{82}$$

By Proposition 11.1, $P_X = P_{X_1} \otimes \cdots \otimes P_{X_1}$ and hence we use Fubini-Tonnelli (cf. Proposition 9.11) and obtain

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^{n} f_i(X_i)\right] &= \int_{\mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_n}} f(x) P_{X_1} \otimes \cdots \otimes P_{X_1}(dx) \\
&= \prod_{i=1}^{n} \left(\int_{\mathbb{R}^{k_i}} f(x_i) P_{X_i}(dx_i) \right) = \prod_{i=1}^{n} \mathbb{E}\big[f_i(X_i)\big].
\end{aligned}$$

For the remaining case, assume that $\mathbb{E}[|f_i(X_i)|] < \infty$ for any $i = 1, \ldots, n$. With f as in (81), we obatin

$$\mathbb{E}[|f(X)|] = \prod_{i=1}^{n} \mathbb{E}\big[|f_i(X_i)|\big] < \infty,$$

by the previous case. Thus, upon Remark 11.4, (82) remains valid and the proposition is proven. □

Proof of Proposition 11.5. By Definition 10.14, the Fourier transform of $P_{X_1} \otimes \cdots \otimes P_{X_n}$ is given by

$$P_{X_1} \widehat{\otimes \cdots \otimes} P_{X_n}(v) = \int_{\mathbb{R}^n} \left(\prod_{i=1}^n \mathrm{e}^{iv_i x_i} \right) P_{X_1} \otimes \cdots \otimes P_{X_n}(d(x_1, \ldots, x_n)).$$

If we write $\mathrm{e}^{iv_i x_i} = \cos(v_i x_i) + i \sin(v_i x_i)$, $i = 1, \ldots, n$, we recall Definition 10.13, apply Fubini-Lebesgue (cf. Proposition 9.14) and obtain

$$P_{X_1} \widehat{\otimes \cdots \otimes} P_{X_n}(v) = \prod_{i=1}^n \left(\int_{\mathbb{R}} \mathrm{e}^{iv_i x_i} \, P_{X_i}(dx_i) \right) = \prod_{i=1}^n \widehat{P}_{X_i}(v_i).$$

This concludes the proof of the proposition since by Proposition 11.1, $X_1, \ldots, X_n$ are independent if and only if the law of X is the product measure on $\mathfrak{B}(\mathbb{R}^n)$. $\square$

Proof of Proposition 11.8. By item (i) of Proposition 11.7, the law of Z is given by

$$P_{X_1} * \cdots * P_{X_n} = P_{X_1} \otimes \cdots \otimes P_{X_n} s^{-1}, \quad s(x_1, \ldots, x_n) = \sum_{i=1}^n x_i.$$

Hence, the characteristic function of Z is given by the Fourier transform of the latter measure, i.e., for any $v \in \mathbb{R}^k$,

$$\begin{aligned}
\Phi_Z(v) &= P_{X_1} \widehat{* \cdots *} P_{X_n}(v) \\
&= \int_{\mathbb{R}^k} \mathrm{e}^{iv^t z} \, P_{X_1} * \cdots * P_{X_n}(dz) \\
&= \int_{\mathbb{R}^k} \mathrm{e}^{iv^t z} \, P_{X_1} \otimes \cdots \otimes P_{X_n} s^{-1}(dz) \\
&= \int_{\mathbb{R}^k \times \cdots \times \mathbb{R}^k} \mathrm{e}^{iv^t (x_1 + \cdots x_n))} \, P_{X_1} \otimes \cdots \otimes P_{X_n}(d(x_1, \ldots, x_n)) \\
&= \prod_{i=1}^n \left(\int_{\mathbb{R}^k} \mathrm{e}^{iv^t x_i} \, P_{X_i}(dx_i) \right) = \prod_{i=1}^n \Phi_{X_i}(v).
\end{aligned}$$

$\square$

C.5 Grouping of independent random vectors

Proposition C.3. *Let $f \colon \Omega \to \Omega^*$. Suppose that $\mathcal{G} \subset \mathcal{P}(\Omega^*)$ is a family of subsets of Ω^*. Then,*

$$\sigma(f) = \{f^{-1}(A) \colon A \in \sigma(\mathcal{G})\} = \sigma(\{f^{-1}(G) \colon G \in \mathcal{G}\})$$

Proof. By Exercise 4.9, we know that $\sigma(f)$ is a σ-field on Ω. Thus, since $\mathcal{G} \subset \sigma(\mathcal{G})$, it follows readily that $\sigma(\{f^{-1}(G) \colon G \in \mathcal{G}\}) \subset \sigma(f)$. With regard to the other inclusion, we note that for any $G \in \mathcal{G}$, $f^{-1}(G) \in \sigma(\{f^{-1}(G) \colon G \in \mathcal{G}\})$. Then, by Proposition 7.1, for any $A \in \sigma(\mathcal{G})$, $f^{-1}(A) \in \sigma(\{f^{-1}(G) \colon G \in \mathcal{G}\})$. That is $\sigma(f) \subset \sigma(\{f^{-1}(G) \colon G \in \mathcal{G}\})$. $\square$

Definition C.2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G}_1, \ldots, \mathcal{G}_n$ be n families of subsets of Ω where for any $i = 1, \ldots, n$, $\mathcal{G}_i \subset \mathcal{F}$, i.e., $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are subsets of measurable sets of Ω. The families $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are said to be independent if*

$$\mathbb{P}(G_{k_1} \cap \cdots \cap G_{k_j}) = \mathbb{P}(G_{k_1}) \cdot \ldots \cdot \mathbb{P}(G_{k_j}),$$

for any choice $G_{k_1} \in \mathcal{G}_{k_1}, \ldots, G_{k_j} \in \mathcal{G}_{k_j}$, $1 \le k_1 < k_2 < \cdots < k_j \le n$, $2 \le j \le n$.

Proposition C.4. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G}_1, \dots, \mathcal{G}_n$ be n families of subsets of Ω where for any $i = 1, \dots, n$, $\mathcal{G}_i \subset \mathcal{F}$. Suppose that $\mathcal{G}_1, \dots, \mathcal{G}_n$ are independent and for any $i = 1, \dots, n$, $\mathcal{G}_i$ is a π-system (cf. Definition B.1). Then, the σ-fields $\sigma(\mathcal{G}_1), \dots, \sigma(\mathcal{G}_n)$ are independent.*

Proof. Define $\mathcal{B}_i = \mathcal{G}_i \cup \{\Omega\}$, $i = 1, \dots, n$. That is, we adjoin the set Ω to each $\mathcal{G}_i$. Then, for any $i = 1, \dots, n$, $\mathcal{B}_i$ is a π-system and $\mathcal{B}_1, \dots, \mathcal{B}_n$ remain independent. We notice that $\sigma(\mathcal{B}_i) = \sigma(\mathcal{G}_i)$ for any $i = 1, \dots, n$. Fix sets $B_2 \in \mathcal{B}_2, \dots, B_n \in \mathcal{B}_n$ and define the class

$$\mathcal{L}_{B_2,\dots,B_n} = \{B \in \mathcal{F} \colon \mathbb{P}(B \cap B_2 \cap \dots \cap B_n) = \mathbb{P}(B) \cdot \mathbb{P}(B_2) \cdot \dots \cdot \mathbb{P}(B_n)\}.$$

Then, $\mathcal{L}_{B_2,\dots,B_n}$ is a λ-system (cf. Definition B.2). Clearly, by Definition C.2, $\Omega \in \mathcal{L}_{B_2,\dots,B_n}$. Further, if $A \in \mathcal{L}_{B_2,\dots,B_n}$, we write $B_2 \cap \dots \cap B_n = C$ and obtain with $\Omega \cap C \setminus (A \cap C) = A^c \cap C$ that $A^c \in \mathcal{L}_{B_2,\dots,B_n}$. Also, if $\{A_i \colon i \in \mathbb{N}\} \subset \mathcal{L}_{B_2,\dots,B_n}$ is disjoint, then with $C = B_2 \cap \dots \cap B_n$

$$\mathbb{P}((\cup_{i\in\mathbb{N}} A_i) \cap C) = \sum_{i\in\mathbb{N}} \mathbb{P}(A_i \cap C) = \mathbb{P}(\cup_{i\in\mathbb{N}} A_i)\mathbb{P}(C).$$

Clearly, $\mathcal{B}_1 \subset \mathcal{L}_{B_2,\dots,B_n}$. Therefore, since $\mathcal{L}_{B_2,\dots,B_n}$ is a λ-system that contains the π-system $\mathcal{B}_1$, it follows by Dynkin's $\pi - \lambda$ theorem (cf. Proposition B.4) that $\sigma(\mathcal{B}_1) = \sigma(\mathcal{G}_1) \subset \mathcal{L}_{B_2,\dots,B_n}$. This shows that the families $\sigma(\mathcal{B}_1), \mathcal{B}_2, \dots, \mathcal{B}_n$ are independent. Then, we recycle the previous argument for the π-systems $\sigma(\mathcal{B}_1), \mathcal{B}_2, \dots, \mathcal{B}_n$ and deduce that

$$\sigma(\mathcal{B}_1), \sigma(\mathcal{B}_2), \mathcal{B}_3, \dots, \mathcal{B}_n$$

are independent. Finally, upon a finite number of iterations, the σ-fields $\sigma(\mathcal{G}_1), \dots, \sigma(\mathcal{G}_n)$ are independent. □

Proof of Proposition 11.6. We define the π-systems

$$\mathcal{G}_1 = \underbrace{\{X_1^{-1}(B_1) \cap \dots \cap X_{n_1}^{-1}(B_{n_1}) \colon B_1 \in \mathfrak{B}(\mathbb{R}^{k_1}), \dots, B_{n_1} \in \mathfrak{B}(\mathbb{R}^{k_{n_1}})\}}_{=\{Y_1^{-1}(B_1\times\dots\times B_{n_1}) \colon B_1\in\mathfrak{B}(\mathbb{R}^{k_1}),\dots,B_{n_1}\in\mathfrak{B}(\mathbb{R}^{k_{n_1}})\}}$$

$$\vdots$$

$$\mathcal{G}_p = \underbrace{\{X_{n_{p-1}+1}^{-1}(B_{n_{p-1}+1}) \cap \dots \cap X_n^{-1}(B_n) \colon B_{n_{p-1}+1} \in \mathfrak{B}(\mathbb{R}^{k_{n_{p-1}+1}}), \dots, B_n \in \mathfrak{B}(\mathbb{R}^{k_n})\}}_{=\{Y_p^{-1}(B_{n_{p-1}+1}\times\dots\times B_n) \colon B_{n_{p-1}+1}\in\mathfrak{B}(\mathbb{R}^{k_{n_{p-1}+1}}),\dots,B_n\in\mathfrak{B}(\mathbb{R}^{k_n})\}}.$$

We notice that since by assumption $X_1, \dots, X_n$ are independent, $\mathcal{G}_1, \dots, \mathcal{G}_p$ are independent according to Definition C.2 (cf. Remark 11.2). By Proposition C.4, $\sigma(\mathcal{G}_1), \dots, \sigma(\mathcal{G}_p)$ are independent as well. Then, we apply Proposition C.3 and conclude that $\sigma(Y_1), \dots, \sigma(Y_p)$ are independent (cf. Definition 11.3). In order to show the remaining part of the proposition, if $f_i \colon \mathbb{R}^{k_{n_i+1}} \times \dots \times \mathbb{R}^{k_{n_{i+1}}} \to \mathbb{R}$, $i = 0, \dots, p-1$, are $\mathfrak{B}(\mathbb{R}^{k_{n_i+1}}) \otimes \dots \otimes \mathfrak{B}(\mathbb{R}^{k_{n_{i+1}}})$ measurable, then for any $B_1, \dots, B_p \in \mathfrak{B}(\mathbb{R})$, since $Y_1, \dots, Y_p$ are independent,

$$\begin{aligned}
\mathbb{P}(T_1 \in B_1, \dots, T_p \in B_p) &= \mathbb{P}(Y_1 \in f_1^{-1}(B_1), \dots, Y_p \in f_p^{-1}(B_p)) \\
&= \mathbb{P}(Y_1 \in f_1^{-1}(B_1)) \cdot \dots \cdot \mathbb{P}(Y_p \in f_p^{-1}(B_p)) \\
&= \mathbb{P}(T_1 \in B_1) \cdot \dots \cdot \mathbb{P}(T_p \in B_p).
\end{aligned}$$

Hence, the random variables $T_1, \dots, T_p$ are independent. □

C.6 On the law of a Gauss vector

Proposition C.5. *Suppose that $\Sigma \in \mathbb{R}^{k\times k}$ is symmetric and positive semidefinite. Then, there exists a symmetric and positive semidefinite matrix B s.t. $BB = \Sigma$.*

Proof. Since Σ is symmetric there exists an orthogonal matrix Q ($Q^tQ = I$) s.t. $\Sigma = QDQ^t$, where D is diagonal s.t. $D_{i,i} = \lambda_i$, where λ_i is an eigenvalue of Σ, $i = 1, \ldots, k$. Since Σ is positive semidefinite, we define $B = QD^{1/2}Q^t$, with

$$D^{1/2} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \ldots & 0 \\ 0 & \sqrt{\lambda_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sqrt{\lambda_k} \end{pmatrix}.$$

Then, we obtain $BB = QD^{1/2}Q^tQD^{1/2}Q^t = \Sigma$. □

Remark C.3. *The matrix B in Proposition C.5 is referred to as a square root of Σ. It is denoted with $\sqrt{\Sigma}$.*

Remark C.4. *If $\Sigma \in \mathbb{R}^{k\times k}$ is symmetric and positive semidefinite, then, $\det\sqrt{\Sigma} = \sqrt{\det\Sigma}$. To see it, notice that $\det\Sigma = \det\sqrt{\Sigma}\sqrt{\Sigma} = \det\sqrt{\Sigma}\det\sqrt{\Sigma} = (\det\sqrt{\Sigma})^2$.*

Proposition C.6. *Let $\Sigma \in \mathbb{R}^{k\times k}$ be symmetric and positive semidefinite and $\mu \in \mathbb{R}^k$. Then, the random vector $X = \mu + \sqrt{\Sigma}N$, with $N \sim \mathcal{N}(0, I)$, is s.t. $X \sim \mathcal{N}(\mu, \Sigma)$.*

Proof. Let $v \in \mathbb{R}^k$. We apply Proposition 11.9 and deduce that

$$\begin{aligned}
\Phi_X(v) &= \mathbb{E}[\mathrm{e}^{iv^tX}] \\
&= \mathbb{E}[\mathrm{e}^{iv^t\mu + iv^t\sqrt{\Sigma}N}] \\
&= \mathrm{e}^{iv^t\mu}\,\mathbb{E}[\mathrm{e}^{iv^t\sqrt{\Sigma}N}] \\
&= \mathrm{e}^{iv^t\mu}\,\mathbb{E}[\mathrm{e}^{i((v^t\sqrt{\Sigma})^t)^tN}] \\
&= \mathrm{e}^{iv^t\mu}\,\Phi_N((v^t\sqrt{\Sigma})^t) = \mathrm{e}^{iv^t\mu}\,\mathrm{e}^{-\frac{v^t\sqrt{\Sigma}I(v^t\sqrt{\Sigma})^t}{2}} = \mathrm{e}^{iv^t\mu}\,\mathrm{e}^{-\frac{v^t\sqrt{\Sigma}\sqrt{\Sigma}^tv}{2}} = \mathrm{e}^{iv^t\mu}\,\mathrm{e}^{-\frac{v^t\Sigma v}{2}}.
\end{aligned}$$

Hence, by Proposition 10.12, $X \sim \mathcal{N}(\mu, \Sigma)$. □

Remark C.5. *Notice that if $N_1, \ldots, N_k$ are k independent random variables s.t. for any $i = 1, \ldots, k$, $N_i \sim \mathcal{N}(0, 1)$, then $N = (N_1, \ldots, N_k)$ is a Gauss vector (cf. item (iii) of Remark 11.11). In particular, $N \sim \mathcal{N}(0, I)$. Thus, the latter proposition shows how to construct (given appropriate μ and Σ) a Gauss vector $X \sim \mathcal{N}(\mu, \Sigma)$, upon a collection of independent standard Gaussian random variables.*

Proof of Proposition 11.11. In order to simplify the notation we set $\Sigma = \Sigma(X)$. Define $\widetilde{X} = \mu + \sqrt{\Sigma}N$, with $N \sim \mathcal{N}(0, I)$. By Proposition C.6, X and $\widetilde{X}$ have the same law. Thus, we may omit the distinction and let $X = \widetilde{X}$. Define $T(w) = \mu + \sqrt{\Sigma}w$, $w \in \mathbb{R}^k$. By Example A.6, $T\colon \mathbb{R}^k \to \mathbb{R}^k$ is differentiable with Jacobian matrix $J_T(w_0) = \sqrt{\Sigma}$ for any $w_0 \in \mathbb{R}^k$ (cf. Proposition A.24). Since $\sqrt{\Sigma}$ is invertible, $\det\sqrt{\Sigma} \neq 0$. Further, $T(\mathbb{R}^k) = \mathbb{R}^k$ since for any $y \in \mathbb{R}^k$, there exists a unique $w_0 \in \mathbb{R}^k$ which solves $T(w_0) = y$. In particular, $T\colon \mathbb{R}^k \to \mathbb{R}^k$ is bijective. Let $f\colon \mathbb{R}^k \to \mathbb{R}$ be any nonnegative and $\mathfrak{B}(\mathbb{R}^k)$ measurable function. We have with Proposition 11.1,

$$\begin{aligned}
\mathbb{E}[f(X)] &= \mathbb{E}[f(T(N)] \\
&= \int_{\mathbb{R}^k} f(T(w))P_{N_1} \otimes \cdots \otimes P_{N_1}(dw) \\
&= \frac{1}{\sqrt{(2\pi)^k}} \int_{\mathbb{R}^k} f(T(w))\,\mathrm{e}^{-\frac{w^tw}{2}}\,dw \\
&= \frac{1}{\sqrt{(2\pi)^k}} \frac{1}{\det\sqrt{\Sigma}} \int_{\mathbb{R}^k} f(T(w))\det\sqrt{\Sigma}\,\mathrm{e}^{-\frac{w^tw}{2}}\,dw.
\end{aligned}$$

Define

$$g(y) = f(y)\,\mathrm{e}^{-\frac{(y-\mu)^t\Sigma^{-1}(y-\mu)}{2}}, \quad y \in \mathbb{R}^k.$$

Then, for any $w \in \mathbb{R}^k$, $g(T(w)) = f(T(w))\,\mathrm{e}^{-(w^t w)/2}$. Notice that $\Sigma^{-1} = (\sqrt{\Sigma}\sqrt{\Sigma})^{-1} = \sqrt{\Sigma}^{-1}\sqrt{\Sigma}^{-1}$. Therefore, by Proposition 9.9, we obtain:

$$\begin{aligned}\mathbb{E}[f(X)] &= \frac{1}{\sqrt{(2\pi)^k}}\frac{1}{\det\sqrt{\Sigma}}\int_{\mathbb{R}^k} g(y)dy\\ &= \frac{1}{\sqrt{(2\pi)^k}}\frac{1}{\det\sqrt{\Sigma}}\int_{\mathbb{R}^k} f(y)\,\mathrm{e}^{-\frac{(y-\mu)^t\Sigma^{-1}(y-\mu)}{2}}\,dy.\end{aligned}$$

Finally, we rely on Remark C.4 and the proposition is proven. □

C.7 On Kolmogorov's Zero-One law and the law of large numbers

We remain in the setting of Section 11.

Definition C.3. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables. The σ-field generated by the latter sequence is defined as*

$$\sigma((X_n)_{n\in\mathbb{N}}) = \sigma(\{(X_1,\ldots,X_k)^{-1}(B)\colon B \in \mathfrak{B}(\mathbb{R}^k),\ k \in \mathbb{N}\}).$$

Proposition C.7. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables. We have that*

$$\sigma((X_n)_{n\in\mathbb{N}}) = \sigma\left(\bigcup_{k=1}^{\infty}\sigma(Y_k)\right), \quad Y_k = (X_1,\ldots,X_k), \quad k \in \mathbb{N}.$$

Proof. Let $A \in \{(X_1,\ldots,X_k)^{-1}(B)\colon B \in \mathfrak{B}(\mathbb{R}^k),\ k \in \mathbb{N}\}$. Then, there exists $k \in \mathbb{N}$ and $B \in \mathfrak{B}(\mathbb{R}^k)$ s.t. $A = Y_k^{-1}(B)$, $Y_k = (X_1,\ldots,X_k)$. By definition of $\sigma(Y_k)$, it follows that $A \in \sigma(Y_k)$. In particular, $A \in \sigma(\cup_{k=1}^{\infty}\sigma(Y_k))$. This shows that $\sigma((X_n)_{n\in\mathbb{N}}) \subset \sigma(\cup_{k=1}^{\infty}\sigma(Y_k))$. Let $A \in \cup_{k=1}^{\infty}\sigma(Y_k)$, then there exists $k \in \mathbb{N}$ s.t. $A \in \sigma(Y_k)$. Thus, there exists $B \in \mathfrak{B}(\mathbb{R}^k)$ s.t. $A = Y_k^{-1}(B)$, $Y_k = (X_1,\ldots,X_k)$. This shows that $A \in \sigma((X_n)_{n\in\mathbb{N}})$. Therefore, $\sigma(\cup_{k=1}^{\infty}\sigma(Y_k)) \subset \sigma((X_n)_{n\in\mathbb{N}})$. □

Proposition C.8. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent random variables (cf. Definition 11.4). Let $Y_k = (X_1,\ldots,X_k)$, $k \in \mathbb{N}$. Then, for any $k \in \mathbb{N}$, $\sigma(Y_k)$ is independent of $\sigma((X_{k+j})_{j\in\mathbb{N}})$.*

Proof. Since $(X_n)_{n\in\mathbb{N}}$ is an independent sequence of random variables, it follows from Proposition 11.6 that for any $j \in \mathbb{N}$, $\sigma(Y_k)$ is independent of $\sigma((X_{k+1},\ldots,X_{k+j}))$. In particular, $\sigma(Y_k)$ is independent of $\cup_{j=1}^{\infty}\sigma((X_{k+1},\ldots,X_{k+j}))$. Since the latter union is a π-system, we apply Proposition C.4 and get that $\sigma(Y_k)$ is independent of $\sigma(\cup_{j=1}^{\infty}\sigma((X_{k+1},\ldots,X_{k+j})))$. Then, we apply Proposition C.7 and the proposition is proven. □

Kolmogorov's Zero-One law (for sequences of random variables) reads as follows:

Proposition C.9. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent random variables. Define the σ-field,*

$$\mathcal{F}_\infty = \bigcap_{k=0}^{\infty}\sigma((X_{k+j})_{j\in\mathbb{N}}).$$

Then, for any $A \in \mathcal{F}_\infty$, $\mathbb{P}(A)$ is either zero or one.

Proof. By Proposition C.8, for any $k \in \mathbb{N}$, $\sigma((X_{k+j})_{j\in\mathbb{N}})$ is independent of $\sigma(Y_k)$. In particular, since $\mathcal{F}_\infty \subset \sigma((X_{k+j})_{j\in\mathbb{N}})$ for any $k \in \mathbb{N}$, we deduce that $\sigma(Y_k)$ is independent of $\mathcal{F}_\infty$ for any $k \in \mathbb{N}$. Thus, we apply Propositions C.4 and C.7 and deduce that $\mathcal{F}_\infty$ is independent of $\sigma((X_n)_{n\in\mathbb{N}})$. Let $A \in \mathcal{F}_\infty$, then, $A \in \sigma((X_n)_{n\in\mathbb{N}})$ and we get that

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2,$$

since $\mathcal{F}_\infty$ is independent of $\sigma((X_n)_{n\in\mathbb{N}})$. This shows that $\mathbb{P}(A)$ is either zero or one. □

Proof of Proposition 12.9. Let $S_0(\omega) = 0$ for any $\omega \in \Omega$ and set $S_n = \sum_{i=1}^n X_i$, $n \in \mathbb{N}$. Consider $a \in \mathbb{R}$ s.t. $a > \mathbb{E}[X_1]$. Define $M_a = \sup_{n\in\mathbb{Z}_+}(S_n - na)$, where $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$. To simplify the notation, we write $M = M_a$. By Proposition 7.9, M is a random variable, i.e., M is $\mathcal{F}$ measurable. Notice that $M = \max\{\sup_{n\in\mathbb{N}}(S_n - na), 0\}$. In particular, M is nonnegative. We aim to show that

$$\mathbb{P}(M < \infty) = 1. \tag{83}$$

Let $k \in \mathbb{N}$. We observe that (cf. Propositions 8.5 and 8.6),

$$\{M < \infty\} = \{\sup_{n\in\mathbb{Z}_+}(S_n - na) < \infty\} = \{\sup_{n\geq k}(S_n - S_k - na) < \infty\}.$$

Then, the event $\{\sup_{n\geq k}(S_n - S_k - na) < \infty\}$ is an element of $\sigma((X_{k+j})_{j\in\mathbb{N}})$ (cf. Proposition 7.13). Since $k \in \mathbb{N}$ was arbitrary, we conclude that $\{M < \infty\} \in \mathcal{F}_\infty$, where $\mathcal{F}_\infty$ is as in Proposition C.9. That is, $\mathbb{P}(M < \infty)$ is either zero or one. To show (83), it is therefore sufficient to show that $\mathbb{P}(M < \infty) > 0$. We show that,

$$\mathbb{P}(M = \infty) < 1, \tag{84}$$

from which $\mathbb{P}(M < \infty) > 0$ follows. Given any $m \in \mathbb{N}$, we define the random variables:

$$\begin{aligned} M_m &= \sup_{0\leq n\leq m}(S_n - na), \\ M'_m &= \sup_{0\leq n\leq m}(S_{n+1} - S_1 - na). \end{aligned}$$

We notice that for any $n \in \mathbb{N}$, $n \geq 2$, since $(X_i)_{i\in\mathbb{N}}$ is an i.i.d. sequence of random variables, the law of $(X_1, \ldots, X_n)$ is the same as the law of $(X_2, \ldots, X_{n+1})$ (recall also Proposition 11.1). Thus, upon Proposition 10.3, for any $m \in \mathbb{N}$, M_m and M'_m have the same law. We further observe that for any $m \in \mathbb{N}$ and $\omega \in \Omega$, $M_m(\omega) \leq M_{m+1}(\omega)$ and $M'_m(\omega) \leq M'_{m+1}(\omega)$. Clearly, $M = \lim_{m\to\infty} M_m$ and we define $M' = \lim_{m\to\infty} M'_m$. Given any $x \in \mathbb{R}$,

$$\mathbb{P}(M' \leq x) = \mathbb{P}\left(\bigcap_{m=1}^{\infty}\{M'_m \leq x\}\right) = \lim_{m\to\infty}\mathbb{P}(M'_m \leq x) = \lim_{m\to\infty}\mathbb{P}(M_m \leq x) = \mathbb{P}(M \leq x).$$

Thus, M and M' have the same law. In addition, we verify that

$$M_{m+1} = M'_m - \min\{a - X_1, M'_m\}.$$

Notice that if $M'_m(\omega) \leq a - X_1(\omega)$, it follows that $M_{m+1}(\omega) = 0$ (recall that M_{m+1} is nonnegative). Given the previous display, we deduce that

$$\mathbb{E}[\min\{a - X_1, M'_m\}] = \mathbb{E}[M'_m] - \mathbb{E}[M_{m+1}] = \mathbb{E}[M_m] - \mathbb{E}[M_{m+1}] \leq 0.$$

Then, since for any $\omega \in \Omega$, $\min\{a - X_1, M'_m\}(\omega) \leq (a - X_1)(\omega) \leq |a - X_1|(\omega)$, and $\mathbb{E}[|a - X_1|] < \infty$, we apply Lebesgue's dominated convergence theorem (cf. Proposition 8.10) and deduce that

$$\mathbb{E}[\min\{a - X_1, M'\}] = \lim_{m\to\infty}\mathbb{E}[\min\{a - X_1, M'_m\}] \leq 0. \tag{85}$$

Suppose now by contradiction that $\mathbb{P}(M = \infty) = 1$. Then, since M' and M have the same law, we must have that $\mathbb{P}(M' = \infty) = 1$. Thus, with $\mathbb{P}$ probability one, $\min\{a - X_1, M'\} = a - X_1$. Thus, by (85), $\mathbb{E}[a - X_1] \leq 0$. This gives a contradiction with the assumption that $a > \mathbb{E}[X_1]$. Hence, we have verified (84) and thus (83) follows. To complete the proof, let $a_n = \mathbb{E}[X_1] + 1/n$, $n \in \mathbb{N}$. Then, for any $n \in \mathbb{N}$, $S_n \leq na_n + M_{a_n}$ and by (83) $\mathbb{P}$ a.s.,

$$\limsup_{n\to\infty} \frac{S_n}{n} \leq \limsup_{n\to\infty} \left(\frac{na_n + M_{a_n}}{n} \right) = \mathbb{E}[X_1].$$

Define $\widetilde{X}_i = -X_i$, $i \in \mathbb{N}$, and $\widetilde{S}_n = \sum_{i=1}^n \widetilde{X}_i$. We repeat the previous arguments and conclude that with $\mathbb{P}$ probability one,

$$\limsup_{n\to\infty} \frac{\widetilde{S}_n}{n} \leq \mathbb{E}[\widetilde{X}_1] = -\mathbb{E}[X_1].$$

Therefore, $\liminf_{n\to\infty} \frac{S_n}{n} \geq \mathbb{E}[X_1]$ $\mathbb{P}$ a.s. In conclusion,

$$\mathbb{P}\left(\liminf_{n\to\infty} \frac{S_n}{n} = \mathbb{E}[X_1] = \limsup_{n\to\infty} \frac{S_n}{n} \right) = 1,$$

and the proposition is proven. □

References

[1] Patrick Billingsley, *Probability and measure*, 4 ed., Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, NJ, 2012.

[2] Jean-François Le Gall, *Measure theory, probability, and stochastic processes*, Graduate Texts in Mathematics, vol. 295, Springer, Cham, 2022.