Lecture Notes
Measure Theory

Winter semester 2021/22

Friedrich-Schiller-Universität Jena

Michael H. Neumann

Important information

- Lecture period: October 18 – February 11

- Examination period: February 14 – March 11
  There will be oral examinations, hopefully in face-to-face mode and within the official examination period. Dates for these examinations will be fixed in good time.

- Contact
  - Office: room 3516
  - Phone: 03641/946260
  - Email: michael.neumann@uni-jena.de

Literature

[1] Billingsley, P. (1986). *Probability and Measure*, 2nd ed. Wiley, New York.

[2] Cohn, D. L. (1997). *Measure Theory*. Birkhäuser, Boston.

[3] Elstrodt, J. (2011). *Maß- und Integrationstheorie*. 7nd Edition. Springer, Heidelberg. (in German)

Contents

0 A few problems to be solved

This course is intended as an introduction to the parts of measure theory necessary for analysis and probability. Here are a few topics that will be treated in this course.

1) *Construction of (probability) measures*

It should be conjectured that most if not all students of this course are familiar with basic concepts of probability theory. Suppose we want to construct a mathematical model (that is, a probability space) for an experiment where each trial consists of tossing a coin infinitely often. It is natural to assume that the outcomes of these coin tosses are independent. (Of course, such an experiment cannot be carried out in practice. It is merely of interest for theoretical considerations.) If we denote heads/tails with 1/0, then $\Omega := \big\{\omega = (\omega_1, \omega_2, \ldots)\colon\ \omega_i \in \{0,1\}\big\}$ describes the set of possible outcomes of such a random experiment. For some subsets of Ω, it is quite easy to determine the corresponding probabilities. For example, let A be the subset of Ω which describes the event that there are exactly k heads during the first n coin tosses. Then

$$A = \big\{(\omega_1, \omega_2, \ldots)\colon\ \omega_1 + \cdots + \omega_n = k\big\}.$$

It should be intuitively clear that the probability of this event is $P(A) = \binom{n}{k} 2^{-n}$. At this point we may ask how the probability of an arbitrary subset A of Ω can be determined. This is well possible as long as A describes the outcome of finitely many coin tosses. If $A = \big\{\omega \in \Omega\colon\ (\omega_1, \ldots, \omega_n) \in C\big\}$, for some $C \in \{0,1\}^n$, then $P(A) = \#C/2^n$, where $\#B$ denotes the cardinality of a generic set B. On the other hand, if the occurrence of A depends on infinitely many coin tosses, then things get much more delicate. It is of course desirable to assign a probability to *all* subsets of Ω. This seems to be difficult, at least from a practical point of view, since Ω also contains subsets of a complicated structure such that their formal descriptions seems to be difficult or even impossible. And, even worse, it can be shown that we cannot define an appropriate probability measure P on *all* subsets of Ω. To see this, consider the mappings $T_n\colon\ \Omega \to \Omega$, $n \in \mathbb{N}$, where

$$T_n\big((\omega_1, \ldots, \omega_{n-1}, \omega_n, \omega_{n+1}, \ldots)\big) \;=\; (\omega_1, \ldots, \omega_{n-1}, 1 - \omega_n, \omega_{n+1}, \ldots) \qquad \forall n \in \mathbb{N}.$$

For $A \subseteq \Omega$, we define $T_n(A) := \big\{T_n(\omega)\colon\ \omega \in A\big\}$. For reasons of symmetry, if there exists a probability measure P which assigns probabilities to all subsets of Ω, then P should satisfy

$$P\big(A\big) \;=\; P\big(T_n(A)\big) \tag{0.0.1}$$

for all $A \in 2^\Omega$ and for all $n \in \mathbb{N}$. (2^Ω denotes the set of all subsets of Ω, the so-called power set.) The following lemma shows that such a probability measure does *not* exist.

Lemma 0.0.1. *A probability measure $P\colon\ 2^\Omega \to [0,1]$ such that (0.0.1) holds true for all $A \in 2^\Omega$ and all $n \in \mathbb{N}$ does* not *exist.*

Proof. We prove this lemma by contradiction. Suppose that a probability measure $P\colon\ 2^\Omega \to [0,1]$ satisfying (0.0.1) for all $A \in 2^\Omega$ and all $n \in \mathbb{N}$ does exist. In what follows we split the set Ω into countably many disjoint subsets of equal probability. This will give the desired contradiction.

The mappings T_n, $n \in \mathbb{N}$, induce the following equivalence relation on Ω. Two elements ω and ω' are said to be equivalent, $\omega \sim \omega'$, if and only if there exist $n_1, \ldots, n_k \in \mathbb{N}$ and $k \in \mathbb{N}$ such that

$$\omega = T_{n_1} \circ \cdots \circ T_{n_k}(\omega').$$

In other words, two elements of Ω are equivalent if they differ in only finitely many terms. It is easy to see that $\sim$ is an equivalence relation on Ω, i.e. the properties of reflexivity ($\omega \sim \omega$ holds for each ω), symmetry ($\omega \sim \omega'$ implies $\omega' \sim \omega$), and transitivity ($\omega \sim \omega'$ and $\omega' \sim \omega''$ imply $\omega \sim \omega''$) are fulfilled. We choose from each equivalence class exactly one element and we denote the set of these elements by A. Furthermore, let

$$\mathcal{T} := \left\{ T_{n_1} \circ \cdots \circ T_{n_k} : \quad n_1, \ldots, n_k \in \mathbb{N},\ k \in \mathbb{N} \right\}$$

be the collection of finite compositions of our mappings T_n. Then

a) $\mathcal{T}$ is countably infinite,

b) $\bigcup_{T \in \mathcal{T}} T(A) = \Omega$,

c) if $S, T \in \mathcal{T}$ and $S \neq T$, then $S(A) \cap T(A) = \emptyset$.
(Suppose that the opposite holds true, that is $S, T \in \mathcal{T}$, $S \neq T$ and $\omega \in S(A) \cap T(A)$ for some ω. Then there exist $\omega_S, \omega_T \in A$ such that $\omega = S(\omega_S) = T(\omega_T)$. This implies that $\omega_T = T \circ S(\omega_S)$, i.e. $\omega_S \sim \omega_T$. Since the set A contains exactly one representative from each equivalence class we obtain that $\omega_S = \omega_T$. This leads to $\omega = S(\omega_S) = T(\omega_S)$ which contradicts $S \neq T$.)

d) for $T = T_{n_1} \circ \cdots \circ T_{n_k} \in \mathcal{T}$,

$$P\big(T(A)\big) = P\big(T_{n_1} \circ \cdots \circ T_{n_k}(A)\big) = P\big(T_{n_1} \circ \cdots \circ T_{n_{k-1}}(A)\big) = \ldots = P(A).$$

We obtain from b) to d) that

$$1 = P(\Omega) = P\Big(\bigcup_{T \in \mathcal{T}} T(A) \Big) = \sum_{T \in \mathcal{T}} P\big(T(A)\big) = \sum_{T \in \mathcal{T}} P(A).$$

This is however impossible since $\mathcal{T}$ is according to a) countably infinite. □

The above example shows that we have to be careful when we intend to construct probability measures or, more generally, measures on uncountable spaces Ω. We will see in Chapter 1 how these difficulties can be overcome. In fact, it will be only necessary to provide an explicit specification of $P(A)$ for sets A with a simple structure. This will suffice to specify a probability measure P on a well-structured family $\mathcal{A}$ of subsets of Ω. As we have seen above, sometimes such a family $\mathcal{A}$ cannot contain all subsets of Ω. However, it will be rich enough for all “practical” purposes.

2) *An extension of the concept of Riemann integrals*

We begin by recalling the (hopefully well-known) definition of the Riemann integral which is named after the German mathematician Georg Friedrich Bernhard Riemann. For $a, b \in \mathbb{R}$, $a < b$, let $[a, b]$ be a closed bounded interval. A **partition** $\mathcal{P}$ of $[a, b]$ is a finite sequence $(a_i)_{i=0,\dots,n}$ of real numbers such that

$$a = a_0 < a_1 < \ldots < a_n = b.$$

Let f be a bounded real-valued function on $[a, b]$. If $\mathcal{P}$ is the partition $(a_i)_{i=0,\dots,n}$ of $[a, b]$, then the **lower sum** $l(f, \mathcal{P})$ corresponding to f and $\mathcal{P}$ is defined to be $\sum_{i=1}^{n} \inf\big\{f(x)\colon\ x \in [a_{i-1}, a_i]\big\}(a_i - a_{i-1})$. Likewise we define the **upper sum** $u(f, \mathcal{P})$ corresponding to f and $\mathcal{P}$ as $\sum_{i=1}^{n} \sup\big\{f(x)\colon\ x \in [a_{i-1}, a_i]\big\}(a_i - a_{i-1})$. Now we define the **lower integral** $\underline{\int}_a^b f(x)\, dx$ of f over $[a, b]$ as the supremum of the lower sums and the **upper integral** $\overline{\int}_a^b f(x)\, dx$ of f over $[a, b]$ as the infimum of the upper sums. It follows immediately that $\underline{\int}_a^b f(x)\, dx \leq \overline{\int}_a^b f(x)\, dx$. If $\underline{\int}_a^b f(x)\, dx = \overline{\int}_a^b f(x)\, dx$, then f is **Riemann integrable** on $[a, b]$, and the common value of $\underline{\int}_a^b f(x)\, dx$ and $\overline{\int}_a^b f(x)\, dx$ is called the **Riemann integral** of f over $[a, b]$ and is denoted by $\int_a^b f(x)\, dx$. It is well-known that a continuous real-valued function f is Riemann integrable over each bounded interval $[a, b]$. There are, however, several annoying deficiencies of this concept. Here are a few of them:

a) *Non-integrability of certain simple functions*

The Dirichlet function is the indicator function $\mathbb{1}_{\mathbb{Q}}$ of the set $\mathbb{Q}$ of rational numbers, i.e. $\mathbb{1}_{\mathbb{Q}}(x) = 1$ if x is a rational number and $\mathbb{1}_{\mathbb{Q}}(x) = 0$ if x is not a rational number. It is named after the German mathematician Peter Gustav Lejeune Dirichlet. (Although his surname Lejeune Dirichlet sounds French, he was born in Düren and grew up in Bonn and Cologne.) If $a < b$, then it is obvious that, for any arbitrary partition $\mathcal{P}$ of $[a, b]$, $l(\mathbb{1}_{\mathbb{Q}}, \mathcal{P}) = 0$ whereas $u(\mathbb{1}_{\mathbb{Q}}, \mathcal{P}) = b - a$. Hence, $\underline{\int}_a^b \mathbb{1}_{\mathbb{Q}}(x)\, dx = 0$ and $\overline{\int}_a^b \mathbb{1}_{\mathbb{Q}}(x)\, dx = b - a$. Although having a simple structure, the function $\mathbb{1}_{\mathbb{Q}}$ is not integrable in the Riemannian sense.

b) *Non-compatibility with limiting operations*

Let $(q_n)_{n\in\mathbb{N}}$ be any enumeration of the rational numbers. Define

$$D_n(x) = \begin{cases} 1 & \text{if } x \in \{q_1, \dots, q_n\}, \\ 0 & \text{otherwise.} \end{cases}$$

Then $\int_0^1 D_n(x)\, dx = 0$, $D_n(x) \underset{n\to\infty}{\longrightarrow} \mathbb{1}_{\mathbb{Q}}(x)$ for all $x \in \mathbb{R}$, but

$$\int_0^1 D_n(x)\, dx \not\to \int_0^1 \mathbb{1}_{\mathbb{Q}}(x)\, dx \qquad \text{as } n \to \infty.$$

c) *Restriction to subsets of* $\mathbb{R}^d$ *as a possible domain of integration*

The notion of the Riemann integral is restricted to certain subsets of $\mathbb{R}$ or $\mathbb{R}^d$ as possible domains of integrations. However, in some cases a wider concept seems to be desirable. For example, suppose that $X\colon\ \Omega \to [0,\infty)$ is a non-negative random variable which is defined on a probability space $(\Omega, \mathcal{A}, P)$. If Ω is finite or countably infinite, then the expectation of X under P is defined and can be expressed as $EX = \sum_{\omega\in\Omega} X(\omega)P\big(\{\omega\}\big)$. With the notion of the Lebesgue integral to be defined in Chapter 2, it can alternatively (and more conveniently) be expressed by such an integral, $\int_\Omega X\,dP$. Here, Ω need not be a subset of some Euclidean space $\mathbb{R}^d$. In fact, any non-empty set Ω is possible.

All of these weaknesses above will be healed by the more general concept of the Lebesgue integral. It will also be shown that the Riemann and Lebesgue integrals coincide if the former integral exists. In this sense, the concept of the Lebesgue integral is an extension but not a redefinition of the Riemann integral.

3) *Conditional distributions*

In any elementary course on probability theory the concept of an (elementary) conditional probability is introduced. If X and Y are random variables on a probability space $(\Omega, \mathcal{A}, P)$, then the conditional probability of the event that $X \in C$ given $Y = y$ is defined by

$$P\big(X \in C \mid Y = y\big) = \begin{cases} \frac{P(X\in C,\, Y=y)}{P(Y=y)} & \text{if } P(Y = y) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

This definition is certainly good enough if Y is a discrete random variable which takes their values in a finite or countably infinite set Ω_Y. In this case, $P^Y\big(\{y\colon\ P(Y = y) = 0\}\big) = 0$, i.e. the meaningless second part of the above definition is not relevant. On the other hand, if for example Y is normally distributed, then $P^Y\big(\{y\colon\ P(Y = y) = 0\}\big) = 1$. Then the above definition leads to $P(X \in C \mid Y = y) = 0$ for *all* $y \in \mathbb{R}$. In this case, the above definition is no longer meaningful and another concept to overcome this deficiency is in order. Based on results in connection with Lebesgue integrals, we also generalize in Section 2.8 the concept of conditional probabilities and distributions. In the case where Y follows a discrete distribution, this alternative definition of conditional probabilities is equivalent to the simple one shown above. Hence, this is also an extension but not a redefinition of the more elementary concept mentioned above.

1 Construction of measures on general spaces

This chapter is devoted to the construction of measures on general spaces Ω. Measures are always defined on well-structured collections of subsets of Ω, so-called σ-algebras. These and similar objects will be introduced in the first subsection of this chapter. Afterwards we turn to the construction of measures on arbitrary spaces. This will include in particular the so-called Lebesgue measure.

1.1 Classes of sets

Before we provide an exact definition of collections of sets on which measures will be defined, we want to suggest some structural properties of these systems. Suppose that we want to find a model for a random experiment where a random quantity X takes values in $\mathbb{R}$ or in some subset of $\mathbb{R}$. In particular, we want to find a function P which describes the probabilities that X takes its value in certain subsets of $\mathbb{R}$. In this case, $\mathbb{R}$ takes the role of our basic space Ω. We obtain some of these probabilities almost for free. For example, the probability that $X \in \emptyset$ is zero since it is impossible that X does not take any value in $\mathbb{R}$. Furthermore, it is sure that $X \in \mathbb{R}$ which means that we have to assign the probability of one to $\mathbb{R}$. (In mathematics, probabilities are always given as numbers between 0 and 1.) But we have even more. If $P(A)$ denotes the probability that X takes any value in the set A, then it is intuitively clear that the probability that X takes a value in the complement A^c of A is $1 - P(A)$. Moreover, if $A_1, A_2, \ldots$ are disjoint subsets of $\mathbb{R}$ and if we know the corresponding probabilities $P(A_1), P(A_2), \ldots$, then we can conclude that the probability that X takes its value in any of the sets $A_1, A_2, \ldots$ is equal to $\sum_{i=1}^{\infty} P(A_i)$. This suggests that probability measures, and more generally arbitrary measures as well, can be defined on collections of sets which contain the empty set $\emptyset$ and the complete space Ω, are closed under complementation and under the formation of unions of disjoint sets. These considerations suggest the following definition of a well-structured system of sets which is named after the Russian-born mathematician Eugene Borisovich Dynkin who emigrated to the United States in 1977.

Definition. Let Ω be a nonempty set, and let $\mathcal{D}$ be a collection of subsets of Ω. Then $\mathcal{D}$ is a **Dynkin system** (sometimes also referred to as λ-system) on Ω if

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

(ii) $\mathcal{D}$ is closed under complementation in Ω, i.e. if $A \in \mathcal{D}$, then $A^c := \Omega \setminus A \in \mathcal{D}$,

(iii) $\mathcal{D}$ is closed under the formation of countable unions of pairwise disjoint sets, i.e. if $A_1, A_2, \ldots$ is a sequence of subsets in $\mathcal{D}$ such that $A_i \cap A_j = \emptyset$ for all $i \neq j$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{D}$.

In probability theory, one is often interested in the probability that all events $A_1, \ldots, A_n$ of a certain finite collection occur, i.e. the probability of $\bigcap_{i=1}^{n} A_i$ has to be determined. It can be shown that a Dynkin-system $\mathcal{D}$ which is closed under the formation of finite intersections contains all countable unions of (not necessarily pairwise disjoint) sets $A_1, A_2, \ldots \in \mathcal{D}$. Actually, this follows from

$$\bigcup_{i=1}^{\infty} A_i \;=\; A_1 \cup \bigcup_{i=2}^{\infty} \big(A_i \setminus (A_1 \cup \cdots \cup A_{i-1})\big) \;=\; A_1 \cup \bigcup_{i=2}^{\infty} \big(A_1^c \cap \cdots \cap A_{i-1}^c \cap A_i\big).$$

In view of this, a collection of sets with these structural properties is usually considered to be the "golden standard". We formalize this by the next definition.

Definition. Let Ω be a nonempty set, and let $\mathcal{A}$ be a collection of subsets of Ω. Then $\mathcal{A}$ is a **σ-algebra** (also σ-field) on Ω if

(i) $\Omega \in \mathcal{A}$,

(ii) $\mathcal{A}$ is closed under complementation in Ω, i.e. if $A \in \mathcal{A}$, then $A^c := \Omega \setminus A \in \mathcal{A}$,

(iii) $\mathcal{A}$ is closed under the formation of countable unions of sets, i.e. if $A_1, A_2, \ldots$ is a sequence of subsets in $\mathcal{A}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

The pair $(\Omega, \mathcal{A})$ is called a **measurable space**. A subset A of Ω which belongs to $\mathcal{A}$ is called **$\mathcal{A}$-measurable** or, if it is clear which σ-algebra is meant, simply measurable.

We turn to a few examples. Let Ω be a non-empty set.

1) The power set $2^{\Omega} = \{A\colon\ A \subseteq \Omega\}$ is the largest σ-algebra on Ω.

2) $\{\emptyset, \Omega\}$ is the smallest σ-algebra on Ω.

3) For $A \subseteq \Omega$, $\{\emptyset, A, A^c, \Omega\}$ is the smallest σ-algebra on Ω which contains the set A.

4) A Dynkin system $\mathcal{D}$ on Ω need not be a σ-algebra. Here is a simple (toy) example: Let $\Omega = \{1, 2, \ldots, 2n\}$ $(n \geq 2)$ and let $\mathcal{D}_e := \{A \in 2^{\Omega}\colon\ \#A = 2k \text{ for some } k \in \{0, 1, \ldots, n\}\}$. ($D_e$ is the collection of subsets of Ω which contain an even number of elements.)
Then $\mathcal{D}_e$ is obviously a Dynkin-system on Ω. However, $A_1 = \{1, 2\} \in \mathcal{D}_e$, $A_2 = \{2, 3\} \in \mathcal{D}_e$, but $A_1 \cap A_2 = \{2\}$ is not contained in $\mathcal{D}_e$. Therefore, $\mathcal{D}_e$ is not a σ-algebra on Ω.

The following lemma collects a few simple but useful properties of σ-algebras.

Lemma 1.1.1. *Let Ω be a nonempty set and let $\mathcal{A}$ be a σ-algebra on Ω. Then*

(i) If $A_1, A_2, \ldots \in \mathcal{A}$, then $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$.

(ii) If $A_1, \ldots, A_n \in \mathcal{A}$, then $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$ and $\bigcap_{i=1}^{n} A_i \in \mathcal{A}$.

(iii) If $A, B \in \mathcal{A}$, then $A \setminus B \in \mathcal{A}$.

Proof. The proof consists mainly of a direct application of the axioms of a σ-algebra.

(i) It follows from one of De Morgan's laws that $\bigcap_{i=1}^{\infty} A_i = \left(\left(\bigcap_{i=1}^{\infty} A_i\right)^c\right)^c = \left(\bigcup_{i=1}^{\infty} A_i^c\right)^c$. Using closure under complementation and under the formation of countable unions we obtain that $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$.

(ii) We choose the sets $A_{n+1}, A_{n+2}, \ldots$ to be equal to A_n. Then $\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ and $\bigcap_{i=1}^{n} A_i = \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$, which proves statement (ii).

(iii) This follows from $A \setminus B = A \cap B^c$.

□

The next proposition provides a basic result for the construction of σ-algebras.

Proposition 1.1.2. *Let Ω be a non-empty set and let $\big(\mathcal{A}_i\big)_{i\in I}$ be a non-empty collection of σ-algebras on Ω, where I is an arbitrary (finite, countably infinite or even uncountable) index set. Then the intersection of these σ-algebras,*

$$\bigcap_{i\in I}\mathcal{A}_i = \big\{A \subseteq \Omega: \quad A \in \mathcal{A}_i \quad \text{for all } i \in I\big\},$$

is a σ-algebra on Ω.

Proof. Exercise. □

Proposition 1.1.2 implies the following result which will be used several times in what follows.

Corollary 1.1.3. *Let Ω be a non-empty set and let $\mathcal{E}$ be a family of subsets of Ω. Then*

$$\sigma\big(\mathcal{E}\big) := \bigcap_{\mathcal{A}:\ \mathcal{A}\ \sigma-\text{algebra on}\ \Omega,\ \mathcal{E}\subseteq\mathcal{A}} \mathcal{A}$$

is the smallest σ-algebra on Ω that contains $\mathcal{E}$. It is called the **σ-algebra generated by** *$\mathcal{E}$.*

Proof. Let $\mathcal{C}$ be the collection of all σ-algebras on Ω that include $\mathcal{E}$. Then $\mathcal{C}$ is non-empty since it contains the power set 2^Ω of Ω. The intersection of the σ-algebras that belong to $\mathcal{C}$ is, according to Proposition 1.1.2, a σ-algebra and contains $\mathcal{E}$. It is the smallest σ-algebra that contains $\mathcal{E}$ since it is included in each σ-algebra that contains $\mathcal{E}$. □

In the following we consider more closely an important σ-algebra on $\mathbb{R}^d$. An appropriate choice of such a collection of subsets has to fulfill two requirements. On the one hand, it should be rich enough such that it contains virtually all subsets of $\mathbb{R}^d$ which are of interest in analysis and probability theory. On the other hand, it should be small enough such that it still allows the construction of so-called measures with certain properties. We will see in Subsection 1.4 of this course that the power set 2^Ω is too large for this purpose. The "golden standard" is given by the so-called **Borel σ-algebra** on $\mathbb{R}^d$, which is defined as follows.

Definition. The **Borel σ-algebra** on $\mathbb{R}^d$ is the σ-algebra on $\mathbb{R}^d$ generated by the collection of open subsets of $\mathbb{R}^d$. It is denoted by $\mathcal{B}(\mathbb{R}^d)$ or $\mathcal{B}^d$.

The following proposition shows that $\mathcal{B}^d$ is also generated by other collections of subsets of $\mathbb{R}^d$.

Proposition 1.1.4. *Denote by $\mathcal{O}^d$, $\mathcal{C}^d$ and $\mathcal{I}^d$ the respective collections of all open subsets, closed subsets, and half-open rectangles that have the form $(a_1, b_1] \times \cdots \times (a_d, b_d]$ such that $a_i \leq b_i$ for all $i = 1, \ldots, d$. Then*

$$\mathcal{B}^d = \sigma\big(\mathcal{O}^d\big) = \sigma\big(\mathcal{C}^d\big) = \sigma\big(\mathcal{I}^d\big).$$

Before we prove this proposition, we derive some general facts about σ-algebras.

Lemma 1.1.5. *Suppose that Ω is a non-empty set.*

(i) If $\mathcal{E}$ and $\mathcal{F}$ are collections of subsets of Ω such that $\mathcal{E} \subseteq \mathcal{F}$, then $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{F})$.

(ii) If $\mathcal{A}$ is a σ-algebra on Ω, then $\sigma(\mathcal{A}) = \mathcal{A}$.

Proof. (i) According to Corollary 1.1.3, $\sigma(\mathcal{E})$ and $\sigma(\mathcal{F})$ can be represented as the intersection of all σ-algebras that include $\mathcal{E}$ and $\mathcal{F}$, respectively. Since $\mathcal{E} \subseteq \mathcal{F}$ it follows that each σ-algebra containing $\mathcal{F}$ contains $\mathcal{E}$ as well, i.e.

$$\big\{\mathcal{A}\colon\ \mathcal{A} \text{ is a } \sigma\text{−algebra on } \Omega,\ \ \mathcal{E} \subseteq \mathcal{A}\big\} \supseteq \big\{\mathcal{A}\colon\ \mathcal{A} \text{ is a } \sigma\text{−algebra on } \Omega,\ \ \mathcal{F} \subseteq \mathcal{A}\big\}.$$

Hence

$$\sigma\big(\mathcal{E}\big) = \bigcap_{\mathcal{A}\colon\ \mathcal{A}\ \sigma\text{−algebra on } \Omega,\ \mathcal{E}\subseteq\mathcal{A}} \mathcal{A} \subseteq \bigcap_{\mathcal{A}\colon\ \mathcal{A}\ \sigma\text{−algebra on } \Omega,\ \mathcal{F}\subseteq\mathcal{A}} \mathcal{A} = \sigma\big(\mathcal{F}\big).$$

(ii) It is clear that $\mathcal{A} \subseteq \sigma(\mathcal{A})$. On the other hand, since $\mathcal{A}$ itself is a σ-algebra that contains $\mathcal{A}$ we obtain from Corollary 1.1.3 that

$$\sigma\big(\mathcal{A}\big) = \bigcap_{\mathcal{A}'\colon\ \mathcal{A}'\ \sigma\text{−algebra on } \Omega,\ \mathcal{A}\subseteq\mathcal{A}'} \mathcal{A}' \subseteq \mathcal{A},$$

which completes the proof.

□

With the results of Lemma 1.1.5, we are in a position to prove Proposition 1.1.4.

Proof of Proposition 1.1.4. We show that $\sigma\big(\mathcal{O}^d\big) \overset{a)}{\supseteq} \sigma\big(\mathcal{C}^d\big) \overset{b)}{\supseteq} \sigma\big(\mathcal{I}^d\big)$, and then that $\sigma\big(\mathcal{I}^d\big) \overset{c)}{\supseteq} \mathcal{B}^d$; this will establish the proposition.

a) Let $C \in \mathcal{C}^d$ be an arbitrary closed subset of $\mathbb{R}^d$. Then its complement C^c is an open set, and so it is in $\sigma(\mathcal{O}^d) = \mathcal{B}^d$. Since $\mathcal{B}^d$ is closed to complementation we obtain that $C \in \mathcal{B}^d$. Hence, $\mathcal{C}^d \subseteq \mathcal{B}^d$, which implies by Lemma 1.1.5 that $\sigma(\mathcal{C}^d) \subseteq \sigma(\mathcal{B}^d) = \mathcal{B}^d$.

b) Let $(a,b] = (a_1,b_1] \times \cdots \times (a_d,b_d] \in \mathcal{I}^d$ be arbitrary. Then

$$(a,b] \;=\; \bigcup_{n=1}^{\infty} [a_1 + 1/n, b_1] \times \cdots \times [a_d + 1/n, b_d],$$

i.e. $(a,b] \in \sigma(\mathcal{C}^d)$. This implies that $\mathcal{I}^d \subseteq \sigma(\mathcal{C}^d)$ and hence, again by Lemma 1.1.5, $\sigma(\mathcal{I}^d) \subseteq \sigma\big(\sigma(\mathcal{C}^d)\big) = \sigma(\mathcal{C}^d)$.

c) Let $O \in \mathcal{O}^d$ be an arbitrary open subset of $\mathbb{R}^d$. We have, for each element $x = (x_1, \ldots, x_d) \in O$, that

$$\{x\} \;=\; \bigcap_{n=1}^{\infty} (x_1 - 1/n, x_1] \times \cdots \times (x_d - 1/n, x_d] \;\in\; \sigma\big(\mathcal{I}^d\big),$$

which implies that

$$O \;=\; \bigcup_{x\colon\, x \in O} \{x\} \;\subseteq\; \bigcup_{x\colon\, x \in O} \bigcap_{n=1}^{\infty} (x_1 - 1/n, x_1] \times \cdots \times (x_d - 1/n, x_d],$$

which seems to suggest that $O \in \sigma(\mathcal{I}^d)$. However, this conclusion is flawed. If the open set O is non-empty it is uncountably infinite, and the union of the sets $\bigcap_{n=1}^{\infty} (x_1 - 1/n, x_1] \times \cdots \times (x_d - 1/n, x_d]$ is taken over an *uncountable* collection of sets in $\sigma(\mathcal{I}^d)$. Since σ-algebras are only closed under the formation of *countable* unions, we cannot conclude from the above arguments that $O = \bigcup_{x\colon\, x \in O} \bigcap_{n=1}^{\infty} (x_1 - 1/n, x_1] \times \cdots \times (x_d - 1/n, x_d] \;\in\; \sigma(\mathcal{I}^d)$.

To overcome this difficulty, we consider the set of half-open intervals with *rational* coordinates,

$$\mathcal{I}^d_{\mathbb{Q}} \;:=\; \big\{ (r_1, s_1] \times \cdots \times (r_d, s_d] \colon \;\; r_i, s_i \in \mathbb{Q},\; r_i \le s_i \text{ for all } i = 1, \ldots, d \big\}.$$

Since each point $x \in O$ is an inner point of the set O we have that $x \in \bigcup_{I \in \mathcal{I}^d_{\mathbb{Q}}\colon\, I \subseteq O} I$, which implies that

$$O \subseteq \bigcup_{I \in \mathcal{I}^d_{\mathbb{Q}}\colon\, I \subseteq O} I \;\subseteq\; O.$$

Hence, we can represent O as a *finite or countably infinite* collection of subsets from $\mathcal{I}^d$. Therefore, $O \in \sigma(\mathcal{I}^d)$, which implies by Lemma 1.1.5

$$\mathcal{B}^d \;=\; \sigma(\mathcal{O}^d) \;\subseteq\; \sigma\big((\mathcal{I}^d)\big) \;=\; \sigma(\mathcal{I}^d).$$

□

1.2 Measures

Let Ω be a non-empty set that is equipped with a well-structured collection $\mathcal{A}$ of subsets, a σ-algebra. In probability theory or mathematical statistics, Ω usually describes the possible outcomes of a random experiment. In this case, one would perhaps like to assign probabilities to all possible outcomes $\omega \in \Omega$. Often one is also interested in the probability that the outcome of the experiment falls in a certain subset $A \in \mathcal{A}$ of Ω. Such probabilities will be described by a suitable function $P\colon \mathcal{A} \to [0,1]$. (Recall that probabilities are expressed by numbers between zero and one, not in percent.)

Apart from probability theory, in the case where Ω is the d-dimensional Euclidean space denoted by $\mathbb{R}^d$, one wants to specify the volume of a subset of $\mathbb{R}^d$. In dimensions 1, 2, or 3, this coincides with the concepts of length, area, or ordinary volume. Let us consider the case of $d = 2$. For some "nice" subsets of $\mathbb{R}^2$, it is obvious how its area has to be specified. For example, if $A = (a_1, b_1] \times (a_2, b_2]$ ($a_i \leq b_i$, $i = 1, 2$), then the area is clearly equal to $(b_1 - a_1)(b_2 - a_2)$. However, a σ-algebra which contains all half-open subsets of $\mathbb{R}^2$ also contains less nice sets that can hardly be described analytically. Nevertheless, it is possible to specify a function $\lambda^2\colon \mathcal{B}^2 \to [0,\infty]$ that assigns the volume to all Borel sets. In what follows we describe how measures with certain properties can be specified on suitable σ-algebras on arbitrary spaces. We start with a formal definition of the notion of measure.

Definition. Let Ω be a non-empty set and let $\mathcal{A}$ be a σ-algebra on Ω. A set function μ from $\mathcal{A}$ to the extended real number line $\bar{\mathbb{R}} := \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$ is a **measure** on $\mathcal{A}$ if it satisfies these conditions:

(i) $\mu(A) \in [0,\infty]$ for all $A \in \mathcal{A}$,

(ii) $\mu(\emptyset) = 0$,

(iii) if $A_1, A_2, \ldots$ is a sequence of disjoint sets from $\mathcal{A}$, then

$$\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mu(A_i).$$

This property is referred to as σ-**additivity**.

The pair $(\Omega, \mathcal{A})$ is called **measurable space** and the triple $(\Omega, \mathcal{A}, \mu)$ is referred to as **measure space**. If $(\Omega, \mathcal{A}, \mu)$ is a measure space, then one often says that μ is a measure on $(\Omega, \mathcal{A})$, or, if the σ-algebra is clear from the context, a measure on Ω.

The measure μ is **finite** or **infinite** as $\mu(\Omega) < \infty$ or $\mu(\Omega) = \infty$. μ is a **probability measure** if $\mu(\Omega) = 1$.

Since measures assume values in the set $[0,\infty]$ consisting of the ordinary non-negative real numbers and the special value ∞, some conventions involving ∞ are called for.

For $x, y \in [0,\infty]$, $x \leq y$ means that $y = \infty$ or else x and y are finite and $x \leq y$ in the ordinary sense. Similarly, $x < y$ means that $y = \infty$ and x is finite or else x and y are both finite and $x < y$ holds in the usual sense.

For a finite or infinite sequence $x, x_1, x_2, \ldots$ in $[0,\infty]$,

$$x = \sum_i x_i$$

means that either (i) $x = \infty$ and $x_i = \infty$ for some i, or (ii) $x = \infty$ and $x_i < \infty$ for all i and $\sum_i x_i$ is an ordinary divergent series, or (iii) $x < \infty$ and $x_i < \infty$ for all i and $x = \sum_i x_i$ holds in the usual sense. In all cases, the order of summation has no effect on the sum.

For an infinite sequence $x, x_1, x_2, \ldots$ in $[0, \infty]$,

$$x_i \nearrow x$$

means in the first place that $x_i \leq x_{i+1} \leq x$ and in the second place that either (i) $x < \infty$ and there is convergence in the usual sense, or (ii) $x = \infty$ and $x_i = \infty$ for some i, or (iii) $x = \infty$ and the x_i are finite reals converging to infinity in the usual sense.

We turn to some examples of measures.

1. Let Ω be a non-empty set, and let $\mathcal{A}$ be a σ-algebra on Ω. Define a function $\mu\colon \mathcal{A} \to [0, \infty]$ such that $\mu(A) = n$ if A is a finite set with n elements, and $\mu(A) = \infty$ if A is an infinite set. Then μ is called **counting measure** on $(\Omega, \mathcal{A})$.

2. Let Ω be a non-empty set, and let $\mathcal{A}$ be a σ-algebra on Ω. Let x be a member of Ω. Define a function $\delta_x\colon \mathcal{A} \to [0, 1]$ such that $\delta_x(A) = 1$ if $x \in \mathcal{A}$, and $\delta_x(A) = 0$ if $x \notin A$. Then δ_x is a probability measure on $(\Omega, \mathcal{A})$ which is called **Dirac measure** concentrated at x.

A few properties of measures are summarized in the following proposition.

Proposition 1.2.1. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, and let $A_1, A_2, \ldots \in \mathcal{A}$. Then*

(i) *μ is finitely additive: If $A_1, \ldots, A_n$ are pairwise disjoint, then*

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

(ii) *μ is monotone: If $A_1 \subseteq A_2$, then*

$$\mu(A_1) \leq \mu(A_2).$$

(iii) *Subtractivity: If $A_1 \subseteq A_2$ and $\mu(A_1) < \infty$, then*

$$\mu(A_2 \setminus A_1) = \mu(A_2) - \mu(A_1).$$

(iv) *Continuity from below: If $A_1 \subseteq A_2 \subseteq \ldots$, then*

$$\mu(A_n) \nearrow \mu\Big(\bigcup_{i=1}^{\infty} A_i\Big).$$

(v) *Continuity from above: If $A_1 \supseteq A_2 \supseteq \ldots$ and $\mu(A_N) < \infty$ for some N, then*

$$\mu(A_n) \searrow \mu\Big(\bigcap_{i=1}^{\infty} A_i\Big).$$

(vi) *σ-subadditivity:*

$$\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) \le \sum_{i=1}^{\infty} \mu(A_i).$$

(vii) *finite subadditivity:*

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

Remark 1.2.2. *For property (v) to hold, the condition that $\mu(A_N) < \infty$ for some N is indeed necessary. To see this, consider the following counterexample. Let $\Omega = \mathbb{N}$, $\mathcal{A} = 2^{\Omega}$, and let μ be counting measure on $(\Omega, \mathcal{A})$. Consider the sets $A_i := \{i, i+1, \dots\}$. Then $\mu(A_i) = \infty$ holds for all i. On the other hand, we have that $\bigcap_{i=1}^{\infty} A_i = \emptyset$, which implies that $\mu(\bigcap_{i=1}^{\infty} A_i) = 0$.*

Likewise, for property (iii) to hold, $\mu(A_1) < \infty$ is a necessary condition. With the sets A_i defined above and the counting measure μ we have that $\mu(A_1 \setminus A_i) = i - 1$. On the other hand, the right-hand side of the equation in (iii) involves "$\infty - \infty$" which is strictly forbidden in measure theory!

Proof of Proposition 1.2.1.
(i) Choose $A_i = \emptyset$ for all $i > n$. Then we obtain from σ-additivity of μ

$$\mu\Big(\bigcup_{i=1}^{n} A_i\Big) = \mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \underbrace{\mu(A_i)}_{=0 \text{ if } i>n} = \sum_{i=1}^{n} \mu(A_i).$$

(ii), (iii) The sets A_1 and $A_2 \setminus A_1$ are disjoint. Therefore it follows from (i) that

$$\mu(A_1) + \mu(A_2 \setminus A_1) = \mu(A_2),$$

which implies (ii) and (iii).
(iv) In order to use σ-additivity we represent $\bigcup_{i=1}^{\infty} A_i$ as a union of pairwise **disjoint** sets. Let $B_1 := A_1$ and, for $i > 1$, $B_i = A_i \setminus A_{i-1}$. Then $A_i = B_1 \cup \dots \cup B_i$ and $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} \bigcup_{j=1}^{i} B_j = \sum_{j=1}^{\infty} B_j$. This implies

$$\begin{aligned}\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) &= \mu\Big(\bigcup_{i=1}^{\infty} B_i\Big) = \sum_{i=1}^{\infty} \mu(B_i) \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} \mu(B_i) = \lim_{n\to\infty} \mu(A_n).\end{aligned}$$

(v) Let $B_i := A_N \setminus A_i$. Then $B_1 \subseteq B_2 \subseteq \dots$ and we obtain from (iv)

$$\mu\Big(\bigcup_{i=1}^{\infty} B_i\Big) = \lim_{i\to\infty} \mu(B_i). \tag{1.2.1a}$$

We have that

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} (A_N \cap A_i^c) = A_N \cap \Big(\bigcup_{i=1}^{\infty} A_i^c\Big) = A_N \cap \Big(\bigcap_{i=1}^{\infty} A_i\Big)^c = A_N \setminus \Big(\bigcap_{i=1}^{\infty} A_i\Big),$$

which implies

$$\mu\Big(\bigcup_{i=1}^{\infty} B_i\Big) \;=\; \mu(A_N) \;-\; \mu\Big(\bigcap_{i=1}^{\infty} A_i\Big). \tag{1.2.1b}$$

For $i \geq N$ we have that $A_N \supseteq A_i$ and $\mu(A_i) < \infty$, which implies by (iii) $\mu(B_i) = \mu(A_N) - \mu(A_i)$. Therefore,

$$\lim_{i\to\infty} \mu(B_i) \;=\; \mu(A_N) \;-\; \lim_{i\to\infty} \mu(A_i). \tag{1.2.1c}$$

(v) now follows from (1.2.1a) to (1.2.1c).
(vi), (vii) As in the proof of (iv), define $B_1 := A_1$ and, for $i > 1$, $B_i = A_i \setminus A_{i-1}$. Then $A_1 \cup \cdots \cup A_i = B_1 \cup \cdots \cup B_i$ and $\bigcup_{i=1}^{\infty} A_i = \sum_{i=1}^{\infty} B_i$. $B_1, B_2, \ldots$ are pairwise disjoint sets, and we obtain

$$\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) \;=\; \mu\Big(\bigcup_{i=1}^{\infty} B_i\Big) \;=\; \sum_{i=1}^{\infty} \underbrace{\mu(B_i)}_{\leq \mu(A_i)} \;\leq\; \sum_{i=1}^{\infty} \mu(A_i)$$

and, analogously,

$$\mu\Big(\bigcup_{i=1}^{n} A_i\Big) \;=\; \mu\Big(\bigcup_{i=1}^{n} B_i\Big) \;=\; \sum_{i=1}^{n} \underbrace{\mu(B_i)}_{\leq \mu(A_i)} \;\leq\; \sum_{i=1}^{n} \mu(A_i).$$

□

In the following we show how measures on general measurable spaces $(\Omega, \mathcal{A})$ can be defined. As long as the space Ω is finite or countably infinite, the specification of a measure is easy and does not require any deep results. This is corroborated by the following proposition.

Proposition 1.2.3. *Let Ω by a non-empty, finite or countably infinite set, and let $p\colon \Omega \to [0, \infty]$ be an arbitrary function. Then there exists a unique measure μ on $(\Omega, 2^{\Omega})$ such that*

$$\mu(\{\omega\}) \;=\; p(\omega) \qquad \forall \omega \in \Omega.$$

Proof. If there exists such a measure μ at all, then if follows from the axiom of σ-additivity that the measure of an arbitrary subset A of Ω is given by

$$\mu(A) \;=\; \sum_{\omega\in A} p(\{\omega\}). \tag{1.2.2}$$

(The order of summation is irrelevant since all summands are nonnegative. This is actually the only possible assignment of a measure to A since the sum is taken over at most countably many terms.) It is now easy to see that the function $\mu\colon 2^{\Omega} \to [0, \infty]$ satisfies all axioms of a measure: μ is obviously a non-negative function and $\mu(\emptyset) = 0$. If $A_1, A_2, \ldots$ are disjoint subsets of Ω, then

$$\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) \;=\; \sum_{\omega\in\bigcup_{i=1}^{\infty} A_i} p(\{\omega\}) \;=\; \sum_{i=1}^{\infty}\sum_{\omega\in A_i} p(\{\omega\}) \;=\; \sum_{i=1}^{\infty} \mu(A_i).$$

(Note that the seconds equality holds since all summands are nonnegative.) □

1.3 A general approach to specify measures on uncountable spaces

We have seen in the introductory part of this course (Lemma 0.0.1 on page 4) that the specification of a measure that fulfills a “wish list” of properties is not always a simple task. As a second example, which will also guide us through the present subsection, we consider the so-called **Lebesgue measure** which is named after the French mathematician Henri Léon Lebesgue. On $\mathbb{R}^d$, equipped with the σ-algebra $\mathcal{B}^d$ of Borel sets, Lebesgue measure λ^d is that measure which assigns to a Borel set A its d-dimensional volume. For $d = 1$, 2, or 3, it coincides with the standard measures of length, area, or ordinary volume.

In the following we approach cautiously to a definition of Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}^d)$. We begin with its specification on sets with a simple structure. Since this specification is done on a collection of sets which not a σ-algebra and is much smaller than $\mathcal{B}^d$ we denote this set function by λ_0^d. Let

$$\mathcal{I}^d \;=\; \big\{(a,b] = (a_1,b_1] \times \cdots \times (a_d,b_d]\colon\; -\infty < a_i \le b_i < \infty\big\}$$

be the collection of half-open rectangles in $\mathbb{R}^d$. Since Lebesgue measure should assign to all suitable subsets of $\mathbb{R}^d$ a number that corresponds to the usual notion of a volume we define

$$\lambda_0^d\big((a,b]\big) \;=\; \prod_{i=1}^{d}(b_i \,-\, a_i) \qquad \forall (a,b] \in \mathcal{I}^d. \tag{1.3.1a}$$

We can even go one step further with our definition. If $A_1, \ldots, A_k$ are pairwise disjoint sets from $\mathcal{I}^d$ then the only possible extension of this definition which does not contradict the property of finite additivity is given by

$$\lambda_0^d\big(A_1 \cup \cdots \cup A_k\big) \;=\; \sum_{i=1}^{k} \lambda_0^d\big(A_i\big). \tag{1.3.1b}$$

This extension is in line with our intention to define a measure which describes the volume of subsets of $\mathbb{R}^d$. There is a question here because $A_1 \cup \cdots \cup A_k$ will have other representations as a finite union of disjoint rectangles. Suppose that $A_1, \ldots, A_k$ and $B_1, \ldots, B_l$ are both collections of pairwise disjoint half-open rectangles, and that $A_1 \cup \cdots \cup A_k = B_1 \cup \cdots \cup B_l$. Then $C_{i,j} := A_i \cap B_j$, $(i = 1, \ldots, k;\; j = 1, \ldots, l)$ is also a collection of disjoint sets from $\mathcal{I}^d$. Since $A_i = C_{i,1} \cup \cdots \cup C_{i,l}$ $(i = 1, \ldots, k)$ and $B_j = C_{1,j} \cup \cdots \cup C_{k,j}$ $(j = 1, \ldots, l)$ it follows from (1.3.1b) that

$$\begin{aligned}
\lambda_0^d\big(A_1 \cup \cdots \cup A_k\big) \;&=\; \sum_{i=1}^{k} \lambda_0^d\big(A_i\big) \;=\; \sum_{i=1}^{k}\sum_{j=1}^{l} \lambda_0^d\big(C_{i,j}\big) \\
&=\; \sum_{j=1}^{l}\sum_{i=1}^{k} \lambda_0^d\big(C_{i,j}\big) \;=\; \sum_{j=1}^{l} \lambda_0^d\big(B_j\big) \;=\; \lambda_0^d\big(B_1 \cup \cdots \cup B_l\big),
\end{aligned} \tag{1.3.2}$$

i.e. the extension (1.3.1b) is indeed consistent in the sense that it does not depend on the particular choice of a representation of a set. With (1.3.1a) and (1.3.1b) we have the desired specification of Lebesgue measure on the set

$$\mathcal{B}_0^d \;:=\; \Big\{\bigcup_{i=1}^{k} A_i\colon\;\; A_1, \ldots, A_k \in \mathcal{I}^d \text{ disjoint}, k \in \mathbb{N}\Big\}.$$

$\mathcal{B}_0^d$ is actually a well-structured collection of subsets of $\mathbb{R}^d$. We give this structure the following definition.

Definition. Let Ω be a non-empty set. A collection $\mathcal{R}$ of subsets of Ω is called a **ring** (ring of sets) if

(i) $\emptyset \in \mathcal{R}$,

(ii) if $A, B \in \mathcal{R}$, then $A \setminus B \in \mathcal{R}$,

(iii) if $A, B \in \mathcal{R}$, then $A \cup B \in \mathcal{R}$.

If additionally $\Omega \in \mathcal{R}$, then $\mathcal{R}$ is also an **algebra** on Ω.

Remark 1.3.1. *If $\mathcal{R}$ is a ring on a non-empty set Ω, and if $A, B \in \mathcal{R}$, then it follows from*

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

that $A \cap B \in \mathcal{R}$, i.e. $\mathcal{R}$ is closed under finite intersections.

It can be easily seen that the collection of sets $\mathcal{B}_0^d$ which consists of all finite unions of half-open intervals is a ring on $\mathbb{R}^d$.

Lemma 1.3.2. *$\mathcal{B}_0^d$ is the smallest ring on Ω which contains $\mathcal{I}^d$.*

Proof. We check that $\mathcal{B}_0^d$ satisfies the axioms of a ring.

(i) $\emptyset \in \mathcal{B}_0^d$ is obvious since, e.g. $\emptyset = (0,0] \times \cdots \times (0,0]$.

(ii) We have to show that, for arbitrary $A, B \in \mathcal{B}_0^d$, $A \setminus B$ can represented as a finite union of disjoint rectangles in $\mathbb{R}^d$. For clarity, the argument is broken into several steps.
a) Let $A = I_1 \times \cdots \times I_d$ and $B = J_1 \times \cdots \times J_d$ be rectangles such that $B \subseteq A$. If $B \neq \emptyset$, then $J_i \subseteq I_i$ and $I_i \setminus J_i$ is a union $J_i' \cup J_i''$ of disjoint intervals (possibly empty). Consider the 3^d disjoint rectangles $U_1 \times \cdots \times U_d$, where for each i U_i is J_i or J_i' or J_i''. One of these rectangles $(J_1 \times \cdots \times J_d)$ is B itself, and $A \setminus B$ is the union of the others. Hence, $A \setminus B \in \mathcal{B}_0^d$.
b) Suppose now that A and B are arbitrary rectangles. Then $A \cap B$ is also a rectangle and it follows from *a)* that $A \setminus B = A \setminus (A \cap B) \in \mathcal{B}_0^d$.
c) Let now $A \in \mathcal{B}_0^d$ and $B \in \mathcal{I}^d$ be arbitrary. Then there exist disjoint rectangles $I_1, \ldots, I_k$ such that $A = I_1 \cup \cdots \cup I_k$. It holds that $A \setminus B = (I_1 \setminus B) \cup \cdots \cup (I_k \setminus B)$, where $I_1 \setminus B, \ldots, I_k \setminus B$ are disjoint sets that are, by *b)*, members of $\mathcal{B}_0^d$. Therefore, $A \setminus B \in \mathcal{B}_0^d$.
d) Now we assume that $A \in \mathcal{B}_0^d$ and $B \in \mathcal{B}_0^d$. Then there exist disjoint rectangles $J_1, \ldots, J_l$ such that $B = J_1 \cup \cdots \cup J_l$. Since $A \setminus B = \Big(\big((A \setminus J_1) \setminus J_2\big) \setminus \ldots \setminus J_{l-1}\Big) \setminus J_l$ we conclude from *c)* that $A \setminus B \in \mathcal{B}_0^d$.

(iii) Let $A, B \in \mathcal{B}_0^d$ be arbitrary. Then $A \cup B = (A \cap B) \cup (A \setminus B) \cup (B \setminus A)$ are disjoint sets in $\mathcal{B}_0^d$. This implies that $A \cup B \in \mathcal{B}_0^d$.

$\square$

In the following we show that the set function $\lambda_0^d\colon \mathcal{B}_0^d \to [0,\infty)$ also satisfies the axioms of a measure. The next definition introduces corresponding notions in a general framework.

Definition. Suppose that $\mathcal{R}$ is a ring on a non-empty set Ω, and that $\mu\colon \mathcal{R} \to [0,\infty]$.

(i) μ is called a **content** if

a) $\mu(\emptyset) = 0$,

b) for disjoint sets $A_1, \dots, A_n \in \mathcal{R}$,

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

(ii) μ is called a **pre-measure** if

a) $\mu(\emptyset) = 0$,

b) for disjoint sets $A_1, A_2, \dots \in \mathcal{R}$ such that $\bigcup_{i=1}^\infty A_i \in \mathcal{R}$,

$$\mu\Big(\bigcup_{i=1}^\infty A_i\Big) = \sum_{i=1}^\infty \mu(A_i).$$

(iii) μ is called **σ-finite** if the exists sets $A_1, A_2, \dots \in \mathcal{R}$ such that $\mu(A_i) < \infty$ for all i and $\bigcup_{i=1}^\infty A_i = \Omega$.

Before we proceed we derive a simple calculation rule which is named after the English mathematician James Joseph Sylvester and the French mathematician Henri Poincaré. This rule is often stated for probability measures on σ-algebras, however, the simpler structures of a ring and a content are sufficient for this.

Proposition 1.3.3. (Inclusion-exclusion principle, formula of Poincaré-Sylvester) *Let μ be a content on a ring $\mathcal{R}$ on Ω. If $A_1, \dots, A_n \in \mathcal{R}$ are such that $\mu(A_i) < \infty$ for $i = 1, \dots, n$, then*

$$\mu\Big(\bigcup_{i=1}^n A_i\Big) = \sum_{k=1}^n (-1)^{k+1} \sum_{(i_1,\dots,i_k)\colon\ 1 \le i_1 < \dots < i_k \le n} \mu\big(A_{i_1} \cap \dots \cap A_{i_k}\big).$$

Proof. Recall that, according to Remark 1.3.1, a ring is intersection-stable, i.e. finite intersections of the sets $A_1, \dots, A_n$ are members of $\mathcal{R}$.

We prove the statement by induction. Let $n = 2$. Since $A_1 \cap A_2$, $A_1 \setminus (A_1 \cap A_2)$, and $A_2 \setminus (A_1 \cap A_2)$ are disjoint sets with union equal to $A_1 \cup A_2$ we obtain

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

Suppose now that the formula holds true for $n = 2$ and $n = k$. Then

$$\begin{aligned}
&\mu\big(A_1 \cup \cdots \cup A_{k+1}\big) \\
&\quad= \mu\big(A_1 \cup \cdots \cup A_k\big) + \mu\big(A_{k+1}\big) - \mu\big((A_1 \cup \cdots \cup A_k) \cap A_{k+1}\big) \\
&\quad= \mu\big(A_1 \cup \cdots \cup A_k\big) + \mu\big(A_{k+1}\big) - \mu\big((A_1 \cap A_{k+1}) \cup \cdots \cup (A_k \cap A_{k+1})\big) \\
&\quad= \sum_{j=1}^{k} (-1)^{j+1} \sum_{(i_1,\ldots,i_j):\ 1\le i_1<\cdots<i_j\le n} P\Big(A_{i_1} \cap \cdots \cap A_{i_j}\Big) \\
&\qquad + (-1)^{1+1}\mu\big(A_{k+1}\big) \\
&\qquad - \sum_{j=1}^{k} (-1)^{j+1} \sum_{(i_1,\ldots,i_j):\ 1\le i_1<\cdots<i_j\le n} P\Big(A_{i_1} \cap \cdots \cap A_{i_j} \cap A_{k+1}\Big).
\end{aligned}$$

Hence, the formula is also true for $n = k + 1$ and the proof is complete. □

The significance of the above concepts will become clear when we extend the set function λ_0^d on $\mathcal{B}_0^d$ to Lebesgue measure λ^d that has to be defined on the collection $\mathcal{B}^d$ of all Borel sets in $\mathbb{R}^d$. Next we show that λ_0^d is actually a σ-additive set function, and hence a pre-measure on $\mathcal{B}_0^d$.

Lemma 1.3.4. *The function $\lambda_0^d\colon \mathcal{B}_0^d \to [0,\infty)$ defined by (1.3.1a) and (1.3.1b) is the unique pre-measure on $\mathcal{B}_0^d$ such that*

$$\lambda_0^d\big((a,b]\big) = \prod_{i=1}^{d} (b_i - a_i) \qquad \textit{for all } a = (a_1,\ldots,a_d),\ b = (b_1,\ldots,b_d),\ a_i \le b_i\ \forall i = 1,\ldots,d.$$

Proof. We have already seen that λ_0^d is the unique content on $\mathcal{B}_0^d$ that satisfies (1.3.1a). It remains to show that λ_0^d is σ-additive on $\mathcal{B}_0^d$.

Let $A_1, A_2, \ldots$ be disjoint sets that belong to $\mathcal{B}_0^d$ and suppose that $A := \bigcup_{i=1}^\infty A_i \in \mathcal{B}_0^d$. We have to show that

$$\lambda_0^d\big(A\big) = \sum_{i=1}^{\infty} \lambda_0^d\big(A_i\big).$$

Since $A_1 \cup \cdots \cup A_n \subseteq A$ it follows from finite additivity of λ_0^d that $\sum_{i=1}^n \lambda_0^d\big(A_i\big) = \lambda_0^d\big(A_1 \cup \cdots \cup A_n\big) \le \lambda_0^d(A)$, which implies that

$$\sum_{i=1}^{\infty} \lambda_0^d\big(A_i\big) = \lim_{n\to\infty} \sum_{i=1}^{n} \lambda_0^d\big(A_i\big) \le \lambda_0^d\big(A\big).$$

It remains to show that

$$\lim_{n\to\infty} \sum_{i=1}^{n} \lambda_0^d\big(A_i\big) \ge \lambda_0^d\big(A\big). \tag{1.3.3}$$

Suppose that the opposite is true, i.e. there exists some $\epsilon > 0$ such that

$$\sum_{i=1}^{n} \lambda_0^d\big(A_i\big) = \lambda_0^d\big(A_1 \cup \cdots \cup A_n\big) \le \lambda_0^d\big(A\big) - \epsilon \quad \forall n \in \mathbb{N}.$$

For $n \in \mathbb{N}$, define $B_n = A \setminus \big(A_1 \cup \cdots \cup A_n\big)$. Since $\bigcup_{i=1}^{\infty} A_i = A$ we obtain that

$$\bigcap_{i=1}^{\infty} B_i \;=\; \emptyset. \tag{1.3.4}$$

$(B_n)_{n\in\mathbb{N}}$ is a non-increasing sequence of sets that belong to $\mathcal{B}_0^d$ and it holds that

$$\lambda_0^d\big(B_n\big) \;=\; \lambda_0^d\big(A\big) \;-\; \lambda_0^d\big(A_1 \cup \cdots \cup A_n\big) \;\geq\; \epsilon \quad \forall n \in \mathbb{N}.$$

For each $n \in \mathbb{N}$, we can choose a set $C_n \in \mathcal{B}_0^d$ such that

$$C_n \;\subseteq\; \overline{C_n} \;\subseteq\; B_n$$

and

$$\lambda_0^d\big(B_n \setminus C_n\big) \;\leq\; \epsilon\, 2^{-n}.$$

Let $D_n := C_1 \cap \cdots \cap C_n$. Then

$$\lambda_0^d\big(B_n \setminus D_n\big) \;\leq\; \sum_{i=1}^{n} \lambda_0^d\big(B_n \setminus C_i\big) \;\leq\; \sum_{i=1}^{n} \lambda_0^d\big(B_i \setminus C_i\big) \;\leq\; \epsilon\big(2^{-1} + \cdots + 2^{-n}\big),$$

which implies that

$$\lambda_0^d\big(D_n\big) \;=\; \lambda_0^d\big(B_n\big) \;-\; \lambda_0^d\big(B_n \setminus D_n\big) \;\geq\; \epsilon\big[1 \;-\; \big(2^{-1} + \cdots + 2^{-n}\big)\big] \;>\; 0.$$

Therefore, $(\overline{D_n})_{n\in\mathbb{N}}$ is a **non-increasing** sequence of **non-empty compact** subsets of $\mathbb{R}^d$. This, however, implies that

$$\bigcap_{n=1}^{\infty} \overline{D_n} \neq \emptyset, \tag{1.3.5}$$

which contradicts (1.3.4). Hence, (1.3.3) follows, which completes the proof.

The proof of (1.3.5) follows from a standard argument for compact sets. Suppose that the opposite is true, i.e. $\bigcap_{i=1}^{\infty} \overline{D_n} = \emptyset$. Then $\overline{D_1} \cap \Big(\bigcap_{n=2}^{\infty} \overline{D_n}\Big) = \emptyset$, which implies that $\overline{D_1} \subseteq \Big(\bigcap_{n=2}^{\infty} \overline{D_n}\Big)^c = \bigcup_{n=2}^{\infty} \overline{D_n}^{\,c}$. This means that the compact set $\overline{D_1}$ is covered by the collection of open sets $\overline{D_2}^{\,c}, \overline{D_3}^{\,c}, \ldots$. By the Heine-Borel theorem we can choose a finite subcover, i.e. there exists some $N \geq 2$ such that $\overline{D_1} \subseteq \bigcup_{n=2}^{N} \overline{D_n}^{\,c}$. Therefore, $\overline{D_N} = \overline{D_1} \cap \cdots \cap \overline{D_N} = \emptyset$, which is a contradiction to the fact that the sets $\overline{D_n}$ are all non-empty. Hence, (1.3.5) is proven. □

Let us summarize what we have achieved so far. We intend to construct a measure on $(\mathbb{R}^d, \mathcal{B}^d)$ which assigns to all subsets that belong to $\mathcal{B}^d$ a number which is consistent with the usual notion of volume in $\mathbb{R}^d$. For sets with a simple structure, the specification was undisputable. In particular, for a rectangle $(a, b] = (a_1, b_1] \times \cdots \times (a_d, b_d]$, the only reasonable choice of its measure is given as $\lambda_0^d\big((a, b]\big) = \prod_{i=1}^{d}(b_i - a_i)$, and for a union of rectangles the required additivity provides a specification of its measure. In fact, Lemma 1.3.2 shows that the corresponding set function $\lambda_0^d\colon \mathcal{B}_0^d \to [0, \infty)$ is a pre-measure on the ring generated by the half-open rectangles in $\mathbb{R}^d$. The system of sets $\mathcal{B}_0^d$ is well-structured but it is clearly not a σ-algebra on $\mathbb{R}^d$. Actually, the σ-algebra $\mathcal{B}^d$ which is generated by the half-open rectangles is much richer than $\mathcal{B}_0^d$ and there are sets in $\mathcal{B}^d$ which cannot be neatly described. In view of this, a direct assignment of a measure to such sets seems to be out of reach. In the following we develop a standard technique for specifying a measure in such a case. It will turn out that an explicit specification of Lebesgue measure for sets not belonging to $\mathcal{B}_0^d$ is not necessary. In fact, it will be shown that there exists a **unique extension** of the pre-measure λ_0^d defined on the ring $\mathcal{B}_0^d$ to a measure λ^d on the σ-algebra $\mathcal{B}^d$. The idea of this extension can be sketched as follows: For an **arbitrary subset** Q of $\mathbb{R}^d$ we define

$$\lambda^*\big(Q\big) \;=\; \inf\Big\{\sum_{i=1}^{\infty}\lambda_0^d(A_i):\quad Q\subseteq\bigcup_{i=1}^{\infty}A_i,\;\; A_1,A_2,\ldots\in\mathcal{B}_0^d\Big\}.$$

It turns out that $\lambda^*\big(A\big)=\lambda_0^d\big(A\big)$ holds for all $A\in\mathcal{B}_0^d$, i.e., λ^* is actually an extension but not a redefinition of λ_0^d. Furthermore, λ^* satisfies the axioms of a measure on a collection of sets that includes $\mathcal{B}^d$. Hence, the restriction of λ^* to this σ-algebra is a measure on $(\mathbb{R}^d,\mathcal{B}^d)$. And finally, this extension turns out to be unique. Therefore, Lebesgue measure we are striving for will be completely specified.

In what follows we develop this method for constructing measures in a general framework. We begin with the following definition.

Definition. Let Ω be a non-empty set, and let 2^Ω be the collection of all subsets of Ω, the so-called power set. A set function $\mu^*\colon\, 2^\Omega\to[0,\infty]$ is an **outer measure** on $(\Omega,2^\Omega)$ if

(i) $\mu^*\big(\emptyset\big) \;=\; 0$,

(ii) if $A\subseteq B\subseteq\Omega$, then $\mu^*\big(A\big)\le\mu^*\big(B\big)$, and

(iii) if $A_1,A_2,\ldots$ are subsets of Ω, then $\mu^*\big(\bigcup_{i=1}^\infty A_i\big)\;\le\;\sum_{i=1}^\infty\mu^*\big(A_i\big)$.

Example
Let Ω be a non-empty set, and let $\eta(A)=0$ if $A=\emptyset$ and $\eta(A)=1$ if $A\neq\emptyset$. Then η is an outer measure on 2^Ω. If Ω contains at least 2 elements, then η is not a measure on 2^Ω.

Lemma 1.3.5. *Let μ be a pre-measure on a ring $\mathcal{R}$ on a non-empty set Ω, and let, for $Q\subseteq\Omega$*

$$\mu^*\big(Q\big) \;=\; \begin{cases} \inf\big\{\sum_{i=1}^\infty\mu(A_i):\quad Q\subseteq\bigcup_{i=1}^\infty A_i,\; A_1,A_2,\ldots\in\mathcal{R}\big\}, & \textit{if } \big\{\ldots\big\} \textit{ is non-empty,}\\ \infty & \textit{if } \big\{\ldots\big\} \textit{ is empty.}\end{cases}$$

Then

(i) $\mu^*(Q) \;=\; \mu(Q)$ *for all* $Q\in\mathcal{R}$,

(ii) μ^* *is an outer measure on* 2^Ω.

Proof. (i) Let Q be an arbitrary set that belongs to $\mathcal{R}$. With $A_1 := Q$ and $A_i = \emptyset$ for all $i \geq 2$ we have $Q \subseteq \bigcup_{i=1}^{\infty} A_i$ and

$$\mu(Q) \;=\; \sum_{i=1}^{\infty} \mu(A_i) \;\geq\; \mu^*(Q).$$

On the other hand, it follows from σ-subadditivity of the pre-measure μ that, for arbitrary sets $A_1, A_2, \ldots \in \mathcal{R}$ such that $Q \subseteq \bigcup_{i=1}^{\infty} A_i$, $\mu(Q) \leq \sum_{i=1}^{\infty} \mu(A_i)$, which implies

$$\mu(Q) \;\leq\; \mu^*(Q).$$

Hence, $\mu^*(Q) = \mu(Q)$ holds true for all $Q \in \mathcal{R}$.

(ii) We verify that μ^* satisfies the axioms of an outer measure.

a) It follows from (i) that $\mu^*(\emptyset) = \mu(\emptyset) = 0$.

b) If $Q_1 \subseteq Q_2 \subseteq \Omega$, then each cover $A_1, A_2, \ldots \in \mathcal{R}$ of Q_2 covers Q_1, too. Hence,

$$\begin{aligned} \mu^*\big(Q_1\big) \;&=\; \inf\Big\{\sum_{i=1}^{\infty} \mu(A_i):\; A_1, A_2, \ldots \in \mathcal{R},\; Q_1 \subseteq \bigcup_{i=1}^{\infty} A_i\Big\} \\ &\leq\; \inf\Big\{\sum_{i=1}^{\infty} \mu(A_i):\; A_1, A_2, \ldots \in \mathcal{R},\; Q_2 \subseteq \bigcup_{i=1}^{\infty} A_i\Big\} \;=\; \mu^*\big(Q_2\big). \end{aligned}$$

c) Let $Q_1, Q_2, \ldots$ be arbitrary subsets of Ω.

*c.*1) If $\mu^*(Q_n) = \infty$ for some $n \in \mathbb{N}$, then

$$\mu^*\Big(\bigcup_{n=1}^{\infty} Q_n\Big) \;\leq\; \sum_{n=1}^{\infty} \mu^*\big(Q_n\big)$$

is obviously fulfilled.

*c.*2) Otherwise, if $\mu^*(Q_n) < \infty$ for all $n \in \mathbb{N}$, we obtain the σ-subadditivity as follows. For arbitrary $\epsilon > 0$, and for all $n \in \mathbb{N}$, there exist sets $A_{n,1}, A_{n,2}, \ldots \in \mathcal{R}$ such that $Q_n \subseteq \bigcup_{i=1}^{\infty} A_{n,i}$ and $\sum_{i=1}^{\infty} \mu(A_{n,i}) \leq \mu^*(Q_n) + 2^{-n}\epsilon$. Now we have

$$\bigcup_{n=1}^{\infty} Q_n \;\subseteq\; \bigcup_{n=1}^{\infty}\bigcup_{i=1}^{\infty} A_{n,i}$$

and

$$\begin{aligned} \mu^*\Big(\bigcup_{n=1}^{\infty} Q_n\Big) \;&\leq\; \sum_{n=1}^{\infty}\sum_{i=1}^{\infty} \mu\big(A_{n,i}\big) \\ &\leq\; \sum_{n=1}^{\infty} \Big(\mu^*\big(Q_n\big) \;+\; \epsilon 2^{-n}\Big) \\ &\leq\; \sum_{n=1}^{\infty} \mu^*\big(Q_n\big) \quad+\quad \epsilon. \end{aligned}$$

This, however, implies that

$$\mu^*\Big(\bigcup_{n=1}^{\infty} Q_n\Big) \;\leq\; \sum_{n=1}^{\infty} \mu^*\big(Q_n\big).$$

Hence, μ^* satisfies all axioms of an outer measure on 2^{Ω}.

□

The definition of the outer measure μ^* given in Lemma 1.3.5 provides an extension of the pre-measure μ. However, it is not guaranteed that μ^* satisfies all axioms of a measure on 2^Ω. We will see later that this is indeed not the case in general. The next theorem shows that we can find a suitable subset $\mathcal{M}_{\mu^*}$ of the power set 2^Ω such that the property of σ-additivity is satisfied on this set. Furthermore, this subset $\mathcal{M}_{\mu^*}$ will be large enough for usual purposes, it contains in particular the σ-algebra $\sigma(\mathcal{R})$ that is generated by the ring $\mathcal{R}$. In the special case where we start with the pre-measure λ_0^d on $\mathcal{B}_0^d$, the corresponding collection $\mathcal{M}_{\lambda^*}$ of sets contains $\sigma(\mathcal{B}_0^d) = \mathcal{B}^d$ which is just the σ-algebra on which an extension of λ_0^d should be defined.

Before we turn to the statement and proof of the announced theorem, we try to provide some intuition about a suitable subset $\mathcal{M}_{\mu^*}$ of 2^Ω. On this collection of sets, the outer measure μ^* must satisfy all axioms of a measure, i.e. in particular the properties of σ-additivity and finite additivity have to hold true. Consider finite additivity. Let $A_1, \dots, A_n$ be disjoint subsets of Ω. We are looking for a suitable condition on these sets which ensures that

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

Suppose that, for any reason, $\mu^*\big(\bigcup_{i=1}^{n-1} A_i\big) = \sum_{i=1}^{n-1} \mu(A_i)$ holds true. Then

$$\begin{aligned}\sum_{i=1}^n \mu^*(A_i) \;&=\; \mu^*\Big(\bigcup_{i=1}^{n-1} A_i\Big) \,+\, \mu^*(A_n) \\ &=\; \mu^*\Big(\big(\bigcup_{i=1}^{n-1} A_i\big) \cap A_n^c\Big) \,+\, \mu^*\Big(\big(\bigcup_{i=1}^{n-1} A_i\big) \cap A_n\Big).\end{aligned}$$

If A_n is such that the right-hand side of this equation is equal to $\mu^*\big(\bigcup_{i=1}^n A_i\big)$, then we obtain the desired equality. This motivates the definition of the collection $\mathcal{A}^*$ of sets in the following theorem.

Theorem 1.3.6. (Carathéodory's extension theorem)
Suppose that $\mathcal{R}$ is a ring on a non-empty set Ω, and that μ is a pre-measure on $\mathcal{R}$. Let $\mu^*\colon\, 2^\Omega \to [0,\infty]$ be the corresponding outer measure, i.e.

$$\mu^*(Q) \;=\; \begin{cases} \inf\big\{\sum_{i=1}^\infty \mu(A_i)\colon \quad Q \subseteq \bigcup_{i=1}^\infty A_i,\; A_1, A_2, \ldots \in \mathcal{R}\big\} & \text{if } \{\ldots\} \text{ is non-empty,} \\ \infty & \text{if } \{\ldots\} \text{ is empty.}\end{cases}$$

Let

$$\mathcal{M}_{\mu^*} \;:=\; \big\{A \in 2^\Omega\colon\;\; \mu^*(Q) \;=\; \mu^*(Q \cap A) \,+\, \mu^*(Q \cap A^c) \quad \forall Q \in 2^\Omega\big\}.$$

Then

(i) $\mathcal{M}_{\mu^*}$ *is a σ-algebra on Ω and $\mu^*\big|_{\mathcal{M}_{\mu^*}}$ is a measure on $(\Omega, \mathcal{M}_{\mu^*})$.*
(A set A that belongs to $\mathcal{M}_{\mu^*}$ is called ***μ^*-measurable****.)*

(ii) $\sigma(\mathcal{R}) \subseteq \mathcal{M}_{\mu^*}$*, i.e. $\mu^*\big|_{\sigma(\mathcal{R})}$ is a measure on $(\Omega, \sigma(\mathcal{R}))$ and $\mu^*(A) \;=\; \mu(A)\;\; \forall A \in \mathcal{R}$.*

This theorem, named after the Greek mathematician Constantin Carathéodory, is one of the main tools for the construction of measures. It will be complemented by the uniqueness theorem (Theorem 1.3.8) which will be stated and proved below. The latter theorem will ensure that a σ-finite measure on a σ-algebra $\mathcal{A}$ is completely specified by its values on an intersection-stable collection of sets $\mathcal{E}$ which generates $\mathcal{A}$, if $\mathcal{E}$ contains sets $E_1, E_2, \ldots$ such that $\bigcup_{n=1}^{\infty} E_n = \Omega$.

Proof of Theorem 1.3.6.

(i) *a*) First we prove that $\mathcal{M}_{\mu^*}$ is a σ-algebra on Ω.

a.1) Since $\mu^*\big(\underbrace{Q\cap\Omega}_{=Q}\big)+\mu^*\big(\underbrace{Q\cap\Omega^c}_{=\emptyset}\big) = \mu^*\big(Q\big)$ for all $Q\subseteq\Omega$ we have that $\Omega\in\mathcal{M}_{\mu^*}$.

a.2) Suppose that $A\in\mathcal{M}_{\mu^*}$. Then, for arbitrary $Q\subseteq\Omega$,

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

which implies that $A^c\in\mathcal{M}_{\mu^*}$.

a.3) We show that $\mathcal{M}_{\mu^*}$ is closed under the formation of **finite** unions. Let $A, B\in\mathcal{M}_{\mu^*}$. Then, for arbitrary $Q\subseteq\Omega$,

$$\begin{aligned}\mu^*\big(Q\big) &= \mu^*\big(Q\cap A\big) + \mu^*\big(Q\cap A^c\big)\\ &= \mu^*\big(Q\cap A\cap B\big) + \mu^*\big(Q\cap A\cap B^c\big)\\ &\quad + \mu^*\big(Q\cap A^c\cap B\big) + \mu^*\big(Q\cap\underbrace{A^c\cap B^c}_{=(A\cup B)^c}\big).\end{aligned}$$

Moreover,

$$\begin{aligned}\mu^*\big(Q\cap(A\cup B)\big) &= \mu^*\big(\underbrace{Q\cap(A\cup B)\cap A}_{=Q\cap A}\big) + \mu^*\big(\underbrace{Q\cap(A\cup B)\cap A^c}_{=Q\cap A^c\cap B}\big)\\ &= \mu^*\big(Q\cap A\cap B\big) + \mu^*\big(Q\cap A\cap B^c\big) + \mu^*\big(Q\cap A^c\cap B\big),\end{aligned}$$

which implies that

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

It follows from *a*.1), *a*.2), and *a*.3) that $\mathcal{M}_{\mu^*}$ is an **algebra** on Ω.

a.4) Now we show that $\mathcal{M}_{\mu^*}$ is closed under the formation of countable unions of **disjoint** sets. Suppose that $A_1, A_2, \ldots\in\mathcal{M}_{\mu^*}$ are disjoint, and that $Q\subseteq\Omega$. Since μ^* is an outer measure we have

$$\mu^*\big(Q\big) \le \mu^*\Big(Q\cap\big(\bigcup_{i=1}^{\infty}A_i\big)\Big) + \mu^*\Big(Q\cap\big(\bigcup_{i=1}^{\infty}A_i\big)^c\Big).$$

Hence it remains to show the reverse inequality,

$$\mu^*\big(Q\big) \ge \mu^*\Big(Q\cap\big(\bigcup_{i=1}^{\infty}A_i\big)\Big) + \mu^*\Big(Q\cap\big(\bigcup_{i=1}^{\infty}A_i\big)^c\Big).$$

We obtain from a.3), for arbitrary $n \in \mathbb{N}$,

$$\begin{aligned} \mu^*(Q) &= \mu^*\Big(Q \cap (\bigcup_{i=1}^n A_i)\Big) + \mu^*\Big(Q \cap (\bigcup_{i=1}^n A_i)^c\Big) \\ &\geq \mu^*\Big(Q \cap (\bigcup_{i=1}^n A_i)\Big) + \mu^*\Big(Q \cap (\bigcup_{i=1}^\infty A_i)^c\Big) \\ &= \sum_{i=1}^n \mu^*(Q \cap A_i) + \mu^*\Big(Q \cap (\bigcup_{i=1}^\infty A_i)^c\Big) \end{aligned} \tag{1.3.6}$$

In fact, the latter equality follows from

$$\begin{aligned} \mu^*\Big(Q \cap (\bigcup_{i=1}^n A_i)\Big) &= \mu^*\Big(Q \cap \underbrace{(\bigcup_{i=1}^n A_i) \cap A_1}_{=A_1}\Big) + \mu^*\Big(Q \cap \underbrace{(\bigcup_{i=1}^n A_i) \cap A_1^c}_{=\bigcup_{i=2}^n A_i}\Big) \\ &= \ldots = \mu^*\big(Q \cap A_1\big) + \ldots + \mu^*\big(Q \cap A_n\big). \end{aligned}$$

Letting in (1.3.6) $n \to \infty$ we obtain

$$\begin{aligned} \mu^*(Q) &\geq \sum_{i=1}^\infty \mu^*(Q \cap A_i) + \mu^*\Big(Q \cap (\bigcup_{i=1}^\infty A_i)^c\Big) \\ &\geq \mu^*\Big(Q \cap (\bigcup_{i=1}^\infty A_i)\Big) + \mu^*\Big(Q \cap (\bigcup_{i=1}^\infty A_i)^c\Big), \end{aligned}$$

as required.

a.5) Now we show that $\mathcal{M}_{\mu^*}$ is closed under the formation of arbitrary unions of sets. Suppose that $A_1, A_2, \ldots \in \mathcal{M}_{\mu^*}$, and that $Q \subseteq \Omega$. Since $\mathcal{M}_{\mu^*}$ is an algebra on Ω, have that $B_1 := A_1$, and $B_i := A_i \setminus (A_1 \cup \ldots \cup A_{i-1})$ for all $i \geq 2$ belong to $\mathcal{M}_{\mu^*}$. Therefore,

$$\bigcup_{i=1}^\infty A_i = \bigcup_{i=1}^\infty B_i \in \mathcal{M}_{\mu^*}.$$

We conclude from *a*.1), *a*.2), and *a*.5) that $\mathcal{M}_{\mu^*}$ is a σ-algebra on Ω.

b) Now we show that μ^* satisfies the axioms of a measure on $\mathcal{M}_{\mu^*}$.
Since μ^* is by Lemma 1.3.5 an outer measure on 2^Ω, $\mu^*(\emptyset) = 0$. Next we show that μ^* is a σ-additive set function on $\mathcal{M}_{\mu^*}$. Let $A_1, A_2, \ldots \in \mathcal{M}_{\mu^*}$ be disjoint sets. Since μ^* is an outer measure we have

$$\mu^*\Big(\bigcup_{i=1}^\infty A_i\Big) \leq \sum_{i=1}^\infty \mu^*(A_i),$$

and it remains to prove the reverse inequality. Finite additivity follows easily since

$$\begin{aligned}
\mu^*\Big(\bigcup_{i=1}^n A_i\Big) &= \mu^*\Big(\underbrace{(\bigcup_{i=1}^n A_i) \cap A_1}_{=A_1}\Big) + \mu^*\Big(\underbrace{(\bigcup_{i=1}^n A_i) \cap A_1^c}_{=\bigcup_{i=2}^n A_i}\Big) \\
&= \mu^*(A_1) + \mu^*\Big(\bigcup_{i=2}^n A_i\Big) \\
&= \ldots = \sum_{i=1}^n \mu^*(A_i).
\end{aligned}$$

This implies that

$$\sum_{i=1}^n \mu^*(A_i) = \mu^*\Big(\bigcup_{i=1}^n A_i\Big) \le \mu^*\Big(\bigcup_{i=1}^\infty A_i\Big) \qquad \forall n \in \mathbb{N},$$

and, therefore,

$$\sum_{i=1}^\infty \mu^*(A_i) = \lim_{n\to\infty} \sum_{i=1}^n \mu^*(A_i) \le \mu^*\Big(\bigcup_{i=1}^\infty A_i\Big).$$

Hence, μ^* is a σ-additive set function $\mathcal{M}_{\mu^*}$.

(ii) It remains to show that $\sigma(\mathcal{R}) \subseteq \mathcal{M}_{\mu^*}$. To this end, it will be enough to show that $\mathcal{R} \subseteq \mathcal{M}_{\mu^*}$.

Let $A \in \mathcal{R}$ be arbitrary. Since μ^* is an outer measure we have that $\mu^*(Q) \le \mu^*(Q \cap A) + \mu^*(Q \cap A^c)$ holds true for all $Q \subseteq \Omega$, we only have to show that

$$\mu^*(Q) \ge \mu^*(Q \cap A) + \mu^*(Q \cap A^c) \qquad \forall Q \subseteq \Omega. \tag{1.3.7}$$

If $\mu^*(Q) = \infty$, then (1.3.7) is trivial. Otherwise, if $\mu^*(Q) < \infty$, we find for arbitrary $\epsilon > 0$ sets $A_1, A_2, \ldots \in \mathcal{R}$ such that $Q \subseteq \bigcup_{i=1}^\infty A_i$ and

$$\sum_{i=1}^\infty \mu(A_i) \le \mu^*(Q) + \epsilon.$$

This implies

$$\begin{aligned}
&\mu^*(Q \cap A) + \mu^*(Q \cap A^c) \\
&\le \sum_{i=1}^\infty \mu(A_i \cap A) + \sum_{i=1}^\infty \mu(A_i \cap A^c) \\
&= \sum_{i=1}^\infty \mu(A_i) \le \mu^*(Q) + \epsilon,
\end{aligned}$$

which proves (1.3.7). Hence, $\mathcal{R} \subseteq \mathcal{M}_{\mu^*}$. Since $\mathcal{M}_{\mu^*}$ is a σ-algebra, we obtain

$$\sigma(\mathcal{R}) \subseteq \sigma(\mathcal{M}_{\mu^*}) = \mathcal{M}_{\mu^*},$$

as required.

Finally, (i) of Lemma 1.3.5 reveals that

$$\mu^*(A) = \mu(A) \qquad \forall A \in \mathcal{R},$$

which completes the proof. □

Theorem 1.3.6 showed that an arbitrary pre-mesure μ on a ring $\mathcal{R}$ can be extended to a measure on the σ-algebra generated by $\mathcal{R}$. This extension was obtained by a restriction of the corresponding outer measure to $\sigma(\mathcal{R})$. On the other hand, this measure was explicitly specified only for sets that belong to $\mathcal{R}$. In the following we will seek conditions that ensure that this extension is unique, which will then make clear that a specification on $\mathcal{R}$ is indeed sufficient. Before we state the uniqueness theorem, we consider a simple example which will guide us to find conditions that ensure uniqueness.

Suppose that $(\Omega, \mathcal{A})$ is a measurable space and that μ and ν are two measures on $\mathcal{A}$ such that $\mu(\Omega) = \nu(\Omega) < \infty$. Then the set $\mathcal{D} := \big\{A \in \mathcal{A} \colon\ \mu(A) = \nu(A)\big\}$ is a Dynkin system. Indeed, we can verify this property:

a) $\mu(\Omega) = \nu(\Omega) \qquad \rightsquigarrow \qquad \Omega \in \mathcal{D}$,

b) if $A \in \mathcal{D}$, i.e. $\mu(A) = \nu(A)$, then

$$\mu(A^c) \;=\; \mu(\Omega) - \mu(A) \;=\; \nu(\Omega) - \nu(A) \;=\; \nu(A^c)$$

$\rightsquigarrow \qquad A^c \in \mathcal{D}$,

c) if $A_1, A_2, \ldots \in \mathcal{D}$ are disjoint sets, then

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

$\rightsquigarrow \qquad \bigcup_{i=1}^{\infty} A_i \in \mathcal{D}$.

Suppose now the these two measures μ and ν coincide on a collection $\mathcal{E}$ of subsets of Ω, $\mathcal{E} \subseteq \mathcal{A}$, i.e. we have that $\mathcal{E} \subseteq \mathcal{D}$. Let

$$\delta\big(\mathcal{E}\big) \;:=\; \bigcap_{\mathcal{D}\colon\ \mathcal{E}\subseteq\mathcal{D},\, \mathcal{D}\ \text{Dynkin}} \mathcal{D}$$

be the **Dynkin system** which is **generated by** $\mathcal{E}$. It follows from $\mathcal{E} \subseteq \mathcal{D}$ that $\delta\big(\mathcal{E}\big) \subseteq \delta\big(\mathcal{D}\big)$. Since $\delta\big(\mathcal{D}\big) = \mathcal{D}$ we can conclude that the two measures μ and ν coincide on $\delta\big(\mathcal{E}\big)$. If, for some reason, $\delta\big(\mathcal{E}\big) = \sigma\big(\mathcal{E}\big)$, then we obtain that the two measures coincide on the σ-algebra generated by $\mathcal{E}$. The following lemma provides a sufficient condition on $\mathcal{E}$ such that this equality holds. This will also make clear that the restriction $\lambda^*\big|_{\mathcal{B}^d}$ is indeed the **unique** extension of λ_0^d to a measure on $\mathcal{B}^d$.

Lemma 1.3.7. *Let Ω be a non-empty set, and let $\mathcal{E}$ be a collection of subsets of Ω which is closed under the formation of finite intersections. ($\mathcal{E}$ is called to be* **intersection-stable***.) Then*

(i) $\delta\big(\mathcal{E}\big)$ *is intersection-stable,*

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

The good set principle
For the proof of this lemma we will use the so-called **good set principle** which can be described as follows. Suppose that we intend to show that each member of a σ-algebra or Dynkin system $\mathcal{A}$ on Ω has some property $\mathbf{P}$. It is often the case that $\mathcal{A}$ is quite complex and that it is difficult or even impossible to describe all sets that belong to $\mathcal{A}$ in a convenient way. One such example is given by the system of Borel sets $\mathcal{B}^d$ on $\mathbb{R}^d$. Then it often turns out that it is impossible to prove directly that all sets that belong to $\mathcal{A}$ have this property $\mathbf{P}$. A possible solution in such a difficult situation can then be obtained in the following way. Suppose that we have a (typically simple) collection of sets $\mathcal{E}$ that generate $\mathcal{A}$, and that we can show that all members of $\mathcal{E}$ have this property $\mathbf{P}$. We may define, without any hesitation, the **"system of good sets"**,

$$\mathcal{G} := \big\{A \subseteq \Omega: \quad A \text{ has property } \mathbf{P}\big\}.$$

If we are able to show that $\mathcal{G}$ is a σ-algebra (or Dynkin sytem) then we can conclude from $\mathcal{E} \subseteq \mathcal{G}$ that

$$\mathcal{A} = \sigma(\mathcal{E}) \subseteq \sigma(\mathcal{G}) = \mathcal{G},$$

or, likewise, $\mathcal{A} = \delta(\mathcal{E}) \subseteq \delta(\mathcal{G}) = \mathcal{G}$. This, however, shows that all sets that belong to $\mathcal{A}$ have property $\mathbf{P}$.

Proof of Lemma 1.3.7.

(i) For each $D \in \delta(\mathcal{E})$, we define a corresponding system of good sets,

$$\mathcal{G}_D := \big\{M \subseteq \Omega: \ M \cap D \in \delta(\mathcal{E})\big\}.$$

We have to show that $\delta(\mathcal{E}) \subseteq \mathcal{G}_D$. This will be accomplished in two steps.

1) *For each $D \in \delta(\mathcal{E})$, the collection $\mathcal{G}_D$ is a Dynkin system:*

   a) Since $\Omega \cap D = D$ we have $\Omega \cap D \in \delta(\mathcal{E})$, hence $\Omega \in \mathcal{G}_D$.

   b) Let $A \in \mathcal{G}_D$, i.e. $A \cap D \in \delta(\mathcal{E})$. In order to show that $A^c \in \mathcal{G}_D$ holds we represent $A^c \cap D$ in an appropriate form:

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

   Since $D^c \in \delta(\mathcal{E})$, $A \cap D \in \delta(\mathcal{E})$, and D^c and $A \cap D$ are disjoint sets we conclude that $A^c \cap D \in \delta(\mathcal{E})$, i.e. $A^c \in \mathcal{G}_D$.

   c) Finally, if $A_1, A_2, \ldots$ are disjoint sets that belong to $\mathcal{G}_D$, then $A_1 \cap D, A_2 \cap D, \ldots$ are also disjoint. Since these sets belong to $\delta(\mathcal{E})$ it follows that $\big(\bigcup_{i=1}^\infty A_i\big) \cap D = \bigcup_{i=1}^\infty (A_i \cap D) \in \delta(\mathcal{E})$, i.e. $\bigcup_{i=1}^\infty A_i \in \mathcal{G}_D$.

2) *$\delta(\mathcal{E})$ is intersection-stable, i.e. $\delta(\mathcal{E}) \cap \delta(\mathcal{E}) \subseteq \delta(\mathcal{E})$:*

   a) First, let $E \in \mathcal{E}$ be arbitrary. Since $\mathcal{E}$ is intersection-stable, we have that $\mathcal{E} \subseteq \mathcal{G}_E$, which implies that $\delta(\mathcal{E}) \subseteq \delta(\mathcal{G}_E) = \mathcal{G}_E$. In other words, we have that

   $$\delta(\mathcal{E}) \cap \mathcal{E} \subseteq \delta(\mathcal{E}).$$

b) This relation is equivalent to $\mathcal{E} \cap \delta(\mathcal{E}) \subseteq \delta(\mathcal{E})$, i.e. for arbitrary $D \in \delta(\mathcal{E})$ we have that $\mathcal{E} \subseteq \mathcal{G}_D$. Using once more the fact that $\mathcal{G}_D$ is a Dynkin system we obtain $\delta(\mathcal{E}) \subseteq \delta(\mathcal{G}_D) = \mathcal{G}_D$. Since $D \in \delta(\mathcal{E})$ was arbitrarily chosen we conclude that

$$\delta(\mathcal{E}) \cap \delta(\mathcal{E}) \subseteq \delta(\mathcal{E}).$$

(ii) We show that $\delta(\mathcal{E})$ is actually a σ-algebra on Ω. Since $\delta(\mathcal{E})$ is a Dynkin system we only have to show that $\delta(\mathcal{E})$ is stable under the formation of arbitrary unions of sets. Let $A_1, A_2, \ldots \in \delta(\mathcal{E})$ be arbitrary. We represent $\bigcup_{i=1}^{\infty} A_i$ as a union of **disjoint** members of $\delta(\mathcal{E})$. To this end, we define $B_1 := A_1$, and for $i \geq 2$ $B_i := A_i \cap A_1^c \cap \cdots \cap A_{i-1}^c$. Since $\delta(\mathcal{E})$ is an intersection-stable Dynkin system we have that the disjoint sets $B_1, B_2, \ldots$ are members of $\delta(\mathcal{E})$, and we obtain that

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i \in \delta(\mathcal{E})$$

□

Theorem 1.3.8. (Uniqueness theorem)
Suppose that Ω is a non-empty set and that $\mathcal{E}$ is an intersection-stable collection of subsets of Ω. Let μ and ν be measures on $\sigma(\mathcal{E})$ such that

(i) $\mu(E) = \nu(E) \qquad \forall E \in \mathcal{E}$,

(ii) there exist sets $E_1 \subseteq E_2 \subseteq \ldots$ that belong to $\mathcal{E}$, $\bigcup_{n=1}^{\infty} E_n = \Omega$, and $\mu(E_n) = \nu(E_n) < \infty \quad \forall n \in \mathbb{N}$.

Then

$$\mu(A) = \nu(A) \qquad \forall A \in \sigma(\mathcal{E}).$$

Proof. We prove that, for all $n \in \mathbb{N}$,

$$\mu(A \cap E_n) = \nu(A \cap E_n) \qquad \forall A \in \sigma(\mathcal{E}). \tag{1.3.8}$$

Since $E_n \nearrow \Omega$ we obtain from continuity from below that

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

holds for all $A \in \sigma(\mathcal{E})$, which completes the proof.

To prove (1.3.8), we define the corresponding system of good sets,

$$\mathcal{G}_n := \{A \in \sigma(\mathcal{E}) : \ \mu(A \cap E_n) = \nu(A \cap E_n)\}.$$

It follows that $\mathcal{G}_n$ is a Dynkin system on Ω. Indeed, we have

a) $\mu(\Omega \cap E_n) = \mu(E_n) = \nu(E_n) = \nu(\Omega \cap E_n)$
$\rightsquigarrow \qquad \Omega \in \mathcal{G}_n$,

b) if $A \in \mathcal{G}_n$, then

$$\mu\big(A^c \cap E_n\big) \;=\; \mu\big(\Omega \cap E_n\big) \;-\; \mu\big(A \cap E_n\big) \;=\; \nu\big(\Omega \cap E_n\big) \;-\; \nu\big(A \cap E_n\big) \;=\; \nu\big(A^c \cap E_n\big)$$

$$\rightsquigarrow \qquad A^c \in \mathcal{G}_n,$$

c) if $A_1, A_2, \ldots \in \mathcal{G}_n$ are disjoint sets, then

$$\begin{aligned}\mu\Big(\big(\bigcup_{i=1}^{\infty} A_i\big) \cap E_n\Big) \;&=\; \mu\Big(\bigcup_{i=1}^{\infty}(A_i \cap E_n)\Big) \;=\; \sum_{i=1}^{\infty} \mu\big(A_i \cap E_n\big) \\ &=\; \sum_{i=1}^{\infty} \nu\big(A_i \cap E_n\big) \;=\; \nu\Big(\bigcup_{i=1}^{\infty}(A_i \cap E_n)\Big) \;=\; \nu\Big(\big(\bigcup_{i=1}^{\infty} A_i\big) \cap E_n\Big),\end{aligned}$$

$$\rightsquigarrow \qquad \textstyle\bigcup_{i=1}^{\infty} A_i \in \mathcal{G}_n.$$

Now we obtain from from $\mathcal{E} \subseteq \mathcal{G}_n$ that $\delta\big(\mathcal{E}\big) \subseteq \delta\big(\mathcal{G}_n\big) = \mathcal{G}_n$. Since $\mathcal{E}$ is intersection-stable we conclude from Lemma 1.3.7 that $\sigma\big(\mathcal{E}\big) = \delta\big(\mathcal{E}\big)$, which proves (1.3.8). □

Remark 1.3.9. *The condition that $\mathcal{E}$ is an intersection-stable collection of sets is essential for the validity of Theorem 1.3.8. To see this, consider the following simple example. Suppose that we want to model the experiment consisting of two coin tosses. If we denote heads/tails with 1/0, then the possible outcomes are $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$. If we assume that the two coin tosses are carried out independently, then we would assign the probabilities $P_1\big(\{(0,0)\}\big) = P_1\big(\{(0,1)\}\big) = P_1\big(\{(1,0)\}\big) = P_1\big(\{(1,1)\}\big) = 1/4$. Otherwise, if the experimenter cheats by tossing the coin only once but reporting either $(0,0)$ or $(1,1)$, then we had to choose a probability measure P_2 such that $P_2\big(\{(0,0)\}\big) = P_2\big(\{(1,1)\}\big) = 1/2$ and $P_2\big(\{(0,1)\}\big) = P_2\big(\{(1,0)\}\big) = 0$.*

Consider the collection of sets

$$\mathcal{E} \;=\; \Big\{\big\{(0,0),(0,1)\big\}, \big\{(0,0),(1,0)\big\}, \big\{(1,1),(0,1)\big\}, \big\{(1,1),(1,0)\big\}\Big\}.$$

Then $\mathcal{E}$ is a collection of sets that is **not** *intersection-stable, $\sigma(\mathcal{E}) = 2^{\Omega}$, and*

$$P_1\big(E\big) \;=\; 1/2 \;=\; P_2\big(E\big) \qquad \forall E \in \mathcal{E},$$

but we have $P_1 \neq P_2$.

1.4 Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}^d)$

Now we turn to a few applications of the methodology of defining measures. We begin with Lebesgue measure. Recall that our intention is to specify a measure on $(\mathbb{R}^d, \mathcal{B}^d)$ which assigns the volume to suitable subsets of $\mathbb{R}^d$. For rectangles $(a, b] = (a_1, b_1] \times \cdots \times (a_d, b_d]$, the volume is given by

$$\lambda_0^d\big((a, b]\big) \;=\; \prod_{i=1}^{d} (b_i - a_i), \tag{1.4.1}$$

and for finite unions of disjoint rectangles $A_1, \ldots, A_k$ by

$$\lambda_0^d\big(A_1 \cup \cdots \cup A_k\big) \;=\; \sum_{i=1}^{k} \lambda_0^d\big(A_i\big).$$

Lemma 1.3.4 states that λ_0^d is the unique pre-measure on the ring $\mathcal{B}_0^d$ generated by the half-open rectangles that fulfills (1.4.1). As an intermediate step to our intended definition of Lebesgue measure we define the outer measure λ^*, where for an **arbitrary subset** Q of $\mathbb{R}^d$,

$$\lambda^*\big(Q\big) \;=\; \inf\Big\{ \sum_{i=1}^{\infty} \lambda_0^d(A_i) \colon \quad Q \subseteq \bigcup_{i=1}^{\infty} A_i,\ A_1, A_2, \ldots \in \mathcal{B}_0^d \Big\}.$$

(Note that the set $\big\{A_1, A_2, \ldots \in \mathcal{B}_0^d \colon\ Q \subseteq \bigcup_{i=1}^{\infty} A_i\big\}$ is always non-empty.) Let

$$\mathcal{M}_{\lambda^*} \;:=\; \big\{ A \in 2^{\Omega} \colon\ \ \lambda^*\big(Q\big) \;=\; \lambda^*\big(Q \cap A\big) \;+\; \lambda^*\big(Q \cap A^c\big) \quad \forall Q \in 2^{\Omega} \big\}.$$

be the collection of all λ^*-measurable subsets of $\mathbb{R}^d$. It follows from Carathéodory's extension theorem (Theorem 1.3.6) that $\mathcal{M}_{\lambda^*}$ is a σ-algebra on $\mathbb{R}^d$ and that $\lambda^*\big|_{\mathcal{M}_{\lambda^*}}$ is a measure on $(\mathbb{R}^d, \mathcal{M}_{\lambda^*})$ such that $\lambda^*\big|_{\mathcal{M}_{\lambda^*}}\big(A\big) = \lambda_0^d\big(A\big)$ for all $A \in \mathcal{B}_0^d$. Furthermore, $\mathcal{M}_{\lambda^*}$ contains $\sigma(\mathcal{B}_0^d) = \mathcal{B}^d$ which implies that $\lambda^*\big|_{\mathcal{B}^d}$ is a measure on $(\mathbb{R}^d, \mathcal{B}^d)$ such that $\lambda^*\big|_{\mathcal{B}^d}\big(A\big) = \lambda_0^d\big(A\big)$ for all $A \in \mathcal{B}_0^d$. In other words, $\lambda^*\big|_{\mathcal{B}^d}$ is an extension of the pre-measure λ_0^d. Since the ring $\mathcal{B}_0^d$ is intersection-table, we obtain from the uniqueness theorem (Theorem 1.3.8) that this is the only possible extension of λ_0^d to a measure on $\sigma(\mathcal{B}_0^d) = \mathcal{B}^d$. Following the terminology in Billingsley [1] and Cohn [2], we call the restriction $\lambda^*\big|_{\mathcal{M}_{\lambda^*}}$ of the outer measure λ^* to the collection of λ^*-measurable sets **Lebesgue measure** and denote it simply by λ^d. The restriction $\lambda^*\big|_{\mathcal{B}^d}$ is also called **Lebesgue measure**, and it too will be denoted by λ^d. We can specify which version of Lebesgue measure we intend by referring, for example, to Lebesgue measure on $(\mathbb{R}^d, \mathcal{M}_{\lambda^*})$ or to Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}^d)$. Note that some other authors distinguish between these two measures. For example, Elstrodt [3] also calls $\lambda^*\big|_{\mathcal{M}_{\lambda^*}}$ Lebesgue measure but its restriction $\lambda^*\big|_{\mathcal{B}^d}$ to Borel sets Lebesgue-Borel measure.

In the following we take a closer look at some properties of Lebesgue measure. As a by-product, this will also provide some insight in the σ-algebras $\mathcal{M}_{\lambda^*}$ and $\mathcal{B}^d$. First we prove translation-invariance of Lebesgue outer measure λ^*, which implies this property for its restrictions, Lebesgue measure on $(\mathbb{R}^d, \mathcal{M}_{\lambda^*})$ and $(\mathbb{R}^d, \mathcal{B}^d)$, respectively. For these implications to make sense, it is required that the σ-algebras $\mathcal{M}_{\lambda^*}$ and $\mathcal{B}^d$ are closed under translations, which will also be shown in the next proposition. For each subset A and each element x of $\mathbb{R}^d$ we shall denote by $A + x$ the subset of $\mathbb{R}^d$ defined by

$$A + x \;=\; \big\{ y + x \colon \quad y \in A \big\}.$$

The set $A + x$ is called the **translate** of A by x.

Proposition 1.4.1.

(i) *For arbitrary* $Q \subseteq \mathbb{R}^d$ *and* $x \in \mathbb{R}^d$,

$$\lambda^*\big(Q\big) \;=\; \lambda^*\big(Q+x\big).$$

(ii) *If* $A \in \mathcal{M}_{\lambda^*}$, $x \in \mathbb{R}^d$, *then* $A + x \in \mathcal{M}_{\lambda^*}$.

(iii) *If* $A \in \mathcal{B}^d$, $x \in \mathbb{R}^d$, *then* $A + x \in \mathcal{B}^d$.

Proof. (i) This statement follows from the obvious translation-invariance of the underlying pre-measure λ_0^d on $\mathcal{B}_0^d$. Let $Q \subseteq \mathbb{R}^d$ and $x \in \mathbb{R}^d$ be arbitrary. We show that

$$\lambda^*\big(Q\big) \;\geq\; \lambda^*\big(Q+x\big). \tag{1.4.2}$$

Case 1: If $\lambda^*(Q) = \infty$, then (1.4.2) is obviously fulfilled.
Case 2: If $\lambda^*(Q) < \infty$, then we find for arbitrary $\epsilon > 0$ sets $A_1, A_2, \ldots \in \mathcal{B}_0^d$ such that $Q \subseteq \bigcup_{i=1}^{\infty} A_i$ and

$$\lambda^*(Q) + \epsilon \;\geq\; \sum_{i=1}^{\infty} \lambda_0^d(A_i).$$

Note that $A_1 + x, A_2 + x, \ldots$ also belong to $\mathcal{B}_0^d$ and $Q + x \subseteq \bigcup_{i=1}^{\infty}(A_i + x)$. Since $\lambda_0^d(A_i) = \lambda_0^d(A_i + x)$ for all $i \in \mathbb{N}$ we obtain that

$$\begin{aligned}
\lambda^*\big(Q\big) \;+\; \epsilon \;\;&\geq\;\; \sum_{i=1}^{\infty} \lambda_0^d\big(A_i\big) \;=\; \sum_{i=1}^{\infty} \lambda_0^d\big(A_i + x\big) \\
&\geq\;\; \inf\Big\{ \sum_{i=1}^{\infty} \lambda_0^d\big(B_i\big) : \quad B_1, B_2, \ldots \in \mathcal{B}_0^d,\; Q + x \subseteq \bigcup_{i=1}^{\infty} B_i \Big\} \\
&=\;\; \lambda^*\big(Q+x\big),
\end{aligned}$$

which proves (1.4.2). The reverse inequality can be shown analogously, which proves (i).

(ii) Let $A \in \mathcal{M}_{\lambda^*}$ be arbitrary. Then, for all $Q \subseteq \mathbb{R}^d$,

$$\lambda^*\big(Q\big) \;=\; \lambda^*\big(Q \cap A\big) \;+\; \lambda^*\big(Q \cap A^c\big).$$

Using the translation-invariance of λ^* we obtain

$$\begin{aligned}
\lambda^*\big(Q\big) \;&=\; \lambda^*\big(Q - x\big) \;=\; \lambda^*\big((Q-x) \cap A\big) \;+\; \lambda^*\big((Q-x) \cap A^c\big) \\
&=\; \lambda^*\big(Q \cap (A + x)\big) \;+\; \lambda^*\big(Q \cap \underbrace{(A^c + x)}_{=\,(A+x)^c}\big) \\
&=\; \lambda^*\big(Q \cap (A + x)\big) \;+\; \lambda^*\big(Q \cap (A + x)^c\big).
\end{aligned}$$

Hence, $A + x \in \mathcal{M}_{\lambda^*}$.

(iii) Define an appropriate **system of good sets**,

$$\mathcal{G} \;:=\; \big\{A \subseteq \mathbb{R}^d : \quad A + x \in \mathcal{B}^d\big\}.$$

If A is an open set, then $A + x$ is open as well. Hence, $\mathcal{O}^d \subseteq \mathcal{G}$.

Furthermore, $\mathcal{G}$ is a σ-algebra on $\mathbb{R}^d$. Indeed,

a) $\Omega + x = \Omega \in \mathcal{B}^d \quad \rightsquigarrow \quad \Omega \in \mathcal{G}$

b) If $A \in \mathcal{G}$, i.e. $A + x \in \mathcal{B}^d$, then $A^c + x = (A + x)^c \in \mathcal{B}^d. \quad \rightsquigarrow \quad A^c \in \mathcal{G}$

c) If $A_1, A_2, \ldots \in \mathcal{G}$, then $A_1 + x, A_2 + x, \ldots \in \mathcal{B}^d$, and so $\big(\bigcup_{i=1}^\infty A_i\big) + x = \bigcup_{i=1}^\infty (A_i + x) \in \mathcal{B}^d. \quad \rightsquigarrow \quad \bigcup_{i=1}^\infty A_i \in \mathcal{G}$

Now we obtain

$$\mathcal{B}^d = \sigma\big(\mathcal{O}^d\big) \subseteq \sigma\big(\mathcal{G}\big) = \mathcal{G},$$

i.e. for all $A \in \mathcal{B}^d$ we have that $A + x \in \mathcal{B}^d$.

□

The property of translation invariance of outer Lebesgue measure λ^* stated in Proposition 1.4.1 allows us to show that there exist subsets of $\mathbb{R}$ that are not Lebesgue measurable.

Corollary 1.4.2. *There is a subset E of the interval $(0,1)$ that is not Lebesgue measurable, i.e. $\mathcal{M}_{\lambda^*} \subset 2^{\mathbb{R}}$.*

Proof. Define a relation $\sim$ on $\mathbb{R}$ by letting $x \sim y$ hold if and only if $x - y$ is rational. It is easy to check that $\sim$ is an equivalence relation: it is reflexive ($x \sim x$ holds for each x), symmetric ($x \sim y$ implies $y \sim x$), and transitive ($x \sim y$ and $y \sim z$ imply $x \sim z$). This relation partitions $\mathbb{R}$ into disjoint equivalence classes which have the respective form $\mathbb{Q}+x$ for some x. Since each equivalence class contains a number in the interval $(0,1)$ we can choose from each class exactly one element that belongs to $(0,1)$. Let E be the collection of these elements. Note that there does not exist a deterministic rule of specifying the elements of this set; here the acceptance of the **axiom of choice** is required.

(The axiom of choice can be formulated as follows: Given any collection $\mathcal{M}$ of pairwise disjoint non-empty subsets of a set Ω, it is possible to assemble a new set containing exactly one element from each member of the given collection. This axiom is often formulated in terms of choice functions, i.e. there exists a function $f\colon \mathcal{M} \to \Omega$ such that $f(M) \in M$ for all $M \in \mathcal{M}$. As mentioned above, the set E is non-constructible, i.e. we cannot find a deterministic rule of specifying the elements of this set. However, the possibility of choosing such a set is taken for granted in ordinary mathematical analysis.)

Let $(r_n)_{n\in\mathbb{N}}$ be an enumeration of the rational numbers in the interval $(-1,1)$, and for each n let $E_n = E + r_n$. It is easy to see that the sets E_n are disjoint. Indeed, if $E_m \cap E_n \neq \emptyset$ for some $m \neq n$, then there are elements e and e' of E such that $e + r_m = e' + r_n$. However, this means that $e \sim e'$, and since E contains only one element from each equivalence class we obtain that $e = e'$. Hence, $r_n = r_m$ which contradicts $m \neq n$. Furthermore, by translation-invariance of Lebesgue outer measure λ^* we have $\lambda^*(E_n) = \lambda^*(E)$ for each $n \in \mathbb{N}$. Suppose now that E is Lebesgue measurable. Then for each n the set E_n is also Lebesgue measurable; see statement (ii) of Proposition 1.4.1. We obtain that

$$\lambda^*\Big(\bigcup_{n=1}^\infty E_n\Big) = \sum_{n=1}^\infty \lambda^*(E_n).$$

Now we shall obtain a contradiction: Since $(0,1) \subseteq \bigcup_{n=1}^\infty E_n$, we have that $\lambda^*\big(\bigcup_{n=1}^\infty E_n\big) \geq \lambda^*((0,1)) = 1$; hence $\lambda^*(E) = 0$ is impossible. On the other hand, $\bigcup_{n=1}^\infty E_n$ is obviously contained in $(-1,2)$, hence $\lambda^*\big(\bigcup_{n=1}^\infty E_n\big) \leq \lambda^*\big((-1,2)\big) = 3$. Therefore, $\lambda^*(E) > 0$ is also impossible and the assumption that E is Lebesgue measurable leads to a contradiction. □

Proposition 1.4.3.
Let λ^* be Lebesgue outer measure on $(\mathbb{R}^d, 2^{\mathbb{R}^d})$ and let $Q \subseteq \mathbb{R}^d$. The following statements are equivalent:

(i) $Q \in \mathcal{M}_{\lambda^*}$.

(ii) *For all $\epsilon > 0$, there exists an open set U_ϵ and a closed set F_ϵ such that*

$$F_\epsilon \subseteq Q \subseteq U_\epsilon \qquad \text{and} \qquad \lambda^d\big(U_\epsilon \setminus F_\epsilon\big) \leq \epsilon.$$

(iii) *There exist sets $A_1, A_2 \in \mathcal{B}^d$ such that*

$$A_1 \subseteq Q \subseteq A_2 \qquad \text{and} \qquad \lambda^d\big(A_2 \setminus A_1\big) = 0.$$

Proof. We prove
a) that (i) implies (ii),
b) that (ii) implies (iii),
and
c) that (iii) implies (i).

a) Let $Q \in \mathcal{M}_{\lambda^*}$ and $\epsilon > 0$ be arbitrary. We will first prove that there exists an open subset U_ϵ of $\mathbb{R}^d$ such that

$$Q \subseteq U_\epsilon \qquad \text{and} \qquad \lambda^*\big(U_\epsilon \setminus Q\big) \leq \epsilon/2. \tag{1.4.3}$$

Since $\mathcal{M}_{\lambda^*}$ is a σ-algebra we have that Q^c is also a member of $\mathcal{M}_{\lambda^*}$, and we conclude from (1.4.3) that there exists an open subset V_ϵ of $\mathbb{R}^d$ such that $Q^c \subseteq V_\epsilon$ and $\lambda^*\big(V_\epsilon \setminus Q^c\big) \leq \epsilon/2$. Then $F_\epsilon := V_\epsilon^c$ is, as a complement of an open set, a closed subset of $\mathbb{R}^d$, $F_\epsilon \subseteq Q$, and $\lambda^*\big(Q \setminus F_\epsilon\big) = \lambda^*\big(V_\epsilon \setminus Q^c\big) \leq \epsilon/2$. Hence,

$$F_\epsilon \subseteq Q \subseteq U_\epsilon,$$

and

$$\lambda^d\big(U_\epsilon \setminus F_\epsilon\big) = \lambda^*\big(U_\epsilon \setminus F_\epsilon\big) = \lambda^*\big(U_\epsilon \setminus Q\big) + \lambda^*\big(Q \setminus F_\epsilon\big) \leq \epsilon,$$

as required.

We turn to the proof of (1.4.3). We consider first a set A that belongs to $\mathcal{B}_0^d$. Then $A = I_1 \cup \cdots \cup I_k$, where $I_i = \big(a_1^{(i)}, b_1^{(i)}\big] \times \cdots \times \big(a_d^{(i)}, b_d^{(i)}\big]$. The rectangle I_i is covered by each of the open sets $U_{n,i} := \big(a_1^{(i)}, b_1^{(i)} + 1/n\big) \times \cdots \times \big(a_d^{(i)}, b_d^{(i)} + 1/n\big)$. Then $U_n := U_{n,1} \cup \cdots \cup U_{n,k}$ is an open set and $U_n \searrow A$. Since $\lambda^d\big(U_n\big) < \infty$ it follows from continuity from above that

$$\lambda^d\big(U_n\big) \searrow \lambda^d\big(A\big) = \lambda^*\big(A\big). \tag{1.4.4}$$

Consider now a set $Q \in \mathcal{M}_{\lambda^*}$.
Case 1: Suppose that $\lambda^*\big(Q\big) < \infty$. Then there exist sets $A_1, A_2, \ldots \in \mathcal{B}_0^d$ such that $Q \subseteq \bigcup_{i=1}^\infty A_i$ and

$$\lambda^*\big(Q\big) \geq \sum_{i=1}^\infty \lambda^d(A_i) - \epsilon/4.$$

Furthermore, it follows from (1.4.4) that there exist open sets $U_1, U_2, \ldots$ such that $A_i \subseteq U_i$ and

$$\lambda^d\big(A_i\big) \geq \lambda^d\big(U_i\big) - \epsilon 2^{-(i+2)} \qquad \text{for all } i \geq 1.$$

Then $U_\epsilon := \bigcup_{i=1}^\infty U_i$ is an open set, $Q \subseteq U_\epsilon$, and we obtain

$$\begin{aligned} \lambda^*\big(Q\big) &\geq \sum_{i=1}^\infty \lambda^d\big(A_i\big) \quad - \quad \epsilon/4 \\ &\geq \sum_{i=1}^\infty \Big(\lambda^d\big(U_i\big) \ - \ \epsilon 2^{-(i+2)}\Big) \quad - \quad \epsilon/4 \\ &\geq \sum_{i=1}^\infty \lambda^d\big(U_i\big) \ - \ \epsilon/2 \ \geq \ \lambda^d\big(U_\epsilon\big) \ - \ \epsilon/2. \end{aligned}$$

Since $Q \in \mathcal{M}_{\lambda^*}$ we have $\lambda^*\big(U_\epsilon\big) = \lambda^*\big(\underbrace{U_\epsilon \cap Q}_{=Q}\big) + \lambda^*\big(U_\epsilon \setminus Q\big)$, and therefore

$$\lambda^*\big(U_\epsilon \setminus Q\big) \ = \ \lambda^*\big(U_\epsilon\big) \ - \ \lambda^*\big(Q\big) \ \leq \ \epsilon/2. \tag{1.4.5}$$

Case 2: It remains to consider the case where $Q \in \mathcal{M}_{\lambda^*}$ and $\lambda^*\big(Q\big) = \infty$. Let $C_n = (-n, n] \times \cdots \times (-n, n] \in \mathbb{R}^d$. Then $Q \cap C_n \in \mathcal{M}_{\lambda^*}$ and $\lambda^*\big(Q \cap C_n\big) \leq (2n)^d$. Hence we obtain from (1.4.5) that there exist open sets $U_{n,\epsilon}$ such that $Q \cap C_n \subseteq U_{n,\epsilon}$ and $\lambda^*\big(U_{n,\epsilon} \setminus (Q \cap C_n)\big) \leq \epsilon 2^{-(n+1)}$. The set $U_\epsilon := \bigcup_{n=1}^\infty U_{n,\epsilon}$ is an open subset of $\mathbb{R}^d$, $\mathbb{Q} \subseteq U_\epsilon$, and

$$\lambda^*\big(U_\epsilon \setminus Q\big) \ \leq \ \sum_{n=1}^\infty \lambda^*\big(U_{n,\epsilon} \setminus (Q \cap C_n)\big) \ \leq \ \epsilon/2.$$

Hence, (1.4.3) is proved for all $Q \in \mathcal{M}_{\lambda^*}$, which completes the proof of the implication $(i) \rightsquigarrow (ii)$.

b) Let $Q \subseteq \mathbb{R}^d$ be such that there exist open sets $U_1, U_2, \ldots$ and closed sets $F_1, F_2, \ldots$ with

$$F_n \ \subseteq \ Q \ \subseteq \ U_n \qquad \text{and} \qquad \lambda^d\big(U_n \setminus F_n\big) \ \leq \ 1/n$$

hold for all $n \in \mathbb{N}$. Let $A_1 = \bigcup_{n=1}^\infty F_n$ and $A_2 = \bigcap_{n=1}^\infty U_n$. With this choice, we have that $A_1, A_2 \in \mathcal{B}^d$,

$$A_1 \ \subseteq \ Q \ \subseteq \ A_2$$

and, since $A_2 \setminus A_1 \subseteq U_n \setminus F_n \ \forall n \in \mathbb{N}$,

$$\lambda^d\big(A_2 \setminus A_1\big) \ \leq \ \lambda^d\big(U_n \setminus F_n\big) \underset{n \to \infty}{\longrightarrow} 0.$$

c) Suppose that $Q \subseteq \mathbb{R}^d$ is such that there exist Borel sets $A_1, A_2 \in \mathcal{B}^d$, where $A_1 \subseteq Q \subseteq A_2$ and $\lambda^d\big(A_2 \setminus A_1\big) = 0$. For $Q \in \mathcal{M}_{\lambda^*}$ to hold, we have to show that, for an arbitrary $R \subseteq \mathbb{R}^d$,

$$\lambda^*\big(R\big) \ = \ \lambda^*\big(R \cap Q\big) \ + \ \lambda^*\big(R \cap Q^c\big).$$

Since $A_1 \in \mathcal{B}^d \subseteq \mathcal{M}_{\lambda^*}$ we have that

$$\lambda^*\big(R\big) \ = \ \lambda^*\big(R \cap A_1\big) \ + \ \lambda^*\big(R \cap A_1^c\big).$$

We obtain by monotonicity of λ^* that

$$\begin{aligned}\lambda^*\big(R\cap A_1\big) &\le \lambda^*\big(R\cap Q\big) \le \lambda^*\big(R\cap A_2\big)\\ &\le \lambda^*\big(R\cap A_1\big) + \lambda^*\big(R\cap (A_2\setminus A_1)\big) \le \lambda^*\big(R\cap A_1\big) + \underbrace{\lambda^*\big(A_2\setminus A_1\big)}_{=0},\end{aligned}$$

and analogously

$$\begin{aligned}\lambda^*\big(R\cap A_2^c\big) &\le \lambda^*\big(R\cap Q^c\big) \le \lambda^*\big(R\cap A_1^c\big)\\ &\le \lambda^*\big(R\cap A_2^c\big) + \lambda^*\big(R\cap (A_1^c\setminus A_2^c)\big) \le \lambda^*\big(R\cap A_2^c\big) + \underbrace{\lambda^*\big(A_1^c\setminus A_2^c\big)}_{=\lambda^*(A_2\setminus A_1)=0}.\end{aligned}$$

Since the left-hand sides of these chains of inequalities coincide with the respective right-hand sides, all of these inequalities are in fact equalities. Hence, $\lambda^*\big(R\cap Q\big) = \lambda^*\big(R\cap A_1\big)$ and $\lambda^*\big(R\cap Q^c\big) = \lambda^*\big(R\cap A_1^c\big)$, which implies that

$$\lambda^*\big(R\big) = \lambda^*\big(R\cap Q\big) + \lambda^*\big(R\cap Q^c\big),$$

as required. Hence, $Q\in\mathcal{M}_{\lambda^*}$.

□

It is sometimes convenient to be able to deal with arbitrary subsets of sets of measure zero. This leads to the following definition.

Definition. Let $(\Omega,\mathcal{A},\mu)$ be a measure space. The measure μ (or the measure space $(\Omega,\mathcal{A},\mu)$) is **complete** if the relations $A\in\mathcal{A}$, $\mu(A)=0$, and $B\subseteq A$ together imply that $B\in\mathcal{A}$. A set B with this property is called **μ-negligible** or **μ-null**.

Lemma 1.4.4. *Let μ^* be an outer measure on $(\Omega,2^\Omega)$ and let $\mathcal{M}_{\mu^*}$ be the σ-algebra of all μ^*-measurable subsets of Ω. Then the restriction of μ^* to $\mathcal{M}_{\mu^*}$ is complete.*

Proof. Let $A\in\mathcal{M}_{\mu^*}$, $\mu^*(A)=0$, and $B\subseteq A$. We have to show that, for arbitrary $Q\subseteq\Omega$,

$$\mu^*\big(Q\big) = \mu^*\big(Q\cap B\big) + \mu^*\big(Q\cap B^c\big). \tag{1.4.6}$$

Since $\mu^*(Q\cap B)\le\mu^*(B)\le\mu^*(A)=0$ we obtain that

$$\mu^*\big(Q\cap B^c\big) \le \mu^*\big(Q\big) \le \mu^*\big(Q\cap B^c\big) + \underbrace{\mu^*\big(Q\cap B\big)}_{=0},$$

which implies that $\mu^*(Q)=\mu^*(Q\cap B^c)$, and so (1.4.6). Hence $B\in\mathcal{M}_{\mu^*}$. □

Lemma 1.4.4 implies in particular that the restriction of Lebesgue outer measure to $\mathcal{M}_{\lambda^*}$, i.e. $\lambda^*\big|_{\mathcal{M}_{\lambda^*}}$, is complete.

If we have an arbitrary, possibly non-complete measure μ on a measurable space $(\Omega,\mathcal{A})$, then it is quite easy to extend this measure to a complete one. The following proposition provides a simple method to achieve this.

Proposition 1.4.5.
Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, and let

$$\mathcal{N} := \big\{N \subseteq \Omega: \quad \exists N_0 \in \mathcal{A},\ \mu(N_0) = 0,\ \text{and } N \subseteq N_0\big\}$$

be the collection of all μ-negligible subsets of Ω. Then

(i) $\widetilde{\mathcal{A}} := \big\{A \cup N: \quad A \in \mathcal{A}, N \in \mathcal{N}\big\}$ *is a σ-algebra on Ω.*

(ii) *The set function $\widetilde{\mu}\colon \widetilde{\mathcal{A}} \to [0, \infty]$ defined by*

$$\widetilde{\mu}\big(A \cup N\big) := \mu\big(A\big) \qquad \forall A \in \mathcal{A},\ \forall N \in \mathcal{N}$$

is the unique extension of μ to a measure on $(\Omega, \widetilde{\mathcal{A}})$. The measure space $\big(\Omega, \widetilde{\mathcal{A}}, \widetilde{\mu}\big)$ is complete and is called the **completion** *of the measure space $(\Omega, \mathcal{A}, \mu)$.*

Proof. (i) It is easy to see that $\widetilde{\mathcal{A}}$ fulfills the axioms of a σ-algebra.

a) Since $\Omega = \underbrace{\Omega}_{\in \mathcal{A}} \cup \underbrace{\emptyset}_{\in \mathcal{N}}$ we have that $\Omega \in \widetilde{\mathcal{A}}$.

b) Let $\widetilde{A} = A \cup N \in \widetilde{\mathcal{A}}$, where $A \in \mathcal{A}$ and $N \in \mathcal{N}$. There exists a μ-null set $N_0 \in \mathcal{A}$ such that $N \subseteq N_0$. Then we can represent the complement of $\widetilde{A}$ as

$$\begin{aligned} \widetilde{A}^c &= \big(A \cup N\big)^c = A^c \cap N^c \\ &= \underbrace{\big(A^c \cap N_0^c\big)}_{\in \mathcal{A}} \cup \underbrace{\big(A^c \cap (N^c \setminus N_0^c)\big)}_{\subseteq N^c \setminus N_0^c = N_0 \setminus N} \end{aligned}$$

Hence, $\widetilde{A}^c \in \widetilde{\mathcal{A}}$.

c) Let $\widetilde{A}_1 = A_1 \cup N_1, \widetilde{A}_2 = A_2 \cup N_2, \ldots \in \widetilde{\mathcal{A}}$, where for each i $A_i \in \mathcal{A}$, $N_i \in \mathcal{N}$. Then

$$\bigcup_{i=1}^{\infty} \widetilde{A}_i = \underbrace{\bigcup_{i=1}^{\infty} A_i}_{\in \mathcal{A}} \cup \underbrace{\bigcup_{i=1}^{\infty} N_i}_{\in \mathcal{N}},$$

which implies that $\bigcup_{i=1}^{\infty} \widetilde{A}_i \in \widetilde{\mathcal{A}}$.

(ii) The proof of (ii) is split into several steps.

a) *Consistency of the definition of $\widetilde{\mu}$*
Suppose that a set $A \in \widetilde{\mathcal{A}}$ has representations $A_1 \cup N_1$ and $A_2 \cup N_2$, where $A_1, A_2 \in \mathcal{A}$ and $N_1, N_2 \in \mathcal{N}$. There exists a μ-null set $C \in \mathcal{A}$ such that $N_2 \subseteq C$, and we obtain

$$\mu\big(A_1\big) \leq \mu\big(A_2 \cup C\big) = \mu\big(A_2\big) + \underbrace{\mu\big(C\big)}_{=0}.$$

Since the reverse inequality can be shown in the same way we have that

$$\mu\big(A_1\big) = \mu\big(A_2\big),$$

i.e. the value of $\widetilde{A}$ does not depend on the chosen representation of the set A.

b) *$\widetilde{\mu}$ is a measure on $\widetilde{\mathcal{A}}$*
While $\widetilde{\mu}(\emptyset) = 0$ is obvious, the σ-additivity of $\widetilde{\mu}$ follows directly from that of μ.

c) *Uniqueness of the extension*
The uniqueness of the extension follows from the required monotonicity of $\widetilde{\mu}$. Indeed, let $\widetilde{A} = A \cup N$, where $A \in \mathcal{A}$ and $N \in \mathcal{N}$. Then there exists $N_0 \in \mathcal{A}$ such that $N \subseteq N_0$ and $\mu(N_0) = 0$. Since $A \subseteq \widetilde{A} \subseteq A \cup N_0$ and $\mu(A) = \mu(A \cup N_0)$, the only possible choice of the value of $\widetilde{\mu}(\widetilde{A})$ which is in line with the required monotonicity is given by $\widetilde{\mu}(\widetilde{A}) = \mu(A)$.

d) *Completeness of* $(\Omega, \widetilde{\mathcal{A}}, \widetilde{\mu})$
Let $N \in \mathcal{N}$ be arbitrary. Since N admits the representation $N = \emptyset \cup N$ we see that $N \in \widetilde{\mathcal{A}}$. Hence, $(\Omega, \widetilde{\mathcal{A}}, \widetilde{\mu})$ is complete.

□

We are now in a position to investigate the relation of the σ-algebras $\mathcal{B}^d$ and $\mathcal{M}_{\lambda^*}$. We know already that $\mathcal{B}^d \subseteq \mathcal{M}_{\lambda^*}$; see (ii) of Theorem 1.3.6. Moreover, it is shown in Proposition 2.1.9 in Cohn [2, page 56] that there exists a Lebesgue measurable subset B of $\mathbb{R}$ such that $\lambda(B) = 0$ that is not a Borel set. The following proposition clarifies the relation between $\mathcal{B}^d$ and $\mathcal{M}_{\lambda^*}$.

Proposition 1.4.6. $\big(\mathbb{R}^d, \mathcal{M}_{\lambda^*}, \lambda^*\big)$ *is the completion of* $\big(\mathbb{R}^d, \mathcal{B}^d, \lambda^d\big)$.

Proof. Let, according to Proposition 1.4.5,

$$\begin{aligned} \mathcal{N} &= \big\{N \subseteq \mathbb{R}^d\colon \quad \exists N_0 \in \mathcal{B}^d,\ \lambda^d(N_0) = 0, \text{ and } N \subseteq N_0\big\}, \\ \widetilde{\mathcal{B}^d} &= \big\{A \cup N\colon \quad A \in \mathcal{B}^d,\ N \in \mathcal{N}\big\}, \end{aligned}$$

and, for $A \in \mathcal{B}^d$, $N \in \mathcal{N}$,

$$\widetilde{\lambda^d}\big(A \cup N\big) = \lambda^d\big(A\big).$$

We show that a) $\widetilde{\mathcal{B}^d} = \mathcal{M}_{\lambda^*}$, and b) $\widetilde{\lambda^d} = \lambda^*\big|_{\mathcal{M}_{\lambda^*}}$.

a) Let $Q \in \mathcal{M}_{\lambda^*}$ be arbitrary. By (iii) of Proposition 1.4.3 there exist sets $A_1, A_2 \in \mathcal{B}^d$ such that

$$A_1 \subseteq Q \subseteq A_2 \qquad \text{and} \qquad \lambda^d\big(A_2 \setminus A_1\big) = 0.$$

Then

$$Q = A_1 \cup \big(Q \cap (A_2 \setminus A_1)\big),$$

where $Q \cap (A_2 \setminus A_1)$ is a λ^d-null set. Hence, $\mathcal{M}_{\lambda^*} \subseteq \widetilde{\mathcal{B}^d}$.

Let now $\widetilde{A} \in \widetilde{\mathcal{B}^d}$. Then there exist $A \in \mathcal{B}^d$ and $N \in \mathcal{N}$ such that $\widetilde{A} = A \cup N$. For arbitrary $Q \subseteq \mathbb{R}^d$, it follows from monotonicity of λ^* that

$$\lambda^*\big(Q \cap A\big) \le \lambda^*\big(Q \cap \widetilde{A}\big) \le \lambda^*\big(Q \cap A\big) + \underbrace{\lambda^*\big(Q \cap N\big)}_{=0},$$

which implies that $\lambda^*(Q \cap A) = \lambda^*(Q \cap \widetilde{A})$. Since $A \in \mathcal{B}^d \subseteq \mathcal{M}_{\lambda^*}$ we obtain

$$\begin{aligned} \lambda^*\big(Q\big) &= \lambda^*\big(Q \cap A\big) + \lambda^*\big(Q \cap A^c\big) \\ &\ge \lambda^*\big(Q \cap \widetilde{A}\big) + \lambda^*\big(Q \cap \widetilde{A}^c\big), \end{aligned}$$

and so

$$\lambda^*(Q) \;=\; \lambda^*(Q \cap \widetilde{A}) \;+\; \lambda^*(Q \cap \widetilde{A}^c).$$

Hence, $\widetilde{\mathcal{B}^d} \subseteq \mathcal{M}_{\lambda^*}$.

b) Let $A \in \mathcal{B}^d$ and $N \in \mathcal{N}$ be arbitrary. Since $\lambda^*(A) \le \lambda^*(A \cup N) \le \lambda^*(A) + \underbrace{\lambda^*(N)}_{=0}$ we obtain that

$$\widetilde{\lambda}^d(A \cup N) \;=\; \lambda^d(A) \;=\; \lambda^*(A) \;=\; \lambda^*(A \cup N).$$

□

1.5 Probability measures

In this section we describe the specification of probability measures on $\mathbb{R}^d$. Recall that a measure P is a **probability measure** on $(\mathbb{R}^d, \mathcal{B}^d)$ if $P(\mathbb{R}^d) = 1$. It should be known from basic courses in probability that there are different types of probability measures. If there exists a countable subset $\{x_1, x_2, \ldots\}$ of $\mathbb{R}^d$ such that $P\big(\{x_1, x_2, \ldots\}\big) = 1$, then P is a discrete probability measure. It can be easily specified by defining the probabilities $P(\{x_i\})$ since the property of σ-additivity implies that

$$P(A) \;=\; \sum_{i\colon x_i \in A} P\big(\{x_i\}\big), \tag{1.5.1}$$

which holds true for all $A \subseteq \mathbb{R}^d$. (Actually, we can use the power set as the corresponding σ-algebra.) Another important case is that of a measure P possessing a (Riemann-integrable) probability density p. In dimension $d = 1$, we then have $P((a, b]) = \int_a^b p(x)\, dx$ for all $a \le b$. Of course, such a measure cannot be specified as in equation (1.5.1) since, for each $x \in \mathbb{R}$, $P(\{x\}) = 0$. A unified description of probability measures on $\mathbb{R}^d$ can be achieved by the corresponding **cumulative distribution functions**. For a given probability measure P on $(\mathbb{R}^d, \mathcal{B}^d)$, its cumulative distribution function $F\colon \mathbb{R}^d \to [0, 1]$ is given by

$$F(x_1, \ldots, x_d) \;:=\; P\big((-\infty, x_1] \times \cdots \times (-\infty, x_d]\big) \qquad \forall (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

If $(x, x+y] = (x_1, x_1+y_1] \times \cdots \times (x_d, x_d+y_d]$ is a rectangle ($y_i \ge 0\ \forall i = 1, \ldots, d$), then P assigns to $(x, x+y]$ the probability

$$P\big((x, x+y]\big) \;=\; \sum_{(\theta_1, \ldots, \theta_d) \in \{0,1\}^d} (-1)^{\sum_{i=1}^d (1-\theta_i)} F\big(x_1 + \theta_1 y_1, \ldots, x_d + \theta_d y_d\big). \tag{1.5.2}$$

To see this, consider the sets $A_i = (-\infty, x_1 + y_1] \times \cdots \times (-\infty, x_{i-1} + y_{i-1}] \times (-\infty, x_i] \times (-\infty, x_{i+1} + y_{i+1}] \times \cdots \times (-\infty, x_d + y_d]$. We obtain by the inclusion-exclusion principle (Proposition 1.3.3) that

$$\begin{aligned} P\big(A_1 \cup \cdots \cup A_d\big) \;&=\; \sum_{k=1}^d (-1)^{k+1} \sum_{1 \le i_1 < \ldots < i_k \le d} P\big(A_{i_1} \cap \cdots \cap A_{i_k}\big) \\ &=\; \sum_{\theta \in \{0,1\}^d,\, \theta \ne (1,\ldots,1)} (-1)^{\sum_{i=1}^d (1-\theta_i)\,+\,1} F\big(x_1 + \theta_1 y_1, \ldots, x_d + \theta_d y_d\big). \end{aligned}$$

Then (1.5.2) follows from

$$P\big((x, x+y]\big) \;=\; P\big((-\infty, x_1 + y_1] \times \cdots \times (-\infty, x_d + y_d]\big) \;-\; P\big(A_1 \cup \cdots \cup A_d\big).$$

The following lemma collects important properties of the cumulative distribution function of a probability measure on $(\mathbb{R}^d, \mathcal{B}^d)$.

Lemma 1.5.1. *Let P be a probability measure on $(\mathbb{R}^d, \mathcal{B}^d)$, and let F be the corresponding cumulative distribution function. Then*

(i) *F is right-continuous in each of its variables, i.e.*

$$\lim_{x \searrow x_i} F\big(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_d\big) \;=\; F\big(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_d\big) \quad \forall (x_1, \ldots, x_d) \in \mathbb{R}^d,$$

(ii) *for all $(x_1, \ldots, x_d) \in \mathbb{R}^d$, $y_1, \ldots, y_d \geq 0$,*

$$\sum_{(\theta_1, \ldots, \theta_d) \in \{0,1\}^d} (-1)^{\sum_{i=1}^d (1-\theta_i)} F\big(x_1 + \theta_1 y_1, \ldots, x_d + \theta_d y_d\big) \;\geq\; 0,$$

(iii) $\lim_{n\to\infty} \sup \big\{F(x_1, \ldots, x_{i-1}, -n, x_{i+1} \ldots, x_d) : \; x_1, \ldots, x_{i-1}, x_{i+1} \ldots, x_d \in \mathbb{R}\big\} \;=\; 0,$ *and* $\lim_{n\to\infty} F(n, \ldots, n) = 1.$

Proof.

(i) Let $(x_1, \ldots, x_n) \in \mathbb{R}^d$ be arbitrary. If $(y_n)_{n\in\mathbb{N}}$ is any sequence such that $y_n \searrow x_i$, and $A_n = (-\infty, x_1] \times \cdots \times (-\infty, x_{i-1}] \times (-\infty, y_n] \times (-\infty, x_{i+1}] \times \cdots \times (-\infty, x_d]$, then $A_n \supseteq A_{n+1}$ for all n and $\bigcap_{n=1}^\infty A_n = A := (-\infty, x_1] \times \cdots \times (-\infty, x_d]$. Since $P(A_n) < \infty$ for all n it follows from continuity from above that

$$\begin{aligned} F\big(x_1, \ldots, x_{i-1}, y_n, x_{i+1}, \ldots, x_d\big) \;&=\; P\big(A_n\big) \underset{n\to\infty}{\longrightarrow} P\big(A\big) \\ &=\; F\big(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_d\big). \end{aligned}$$

(ii) Follows from (1.5.2).

(iii) We obtain from continuity from above that

$$\begin{aligned} & F\big(x_1, \ldots, x_{i-1}, -n, x_{i+1} \ldots, x_d\big) \\ &\quad=\; P\big((-\infty, x_1] \times \cdots \times (-\infty, x_{i-1}] \times (-\infty, -n] \times (-\infty, x_{i+1}] \times \cdots \times (-\infty, x_d]\big) \\ &\quad\leq\; P\big((-\infty, \infty) \times \cdots \times (-\infty, \infty) \times (-\infty, -n] \times (-\infty, \infty) \times \cdots \times (-\infty, \infty)\big) \\ &\quad\underset{n\to\infty}{\longrightarrow}\; P\big(\mathbb{R}^{i-1} \times \emptyset \times \mathbb{R}^{d-i}\big) \;=\; P\big(\emptyset\big) \;=\; 0. \end{aligned}$$

From continuity from below we get

$$\begin{aligned} F\big(n, \ldots, n\big) \;&=\; P\big((-\infty, n] \times \cdots \times (-\infty, n]\big) \\ &\underset{n\to\infty}{\longrightarrow}\; P\big(\mathbb{R}^d\big) \;=\; 1. \end{aligned}$$

□

A remarkable fact is that the converse of Lemma 1.5.1 is also true, i.e. its conclusion and hypothesis can be switched.

Theorem 1.5.2. *Suppose that a function $F\colon \mathbb{R}^d \to [0,1]$ satisfies (i) to (iii) in Lemma 1.5.1. Then there exists a unique probability measure P on $(\mathbb{R}^d, \mathcal{B}^d)$ such that*

$$P\big((-\infty, x_1] \times \cdots \times (-\infty, x_d]\big) \;=\; F\big(x_1, \ldots, x_d\big) \qquad \forall (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

Proof. We proceed as in the case of the specification of Lebesgue measure. First we assign probabilities to the collection of half-open rectangles. If $(x_1, \ldots, x_d) \in \mathbb{R}^d$, $y_1, \ldots, y_d \geq 0$, then we set in line with (1.5.2)

$$P_0\big((x, x+y]\big) \;:=\; \sum_{(\theta_1,\ldots,\theta_d)\in\{0,1\}^d} (-1)^{\sum_{i=1}^d (1-\theta_i)} F\big(x_1 + \theta_1 y_1, \ldots, x_d + \theta_d y_d\big).$$

Here it becomes clear that condition (ii) is relevant since otherwise the non-negativity would be violated. This definition can be readily extended to sets belonging to $\mathcal{B}_0^d$. If $A \in \mathcal{B}_0^d$, then there exist pairwise disjoint rectangles $I_1, \ldots, I_k$ such that $A = I_1 \cup \cdots \cup I_k$, and we set

$$P_0\big(A\big) \;:=\; \sum_{i=1}^{k} P_0\big(I_i\big).$$

It can be shown analogously to (1.3.2) that the value assigned to A is independent of the chosen representation of this set; hence this definition is "consistent". We can check in literally the same way as in the proof of Lemma 1.3.4 that P_0 is a pre-measure on $\mathcal{B}_0^d$. For this, right-continuity of F will prove to be important. By Carathéodory's theorem (Theorem 1.3.6), this pre-measure can be extended to a measure P on the σ-algebra $\sigma(\mathcal{B}_0^d) = \mathcal{B}^d$. Moreover, since $\mathcal{B}_0^d$ is intersection-stable, it follows from the uniqueness theorem (Theorem 1.3.8) that this extension is unique. Furthermore, it follows from (iii) that

$$\begin{aligned}
P\big(\mathbb{R}^d\big) \;&=\; \lim_{n\to\infty} P\big((-n,n] \times \cdots \times (-n,n]\big) \\
&=\; \underbrace{\lim_{n\to\infty} F\big(n, \ldots, n\big)}_{=1} \\
&\quad + \sum_{(\theta_1,\ldots,\theta_d)\in\{0,1\}^d,\, \theta\neq(1,\ldots,1)} (-1)^{\sum_{i=1}^d (1-\theta_i)} \underbrace{\lim_{n\to\infty} F\big(-n + 2\theta_1 n, \ldots, -n + 2\theta_d n\big)}_{=0} \\
&=\; 1.
\end{aligned}$$

Hence, P is a probability measure on $(\mathbb{R}^d, \mathcal{B}^d)$. Finally,

$$\begin{aligned}
&P\big((-\infty, x_1] \times \cdots \times (-\infty, x_d]\big) \\
&\;=\; \lim_{n\to\infty} P\big((-n, x_1] \times \cdots \times (-n, x_d]\big) \\
&\;=\; \lim_{n\to\infty} \sum_{(\theta_1,\ldots,\theta_d)\in\{0,1\}^d} (-1)^{\sum_{i=1}^d (1-\theta_i)} F\big(-n + \theta_1(n+x_1), \ldots, -n + \theta_d(n+x_d)\big) \\
&\;=\; F\big(x_1, \ldots, x_d\big) \\
&\quad + \sum_{(\theta_1,\ldots,\theta_d)\in\{0,1\}^d,\, \theta\neq(1,\ldots,1)} (-1)^{\sum_{i=1}^d (1-\theta_i)} \underbrace{\lim_{n\to\infty} F\big(-n + \theta_1(n+x_1), \ldots, -n + \theta_d(n+x_d)\big)}_{=0},
\end{aligned}$$

as required. □

2 The Lebesgue integral

We have seen in the introductory part of this course (Chapter 0) that the concept of the Riemann integral has several shortcomings:

- Even quite elementary functions such as the indicator function $\mathbb{1}_{\mathbb{Q}}$ of the set $\mathbb{Q}$ of rational numbers are not integrable in the Riemannian sense.
- For a sequence $(f_n)_{n\in\mathbb{N}}$ of functions that are Riemann integrable on the interval $[a,b]$, it is not guaranteed that $f_n(x) \underset{n\to\infty}{\longrightarrow} f(x)$ for all $x \in [a,b]$ implies that $\int_a^b f_n(x)\,dx \underset{n\to\infty}{\longrightarrow} \int_a^b f(x)\,dx$. Even worse, the existence of the latter integral is not guaranteed.
- The expected value of a real-valued random variable X with a sufficiently regular density p can be expressed by an (improper) Riemann integral, i.e. $EX = \int_{-\infty}^{\infty} x\,p(x)\,dx$. This is e.g. the case for a normally distributed random variable $X \sim N(\mu,\sigma^2)$, where $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/\sigma^2}$ and $EX = \int_{-\infty}^{\infty} x\,p(x)\,dx = \mu$. On the other hand, the expected value of a random variable Y with a binomial distribution, i.e. $P(Y=k) = \binom{n}{k} p^k(1-p)^{n-k}$ for $k = 0,1,\ldots,n$, cannot be expressed by a Riemann integral.

In this section we introduce a more general concept of an integral, the so-called Lebesgue integral. We define, on a measure space $(\Omega,\mathcal{A},\mu)$, an integral $\int_A f\,d\mu$, where $A \in \mathcal{A}$ is a measurable set and $f\colon\,\Omega \to \mathbb{R}$ is a sufficiently regular function. It will be shown that the value of this integral coincides with that of the Riemann integral $\int_a^b f(x)\,dx$ if $A = [a,b]$, if the function f is integrable over $[a,b]$ in the Riemannian sense and if the integrator μ is Lebesgue measure. In this sense, the notion of the Lebesgue integral is an extension but not a re-definition of the Riemann integral.

It will be shown that the shortcomings of the Riemann integral mentioned above are largely healed by this new concept. It will turn out that practically all non-negative functions are Lebesgue integrable, which leads to a hassle-free work with this concept. Moreover, we will see that there exist simple sufficient conditions such that the pointwise convergence of a sequence $(f_n)_{n\in\mathbb{N}}$ of functions implies the convergence of the corresponding integrals. And finally, the concept of the Lebesgue integral allows for a unique representation of expected values of random variables, no matter whether they posses a density or not.

2.1 Definition of the Lebesgue integral of "simple functions"

We begin with the definition of the Lebesgue integral. Let $(\Omega,\mathcal{A},\mu)$ be an arbitrary measure space. If $A \in \mathcal{A}$ and $\mathbb{1}_A$ is the indicator function of this set, then

$$\int_\Omega \mathbb{1}_A\,d\mu \;:=\; \mu(A)$$

is the integral of $\mathbb{1}_A$ with respect to μ. This definition can be extended to so-called $\mathcal{A}$-simple functions. A function $s\colon\,\Omega \to [0,\infty)$ is called to be an **$\mathcal{A}$-simple function** if there exist non-negative real numbers $\alpha_1,\ldots,\alpha_k$ and disjoint sets $A_1,\ldots,A_k \in \mathcal{A}$ such that

$$s \;=\; \sum_{i=1}^{k} \alpha_i \mathbb{1}_{A_i}.$$

In this case, we define the integral of s with respect to μ to be

$$\int_\Omega s\,d\mu := \sum_{i=1}^k \alpha_i \mu\big(A_i\big).$$

Before we proceed we need to check that the value of this integral depends only on s and not on the particular choice of $\alpha_1,\dots,\alpha_k$ and $A_1,\dots,A_k \in \mathcal{A}$. Suppose that there exist $\alpha_1,\dots,\alpha_k \geq 0$, $\beta_1,\dots,\beta_l \geq 0$, disjoint sets $A_1,\dots,A_k \in \mathcal{A}$, and disjoint sets $B_1,\dots,B_l \in \mathcal{A}$ such that

$$\sum_{i=1}^k \alpha_i \mathbb{1}_{A_i}(\omega) = \sum_{j=1}^l \beta_j \mathbb{1}_{B_j}(\omega) \qquad \forall \omega \in \Omega.$$

Let $\alpha_0 = \beta_0 = 0$ and $A_0 = \Omega \setminus \big(A_1 \cup \dots \cup A_k\big)$, $B_0 = \Omega \setminus \big(B_1 \cup \dots \cup B_l\big)$. Then $A_i = \bigcup_{j=0}^l A_i \cap B_j$, $B_j = \bigcup_{i=0}^k B_j \cap A_i$, and it follows from additivity of μ and the fact that $\alpha_i = \beta_j$ if $A_i \cap B_j \neq \emptyset$ that

$$\begin{aligned}\sum_{i=1}^k \alpha_i \mu\big(A_i\big) &= \sum_{i=0}^k \alpha_i \mu\big(A_i\big) = \sum_{i=0}^k \sum_{j=0}^l \alpha_i \mu\big(A_i \cap B_j\big) \\ &= \sum_{i=0}^k \sum_{j=0}^l \beta_j \mu\big(A_i \cap B_j\big) = \sum_{j=0}^l \beta_j \mu\big(B_j\big) = \sum_{j=1}^l \beta_j \mu\big(B_j\big).\end{aligned}$$

Hence, $\int_\Omega s\,d\mu$ does not depend on the chosen representation of s.

At this point we see the crucial advantage of the definition of the Lebesgue integral. Let $a, b \in \mathbb{R}$, $a < b$ and $s = \mathbb{1}_{\mathbb{Q}\cap[a,b]}$. Then we obtain that the lower Riemann integral $\underline{\int}_a^b s(x)\,dx$ is equal to zero while the upper Riemann integral $\overline{\int}_a^b s(x)\,dx$ is equal to $b - a$. Therefore, the function s is not integrable in the Riemannian sense. On the other hand, s is a $\mathcal{B}$-simple function and its Lebesgue integral with respect to Lebesgue measure $\int_\mathbb{R} s\,d\lambda$ exists and is equal to $\lambda\big(\mathbb{Q} \cap [a,b]\big) = 0$. This is indeed what we could expect since the interval $[a,b]$ is "dominated" by irrational numbers, $\lambda(\mathbb{Q}^c \cap [a,b]) = b-a > 0 = \lambda(\mathbb{Q}\cap[a,b])$. The essential difference between the Riemann and the Lebesgue integral is that the former is based on partitions $a_0,\dots,a_n$ of the domain of integration $[a,b]$ whereas the latter is based on a decomposition of this interval (here $Q \cap [a,b], Q^c \cap [a,b]$) which is **tailor-made** for the function to be integrated.

Before we proceed to the next stage of our construction, we verify a few properties of integrals of simple functions.

Proposition 2.1.1. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let f and g be $\mathcal{A}$-simple functions. Then*

(i) if $c \geq 0$, then $\int_\Omega cf\, d\mu \;=\; c\, \int_\Omega f\, d\mu$,

(ii) $\int_\Omega (f+g)\, d\mu \;=\; \int_\Omega f\, d\mu \,+\, \int_\Omega g\, d\mu$,

(iii) if $f(\omega) \leq g(\omega)$ holds for all $\omega \in \Omega$, then $\int_\Omega f\, d\mu \;\leq\; \int_\Omega g\, d\mu$.

Proof. Suppose that $f = \sum_{i=1}^k \alpha_i \mathbb{1}_{A_i}$, where $\alpha_1, \ldots, \alpha_k \geq 0$ and $A_1, \ldots, A_k$ are disjoint sets that belong to $\mathcal{A}$, and that $g = \sum_{j=1}^l \beta_j \mathbb{1}_{B_j}$, where $\beta_1, \ldots, \beta_l \geq 0$ and $B_1, \ldots, B_l$ are disjoint sets that belong to $\mathcal{A}$.

(i) The function cf has a representation as $cf(\omega) \;=\; \sum_{i=1}^k (c\alpha_i)\mathbb{1}_{A_i}(\omega)$, which leads to

$$\int_\Omega (cf)\, d\mu \;=\; \sum_{i=1}^k (c\alpha_i)\mu\big(A_i\big) \;=\; c\cdot \sum_{i=1}^k \alpha_i \mu\big(A_i\big) \;=\; c\, \int_\Omega f\, d\mu.$$

(ii) Suppose in addition that the respective representations of f and g are such that $\bigcup_{i=1}^k A_i = \bigcup_{j=1}^l B_j$. (Otherwise we add α_0, $A_0 = \Omega \setminus \big(A_1 \cup \cdots \cup A_k\big)$ and/or $\beta_0 = 0$, $B_0 = \Omega \setminus \big(B_1 \cup \cdots \cup B_l\big)$ to these representations.) Then $\mathbb{1}_{A_i}(\omega) = \sum_j \mathbb{1}_{A_i \cap B_j}(\omega)$ and $\mathbb{1}_{B_j}(\omega) = \sum_i \mathbb{1}_{A_i \cap B_j}(\omega)$, and hence

$$\begin{aligned} f(\omega) \,+\, g(\omega) &= \sum_i \alpha_i \mathbb{1}_{A_i}(\omega) \,+\, \sum_j \beta_j \mathbb{1}_{B_j}(\omega) \\ &= \sum_{i,j} \alpha_i \mathbb{1}_{A_i \cap B_j}(\omega) \,+\, \sum_{i,j} \beta_j \mathbb{1}_{A_i \cap B_j}(\omega) \\ &= \sum_{i,j} \big(\alpha_i \,+\, \beta_j\big) \mathbb{1}_{A_i \cap B_j}(\omega) \end{aligned}$$

holds for all $\omega \in \Omega$. Since $A_i = \bigcup_j A_i \cap B_j$ and $B_j = \bigcup_i A_i \cap B_j$ we obtain

$$\begin{aligned} \int_\Omega \big(f \,+\, g\big)\, d\mu &= \sum_{i,j} \big(\alpha_i \,+\, \beta_j\big)\mu\big(A_i \cap B_j\big) \\ &= \sum_{i,j} \alpha_i \mu\big(A_i \cap B_j\big) \,+\, \sum_{i,j} \beta_j \mu\big(A_i \cap B_j\big) \\ &= \sum_i \alpha_i \mu\big(A_i\big) \,+\, \sum_j \beta_j \mu\big(B_j\big) \;=\; \int_\Omega f\, d\mu \,+\, \int_\Omega g\, d\mu. \end{aligned}$$

(iii) Suppose that $f(\omega) \leq g(\omega)$ holds for all $\omega \in \Omega$. Since $g - f$ is also an $\mathcal{A}$-simple function we obtain from (ii) that

$$\begin{aligned} \int_\Omega g\, d\mu &= \int_\Omega \big[f \,+\, (g-f)\big]\, d\mu \\ &= \int_\Omega f\, d\mu \,+\, \underbrace{\int_\Omega (g-f)\, d\mu}_{\geq 0} \;\geq\; \int_\Omega f\, d\mu. \end{aligned}$$

(Note that "$\infty \geq \infty$" is not excluded here.)

□

2.2 Measurable functions

In what follows we extend the definition of the Lebesgue integral, which is so far restricted to simple functions $s\colon \Omega \to [0,\infty)$, to a broader class of functions. If $(\Omega,\mathcal{A},\mu)$ is a measure space and $f\colon \Omega \to [0,\infty)$ is an arbitrary non-negative function, then we could approximate this function from below by $\mathcal{A}$-simple functions, and define the integral of f with respect to μ as

$$\int_\Omega f\,d\mu \;=\; \sup\Big\{\int_\Omega s\,d\mu\colon\; s \text{ is an } \mathcal{A}\text{-simple function, } s \le f\Big\}.$$

Such a definition coincides with our previous definition of the integral of a simple function, and for $f = \mathbb{1}_{\mathbb{Q}}$ we obtain the heuristically expected result $\int_\Omega f\,d\lambda = 0$. On the other hand, it is of course desirable that the concept of the Lebesgue integral has properties such a additivity and translation-invariance. Recall that Corollary 1.4.2 states that there exists a subset E of the interval $(0,1)$ that is not Lebesgue measurable. Furthermore, if $(r_n)_{n\in\mathbb{N}}$ is an enumeration of the rational numbers that are contained in $(-1,1)$, then $(0,1) \subseteq \bigcup_{n=1}^\infty (E + r_n) \subseteq (-1,2)$. If the Lebesgue integral has indeed the property of being translation-invariant, then we obtain that $\int_\mathbb{R} \mathbb{1}_{E+r_n}\,d\lambda = \int_\mathbb{R} \mathbb{1}_E\,d\lambda$ for all $n \in \mathbb{N}$. On the other hand, the function $f := \sum_{n=1}^\infty \mathbb{1}_{E+r_n}$ has the properties that $f(x) = 1$ for all $x \in (0,1)$, $f(x) = 0$ for all $x \in (-\infty,-1] \cup [2,\infty)$, and $f(x) \in \{0,1\}$ for all $x \in (-1,0] \cup [1,2)$. Therefore, the value of the integral $\int_\mathbb{R} f\,d\lambda$ should lie between 1 and 3, which is in contradiction to the desirable property that $\int_\mathbb{R} f\,d\lambda = \sum_{n=1}^\infty \int_\mathbb{R} \mathbb{1}_{E+r_n}\,d\lambda$. In view of this, the collection of functions which allow a meaningful definition of integrals has to be restricted to sufficiently regular ones. The notion of measurable functions will satisfy this requirement. We begin with a definition of this property in a general context.

Definition. Let $(\Omega,\mathcal{A})$ and $(\Omega',\mathcal{A}')$ be measurable spaces. A mapping $f\colon \Omega \to \Omega'$ is said to be **$(\mathcal{A}-\mathcal{A}')$-measurable** if

$$f^{-1}(A') \;:=\; \big\{\omega \in \Omega\colon\;\; f(\omega) \in A'\big\} \;\in\; \mathcal{A} \qquad \forall A' \in \mathcal{A}'.$$

Note that the inverse function of f, which is usually also denoted by f^{-1}, need not exist. Actually, $f^{-1}(B)$ denotes the so-called **inverse image** of a generic set $B \subseteq \Omega'$. For a real-valued function f, the image space Ω' is the real line $\mathbb{R}$, and in this case the collection of Borel sets $\mathcal{B}$ is always tacitly understood to play the role of $\mathcal{A}'$. An $(\mathcal{A}-\mathcal{B})$-measurable real-valued function f is simply called $\mathcal{A}$-measurable. In the context of probability theory, the corresponding measure space is a probability space $(\Omega,\mathcal{A},P)$, and a real $\mathcal{A}$-measurable function $X\colon \Omega \to \mathbb{R}$ is called a **random variable**. The point of the definition is to ensure that $\{\omega \in \Omega\colon\; X(\omega) \in B\}$ has a measure or probability $P(\{\omega \in \Omega\colon\; X(\omega) \in B\})$ for all sufficiently regular sets B, i.e. for all Borel sets $B \in \mathcal{B}$.

We turn to a few simple examples of real-valued measurable functions. Let $(\Omega,\mathcal{A})$ be a measurable space.

1) A constant function $f\colon \Omega \to \mathbb{R}$ is $\mathcal{A}$-measurable.
Indeed, if $f(\omega) = c$ for all $\omega \in \Omega$, and A' is an arbitrary subset of $\mathbb{R}$, then

$$f^{-1}(A') \;=\; \begin{cases} \Omega & \text{if } c \in A', \\ \emptyset & \text{if } c \notin A'. \end{cases}$$

2) If $A \in \mathcal{A}$, then the indicator function $\mathbb{1}_A$ is $\mathcal{A}$-measurable.
To see this, we consider again the set of all possible inverse images. If A' is an arbitrary subset of $\mathbb{R}$, then

$$\mathbb{1}_A^{-1}(A') = \begin{cases} \Omega & \text{if } 0, 1 \in A', \\ \emptyset & \text{if } 0 \notin A',\ 1 \notin A', \\ A & \text{if } 1 \in A',\ 0 \notin A', \\ A^c & \text{if } 0 \in A',\ 1 \notin A'. \end{cases}$$

Our hypothesis of $A \in \mathcal{A}$ implies that $\{\emptyset, A, A^c, \Omega\} \subseteq \mathcal{A}$, which means that $\mathbb{1}_A$ is $\mathcal{A}$-measurable.

3) An $\mathcal{A}$-simple function $s\colon \Omega \to [0, \infty)$ is $\mathcal{A}$-measurable.
Let $s = \sum_{i=1}^k \alpha_i \mathbb{1}_{A_i}$, where $\alpha_1, \dots, \alpha_k > 0$ and $A_1, \dots, A_k$ are disjoint sets that belong to $\mathcal{A}$. If $\alpha_i \neq \alpha_j$ for $i \neq j$, then the collection of inverse images is given by the $\sigma(\{A_1, \dots, A_k\})$. If $\alpha_i = \alpha_j$ for some $i \neq j$, then the collection of inverse images is even smaller, it is in fact a proper subset of $\sigma(\{A_1, \dots, A_k\})$. In both cases, $s^{-1}(A') \in \mathcal{A}$ for all $A' \subseteq \mathbb{R}$.

The above definition of an $(\mathcal{A} - \mathcal{A}')$-measurable mapping means that the inverse images $f^{-1}(A')$ must belong to $\mathcal{A}$ for all $A' \in \mathcal{A}'$. A direct verification of such a property might be quite cumbersome if not impossible in cases where the σ-algebra $\mathcal{A}'$ is rich and contains sets with quite a complicated structure. The next lemma provides some tools for checking measurability in a convenient way.

Lemma 2.2.1. *Let $(\Omega, \mathcal{A})$, $(\Omega', \mathcal{A}')$, and $(\Omega'', \mathcal{A}'')$ be measurable spaces, and let $f\colon \Omega \to \Omega'$ and $g\colon \Omega' \to \Omega''$ be mappings.*

(i) If $f^{-1}(E') \in \mathcal{A}'$ holds for all $E' \in \mathcal{E}'$, where $\mathcal{E}'$ is a collection of subsets of Ω' such that $\sigma(\mathcal{E}') = \mathcal{A}'$, then the mapping f is $(\mathcal{A} - \mathcal{A}')$-measurable.

(ii) If f is $(\mathcal{A} - \mathcal{A}')$-measurable and g is $(\mathcal{A}' - \mathcal{A}'')$-measurable, then the composition $g \circ f\colon \Omega \to \Omega''$ ($g \circ f(\omega) = g(f(\omega))\ \forall \omega \in \Omega$) is $(\mathcal{A} - \mathcal{A}'')$-measurable.

Proof.

(i) We use the good set principle and define the system of good sets,

$$\mathcal{G} := \{A' \subseteq \Omega'\colon\ f^{-1}(A') \in \mathcal{A}\}.$$

The set $\mathcal{G}$ is a σ-algebra on Ω'. Indeed, we have:

a) $f^{-1}(\Omega') = \Omega \in \mathcal{A}$, hence $\Omega' \in \mathcal{G}$.

b) If $A' \in \mathcal{G}$, then $f^{-1}(A') \in \mathcal{A}$, and so $f^{-1}(A'^c) = \left(f^{-1}(A')\right)^c \in \mathcal{A}$, which means that $A'^c \in \mathcal{G}$.

c) If $A'_1, A'_2, \dots \in \mathcal{G}$, then $f^{-1}(A'_1), f^{-1}(A'_2), \dots \in \mathcal{A}$, and hence $f^{-1}\left(\bigcup_{i=1}^\infty A'_i\right) = \bigcup_{i=1}^\infty f^{-1}(A'_i) \in \mathcal{A}$. This implies that $\bigcup_{i=1}^\infty A'_i \in \mathcal{G}$.

Since by assumption $\mathcal{E}' \subseteq \mathcal{G}$ we therefore obtain that

$$\mathcal{A}' = \sigma(\mathcal{E}') \subseteq \sigma(\mathcal{G}) = \mathcal{G}.$$

Hence, the mapping $f\colon \Omega \to \Omega'$ is $(\mathcal{A} - \mathcal{A}')$-measurable.

(ii) Let $A'' \in \mathcal{A}''$ be arbitrary. Since g is $(\mathcal{A}' - \mathcal{A}'')$-measurable we have that $g^{-1}(A'') \in \mathcal{A}'$, and since f is $(\mathcal{A} - \mathcal{A}')$-measurable we obtain $(g \circ f)^{-1}(A'') = f^{-1}\big(g^{-1}(A'')\big) \in \mathcal{A}$. Hence, $g \circ f$ is $(\mathcal{A} - \mathcal{A}'')$-measurable.

□

Lemma 2.2.1 allows us to prove almost effortlessly the useful result that that all continuous functions $f\colon \mathbb{R}^d \to \mathbb{R}^{d'}$ are $(\mathcal{B}^d - \mathcal{B}^{d'})$-measurable.

Corollary 2.2.2. *If $f\colon \mathbb{R}^d \to \mathbb{R}^{d'}$ is a continuous function, then it is $(\mathcal{B}^d - \mathcal{B}^{d'})$-measurable.*

Proof. Let O' be an arbitrary open subset of $\mathbb{R}^{d'}$. Then $f^{-1}(O')$ is an open subset of $\mathbb{R}^d$. Hence, all inverse images of open sets belong to $\mathcal{B}^d$. Since the open sets generate $\mathcal{B}^{d'}$ it follows from part (i) of Lemma 2.2.1 that f is $(\mathcal{B}^d - \mathcal{B}^{d'})$-measurable. □

For a mapping $f\colon \Omega \to \mathbb{R}^d$ carrying Ω into $\mathbb{R}^d$, f must have the form

$$f(\omega) = \big(f_1(\omega), \ldots, f_d(\omega)\big),$$

where $f_1, \ldots, f_d$ are real-valued functions. In probabilistic contexts, a measurable mapping into $\mathbb{R}^d$ is called a random vector. Using once more Lemma 2.2.1 we see that measurability of f follows from that of the component functions.

Corollary 2.2.3. *Let $(\Omega, \mathcal{A})$ be a measurable space, and let $f_i\colon \Omega \to \mathbb{R}$ $(i = 1, \ldots, d)$, $f(\omega) = \big(f_1(\omega), \ldots, f_d(\omega)\big)$ $\forall \omega \in \Omega$.*

Then $f\colon \Omega \to \mathbb{R}^d$ is $(\mathcal{A} - \mathcal{B}^d)$-measurable if and only if each component function $f_i\colon \Omega \to \mathbb{R}$ is $(\mathcal{A} - \mathcal{B})$-measurable.

Proof.

($\Longrightarrow$) Let f be $(\mathcal{A} - \mathcal{B}^d)$-measurable and let $B \in \mathcal{B}$ be arbitrary. Then

$$f_i^{-1}\big(B\big) = f^{-1}\big(\mathbb{R}^{i-1} \times B \times \mathbb{R}^{d-i}\big).$$

Since $\mathbb{R}^{i-1} \times B \times \mathbb{R}^{d-i} \in \mathcal{B}^d$ we obtain from $(\mathcal{A} - \mathcal{B}^d)$-measurability of f that $f_i^{-1}\big(B\big) \in \mathcal{A}$, i.e. f_i is $(\mathcal{A} - \mathcal{B})$-measurable.

($\Longleftarrow$) Let $f_1, \ldots, f_d$ be $(\mathcal{A}-\mathcal{B})$-measurable. Then, for $a_i, b_i \in \mathbb{R}$, $a_i \leq b_i$, $f_i^{-1}\big((a_i, b_i]\big) \in \mathcal{A}$. Therefore,

$$f^{-1}\big((a_1, b_1] \times \cdots \times (a_d, b_d]\big) = \bigcap_{i=1}^{d} f_i^{-1}\big((a_i, b_i]\big) \in \mathcal{A}.$$

Since the half-open rectangles generate $\mathcal{B}^d$ we conclude by Lemma 2.2.1 that f is $(\mathcal{A} - \mathcal{B}^d)$-measurable.

□

If $f_1, \dots, f_d \colon \Omega \to \mathbb{R}$ are $(\mathcal{A}-\mathcal{B})$-measurable, then we obtain from Corollaries 2.2.2 and 2.2.3 that $\omega \mapsto \sum_{i=1}^d f_i(\omega)$, $\omega \mapsto \prod_{i=1}^d f_i(\omega)$, and $\omega \mapsto \max\{f_1(\omega), \dots, f_d(\omega)\}$ are also $(\mathcal{A}-\mathcal{B})$-measurable. Indeed, $\omega \mapsto f(\omega) = (f_1(\omega), \dots, f_d(\omega))$ is by Corollary 2.2.3 $(\mathcal{A}-\mathcal{B}^d)$-measurable. Since $(x_1, \dots, x_d) \mapsto \sum_{i=1}^d x_i$, $(x_1, \dots, x_d) \mapsto \prod_{i=1}^d x_i$, and $(x_1, \dots, x_d) \mapsto \max\{x_1, \dots, x_d\}$ are continuous and so by Corollary 2.2.2 $(\mathcal{B}^d-\mathcal{B})$-measurable we obtain from Lemma 2.2.1 that the above functions are $(\mathcal{A}-\mathcal{B})$-measurable.

In the following we will also consider suprema, infima, and limits of measurable functions. Then, but also in other contexts, it will be convenient to admit the artificial values ∞ and $-\infty$, i.e. we allow functions to take values in the **extended real line** $\bar{\mathbb{R}} := \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$. A function $f\colon \Omega \to \bar{\mathbb{R}}$ is called **extended real-valued function**. An appropriate σ-algebra on $\bar{\mathbb{R}}$ is given by $\bar{\mathcal{B}} := \sigma\big(\mathcal{B} \cup \{\infty\} \cup \{-\infty\}\big)$. If $(\Omega, \mathcal{A})$ is a measurable space, then a function $f\colon \Omega \to \bar{\mathbb{R}}$ is called $\mathcal{A}$-measurable if $f^{-1}(B) \in \mathcal{A}$ holds for all $B \in \bar{\mathcal{B}}$. For this condition to hold, it suffices that $f^{-1}(B) \in \mathcal{A} \quad \forall B \in \mathcal{B}$, $f^{-1}(\{\infty\}) \in \mathcal{A}$, and $f^{-1}(\{-\infty\}) \in \mathcal{A}$.

Proposition 2.2.4. *Let $(\Omega, \mathcal{A})$ be a measurable space and let $(f_n)_{n\in\mathbb{N}}$ be a sequence of extended real-valued $\mathcal{A}$-measurable functions on Ω. Then*

(i) *The functions $\sup_n f_n$ and $\inf_n f_n$ are $\mathcal{A}$-measurable.*

(ii) *The functions $\limsup_n f_n$ and $\liminf_n f_n$ are $\mathcal{A}$-measurable.*

(iii) *If $\limsup_n f_n(\omega) = \liminf_n f_n(\omega) \ \ \forall \omega \in \Omega$, then $\lim_n f_n(\omega)$ exists for all $\omega \in \Omega$ and $\lim_n f_n$ is an extended real-valued $\mathcal{A}$-measurable function.*

Proof.

(i) First of all, it is not difficult to see that $\big\{[-\infty, x]\colon\ x \in \mathbb{R}\big\}$ generates $\bar{\mathcal{B}}$. Indeed, we have that $\{-\infty\} = \bigcap_{n\in\mathbb{N}}[-\infty, -n]$, $\{\infty\} = \bar{\mathbb{R}} \setminus \bigcup_{n\in\mathbb{N}}[-\infty, n]$, and the sets $(-\infty, x] = [-\infty, x] \setminus \{-\infty\}$ generate $\bar{\mathcal{B}}$. Therefore, it follows from the identity

$$\big\{\omega \in \Omega\colon\ \sup_n f_n(\omega) \le x\big\} = \bigcap_{n=1}^{\infty} \big\{\omega \in \Omega\colon\ f_n(\omega) \le x\big\}$$

by (i) of Lemma 2.2.1 that $\sup_n f_n$ is $\mathcal{A}$-measurable.

To prove measurability of $\inf_n f_n$, note that $\big\{[x, \infty]\colon\ x \in \mathbb{R}\big\}$ also generates $\bar{\mathcal{B}}$. Since

$$\big\{\omega \in \Omega\colon\ \inf_n f_n(\omega) \ge x\big\} = \bigcap_{n=1}^{\infty} \big\{\omega \in \Omega\colon\ f_n(\omega) \ge x\big\}$$

we obtain by (i) of Lemma 2.2.1 that $\inf_n f_n$ is $\mathcal{A}$-measurable.

(ii) Note that

$$\limsup_n f_n = \inf_k \sup_{n\ge k} f_n \qquad \text{and} \qquad \liminf_n f_n = \sup_k \inf_{n\ge k} f_n.$$

Part (i) of this proposition implies first that $\sup_{n\ge k} f_n$ and $\inf_{n\ge k} f_n$ are $\mathcal{A}$-measurable, and then that $\inf_k \sup_{n\ge k} f_n$ and $\sup_k \inf_{n\ge k} f_n$ are $\mathcal{A}$-measurable.

(iii) This follows directly from (ii).

□

The following result will be repeatedly used in connection with Lebesgue integrals of $\mathcal{A}$-measurable functions.

Proposition 2.2.5. *Let* $(\Omega, \mathcal{A})$ *be a measurable space, and let* $f\colon\ \Omega \to [0, \infty]$ *be a non-negative* $(\mathcal{A} - \bar{\mathcal{B}})$*-measurable function. Then there exists a sequence* $(s_n)_{n\in\mathbb{N}}$ *of* $\mathcal{A}$*-simple functions that satisfy*

(i) $\quad s_n(\omega) \le s_{n+1}(\omega) \le f(\omega) \qquad \forall \omega \in \Omega,\ \forall n \in \mathbb{N},$

(ii) $\quad s_n(\omega) \underset{n\to\infty}{\longrightarrow} f(\omega) \qquad \forall \omega \in \Omega.$

Proof. For $n \in \mathbb{N}$, we define

$$\begin{aligned} A_{n,i} &= f^{-1}\big([(i-1)/2^n, i/2^n)\big), \qquad \text{for } i = 1, 2, \ldots, n2^n, \\ A_n &= f^{-1}\big([n, \infty]\big). \end{aligned}$$

The measurability of f implies that these sets belong to $\mathcal{A}$. For each $n \in \mathbb{N}$ we define the simple functions

$$s_n = \sum_{i=1}^{n2^n} (i-1)/2^n\, \mathbb{1}_{A_{n,i}} \quad + \quad n\, \mathbb{1}_{A_n}.$$

With this choice, we have obviously that

$$s_n(\omega) \nearrow f(\omega) \qquad \forall \omega \in \Omega.$$

□

2.3 The Lebesgue integral of measurable functions

We are now in a position to define the Lebesgue integral of arbitrary measurable functions. We begin with **non-negative** functions.

Definition. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let $f\colon\ \Omega \to [0, \infty]$ be an $(\mathcal{A} - \bar{\mathcal{B}})$-measurable function. Then

$$\int_\Omega f\, d\mu := \sup\Big\{ \int_\Omega s\, d\mu\colon \quad s\ \ \mathcal{A} - \text{simple function},\ s \le f \Big\}$$

denotes the **Lebesgue integral** of f with respect to μ.

It is easy to see that for $\mathcal{A}$-simple functions this agrees with the previous definition. The next proposition shows that the above integral can alternatively be defined as the limit of the integrals of an **arbitrary** non-decreasing sequence of $\mathcal{A}$-simple functions that converge to f. This alternative representation will allow us to prove a few properties of the integral collected in Proposition 2.3.2 below.

Proposition 2.3.1. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, let $f\colon\ \Omega \to [0,\infty]$ be an $(\mathcal{A} - \bar{\mathcal{B}})$-measurable function, and let $(s_n)_{n\in\mathbb{N}}$ be a sequence of $\mathcal{A}$-simple functions such that*

$$s_n(\omega) \nearrow f(\omega) \qquad \forall \omega \in \Omega.$$

Then

$$\int_\Omega f\, d\mu \;=\; \lim_{n\to\infty} \int_\Omega s_n\, d\mu.$$

Proof. According to the definition of the integral, there exists a sequence $(r_m)_{m\in\mathbb{N}}$ of $\mathcal{A}$-simple functions such that $r_m \le f\ \forall m \in \mathbb{N}$ and

$$\int_\Omega f\, d\mu \;=\; \lim_{m\to\infty} \int_\Omega r_m\, d\mu.$$

We show that, for arbitrary **fixed** $m \in \mathbb{N}$,

$$\int_\Omega r_m\, d\mu \;\le\; \lim_{n\to\infty} \int_\Omega s_n\, d\mu, \tag{2.3.1}$$

which implies that

$$\int_\Omega f\, d\mu \;=\; \lim_{m\to\infty} \int_\Omega r_m\, d\mu \;\le\; \lim_{n\to\infty} \int_\Omega s_n\, d\mu,$$

and hence

$$\int_\Omega f\, d\mu \;=\; \lim_{n\to\infty} \int_\Omega s_n\, d\mu.$$

("$<$" is impossible since $s_n \le f\ \forall n \in \mathbb{N}$.)

The idea of the proof of (2.3.1) may be sketched as follows. Let $r_m = \sum_{i=1}^k \alpha_i \mathbb{1}_{A_i}$, where $\alpha_1, \ldots, \alpha_k \ge 0$ and $A_1, \ldots, A_k \in \mathcal{A}$ and let

$$\Omega_n \;=\; \big\{\omega \in \Omega\colon\ r_m(\omega) \le s_n(\omega)\big\}.$$

Since

$$\big\{\omega \in \Omega\colon\ r_m(\omega) > s_n(\omega)\big\} \;=\; \bigcup_{r\in\mathbb{Q}} \big\{\omega \in \Omega\colon\ r_m(\omega) > r\big\} \cap \big\{\omega \in \Omega\colon\ s_n(\omega) < r\big\} \;\in\; \mathcal{A}$$

we see that $\Omega_n = \big\{\omega \in \Omega\colon\ r_m(\omega) > s_n(\omega)\big\}^c \in \mathcal{A}$. Suppose first that $\Omega_n \nearrow \Omega$. Since the measure μ is continuous from below it follows that $\mu\big(A_i \cap \Omega_n\big) \underset{n\to\infty}{\longrightarrow} \mu\big(A_i\big)$, and we obtain that

$$\begin{aligned}\int_\Omega r_m\, d\mu \;&=\; \sum_{i=1}^k \alpha_i\, \mu\big(A_i\big) \;=\; \lim_{n\to\infty} \sum_{i=1}^k \alpha_i\, \mu\big(A_i \cap \Omega_n\big) \\ &=\; \lim_{n\to\infty} \int_\Omega \underbrace{r_m \mathbb{1}_{\Omega_n}}_{\le s_n}\, d\mu \;\le\; \lim_{n\to\infty} \int_\Omega s_n\, d\mu.\end{aligned}$$

However, it is not guaranteed that $\Omega_n \nearrow \Omega$ and we have to refine our idea of proof.

Let $\epsilon > 0$ be arbitrary and let

$$\Omega_{n,\epsilon} \;=\; \big\{\omega \in \Omega\colon\ r_m(\omega) \le (\mathbf{1} + \epsilon) s_n(\omega)\big\}.$$

Then we obtain in complete analogy to the considerations above that $\Omega_{n,\epsilon} \in \mathcal{A}$. Now we have indeed that $\Omega_{n,\epsilon} \nearrow \Omega$ and we obtain

$$\begin{aligned} \int_\Omega r_m \, d\,\mu &= \sum_{i=1}^k \alpha_i \, \mu\big(A_i\big) = \lim_{n\to\infty} \sum_{i=1}^k \alpha_i \, \mu\big(A_i \cap \Omega_{n,\epsilon}\big) \\ &= \lim_{n\to\infty} \int_\Omega \underbrace{r_m \mathbb{1}_{\Omega_{n,\epsilon}}}_{\le (1+\epsilon) s_n} \, d\mu \le (1+\epsilon) \lim_{n\to\infty} \int_\Omega s_n \, d\mu, \end{aligned}$$

which implies (2.3.1). □

The next proposition collects a few elementary properties of the integral of non-negative measurable functions.

Proposition 2.3.2. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, let $f, g \colon \Omega \to [0, \infty]$ be non-negative $(\mathcal{A} - \bar{\mathcal{B}})$-measurable functions, and let $\alpha \in [0, \infty]$. Then*

(i) $\int_\Omega \alpha \, f \, d\mu = \alpha \int_\Omega f \, d\mu$,

(ii) $\int_\Omega (f+g) \, d\mu = \int_\Omega f \, d\mu + \int_\Omega g \, d\mu$,

(iii) *if $f(\omega) \le g(\omega)$ holds for all $\omega \in \Omega$, then* $\int_\Omega f \, d\mu \le \int_\Omega g \, d\mu$,

(iv) $\mu\big(\{\omega\colon f(\omega) > 0\}\big) = 0$ *if and only if* $\int_\Omega f \, d\mu = 0$.

Proof.

(i),(ii) It follows from Proposition 2.2.5 that there exist non-decreasing sequences $(s_n)_{n\in\mathbb{N}}$ and $(t_n)_{n\in\mathbb{N}}$ of $\mathcal{A}$-simple functions such that $f = \lim_{n\to\infty} s_n$ and $g = \lim_{n\to\infty} t_n$. Then $(\alpha \, s_n)_{n\in\mathbb{N}}$ and $(s_n + t_n)_{n\in\mathbb{N}}$ are non-decreasing sequences of $\mathcal{A}$-simple functions such that $\alpha \, f = \lim_{n\to\infty} \alpha \, s_n$ and $f + g = \lim_{n\to\infty}(s_n + t_n)$. We obtain from Proposition 2.3.1 that $\int_\Omega s_n \, d\mu \underset{n\to\infty}{\longrightarrow} \int_\Omega f \, d\mu$ and $\int_\Omega t_n \, d\mu \underset{n\to\infty}{\longrightarrow} \int_\Omega g \, d\mu$ as well as $\int_\Omega \alpha \, s_n \, d\mu \underset{n\to\infty}{\longrightarrow} \int_\Omega \alpha \, f \, d\mu$ and $\int_\Omega (s_n + t_n) \, d\mu \underset{n\to\infty}{\longrightarrow} \int_\Omega (f+g) g \, d\mu$. Linearity of the integral of simple functions stated in Proposition 2.1.1 yields that

$$\int_\Omega \alpha \, f \, d\mu = \lim_{n\to\infty} \int_\Omega \alpha \, s_n \, d\mu = \lim_{n\to\infty} \alpha \int_\Omega s_n \, d\mu = \alpha \int_\Omega f \, d\mu$$

and

$$\int_\Omega (f+g) \, d\mu = \lim_{n\to\infty} \int_\Omega (s_n + t_n) \, d\mu = \lim_{n\to\infty} \int_\Omega s_n \, d\mu + \lim_{n\to\infty} \int_\Omega t_n \, d\mu = \int_\Omega f \, d\mu + \int_\Omega g \, d\mu.$$

(iii) Note that the collection of $\mathcal{A}$-simple functions s that satisfy $s \le f$ is included in the collection of $\mathcal{A}$-simple functions s that satisfy $s \le g$. Therefore we obtain that

$$\begin{aligned} \int_\Omega f \, d\mu &= \sup\Big\{ \int_\Omega s \, d\mu\colon \quad s \ \mathcal{A} - \text{simple}, \ s \le f \Big\} \\ &\le \sup\Big\{ \int_\Omega s \, d\mu\colon \quad s \ \mathcal{A} - \text{simple}, \ s \le g \Big\} = \int_\Omega g \, d\mu. \end{aligned}$$

(iv) Let $A := \{\omega \in \Omega\colon\ f(\omega) > 0\}$.

($\Longrightarrow$) Suppose that $\mu(A) = 0$. We obtain for an arbitrary $\mathcal{A}$-simple function s satisfying $s \le f$ that $\mu\big(\{\omega\colon\ s(\omega) > 0\}\big) = 0$, which implies that $\int_\Omega s\, d\mu = 0$. Therefore we obtain that $\int_\Omega f\, d\mu = 0$.

($\Longleftarrow$) Now suppose that $\mu(A) > 0$. Let $A_n := \{\omega \in \Omega\colon\ f(\omega) \ge 1/n\}$. Since $A_n \nearrow A$ we obtain from continuity from below that $\mu(A_n) \nearrow \mu(A)$. Hence, there exists some $N \in \mathbb{N}$ such that $\mu(A_N) > 0$. Let $s = (1/N)\mathbb{1}_{A_N}$. Then $s \le f$ and we obtain that

$$\int_\Omega f\, d\mu \ \ge\ \int_\Omega s\, d\mu \ \ge\ (1/N)\,\mu(A_N) \ >\ 0.$$

□

Now we extend the notion of the Lebesgue integral to extended real-valued functions which are not necessarily non-negative. Let $(\Omega, \mathcal{A}, \mu)$ be an arbitrary measure space. For general $f\colon\ \Omega \to \bar{\mathbb{R}} = [-\infty, \infty]$, consider its **positive part**,

$$f^+(\omega) \ =\ \begin{cases} f(\omega) & \text{if } 0 \le f(\omega) \le \infty, \\ 0 & \text{if } -\infty \le f(\omega) \le 0, \end{cases}$$

and its **negative part**,

$$f^-(\omega) \ =\ \begin{cases} -f(\omega) & \text{if } -\infty \le f(\omega) \le 0, \\ 0 & \text{if } 0 \le f(\omega) \le \infty. \end{cases}$$

Then

$$f \ =\ f^+ \ -\ f^-,$$

where f^+ and f^- are **both non-negative** functions. If f is $(\mathcal{A} - \bar{\mathcal{B}})$-measurable, then it follows from part (i) of Proposition 2.2.4 that f^+ and f^- are also $(\mathcal{A} - \bar{\mathcal{B}})$-measurable.

Definition. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let $f\colon\ \Omega \to \bar{\mathbb{R}}$ be an $(\mathcal{A}\!-\!\bar{\mathcal{B}})$-measurable function.

(i) If at least one of $\int_\Omega f^+\, d\mu$ and $\int_\Omega f^-\, d\mu$ is finite, then the integral of f is said to **exist** and is defined by

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

If both $\int_\Omega f^+\, d\mu$ and $\int_\Omega f^-\, d\mu$ are finite, then $\int_\Omega f\, d\mu$ is finite and f is called **integrable** (or **μ-integrable**, or **summable**).

(ii) If $A \in \mathcal{A}$, then $f\mathbb{1}_A$ is $(\mathcal{A} - \bar{\mathcal{B}})$-measurable and the integral of f over A is defined by

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

provided the integral on the right-hand side exists.

(iii) A complex-valued function $f\colon \Omega \to \mathbb{C}$ is said to be $\mathcal{A}$-measurable if both its real part $\mathrm{Re}(f)$ and its imaginary part $\mathrm{Im}(f)$ are $(\mathcal{A}-\mathcal{B})$-measurable. If $A \in \mathcal{A}$ and if both $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are μ-integrable over A, then the integral of f over A is defined by

$$\int_A f\,d\mu := \int_A \mathrm{Re}(f)\,d\mu + i\int_A \mathrm{Im}(f)\,d\mu.$$

Remark 2.3.3. *The concept of the Lebesgue integral allows for a unified definition of the expected value of a random variable, no matter if it has a discrete or a continuous distribution. Let $(\Omega, \mathcal{A}, P)$ be a probability space. $X\colon \Omega \to \bar{\mathbb{R}}$ is said to be a* **random variable** *if it is $(\mathcal{A}-\bar{\mathcal{B}})$-measurable. If at least one of $\int_\Omega X^+\,dP$ and $\int_\Omega X^-\,dP$ is finite, then the* **expected value** *of X exists and it is defined by*

$$EX := \int_\Omega X\,dP = \int_\Omega X^+\,dP - \int_\Omega X^-\,dP.$$

The following proposition provides a simple criterion for the integrability of a function f.

Proposition 2.3.4. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, and let $f\colon \Omega \to \bar{\mathbb{R}}$ be an $(\mathcal{A}-\bar{\mathcal{B}})$-measurable function.*

Then f is μ-integrable if and only $|f|$ is μ-integrable. If these functions are integrable then,

$$\Big|\int_\Omega f\,d\mu\Big| \le \int_\Omega |f|\,d\mu.$$

Proof. First of all, if f is $(\mathcal{A}-\bar{\mathcal{B}})$-measurable, then both f^+ and f^- are $(\mathcal{A}-\bar{\mathcal{B}})$-measurable which implies that $|f| = f^+ + f^-$ is also $(\mathcal{A}-\bar{\mathcal{B}})$-measurable.

Recall that by definition f is μ-integrable if and only if both f^+ and f^- are μ-integrable. On the other hand, part (ii) of Proposition 2.3.2 implies that $|f| = f^+ + f^-$ is μ-integrable if and only if both f^+ and f^- are μ-integrable. Thus the integrability of f is equivalent to the integrability of $|f|$.

In case f and $|f|$ are integrable, we obtain

$$\begin{aligned}\Big|\int_\Omega f\,d\mu\Big| &= \Big|\int_\Omega f^+\,d\mu - \int_\Omega f^-\,d\mu\Big| \\ &\le \int_\Omega f^+\,d\mu + \int_\Omega f^-\,d\mu = \int_\Omega |f|\,d\mu.\end{aligned}$$

□

Definition. Let $(\Omega, \mathcal{A}, \mu)$ be a measurable space. Then

$$L^1(\mu) := \Big\{f\colon \Omega \to \mathbb{R}\colon \quad f \text{ is } (\mathcal{A}-\mathcal{B})\text{-measurable}, \int_\Omega |f|\,d\mu < \infty\Big\}$$

denotes the set of all **real-valued** (rather than extended real-valued) μ-integrable functions.

The next proposition generalizes Proposition 2.3.2.

Proposition 2.3.5. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, let $f, g \in L^1(\mu)$, and let $\alpha \in \mathbb{R}$. Then*

(i) $\alpha f \in L^1(\mu)$ *and* $\int_\Omega \alpha f \, d\mu = \alpha \int_\Omega f \, d\mu$,

(ii) $f + g \in L^1(\mu)$ *and* $\int_\Omega (f+g) \, d\mu = \int_\Omega f \, d\mu + \int_\Omega g \, d\mu$,

(iii) *if $f(\omega) \le g(\omega)$ holds for all $\omega \in \Omega$, then* $\int_\Omega f \, d\mu \le \int_\Omega g \, d\mu$.

Proof.

(i) If $\alpha \ge 0$, then $(\alpha f)^+ = \alpha f^+$ and $(\alpha f)^- = \alpha f^-$; thus $(\alpha f)^+$ and $(\alpha f)^-$, and hence αf, are μ-integrable. Then

$$\begin{aligned} \int_\Omega \alpha f \, d\mu &= \int_\Omega (\alpha f)^+ \, d\mu - \int_\Omega (\alpha f)^- \, d\mu \\ &= \alpha \int_\Omega f^+ \, d\mu - \alpha \int_\Omega f^- \, d\mu = \alpha \int_\Omega f \, d\mu. \end{aligned}$$

If $\alpha < 0$, then $(\alpha f)^+ = -\alpha f^-$ and $(\alpha f)^- = -\alpha f^+$; thus $(\alpha f)^+$ and $(\alpha f)^-$, and hence αf, are μ-integrable. Then

$$\begin{aligned} \int_\Omega \alpha f \, d\mu &= \int_\Omega (\alpha f)^+ \, d\mu - \int_\Omega (\alpha f)^- \, d\mu \\ &= (-\alpha) \int_\Omega f^- \, d\mu - (-\alpha) \int_\Omega f^+ \, d\mu = \alpha \int_\Omega f \, d\mu. \end{aligned}$$

(ii) Let $h := f + g$. Since $h^+ \le f^+ + g^+$ and $h^- \le f^- + g^-$ we have that $\int_\Omega h^+ \, d\mu \le \int_\Omega f^+ \, d\mu + \int_\Omega g^+ \, d\mu < \infty$ and $\int_\Omega h^- \, d\mu \le \int_\Omega f^- \, d\mu + \int_\Omega g^- \, d\mu < \infty$. Hence h is μ-integrable.

It follows from $h^+ - h^- = f^+ - f^- + g^+ - g^-$ that

$$h^+ + f^- + g^- = h^- + f^+ + g^+.$$

This implies by part (ii) of Proposition 2.3.2 that

$$\int_\Omega h^+ \, d\mu + \int_\Omega f^- \, d\mu + \int_\Omega g^- \, d\mu = \int_\Omega h^- \, d\mu + \int_\Omega f^+ \, d\mu + \int_\Omega g^+ \, d\mu.$$

Since f, g, and h are μ-integrable, all of the above integrals are finite and we obtain

$$\begin{aligned} \int_\Omega (f+g) \, d\mu &= \int_\Omega h^+ \, d\mu - \int_\Omega h^- \, d\mu \\ &= \int_\Omega f^+ \, d\mu - \int_\Omega f^- \, d\mu + \int_\Omega g^+ \, d\mu - \int_\Omega g^- \, d\mu \\ &= \int_\Omega f \, d\mu + \int_\Omega g \, d\mu. \end{aligned}$$

(iii) If $f(\omega) \le g(\omega)$ holds for all $\omega \in \Omega$, then $g - f$ is a non-negative μ-integrable function. Therefore $\int_\Omega (g-f) \, d\mu \ge 0$, which implies

$$\int_\Omega g \, d\mu - \int_\Omega f \, d\mu = \int_\Omega (g-f) \, d\mu \ge 0.$$

□

In probability theory, one of the most important concepts is that of a random variable. Suppose that $(\Omega, \mathcal{A}, P)$ is a probability space. In this context, $(\mathcal{A}-\mathcal{B})$-measurable functions $X_1, \ldots, X_n\colon \Omega \to \mathbb{R}$ are called a random variables. The concept of random variables is indispensable when several random effects should be modeled simultaneously, for example stock prices at different days. The point of this construction is that both the random behavior of each of these aspects as well as their interoperation is well-defined since all of these random variables are functions that "reside" on one and the same basic space Ω and the probability measure P on $(\Omega, \mathcal{A})$ defines their joint random behavior. On the other hand, if we are merely interested in properties related to each single random variable, the focus will be directed on the random behavior of the **image** of the above functions, e.g. $X_1(\omega)$. Its random behavior is defined by the probability measure P since this assigns probabilities to sets $\{\omega \in \Omega\colon\ X_1(\omega) \in B\}$ for all $B \in \mathcal{B}$. However, it is then more convenient, if we need not resort to the measure P defined on the space $(\Omega, \mathcal{A})$, and if the random behavior of $X(\omega)$ is described by a probability measure on the image space $(\mathbb{R}, \mathcal{B})$ of X_1. This is accomplished by the following definition.

Definition. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, let $(\widetilde{\Omega}, \widetilde{\mathcal{A}})$ be a measurable space, and let $f\colon \Omega \to \widetilde{\Omega}$ be an $(\mathcal{A}-\widetilde{\mathcal{A}})$-measurable function. Then the set function $\mu^f\colon \widetilde{\mathcal{A}} \to [0,\infty]$ defined by

$$\mu^f(B) \;:=\; \mu\big(f^{-1}(B)\big) \;=\; \mu\big(\{\omega \in \Omega\colon\ f(\omega) \in B\}\big) \qquad \forall B \in \widetilde{\mathcal{A}}$$

is called the **image of** μ under f.

It is easy to see that the set function μ^f satisfies the axioms of a measure on $(\widetilde{\Omega}, \widetilde{\mathcal{A}})$. The following proposition shows that integrals over Ω with respect to μ can be equivalently represented as integral over the image space $\widetilde{\Omega}$ with respect to μ^f.

Proposition 2.3.6. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, let $(\widetilde{\Omega}, \widetilde{\mathcal{A}})$ be a measurable space, and let $f\colon \Omega \to \widetilde{\Omega}$ be an $(\mathcal{A}-\widetilde{\mathcal{A}})$-measurable function. Furthermore, let $g\colon \widetilde{\Omega} \to \bar{\mathbb{R}}$ be a $(\widetilde{\mathcal{A}}-\bar{\mathcal{B}})$ measurable function.*

Then the integral $\int_{\widetilde{\Omega}} g\, d\mu^f$ exists if and only if the integral $\int_\Omega (g \circ f)\, d\mu$ exists. If both integrals exist, then

$$\int_{\widetilde{\Omega}} g\, d\mu^f \;=\; \int_\Omega (g \circ f)\, d\mu.$$

Proof.
We consider first the case of a $\widetilde{\mathcal{A}}$-simple function $g = \sum_{i=1}^k \alpha_i \mathbb{1}_{B_i}$, where $\alpha_1, \ldots, \alpha_k \geq 0$, $B_1, \ldots, B_k \in \widetilde{\mathcal{A}}$, and $k \in \mathbb{N}$. Then $(g \circ f)(\omega) = g\big(f(\omega)\big) = \sum_{i=1}^k \alpha_i \mathbb{1}_{B_i}\big(f(\omega)\big)$ and we obtain that

$$\begin{aligned}
\int_\Omega (g \circ f)\, d\mu \;&=\; \sum_{i=1}^k \alpha_i \underbrace{\int_\Omega \mathbb{1}_{B_i}\big(f(\omega)\big)\, d\mu(\omega)}_{=\mu(\{\omega\colon\ f(\omega)\in B_i\})} \\
&=\; \sum_{i=1}^k \alpha_i \mu^f(B_i) \;=\; \sum_{i=1}^k \alpha_i \int_{\widetilde{\Omega}} \mathbb{1}_{B_i}\, d\mu^f \\
&=\; \int_{\widetilde{\Omega}} g\, d\mu^f. \qquad (2.3.2)
\end{aligned}$$

Now we consider the general case. Since g is by assumption $(\widetilde{\mathcal{A}} - \bar{\mathcal{B}})$-measurable and since it then follows from part (ii) of Lemma 2.2.1 that $g \circ f$ is $(\mathcal{A} - \bar{\mathcal{B}})$-measurable, the functions g^+ and g^- are $(\widetilde{\mathcal{A}} - \bar{\mathcal{B}})$-measurable and $(g \circ f)^+$ and $(g \circ f)^-$ are $(\mathcal{A} - \bar{\mathcal{B}})$-measurable. Let $(g_n)_{n \in \mathbb{N}}$ be a sequence of $\widetilde{\mathcal{A}}$-simple functions such that $g_n \nearrow g^+$. Then $(g_n \circ f) \nearrow (g \circ f)^+$ and it follows from (2.3.2) that

$$\int_{\widetilde{\Omega}} g^+ \, d\mu^f \;=\; \lim_{n \to \infty} \int_{\widetilde{\Omega}} g_n \, d\mu^f \;=\; \lim_{n \to \infty} \int_{\Omega} (g_n \circ f) \, d\mu \;=\; \int_{\Omega} (g \circ f)^+ \, d\mu.$$

Likewise, we obtain that

$$\int_{\widetilde{\Omega}} g^- \, d\mu^f \;=\; \int_{\Omega} (g \circ f)^- \, d\mu.$$

If one of the integrals on the left-hand sides is finite, the corresponding integral on the right-hand sides is finite as well, and both integrals $\int_{\widetilde{\Omega}} g \, d\mu^f$ and $\int_{\Omega} (g \circ f) \, d\mu$ exist and are equal. □

In probability theory, the above concepts play an important role. Let $(\Omega, \mathcal{A}, P)$ be an arbitrary probability space, and let $X \colon \Omega \to \mathbb{R}$ be a random variable. Then P^X is called the **distribution of** X **under** P. If $g \colon \mathbb{R} \to \mathbb{R}$ is $(\mathcal{B} - \mathcal{B})$-measurable and if the expected value of $g(X)$ exists, then it can be expressed either by $\int_{\Omega} g(X(\omega)) \, dP(\omega)$ or $\int_{\mathbb{R}} g(x) \, dP^X(x)$.

2.4 Limit theorems

In this section we state and prove the basic limit theorems of integration theory. We begin with results for sequences of non-negative measurable functions. The first of them is a result due to the Italian mathematician Beppo Levi, who proved a slight generalization in 1906 of an earlier result by Henri Lebesgue.

Theorem 2.4.1. (Monotone convergence theorem, Beppo Levi's theorem) *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, and let $(f_n)_{n\in\mathbb{N}}$ be a sequence of non-negative extended real-valued $\mathcal{A}$-measurable functions on Ω such that $f_n(\omega) \le f_{n+1}(\omega)$ holds for all $\omega \in \Omega$ and all $n \in \mathbb{N}$.*

Then $f\colon \Omega \to [0,\infty]$ defined by $f(\omega) := \lim_{n\to\infty} f_n(\omega)$ $\forall \omega \in \Omega$ is an $(\mathcal{A} - \bar{\mathcal{B}})$-measurable function and it holds that

$$\int_\Omega f_n(\omega)\, d\mu \underset{n\to\infty}{\longrightarrow} \int_\Omega f\, d\mu.$$

Proof. The monotonicity of the integral (part (iii) of Proposition 2.3.5) implies that $\int_\Omega f_n\, d\mu \le \int_\Omega f_{n+1}\, d\mu$ holds for all $n \in \mathbb{N}$. Hence the series $\big(\int_\Omega f_n\, d\mu\big)_{n\in\mathbb{N}}$ converges (perhaps to $+\infty$). Since $f_n \le f$ we obtain, again by the monotonicity of the integral, that

$$\int_\Omega f_n\, d\mu \le \int_\Omega f\, d\mu \qquad \forall n \in \mathbb{N},$$

which implies that

$$\lim_{n\to\infty} \int_\Omega f_n\, d\mu \le \int_\Omega f\, d\mu.$$

It remains to prove the reverse inequality. To this end, let $s = \sum_{i=1}^k \alpha_i \mathbb{1}_{A_i}$ be an $\mathcal{A}$-simple function such that $s(\omega) \le f(\omega)$ $\forall \omega \in \Omega$. Let $\epsilon > 0$ be arbitrary, and let

$$E_n := \big\{\omega \in \Omega\colon \quad s(\omega) \le (1+\epsilon) f_n(\omega)\big\}.$$

Then $E_n \in \mathcal{A}$ and $E_n \nearrow \Omega$. Indeed, since $f_n(\omega) \nearrow f(\omega)$, $s(\omega) \le f(\omega)$, and $s(\omega)$ is finite we see that $\omega \in E_N$ for N sufficiently large.[1] By continuity from below we have that $\mu(A_i \cap E_n) \underset{n\to\infty}{\longrightarrow} \mu(A_i)$, which implies that

$$\begin{aligned}\int_\Omega s\, d\mu &= \sum_{i=1}^k \alpha_i \mu\big(A_i\big) = \lim_{n\to\infty} \sum_{i=1}^k \alpha_i \mu\big(A_i \cap E_n\big) \\ &= \lim_{n\to\infty} \int_\Omega \underbrace{s \cdot \mathbb{1}_{E_n}}_{\le (1+\epsilon) f_n}\, d\mu \le (1+\epsilon) \lim_{n\to\infty} \int_\Omega f_n\, d\mu.\end{aligned}$$

This implies that

$$\int_\Omega s\, d\mu \le \lim_{n\to\infty} \int_\Omega f_n\, d\mu,$$

[1] Note that the sets $E_n' = \{\omega\colon f(\omega) \le (1+\epsilon) f_n(\omega)\}$ do not necessarily converge to Ω. We have that $\omega \notin \bigcup_{n\in\mathbb{N}} E_n'$ if $f_n(\omega) < \infty$ $\forall n \in \mathbb{N}$ but $f(\omega) = \infty$. This pitfall is avoided by our reference to the simple function s.

and hence

$$\int_\Omega f\,d\mu \;=\; \sup\Big\{\int_\Omega s\,d\mu\colon\;\; s \text{ simple}, s\le f\Big\} \;\le\; \lim_{n\to\infty}\int_\Omega f_n\,d\mu.$$

This completes the proof. □

Recall that Proposition 2.3.2 states **finite** additivity for integrals of non-negative functions. In conjunction with the Monotone convergenc theorem we can extend this result to **countable** additivity.

Corollary 2.4.2. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, and let $(f_n)_{n\in\mathbb{N}}$ be a sequence of extended real-valued $\mathcal{A}$-measurable functions on Ω. Then*

$$\int_\Omega \sum_{i=1}^{\infty} f_i\,d\mu \;=\; \sum_{i=1}^{\infty}\int_\Omega f_i\,d\mu.$$

Proof. Let

$$g_n \;:=\; f_1+\cdots+f_n.$$

Then $(g_n)_{n\in\mathbb{N}}$ is a non-decreasing sequence of $(\mathcal{A}-\bar{\mathcal{B}})$-measurable functions, $g_n \nearrow g := \sum_{i=1}^{\infty} f_i$, and it follows from the monotone convergence theorem (Theorem 2.4.1) that

$$\int_\Omega \sum_{i=1}^{\infty} f_i\,d\mu \;=\; \int_\Omega g\,d\mu \;=\; \lim_{n\to\infty}\int_\Omega g_n\,d\mu.$$

On the other hand, it follows from part (ii) of Proposition 2.3.2 that

$$\int_\Omega g_n\,d\mu \;=\; \sum_{i=1}^{n}\int_\Omega f_i\,d\mu,$$

which implies that

$$\lim_{n\to\infty}\int_\Omega g_n\,d\mu \;=\; \lim_{n\to\infty}\sum_{i=1}^{n}\int_\Omega f_i\,d\mu \;=\; \sum_{i=1}^{\infty}\int_\Omega f_i\,d\mu.$$

□

Corollary 2.4.2 can be used to construct a large class of measures.

Theorem 2.4.3. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, and let $f, g\colon\, \Omega \to [0,\infty]$ be $(\mathcal{A}-\bar{\mathcal{B}})$-measurable functions. Then*

(i) $\nu\colon\, \mathcal{A} \to [0,\infty]$ *defined by*

$$\nu(A) \;:=\; \int_A f\,d\mu \qquad \forall A\in\mathcal{A}$$

is a measure on $(\Omega, \mathcal{A})$*,*

(ii) $\int_\Omega g\,d\nu \;=\; \int_\Omega g\cdot f\,d\mu.$

The function f is said to be a **density** of ν w.r.t. μ. We have in particular that $\mu(A) = 0$ implies $\nu(A) = 0$ and we say that the measure ν is **absolutely continuous** with respect to the measure μ ($\nu \ll \mu$). We shall see later that the converse statement also holds true: If ν is an arbitrary and μ a σ-finite measure on a measurable space $(\Omega, \mathcal{A})$ and if ν is absolutely continuous w.r.t. μ, then ν has a density f w.r.t. μ, i.e. there exists an $(\mathcal{A} - \bar{\mathcal{B}})$-measurable function $f\colon \Omega \to [0, \infty]$ such that $\nu(A) = \int_A f\, d\mu$ holds for all $A \in \mathcal{A}$. The latter results is also of great importance in probability theory; it will be the basis for an advanced definition of conditional probabilities.

Proof of Theorem 2.4.3.

(i) It is easy to check that the set function ν satisfies the axioms of a measure. Indeed, we have that

$$\nu(\emptyset) = \int_\emptyset f\, d\mu = \int_\Omega \underbrace{f \cdot \mathbb{1}_\emptyset}_{\equiv 0}\, d\mu = 0$$

and, for pairwise disjoint sets $A_1, A_2, \ldots \in \mathcal{A}$,

$$\begin{aligned} \nu\Big(\bigcup_{i=1}^\infty A_i\Big) &= \int_\Omega f \cdot \mathbb{1}_{\bigcup_{i=1}^\infty A_i}\, d\mu \\ &= \int_\Omega \sum_{i=1}^\infty f \cdot \mathbb{1}_{A_i}\, d\mu \\ &= \sum_{i=1}^\infty \int_\Omega f \cdot \mathbb{1}_{A_i}\, d\mu = \sum_{i=1}^\infty \nu(A_i). \end{aligned}$$

(ii) Let $s = \sum_{i=1}^k \alpha_i \mathbb{1}_{A_i}$ be an $\mathcal{A}$-simple function. Then

$$\begin{aligned} \int_\Omega s\, d\nu &= \sum_{i=1}^k \alpha_i\, \nu(A_i) = \sum_{i=1}^k \alpha_i \int_{A_i} f\, d\mu \\ &= \sum_{i=1}^k \alpha_i \int_\Omega f \cdot \mathbb{1}_{A_i}\, d\mu = \int_\Omega s \cdot f\, d\mu. \end{aligned}$$

Now let $g\colon \Omega \to [0, \infty]$ be an arbitrary $(\mathcal{A} - \bar{\mathcal{B}})$-measurable function. Then there exists a sequence $\big(s_n\big)_{n\in\mathbb{N}}$ of $\mathcal{A}$-simple functions such that $s_n \nearrow g$. This implies that $s_n \cdot f \nearrow g \cdot f$, and so

$$\int_\Omega g\, d\nu = \lim_{n\to\infty} \int_\Omega s_n\, d\nu = \lim_{n\to\infty} \int_\Omega s_n \cdot f\, d\mu = \int_\Omega g \cdot f\, d\mu.$$

□

The next result is an immediate consequence of the monotone convergence theorem. It is often used to provide an upper bound for the value of the integral of a function f that can be represented as the limit or limit inferior of a sequence of functions. This theorem is named after the French mathematician Pierre Fatou.

Theorem 2.4.4. (Fatou's lemma)
Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, and let $(f_n)_{n\in\mathbb{N}}$ be a sequence of $[0,\infty]$-valued $(\mathcal{A}-\bar{\mathcal{B}})$-measurable functions on Ω. Then

$$\int_\Omega \liminf_{n\to\infty} f_n \, d\mu \;\leq\; \liminf_{n\to\infty} \int_\Omega f_n \, d\mu.$$

Proof. For each positive integer n, let

$$g_n \;:=\; \inf_{k\geq n} f_k.$$

Then $g_n \nearrow \liminf_{n\to\infty} f_n$ and we obtain from the monotone convergence theorem (Theorem 2.4.1) that

$$\int_\Omega \liminf_{n\to\infty} f_n \, d\mu \;=\; \int_\Omega \lim_{n\to\infty} g_n \, d\mu \;=\; \lim_{n\to\infty} \int_\Omega g_n \, d\mu.$$

On the other hand, we have that $g_n \leq f_n$, which implies $\int_\Omega g_n \, d\mu \leq \int_\Omega f_n d\mu$ for all $n \in \mathbb{N}$, and therefore

$$\lim_{n\to\infty} \int_\Omega g_n \, d\mu \;=\; \liminf_{n\to\infty} \int_\Omega g_n \, d\mu \;\leq\; \liminf_{n\to\infty} \int_\Omega f_n \, d\mu.$$

□

The next theorem provides sufficient conditions under which pointwise convergence of a sequence of functions implies convergence of the corresponding integrals. Its power and utility are two of the primary theoretical advantages of Lebesgue integration over Riemann integration.

Theorem 2.4.5. (Lebesgue's dominated convergence theorem)
Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, and let f and $f_1, f_2, \ldots$ be $[-\infty,\infty]$-valued $(\mathcal{A}-\bar{\mathcal{B}})$-measurable functions such that

$$f_n(\omega) \underset{n\to\infty}{\longrightarrow} f(\omega) \tag{2.4.1}$$

holds μ-almost everywhere, i.e. for all $\omega \in \Omega \setminus N$, where $N \in \mathcal{A}$ and $\mu(N) = 0$. Furthermore, suppose that there exists a $[0,\infty]$-valued $(\mathcal{A}-\bar{\mathcal{B}})$-measurable function g such that $\int_\Omega g \, d\mu < \infty$ and

$$\big|f_n(\omega)\big| \;\leq\; g(\omega) \qquad \forall \omega \in \Omega\setminus N,\ \forall n \in \mathbb{N} \tag{2.4.2}$$

and

$$\int_\Omega g \, d\mu \;<\; \infty. \tag{2.4.3}$$

Then

$$\int_\Omega f_n \, d\mu \underset{n\to\infty}{\longrightarrow} \int_\Omega f \, d\mu.$$

Proof. Since g is μ-integrable we have that $\mu\big(\{\omega\colon\, g(\omega) = \infty\}\big) = 0$. Let $\widetilde{N} = N \cup \{\omega\colon\, g(\omega) = \infty\}$. Then $\widetilde{N} \in \mathcal{A}$ and $\mu\big(\widetilde{N}\big) = 0$. We have that $2g(\omega) - |f_n(\omega) - f(\omega)| \geq 0$ and $f_n(\omega) \underset{n\to\infty}{\longrightarrow} f(\omega)$ holds for all $\omega \in \Omega \setminus \widetilde{N}$. Therefore, we obtain from Fatou's lemma (Theorem 2.4.4) that

$$\begin{aligned}\int_{\Omega\setminus\widetilde{N}} 2g\, d\mu \;&=\; \int_{\Omega\setminus\widetilde{N}} \liminf_n \big(2g - |f_n - f|\big)\, d\mu \\ &\leq\; \liminf_n \int_{\Omega\setminus\widetilde{N}} \big(2g - |f_n - f|\big)\, d\mu \\ &=\; \liminf_n \Big\{ \int_{\Omega\setminus\widetilde{N}} 2g\, d\mu \;-\; \int_{\Omega\setminus\widetilde{N}} |f_n - f|\, d\mu \Big\} \\ &=\; \int_{\Omega\setminus\widetilde{N}} 2g\, d\mu \;-\; \limsup_n \int_{\Omega\setminus\widetilde{N}} |f_n - f|\, d\mu.\end{aligned}$$

Since $\int_{\Omega\setminus\overline{N}} 2g\, d\mu < \infty$ we therefore obtain

$$\limsup_n \int_{\Omega\setminus\widetilde{N}} |f_n - f|\, d\mu \;=\; 0.$$

It follows from (2.4.2) and (2.4.3) that f and $f_1, f_2, \ldots$ are μ-integrable. Since $\mu\big(\widetilde{N}\big) = 0$ we obtain

$$\begin{aligned}\Big|\int_\Omega f_n\, d\mu - \int_\Omega f\, d\mu\Big| \;&=\; \Big|\int_{\Omega\setminus\widetilde{N}} f_n\, d\mu - \int_{\Omega\setminus\widetilde{N}} f\, d\mu\Big| \\ &=\; \Big|\int_{\Omega\setminus\widetilde{N}} f_n - f\, d\mu\Big| \\ &\leq\; \int_{\Omega\setminus\widetilde{N}} |f_n - f|\, d\mu \underset{n\to\infty}{\longrightarrow} 0,\end{aligned}$$

which completes the proof. □

2.5 Riemann integral vs. Lebesgue integral

In this section we show that the Lebesgue integral of a function f over an interval $[a,b]$ with Lebesgue measure λ as integrator coincides with the Riemann integral, provided the latter exists. In this sense, the Lebesgue integral can be regarded as an extension of the Riemann integral. Before we state and prove an exact result, we briefly recall how the Riemann integral is defined.

For $a, b \in \mathbb{R}$, $a < b$, let $[a,b]$ be a closed bounded interval. A **partition** $\mathcal{P}$ of $[a,b]$ is a finite sequence $(a_i)_{i=0,\ldots,n}$ of real numbers such that

$$a \;=\; a_0 \;<\; a_1 \;<\; \ldots \;<\; a_n \;=\; b.$$

Let f be a bounded real-valued function on $[a,b]$. If $\mathcal{P}$ is the partition $(a_i)_{i=0,\ldots,n}$ of $[a,b]$, then the **lower sum** $l(f,\mathcal{P})$ corresponding to f and $\mathcal{P}$ is defined to be $\sum_{i=1}^n \inf\big\{f(x)\colon\, x \in [a_{i-1}, a_i]\big\}(a_i - a_{i-1})$. Likewise we define the **upper sum** $u(f,\mathcal{P})$ corresponding to f and $\mathcal{P}$ as $\sum_{i=1}^n \sup\big\{f(x)\colon\, x \in [a_{i-1}, a_i]\big\}(a_i - a_{i-1})$. Now we define the **lower integral** $\underline{\int}_a^b f(x)\,dx$ of f over $[a,b]$ as the supremum of the lower sums and the **upper integral** $\overline{\int}_a^b f(x)\,dx$ of f over $[a,b]$ as the infimum of the upper sums. It follows immediately that $\underline{\int}_a^b f(x)\,dx \leq \overline{\int}_a^b f(x)\,dx$. If $\underline{\int}_a^b f(x)\,dx = \overline{\int}_a^b f(x)\,dx$, then f is **Riemann integrable** on $[a,b]$, and the common value of $\underline{\int}_a^b f(x)\,dx$ and $\overline{\int}_a^b f(x)\,dx$ is called the **Riemann integral** of f over $[a,b]$ and is denoted by $\int_a^b f(x)\,dx$. It is well-known that a continuous real-valued function f is Riemann integrable over each bounded interval $[a,b]$. The next theorem provides a necessary and sufficient condition for the Riemann integrability of a function and states that the Lebesgue and Riemann integrals coincide, provided the latter exists.

Theorem 2.5.1. *Let $a,b \in \mathbb{R}$, $a < b$, and let f be a bounded real-valued function on $[a,b]$. Then*

(i) *f is Riemann integrable over $[a,b]$*
$\iff$
the set D_f of discontinuity points of f has Lebesgue measure 0.

(ii) *If f is Riemann integrable over $[a,b]$, then*

$$\int_a^b f(x)\,dx = \int_{[a,b]} f\,d\lambda,$$

i.e. the Riemann and Lebesgue integrals of f coincide.
(Note that f need not be $(\mathcal{B}-\mathcal{B})$-measurable. Rather it will be $(\mathcal{M}_{\lambda^*}-\mathcal{B})$-measurable and $\int_{[a,b]} f\,d\lambda$ has to be understood as the Lebesgue integral w.r.t. Lebesgue measure $\lambda^*\big|_{\mathcal{M}_{\lambda^*}}$ (the completion of $\lambda^*\big|_{\mathcal{B}}$).)

Proof.

(i) $(\Longrightarrow)$
Suppose that f is Riemann integrable over $[a,b]$. Then for each positive integer n, we can choose a partition $\mathcal{P} = (a_{n,i})_{i=0,\dots,N_n}$ of $[a,b]$, such that $u(f,\mathcal{P}_n)-l(f,\mathcal{P}_n) \leq 1/n$. By replacing these partitions with finer partitions if necessary, we can assume for each n that $\mathcal{P}_{n+1}$ is a refinement of $\mathcal{P}_n$, i.e. each term of $(a_{n,i})_{i=0,\dots,N_n}$ appears among the terms of $(a_{n+1,i})_{i=0,\dots,N_{n+1}}$. We define for each $n \in \mathbb{N}$

$$\begin{aligned} \beta_{n,i} &:= \inf\big\{f(x)\colon\ x \in [a_{n,i-1}, a_{n,i}]\big\}, \\ \gamma_{n,i} &:= \sup\big\{f(x)\colon\ x \in [a_{n,i-1}, a_{n,i}]\big\} \qquad (i = 1,\dots,N_n) \end{aligned}$$

and

$$\begin{aligned} g_n &:= \beta_{n,1}\mathbb{1}_{[a_{n,0},a_{n,1}]} + \sum_{i=2}^{N_n} \beta_{n,i}\mathbb{1}_{(a_{n,i-1},a_{n,i}]}, \\ h_n &:= \gamma_{n,1}\mathbb{1}_{[a_{n,0},a_{n,1}]} + \sum_{i=2}^{N_n} \gamma_{n,i}\mathbb{1}_{(a_{n,i-1},a_{n,i}]}. \end{aligned}$$

The functions g_n and h_n are $\mathcal{B}$-simple functions and therefore $(\mathcal{B}-\mathcal{B})$-measurable. We have that

$$\begin{aligned} l(f,\mathcal{P}_n) &= \sum_{i=1}^{N_n} \beta_{n,i}\big(a_{n,i}-a_{n,i-1}\big) = \int_{[a,b]} g_n\,d\lambda,\\ u(f,\mathcal{P}_n) &= \sum_{i=1}^{N_n} \gamma_{n,i}\big(a_{n,i}-a_{n,i-1}\big) = \int_{[a,b]} h_n\,d\lambda. \end{aligned}$$

$(g_n)_{n\in\mathbb{N}}$ is an non-decreasing sequence of $\mathcal{B}$-simple functions and it holds that $g_n \nearrow g$, where g is $(\mathcal{B}-\mathcal{B})$-measurable. Likewise, $(h_n)_{n\in\mathbb{N}}$ is an non-increasing sequence of $\mathcal{B}$-simple functions and it holds that $h_n \searrow h$, where h is $(\mathcal{B}-\mathcal{B})$-measurable. It follows from Lebesgue's dominated convergence theorem (Recall that f is bounded.) that

$$0 = \lim_{n\to\infty}\Big(u(f,\mathcal{P}_n)-l(f,\mathcal{P}_n)\Big) = \lim_{n\to\infty}\int_{[a,b]}\big(h_n-g_n\big)\,d\lambda = \int_{[a,b]}\big(h-g\big)\,d\lambda.$$

which implies that

$$\lambda\big(\{x\in[a,b]\colon\ h(x)\neq g(x)\}\big)=0.$$

Note that if $h(x)=g(x)$ and if x is a point in $[a,b]$ that appears in none of the partitions $\mathcal{P}_n$, then f is continuous in x. Therefore,

$$\{x\in[a,b]\colon\ h(x)\neq g(x)\}\ \subseteq\ D_f\ \subseteq\ \{x\in[a,b]\colon\ h(x)\neq g(x)\}\cup\bigcup_{n\in\mathbb{N}}\mathcal{P}_n.$$

Since $\bigcup_{n\in\mathbb{N}}\mathcal{P}_n = \{x\in[a,b]\colon\ x \text{ appears in } \mathcal{P}_n \text{ for some } n\}$ is countable we conclude that $D_f\in\mathcal{B}$ and $\lambda\big(D_f\big)=0$.

$(\Longleftarrow)$

Now suppose that the set D_f of discontinuity points of f has Lebesgue measure 0. For each n let $\mathcal{P}_n$ be the partition of $[a,b]$ that divides $[a,b]$ into 2^n subintervals of equal length, i.e. $\mathcal{P}_n=(a_{n,i})_{i=0,\dots,2^n}$, where $a_i=a+i2^{-n}(b-a)$. Use these partitions $\mathcal{P}_n$ to construct functions g_n and h_n as in the first part of the proof. The relations $f(x)=\lim_n g_n(x)$ and $f(x)=\lim_n h_n(x)$ clearly hold at each x at which f is continuous, and so at almost every x in $[a,b]$. Thus $\lim_n(h_n-g_n)=0$ holds almost everywhere, and so, since $\int_{[a,b]} g_n\,d\lambda = l(f,\mathcal{P}_n)$ and $\int_{[a,b]} h_n\,d\lambda = u(f,\mathcal{P}_n)$, the dominated convergence theorem implies that

$$\lim_n\big(u(f,\mathcal{P}_n)\ -\ l(f,\mathcal{P}_n)\big)\ =\ 0.$$

Hence, f is Riemann integrable over $[a,b]$.

(ii) Suppose that f is Riemann integrable over $[a,b]$. Let the functions g_n, h_n, g, and h be defined as in the first half of the proof of (i). Note also that $g\le f\le h$ and so f differs from g only on a set E_f of Lebesgue measure zero. This λ-null set is not necessarily a Borel set, however it belongs to the completion $\mathcal{M}_{\lambda^*}$ of $\mathcal{B}$. (Recall it is shown in Proposition 2.1.9 in Cohn [2, page 56] that there exists a Lebesgue measurable subset E of $\mathbb{R}$ such that $\lambda(E)=0$ that is not a Borel set.) Let $B\in\mathcal{B}$ be arbitrary. Since $\big\{\omega\colon\ f(\omega)\in B\big\}=\big(\big\{\omega\colon\ g(\omega)\in B\big\}\cap E_f^c\big)\cup\underbrace{\big(\big\{\omega\colon\ f(\omega)\in B\big\}\cap E_f\big)}_{\subseteq E_f}$

we see that the function f is Lebesgue measurable, i.e. $(\mathcal{M}_{\lambda^*} - \mathcal{B})$-measurable. Hence,

$$\int_a^b f(x)\,dx \;=\; \int_{[a,b]} f\,d\lambda^*\big|_{\mathcal{M}_{\lambda^*}}.$$

Since two $(\mathcal{A}{-}\mathcal{B})$-measurable functions f and g defined on a measure space $(\Omega, \mathcal{A}, \mu)$ that are equal μ-almost everywhere have the same integral, there is a convention that for a function f which is only defined outside a μ-zero set Ω_0, the integral $\int_A f\,d\mu$ $(A \in \mathcal{A})$ is defined as $\int_{A\setminus\Omega_0} f\,d\mu$. Following this convention, we could also express the above Riemann integral by

$$\int_{[a,b]\setminus E_f} f\,d\lambda^*\big|_{\mathcal{B}},$$

which is then also written as $\int_{[a,b]} f\,d\lambda^*\big|_{\mathcal{B}}$.

□

We have seen that Riemann integrability of a bounded function over a compact interval implies Lebesgue integrability and that then these integrals coincide. Hence, the concept of the Lebesgue integral can be regarded as a generalization of the Riemann integral. This implication does not hold for **improper** integrals in general.

Example. Consider the function $x \mapsto \sin(x)/x$. Then the (improper) Riemann integral of this function over $[0,\infty)$ exists and is finite. Indeed, we have for each $N \in \mathbb{N}$

$$\int_0^{N\pi} \frac{\sin(x)}{x}\,dx \;=\; \sum_{k=1}^{N} \underbrace{\int_{(k-1)\pi}^{k\pi} \frac{\sin(x)}{x}}_{=:I_k},$$

where $\big(I_k\big)_{k\in\mathbb{N}}$ is an alternating sequence and $|I_k| \searrow 0$. Since in addition $\int_{N\pi}^{(N+1)\pi} |sin(x)/x|\,dx \underset{N\to\infty}{\longrightarrow} 0$ we obtain that $\lim_{b\to\infty} \int_0^b \sin(x)/x\,dx$ exists and is finite. Then the Riemann integral is defined by

$$\int_0^{\infty} \frac{\sin(x)}{x}\,dx \;:=\; \lim_{b\to\infty} \int_0^b \frac{\sin(x)}{x}\,dx.$$

On the other hand, since

$$\begin{aligned}
\int_{[0,\infty)} \Big(\frac{\sin(x)}{x}\Big)^{+} d\lambda(x) \;&=\; \sum_{k=0}^{\infty} \int_{2k\pi}^{(2k+1)\pi} \frac{\sin(x)}{x}\,d\lambda(x) \\
&\geq\; \sum_{k=0}^{\infty} \frac{1}{(2k+1)\pi} \int_0^{\pi} \sin(x)\,d\lambda(x) \;=\; \infty
\end{aligned}$$

and

$$\begin{aligned}
\int_{[0,\infty)} \Big(\frac{\sin(x)}{x}\Big)^{-} d\lambda(x) \;&=\; \sum_{k=1}^{\infty} \int_{(2k-1)\pi}^{2k\pi} \Big|\frac{\sin(x)}{x}\Big|\,d\lambda(x) \\
&\geq\; \sum_{k=1}^{\infty} \frac{1}{2k\pi} \int_{\pi}^{2\pi} \big|\sin(x)\big|\,d\lambda(x) \;=\; \infty
\end{aligned}$$

the Lebesgue integral of this function over $[0,\infty)$ does **not** exist.

2.6 Product spaces and product measures

This section is devoted to measures and integrals on product spaces. Let $(\Omega_1, \mathcal{A}_1, \mu_1)$ and $(\Omega_2, \mathcal{A}_2, \mu_2)$ be measure spaces, and let $\Omega_1 \times \Omega_2$ be the Cartesian product of the sets Ω_1 and Ω_2. We pursue two major goals in this section:

(i) We shall construct a measure $\mu_1 \otimes \mu_2$ on $\Omega_1 \times \Omega_2$, equipped with a suitable σ-algebra, such that

$$(\mu_1 \otimes \mu_2)\big(A_1 \times A_2\big) \;=\; \mu_1\big(A_1\big) \cdot \mu_2\big(A_2\big) \qquad \forall A_1 \in \mathcal{A}_1,\, \forall A_2 \in \mathcal{A}_2.$$

(ii) We shall prove that, under suitable conditions,

$$\begin{aligned}\int_{\Omega_1 \times \Omega_2} f\, d(\mu_1 \otimes \mu_2) \;&=\; \int_{\Omega_2} \left[\int_{\Omega_1} f(\omega_1, \omega_2)\, d\mu_1(\omega_1)\right] d\mu_2(\omega_2) \\ &=\; \int_{\Omega_1} \left[\int_{\Omega_2} f(\omega_1, \omega_2)\, d\mu_2(\omega_2)\right] d\mu_1(\omega_1).\end{aligned}$$

To this end, we define on $\Omega_1 \times \Omega_2$ the so-called **product σ-algebra** $\mathcal{A}_1 \otimes \mathcal{A}_2$, which is given by

$$\mathcal{A}_1 \otimes \mathcal{A}_2 \;:=\; \sigma\big(\{A_1 \times A_2\colon\; A_1 \in \mathcal{A}_1,\, A_2 \in \mathcal{A}_2\}\big).$$

As an example, consider the space $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. On $\mathbb{R}$, the standard choice of a σ-algebra is the Borel σ-algebra $\mathcal{B}$ which is generated by the collection $\mathcal{I}^1$ of half-open intervals. Likewise, the Borel σ-algebra $\mathcal{B}^2$ on $\mathbb{R}^2$ is generated by the collection of half-open rectangles,

$$\mathcal{B}^2 \;=\; \sigma\big(\mathcal{I}^2\big) \;=\; \sigma(\{(a_1, b_1] \times (a_2, b_2]\colon\; a_1, a_2, b_1, b_2 \in \mathbb{R},\; a_i \le b_i\}\big).$$

On the other hand, the product of $\mathcal{B}$ with itself is equal to

$$\mathcal{B} \otimes \mathcal{B} \;=\; \sigma\big(\{A_1 \times A_2\colon\; A_1, A_2 \in \mathcal{B}\}\big).$$

It is not difficult to see that the product σ-algebra $\mathcal{B} \otimes \mathcal{B}$ is equal to $\mathcal{B}^2$. Since the generator $\{(a_1, b_1] \times (a_2, b_2]\colon\; a_1, a_2, b_1, b_2 \in \mathbb{R},\; a_i \le b_i\}$ of $\mathcal{B}^2$ is included in the generator $\{A_1 \times A_2\colon\; A_1, A_2 \in \mathcal{B}\}$ of $\mathcal{B} \otimes \mathcal{B}$ it follows that

$$\mathcal{B}^2 \;\subseteq\; \mathcal{B} \otimes \mathcal{B}. \tag{2.6.1a}$$

it remains to show the reverse inclusion, i.e.

$$\mathcal{B} \otimes \mathcal{B} \;=\; \sigma\big(\{A_1 \times A_2\colon\; A_1, A_2 \in \mathcal{B}\}\big) \;\subseteq\; \mathcal{B}^2. \tag{2.6.1b}$$

We first show that, for arbitrary $A_1, A_2 \in \mathcal{B}$, $A_1 \times A_2$ belongs to $\mathcal{B}^2$. Consider the projections π_1 and π_2 of $\mathbb{R}^2$ onto $\mathbb{R}$ defined by $\pi_1(x, y) = x$ and $\pi_2(x, y) = y$. Since $\pi_i\colon \mathbb{R}^2 \to \mathbb{R}$ are continuous functions, and hence $(\mathcal{B}^2 - \mathcal{B})$-measurable, we obtain that $\pi_1^{-1}(A_1) = A_1 \times \mathbb{R} \in \mathcal{B}^2$ and $\pi_2^{-1}(A_2) = \mathbb{R} \times A_2 \in \mathcal{B}^2$. Therefore, $A_1 \times A_2 \in \mathcal{B}^2$, which implies $\mathcal{B} \otimes \mathcal{B} = \sigma\big(\{A_1 \times A_2\colon\; A_1, A_2 \in \mathcal{B}\}\big) \subseteq \sigma\big(\mathcal{B}^2\big) = \mathcal{B}^2$.

In this section we shall first construct the product measure $\mu_1 \otimes \mu_2$ on $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ which is such that

$$(\mu_1 \otimes \mu_2)\big(A_1 \times A_2\big) \;=\; \mu_1\big(A_1\big) \cdot \mu_2\big(A_2\big) \qquad \forall A_1 \in \mathcal{A}_1,\, \forall A_2 \in \mathcal{A}_2.$$

For an arbitrary set $E \in \mathcal{A}_1 \otimes \mathcal{A}_2$, we shall define

$$\begin{aligned} \mu_1 \otimes \mu_2(E) &= \int_{\Omega_1} \mu_2\big(\{\omega_2 \in \Omega_2 \colon\ (\omega_1, \omega_2) \in E\}\big)\, d\mu_1(\omega_1) \\ &= \int_{\Omega_2} \mu_1\big(\{\omega_1 \in \Omega_1 \colon\ (\omega_1, \omega_2) \in E\}\big)\, d\mu_2(\omega_2). \end{aligned}$$

This requires in particular that for each $\omega_1 \in \Omega_1$ and for each $\omega_2 \in \Omega_2$ the so-called **sections** E_{ω_1} and E^{ω_2} of E which are given by

$$E_{\omega_1} = \big\{\omega_2 \in \Omega_2 \colon\ (\omega_1, \omega_2) \in E\big\}$$

and

$$E^{\omega_2} = \big\{\omega_1 \in \Omega_1 \colon\ (\omega_1, \omega_2) \in E\big\}$$

belong to $\mathcal{A}_2$ and $\mathcal{A}_1$, respectively. Furthermore, for the above iterated integrals to exist, the functions $\omega_2 \mapsto \mu_1(E^{\omega_2})$ and $\omega_1 \mapsto \mu_2(E_{\omega_1})$ must be $(\mathcal{A}_2 - \bar{\mathcal{B}})$-measurable and $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable, respectively. The following lemma shows that these technical requirements are fulfilled.

Lemma 2.6.1. *Suppose that $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ are measurable spaces and that $E \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Then*

(i) *For each $\omega_1 \in \Omega_1$ and each $\omega_2 \in \Omega_2$,*

$$E_{\omega_1} \in \mathcal{A}_2 \qquad \textit{and} \qquad E^{\omega_2} \in \mathcal{A}_1.$$

(ii) *If μ_1 and μ_2 are σ-finite measures on $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$, respectively, then*

$$\omega_1 \mapsto \mu_2\big(E_{\omega_1}\big) \quad \textit{is} \quad (\mathcal{A}_1 - \bar{\mathcal{B}})\textit{-measurable}$$

and

$$\omega_2 \mapsto \mu_1\big(E^{\omega_2}\big) \quad \textit{is} \quad (\mathcal{A}_2 - \bar{\mathcal{B}})\textit{-measurable.}$$

Proof.

(i) We prove only the first statement since the second can be proved analogously. For $\omega_1 \in \Omega_1$, define the corresponding system of good sets by

$$\mathcal{G}_{\omega_1} := \big\{E \subseteq \Omega_1 \times \Omega_2 \colon \quad E_{\omega_1} \in \mathcal{A}_2\big\}.$$

Then

a) For $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$,

$$\big(A_1 \times A_2\big)_{\omega_1} = \begin{cases} A_2 & \text{if } \omega_1 \in A_1, \\ \emptyset & \text{if } \omega_1 \notin A_1. \end{cases}$$

Therefore,

$$\big\{A_1 \times A_2 \colon\ A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2\big\} \subseteq \mathcal{G}_{\omega_1}.$$

b) $\mathcal{G}_{\omega_1}$ is a σ-algebra on $\Omega_1 \times \Omega_2$. Indeed, we have that

* $\big(\Omega_1 \times \Omega_2\big)_{\omega_1} = \Omega_2 \in \mathcal{A}_2$, hence $\Omega_1 \times \Omega_2 \in \mathcal{G}_{\omega_1}$.
* If $A \in \mathcal{G}_{\omega_1}$, then $A_{\omega_1} \in \mathcal{A}_2$, and so $\big(A^c\big)_{\omega_1} = \big(A_{\omega_1}\big)^c \in \mathcal{A}_2$. This yields $A^c \in \mathcal{G}_{\omega_1}$.
* Finally, if $A_1, A_2, \ldots \in \mathcal{G}_{\omega_1}$, then $(A_1)_{\omega_1}, (A_2)_{\omega_1}, \ldots \in \mathcal{A}_2$, and so $\big(\bigcup_{i=1}^\infty A_i\big)_{\omega_1} = \bigcup_{i=1}^\infty (A_i)_{\omega_1} \in \mathcal{A}_2$. Hence, $\bigcup_{i=1}^\infty A_i \in \mathcal{G}_{\omega_1}$.

Therefore,

$$\mathcal{A}_1 \otimes \mathcal{A}_2 \;=\; \sigma\big(\{A_1 \times A_2 \colon\; A_1 \in \mathcal{A}_1,\; A_2 \in \mathcal{A}_2\}\big) \overset{(a)}{\subseteq} \sigma\big(\mathcal{G}_{\omega_1}\big) \overset{(b)}{=} \mathcal{G}_{\omega_1},$$

which proves (i).

(ii) We prove again only the first statement. The second one can be proved analogously. First of all, it follows from part (i) that $E_{\omega_1} \in \mathcal{A}_2$, and hence that $\mu_2\big(E_{\omega_1}\big)$ is defined. For $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$,

$$\mu_2\big((A_1 \times A_2)_{\omega_1}\big) \;=\; \mu_2(A_2)\,\mathbb{1}_{A_1}(\omega_1).$$

Hence, the mapping $\omega_1 \mapsto \mu_2\big((A_1 \times A_2)_{\omega_1}\big)$ is $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable.

Suppose first that $\mu_2(\Omega_2) < \infty$. To prove measurability of the mapping $\omega_1 \mapsto \mu_2\big(E_{\omega_1}\big)$ for all $E \in \mathcal{A}_1 \otimes \mathcal{A}_2$, we consider the system of good sets

$$\mathcal{G} \;:=\; \big\{E \in \mathcal{A}_1 \otimes \mathcal{A}_2 \colon \quad \omega_1 \mapsto \mu_2\big(E_{\omega_1}\big) \quad \text{is} \quad (\mathcal{A}_1 - \bar{\mathcal{B}})\text{-measurable}\big\}.$$

We show that $\mathcal{G}$ is a Dynkin system on $\Omega_1 \times \Omega_2$:

a) It follows from the above considerations that $\Omega_1 \times \Omega_2 \in \mathcal{G}$.

b) Suppose that $E \in \mathcal{G}$, i.e. the mapping $\omega_1 \mapsto \mu_2\big(E_{\omega_1}\big)$ is $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable. Since $\mu_2\big((E^c)_{\omega_1}\big) = \mu_2\big((E_{\omega_1})^c\big) = \mu_2\big((\Omega_1 \times \Omega_2)_{\omega_1}\big) - \mu_2\big(E_{\omega_1}\big)$ we see that $\omega_1 \mapsto \mu_2\big((E^c)_{\omega_1}\big)$ is also $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable[2]. Hence $E^c \in \mathcal{G}$.

c) If $E_1, E_2, \ldots$ are disjoint sets that belong to $\mathcal{G}$, then the mappings $\omega_1 \mapsto \mu_2\big((E_i)_{\omega_1}\big)$ $(i \in \mathbb{N})$ are $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable. For each ω_1, the sections $(E_1)_{\omega_1}$, $(E_2)_{\omega_1}$, ... are disjoint sets that belong to $\mathcal{A}_2$, and we obtain that

$$\mu_2\Big(\big(\bigcup_{i=1}^\infty E_i\big)_{\omega_1}\Big) \;=\; \mu_2\Big(\bigcup_{i=1}^\infty (E_i)_{\omega_1}\Big) \;=\; \sum_{i=1}^\infty \mu_2\big((E_i)_{\omega_1}\big).$$

Hence, the mapping $\omega_1 \mapsto \mu_2\big(\big(\bigcup_{i=1}^\infty E_i\big)_{\omega_1}\big)$ is $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable, i.e. $\bigcup_{i=1}^\infty E_i \in \mathcal{G}$.

It follows from a) to c) that $\mathcal{G}$ is a Dynkin system on $\Omega_1 \times \Omega_2$. Since $\big\{A_1 \times A_2 \colon\; A_1 \in \mathcal{A}_1,\; A_2 \in \mathcal{A}_2\big\}$ is an intersection-stable collection of sets that is contained in $\mathcal{G}$ we obtain from Lemma 1.3.7 that

$$\begin{aligned}\mathcal{A}_1 \otimes \mathcal{A}_2 \;&=\; \sigma\big(\{A_1 \times A_2 \colon\; A_1 \in \mathcal{A}_1,\; A_2 \in \mathcal{A}_2\}\big) \\ &=\; \delta\big(\{A_1 \times A_2 \colon\; A_1 \in \mathcal{A}_1,\; A_2 \in \mathcal{A}_2\}\big) \;\subseteq\; \delta\big(\mathcal{G}\big) \;=\; \mathcal{G}.\end{aligned}$$

[2]Here we use the fact that μ_2 is finite.

Now suppose that μ_2 is only σ-finite rather than finite. Then there exists a sequence of sets $(F_n)_{n\in\mathbb{N}}$ such that $F_n \in \mathcal{A}_2$, $F_n \subseteq F_{n+1}$, and $\mu_2(F_n) < \infty$ hold for all $n \in \mathbb{N}$, and that $\bigcup_{n=1}^{\infty} F_n = \Omega_2$. Then $\mu_{2,n}$ defined by

$$\mu_{2,n}\big(A\big) \;=\; \mu_2\big(A \cap F_n\big) \qquad \forall A \in \mathcal{A}_2$$

is a finite measure and it follows that $\omega_1 \mapsto \mu_{2,n}\big(E_{\omega_1}\big)$ is $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable for all $E \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Since $E_{\omega_1} \cap F_n \nearrow E_{\omega_1}$ it follows from continuity from below that $\mu_{2,n}(E_{\omega_1}) \nearrow \mu_2(E_{\omega_1})$. Hence, $\omega_1 \mapsto \mu_2(E_{\omega_1})$ is the limit of $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable mappings, and therefore also $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable.

□

Theorem 2.6.2. *Let $(\Omega_1, \mathcal{A}_1, \mu_1)$ and $(\Omega_2, \mathcal{A}_2, \mu_2)$ be σ-finite measure spaces. Then there exists a unique measure $\mu_1 \otimes \mu_2$ on $\mathcal{A}_1 \otimes \mathcal{A}_2$ such that*

$$(\mu_1 \otimes \mu_2)\big(A_1 \times A_2\big) \;=\; \mu_1\big(A_1\big) \cdot \mu_2\big(A_2\big) \qquad \forall A_1 \in \mathcal{A}_1, \forall A_2 \in \mathcal{A}_2. \tag{2.6.2}$$

The measure $\mu_1 \otimes \mu_2$ is called the **product** *of μ_1 and μ_2.*

Furthermore, the measure under $\mu_1 \otimes \mu_2$ of an arbitrary set $E \in \mathcal{A}_1 \otimes \mathcal{A}_2$ is given by

$$(\mu_1 \otimes \mu_2)\big(E\big) \;=\; \int_{\Omega_1} \mu_2\big(E_{\omega_1}\big)\, d\mu_1(\omega_1) \;=\; \int_{\Omega_2} \mu_1\big(E^{\omega_2}\big)\, d\mu_2(\omega_2). \tag{2.6.3}$$

Proof. Let $E \in \mathcal{A}_1 \otimes \mathcal{A}_2$ be arbitrary. Recall that it follows from part (i) of Lemma 2.6.1 that $E_{\omega_1} \in \mathcal{A}_2 \;\; \forall \omega_1 \in \Omega_1$ and $E^{\omega_2} \in \mathcal{A}_1 \;\; \forall \omega_2 \in \Omega_2$. Hence $\mu_2(E_{\omega_1})$ and $\mu_1(E^{\omega_2})$ are defined. Furthermore, part (ii) of Lemma 2.6.1 states that the mappings $\omega_1 \mapsto \mu_2(E_{\omega_1})$ and $\omega_2 \mapsto \mu_1(E^{\omega_2})$ are $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable and $(\mathcal{A}_2 - \bar{\mathcal{B}})$-measurable, respectively. Therefore, the integrals in equation (2.6.3) exist. Thus we can define functions $(\mu_1 \otimes \mu_2)_1$ and $(\mu_1 \otimes \mu_2)_2$ on $\mathcal{A}_1 \otimes \mathcal{A}_2$ by

$$(\mu_1 \otimes \mu_2)_1\big(E\big) \;:=\; \int_{\Omega_2} \mu_2\big(E_{\omega_1}\big)\, d\mu_1(\omega_1)$$

and

$$(\mu_1 \otimes \mu_2)_2\big(E\big) \;:=\; \int_{\Omega_1} \mu_1\big(E^{\omega_2}\big)\, d\mu_2(\omega_2).$$

It is easy to see that $(\mu_1 \otimes \mu_2)_1$ and $(\mu_1 \otimes \mu_2)_2$ are measures on $\mathcal{A}_1 \otimes \mathcal{A}_2$. We show this for $(\mu_1 \otimes \mu_2)_1$. $(\mu_1 \otimes \mu_2)_1$ is obviously a non-negative set function and

$$(\mu_1 \otimes \mu_2)_1\big(\emptyset\big) \;=\; \int_{\Omega_2} \underbrace{\mu_2\big(\emptyset_{\omega_1}\big)}_{=0}\, d\mu_1(\omega_1) \;=\; 0.$$

If $E_1, E_2, \dots$ are arbitrary disjoint sets that belong to $\mathcal{A}_1 \otimes \mathcal{A}_2$, then $(E_1)_{\omega_1}, (E_2)_{\omega_1}, \dots$ are disjoint sets in $\mathcal{A}_2$, and we obtain from Corollary 2.4.2

$$\begin{aligned}
(\mu_1 \otimes \mu_2)_1\Big(\bigcup_{i=1}^{\infty} E_i\Big) &= \int_{\Omega_2} \mu_2\Big(\big(\bigcup_{i=1}^{\infty} E_i\big)_{\omega_1}\Big)\, d\mu_1(\omega_1) \\
&= \int_{\Omega_2} \sum_{i=1}^{\infty} \mu_2\big((E_i)_{\omega_1}\big)\, d\mu_1(\omega_1) \\
&\overset{Cor.\ 2.4.2}{=} \sum_{i=1}^{\infty} \int_{\Omega_2} \mu_2\big((E_i)_{\omega_1}\big)\, d\mu_1(\omega_1) \\
&= \sum_{i=1}^{\infty} (\mu_1 \otimes \mu_2)_1\big(E_i\big).
\end{aligned}$$

We can show similarly that $(\mu_1 \otimes \mu_2)_2$ satisfies the axioms of a measure on $\mathcal{A}_1 \otimes \mathcal{A}_2$.

If $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$, then

$$\begin{aligned}
\int_{\Omega_1} \underbrace{\mu_2\big((A_1 \times A_2)_{\omega_1}\big)}_{=\,\mu_2(A_2)\, \mathbb{1}_{A_1}(\omega_1)}\, d\mu_1(\omega_1) &= \mu_2\big(A_2\big) \cdot \int_{\Omega_2} \mathbb{1}_{A_1}(\omega_1)\, d\mu_1(\omega_1) \\
&= \mu_1\big(A_1\big) \cdot \mu_2\big(A_2\big)
\end{aligned}$$

and, likewise,

$$\int_{\Omega_2} \mu_1\big((A_1 \times A_2)^{\omega_2}\big)\, d\mu_2(\omega_2) \;=\; \mu_1\big(A_1\big) \cdot \mu_2\big(A_2\big).$$

Therefore, $(\mu_1 \otimes \mu_2)_1$ as well as $(\mu_1 \otimes \mu_2)_2$ satisfy (2.6.2).

As for uniqueness, let μ be an arbitrary measure on $\mathcal{A}_1 \otimes \mathcal{A}_2$ such that

$$\mu\big(A_1 \times A_2\big) \;=\; \mu_1\big(A_1\big) \cdot \mu_2\big(A_2\big) \qquad \forall A_1 \in \mathcal{A}_1,\ \forall A_2 \in \mathcal{A}_2.$$

Since μ_1 and μ_2 are σ-finite measures there exist sequences $(E_n)_{n\in\mathbb{N}}$ in $\mathcal{A}_1$ and $(F_n)_{n\in\mathbb{N}}$ in $\mathcal{A}_2$ such that $\mu_1(E_n) < \infty\ \forall n$ and $E_n \nearrow \Omega_1$ as well as $\mu_2(F_n) < \infty\ \forall n$ and $F_n \nearrow \Omega_2$. Then $E_n \times F_n \nearrow \Omega_1 \times \Omega_2$ and

$$\mu\big(E_n \times F_n\big) \;=\; (\mu_1 \otimes \mu_2)_1\big(E_n \times F_n\big) \;=\; (\mu_1 \otimes \mu_2)_2\big(E_n \times F_n\big) \;<\; \infty \qquad \forall n \in \mathbb{N}.$$

Since the measures μ, $(\mu_1 \otimes \mu_2)_1$, and $(\mu_1 \otimes \mu_2)_2$ are equal on the intersection-stable collection of sets $\big\{A_1 \times A_2\colon\ A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2\big\}$, it follows from the uniqueness Theorem (Theorem 1.3.8) that

$$\mu\big(E\big) \;=\; (\mu_1 \otimes \mu_2)_1\big(E\big) \;=\; (\mu_1 \otimes \mu_2)_2\big(E\big) \quad \forall E \in \sigma\big(\{A_1 \times A_2\colon\ A_1 \in \mathcal{A}_1\ A_2 \in \mathcal{A}_2\}\big) \;=\; \mathcal{A}_1 \otimes \mathcal{A}_2.$$

□

Now we turn to the announced result for integrals on product spaces. Let $(\Omega_1, \mathcal{A}_1, \mu_1)$ and $(\Omega_2, \mathcal{A}_2, \mu_2)$ be σ-finite measure spaces. Note that the relation (2.6.2) in Theorem 2.6.2 can be represented in the form

$$\begin{aligned}\int_{\Omega_1\times\Omega_2} \mathbb{1}_E \, d(\mu_1 \otimes \mu_2) &= \int_{\Omega_2} \Big[\underbrace{\int_{\Omega_1} \mathbb{1}_E(\omega_1,\omega_2)\, d\mu_1(\omega_1)}_{=\mu_1(E^{\omega_2})} \Big] \, d\mu_2(\omega_2) \\ &= \int_{\Omega_1} \Big[\underbrace{\int_{\Omega_2} \mathbb{1}_E(\omega_1,\omega_2)\, d\mu_2(\omega_2)}_{=\mu_2(E_{\omega_1})} \Big] \, d\mu_1(\omega_1).\end{aligned}$$

For a non-negative $(\mathcal{A}_1 \otimes \mathcal{A}_2 - \bar{\mathcal{B}})$-measurable function $f\colon \Omega_1 \times \Omega_2 \to \bar{\mathbb{R}}$, we shall prove that

$$\begin{aligned}\int_{\Omega_1\times\Omega_2} f \, d(\mu_1 \otimes \mu_2) &= \int_{\Omega_2} \Big[\int_{\Omega_1} f(\omega_1,\omega_2)\, d\mu_1(\omega_1) \Big] \, d\mu_2(\omega_2) \\ &= \int_{\Omega_1} \Big[\int_{\Omega_2} f(\omega_1,\omega_2)\, d\mu_2(\omega_2) \Big] \, d\mu_1(\omega_1).\end{aligned}$$

Before we state and prove corresponding results, we convince ourselves that the above iterated integrals exist, i.e. we verify in particular that for each $\omega_2 \in \Omega_2$ the so-called **sections** $f^{\omega_2}\colon \Omega_1 \to \bar{\mathbb{R}}$ defined by $f^{\omega_2}(\omega_1) = f(\omega_1, \omega_2)$ are $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable, and that for each $\omega_1 \in \Omega_1$ the sections $f_{\omega_1}\colon \Omega_2 \to \bar{\mathbb{R}}$ defined by $f_{\omega_1}(\omega_2) = f(\omega_1, \omega_2)$ are $(\mathcal{A}_2 - \bar{\mathcal{B}})$-measurable. Furthermore, for the existence of the outer integrals it is required that the functions $\omega_2 \mapsto \int_{\Omega_1} f^{\omega_2}(\omega_1)\, d\mu_1(\omega_1)$ and $\omega_1 \mapsto \int_{\Omega_2} f_{\omega_1}(\omega_2)\, d\mu_2(\omega_2)$ are $(\mathcal{A}_2 - \bar{\mathcal{B}})$-measurable and $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable, respectively. The next lemma clarifies these technical details.

Lemma 2.6.3.
Let $(\Omega_1, \mathcal{A}_1, \mu_1)$ and $(\Omega_2, \mathcal{A}_2, \mu_2)$ be σ-finite measure spaces, and let $f\colon \Omega_1 \times \Omega_2 \to [0, \infty]$ be an $(\mathcal{A}_1 \otimes \mathcal{A}_2 - \bar{\mathcal{B}})$-measurable function. Then

(i) *for each $\omega_2 \in \Omega_2$ the sections $f^{\omega_2}\colon \Omega_1 \to [0, \infty]$ are $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable and for each $\omega_1 \in \Omega_1$ the sections $f_{\omega_1}\colon \Omega_2 \to [0, \infty]$ are $(\mathcal{A}_2 - \bar{\mathcal{B}})$-measurable, and*

(ii) *the function $\omega_2 \mapsto \int_{\Omega_1} f^{\omega_2}\, d\mu_1$ is $(\mathcal{A}_2 - \bar{\mathcal{B}})$-measurable and the function $\omega_1 \mapsto \int_{\Omega_2} f_{\omega_1}\, d\mu_2$ is $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable.*

Proof. The proof will be split up into three steps.

a) First we consider the case that $f = \mathbb{1}_E$, where $E \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Then the sections f^{ω_2} and f_{ω_1} are the respective characteristic functions $\mathbb{1}_{E^{\omega_2}}$ and $\mathbb{1}_{E_{\omega_1}}$. We have from part (i) of Lemma 2.6.1 that $E^{\omega_2} \in \mathcal{A}_1$ and $E_{\omega_1} \in \mathcal{A}_2$, which implies that $f^{\omega_2} = \mathbb{1}_{E^{\omega_2}}$ is $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable and $f_{\omega_1} = \mathbb{1}_{E_{\omega_1}}$ is $(\mathcal{A}_2 - \bar{\mathcal{B}})$-measurable.

   Since $\int_{\Omega_1} \mathbb{1}_{E^{\omega_2}}\, d\mu_1 = \mu_1(E^{\omega_2})$ it follows from part (ii) of Lemma 2.6.1 that $\omega_2 \mapsto \int_{\Omega_1} f^{\omega_2}\, d\mu_1$ is $(\mathcal{A}_2 - \bar{\mathcal{B}})$-measurable. Likewise, since $\int_{\Omega_2} \mathbb{1}_{E_{\omega_1}}\, d\mu_2 = \mu_2(E_{\omega_1})$ we obtain that $\omega_1 \mapsto \int_{\Omega_2} f_{\omega_1}\, d\mu_2$ is $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable.

b) Now suppose that f is an $(\mathcal{A}_1 \otimes \mathcal{A}_2)$-simple function, i.e. $f = \sum_{i=1}^k \alpha_i \mathbb{1}_{E_i}$, where $\alpha_1, \dots, \alpha_k \geq 0$ and $E_1, \dots, E_k \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Then $f^{\omega_2} = \sum_{i=1}^k \alpha_i \mathbb{1}_{(E_i)^{\omega_2}}$ and $f_{\omega_1} = \sum_{i=1}^k \alpha_i \mathbb{1}_{(E_i)_{\omega_1}}$. Hence, it follows from Lemma 2.2.1 and its Corollaries 2.2.2 and 2.2.3 that f^{ω_2} is $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable and f_{ω_1} is $(\mathcal{A}_2 - \bar{\mathcal{B}})$-measurable.

   It follows from (i) and (ii) of Proposition 2.3.2 that $\int_{\Omega_1} f^{\omega_2}\, d\mu_1 = \sum_{i=1}^k \alpha_i \int_{\Omega_1} \mathbb{1}_{(E_i)^{\omega_2}}\, d\mu_1$; hence $\omega_2 \mapsto \int_{\Omega_1} f^{\omega_2}\, d\mu_1$ is $(\mathcal{A}_2 - \bar{\mathcal{B}})$-measurable. Likewise we obtain that $\omega_1 \mapsto \int_{\Omega_2} f_{\omega_1}\, d\mu_2$ is $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable.

c) Finally, let $f\colon \Omega_1 \times \Omega_2 \to [0,\infty]$ be an arbitrary non-negative $(\mathcal{A}_1 \otimes \mathcal{A}_2 - \bar{\mathcal{B}})$-measurable function. Then there exists a sequence $(f_n)_{n\in\mathbb{N}}$ of $(\mathcal{A}_1 \otimes \mathcal{A}_2)$-simple functions such that $f_n \nearrow f$. Since $(f_n)^{\omega_2} \nearrow f^{\omega_2}$ and $(f_n)_{\omega_1} \nearrow f_{\omega_1}$ we obtain from (iii) of Proposition 2.2.4 that f^{ω_2} is $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable and f_{ω_1} is $(\mathcal{A}_2 - \bar{\mathcal{B}})$-measurable.

   It follows from Beppo Levi's theorem (Theorem 2.4.1) that $\int_{\Omega_1} (f_n)^{\omega_2}\, d\mu_1 \nearrow \int_{\Omega_1} f^{\omega_2}\, d\mu_1$ and $\int_{\Omega_1} (f_n)_{\omega_1}\, d\mu_2 \nearrow \int_{\Omega_1} f_{\omega_1}\, d\mu_2$. Hence, $\omega_2 \mapsto \int_{\Omega_1} f^{\omega_2}\, d\mu_1$ is $(\mathcal{A}_2 - \bar{\mathcal{B}})$-measurable and $\omega_1 \mapsto \int_{\Omega_2} f_{\omega_1}\, d\mu_2$ is $(\mathcal{A}_1 - \bar{\mathcal{B}})$-measurable.

□

Now we are in a position to state and prove our main results in this subsection, Tonelli's theorem and Fubini's theorem. These results give conditions under which it is possible to compute an integral over a product domain by using an iterated integral. Furthermore, it allows the order of integration to be changed in certain iterated integrals, which is a useful tool for the practical computation of such integrals. Fubini's theorem was proved in 1907 by the Italian mathematician Guido Fubini. Tonelli's theorem is a variation of Fubini's theorem and applies to non-negative rather integrable functions. It was proved by the Italian mathematician Leonida Tonelli in 1909. We begin with the latter theorem since it can be used as a basis for a proof of Fubini's theorem.

Theorem 2.6.4. (Tonelli's theorem)
Let $(\Omega_1, \mathcal{A}_1, \mu_1)$ and $(\Omega_2, \mathcal{A}_2, \mu_2)$ be σ-finite measure spaces, and let $f\colon \Omega_1 \times \Omega_2 \to [0,\infty]$ be an $(\mathcal{A}_1 \otimes \mathcal{A}_2 - \bar{\mathcal{B}})$-measurable function. Then

$$\int_{\Omega_1\times\Omega_2} f\, d(\mu_1 \otimes \mu_2) = \int_{\Omega_2} \left[\int_{\Omega_1} f^{\omega_2}(\omega_1)\, d\mu_1(\omega_1) \right] d\mu_2(\omega_2) \tag{2.6.4a}$$

$$= \int_{\Omega_1} \left[\int_{\Omega_2} f_{\omega_1}(\omega_2)\, d\mu_2(\omega_2) \right] d\mu_1(\omega_1). \tag{2.6.4b}$$

Proof. First of all, it follows from Lemma 2.6.3 that all of the above integrals exist.

Suppose first that f is an $(\mathcal{A}_1 \otimes \mathcal{A}_2)$-simple function, i.e. $f = \sum_{i=1}^k \alpha_i \mathbb{1}_{E_i}$, where $\alpha_1, \dots, \alpha_k \geq 0$ and $E_1, \dots, E_k \in \mathcal{A}_1 \otimes \mathcal{A}_2$. We obtain from Proposition 2.3.2

$$\int_{\Omega_1\times\Omega_2} f\, d(\mu_1 \otimes \mu_2) = \sum_{i=1}^k \alpha_i \int_{\Omega_1\times\Omega_2} \mathbb{1}_{E_i}\, d(\mu_1 \otimes \mu_2) = \sum_{i=1}^k \alpha_i\, (\mu_1 \otimes \mu_2)(E_i)$$

and

$$\int_{\Omega_2}\left[\int_{\Omega_1} f^{\omega_2}(\omega_1)\,d\mu_1(\omega_1)\right]d\mu_2(\omega_2) \;=\; \int_{\Omega_2}\left[\sum_{i=1}^{k}\alpha_i \underbrace{\int_{\Omega_1}\mathbb{1}_{(E_i)^{\omega_2}}(\omega_1)\,d\mu_1(\omega_1)}_{=\mu_1\left((E_i)^{\omega_2}\right)}\right]d\mu_2(\omega_2)$$
$$\;=\; \sum_{i=1}^{k}\alpha_i\int_{\Omega_2}\mu_1\big((E_i)^{\omega_2}\big)\,d\mu_2(\omega_2).$$

Since by equation (2.6.3) in Theorem 2.6.2 the relation $(\mu_1 \otimes \mu_2)\big(E_i\big) = \int_{\Omega_2}\mu_1\big((E_i)^{\omega_2}\big)\,d\mu_2(\omega_2)$ follows we obtain that

$$\int_{\Omega_1\times\Omega_2} f\,d(\mu_1\otimes\mu_2) \;=\; \int_{\Omega_2}\left[\int_{\Omega_1} f^{\omega_2}(\omega_1)\,d\mu_1(\omega_1)\right]d\mu_2(\omega_2). \tag{2.6.5}$$

The relation

$$\int_{\Omega_1\times\Omega_2} f\,d(\mu_1\otimes\mu_2) \;=\; \int_{\Omega_1}\left[\int_{\Omega_2} f_{\omega_1}(\omega_2)\,d\mu_2(\omega_2)\right]d\mu_1(\omega_1) \tag{2.6.6}$$

can be shown analogously.

Finally, let $f\colon \Omega_1\times\Omega_2 \to [0,\infty]$ be an arbitrary non-negative $(\mathcal{A}_1\otimes\mathcal{A}_2-\bar{\mathcal{B}})$-measurable function. Then there exists a sequence $(f_n)_{n\in\mathbb{N}}$ of $(\mathcal{A}_1\otimes\mathcal{A}_2)$-simple functions such that $f_n \nearrow f$. It follows from Beppo Levi's theorem (Theorem 2.4.1) that

$$\int_{\Omega_1\times\Omega_2} f_n\,d(\mu_1\otimes\mu_2) \;\nearrow\; \int_{\Omega_1\times\Omega_2} f\,d(\mu_1\otimes\mu_2). \tag{2.6.7}$$

We have seen at the end of the proof of Lemma 2.6.3 that $\int_{\Omega_1}(f_n)^{\omega_2}\,d\mu_1 \nearrow \int_{\Omega_1} f^{\omega_2}\,d\mu_1$ and $\int_{\Omega_1}(f_n)_{\omega_1}\,d\mu_2 \nearrow \int_{\Omega_1} f_{\omega_1}\,d\mu_2$. This implies, again by Beppo Levi's theorem

$$\int_{\Omega_2}\left[\int_{\Omega_1}(f_n)^{\omega_2}(\omega_1)\,d\mu_1(\omega_1)\right]d\mu_2(\omega_2) \;\nearrow\; \int_{\Omega_2}\left[\int_{\Omega_1} f^{\omega_2}(\omega_1)\,d\mu_1(\omega_1)\right]d\mu_2(\omega_2) \tag{2.6.8}$$

and

$$\int_{\Omega_1}\left[\int_{\Omega_2}(f_n)_{\omega_1}(\omega_2)\,d\mu_2(\omega_2)\right]d\mu_1(\omega_1) \;\nearrow\; \int_{\Omega_1}\left[\int_{\Omega_2} f_{\omega_1}(\omega_2)\,d\mu_2(\omega_2)\right]d\mu_1(\omega_1). \tag{2.6.9}$$

From (2.6.5), (2.6.7), and (2.6.8) we obtain (2.6.4a), and from (2.6.6), (2.6.7), and (2.6.9) we obtain (2.6.4b). □

Note that (2.6.4a) and (2.6.4b) are applicable to each non-negative $(\mathcal{A}_1\otimes\mathcal{A}_2-\bar{\mathcal{B}})$-measurable function, integrable or not. Thus one can often determine whether an $(\mathcal{A}_1\otimes\mathcal{A}_2-\bar{\mathcal{B}})$-measurable function f is integrable by using Theorem 2.6.4 to calculate $\int_{\Omega_1\times\Omega_2}|f|\,d(\mu_1\otimes\mu_2)$.

Now we turn to the second main result in this section, Fubini's theorem, which applies to $(\mathcal{A}_1\otimes\mathcal{A}_2-\bar{\mathcal{B}})$-measurable functions f that are not necessarily non-negative. Before we state and prove this theorem, we take a brief look at possible obstacles to obtain such a result. One of the relations we intend to prove is

$$\int_{\Omega_1\times\Omega_2} f\,d(\mu_1\otimes\mu_2) \;=\; \int_{\Omega_2}\left[\int_{\Omega_1} f^{\omega_2}(\omega_1)\,d\mu_1(\omega_1)\right]d\mu_2(\omega_2). \tag{2.6.10}$$

This requires that f is $(\mathcal{A}_1 \otimes \mathcal{A}_2 - \bar{\mathcal{B}})$-measurable, what we assume from here on. Provided that the integral on the left-hand side of (2.6.10) exists, we have

$$\int_{\Omega_1 \times \Omega_2} f \, d(\mu_1 \otimes \mu_2) \;=\; \int_{\Omega_1 \times \Omega_2} f^+ \, d(\mu_1 \otimes \mu_2) \;-\; \int_{\Omega_1 \times \Omega_2} f^- \, d(\mu_1 \otimes \mu_2).$$

Furthermore, it follows by Theorem 2.6.4 that

$$\int_{\Omega_1 \times \Omega_2} f^+ \, d(\mu_1 \otimes \mu_2) \;=\; \int_{\Omega_2} \left[\int_{\Omega_1} (f^+)^{\omega_2}(\omega_1) \, d\mu_1(\omega_1) \right] d\mu_2(\omega_2)$$

as well as

$$\int_{\Omega_1 \times \Omega_2} f^- \, d(\mu_1 \otimes \mu_2) \;=\; \int_{\Omega_2} \left[\int_{\Omega_1} (f^-)^{\omega_2}(\omega_1) \, d\mu_1(\omega_1) \right] d\mu_2(\omega_2).$$

Hence, relation (2.6.10) follows if

$$\begin{aligned} &\int_{\Omega_2} \left[\int_{\Omega_1} f^{\omega_2}(\omega_1) \, d\mu_1(\omega_1) \right] d\mu_2(\omega_2) \\ &= \int_{\Omega_2} \left[\int_{\Omega_1} (f^+)^{\omega_2}(\omega_1) \, d\mu_1(\omega_1) \right] d\mu_2(\omega_2) \;-\; \int_{\Omega_2} \left[\int_{\Omega_1} (f^-)^{\omega_2}(\omega_1) \, d\mu_1(\omega_1) \right] d\mu_2(\omega_2) \\ &= \int_{\Omega_2} \left[\int_{\Omega_1} (f^{\omega_2})^+ \, d\mu_1 \right] d\mu_2 \;-\; \int_{\Omega_2} \left[\int_{\Omega_1} (f^{\omega_2})^- \, d\mu_1(\omega_1) \right] d\mu_2. \end{aligned}$$

This, however, requires the inner integral $\int_{\Omega_1} f^{\omega_2} \, d\mu_1$ to be defined, i.e. at least one of the integrals $\int_{\Omega_1} (f^{\omega_2})^+ \, d\mu_1$ and $\int_{\Omega_1} (f^{\omega_2})^- \, d\mu_1$ should be finite. This is not guaranteed for all $\omega_2 \in \Omega_2$, however, if

$$\int_{\Omega_1 \times \Omega_2} |f| \, d(\mu_1 \otimes \mu_2) \;<\; \infty,$$

then we obtain from $\int_{\Omega_1 \times \Omega_2} |f| \, d(\mu_1 \otimes \mu_2) \;=\; \int_{\Omega_2} \left[\int_{\Omega_1} |f|^{\omega_2} \, d\mu_1 \right] d\mu_2$ that

$$\mu_2\Big(\underbrace{\Big\{ \omega_2 \in \Omega_2 \colon \quad \int_{\Omega_1} |f|^{\omega_2} \, d\mu_1 = \infty \Big\}}_{=: \Omega_{2,0}} \Big) \;=\; 0.$$

A similar argument shows that

$$\mu_1\Big(\underbrace{\Big\{ \omega_1 \in \Omega_1 \colon \quad \int_{\Omega_2} |f|_{\omega_1} \, d\mu_2 = \infty \Big\}}_{=: \Omega_{1,0}} \Big) \;=\; 0.$$

Since $(\mu_1 \otimes \mu_2)\big(\Omega_1 \times \Omega_{2,0}\big) = \mu_1\big(\Omega_1\big) \cdot \mu_2\big(\Omega_{2,0}\big) = 0$ and $(\mu_1 \otimes \mu_2)\big(\Omega_{1,0} \times \Omega_2\big) = \mu_1\big(\Omega_{1,0}\big) \cdot \mu_2\big(\Omega_2\big) = 0$ we obtain that

$$\int_{\Omega_1 \times \Omega_2} f \, d(\mu_1 \otimes \mu_2) \;=\; \int_{\Omega_1 \times (\Omega_2 \setminus \Omega_{2,0})} f \, d(\mu_1 \otimes \mu_2) \;=\; \int_{(\Omega_1 \setminus \Omega_{1,0}) \times \Omega_2} f \, d(\mu_1 \otimes \mu_2),$$

which leads to the following theorem.

Theorem 2.6.5. (Fubini's theorem)
Let $(\Omega_1, \mathcal{A}_1, \mu_1)$ *and* $(\Omega_2, \mathcal{A}_2, \mu_2)$ *be* σ*-finite measure spaces, and let* $f\colon \Omega_1 \times \Omega_2 \to [0,\infty]$ *be an* $(\mathcal{A}_1 \otimes \mathcal{A}_2 - \bar{\mathcal{B}})$*-measurable and* $\mu_1 \otimes \mu_2$*-integrable function. Then*

$$\int_{\Omega_1\times\Omega_2} f\, d(\mu_1 \otimes \mu_2) \;=\; \int_{\Omega_2\setminus\Omega_{2,0}} \left[\int_{\Omega_1} f^{\omega_2}(\omega_1)\, d\mu_1(\omega_1)\right] d\mu_2(\omega_2)$$
$$=\; \int_{\Omega_1\setminus\Omega_{1,0}} \left[\int_{\Omega_2} f_{\omega_1}(\omega_2)\, d\mu_2(\omega_2)\right] d\mu_1(\omega_1),$$

where $\Omega_{1,0} = \left\{\omega_1 \in \Omega_1\colon\; \int_{\Omega_2} |f|_{\omega_1}\, d\mu_2 = \infty\right\}$ *and* $\Omega_{2,0} = \left\{\omega_2 \in \Omega_2\colon\; \int_{\Omega_1} |f|^{\omega_2}\, d\mu_1 = \infty\right\}$*.*

Note that functions that are equal almost everywhere have the same integral. Therefore, the theory of integration can be extended to functions that are defined almost everywhere. Using such a convention the above iterated integrals can also be written as $\int_{\Omega_2}\left[\int_{\Omega_1} f^{\omega_2}(\omega_1)\, d\mu_1(\omega_1)\right] d\mu_2(\omega_2)$ and $\int_{\Omega_1}\left[\int_{\Omega_2} f_{\omega_1}(\omega_2)\, d\mu_2(\omega_2)\right] d\mu_1(\omega_1)$.

Application of Fubini's theorem usually follows a two-step procedure that parallels its proof. Note that it follows from $(\mathcal{A}_1 \otimes \mathcal{A}_2 - \bar{\mathcal{B}})$-measurability of f that $|f|$ is also $(\mathcal{A}_1 \otimes \mathcal{A}_2 - \bar{\mathcal{B}})$-measurable. Hence, Tonelli's theorem is applicable and it holds that $\int_{\Omega_1\times\Omega_2} |f|\, d(\mu_1\otimes\mu_2) = \int_{\Omega_2}\left[\int_{\Omega_1} |f|^{\omega_2}\, d\mu_1\right] d\mu_2 = \int_{\Omega_1}\left[\int_{\Omega_2} |f|_{\omega_1}\, d\mu_2\right] d\mu_1$. Usually one of these iterated integrals is computed (or estimated above). If the result is finite, then the double integral (integral with respect to $\mu_1 \otimes \mu_2$) of $|f|$ must be finite, so that f is integrable with respect to $\mu_1 \otimes \mu_2$; then the value of the double integral of f is found by computing one of the iterated integrals of f. If the integral of $|f|$ is infinite, f is not $\mu_1 \otimes \mu_2$-integrable.

Now we turn to an important application, to the so-called convolution of probability measures. Suppose that X_1 and X_2 are independent $\mathbb{R}^d$-valued random vectors that are defined on a common probability space $(\Omega, \mathcal{A}, P)$. Then P^{X_i} denotes the distribution of X_i under the probability measure P, i.e.

$$P^{X_i}(B) \;=\; P\big(\{\omega \in \Omega\colon \quad X_i(\omega) \in B\}\big) \qquad \forall B \in \mathcal{B}^d.$$

The **convolution** $P^{X_1} * P^{X_2}$ of P^{X_1} and P^{X_2} is defined as the distribution of $X_1 + X_2$ under P, i.e.

$$\big(P^{X_1} * P^{X_2}\big)(B) \;=\; P\big(\{\omega \in \Omega\colon \quad X_1(\omega) + X_2(\omega) \in B\}\big) \qquad \forall B \in \mathcal{B}^d.$$

In what follows we derive formulas which allow an easy explicit computation of $P^{X_1} * P^{X_2}$. Since X_1 and X_2 are $(\mathcal{A} - \mathcal{B}^d)$-measurable it follows from Corollary 2.2.3 that the random vector $X = \binom{X_1}{X_2}$ is $(\mathcal{A} - \mathcal{B}^{2d})$-measurable. Since X_1 and X_2 are assumed to be independent under P we obtain, for arbitrary Borel sets B_1 and B_2,

$$P^X\big(B_1 \times B_2\big) \;=\; P\big(\{\omega\colon \quad X_1(\omega) \in B_1 \;\text{ and }\; X_2(\omega) \in B_2\}\big) \;=\; P^{X_1}\big(B_1\big) \cdot P^{X_2}\big(B_2\big).$$

Hence, P^X is the product of P^{X_1} and P^{X_2} and it follows from Theorem 2.6.2 that

$$P^X(B) \;=\; \big(P^{X_1} \otimes P^{X_2}\big)(B) \;=\; \int_{\mathbb{R}^d} P^{X_1}(B^{x_2})\, dP^{X_2}(x_2) \qquad \forall B \in \mathcal{B}^{2d}.$$

(Note that, according to the discussion at the beginning of Section 2.6, $\mathcal{B}^d \otimes \mathcal{B}^d = \mathcal{B}^{2d}$.)

Now we consider, for arbitrary $C \in \mathcal{B}^d$, the set $B := \left\{\binom{x_1}{x_2} \colon x_1 + x_2 \in C\right\}$. Then $B \in \mathcal{B}^{2d}$ and $B^{x_2} = \{x_1 \colon x_1 + x_2 \in C\} = C - x_2$. If the probability measures P^{X_1} and P^{X_2} have respective densities p^{X_1} and p^{X_2} with respect to Lebesgue measure λ^d, then

$$P^{X_1+X_2}\big(C\big) \;=\; P^X\big(B\big) \;=\; \int_{\mathbb{R}^d} \left[\int_{C-x_2} p^{X_1}(x_1)\, d\lambda^d(x_1) \right] p^{X_2}(x_2)\, d\lambda^d(x_2).$$

Consider the mapping $x \mapsto T(x) := x - x_2$. Since Lebesgue measure is translation-invariant we have that $\lambda^d = (\lambda^d)^T$, and we obtain by Proposition 2.3.6 that

$$\begin{aligned}
\int_{C-x_2} p^{X_1}(x_1)\, d\lambda^d(x_1) \;&=\; \int_{\mathbb{R}^d} \mathbb{1}_C(x_1 + x_2) p^{X_1}(x_1)\, d\lambda^d(x_1) \\
&=\; \int_{\mathbb{R}^d} \mathbb{1}_C(x_1 + x_2) p^{X_1}(x_1)\, d(\lambda^d)^T(x_1) \\
&=\; \int_{\mathbb{R}^d} \underbrace{\mathbb{1}_C \circ T(x_1 + x_2)}_{= \mathbb{1}_C(x_1)} \underbrace{p^{X_1} \circ T(x_1)}_{= p^{X_1}(x_1 - x_2)}\, d\lambda^d(x_1) \;=\; \int_C p^{X_1}(x_1 - x_2)\, d\lambda^d(x_1).
\end{aligned}$$

Since $\binom{x_1}{x_2} \mapsto p^{X_1}(x_1 - x_2) p^{X_2}(x_2)$ is a non-negative and $(\mathcal{B}^{2d} - \mathcal{B}^d)$-measurable function, Tonelli's theorem implies that

$$\begin{aligned}
P^{X_1+X_2}\big(C\big) \;&=\; \int_{\mathbb{R}^d} \left[\int_C p^{X_1}(x_1 - x_2)\, d\lambda^d(x_1) \right] p^{X_2}(x_2)\, d\lambda^d(x_2) \\
&=\; \int_{\mathbb{R}^d} \left[\int_C p^{X_1}(x_1 - x_2) p^{X_2}(x_2)\, d\lambda^d(x_1) \right] d\lambda^d(x_2) \\
&=\; \int_C \left[\int_{\mathbb{R}^d} p^{X_1}(x_1 - x_2) p^{X_2}(x_2)\, d\lambda^d(x_2) \right] d\lambda^d(x_1) \qquad \forall C \in \mathcal{B}^d.
\end{aligned}$$

Hence, $P^{X_1+X_2}$ has a density $p^{X_1+X_2}$ w.r.t. Lebesgue measure λ^d and

$$p^{X_1+X_2}(x) \;=\; \int_{\mathbb{R}^d} p^{X_1}(x - y) p^{X_2}(y)\, d\lambda^d(y) \qquad \lambda^d\text{-almost everywhere.}$$

2.7 Existence of densities - the Radon-Nikodym theorem

Suppose that μ is a measure on a measurable space $(\Omega, \mathcal{A})$ and that $f\colon \Omega \to [0,\infty]$ is an $(\mathcal{A}-\bar{\mathcal{B}})$-measurable function. Then it follows from part (i) of Theorem 2.4.3 that the set function $\nu\colon \mathcal{A} \to [0,\infty]$ defined by

$$\nu\big(A\big) \;=\; \int_A f\, d\mu \qquad \forall A \in \mathcal{A}$$

is also a measure on $(\Omega, \mathcal{A})$. The function f is said to be a **density** of ν with respect to μ. It follows from the definition of ν that $\mu(A) = 0$ for a set $A \in \mathcal{A}$ implies that $\nu(A) = 0$, i.e. the measure ν is **absolutely continuous** with respect to μ, and we write $\nu \ll \mu$. In what follows we prove the important result that the converse statement is also true: if μ is a σ-finite measure and $\nu \ll \mu$, then the measure ν has a density w.r.t μ. First we consider the case where the measures ν and μ are finite.

Proposition 2.7.1. *Let μ and ν be finite measures on a measure space $(\Omega, \mathcal{A})$ such that $\nu \ll \mu$. Then there exist an $(\mathcal{A} - \mathcal{B})$-measurable function $f\colon \Omega \to [0,\infty)$, a* ***density****, such that*

$$\nu\big(A\big) \;=\; \int_A f\, d\mu \qquad \forall A \in \mathcal{A}.$$

For two such densities f_1 and f_2,

$$\mu\big(\{\omega\colon\; f_1(\omega) \neq f_2(\omega)\}\big) \;=\; 0.$$

Proof. The density f will be obtained by an "approximation from below". Let

$$\mathcal{G} \;:=\; \Big\{g\colon\; \Omega \to [0,\infty) \Big|\, g \text{ is } (\mathcal{A}-\mathcal{B})\text{-measurable and } \int_A g\, d\mu \le \nu(A) \;\; \forall A \in \mathcal{A}\Big\}$$

be the collection of "candidate functions". The function f we are seeking will be a "greatest element" in $\mathcal{G}$. To identify such an element, we focus on $\alpha := \sup\big\{\, \int_\Omega g\, d\mu\colon\, g \in \mathcal{G}\big\}$. Then there exists a sequence $(g_n)_{n\in\mathbb{N}}$ of functions in $\mathcal{G}$ such that

$$\int_\Omega g_n\, d\mu \underset{n\to\infty}{\longrightarrow} \alpha.$$

Note that if g and g' lie in $\mathcal{G}$, then $\max\{g,g'\}$ lies in $\mathcal{G}$ as well. Indeed, $\max\{g,g'\}$ is also $(\mathcal{A}-\mathcal{B})$-measurable and it holds that

$$\begin{aligned}\int_A \max\{g,g'\}\, d\mu \;&=\; \int_{A\cap\{\omega\colon\, g(\omega)\ge g'(\omega)\}} g\, d\mu \,+\, \int_{A\cap\{\omega\colon\, g(\omega)<g'(\omega)\}} g'\, d\mu \\ &\le\; \nu\big(A\cap\{\omega\colon\, g(\omega)\ge g'(\omega)\}\big) \,+\, \nu\big(A\cap\{\omega\colon\, g(\omega)< g'(\omega)\}\big) \;=\; \nu\big(A\big).\end{aligned}$$

Let $g_n' := \max\{g_1,\dots,g_n\}$. Then g_n' belongs to $\mathcal{G}$. Since $(g_n')_{n\in\mathbb{N}}$ is a non-decreasing sequence of $(\mathcal{A}-\mathcal{B})$-measurable functions there exists an $(\mathcal{A}-\bar{\mathcal{B}})$-measurable function $g\colon \Omega \to [0,\infty]$ such that $g_n' \nearrow g$. It follows from Beppo Levi's theorem (Theorem 2.4.1) that

$$\int_A g\, d\mu \;=\; \lim_{n\to\infty}\int_A g_n'\, d\mu \;\le\; \nu\big(A\big) \qquad \forall A \in \mathcal{A}. \tag{2.7.1}$$

If

$$\int_\Omega g\,d\mu \;=\; \nu(\Omega), \tag{2.7.2}$$

then we also obtain that

$$\int_A g\,d\mu \;=\; \int_\Omega g\,d\mu \;-\; \int_{A^c} g\,d\mu \;\geq\; \nu(\Omega) \;-\; \nu(A^c) \;=\; \nu(A). \tag{2.7.3}$$

(2.7.1) and (2.7.3) imply that

$$\int_A g\,d\mu \;=\; \nu(A) \qquad \forall A \in \mathcal{A}.$$

We postpone the verification of relation (2.7.2) to the end of this proof and proceed directly with the remaining steps. It can still be the case that g attains the value ∞. However,

$$\int_{\{\omega\colon\, g(\omega)=\infty\}} g\,d\mu \;=\; \infty \cdot \mu\big(\{\omega\colon\; g(\omega) = \infty\}\big).$$

Since ν is a finite measure we conclude that $\mu(\{\omega\colon\, g(\omega) = \infty\}) = 0$. Hence, the function $f\colon \Omega \to [0,\infty)$ defined by

$$f(\omega) \;:=\; \begin{cases} g(\omega) & \text{if } g(\omega) < \infty, \\ 0 & \text{if } g(\omega) = \infty \end{cases}$$

is $(\mathcal{A}-\mathcal{B})$-measurable and satisfies

$$\int_A f\,d\mu \;=\; \nu(A) \qquad \forall A \in \mathcal{A}.$$

Hence, f it is a density of ν w.r.t. μ.

If f_1 and f_2 are two such densities, then

$$\int_{\{\omega\colon\, f_1(\omega)>f_2(\omega)\}} (f_1 - f_2)\,d\mu \;=\; \underbrace{\int_{\{\omega\colon\, f_1(\omega)>f_2(\omega)\}} f_1\,d\mu}_{=\nu(\{\omega\colon\, f_1(\omega)>f_2(\omega)\})} \;-\; \underbrace{\int_{\{\omega\colon\, f_1(\omega)>f_2(\omega)\}} f_2\,d\mu}_{=\nu(\{\omega\colon\, f_1(\omega)>f_2(\omega)\})} \;=\; 0,$$

which implies by (iv) of Proposition 2.3.2 that

$$\mu\big(\{\omega\colon\; f_1(\omega) > f_2(\omega)\}\big) \;=\; 0.$$

The relation $\mu\big(\{\omega\colon\; f_1(\omega) < f_2(\omega)\}\big) = 0$ can be proved analogously which shows that two densities f_1 and f_2 are equal up to a μ-null set.

Now we turn to the proof of (2.7.2). Assume the contrary, i.e.

$$\int_\Omega g\,d\mu \;<\; \nu(\Omega). \tag{2.7.4}$$

The set function $\rho\colon \mathcal{A} \to [0,\infty)$ defined by $\rho(A) = \nu(A) - \int_A g\,d\mu$ is a measure on $(\Omega,\mathcal{A})$. Since μ is a finite measure there exists some $\beta > 0$ such that

$$\nu(\Omega) \;-\; \int_\Omega g\,d\mu \;-\; \beta\,\mu(\Omega) \;>\; 0.$$

It follows from Lemma 2.7.2 below that there exists some $\Omega_+ \in \mathcal{A}$ such that

$$\nu(\Omega_+) - \int_{\Omega_+} g \, d\mu - \beta \, \mu(\Omega_+) \geq \nu(\Omega) - \int_{\Omega} g \, d\mu - \beta \, \mu(\Omega)$$

and

$$\nu(A) \geq \int_A g \, d\mu + \beta \, \mu(A) \geq 0 \qquad \forall A \in \mathcal{A} \cap \Omega_+.$$

Let

$$\widetilde{g}(\omega) := g(\omega) + \beta \cdot \mathbb{1}_{\Omega_+}(\omega).$$

Then, for $A \in \mathcal{A}$,

$$\begin{aligned} \int_A \widetilde{g} \, d\mu &= \int_A g \, d\mu + \beta \int_A \mathbb{1}_{\Omega_+} \, d\mu \\ &= \int_{A \cap \Omega_+} g \, d\mu + \beta \, \mu(A \cap \Omega_+) + \int_{A \cap \Omega_+^c} g \, d\mu \\ &\leq \nu(A \cap \Omega_+) + \nu(A \cap \Omega_+^c) = \nu(A). \end{aligned}$$

Hence, $\widetilde{g} \in \mathcal{G}$. On the other hand,

$$\int_\Omega \widetilde{g} \, d\mu = \int_\Omega g \, d\mu + \beta \, \mu(\Omega_+) = \alpha + \beta \, \mu(\Omega_+).$$

Since $\nu \ll \mu$ we also have $\rho \ll \mu$ and $\rho(\Omega_+) > 0$ implies $\mu(\Omega_+) > 0$, and so $\int_\Omega \widetilde{g} \, d\mu > \alpha$, which leads to a contradiction. Hence, our assumption (2.7.4) is wrong and (2.7.2) holds true. The proof is therefore complete. □

In the course of the proof of Proposition 2.7.1 we used an auxiliary result which will be stated and proved now.

Lemma 2.7.2. *Let σ and τ be finite measures on a measurable space $(\Omega, \mathcal{A})$ and let $\rho := \sigma - \tau$. (ρ is a so-called* **finite signed measure** *(a.k.a.* **real measure***) on $(\Omega, \mathcal{A})$.) Then there exists a set $\Omega_+ \in \mathcal{A}$ such that*

(i) $\rho(\Omega_+) \geq \rho(\Omega)$,

(ii) $\rho(A) \geq 0 \qquad \forall A \in \mathcal{A} \cap \Omega_+ := \{A \cap \Omega_+ : \ A \in \mathcal{A}\}$.

Proof. The idea of the proof is not far to seek: We repeatedly cut out sets which violate (ii). When doing so, we obtain sets $\Omega_n \in \mathcal{A}$, $\Omega =: \Omega_0 \supseteq \Omega_1 \supseteq \Omega_2 \supseteq \ldots$, such that

$$\rho(\Omega_n) \geq \rho(\Omega_{n-1}) \tag{2.7.5}$$

and

$$\rho(A) \geq -\frac{1}{n} \qquad \forall A \in \mathcal{A} \cap \Omega_n. \tag{2.7.6}$$

Then we define

$$\Omega_+ := \bigcap_{n=1}^{\infty} \Omega_n.$$

For each set $A \in \mathcal{A} \cap \Omega_+$ we also have that $A \in \mathcal{A} \cap \Omega_n\ \forall n \in \mathbb{N}$, and so

$$\rho\big(A\big) \geq 0.$$

Furthermore, it follows from continuity from above that

$$\rho\big(\Omega_+\big) = \sigma\big(\Omega_+\big) - \tau\big(\Omega_+\big) = \lim_{n\to\infty} \sigma\big(\Omega_n\big) - \lim_{n\to\infty} \tau\big(\Omega_n\big) = \lim_{n\to\infty} \rho\big(\Omega_n\big) \geq \rho\big(\Omega\big).$$

Now we describe how the sets Ω_n can be found. Suppose that sets $\Omega_1, \ldots, \Omega_{n-1}$ satisfying (2.7.5) and (2.7.6) exist. Let $\Omega_{n,0} := \Omega_{n-1}$. If

$$\rho\big(A\big) \geq -\frac{1}{n} \qquad \forall A \in \mathcal{A} \cap \Omega_{n,0},$$

then we choose $\Omega_n = \Omega_{n,0}$. Otherwise, there exists a set $A_{n,1} \in \mathcal{A} \cap \Omega_{n,0}$ such that $\rho(A_{n,1}) < -1/n$ and we define $\Omega_{n,1} := \Omega_{n,0} \setminus A_{n,1}$. In this case, it follows that

$$\rho\big(\Omega_{n,1}\big) = \rho\big(\Omega_{n,0}\big) - \rho\big(A_{n,1}\big) > \rho\big(\Omega_{n-1}\big) + \frac{1}{n}.$$

Now we proceed as before, with $\Omega_{n,1}$ in place of $\Omega_{n,0}$: If $\rho(A) \geq -1/n$ for all $A \in \mathcal{A} \cap \Omega_{n,1}$, then we choose $\Omega_n = \Omega_{n,1}$. Otherwise, there exists a set $A_{n,2} \in \mathcal{A} \cap \Omega_{n,1}$ such that $\rho(A_{n,2}) < -1/n$. We cut out this set and define $\Omega_{n,2} = \Omega_{n,1} \setminus A_{n,2}$. Then

$$\rho\big(\Omega_{n,2}\big) = \rho\big(\Omega_{n,1}\big) - \rho\big(A_{n,2}\big) > \rho\big(\Omega_{n-1}\big) + \frac{2}{n}.$$

After cutting out k sets $A_{n,1}, \ldots, A_{n,k}$ we obtain a set $\Omega_{n,k}$ and it holds that

$$\rho\big(\Omega_{n,k}\big) = \rho\big(\Omega_{n,k-1}\big) - \rho\big(A_{n,k}\big) > \rho\big(\Omega_{n-1}\big) + \frac{k}{n}.$$

Since the measure σ is finite, this process will end after a finite number k_n of steps, where $k_n \leq \sigma(\Omega)/n$. We define $\Omega_n := \Omega_{n,k_n}$. It follows that (2.7.5) and (2.7.6) are fulfilled. □

Now we generalize Proposition 2.7.1 and turn to the main result of this section. This theorem is named after the Austrian mathematician Johann Radon, who proved the theorem for the special case where the underlying space is $\mathbb{R}^n$ in 1913, and after the Polish mathematician Otto Marcin Nikodym who proved the general case in 1930.

Theorem 2.7.3. (Radon-Nikodym theorem)
Let $(\Omega, \mathcal{A})$ be a measurable space, and let ν be an arbitrary and μ be a σ-finite measure on $(\Omega, \mathcal{A})$. If ν is absolutely continuous w.r.t. μ, then there exists an $(\mathcal{A} - \bar{\mathcal{B}})$-measurable function $f\colon \Omega \to [0, \infty]$ such that

$$\nu\big(A\big) = \int_A f\, d\mu \qquad \forall A \in \mathcal{A}.$$

The density f is called the **Radon-Nikodym derivative** *of ν with respect to μ and is often denoted $d\nu/d\mu$.*

For two such densities f_1 and f_2,

$$\mu\big(\{\omega\colon\ f_1(\omega) \neq f_2(\omega)\}\big) = 0.$$

Proof. The proof is split into two steps.

(i) (μ *finite*)
First consider the case where ν is an arbitrary and μ a **finite** measure. If $\nu(\Omega) < \infty$, then the existence of real-valued density f follows from Proposition 2.7.1. It remains to consider the case of $\nu(\Omega) = \infty$. We show that there exist disjoint sets $\Omega_0, \Omega_1, \Omega_2, \ldots \in \mathcal{A}$ such that

$$\nu(\Omega_n) \,<\, \infty \qquad \forall n \in \mathbb{N} \tag{2.7.7}$$

and, for $\Omega_0 := \Omega \setminus \big(\bigcup_{n=1}^{\infty} \Omega_n\big)$: if $A \in \mathcal{A} \cap \Omega_0$, then either

$$\mu\big(A\big) \,=\, \nu\big(A\big) \,=\, 0 \qquad \text{or} \qquad \mu\big(A\big) > 0,\ \nu\big(A\big) = \infty. \tag{2.7.8}$$

On each of these subsets, we find a clear guideline how f has to be chosen.

The following collection of sets contains candidates for the sets $\Omega_1, \Omega_2, \ldots$:

$$\mathcal{Q} \,:=\, \big\{A \in \mathcal{A}\colon\ \nu(A) < \infty\big\}.$$

To exhaust the part of Ω where ν is σ-finite, we choose a sequence of sets $(A_n)_{n\in\mathbb{N}}$ from $\mathcal{Q}$ such that

$$\mu\big(A_n\big) \underset{n\to\infty}{\longrightarrow} \alpha \,:=\, \sup\big\{\mu(A)\colon\ A \in \mathcal{Q}\big\}.$$

We define

$$\Omega_1 \,=\, A_1, \quad \Omega_n \,=\, A_n \setminus (A_1 \cup \cdots \cup A_{n-1}) \ \ (n \geq 2), \quad \Omega_0 \,=\, \Omega \setminus \bigcup_{n=1}^{\infty} \Omega_n.$$

It holds that

$$\nu\big(\Omega_n\big) \,<\, \infty \qquad \forall n \in \mathbb{N},$$

i.e. (2.7.7) is fulfilled.

It remains to check that (2.7.8) is satisfied. Let $A \in \mathcal{A} \cap \Omega_0$ be arbitrary. If $\nu(A) = \infty$, then it follows from $\nu \ll \mu$ that $\mu(A) > 0$. Otherwise, if $\nu(A) < \infty$, we have that $A \cup A_n \in \mathcal{Q}$ for all $n \in \mathbb{N}$, and since A and A_n are disjoint,

$$\alpha \,\geq\, \mu\big(A \cup A_n\big) \,=\, \mu\big(A\big) \,+\, \underbrace{\mu\big(A_n\big)}_{\underset{n\to\infty}{\longrightarrow}\ \alpha},$$

which implies that $\mu(A) = 0$. Again from $\nu \ll \mu$, we obtain that $\nu(A) = 0$. Therefore, (2.7.8) is also fulfilled and the sets $\Omega_0, \Omega_1, \Omega_2, \ldots$ are as required.

For each $n \in \mathbb{N}$, it follows from Proposition 2.7.1 that there exists an $(\mathcal{A}\cap\Omega_n - \mathcal{B})$-measurable function $f_n\colon \Omega_n \to [0, \infty)$ such that

$$\nu\big(A\big) \,=\, \int_A f_n\, d\mu \qquad \forall A \in \mathcal{A} \cap \Omega_n.$$

Furthermore, for two such functions $f_{n,1}$ and $f_{n,2}$,

$$\mu\big(\{\omega\colon\ f_{n,1}(\omega) \neq f_{n,2}(\omega)\}\big) \,=\, 0.$$

For $A \in \mathcal{A} \cap \Omega_0$, it follows from

$$\nu(A) = \begin{cases} \infty & \text{if } \mu(A) > 0, \\ 0 & \text{if } \mu(A) = 0 \end{cases}$$

that f_0 given by $f_0(\omega) = \infty$ for all $\omega \in \Omega_0$ is such that

$$\nu(A) = \int_A f_0 \, d\mu \qquad \forall A \in \mathcal{A} \cap \Omega_0.$$

Moreover, if $\mu(\{\omega \in \Omega_0 \colon f_0'(\omega) < \infty\}) > 0$, then there exists some $c < \infty$ such that $\mu(\{\omega \in \Omega_0 \colon f_0'(\omega) \leq c\}) > 0$. Then $\nu(\{\omega \in \Omega_0 \colon f_0'(\omega) \leq c\}) = \infty$. On the other hand, $\int_{\{\omega \in \Omega_0 \colon f_0'(\omega) \leq c\}} f_0' \, d\mu \leq c\mu(\Omega) < \infty$, which shows that f_0' cannot be a density of ν w.r.t. μ on $\mathcal{A} \cap \Omega_0$.

To summarize, the function

$$f := \sum_{n=0}^{\infty} \mathbb{1}_{\Omega_n} f_n$$

is $(\mathcal{A} - \bar{\mathcal{B}})$-measurable and it holds that

$$\begin{aligned} \nu(A) &= \sum_{n=0}^{\infty} \nu(A \cap \Omega_n) = \sum_{n=0}^{\infty} \int_{A \cap \Omega_n} f_n \, d\mu \\ &= \sum_{n=0}^{\infty} \int_A \mathbb{1}_{\Omega_n} f_n \, d\mu = \int_A f \, d\mu. \end{aligned}$$

Therefore, f is the seeked density of ν w.r.t. μ. If f_1 and f_2 are two such densities, then it follows from the considerations above that $\mu(\{\omega \in \Omega_n \colon f_1(\omega) \neq f_2(\omega)\}) = 0$ for all $n \geq 0$, which implies that these two densities coincide up to a set of μ-measure 0.

(ii) (*μ σ-finite*)
It remains consider the case where μ is σ-finite but not finite. Then there exist pairwise disjoint sets $\Omega_1', \Omega_2', \ldots \in \mathcal{A}$ such that $\bigcup_{n=1}^{\infty} \Omega_n' = \Omega$ and

$$\mu(\Omega_n') < \infty \qquad \forall n \in \mathbb{N}.$$

For each $n \in \mathbb{N}$, it follows from part (i) of this proof that there exists an $(\mathcal{A} \cap \Omega_n' - \bar{\mathcal{B}})$-measurable function $f_n' \colon \Omega_n' \to [0, \infty]$ such that

$$\nu(A) = \int_A f_n' \, d\mu \qquad \forall A \in \mathcal{A} \cap \Omega_n'.$$

Moreover, for two such densities $f_{n,1}'$ and $f_{n,2}'$,

$$\mu(\{\omega \in \Omega_n' \colon f_{n,1}'(\omega) \neq f_{n,2}'(\omega)\}) = 0.$$

The function $f \colon \Omega \to [0, \infty]$ defined by $f(\omega) = \sum_{n=1}^{\infty} \mathbb{1}_{\Omega_n'}(\omega) f_n'(\omega) \quad \forall \omega \in \Omega$ is $(\mathcal{A} - \bar{\mathcal{B}})$-measurable and satisfies, for each $A \in \mathcal{A}$,

$$\begin{aligned} \nu(A) &= \sum_{n=1}^{\infty} \nu(A \cap \Omega_n') = \sum_{n=1}^{\infty} \int_{A \cap \Omega_n'} f_n' \, d\mu \\ &= \sum_{n=1}^{\infty} \int_A \mathbb{1}_{\Omega_n'} f_n' \, d\mu = \int_A f \, d\mu. \end{aligned}$$

Finally, for two such densities f_1 and f_2,

$$\mu\big(\{\omega \in \Omega\colon\ f_1(\omega) \neq f_2(\omega)\}\big) \;=\; \sum_{n=1}^{\infty} \mu\big(\{\omega \in \Omega'_n\colon\ f_1(\omega) \neq f_2(\omega)\}\big) \;=\; 0.$$

□

2.8 An application in probability theory: Conditional distributions

In this section we extend the elementary concept of conditional probability which is usually taught in basic courses on probability theory. Let $(\Omega, \mathcal{A}, P)$ be a probability space, and let A and B be events, i.e. sets belonging to $\mathcal{A}$. If $P(B) > 0$, then

$$P\big(A \mid B\big) \;:=\; \frac{P(A \cap B)}{P(B)}$$

is the conditional probability of A given B. If $P(B) = 0$, then $P(A \mid B)$ is undefined or is simply set to 0. Likewise, if $(\Omega_X, \mathcal{A}_X)$ and $(\Omega_Y, \mathcal{A}_Y)$ are two measure spaces, and if $X\colon\ \Omega \to \Omega_X$ and $Y\colon\ \Omega \to \Omega_Y$ are $(\mathcal{A}{-}\mathcal{A}_X)$- respectively $(\mathcal{A}{-}\mathcal{A}_Y)$-measurable mappings (random variables), then for $C \in \mathcal{A}_X$,

$$P\big(X \in C \mid Y = y\big) \;:=\; \frac{P(X \in C,\ Y = y)}{P(Y = y)}$$

is the conditional probability of the event that $\{\omega\colon X(\omega) \in C\}$ given $\{\omega\colon Y(\omega) = y\}$, provided that the set $\{\omega\colon Y(\omega) = y\}$ is measurable and the probability of this event is positive. If $P(\{\omega\colon Y(\omega) = y\}) = 0$, then $P\big(X \in C \mid Y = y\big)$ is undefined or set to 0. This definition is sufficient if Ω_Y is a finite or countably infinite set. Let $N_Y := \{y \in \Omega_Y\colon\ P(Y = y) = 0\}$ be that subset of Ω_Y on which $P\big(X \in C \mid Y = y\big)$ is not given a meaningful definition. As usual in case of a finite or countably infinite set Ω_Y, let $\mathcal{A}_Y$ be chosen as the power set 2^{Ω_Y} of Ω_Y. Since

$$P\big(\{\omega\colon\ Y(\omega) \in N_Y\}\big) \;=\; \sum_{y \in N_Y} P(Y = y) \;=\; 0$$

we see that we have with probability 1 a meaningful definition of the conditional probabilities $P\big(X \in C \mid Y = y\big)$. On the other hand, if the random variable Y we condition on has a continuous distribution, for example $Y \sim \mathcal{N}(0,1)$, then $P(Y = y) = 0$ for all $y \in \Omega_Y$. This means that the above definition of conditional probabilities does not really help; in fact, we do not have a meaningful definition for all $y \in \Omega_Y$.

In what follows we want to give a meaning to the expression $P\big(X \in C \mid Y = y\big)$, even for cases where $P(Y = y) = 0$ for all $y \in \Omega_Y$. Before we present an improved definition of conditional probability, we stick once more to the case of a discrete random variable Y and derive a relation which guides us to our intended definition.

Lemma 2.8.1. *Let* $(\Omega, \mathcal{A}, P)$ *be a probability space, and let* $(\Omega_X, \mathcal{A}_X)$ *and* $(\Omega_Y, \mathcal{A}_Y)$ *be two measurable spaces, where* Ω_Y *is finite or countably infinite and* $\mathcal{A}_Y = 2^{\Omega_Y}$*. Suppose that* $X\colon \Omega \to \Omega_X$ *and* $Y\colon \Omega \to \Omega_Y$ *are* $(\mathcal{A} - \mathcal{A}_X)$*- respectively* $(\mathcal{A} - \mathcal{A}_Y)$*-measurable mappings,* $\{\omega\colon Y(\omega) = y\} \in \mathcal{A}\ \forall y \in \Omega_Y$*, and let* $C \in \mathcal{A}_X$*.*

Let $\mu_C\colon \Omega_Y \to [0,1]$ *be such that*

$$P\big(X \in C,\ Y \in D\big) = \sum_{y \in D} \mu_C\big(y\big) P\big(Y = y\big) \qquad \forall D \in \mathcal{A}_Y. \tag{2.8.1}$$

Then

$$P^Y\big(\{y \in \Omega_Y\colon\ \ \mu_C(y) \neq P(X \in C \mid Y = y)\}\big) = 0. \tag{2.8.2}$$

Proof. Let $y \in \Omega_Y$ be such that $P(Y = y) > 0$. Choosing $D = \{y\}$ in (2.8.1) we obtain that

$$P\big(X \in C,\ Y = y\big) = \mu_C\big(y\big) P\big(Y = y\big),$$

which implies that

$$\mu_C\big(y\big) = P\big(X \in C \big| Y = y\big).$$

Therefore, $\mu_C(y) \neq P(X \in C \mid Y = y)$ implies $P(Y = y) = 0$, and we obtain that

$$\begin{aligned} P^Y\big(\{y \in \Omega_Y\colon\ \ \mu_C(y) \neq P(X \in C \mid Y = y)\}\big) &\leq P^Y\big(\{y \in \Omega_Y\colon\ P(Y = y) = 0\}\big) \\ &= \sum_{y\colon P(Y=y)=0} P^Y\big(\{y\}\big) = 0. \end{aligned}$$

□

We have seen that (2.8.1) provides an equivalent definition of (elementary) conditional probability. This observation suggests an extension of this concept which covers practically all cases of interest.

Theorem 2.8.2. *Let* $(\Omega, \mathcal{A}, P)$ *be a probability space, and let* $(\Omega_X, \mathcal{A}_X)$ *and* $(\Omega_Y, \mathcal{A}_Y)$ *be two measurable spaces. Suppose that* $X\colon \Omega \to \Omega_X$ *and* $Y\colon \Omega \to \Omega_Y$ *are* $(\mathcal{A} - \mathcal{A}_X)$*-respectively* $(\mathcal{A} - \mathcal{A}_Y)$*-measurable mappings, and let* $C \in \mathcal{A}_X$*.*

Then there exists an $(\mathcal{A}_Y - \mathcal{B})$*-measurable function* $\mu_C\colon \Omega_Y \to [0,1]$ *such that*

$$P\big(X \in C,\ Y \in D\big) = \int_D \mu_C(y)\, dP^Y(y) \qquad \forall D \in \mathcal{A}_Y. \tag{2.8.3}$$

$P(X \in C \mid Y = y) := \mu_C(y)$ *is the conditional probability of* $X \in C$ *given* $Y = y$*.*

Proof. Let C be an arbitrary set that belongs to $\mathcal{A}_X$. We consider the set function $\nu_C\colon \mathcal{A}_Y \to [0,1]$ defined by

$$\nu_C\big(D\big) := P\big(X \in C,\ Y \in D\big) \qquad \forall D \in \mathcal{A}_Y.$$

ν_C is a measure on $(\Omega_Y, \mathcal{A}_Y)$. ($\nu_C$ is non-negative, satisfies $\nu_C(\emptyset) = 0$, and is σ-additive.) Furthermore, since $\nu_C(D) \leq P^Y(D)$ holds for all $D \in \mathcal{A}_Y$ we have that

$$\nu_C \ll P^Y.$$

It follows from Proposition 2.7.1 that there exists an $(\mathcal{A}_Y - \mathcal{B})$-measurable function $\widetilde{\mu}_C\colon \Omega_Y \to [0,\infty)$ such that

$$P\big(X \in C,\; Y \in D\big) \;=\; \int_D \widetilde{\mu}_C(y)\, dP^Y(y) \qquad \forall D \in \mathcal{A}_Y.$$

Let $E := \{y \in \Omega_Y\colon\; \widetilde{\mu}_C(y) > 1\}$. Then $E \in \mathcal{A}_Y$ and we have

$$\int_E \widetilde{\mu}_C(y)\, dP^Y(y) \;=\; P\big(X \in C,\; Y \in E\big) \;\le\; P^Y\big(E\big),$$

which implies that

$$\int_E \underbrace{\big[\widetilde{\mu}_C(y) \,-\, 1\big]}_{>0\ \forall y \in E}\, dP^Y(y) \;=\; 0.$$

Hence, we obtain from (iv) in Proposition 2.3.2 that

$$P^Y\big(E\big) \;=\; 0.$$

We define

$$\mu_C(y) \;:=\; \begin{cases} \widetilde{\mu}_C(y) & \text{if } \widetilde{\mu}_C(y) \le 1, \\ 0 & \text{if } \widetilde{\mu}_C(y) > 1 \end{cases}.$$

Then $\mu_C\colon\; \Omega_Y \to [0,1]$ is $(\mathcal{A}_Y - \mathcal{B})$-measurable and since μ_C is equal to $\widetilde{\mu}_C$ P^Y-almost surely, we obtain that

$$P\big(X \in C,\; Y \in D\big) \;=\; \int_D \mu_C(y)\, dP^Y(y) \qquad \forall D \in \mathcal{A}_Y,$$

as required. □

The determination of conditional probability according to (2.8.3) often requires a guess, and then it can be checked if the system of equations (2.8.3) is satisfied. On the other hand, in the spacial case where the random variables X and Y have a joint density, there is a simple algorithm for computing conditional probabilities.

Suppose that X and Y are real-valued random variables on a probability space $(\Omega, \mathcal{A}, P)$. Then the random vector $\binom{X}{Y}$ is $(\mathcal{A} - \mathcal{B}^2)$-measurable. Suppose further that the distribution of $\binom{X}{Y}$ has a density $p_{X,Y}$ with respect to Lebesgue measure λ^2. We obtain from Tonelli's theorem (Theorem 2.6.4) that, for each $A \in \mathcal{B}$,

$$\begin{aligned} P\big(X \in A\big) \;&=\; P\big(X \in A,\, Y \in \mathbb{R}\big) \;=\; \int_{A \times \mathbb{R}} p_{X,Y}(x,y)\, \lambda^2(dx, dy) \\ &=\; \int_A \left[\int_{\mathbb{R}} p_{X,Y}(x,y)\, d\lambda(y) \right] d\lambda(x), \end{aligned} \tag{2.8.4}$$

where $x \mapsto p_X(x) := \int_{\mathbb{R}} p_{X,Y}(x,y)\, d\lambda(y)$ is $(\mathcal{A} - \mathcal{B})$-measurable. It follows from (2.8.4) that p_X is a density of P^X w.r.t. λ. In relation to the **joint density** $p_{X,Y}$ of X and Y, p_X is called **marginal density** of X. Likewise we can see that the random variable Y has a marginal density p_Y, which is given by

$$p_Y(y) \;=\; \int_{\mathbb{R}} p_{X,Y}(x,y)\, d\lambda(x).$$

To summarize, a marginal density can be obtained from the joint density of two (or more) random variables by integrating this joint density with respect to the remaining components.

Using (2.8.3) we can compute conditional probabilities as follows. Let $C \in \mathcal{B}$ be arbitrary. Then, for each $D \in \mathcal{B}$,

$$\begin{aligned}
P\big(X \in C, Y \in D\big) &= \int_{C\times D} p_{X,Y}\, d\lambda^2 \\
&= \int_D \Big[\int_C p_{X,Y}(x,y)\, d\lambda(x)\Big]\, d\lambda(y) \\
&= \int_{D\cap\{y\colon p_Y(y)>0\}} \Big[\int_C \frac{p_{X,Y}(x,y)}{p_Y(y)}\, d\lambda(x)\Big] p_Y(y)\, d\lambda(y) \\
&\quad + \int_{D\cap\{y\colon p_Y(y)=0\}} \Big[\underbrace{\int_C p_{X,Y}(x,y)\, d\lambda(x)}_{=0=\int_C 0\, d\lambda(x)}\Big]\, d\lambda(y) \\
&= \int_D \Big[\int_C p_{X|Y=y}(x)\, d\lambda(x)\Big] dP^Y(y),
\end{aligned}$$

where

$$p_{X|Y=y}(x) := \begin{cases} \frac{p_{X,Y}(x,y)}{p_Y(y)} & \text{if } p_Y(y) > 0, \\ 0 & \text{if } p_Y(y) = 0 \end{cases}.$$

Hence, a version of the conditional probability of $X \in C$ given $Y = y$ is given by

$$P\big(X \in C \mid Y = y\big) = \int_C p_{X|Y=y}(x)\, d\lambda(x) \qquad \forall C \in \mathcal{B}.$$

If $p_Y(y) > 0$, then the corresponding set function $P\big(X \in \cdot \mid Y = y\big)\colon \mathcal{B} \to [0,1]$ is a probability measure on $(\mathbb{R}, \mathcal{B})$. Indeed, it follows from part (i) of Theorem 2.4.3 that $P\big(X \in \cdot \mid Y = y\big)$ is a measure on $(\mathbb{R}, \mathcal{B})$. Moreover,

$$P\big(X \in \mathbb{R} \mid Y = y\big) = \frac{1}{p_Y(y)} \int_{\mathbb{R}} p_{X,Y}(x,y)\, d\lambda(x) = 1.$$

If $p_Y(y) = 0$, then $P\big(X \in \cdot \mid Y = y\big)$ is not a probability measure. However, this deficiency is practically irrelevant since

$$P^Y\big(\{y\colon p_Y(y) = 0\}\big) = \int_{\{y\colon p_Y(y)=0\}} p_Y(y)\, d\lambda(y) = 0.$$

Example
Suppose that

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right), \tag{2.8.5}$$

where $\Sigma := \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$ is assumed to be positive definite. Then $\binom{X_1}{X_2}$ has a density p_{X_1,X_2} w.r.t. Lebesgue measure λ^2,

$$p_{X_1,X_2}(x_1,x_2) \;=\; \frac{1}{2\pi\sqrt{\det(\Sigma)}} \exp\left\{ -\frac{1}{2} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix} \right\}. \tag{2.8.6}$$

We might guess what the marginal distribution of X_1 is: Since $\binom{X_1}{X_2}$ is normally distributed is natural to assume that X_1 also follws a normal distribution. If so, then the corresponding parameters can be read off from relation (2.8.5), which suggests that $X_1 \sim \mathcal{N}(\mu_1, \sigma_{11})$. To prove this fact, we first derive an appropriate representation of the covariance matrix Σ of $\binom{X_1}{X_2}$. Let $D := \sigma_{22} - \sigma_{21}\sigma_{11}^{-1}\sigma_{12}$. Then $\det(\Sigma) = \sigma_{11}D$ and we obtain that

$$\begin{aligned}
\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}^{-1} &= \frac{1}{\det(\Sigma)} \begin{pmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{21} & \sigma_{11} \end{pmatrix} \\
&= \frac{1}{D} \begin{pmatrix} \sigma_{22}\sigma_{11}^{-1} & -\sigma_{12}\sigma_{11}^{-1} \\ -\sigma_{21}\sigma_{11}^{-1} & \sigma_{11}\sigma_{11}^{-1} \end{pmatrix} \\
&= \begin{pmatrix} \sigma_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \;+\; \frac{1}{D} \begin{pmatrix} \sigma_{21}\sigma_{11}^{-1}\sigma_{12}\sigma_{11}^{-1} & -\sigma_{12}\sigma_{11}^{-1} \\ -\sigma_{21}\sigma_{11}^{-1} & \sigma_{11}\sigma_{11}^{-1} \end{pmatrix}.
\end{aligned}$$

Therefore, we can decompose p_{X_1,X_2} given by (2.8.6) as

$$\begin{aligned}
p_{X_1,X_2}(x_1,x_2) &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}}} \exp\left\{ -\frac{1}{2} \frac{(x_1-\mu_1)^2}{\sigma_{11}} \right\} \\
&\quad \cdot \frac{1}{\sqrt{2\pi\, D}} \exp\left\{ -\frac{1}{2D} \big(x_2 - \mu_2 - \sigma_{21}\sigma_{11}^{-1}(x_1-\mu_1)\big)^2 \right\},
\end{aligned}$$

which implies that

$$\begin{aligned}
p_{X_1}(x_1) &= \int_{\mathbb{R}} p_{X_1,X_2}(x_1,x_2)\, d\lambda(x_2) \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}}} \exp\left\{ -\frac{1}{2} \frac{(x_1-\mu_1)^2}{\sigma_{11}} \right\} \\
&\quad \cdot \underbrace{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\, D}} \exp\left\{ -\frac{1}{2D} \big(x_2 - \mu_2 - \sigma_{21}\sigma_{11}^{-1}(x_1-\mu_1)\big)^2 \right\} d\lambda(x_2)}_{=1}.
\end{aligned}$$

This shows that $X_1 \sim \mathcal{N}(\mu_1, \sigma_{11})$.

The conditional distribution of X_2 given $X_1 = x_1$ has a density $p_{X_2|X_1=x_1}$ given by

$$p_{X_2|X_1=x_1}(x_2) \;=\; \frac{p_{X_1,X_2}(x_1,x_2)}{p_{X_1}(x_1)} \;=\; \frac{1}{\sqrt{2\pi\, D}} \exp\left\{ -\frac{1}{2D} \big(x_2 - \mu_2 - \sigma_{21}\sigma_{11}^{-1}(x_1-\mu_1)\big)^2 \right\}.$$

Hence,

$$P\big(X_2 \in \cdot \mid X_1 = x_1\big) \;=\; \mathcal{N}\big(\mu_2 + \sigma_{21}\sigma_{11}^{-1}(x_1-\mu_1),\; \sigma_{22} - \sigma_{21}\sigma_{11}^{-1}\sigma_{12}\big).$$

Theorem 2.8.2 is about the existence of conditional probability. In fact, if $(\Omega, \mathcal{A}, P)$ is a probability space, $(\Omega_X, \mathcal{A}_X)$ and $(\Omega_Y, \mathcal{A}_Y)$ are two measure spaces, and $X\colon\ \Omega \to \Omega_X$ and $Y\colon\ \Omega \to \Omega_Y$ are $(\mathcal{A} - \mathcal{A}_X)$- respectively $(\mathcal{A} - \mathcal{A}_Y)$-measurable mappings, then for $C \in \mathcal{A}_X$ the conditional probability $P\big(X \in C \big| Y = y\big)$ is well-defined as a solution to (2.8.3). Furthermore, it follows from Proposition 2.7.1 that, for two such conditional probabilities $P_1\big(X \in C \big| Y = y\big)$ and $P_2\big(X \in C \big| Y = y\big)$,

$$P^Y\big(\{y \in \Omega_Y\colon\ P_1\big(X \in C \big| Y = y\big) \neq P_2\big(X \in C \big| Y = y\big)\}\big) \;=\; 0.$$

In other words, for any fixed $C \in \mathcal{A}_X$, the conditional probability of $X \in C$ given $Y = y$ is uniquely defined up to P^Y-null sets. This, however, does not mean that one can be satisfied with such a definition when we consider a conditional **distribution** given by the set function $P\big(X \in \cdot \,\big| Y = y\big)\colon \mathcal{A}_X \to [0,1]$. The following example shows what can go wrong with the definition given in Theorem 2.8.2.

We consider the probability space $(\Omega, \mathcal{A}, P)$, where $\Omega = [0,1]$, $\mathcal{A} = \mathcal{B} \cap \Omega = \{B \cap \Omega\colon\ B \in \mathcal{B}\}$, and $P = \text{Uniform}[0,1]$. We define random variables X and Y by $X(\omega) = Y(\omega) := \omega \ \forall \omega \in \Omega$. Since the event $\{\omega\colon\ Y(\omega) = y\}$ implies that $X(\omega) = y$, it is natural to guess that, for $C \in \mathcal{A}$, conditional probabilities are given by

$$P\big(X \in C \big| Y = y\big) \;=\; \delta_y(C) \;=\; \begin{cases} 1 & \text{if } y \in C, \\ 0 & \text{if } y \notin C \end{cases} .$$

(δ_y is the so-called Dirac measure at y.)
Indeed, we have that

$$P\big(X \in C,\, Y \in D\big) \;=\; P\big(Y \in C \cap D\big) \;=\; \int_D \underbrace{\delta_y(C)}_{= \mathbb{1}_C(y)} \, dP^Y(y) \qquad \forall D \in \mathcal{A}.$$

We define a second family of set functions $P'(X \in \cdot \mid Y = y)$ by

$$P'\big(X \in C \big| Y = y\big) \;:=\; \begin{cases} 1 & \text{if } C = [0,y), \\ P\big(X \in C \big| Y = y\big) & \text{if } C \neq [0,y) \end{cases} .$$

Then $P'\big(X \in C \big| Y = y\big) = P\big(X \in C \big| Y = y\big)$ for all sets $C \in \mathcal{A}$ that are not of the form $[0,y)$ for some $y \in [0,1]$. On the other hand, if $C = [0, \bar{y})$, then $P'\big(X \in C \big| Y = y\big) \neq P\big(X \in C \big| Y = y\big)$ only if $y = \bar{y}$, that is $P^Y(\{y\colon\ P'\big(X \in C \big| Y = y\big) \neq P\big(X \in C \big| Y = y\big)\}) = 0$. Therefore, $P'\big(X \in C \big| Y = y\big)$ is also a version of the conditional probability of $X \in C$ given $Y = y$. Nevertheless, although $P\big(X \in \cdot \,\big| Y = y\big) = \delta_y$ is a probability measure for all $y \in [0,1]$, this property is not shared by $P'\big(X \in \cdot \big| Y = y\big)$ since

$$P'\big(X \in [0,y] \big| Y = y\big) \;=\; 1 \;\neq\; P'\big(X \in [0,y) \big| Y = y\big) \;+\; P'\big(X \in \{y\} \big| Y = y\big) \;=\; 2.$$

Furthermore, it also follows for the conditional **distributions** that

$$P'\big(X \in \cdot \,\big| Y = y\big) \;\neq\; P\big(X \in \cdot \,\big| Y = y\big) \qquad \forall y \in [0,1],$$

i.e., the two versions of conditional distributions are different with probability 1. Hence, when we are interested in conditional distributions rather than conditional probabilities, these shortcomings exemplified by $P'(X \in \cdot \mid Y = \cdot)$ should be rectified.

Definition. Let $(\Omega, \mathcal{A}, P)$ be a probability space, and let $(\Omega_X, \mathcal{A}_X)$ and $(\Omega_Y, \mathcal{A}_Y)$ be two measurable spaces. Suppose that $X\colon \Omega \to \Omega_X$ and $Y\colon \Omega \to \Omega_Y$ are $(\mathcal{A} - \mathcal{A}_X)$-respectively $(\mathcal{A} - \mathcal{A}_Y)$-measurable mappings.
Then $P(X \in \cdot \mid Y = \cdot)\colon \mathcal{A}_X \times \Omega_Y \to [0,1]$ is called **regular conditional distribution** if

(i) For each $y \in \Omega_Y$, $P(X \in \cdot \mid Y = y)$ is a probability measure on $(\Omega_X, \mathcal{A}_X)$.

(ii) For each $C \in \mathcal{A}_X$, $P(X \in C \mid Y = \cdot)$ is a version of the conditional probability $X \in C$ given Y.

The following theorem shows for the case of an $\mathbb{R}^d$-valued random variable X that such a regular conditional distribution always exists. Moreover, two such regular conditional distributions are equal except on a set of probability 0. Taking the possible non-uniqueness into account, a specific such function will be called a **version** of the regular conditional distribution.

Theorem 2.8.3.
Let $(\Omega, \mathcal{A}, P)$ be a probability space, and let $(\Omega_Y, \mathcal{A}_Y)$ be a measurable space. Suppose that $X\colon \Omega \to \mathbb{R}^d$ is $(\mathcal{A} - \mathcal{B}^d)$-measurable and that $Y\colon \Omega \to \Omega_Y$ is $(\mathcal{A} - \mathcal{A}_Y)$-measurable.
Then

(i) there exists (a version) $P\big(X \in \cdot \,\big|Y = \cdot\big)$ of the regular conditional distribution of X given Y,

(ii) if $P_1\big(X \in \cdot \,\big|Y = \cdot\big)$ and $P_2\big(X \in \cdot \,\big|Y = \cdot\big)$ are two versions of a regular conditional probability, then

$$P^Y\Big(\big\{y \in \Omega_Y\colon\ P_1\big(X \in \cdot \,\big|Y = y\big) \neq P_2\big(X \in \cdot \,\big|Y = y\big)\big\}\Big) \;=\; 0.$$

Before we turn to the proof of this theorem for the special case where X is a real-valued random variable, we provide an auxiliary result that will be used in the course of this proof.

Lemma 2.8.4. *Let $(\Omega, \mathcal{A}, P)$ be a probability space, and let $(\Omega_Y, \mathcal{A}_Y)$ be a measurable space. Suppose that $X\colon \Omega \to \mathbb{R}$ is $(\mathcal{A} - \mathcal{B})$-measurable, that $Y\colon \Omega \to \Omega_Y$ is $(\mathcal{A} - \mathcal{A}_Y)$-measurable, and let $P_0(X \in \cdot \mid Y = \cdot)$ be any version of the (not necessarily regular) conditional distribution. For each rational r, let $F_0(r, y) := P_0(X \in (-\infty, r], Y = y)$.*
Then $F_0(\cdot, y)$ satisfies the defining properties of a probability distribution function outside a P^Y-null set $N \in \mathcal{A}_Y$, i.e. for $y \in \Omega_Y \backslash N$

a) $F_0(\cdot, y)$ is monotonically non-decreasing on $\mathbb{Q}$, i.e.

$$F_0\big(r, y\big) \;\leq\; F_0(s, y) \qquad \forall r, s \in \mathbb{Q},\ r < s. \tag{2.8.7}$$

b) $F_0(\cdot, y)$ is right-continuous on $\mathbb{Q}$, i.e.

$$F_0\big(r, y\big) \;=\; \lim_{n\to\infty} F_0\big(r_n, y\big) \qquad \textit{for all sequences } (r_n)_{n\in\mathbb{Q}},\ r_n \in \mathbb{Q},\ r_n \searrow r. \tag{2.8.8}$$

c) $F_0(\cdot, y)$ assigns value 1 to $\mathbb{R}$, i.e.

$$\lim_{n\to\infty} F_0\big(-n, y\big) \;=\; 0, \qquad \lim_{n\to\infty} F_0\big(n, y\big) \;=\; 1. \tag{2.8.9}$$

Proof. We show that there exist sets of probability zero such that, outside these sets, (2.8.7), (2.8.8), and (2.8.9) are satisfied.

a) Let for $r, s \in \mathbb{Q}$, $r < s$, $A_{rs} := \{y \in \Omega_Y\colon\; F_0(r,y) > F_0(s,y)\}$. Then

$$\begin{aligned} 0 \;&\geq\; P\big(X \leq r,\, Y \in A_{rs}\big) \;-\; P\big(X \leq s,\, Y \in A_{rs}\big) \\ &=\; \int_{A_{rs}} \underbrace{\big(F_0(r,y) \;-\; F_0(s,y)\big)}_{>0\ \ \forall y \in A_{rs}} dP^Y(y), \end{aligned}$$

which implies that $P^Y(A_{rs}) = 0$. Let $A := \bigcup_{r,s\in\mathbb{Q},\, r<s} A_{rs}$. Then (2.8.7) is satisfied for each $y \in A^c$, where $A \in \mathcal{A}_Y$ and $P^Y(A) = 0$.

b) Since $F_0(\cdot, y)$ is monotonically non-decreasing for $y \in A^c$, it suffices to consider the sets $B_r := \{y \in A^c\colon\; F_0\big(r,y\big) < \lim_{n\to\infty} F_0\big(r+1/n, y\big)\}$. Since P is continuous from above we obtain

$$\begin{aligned} 0 \;&=\; \lim_{n\to\infty} P\big(X \leq r + 1/n,\, Y \in B_r\big) \;-\; P\big(X \leq r,\, Y \in B_r\big) \\ &=\; \lim_{n\to\infty} \int_{B_r} F_0(r+1/n, y) \;-\; F_0(r,y)\, dP^Y(y) \\ &=\; \int_{B_r} \underbrace{\lim_{n\to\infty} F_0(r+1/n, y) \;-\; F_0(r,y)}_{>0\ \ \forall y \in B_r}\, dP^Y(y), \end{aligned}$$

which implies that $P^Y(B_r) = 0$. Hence, (2.8.8) is satisfied for each $y \in B^c$, where $B := \big(A \cup \big(\bigcup_{r\in\mathbb{Q}} B_r\big)\big) \in \mathcal{A}_Y$ and $P^Y(B) = 0$.

c) We consider the set $C := \big\{y \in A^c\colon\; \lim_{n\to\infty} F_0(n,y) - F_0(-n,y) < 1\big\}$. Since P is continuous from below and since $P(X \in (-n,n] \mid Y = y) = F_0(n,y) - F_0(-n,y)$ we obtain that

$$\begin{aligned} 0 \;&=\; P\big(Y \in C\big) \;\;-\;\; \lim_{n\to\infty} P\big(X \in (-n,n], Y \in C\big) \\ &=\; \int_C dP^Y(y) \;\;-\;\; \lim_{n\to\infty} \int_C F_0(n,y) - F_0(-n,y)\, dP^Y(y) \\ &=\; \int_C \underbrace{1 \;-\; \lim_{n\to\infty} F_0(n,y) - F_0(-n,y)}_{>0\,\forall y \in C}\, dP^Y(y), \end{aligned}$$

which implies that $P^Y(C) = 0$.

To summarize, (2.8.7), (2.8.8), and (2.8.9) together are satisfied for y outside the set $N := A \cup B \cup C$, where $N \in \mathcal{A}_Y$ and $P^Y(N) = 0$. □

Now we turn to the proof of the theorem.

Proof of Theorem 2.8.3.
(i) We restrict ourselves to the case of $d = 1$, i.e. X being a real-valued random variable. The case of an $\mathbb{R}^d$-valued random variable X can be treated similarly, however, the notation gets more cumbersome.

Let $P_0(X \in \cdot \mid Y = \cdot)$ be an arbitrary (not necessarily regular) conditional distribution whose existence follows from Theorem 2.8.2 and let, for $r \in \mathbb{Q}$, $F_0(r, y) := P_0(X \in (-\infty, r] \mid Y = y)$. Recall from Lemma 2.8.4 that there exists a set $N \in \mathcal{A}_Y$ such that $P^Y(N) = 0$ and that for $y \notin N$ the function $F_0(\cdot, y)\colon \mathbb{Q} \to [0, 1]$ shares the properties of a probability distribution function. We use F_0 as a starting point to construct a version of the regular conditional distribution on the complement of N in a meaningful way.

For $y \in N^c$ we extend $F_0(\cdot, y)$ to all of $\mathbb{R}$ be setting

$$F\big(x, y\big) \;:=\; \inf\big\{F_0(r, y)\colon\; r \in \mathbb{Q},\, x < r\big\}.$$

It follows in particular from (2.8.7) and (2.8.8) that

$$F\big(r, y\big) \;=\; F_0\big(r, y\big) \qquad \forall x \in \mathbb{Q}.$$

Next we show that for each $y \in N^c$ the function $F(\cdot, y)$ satisfies relations (i) to (iii) in Lemma 1.5.1. Indeed, we then have

a) $F(\cdot, y)$ is monotonically non-decreasing.
Indeed, if $x_1 < x_2$, then $\{r \in \mathbb{Q}\colon x_1 < r\} \supset \{r \in \mathbb{Q}\colon x_2 < r\}$, and so $F(x_1, y) = \inf\{F_0(r, y)\colon\; r \in \mathbb{Q},\, x_1 < r\} \le \inf\{F_0(r, y)\colon\; r \in \mathbb{Q},\, x_2 < r\} = F(x_2, y)$.

b) $F(\cdot, y)$ is right-continuous.
To see this, let $x \in \mathbb{R}$ be arbitrary and let $(x_n)_{n\in\mathbb{N}}$ be any sequence of real numbers such that $x_n \searrow x$.

We choose an **accompanying sequence** $(r_n)_{n\in\mathbb{N}}$ of rational numbers such that $x_n < r_n$ and $r_n \searrow x$. Then, by monotonicity of $F(\cdot, y)$,

$$F\big(x, y\big) \;\le\; F\big(x_n, y\big) \;\le\; F\big(r_n, y\big) \;=\; F_0\big(r_n, y\big).$$

On the other hand, we have that $F_0(r_n, y) \xrightarrow[n\to\infty]{} F(x, y)$, which leads to

$$F\big(x_n, y\big) \xrightarrow[n\to\infty]{} F\big(x, y\big).$$

c) The relations

$$\lim_{n\to\infty} F\big(-n, y\big) \;=\; 0, \qquad \lim_{n\to\infty} F\big(n, y\big) \;=\; 1$$

follow immediately from (2.8.9).

Hence, if $y \in N^c$, then $F(\cdot, y)$ satisfies relations (i) to (iii) in Lemma 1.5.1, and we obtain by Theorem 1.5.2 that there exists an associated probability measure $P(X \in \cdot \mid Y = y)$ on $(\mathbb{R}, \mathcal{B})$ such that $P((-\infty, x] \mid y) = F(x, y)$ for all $x \in \mathbb{R}$.

For $y \in N$, take $P(X \in \cdot \mid Y = y) = P'(\cdot)$, where P' is any arbitrary but fixed probability measure on $(\mathbb{R}, \mathcal{B})$. With this choice, $P(X \in \cdot \mid Y = y)$ is for each $y \in \Omega_Y$ a probability measure.

Next we show that $P(X \in \cdot \mid Y = \cdot)$ is a version of the conditional distribution, i.e. according to the definition of conditional probability, we have to show that for each $C \in \mathcal{B}$,

d) $\quad y \mapsto P\big(X \in C \big| Y = y\big)$ is $(\mathcal{A}_Y - \mathcal{B})$-measurable

and

e) $\quad P\big(X \in C,\, Y \in D\big) \;=\; \int_D P\big(X \in C \big| Y = y\big)\, dP^Y(y) \qquad \forall D \in \mathcal{A}_Y.$

Proof of d) For $r \in \mathbb{Q}$ we have that

$$P\big(X \in (-\infty, r]\big|Y = y\big) \;=\; \begin{cases} P_0\big(X \in (-\infty, r]\big|Y = y\big) & \text{if } y \in N^c, \\ P'\big((-\infty, r]\big) & \text{if } y \in N \end{cases} \qquad (2.8.10)$$

Hence, $y \mapsto P\big(X \in (-\infty, r]\big|Y = y\big)$ is $(\mathcal{A}_Y - \mathcal{B})$-measurable. To show measurability of $y \mapsto P\big(X \in C\big|Y = y\big)$ for all $C \in \mathcal{B}$, we define the system of good sets

$$\mathcal{D} \;:=\; \Big\{C \in \mathcal{B}\colon\; y \mapsto P\big(X \in C\big|Y = y\big) \text{ is } (\mathcal{A}_Y - \mathcal{B}) - \text{measurable}\Big\}.$$

Since $\mathcal{D}$ is a Dynkin system containing the $\cap$-stable collection of sets $\{(-\infty, r]\colon r \in \mathbb{Q}\}$ we see that $\mathcal{B} = \delta(\{(-\infty, r]\colon r \in \mathbb{Q}\}) \subseteq \mathcal{D}$.

Proof of e) Note that for arbitrary fixed $D \in \mathcal{A}_Y$

$$P\big(X \in \cdot\,,\, Y \in D\big)$$

and

$$\rho_D(\,\cdot\,) \;:=\; \int_D P\big(X \in \cdot\,\big|Y = y\big)\, dP^Y(y)$$

are finite measures on $(\mathbb{R}, \mathcal{B})$. Moreover, it follows from (2.8.10) for each $r \in \mathbb{Q}$ that

$$P\big(X \in (-\infty, r],\, Y \in D\big) \;=\; \int_D P\big(X \in (-\infty, r]\big|Y = y\big)\, dP^Y(y) \qquad \forall D \in \mathcal{A}_Y.$$

Hence, the measures $P\big(X \in \cdot\,, Y \in D\big)$ and $\rho_D(\cdot)$ coincide on the $\cap$-stable collection of sets $\{(-\infty, r]\colon r \in \mathbb{Q}\}$. It follows from the uniqueness theorem (Theorem 1.3.8) that these two measures also coincide on $\sigma\big(\{(-\infty, r]\colon r \in \mathbb{Q}\}\big)$, i.e. for each $D \in \mathcal{A}_Y$

$$P\big(X \in C,\, Y \in D\big) \;=\; \int_D P\big(X \in C\big|Y = y\big)\, dP^Y(y) \qquad \forall C \in \mathcal{B},$$

as required.

(ii) Suppose that $P_1\big(X \in \cdot\,\big|Y = \cdot\big)$ and $P_2\big(X \in \cdot\,\big|Y = \cdot\big)$ are two versions of a regular conditional distribution. Let

$$N_r \;=\; \Big\{y \in \Omega_Y\colon\; P_1\big(X \in (-\infty, r]\big|Y = y\big) \neq P_2\big(X \in (-\infty, r]\big|Y = y\big)\Big\}.$$

Then $N_r \in \mathcal{A}_Y$ and $P^Y(N_r) = 0$. Let $N^{\neq} := \bigcup_{r \in \mathbb{Q}} N_r$. It follows that $N^{\neq} \in \mathcal{A}_Y$, $P^Y(N^{\neq}) \le \sum_{r \in \mathbb{Q}} P(N_r) = 0$. If $y \notin N^{\neq}$, then

$$P_1\big(X \in (-\infty, r]\big|Y = y\big) \;=\; P_2\big(X \in (-\infty, r]\big|Y = y\big) \qquad \forall r \in \mathbb{Q},$$

that is, these two probability measures coincide on the collection of sets $\{(-\infty, r]\colon r \in \mathbb{Q}\}$. It follows again from the uniqueness theorem (Theorem 1.3.8) that these two measures also coincide on $\sigma\big(\{(-\infty, r]\colon r \in \mathbb{Q}\}\big)$, i.e.

$$P_1\big(X \in C\big|Y = y\big) \;=\; P_2\big(X \in C\big|Y = y\big) \qquad \forall C \in \mathcal{B}.$$

□

Now we suppose that X is an integrable, extended real-valued random variable and that $Y\colon \Omega \to \mathcal{A}_Y$ is an arbitrary $(\mathcal{A} - \mathcal{A}_Y)$-measurable random variable on a common probability space $(\Omega, \mathcal{A}, P)$. Then there exists an $(\mathcal{A}_Y - \mathcal{B})$ measurable function $\mu\colon \Omega_Y \to \mathbb{R}$ such that

$$E\big[X\,\mathbb{1}(Y \in D)\big] \;=\; \int_D \mu(y)\,dP^Y(y) \qquad \forall D \in \mathcal{A}_Y. \tag{2.8.11}$$

$E(X \mid Y = y) := \mu(y)$ is a version of the **conditional expected value** of X given $Y = y$. For two such functions μ_1 and μ_2 we have that $P^Y\big(\{y\colon\; \mu_1(y) \neq \mu_2(y)\}\big) = 0$.

Indeed, consider first the case where $X\colon \Omega \to [0,\infty]$ is a non-negative random variable such that $EX < \infty$. The set function $\nu\colon \mathcal{A}_Y \to [0,\infty)$ given by $\nu(D) := E\big[X\,\mathbb{1}(Y \in D)\big]$ is a finite measure and it holds that $\nu \ll P^Y$. Hence, it follows from Proposition 2.7.1 (Alternatively we could use the Radon-Nikodym theorem.) that there exists an $(\mathcal{A}_Y - \mathcal{B})$-measurable function $\mu\colon \Omega_Y \to \mathbb{R}$ such that (2.8.11) is satisfied.

If $X\colon \Omega \to \bar{\mathbb{R}}$ is a extended real-valued random variable such that $E|X| < \infty$, then X^+ and X^- are both non-negative and integrable random variables. Hence, ν^+ and ν^- defined by $\nu^+(D) = E\big[X^+\mathbb{1}(Y \in D)\big]$ and $\nu^-(D) = E\big[X^-\mathbb{1}(Y \in D)\big]$ are both finite measures on $(\Omega_Y, \mathcal{A}_Y)$ that are absolutely continuous with respect to P^Y. Using once more Proposition 2.7.1 we obtain that there exist $(\mathcal{A}_Y - \mathcal{B})$-measurable functions $\mu^+, \mu^-\colon \Omega_Y \to \mathbb{R}$ such that

$$E\big[X^{\pm}\,\mathbb{1}(Y \in D)\big] \;=\; \int_D \mu^{\pm}(y)\,dP^Y(y) \qquad \forall D \in \mathcal{A}_Y.$$

Then $E(X \mid Y = y) := \mu^+(y) - \mu^-(y)$ is a version of the conditional expected value of X given $Y = y$.

Finally, let μ_1 and μ_2 be two such functions. We define $D^{>} := \{y\colon\; \mu_1(y) > \mu_2(y)\}$ and $D^{<} := \{y\colon\; \mu_1(y) < \mu_2(y)\}$. Then

$$\int_{D^>} \underbrace{\mu_1(y) - \mu_2(y)}_{>0\ \ \forall y \in D^>}\,dP^Y(y) \;=\; E\big[X\,\mathbb{1}(Y \in D^>)\big] \;-\; E\big[X\,\mathbb{1}(Y \in D^>)\big] \;=\; 0$$

as well as

$$\int_{D^<} \underbrace{\mu_2(y) - \mu_1(y)}_{>0\ \ \forall y \in D^<}\,dP^Y(y) \;=\; E\big[X\,\mathbb{1}(Y \in D^<)\big] \;-\; E\big[X\,\mathbb{1}(Y \in D^<)\big] \;=\; 0,$$

which shows that μ_1 and μ_2 are equal with probability 1.

In probability theory, and in particular in connection with so-called stochastic processes, it is often convenient to condition on sub-σ-algebras. Suppose that $(\Omega, \mathcal{A}, P)$ is a probability space, let $X\colon\; \Omega \to \bar{\mathbb{R}}$ be an $(\mathcal{A} - \bar{\mathcal{B}})$-measurable function (a random variable) such that $E|X| < \infty$, and let $\widetilde{\mathcal{A}} \subseteq \mathcal{A}$ be a sub-σ-algebra of $\mathcal{A}$. Then $E\big(X \mid \widetilde{\mathcal{A}}\big)\colon\; \Omega \to \mathbb{R}$ is defined to be an $(\widetilde{\mathcal{A}} - \mathcal{B})$-measurable function such that

$$E\big[X\,\mathbb{1}_D\big] \;=\; \int_D E\big(X \mid \widetilde{\mathcal{A}}\big)(\omega)\,dP(\omega) \qquad \forall D \in \widetilde{\mathcal{A}}. \tag{2.8.12}$$

Indeed, if X is a non-negative random variable, then $\nu\colon \widetilde{\mathcal{A}} \to [0,\infty)$ defined by $\nu(D) = E\big[X\,\mathbb{1}_D\big]$ and $P\big|_{\widetilde{\mathcal{A}}}$ defined by $P\big|_{\widetilde{\mathcal{A}}}(D) = P(D)$ for all $D \in \widetilde{\mathcal{A}}$ are both finite measures on

$(\Omega, \widetilde{\mathcal{A}})$ and it holds that $\nu \ll P\big|_{\widetilde{\mathcal{A}}}$. The existence of $E\big(X \mid \widetilde{\mathcal{A}}\big)$ satisfying (2.8.12) follows once again from Proposition 2.7.1. If X is an integrable, not necessarily non-negative random variable, then (2.8.12) follows after a decomposition of X into X^+ and X^-.

If now $Y\colon\ \Omega \to \Omega_Y$ is $(\mathcal{A} - \mathcal{A}_Y)$-measurable, let

$$\sigma\big(Y\big) := \Big\{Y^{-1}(A)\colon\ A \in \mathcal{A}_Y\Big\}$$

be the σ**-algebra generated by** Y. Then, for each $D \in \sigma\big(Y\big)$, there exists some $A \in \mathcal{A}_Y$ such that $D = Y^{-1}(A)$, and it follows from Proposition 2.3.6 that

$$\begin{aligned} E\big[X\,\mathbb{1}_D\big] &= E\big[X\,\mathbb{1}(Y \in A)\big] \\ &= \int_A E\big(X\big|Y = y\big)\,dP^Y(y) \\ &\overset{Prop.\ 2.3.6}{=} \int_D E\big(X \mid Y = Y(\omega)\big)\,dP(\omega). \end{aligned}$$

Hence, $E\big(X \mid Y = Y(\omega)\big) = E\big(X \mid \sigma(Y)\big)(\omega)$ holds for almost all ω. Note that $E\big(X \mid \sigma(Y)\big)$ is a random variable whereas $E\big(X \mid \sigma(Y)\big)(\omega) = E\big(X \mid Y = Y(\omega)\big)$ is one of its realizations.