Probability Theory

Naotaka Kajino

WS 2011/2012, Universität Bielefeld

Preface

This is a lecture note for the lecture course “Probability Theory” in the University of Bielefeld (240111, WS 2011/2012).

Several theorems and exercises are adopted from an unpublished lecture note [6] on measure theory by Professor Jun Kigami in Kyoto University, and some other problems are borrowed from an unpublished lecture note by Professor Grigor’yan in the University of Bielefeld. The author would like to express his deepest gratitude toward Professor Kigami and Professor Grigor’yan for their permission to quote their unpublished notes in this lecture note.

Contents

Chapter 0

Prologue

It is assumed that the reader is already familiar with elementary probability theory, e.g. calculation of probabilities of events resulting from coin flipping or dice. The purpose of this course is to provide a rigorous mathematical background of probability theory. Modern probability theory, as a part of mathematics, is developed on the basis of *measure theory*, which will be treated in the first half of this course.

0.1 Introduction

How to formulate "probability" rigorously?

Let us consider the situation where we throw a dice and see the outcome X. X is a "random variable" taking values in $\{1, 2, 3, 4, 5, 6\}$, and each side of the dice appears with "probability" $1/6$; $\mathbb{P}[X = k] = 1/6$ for $k \in \{1, 2, 3, 4, 5, 6\}$.[1] Of course we can consider the "probabilities" of other "events"; for example, $\mathbb{P}[X \text{ is odd}] = 1/2$, $\mathbb{P}[X \text{ is divisible by 3}] = 1/3$, $\mathbb{P}[X \text{ is a prime number}] = 1/2$.

We have used the terms "probability", "random variable" and "event", which are fundamental notions in probability theory. These phrases, however, are used only in very naive manners and their mathematical meanings are still unclear. We would like to give a rigorous mathematical formulation to these notions, in order to treat probability theory as a part of mathematics.

Here is an idea of how to formulate "probability" mathematically: let Ω be the collection of all possible "cases". Suppose that there is a *function* $\mathbb{P}$*, which assigns to each subset* Ω_0 *of* Ω *a real number* $\mathbb{P}[\Omega_0] \in [0, 1]$, interpreted as the "probability" of Ω_0. A "random variable" X should tell us a number $X(\omega) \in \mathbb{R}$ for each "case" $\omega \in \Omega$, and such X is nothing but a *function* $X : \Omega \to \mathbb{R}$ *on* Ω. For example, in the above situation of a dice,

- $\Omega = \{1, 2, 3, 4, 5, 6\}$,
- $\mathbb{P}[A] = \#A/6$ for $A \subset \Omega$, where $\#A$ denotes the number of elements of A.

[1] It is implicitly assumed that all sides of the dice are equally likely to appear.

- The outcome X of the dice is the function $X : \Omega \to \mathbb{R}$ given by $X(k) = k$.

Let A be an "event". In each "case" $\omega \in \Omega$, either the "event" A occurs or it does not occur, and the set $\Omega_A := \{\omega \in \Omega \mid A \text{ occurs in the "case" } \omega\}$ represents precisely when A occurs. Then the *"probability of A"* should be $\mathbb{P}[\Omega_A]$. In this way, each "event" A is represented by the corresponding set Ω_A of "cases" where it occurs, and then it seems natural to identify Ω_A with the "event" A. In other words, an "event" should be a *subset of* Ω. In the above example of a dice, the three events "X is odd", "X is divisible by 3" and "X is a prime number" correspond to $\{\omega \in \Omega \mid X(\omega) \text{ is odd}\} = \{1, 3, 5\}$, $\{\omega \in \Omega \mid X(\omega) \text{ is divisible by } 3\} = \{3, 6\}$ and $\{\omega \in \Omega \mid X(\omega) \text{ is a prime number}\} = \{2, 3, 5\}$, respectively.

In summary, a rigorous mathematical formulation of "probability" will require

- a set Ω, called the *sample space*, and
- a $[0, 1]$-valued function $\mathbb{P}$, whose argument is an *event* (a subset of Ω) and whose values are the *probabilities of events*,

and then the outcome of a random trial is represented by

- a *random variable* X, which is a function $X : \Omega \to \mathbb{R}$ on Ω.

Required properties of a "probability" and its domain

In order for the above $[0, 1]$-valued function $\mathbb{P}$ to be considered as a "probability", of course it has to possess certain properties. First, we need to specify the conditions to be satisfied by the *domain $\mathcal{F}$ of $\mathbb{P}$*, which is a subset of 2^{Ω}[2] and is the collection of *sets whose probabilities are defined*. Here is a list of properties which $\mathcal{F}$ is desired to have:[3]

- $\emptyset, \Omega \in \mathcal{F}$, where $\emptyset$ denotes the *empty set*.
- If $A \in \mathcal{F}$ then $A^c := \Omega \setminus A \in \mathcal{F}$. If $A, B \in \mathcal{F}$ then $A \setminus B \in \mathcal{F}$.
- If $n \in \mathbb{N}$ and $\{A_i\}_{i=1}^n \subset \mathcal{F}$ then $A_1 \cup \cdots \cup A_n \in \mathcal{F}$ and $A_1 \cap \cdots \cap A_n \in \mathcal{F}$.

In fact, the third condition is still too weak for theoretical purposes, and instead $\mathcal{F}$ will be required to satisfy the following stronger condition:

- If $\{A_n\}_{n=1}^\infty \subset \mathcal{F}$ then $\bigcup_{n=1}^\infty A_n \in \mathcal{F}$ and $\bigcap_{n=1}^\infty A_n \in \mathcal{F}$.

Such a subset $\mathcal{F} \subset 2^{\Omega}$ is called a *σ-algebra in* Ω, and each $A \in \mathcal{F}$ is called an *event*.

At this point one might wonder why we have to consider not 2^{Ω} but a subset $\mathcal{F}$ of 2^{Ω}. In fact, when we consider the probabilities of events involving *infinitely many* random trials, we need to choose an *uncountable* set as the sample space Ω[4] and then 2^{Ω} is too large to be the domain of a natural "probability" $\mathbb{P}$. Why 2^{Ω} is "too large" will become clear during the first half of this course.

[2] 2^{Ω} denotes the *power set of* Ω: $2^{\Omega} := \{A \mid A \subset \Omega\}$, i.e. the set *consisting of all subsets of* Ω.

[3] A subset $\mathcal{F} \subset 2^{\Omega}$ satisfying these three conditions is called an *algebra in* Ω.

[4] For example, a natural choice of Ω for the trial of throwing a dice infinitely many times is to take $\Omega := \{1, 2, 3, 4, 5, 6\}^{\mathbb{N}}$, which is an uncountable set.

As explained above, a "probability" $\mathbb{P}$ is required to be defined on a σ-algebra $\mathcal{F}$ in Ω. Then what properties should $\mathbb{P}$ have? Here are conditions to be satisfied by a "probability" $\mathbb{P}$:

- $\mathbb{P}[\Omega] = 1$.
- $\mathbb{P}[\emptyset] = 0$.
- If $n \in \mathbb{N}$, $\{A_i\}_{i=1}^n \subset \mathcal{F}$ and $A_i \cap A_j = \emptyset$ for any $i, j \in \{1, \dots, n\}$ with $i \neq j$, then $\mathbb{P}[A_1 \cup \cdots \cup A_n] = \mathbb{P}[A_1] + \cdots + \mathbb{P}[A_n]$.

The third property is called the *finite additivity*, which is still insufficient for theoretical purposes and has to be replaced by the following *countable additivity*:

- If $\{A_n\}_{n=1}^\infty \subset \mathcal{F}$ and $A_i \cap A_j = \emptyset$ for any $i, j \in \mathbb{N}$ with $i \neq j$, then $\mathbb{P}\big[\bigcup_{n=1}^\infty A_n\big] = \sum_{n=1}^\infty \mathbb{P}[A_n]$.

A function $\mathbb{P} : \mathcal{F} \to [0, 1]$ which is defined on a σ-algebra $\mathcal{F}$ and satisfies the above conditions is called a *probability measure*, and the triple $(\Omega, \mathcal{F}, \mathbb{P})$ of a set Ω, a σ-algebra $\mathcal{F}$ in Ω and a probability measure $\mathbb{P}$ on $\mathcal{F}$ is called a *probability space*. This is the correct mathematical formulation of the notion of probability.

Note that the "volume" functions, e.g. the "length" of subsets of $\mathbb{R}$, the "area" of subsets of $\mathbb{R}^2$ and the "volume" of subsets of $\mathbb{R}^3$, are also desired to satisfy these conditions except $\mathbb{P}[\Omega] = 1$. Such a function (i.e. a countably additive non-negative function on a σ-algebra) is called a *measure*, which is the correct mathematical formulation of the notion of volume.

Random variables and expectation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. As described above, the outcome of a random trial is represented by a random variable, which is a function $X : \Omega \to \mathbb{R}$. Once a random variable X is given, it is natural to consider its *expectation* (or *mean*) $\mathbb{E}[X]$. Mathematically, it is a synonym for the *integral of X with respect to* $\mathbb{P}$:

$$\mathbb{E}[X] = \int_\Omega X d\mathbb{P}. \tag{0.1}$$

In order for $\mathbb{E}[X]$ to be defined, X has to be suitably related with $\mathcal{F}$. For example, if X takes its values in the set $\mathbb{N}$ of positive integers, then $\mathbb{E}[X]$ should be given by

$$\mathbb{E}[X] = \sum_{n=1}^\infty n \cdot \mathbb{P}[X = n],$$

where $\{X = n\} = \{\omega \in \Omega \mid X(\omega) = n\} = X^{-1}(n)$ is required to belong to $\mathcal{F}$. Such a function X is called *$\mathcal{F}$-measurable*, and *only* $\mathcal{F}$-measurable functions on Ω are (and deserve to be) called *random variables*. The precise definition of $\mathcal{F}$-measurable functions is given in Section 1.2, and integration with respect to a measure will be defined in Section 1.3.

The role of the countable additivity of $\mathbb{P}$ becomes clear when we consider a sequence $\{X_n\}_{n=1}^{\infty}$ of random variables. Suppose that $\{X_n(\omega)\}_{n=1}^{\infty}$ converges to $X(\omega) \in \mathbb{R}$ for any $\omega \in \Omega$. Then since $\mathcal{F}$ is a σ-algebra, $X : \Omega \to \mathbb{R}$ is shown to be $\mathcal{F}$-measurable (and hence it is also a random variable), and the countable additivity of $\mathbb{P}$ assures that, under certain reasonable conditions on $\{X_n\}_{n=1}^{\infty}$,

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X], \qquad \text{that is,} \qquad \lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}\Big[\lim_{n\to\infty} X_n\Big]. \tag{0.2}$$

(0.2) asserts the possibility of *interchange of the order of limit and integral*, which often plays fundamental roles in analysis! In measure theory, this type of assertions are called *convergence theorems*. The properties of σ-algebras and measures make the conditions for convergence theorems much simpler than those in classical calculus, where one usually assumes the *uniform convergence* of the sequence of functions. The precise statements of convergence theorems will be presented in Section 1.3 below.

0.2 Some Basic Facts and Notations

Here we collect some basic facts and notations which the reader is assumed to be familiar with. By an equation of the form

$$A := B$$

we mean that *A is defined by B*.

As usual, $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ denote the set of natural numbers, integers, rational numbers, real numbers and complex numbers, respectively. Here our convention is that $\mathbb{N}$ does NOT contain 0, so that $\mathbb{N} = \{1, 2, 3, \ldots\}$.

Let X be a set. 2^X denotes the *power set* of X, i.e. $2^X := \{A \mid A \subset X\}$, as noted before. X is called *countably infinite* if and only if there exists a bijection $\varphi : \mathbb{N} \to X$, and X is called *countable* if and only if it is either finite or countably infinite. A set which is not countable is called *uncountable*. Clearly $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$ are countable, and it is easy to see that $X_1 \times \cdots \times X_n$ is countable if $n \in \mathbb{N}$ and X_i is a countable set for each $i \in \{1, \ldots, n\}$. On the other hand, $\mathbb{R}$, $\mathbb{C}$ and $A^{\mathbb{N}}$, where A is any set with at least 2 elements, are shown to be uncountable.

Let X, Y be sets, let $f : X \to Y$ be a map and let $A \subset X$. Then the map $f|_A : A \to Y$ defined by $f|_A(x) := f(x)$ is called the *restriction of f to A*.

0.3 The Extended Real Line $[-\infty, \infty]$

In measure theory, it is essential to consider functions with values in the *extended real line*. Here we collect basic definitions and facts concerning the extended real line.

Definition 0.1. (1) Let ∞ and $-\infty$ be two elements distinct from real numbers. The *extended real line* is defined as the set $[-\infty, \infty] := \{-\infty\} \cup \mathbb{R} \cup \{\infty\}$. The canonical order relation $\leq$ on $\mathbb{R}$ is naturally extended to $[-\infty, \infty]$ by defining $a \leq \infty$ and $-\infty \leq a$ for any $a \in [-\infty, \infty]$. For $a, b \in [-\infty, \infty]$, we write $a < b$ if and only if $a \leq b$ and

$a \neq b$, as usual. For $a, b \in [-\infty, \infty]$, we set $(a, b) := \{x \in [-\infty, \infty] \mid a < x < b\}$ and $[a, b] := \{x \in [-\infty, \infty] \mid a \leq x \leq b\}$, and $(a, b]$ and $[a, b)$ are similarly defined.
(2) We say that a sequence $\{a_n\}_{n=1}^{\infty} \subset [-\infty, \infty]$ *converges to* ∞ (resp. *to* $-\infty$)[5], and write $\lim_{n\to\infty} a_n = \infty$ (resp. $\lim_{n\to\infty} a_n = -\infty$), if and only if for any $b \in \mathbb{R}$ there exists $N \in \mathbb{N}$ such that $a_n \geq b$ (resp. $a_n \leq b$) for any $n \geq N$.

The convergence of $\{a_n\}_{n=1}^{\infty}$ to a real number $a \in \mathbb{R}$ is defined in the usual manner.

Below we state basic definitions and facts concerning $[-\infty, \infty]$ without proofs. Their proofs are left to the reader as exercises.

Proposition 0.2. *Let $A \subset [-\infty, \infty]$ be non-empty. Then the supremum (least upper bound)* $\sup A$ *and the infimum (greatest lower bound)* $\inf A$ *of A exist in* $[-\infty, \infty]$.[6]

Proposition 0.3. *Let $\{a_n\}_{n=1}^{\infty} \subset [-\infty, \infty]$ and suppose that $a_n \leq a_{n+1}$ for any $n \in \mathbb{N}$. Then* $\lim_{n\to\infty} a_n = \sup_{n\geq 1} a_n$.

Definition 0.4. For $\{a_n\}_{n=1}^{\infty} \subset [-\infty, \infty]$, we define its *upper limit* $\limsup_{n\to\infty} a_n$ and its *lower limit* $\liminf_{n\to\infty} a_n$ by

$$\limsup_{n\to\infty} a_n := \inf_{n\geq 1}\left(\sup_{k\geq n} a_k\right), \qquad \liminf_{n\to\infty} a_n := \sup_{n\geq 1}\left(\inf_{k\geq n} a_k\right).$$

Clearly $\liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n$. Since the set $\{a_k \mid k \geq n\}$ is decreasing in n, $\sup_{k\geq n} a_k$ is non-increasing in n and $\inf_{k\geq n} a_k$ is non-decreasing in n, so that

$$\lim_{n\to\infty} \sup_{k\geq n} a_k = \limsup_{n\to\infty} a_n, \qquad \lim_{n\to\infty} \inf_{k\geq n} a_k = \liminf_{n\to\infty} a_n.$$

Proposition 0.5. *Let $\{a_n\}_{n=1}^{\infty} \subset [-\infty, \infty]$. Then $\lim_{n\to\infty} a_n$ exists in $[-\infty, \infty]$ (i.e. $\lim_{n\to\infty} a_n = a$ for some $a \in [-\infty, \infty]$) if and only if*

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

Moreover, if $\lim_{n\to\infty} a_n$ exists in $[-\infty, \infty]$ then $\limsup_{n\to\infty} a_n = \lim_{n\to\infty} a_n$.

Definition 0.6. The addition $+$ and the product $\cdot$ in $\mathbb{R}$ are extended to $[-\infty, \infty]$ by setting

$$\begin{aligned} a + \infty = \infty + a &:= \infty && \text{for } a \in (-\infty, \infty], \\ a + (-\infty) = -\infty + a &:= -\infty && \text{for } a \in [-\infty, \infty), \\ a \cdot \infty = \infty \cdot a &:= \begin{cases} \infty & \text{if } a \in (0, \infty], \\ 0 & \text{if } a = 0, \\ -\infty & \text{if } a \in [-\infty, 0), \end{cases} \end{aligned}$$

[5] "resp." is an abbreviation for "respectively".

[6] The supremum and infimum in $[-\infty, \infty]$ are defined in the same way as those in $\mathbb{R}$. To be precise, the *supremum* of $A \subset [-\infty, \infty]$ is a number $M \in [-\infty, \infty]$ such that $a \leq M$ for any $a \in A$ and $M \leq b$ whenever $b \in [-\infty, \infty]$ satisfies $a \leq b$ for any $a \in A$. Such M, if exists, is clearly unique. The *infimum* of A is similarly defined and, if exists, unique. Proposition 0.2 asserts that they *always* exist.

$$a\cdot(-\infty)=(-\infty)\cdot a:=\begin{cases}-\infty & \text{if } a\in(0,\infty],\\ 0 & \text{if } a=0,\\ \infty & \text{if } a\in[-\infty,0).\end{cases}$$

We also set $-(\infty):=-\infty$ and $-(-\infty):=\infty$.

Note that $\infty+(-\infty)$ and $-\infty+\infty$ are NOT defined. It may look strange to define $0\cdot\infty:=0$, but with this convention we can easily verify the following proposition.

Proposition 0.7 (Arithmetic in $[0,\infty]$)**.** (1) *Let* $a,b,c\in[0,\infty]$. *Then*

$$\begin{gathered}a+0=0+a=a,\quad a+b=b+a,\quad (a+b)+c=a+(b+c),\\ a\cdot 1=1\cdot a=a,\quad ab=ba,\quad (ab)c=a(bc),\\ a(b+c)=ab+ac,\quad (a+b)c=ac+bc.\end{gathered}$$

(2) *If* $\{a_n\}_{n=1}^\infty,\{b_n\}_{n=1}^\infty\subset[0,\infty]$ *satisfy* $a_n\le a_{n+1}$ *and* $b_n\le b_{n+1}$ *for any* $n\in\mathbb{N}$, *then*

$$\lim_{n\to\infty}a_nb_n=\Bigl(\lim_{n\to\infty}a_n\Bigr)\Bigl(\lim_{n\to\infty}b_n\Bigr).$$

Remark 0.8. It also holds that $a\cdot 1=1\cdot a=a$, $ab=ba$ and $(ab)c=a(bc)$ for any $a,b,c\in[-\infty,\infty]$.

Definition 0.9. The *sum* $\sum_{n=1}^\infty a_n$ of a non-negative sequence $\{a_n\}_{n=1}^\infty\subset[0,\infty]$ is defined as

$$\sum_{n=1}^\infty a_n:=\lim_{n\to\infty}\sum_{i=1}^n a_i=\sup_{n\in\mathbb{N}}\sum_{i=1}^n a_i=\sup_{A\subset\mathbb{N}:\text{ finite}}\sum_{n\in A}a_n.^7 \tag{0.3}$$

Note that, by the last equality in (0.3), the sum $\sum_{n=1}^\infty a_n$ of $\{a_n\}_{n=1}^\infty\subset[0,\infty]$ remains the same even if the order of $\{a_n\}_{n=1}^\infty$ is changed.

Proposition 0.10. *Let* $\{a_{n,k}\}_{n,k=1}^\infty\subset[0,\infty]$, *and let* $\mathbb{N}\ni\ell\mapsto(n_\ell,k_\ell)\in\mathbb{N}\times\mathbb{N}$ *be a bijection. Then*

$$\begin{aligned}\sum_{n=1}^\infty\sum_{k=1}^\infty a_{n,k}&=\sum_{k=1}^\infty\sum_{n=1}^\infty a_{n,k}=\sum_{\ell=1}^\infty a_{n_\ell,k_\ell}\\ &=\sup_{A\subset\mathbb{N}\times\mathbb{N}:\text{ finite}}\sum_{(n,k)\in A}a_{n,k}=:\sum_{n,k=1}^\infty a_{n,k}.\end{aligned}\tag{0.4}$$

0.4 Topology of Subsets of $\mathbb{R}^d$

We assume the reader to be familiar with the notions of open and closed subsets of the Euclidean spaces and that of continuity of maps between those sets, but it is sometimes

[7] The sum $\sum_{n\in A}a_n$ for $A=\emptyset$ is set to be 0.

useful to present the same notions in a slightly more general setting. Here we restate those topological notions for a general subset of the Euclidean spaces.

Let $d \in \mathbb{N}$. The Euclidean inner product and norm on $\mathbb{R}^d$ are denoted by $\langle \cdot, \cdot \rangle$ and $|\cdot|$, respectively: For $x, y \in \mathbb{R}^d$, $x = (x_1, \dots, x_d)$, $y = (y_1, \dots, y_d)$,

$$\langle x, y \rangle := x_1 y_1 + \cdots + x_d y_d, \qquad |x| := \sqrt{\langle x, x \rangle} = \sqrt{x_1^2 + \cdots + x_d^2}.$$

Also for $r \in (0, \infty)$ we set $B_d(x, r) := \{y \in \mathbb{R}^d \mid |y - x| < r\}$. $A \subset \mathbb{R}^d$ is called bounded if and only if $A \subset B_d(0, r)$ for some $r \in (0, \infty)$. Recall that $U \subset \mathbb{R}^d$ is called an *open subset of* $\mathbb{R}^d$ or simply *open in* $\mathbb{R}^d$ if and only if every $x \in U$ admits $\varepsilon \in (0, \infty)$ such that $B_d(x, \varepsilon) \subset U$, and that $F \subset \mathbb{R}^d$ is called *a closed subset of* $\mathbb{R}^d$ or simply *closed in* $\mathbb{R}^d$ if and only if $\mathbb{R}^d \setminus F$ is open in $\mathbb{R}^d$.

We would like to generalize these notions to the case where the whole space is not $\mathbb{R}^d$ but a subset $S \subset \mathbb{R}^d$. This is done in the following manner. Let us fix a subset S of $\mathbb{R}^d$ in the rest of this section. For $x \in S$ and $r \in (0, \infty)$, we set $B_S(x, r) := B_d(x, r) \cap S = \{y \in S \mid |y - x| < r\}$.

Definition 0.11. (1) $U \subset S$ is called an *open subset of* S or simply *open in* S if and only if every $x \in U$ admits $\varepsilon \in (0, \infty)$ such that $B_S(x, \varepsilon) \subset U$.
(2) $F \subset S$ is called a *closed subset of* S or simply *closed in* S if and only if $S \setminus F$ is open in S.

In this definition, the set $B_S(x, \varepsilon) = \{y \in S \mid |y - x| < \varepsilon\}$ plays the role of the ε-neighborhood of x. Note that these notions *depend heavily on the whole space* S. For example, $[0, 1)$ is open in $[0, 1]$ but not in $\mathbb{R}$.

We have the following simple description of open and closed subsets of S.

Proposition 0.12. *Let $A \subset S$.*
(1) *A is open in S if and only if $A = U \cap S$ for some open subset U of $\mathbb{R}^d$.*
(2) *A is closed in S if and only if $A = F \cap S$ for some closed subset F of $\mathbb{R}^d$.*

Proof. (1) "if" part is clear. Conversely suppose A is open in S. Define

$$\mathcal{I} := \{(x, \varepsilon) \in A \times (0, \infty) \mid B_S(x, \varepsilon) \subset A\}, \qquad U := \bigcup_{(x,\varepsilon) \in \mathcal{I}} B_d(x, \varepsilon).$$

Then U is open in $\mathbb{R}^d$ and $U \cap S = \bigcup_{(x,\varepsilon) \in \mathcal{I}} B_S(x, \varepsilon) \subset A$. On the other hand, since A is open in S, for any $x \in A$ there exists $\varepsilon \in (0, \infty)$ such that $B_S(x, \varepsilon) \subset A$, i.e. $(x, \varepsilon) \in \mathcal{I}$, and then $x \in B_d(x, \varepsilon) \cap A \subset U \cap S$. Thus $A \subset U \cap S$ and $A = U \cap S$.
(2) This is immediate from (1) and the definition of closed subsets of S. □

The continuity of a map is also defined in the usual way.

Definition 0.13. Let $k \in \mathbb{N}$. A map $f : S \to \mathbb{R}^k$ is called *continuous* if and only if for any $x \in S$ and any $\varepsilon \in (0, \infty)$ there exists $\delta \in (0, \infty)$ such that $|f(y) - f(x)| < \varepsilon$ for any $y \in B_S(x, \delta)$.

There are several equivalent ways of stating the continuity of a map, as follows.

Proposition 0.14. *Let $k \in \mathbb{N}$ and let $f : S \to \mathbb{R}^k$. Then f is continuous if and only if any one of the following conditions are satisfied.*
(1) *$f^{-1}(U)$ is open in S for any open subset U of $\mathbb{R}^k$.*
(2) *$f^{-1}(F)$ is closed in S for any closed subset F of $\mathbb{R}^k$.*

Proof. The conditions (1) and (2) are clearly equivalent. If (1) holds, then for $x \in S$ and $\varepsilon \in (0,\infty)$, $f^{-1}(B_k(x,\varepsilon))$ is open in S and contains x and hence $B_S(x,\delta) \subset f^{-1}(B_k(x,\varepsilon))$ for some $\delta \in (0,\infty)$. Thus $|f(y) - f(x)| < \varepsilon$ for any $y \in B_S(x,\delta)$, and f is continuous. Conversely suppose f is continuous, and let $U \subset \mathbb{R}^k$ be open in $\mathbb{R}^k$. Then for $x \in f^{-1}(U)$, there exists $\varepsilon \in (0,\infty)$ such that $B_k(f(x),\varepsilon) \subset U$, and then $|f(y) - f(x)| < \varepsilon$ for any $y \in B_S(x,\delta)$ for some $\delta \in (0,\infty)$ by the continuity of f. Thus $B_S(x,\delta) \subset f^{-1}(B_k(f(x),\varepsilon)) \subset f^{-1}(U)$ and $f^{-1}(U)$ is open in S. □

At the last of this section, we recall a basic result from multivariable calculus, which concerns the compactness of subsets of $\mathbb{R}^d$.

Definition 0.15. S is called *compact* if and only if for **any** family $\{U_\lambda\}_{\lambda\in\Lambda}$ of open subsets of $\mathbb{R}^d$ with $S \subset \bigcup_{\lambda\in\Lambda} U_\lambda$, there exists a **finite** subset Λ_0 of Λ such that $S \subset \bigcup_{\lambda\in\Lambda_0} U_\lambda$.

Theorem 0.16. *S is compact if and only if it is closed in $\mathbb{R}^d$ and bounded.*

Proof. Suppose S is compact. Then $\{B_d(0,n)\}_{n=1}^\infty$ is a family of open subsets of $\mathbb{R}^d$ with $S \subset \mathbb{R}^d = \bigcup_{n=1}^\infty B_d(0,n)$ and hence $S \subset \bigcup_{n\in I} B_d(0,n)$ for some finite set $I \subset \mathbb{N}$ by compactness. Setting $n := \max I$, we obtain $S \subset B_d(0,n)$, i.e. S is bounded. To prove that $\mathbb{R}^d \setminus S$ is open in $\mathbb{R}^d$, let $x \in \mathbb{R}^d \setminus S$. Then

$$S \subset \bigcup_{y\in S} B_d\left(y, \frac{|y-x|}{2}\right) \quad \text{and hence by compactness,} \quad S \subset \bigcup_{y\in F} B_d\left(y, \frac{|y-x|}{2}\right)$$

for some finite set $F \subset S$. Let $r := \min_{y\in F} \frac{|y-x|}{2}$. Then $B_d(x,r) \cap B_d\big(y, \frac{|y-x|}{2}\big) = \emptyset$ for any $y \in F$, which and $S \subset \bigcup_{y\in F} B_d\big(y, \frac{|y-x|}{2}\big)$ imply $B_d(x,r) \cap S = \emptyset$, i.e. $B_d(x,r) \subset \mathbb{R}^d \setminus S$. Thus $\mathbb{R}^d \setminus S$ is open in $\mathbb{R}^d$ and S is closed in $\mathbb{R}^d$.

For the converse, assume that S is closed in $\mathbb{R}^d$ and bounded. Suppose S is not compact, so that there exists a family $\{U_\lambda\}_{\lambda\in\Lambda}$ of open subsets of $\mathbb{R}^d$ with $S \subset \bigcup_{\lambda\in\Lambda} U_\lambda$ such that $S \not\subset \bigcup_{\lambda\in\Lambda_0} U_\lambda$ for any finite sunset Λ_0 of Λ. Since

$$\bigcup_{\lambda\in\Lambda} U_\lambda = \bigcup \big\{ B_d(x,r) \bigm| x \in \mathbb{Q}^d,\ r \in \mathbb{Q}\cap(0,\infty),\ B_d(x,r) \subset U_\lambda \text{ for some } \lambda \in \Lambda \big\}$$

and the family of balls $B_d(x,r)$ in the right-hand side is countable, we may assume that Λ is countably infinite, or more specifically, $\Lambda = \mathbb{N}$. Choose $x_n \in S\setminus\bigcup_{i=1}^n U_i$ for each $n \in \mathbb{N}$. Then $\{x_n\}_{n=1}^\infty \subset S$, and since S is bounded, the Bolzano-Weierstrass theorem implies that there exist $x \in \mathbb{R}^d$ and a strictly increasing sequence $\{n_k\}_{k=1}^\infty \subset \mathbb{N}$ such that $\lim_{k\to\infty} x_{n_k} = x$. For each $n \in \mathbb{N}$ and for sufficiently large k, x_{n_k} belongs to $S \setminus \bigcup_{i=1}^n U_i$. Since this set is closed in $\mathbb{R}^d$, it follows that $x \in S \setminus \bigcup_{i=1}^n U_i$ for any $n \in \mathbb{N}$ and hence that $x \in S \setminus \bigcup_{n=1}^\infty U_n$, which contradicts $S \subset \bigcup_{n=1}^\infty U_n$. Therefore S is compact. □

Exercises

Problem 0.1. Prove Propositions 0.2, 0.3 and 0.5.

Problem 0.2. (1) Let $A \subset [-\infty, \infty]$ be non-empty. Prove that $\sup(-A) = -\inf A$, where $-A := \{-a \mid a \in A\}$.
(2) Let $\{a_n\}_{n=1}^{\infty} \subset [-\infty, \infty]$. Prove that $\limsup_{n\to\infty}(-a_n) = -\liminf_{n\to\infty} a_n$.

Problem 0.3. Let $\{a_n\}_{n=1}^{\infty}, \{b_n\}_{n=1}^{\infty} \subset [-\infty, \infty]$.
(1) Suppose $a_n \le b_n$ for any $n \in \mathbb{N}$. Prove that

$$\limsup_{n\to\infty} a_n \le \limsup_{n\to\infty} b_n \qquad \text{and} \qquad \liminf_{n\to\infty} a_n \le \liminf_{n\to\infty} b_n.$$

(2) Suppose that $\{\limsup_{n\to\infty} a_n, \limsup_{n\to\infty} b_n\} \neq \{\infty, -\infty\}$ and that $\{a_n, b_n\} \neq \{\infty, -\infty\}$ for any $n \in \mathbb{N}$. Prove that

$$\limsup_{n\to\infty}(a_n + b_n) \le \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \tag{0.5}$$

and that the equality holds in (0.5) if $\lim_{n\to\infty} a_n$ exists in $[-\infty, \infty]$. Give an example of $\{a_n\}_{n=1}^{\infty}, \{b_n\}_{n=1}^{\infty} \subset [0, 1]$ for which the strict inequality holds in (0.5).

Problem 0.4. Prove Proposition 0.7 and the assertion in Remark 0.8.

Problem 0.5. Prove the latter two equalities in (0.3) and the first three equalities in (0.4).

Part I

Measure Theory

Chapter 1

Measure and Integration

In this chapter, we introduce the notion of (countably additive) measures and develop the theory of integration with respect to measures. We follow the presentation of [7, Chapter 1] for the most part of this chapter.

1.1 σ-Algebras and Measures

We start with the definition of σ-algebras.

Definition 1.1 (σ-algebras)**.** (1) Let X be a set and let $\mathcal{M} \subset 2^X$. $\mathcal{M}$ is called a *σ-algebra in X* (or a *σ-field in X*) if and only if it possesses the following properties:

(σ1) $\emptyset \in \mathcal{M}$.

(σ2) If $A \in \mathcal{M}$ then $A^c \in \mathcal{M}$, where $A^c := X \setminus A$.

(σ3) If $\{A_n\}_{n=1}^{\infty} \subset \mathcal{M}$ then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{M}$.

(2) The pair $(X, \mathcal{M})$ of a set X and a σ-algebra $\mathcal{M}$ in X is called a *measurable space*, and then a set $A \in \mathcal{M}$ is often called a *measurable set* in X.

Proposition 1.2. *Let* $(X, \mathcal{M})$ *be a measurable space. Then*
(1) $X \in \mathcal{M}$.
(2) *If* $\{A_n\}_{n=1}^{\infty} \subset \mathcal{M}$ *then* $\bigcap_{n=1}^{\infty} A_n \in \mathcal{M}$.
(3) *If* $n \in \mathbb{N}$ *and* $\{A_i\}_{i=1}^{n} \subset \mathcal{M}$ *then* $A_1 \cup \cdots \cup A_n \in \mathcal{M}$ *and* $A_1 \cap \cdots \cap A_n \in \mathcal{M}$.
(4) *If* $A, B \in \mathcal{M}$ *then* $A \setminus B \in \mathcal{M}$.

Proof. (1) $X = \emptyset^c \in \mathcal{M}$ by (σ1) and (σ2).
(2) Since $\{A_n^c\}_{n=1}^{\infty} \subset \mathcal{M}$ by (σ2), $\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c\right)^c \in \mathcal{M}$ by (σ3) and (σ2).
(3) Setting $A_i = \emptyset$ for $i \geq n+1$ and an application of (σ3) yield $A_1 \cup \cdots \cup A_n \in \mathcal{M}$. Then $A_1 \cap \cdots \cap A_n \in \mathcal{M}$ follows in exactly the same way as (2).
(4) Since $B^c \in \mathcal{M}$ by (σ2), $A \setminus B = A \cap B^c \in \mathcal{M}$ by (3). □

Definition 1.3 (Measures)**.** Let $(X, \mathcal{M})$ be a measurable space.
(1) A function $\mu : \mathcal{M} \to [0, \infty]$ is called a *measure on* $\mathcal{M}$ (or *on* $(X, \mathcal{M})$) if and only if $\mu(\emptyset) = 0$ and μ is *countably additive*, that is,

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n) \tag{1.1}$$

whenever $\{A_n\}_{n=1}^{\infty} \subset \mathcal{M}$ and $A_i \cap A_j = \emptyset$ for any $i, j \in \mathbb{N}$ with $i \neq j$. If $\mu(X) = 1$ in addition, then μ is called a *probability measure*.
(2) The triple $(X, \mathcal{M}, \mu)$ of a set X, a σ-algebra $\mathcal{M}$ in X and a measure μ on $\mathcal{M}$ is called a *measure space*. If μ is a probability measure in addition, then $(X, \mathcal{M}, \mu)$ is called a *probability space*.

Proposition 1.4. *Let $(X, \mathcal{M}, \mu)$ be a measure space.*
(1) *If $n \in \mathbb{N}$, $\{A_i\}_{i=1}^{n} \subset \mathcal{M}$ and $A_i \cap A_j = \emptyset$ for any $i, j \in \{1, \ldots, n\}$ with $i \neq j$, then $\mu(A_1 \cup \cdots \cup A_n) = \mu(A_1) + \cdots + \mu(A_n)$.*
(2) *If $A, B \in \mathcal{M}$ and $A \subset B$ then $\mu(A) \leq \mu(B)$.*
(3) *If $\{A_n\}_{n=1}^{\infty} \subset \mathcal{M}$ satisfies $A_n \subset A_{n+1}$ for any $n \in \mathbb{N}$, then $\lim_{n\to\infty} \mu(A_n) = \mu\big(\bigcup_{n=1}^{\infty} A_n\big)$.*
(4) *If $\{A_n\}_{n=1}^{\infty} \subset \mathcal{M}$ satisfies $A_n \supset A_{n+1}$ for any $n \in \mathbb{N}$ and $\mu(A_1) < \infty$, then $\lim_{n\to\infty} \mu(A_n) = \mu\big(\bigcap_{n=1}^{\infty} A_n\big)$.*

Proof. (1) This follows by letting $A_i := \emptyset$ for $i \geq n + 1$ in (1.1) and using $\mu(\emptyset) = 0$.
(2) Since $B = A \cup (B \setminus A)$ and $A \cap (B \setminus A) = \emptyset$, (1) yields $\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A)$.
(3) Set $B_1 := A_1$ and $B_n := A_n \setminus A_{n-1}$ for $n \geq 2$. Then $B_n \in \mathcal{M}$, $B_i \cap B_j = \emptyset$ for $i, j \in \mathbb{N}$ with $i \neq j$, and $A_n = B_1 \cup \cdots \cup B_n$, so that $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$. Hence

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n) = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(B_i) = \lim_{n\to\infty} \mu(A_n)$$

by (1.1) and (1) above.
(4) Set $C_n := A_1 \setminus A_n$. Then $C_n \in \mathcal{M}$ and $C_n \subset C_{n+1}$ for any $n \in \mathbb{N}$, and $\bigcup_{n=1}^{\infty} C_n = A_1 \setminus \big(\bigcap_{n=1}^{\infty} A_n\big)$. Therefore $\mu(A_n) + \mu(C_n) = \mu(A_1) = \mu\big(\bigcap_{n=1}^{\infty} A_n\big) + \mu\big(\bigcup_{n=1}^{\infty} C_n\big)$ by (1), and hence $\mu(A_1) < \infty$ and (3) together yield

$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \mu(A_1) - \lim_{n\to\infty} \mu(C_n) = \lim_{n\to\infty} \Big(\mu(A_1) - \mu(C_n)\Big) = \lim_{n\to\infty} \mu(A_n).$$

This completes the proof. □

Here are some simple examples of measures.

Example 1.5. Let X be a set. Note that 2^X is clearly a σ-algebra in X.
(1) For $A \subset X$, let $\#A$ denote its cardinality, i.e. $\#A$ is the number of the elements of A if A is a finite set and otherwise $\#A := \infty$. The function $\# : 2^X \to [0, \infty]$ is easily seen to be a measure on $(X, 2^X)$ and called the *counting measure on* X.
(2) Fix $x \in X$, and define $\delta_x : 2^X \to [0, 1]$ by $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ if $x \notin A$. Then δ_x is a measure on $(X, 2^X)$ and called the *unit mass at* x.

For measures on countable sets, we have the following clear picture.

Example 1.6. Let X be a *countable* (i.e. either finite or countably infinite) set. Then any $[0,\infty]$-valued function $\varphi : X \to [0,\infty]$ defines a measure μ_φ on $(X, 2^X)$ given by

$$\mu_\varphi(A) := \sum_{x\in A} \varphi(x). \tag{1.2}$$

Conversely, for any measure μ on $(X, 2^X)$, there exists a unique $\varphi : X \to [0,\infty]$ such that $\mu = \mu_\varphi$; it suffices to set $\varphi(x) := \mu(\{x\})$. In other words, *a measure on a countable set is completely characterized by its values on one-point sets.*[1]

The construction of interesting measures requires some (heavy) task and will be treated in Chapter 2. Here we present two fundamental examples, for which we need the following proposition.

Proposition 1.7. *Let X be a set.*
(1) *Let Λ be a non-empty set and suppose that $\mathcal{M}_\lambda$ is a σ-algebra in X for each $\lambda \in \Lambda$. Then $\bigcap_{\lambda\in\Lambda} \mathcal{M}_\lambda$ is a σ-algebra in X.*
(2) *Let $\mathcal{A} \subset 2^X$ and set*

$$\sigma_X(\mathcal{A}) := \bigcap_{\mathcal{M}:\ \sigma\text{-algebra in } X,\ \mathcal{A}\subset\mathcal{M}} \mathcal{M}. \tag{1.3}$$

Then $\sigma_X(\mathcal{A})$ is the smallest σ-algebra in X that includes $\mathcal{A}$.

$\sigma_X(\mathcal{A})$ in (1.3) is called the σ-algebra in X *generated by* $\mathcal{A}$, and it is simply denoted as $\sigma(\mathcal{A})$ when no confusion can occur.

Proof. (1) We verify that $\bigcap_{\lambda\in\Lambda} \mathcal{M}_\lambda$ satisfies the conditions (σ1), (σ2) and (σ3) in Definition 1.1-(1). $\emptyset \in \mathcal{M}_\lambda$ for any $\lambda \in \Lambda$ and hence $\emptyset \in \bigcap_{\lambda\in\Lambda} \mathcal{M}_\lambda$. If $A \in \bigcap_{\lambda\in\Lambda} \mathcal{M}_\lambda$ then $A \in \mathcal{M}_\lambda$ and hence $A^c \in \mathcal{M}_\lambda$ for any $\lambda \in \Lambda$. Thus $A^c \in \bigcap_{\lambda\in\Lambda} \mathcal{M}_\lambda$. If $\{A_n\}_{n=1}^\infty \subset \bigcap_{\lambda\in\Lambda} \mathcal{M}_\lambda$, then for any $\lambda \in \Lambda$, $\{A_n\}_{n=1}^\infty \subset \mathcal{M}_\lambda$ and hence $\bigcup_{n=1}^\infty A_n \in \mathcal{M}_\lambda$. Thus $\bigcup_{n=1}^\infty A_n \in \bigcap_{\lambda\in\Lambda} \mathcal{M}_\lambda$.
(2) Since 2^X is a σ-algebra in X including $\mathcal{A}$, we can take the intersection given in (1.3) to define $\sigma_X(\mathcal{A})$. Then (1) shows that $\sigma_X(\mathcal{A})$ is a σ-algebra in X. By definition, $\mathcal{A} \subset \sigma_X(\mathcal{A})$, and $\sigma_X(\mathcal{A}) \subset \mathcal{M}$ for any σ-algebra $\mathcal{M}$ in X with $\mathcal{A} \subset \mathcal{M}$. □

Example 1.8 (Borel σ-algebra and Lebesgue measure on $\mathbb{R}^d$)**.** Let $d \in \mathbb{N}$. We define the *Borel σ-algebra* $\mathcal{B}(\mathbb{R}^d)$ of $\mathbb{R}^d$ to be the σ-algebra in $\mathbb{R}^d$ generated by its open subsets, i.e.

$$\mathcal{B}(\mathbb{R}^d) := \sigma\bigl(\{U \subset \mathbb{R}^d \mid U \text{ is open in } \mathbb{R}^d\}\bigr). \tag{1.4}$$

Then each $A \in \mathcal{B}(\mathbb{R}^d)$ is called a *Borel set of* $\mathbb{R}^d$. In fact, as stated in the following proposition, $\mathcal{B}(\mathbb{R}^d)$ is generated by d-dimensional intervals. As we will see in the

[1] Here we could consider a σ-algebra $\mathcal{M}$ in X which differs from 2^X, but then for some $x \in X$ we would have $\{x\} \notin \mathcal{M}$ (the one-point set $\{x\}$ is *not* measurable), which looks very weird for a countable set X. This is why we considered measures on 2^X only.

course of this lecture, $\mathcal{B}(\mathbb{R}^d)$ is the right σ-algebra to be considered when dealing with measures on $\mathbb{R}^d$ and $\mathbb{R}^d$-valued functions.

Later we will see many examples of measures defined on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, but here we present only the most standard and most important one: *there exists a unique measure* m_d *on* $\mathcal{B}(\mathbb{R}^d)$ *such that for any* d*-dimensional interval* $[a_1, b_1] \times \cdots \times [a_d, b_d]$,

$$\mathrm{m}_d\bigl([a_1, b_1] \times \cdots \times [a_d, b_d]\bigr) = (b_1 - a_1) \cdots (b_d - a_d). \tag{1.5}$$

m_d is called the *Lebesgue measure on* $\mathbb{R}^d$.[2] This is the mathematically correct formulation of the notion of *"*d*-dimensional volume"*; m_1, m_2 and m_3 represent *length*, *area* and *volume*, respectively.

We need rather long preparations for the proof of the existence and uniqueness, especially existence, of such a measure and we will treat it in the next chapter.

Proposition 1.9. *Let* $d \in \mathbb{N}$ *and define*

$$\mathcal{F}_d := \bigl\{[a_1, b_1] \times \cdots \times [a_d, b_d] \bigm| a_k, b_k \in \mathbb{R}, a_k \le b_k \text{ for } 1 \le k \le d\bigr\} \cup \{\emptyset\}, \tag{1.6}$$

$$\mathcal{F}_d^{\mathbb{Q}} := \bigl\{[a_1, b_1] \times \cdots \times [a_d, b_d] \bigm| a_k, b_k \in \mathbb{Q}, a_k \le b_k \text{ for } 1 \le k \le d\bigr\} \cup \{\emptyset\}. \tag{1.7}$$

Then $\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{F}_d) = \sigma\bigl(\mathcal{F}_d^{\mathbb{Q}}\bigr)$.

Proof. $\mathcal{F}_d^{\mathbb{Q}} \subset \mathcal{F}_d$ by definition, and we also have $\mathcal{F}_d \subset \mathcal{B}(\mathbb{R}^d)$ since any $I \in \mathcal{F}_d$ is closed in $\mathbb{R}^d$ and hence $I^c \in \mathcal{B}(\mathbb{R}^d)$. Thus $\sigma\bigl(\mathcal{F}_d^{\mathbb{Q}}\bigr) \subset \sigma(\mathcal{F}_d) \subset \mathcal{B}(\mathbb{R}^d)$. Let U be an open subset of $\mathbb{R}^d$. For the proof of $\mathcal{B}(\mathbb{R}^d) \subset \sigma\bigl(\mathcal{F}_d^{\mathbb{Q}}\bigr)$, it suffices to show $U \in \sigma\bigl(\mathcal{F}_d^{\mathbb{Q}}\bigr)$. Set

$$\mathcal{A} := \bigl\{I \in \mathcal{F}_d^{\mathbb{Q}} \bigm| I \subset U\bigr\}.$$

Since $\mathcal{F}_d^{\mathbb{Q}}$ is countable, so is $\mathcal{A}$ and hence $\bigcup_{I \in \mathcal{A}} I \in \sigma\bigl(\mathcal{F}_d^{\mathbb{Q}}\bigr)$. Clearly $\bigcup_{I \in \mathcal{A}} I \subset U$. On the other hand, any $x \in U$ admits $I \in \mathcal{A}$ such that $x \in I$; indeed, since U is open, there exists $r \in (0, \infty)$ such that $B(x, \sqrt{d}r) = \{y \in \mathbb{R}^d \mid |y - x| < \sqrt{d}r\} \subset U$. If we choose $a_k, b_k \in \mathbb{Q}$ so that $x_k - r < a_k \le x_k \le b_k < x_k + r$, where $x = (x_1, \ldots, x_d)$, then $I := [a_1, b_1] \times \cdots \times [a_d, b_d]$ satisfies $x \in I$, $I \subset U$ and hence $I \in \mathcal{A}$. Therefore $x \in \bigcup_{I \in \mathcal{A}} I$, thus $U \subset \bigcup_{I \in \mathcal{A}} I$ and hence $U = \bigcup_{I \in \mathcal{A}} I \in \sigma\bigl(\mathcal{F}_d^{\mathbb{Q}}\bigr)$. □

The following lemma is sometimes useful.

Lemma 1.10. *Let* X *be a set, let* $\mathcal{A} \subset 2^X$ *and let* $Y \subset X$. *Define* $\mathcal{A}|_Y \subset 2^Y$ *by*

$$\mathcal{A}|_Y := \{A \cap Y \mid A \in \mathcal{A}\}. \tag{1.8}$$

(1) *If* $\mathcal{A}$ *is a* σ*-algebra in* X*, then* $\mathcal{A}|_Y$ *is a* σ*-algebra in* Y.
(2) $\sigma_Y(\mathcal{A}|_Y) = \sigma_X(\mathcal{A})|_Y$.

[2] More precisely, the *completion of* m_d, which is an extension of m_d to a certain larger σ-algebra, is usually called the Lebesgue measure on $\mathbb{R}^d$; see Theorem 1.37 below for the notion of completion.

Proof. (1) Suppose $\mathcal{A}$ is a σ-algebra in X. Then $\emptyset = \emptyset \cap Y \in \mathcal{A}|_Y$, and $Y \setminus (A \cap Y) = A^c \cap Y \in \mathcal{A}|_Y$ for any $A \in \mathcal{A}$. If $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$, then $\bigcup_{n=1}^\infty (A_n \cap Y) = Y \cap \bigcup_{n=1}^\infty A_n \in \mathcal{A}|_Y$. Thus $\mathcal{A}|_Y$ is a σ-algebra in Y.
(2) $\sigma_Y(\mathcal{A}|_Y) \subset \sigma_X(\mathcal{A})|_Y$ follows since $\mathcal{A}|_Y \subset \sigma_X(\mathcal{A})|_Y$ and $\sigma_X(\mathcal{A})|_Y$ is a σ-algebra in Y by (1). For the converse, let $\mathcal{B} := \{A \subset X \mid A \cap Y \in \sigma_Y(\mathcal{A}|_Y)\}$. Then $\mathcal{A} \subset \mathcal{B}$, and it is immediate to see that $\mathcal{B}$ is a σ-algebra in X. Thus $\sigma_X(\mathcal{A}) \subset \mathcal{B}$, that is, $\sigma_X(\mathcal{A})|_Y \subset \sigma_Y(\mathcal{A}|_Y)$. □

Example 1.11 (Borel σ-algebra in subsets of $\mathbb{R}^d$). Let $d \in \mathbb{N}$ and $S \subset \mathbb{R}^d$. Then the *Borel σ-algebra $\mathcal{B}(S)$ of S* is defined in the same way as that of $\mathbb{R}^d$, i.e.

$$\mathcal{B}(S) := \sigma\big(\{U \subset S \mid U \text{ is open in } S\}\big), \tag{1.9}$$

and each $A \in \mathcal{B}(S)$ is called a *Borel set of S*. Since Proposition 0.12 means that

$$\{U \subset S \mid U \text{ is open in } S\} = \{U \subset \mathbb{R}^d \mid U \text{ is open in } \mathbb{R}^d\}|_S,$$

an application of Lemma 1.10 shows that

$$\mathcal{B}(S) = \mathcal{B}(\mathbb{R}^d)|_S. \tag{1.10}$$

In particular, if $S \in \mathcal{B}(\mathbb{R}^d)$, then $\mathcal{B}(S) = \{A \in \mathcal{B}(\mathbb{R}^d) \mid A \subset S\} \subset \mathcal{B}(\mathbb{R}^d)$.

Example 1.12 (Bernoulli measures). Let $\Omega := \{0,1\}^{\mathbb{N}} = \big\{(\omega_n)_{n=1}^\infty \mid \omega_n \in \{0,1\}\big\}$. If we write 0 for tails of a coin flip and 1 for heads, then the outcome of infinitely many coin flips is represented by a sequence $\omega = (\omega_n)_{n=1}^\infty \in \Omega$, where ω_n corresponds to the n-th outcome, and therefore Ω is a natural choice of the sample space for infinitely many coin flips.

Which σ-algebra should we equip Ω with? An obvious requirement is that any "event" *determined only by the outcomes of finitely many flips*, i.e. any subset of the form $A_n \times \{0,1\}^{\mathbb{N}\setminus\{1,\ldots,n\}}$ with $A_n \subset \{0,1\}^n$, should be measurable. Therefore an easy choice is to consider the following σ-algebra $\mathcal{F}$:

$$\mathcal{F} := \sigma\Big(\big\{A_n \times \{0,1\}^{\mathbb{N}\setminus\{1,\ldots,n\}} \,\big|\, n \in \mathbb{N},\ A_n \subset \{0,1\}^n\big\}\Big). \tag{1.11}$$

$\mathcal{F}$ is actually the right σ-algebra in Ω to be considered, and we can construct a natural probability measure on $\mathcal{F}$ which represents the randomness of infinitely many flips of a coin: *for any $p \in [0,1]$,*[3] *there exists a unique probability measure $\mathbb{P}_p$ on $\mathcal{F}$ such that*[4]

$$\mathbb{P}_p\big[\{(\omega_i)_{i=1}^n\} \times \{0,1\}^{\mathbb{N}\setminus\{1,\ldots,n\}}\big] = \prod_{i=1}^n p^{\omega_i}(1-p)^{1-\omega_i} \tag{1.12}$$

for any $n \in \mathbb{N}$ and any $(\omega_i)_{i=1}^n \in \{0,1\}^n$. $\mathbb{P}_p$ is called the *Bernoulli measure on Ω of probability p*. The proof of its existence and uniqueness is postponed until later chapters.

[3]The number p corresponds to the probability of heads at each flip.

[4]Here $0^0 := 0$.

1.2 Measurable and Simple Functions

In this section, we define measurable functions and present their basic properties. Throughout this section, we fix a measurable space $(X, \mathcal{M})$.

Definition 1.13 (Measurable functions)**.** A function $f : X \to [-\infty, \infty]$ is called $\mathcal{M}$-*measurable* if and only if $f^{-1}(A) \in \mathcal{M}$ for any $A \in \mathcal{B}(\mathbb{R})$ and for $A = \{\infty\}, \{-\infty\}$.

Proposition 1.14. *A function $f : X \to [-\infty, \infty]$ is $\mathcal{M}$-measurable if and only if $f^{-1}\big((a, \infty]\big) \in \mathcal{M}$ for any $a \in \mathbb{Q}$ (or equivalently, for any $a \in \mathbb{R}$).*

Proof. If f is $\mathcal{M}$-measurable, then $f^{-1}\big((a,\infty]\big) = f^{-1}\big((a,\infty)\big) \cup f^{-1}(\infty) \in \mathcal{M}$ for any $a \in \mathbb{R}$. For the converse, suppose $f^{-1}\big((a,\infty]\big) \in \mathcal{M}$ for any $a \in \mathbb{Q}$. Then $f^{-1}(\infty) = \bigcap_{n\in\mathbb{N}} f^{-1}\big((n,\infty]\big) \in \mathcal{M}$, $f^{-1}\big((-\infty,\infty]\big) = \bigcup_{n\in\mathbb{N}} f^{-1}\big((-n,\infty]\big) \in \mathcal{M}$ and hence $f^{-1}(-\infty) = f^{-1}\big((-\infty,\infty]\big)^c \in \mathcal{M}$. We claim that

$$\mathcal{A} := \{A \subset \mathbb{R} \mid f^{-1}(A) \in \mathcal{M}\} \text{ is a } \sigma\text{-algebra in } \mathbb{R} \text{ and satisfies } \mathcal{F}_1^{\mathbb{Q}} \subset \mathcal{A}, \tag{1.13}$$

where $\mathcal{F}_1^{\mathbb{Q}}$ is given by (1.7) with $d = 1$. Then (1.13) yields $\mathcal{B}(\mathbb{R}) = \sigma\big(\mathcal{F}_1^{\mathbb{Q}}\big) \subset \mathcal{A}$.

Let us verify (1.13). $f^{-1}(\emptyset) = \emptyset \in \mathcal{M}$, $f^{-1}(\mathbb{R}) = \big(f^{-1}(\infty) \cup f^{-1}(-\infty)\big)^c \in \mathcal{M}$ and hence $\emptyset, \mathbb{R} \in \mathcal{A}$. If $A \in \mathcal{A}$ then $f^{-1}(\mathbb{R} \setminus A) = f^{-1}(\mathbb{R}) \setminus f^{-1}(A) \in \mathcal{M}$ and hence $\mathbb{R} \setminus A \in \mathcal{A}$. If $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ then $f^{-1}\big(\bigcup_{n=1}^\infty A_n\big) = \bigcup_{n=1}^\infty f^{-1}(A_n) \in \mathcal{M}$ and therefore $\bigcup_{n=1}^\infty A_n \in \mathcal{A}$. Thus $\mathcal{A}$ is a σ-algebra in $\mathbb{R}$. For $a, b \in \mathbb{Q}$ with $a \le b$, $f^{-1}\big([a,\infty]\big) = \bigcap_{n=1}^\infty f^{-1}\big((a - 1/n, \infty]\big) \in \mathcal{M}$ and therefore $f^{-1}\big([a,b]\big) = f^{-1}\big([a,\infty]\big) \setminus f^{-1}\big((b,\infty]\big) \in \mathcal{M}$. Thus $[a,b] \in \mathcal{A}$, proving $\mathcal{F}_1^{\mathbb{Q}} \subset \mathcal{A}$ and (1.13). □

Proposition 1.15. *Let $f, g : X \to [-\infty, \infty]$ be $\mathcal{M}$-measurable.*
(1) The function $f + g : X \to [-\infty,\infty]$, $(f+g)(x) := f(x) + g(x)$, is $\mathcal{M}$-measurable, provided $\{f(x), g(x)\} \neq \{\infty, -\infty\}$ for any $x \in X$[5].
(2) The function $fg : X \to [-\infty,\infty]$, $(fg)(x) := f(x)g(x)$, is $\mathcal{M}$-measurable.

Proof. (1) For any $a \in \mathbb{R}$ we have

$$(f+g)^{-1}\big((a,\infty]\big) = \bigcup_{r,s\in\mathbb{Q},\, r+s>a} f^{-1}\big((r,\infty]\big) \cap g^{-1}\big((s,\infty]\big) \in \mathcal{M},$$

and hence $f + g$ is $\mathcal{M}$-measurable by Proposition 1.14.
(2) It holds that for any $a \in [0,\infty)$,

$$\begin{aligned}&(fg)^{-1}\big((a,\infty]\big)\\ &= \bigcup_{\substack{r,s\in\mathbb{Q}\cap(0,\infty)\\ rs>a}} \Big(f^{-1}\big((r,\infty]\big) \cap g^{-1}\big((s,\infty]\big)\Big) \cup \Big(f^{-1}\big([-\infty,-r)\big) \cap g^{-1}\big([-\infty,-s)\big)\Big)\end{aligned}$$

and therefore $(fg)^{-1}\big((a,\infty]\big) \in \mathcal{M}$. On the other hand, for any $a \in (-\infty, 0)$,

$$(fg)^{-1}\big((a,0)\big)$$

[5] that is, provided neither "$\infty + (-\infty)$" nor "$-\infty + \infty$" appears in the sum $f(x) + g(x)$

$$= \bigcup_{\substack{r,s\in\mathbb{Q}\cap(0,\infty)\\ rs<|a|}} \Big(f^{-1}\big((0,r)\big)\cap g^{-1}\big((-s,0)\big)\Big)\cup\Big(f^{-1}\big((-r,0)\big)\cap g^{-1}\big((0,s)\big)\Big)$$
$$\in \mathcal{M}$$

and hence $(fg)^{-1}\big((a,\infty]\big) = (fg)^{-1}\big((a,0)\big)\cup(fg)^{-1}(0)\cup(fg)^{-1}\big((0,\infty]\big)\in\mathcal{M}$; note that $(fg)^{-1}(0) = f^{-1}(0)\cup g^{-1}(0)\in\mathcal{M}$. Now Proposition 1.14 implies that fg is $\mathcal{M}$-measurable. ☐

For a sequence $\{f_n\}_{n=1}^{\infty}$ of $[-\infty,\infty]$-valued functions on X, we define $[-\infty,\infty]$-valued functions $\sup_{n\geq 1} f_n$, $\inf_{n\geq 1} f_n$, $\limsup_{n\to\infty} f_n$ and $\liminf_{n\to\infty} f_n$ on X by

$$\Big(\sup_{n\geq 1} f_n\Big)(x) := \sup_{n\geq 1}\big(f_n(x)\big), \qquad \Big(\limsup_{n\to\infty} f_n\Big)(x) := \limsup_{n\to\infty}\big(f_n(x)\big),$$
$$\Big(\inf_{n\geq 1} f_n\Big)(x) := \inf_{n\geq 1}\big(f_n(x)\big), \qquad \Big(\liminf_{n\to\infty} f_n\Big)(x) := \liminf_{n\to\infty}\big(f_n(x)\big).$$

Proposition 1.16. *Let $f_n : X\to[-\infty,\infty]$ be $\mathcal{M}$-measurable for each $n\in\mathbb{N}$. Then $\sup_{n\geq 1} f_n$, $\inf_{n\geq 1} f_n$, $\limsup_{n\to\infty} f_n$ and $\liminf_{n\to\infty} f_n$ are all $\mathcal{M}$-measurable.*

Proof. For any $a\in\mathbb{R}$, $(\sup_{n\geq 1} f_n)^{-1}\big((a,\infty]\big) = \bigcup_{n=1}^{\infty} f_n^{-1}\big((a,\infty]\big)\in\mathcal{M}$, and hence $\sup_{n\geq 1} f_n$ is $\mathcal{M}$-measurable by Proposition 1.14. Then $\inf_{n\geq 1} f_n = -\sup_{n\geq 1}(-f_n)$ is also $\mathcal{M}$-measurable by Proposition 1.15-(2). In particular, $\sup_{k\geq n} f_n$ and $\inf_{k\geq n} f_k$ are $\mathcal{M}$-measurable for any $n\in\mathbb{N}$ and so are $\limsup_{n\to\infty} f_n = \inf_{n\geq 1}(\sup_{k\geq n} f_k)$ and $\liminf_{n\to\infty} f_n = \sup_{n\geq 1}(\inf_{k\geq n} f_k)$. ☐

The following lemma is useful in verifying measurability of basic functions.

Lemma 1.17. *Let $d\in\mathbb{N}$ and let $S\subset\mathbb{R}^d$. If $f : S\to\mathbb{R}$ is continuous, then f is $\mathcal{B}(S)$-measurable.*

A $\mathcal{B}(S)$-measurable function on S is also referred to as a *Borel measurable* function. Lemma 1.17 asserts that *every $\mathbb{R}$-valued continuous function is Borel measurable*.

Proof. Let $\mathcal{A} := \{A\subset\mathbb{R}\mid f^{-1}(A)\in\mathcal{B}(S)\}$. We easily see that $\mathcal{A}$ is a σ-algebra in $\mathbb{R}$, and any open subset U of $\mathbb{R}$ belongs to $\mathcal{A}$ since $f^{-1}(U)$ is open in S by the continuity of f. Thus $\mathcal{B}(\mathbb{R})\subset\mathcal{A}$, which means that f is $\mathcal{B}(S)$-measurable. ☐

For $E\subset X$, we define $\mathbf{1}_E : X\to\mathbb{R}$ by

$$\mathbf{1}_E(x) := \begin{cases} 1 & \text{if } x\in E,\\ 0 & \text{if } x\notin E.\end{cases} \tag{1.14}$$

$\mathbf{1}_E$ is called the *indicator function*[6] of E. It is easy to see that $\mathbf{1}_E$ is $\mathcal{M}$-measurable if and only if $E\in\mathcal{M}$.

[6] $\mathbf{1}_E$ is usually called the *characteristic function* of E, but in the context of probability theory, this phrase is kept for the Fourier transform of probability measures on $\mathbb{R}^d$. See Chapter 4 for details.

Definition 1.18 (Simple functions)**.** $s : X \to \mathbb{R}$ is called $\mathcal{M}$-*simple* if and only if it is $\mathcal{M}$-measurable and its range $s(X)$ is a finite set.

Note that ∞ *and* $-\infty$ *are explicitly excluded* from the values of simple functions. Since an $\mathcal{M}$-simple function s is written as $s = \sum_{a \in s(X)} a \mathbf{1}_{s^{-1}(a)}$ with $s^{-1}(a) \in \mathcal{M}$, we easily see from Proposition 1.15 that $s : X \to \mathbb{R}$ is $\mathcal{M}$-simple if and only if

$$s = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i} \qquad \text{for some } n \in \mathbb{N}, \{a_i\}_{i=1}^{n} \subset \mathbb{R} \text{ and } \{A_i\}_{i=1}^{n} \subset \mathcal{M}. \tag{1.15}$$

Proposition 1.19. *Let $f : X \to [0, \infty]$ be $\mathcal{M}$-measurable. Then there exists a sequence $\{s_n\}_{n=1}^{\infty}$ of $\mathcal{M}$-simple functions on X such that for each $x \in X$,*

(S1) $0 \le s_n(x) \le s_{n+1}(x)$ *for any* $n \in \mathbb{N}$,

(S2) $\lim_{n\to\infty} s_n(x) = f(x)$.

Proof. For $n \in \mathbb{N}$, define $s_n : X \to [0, \infty)$ by

$$s_n := \sum_{i=1}^{n \cdot 2^n} \frac{i-1}{2^n} \mathbf{1}_{\{\frac{i-1}{2^n} \le f < \frac{i}{2^n}\}} + n \mathbf{1}_{\{f \ge n\}}, \tag{1.16}$$

where $\{\frac{i-1}{2^n} \le f < \frac{i}{2^n}\} := f^{-1}\big([\frac{i-1}{2^n}, \frac{i}{2^n})\big)$ and $\{f \ge n\} := f^{-1}\big([n, \infty]\big)$. These sets belong to $\mathcal{M}$ by the $\mathcal{M}$-measurability of f and hence s_n is $\mathcal{M}$-simple.

Let $x \in X$. It is easy to verify (S1). If $f(x) < \infty$ then $f(x) - 2^{-n} < s_n(x) \le f(x)$ for $n \in \mathbb{N}$ with $n > f(x)$, and if $f(x) = \infty$ then $s_n(x) = n$ for any $n \in \mathbb{N}$. Thus $\lim_{n\to\infty} s_n(x) = f(x)$ in both cases, proving (S2). □

1.3 Integration and Convergence Theorems

In this section, we define integration with respect to measures and prove fundamental convergence theorems. Throughout this section, we fix a measure space $(X, \mathcal{M}, \mu)$.

1.3.1 Non-negative functions

First we define integration of non-negative simple functions. Recall our convention that $0 \cdot \infty = \infty \cdot 0 := 0$.

Definition 1.20 (Integration of non-negative simple functions)**.** Let $s : X \to [0, \infty)$ be $\mathcal{M}$-simple. We define its μ-*integral* $\int_X s\,d\mu$ *on* X by

$$\int_X s\,d\mu := \sum_{a \in s(X)} a\mu\big(s^{-1}(a)\big). \tag{1.17}$$

Lemma 1.21. *Let $s, t : X \to [0, \infty)$ be $\mathcal{M}$-simple and let $\alpha, \beta \in [0, \infty)$. Then*

$$\int_X (\alpha s + \beta t)\,d\mu = \alpha \int_X s\,d\mu + \beta \int_X t\,d\mu. \tag{1.18}$$

Proof. Note that $\alpha s + \beta t$ is also $[0, \infty)$-valued and $\mathcal{M}$-simple, so that the left-hand side of (1.18) is defined. It is easy to see that $\int_X \alpha s d\mu = \alpha \int_X s d\mu$; indeed, $\int_X \alpha s d\mu = 0 = \alpha \int_X s d\mu$ for $\alpha = 0$, and if $\alpha \in (0, \infty)$ then $(\alpha s)^{-1}(\alpha a) = s^{-1}(a)$ for $a \in [0, \infty)$ and hence

$$\int_X \alpha s d\mu = \sum_{a \in s(X)} (\alpha a)\mu\bigl(s^{-1}(a)\bigr) = \alpha \sum_{a \in s(X)} a\mu\bigl(s^{-1}(a)\bigr) = \alpha \int_X s d\mu.$$

Thus it suffices to show $\int_X (s + t) d\mu = \int_X s d\mu + \int_X t d\mu$. For $a \in (s + t)(X)$ we have

$$(s + t)^{-1}(a) = \bigcup_{\substack{(b,c) \in s(X) \times t(X) \\ b + c = a}} s^{-1}(b) \cap t^{-1}(c)$$

where $s^{-1}(b) \cap t^{-1}(c)$ are mutually disjoint, and therefore

$$\begin{aligned}
&\int_X (s + t) d\mu \\
&= \sum_{a \in (s+t)(X)} a\mu\bigl((s + t)^{-1}(a)\bigr) \\
&= \sum_{a \in (s+t)(X)} \sum_{\substack{(b,c) \in s(X) \times t(X) \\ b + c = a}} a\mu\bigl(s^{-1}(b) \cap t^{-1}(c)\bigr) \\
&= \sum_{b \in s(X)} \sum_{c \in t(X)} (b + c)\mu\bigl(s^{-1}(b) \cap t^{-1}(c)\bigr) \\
&= \sum_{b \in s(X)} b \sum_{c \in t(X)} \mu\bigl(s^{-1}(b) \cap t^{-1}(c)\bigr) + \sum_{c \in t(X)} c \sum_{b \in s(X)} \mu\bigl(s^{-1}(b) \cap t^{-1}(c)\bigr) \\
&= \sum_{b \in s(X)} b\mu\bigl(s^{-1}(b)\bigr) + \sum_{c \in t(X)} c\mu\bigl(t^{-1}(c)\bigr) \\
&= \int_X s d\mu + \int_X t d\mu,
\end{aligned}$$

which completes the proof. □

Definition 1.22 (Integration of non-negative functions)**.** Let $f : X \to [0, \infty]$ be $\mathcal{M}$-measurable. We define its *μ-integral $\int_X f d\mu$ on X* by

$$\int_X f d\mu := \sup\left\{ \int_X s d\mu \;\middle|\; s : X \to \mathbb{R},\, s \text{ is } \mathcal{M}\text{-simple and } 0 \le s \le f \text{ on } X \right\}. \tag{1.19}$$

Note that (1.19) is consistent with (1.17) for non-negative $\mathcal{M}$-simple functions; indeed, the supremum in (1.19) is attained by f if $f : X \to [0, \infty]$ is itself $\mathcal{M}$-simple, since we see from Lemma 1.21 that $\int_X s d\mu \le \int_X s d\mu + \int_X (t - s) d\mu = \int_X t d\mu$ for $\mathcal{M}$-simple functions $s, t : X \to [0, \infty)$ with $s \le t$ on X.

The following lemma is immediate from (1.19).

Lemma 1.23. *If* $f, g : X \to [0, \infty]$ *are* $\mathcal{M}$*-measurable and* $f \leq g$ *on* X*, then* $\int_X f d\mu \leq \int_X g d\mu$.

Now we are in the stage of presenting the first fundamental convergence theorem.

Theorem 1.24 (Monotone convergence theorem, MCT)**.** *Let* $f_n : X \to [0, \infty]$ *be* $\mathcal{M}$*-measurable for each* $n \in \mathbb{N}$ *and suppose* $f_n(x) \leq f_{n+1}(x)$ *for any* $n \in \mathbb{N}$, $x \in X$. *Then* $f : X \to [0, \infty]$ *defined by* $f(x) := \lim_{n\to\infty} f_n(x)$ *is* $\mathcal{M}$*-measurable, and*

$$\lim_{n\to\infty} \int_X f_n d\mu = \int_X f d\mu. \tag{1.20}$$

Proof. Since $f = \lim_{n\to\infty} f_n = \sup_{n\geq 1} f_n$ by Proposition 0.3, f is $\mathcal{M}$-measurable by Proposition 1.16. Let $n \in \mathbb{N}$. Then $0 \leq \int_X f_n d\mu \leq \int_X f_{n+1} d\mu \leq \int_X f d\mu$ by Lemma 1.23, and letting $n \to \infty$ yields $\lim_{n\to\infty} \int_X f_n d\mu \leq \int_X f d\mu$.

For the converse inequality, let $s : X \to \mathbb{R}$ be $\mathcal{M}$-simple and satisfy $0 \leq s \leq f$ on X. Let $c \in (0, 1)$ and define

$$A_n := \{x \in X \mid f_n(x) \geq cs(x)\} = (f_n - cs)^{-1}\big([0, \infty]\big), \qquad n \in \mathbb{N}.$$

Then for $n \in \mathbb{N}$, $A_n \in \mathcal{M}$ by Proposition 1.15, and $A_n \subset A_{n+1}$. Furthermore we have $\bigcup_{n=1}^{\infty} A_n = X$; indeed, if $x \in s^{-1}(0)$ then $f_1(x) \geq 0 = cs(x)$ and hence $x \in A_1$, and if $x \in s^{-1}\big((0, \infty)\big)$ then $f(x) = \sup_{n\geq 1} f_n(x) \geq s(x) > cs(x)$ and therefore $f_n(x) > cs(x)$ for some $n \in \mathbb{N}$. Now for $n \in \mathbb{N}$, by Lemma 1.23 and Lemma 1.21 we have

$$\begin{aligned} \int_X f_n d\mu \geq \int_X f_n \mathbf{1}_{A_n} d\mu \geq \int_X cs\mathbf{1}_{A_n} d\mu &= \int_X \sum_{a\in s(X)} ca\mathbf{1}_{s^{-1}(a)\cap A_n} d\mu \\ &= \sum_{a\in s(X)} ca \int_X \mathbf{1}_{s^{-1}(a)\cap A_n} d\mu = c \sum_{a\in s(X)} a\mu\big(s^{-1}(a) \cap A_n\big). \end{aligned} \tag{1.21}$$

Since $\lim_{n\to\infty} \mu\big(s^{-1}(a) \cap A_n\big) = \mu\big(s^{-1}(a)\big)$ by Proposition 1.4-(3), letting $n \to \infty$ in (1.21) yields $\lim_{n\to\infty} \int_X f_n d\mu \geq c \sum_{a\in s(X)} a\mu\big(s^{-1}(a)\big) = c \int_X s d\mu$, where $c \in (0, 1)$ is arbitrary, and hence $\lim_{n\to\infty} \int_X f_n d\mu \geq \int_X s d\mu$. Finally, taking the supremum over $\mathcal{M}$-simple s with $0 \leq s \leq f$ shows $\lim_{n\to\infty} \int_X f_n d\mu \geq \int_X f d\mu$. □

Proposition 1.25. *Let* $f, g : X \to [0, \infty]$ *be* $\mathcal{M}$*-measurable and let* $\alpha, \beta \in [0, \infty]$. *Then*

$$\int_X (\alpha f + \beta g) d\mu = \alpha \int_X f d\mu + \beta \int_X g d\mu. \tag{1.22}$$

Proof. Let $\{s_n\}_{n=1}^{\infty}$ and $\{t_n\}_{n=1}^{\infty}$ be sequences of non-decreasing non-negative $\mathcal{M}$-simple functions on X converging to f and g, respectively, as given in Proposition 1.19. Then by virtue of Lemma 1.21, Theorem 1.24 yields

$$\int_X \alpha f d\mu = \lim_{n\to\infty} \int_X \min\{\alpha, n\} s_n d\mu = \lim_{n\to\infty} \min\{\alpha, n\} \int_X s_n d\mu = \alpha \int_X f d\mu$$

and

$$\begin{aligned}\int_X (f+g)d\mu &= \lim_{n\to\infty}\int_X (s_n+t_n)d\mu \\ &= \lim_{n\to\infty}\left(\int_X s_n d\mu + \int_X t_n d\mu\right) = \int_X f d\mu + \int_X g d\mu,\end{aligned}$$

which together imply (1.22). □

Proposition 1.26. *Let $f_n : X \to [0,\infty]$ be $\mathcal{M}$-measurable for each $n \in \mathbb{N}$. Then*

$$\int_X \left(\sum_{n=1}^{\infty} f_n\right) d\mu = \sum_{n=1}^{\infty}\int_X f_n d\mu. \tag{1.23}$$

Proof. Since $\sum_{i=1}^{n} f_i(x)$ is non-decreasing in $n \in \mathbb{N}$ and converges to $\sum_{m=1}^{\infty} f_m(x)$ as $n \to \infty$ for any $x \in X$, Theorem 1.24 and Proposition 1.25 yield

$$\int_X \left(\sum_{n=1}^{\infty} f_n\right) d\mu = \lim_{n\to\infty}\int_X \left(\sum_{i=1}^{n} f_i\right) d\mu = \lim_{n\to\infty}\sum_{i=1}^{n}\int_X f_i d\mu = \sum_{n=1}^{\infty}\int_X f_n d\mu,$$

completing the proof. □

Here is another important limit theorem for integrals of non-negative functions.

Theorem 1.27 (Fatou's lemma)**.** *Let $f_n : X \to [0,\infty]$ be $\mathcal{M}$-measurable for each $n \in \mathbb{N}$. Then*

$$\int_X \left(\liminf_{n\to\infty} f_n\right) d\mu \le \liminf_{n\to\infty}\int_X f_n d\mu. \tag{1.24}$$

Proof. Let $m, n \in \mathbb{N}$, $m \ge n$. Since $\inf_{k\ge n} f_k \le f_m$ on X, $\int_X \left(\inf_{k\ge n} f_k\right) d\mu \le \int_X f_m d\mu$ by Lemma 1.23 and hence

$$\int_X \left(\inf_{k\ge n} f_k\right) d\mu \le \inf_{k\ge n}\int_X f_k d\mu. \tag{1.25}$$

Since $\inf_{k\ge n} f_k(x)$ is non-decreasing in n and converges to $\liminf_{m\to\infty} f_m(x)$ as $n \to \infty$ for any $x \in X$, (1.24) follows by using Theorem 1.24 to let $n \to \infty$ in (1.25). □

1.3.2 $[-\infty,\infty]$-valued functions

Definition 1.28. For $f : X \to [-\infty,\infty]$, we define $f^+, f^- : X \to [0,\infty]$ by

$$f^+(x) := \max\{f(x), 0\} \qquad \text{and} \qquad f^-(x) := -\min\{f(x), 0\}, \tag{1.26}$$

so that $f = f^+ - f^-$ and $|f| = f^+ + f^-$ (of course we set $|\infty| := |-\infty| := \infty$).

By Propositions 1.15 and 1.16, if f is $\mathcal{M}$-measurable then so are f^+, f^- and $|f|$.

Definition 1.29 (Integration of $[-\infty,\infty]$-valued functions)**.** Let $f : X \to [-\infty,\infty]$ be $\mathcal{M}$-measurable and let $A \in \mathcal{M}$. We say that f *admits the μ-integral on A* if and only if $\min\bigl\{\int_X f^+ \mathbf{1}_A d\mu, \int_X f^- \mathbf{1}_A d\mu\bigr\} < \infty$, and in this case its *$\mu$-integral $\int_A f d\mu$ on A* is defined by

$$\int_A f d\mu := \int_X f^+ \mathbf{1}_A d\mu - \int_X f^- \mathbf{1}_A d\mu. \tag{1.27}$$

Moreover, f is called *μ-integrable on A* if and only if $\int_X |f| \mathbf{1}_A d\mu < \infty$. When $A = X$, the part *"on X"* will be omitted from these phrases. Finally, we set

$$\mathcal{L}^1(X,\mathcal{M},\mu) := \{f : X \to \mathbb{R} \mid f \text{ is } \mathcal{M}\text{-measurable and } \mu\text{-integrable}\}, \tag{1.28}$$

which will be simply written as $\mathcal{L}^1(X,\mu)$ or $\mathcal{L}^1(\mu)$ when no confusion can occur.

Note that (1.27) with $A = X$ is consistent with (1.19) for non-negative functions, since $f^+ = f$ and $f^- = 0$ for $\mathcal{M}$-measurable $f : X \to [0,\infty]$. Note also that f is μ-integrable on A if and only if f admits the μ-integral on A and $\int_A f d\mu \in \mathbb{R}$.

Notation. The integral $\int_A f d\mu$ is often written in slightly different notations, e.g.

$$\int_A f(x) d\mu(x) := \int_A f(x) \mu(dx) := \int_A f d\mu. \tag{1.29}$$

These alternative notations are used especially when it should be made clear in which variable the integral is taken.[7]

Proposition 1.30. *Let $f : X \to [-\infty,\infty]$ be $\mathcal{M}$-measurable.*
(1) *Let $A \in \mathcal{M}$ satisfy $\mu(A) = 0$. Then f is μ-integrable on A and $\int_A f d\mu = 0$.*
(2) *If $A, B \in \mathcal{M}$ and f is μ-integrable on A and B, then f is μ-integrable on $A \cup B$. If $A \cap B = \emptyset$ in addition, then*

$$\int_{A\cup B} f d\mu = \int_A f d\mu + \int_B f d\mu. \tag{1.30}$$

(3) *If f is μ-integrable, then $\mu\bigl(f^{-1}(\infty) \cup f^{-1}(-\infty)\bigr) = 0$.*

Proof. (1) It suffices to show $\int_X |f| \mathbf{1}_A d\mu = 0$. Let $s : X \to \mathbb{R}$ be $\mathcal{M}$-simple and satisfy $0 \le s \le |f| \mathbf{1}_A$ on X. Then for any $a \in s(X) \setminus \{0\}$, $s^{-1}(a) \subset A$ and hence $\mu\bigl(s^{-1}(a)\bigr) = 0$. Thus $\int_X s d\mu = 0$ for any such s and therefore $\int_X |f| \mathbf{1}_A d\mu = 0$.
(2) By Proposition 1.25 and $\mathbf{1}_{A\cup B} \le \mathbf{1}_A + \mathbf{1}_B$,

$$\int_X |f| \mathbf{1}_{A\cup B} d\mu \le \int_X |f| (\mathbf{1}_A + \mathbf{1}_B) d\mu = \int_X |f| \mathbf{1}_A d\mu + \int_X |f| \mathbf{1}_B d\mu < \infty.$$

If $A \cap B = \emptyset$, then $\mathbf{1}_{A\cup B} = \mathbf{1}_A + \mathbf{1}_B$, which together with Proposition 1.25 immediately shows (1.30).
(3) Set $A := f^{-1}(\infty) \cup f^{-1}(-\infty)$ and let $n \in \mathbb{N}$. Then $|f| \ge |f| \mathbf{1}_A \ge n \mathbf{1}_A$ on X and hence $n\mu(A) = \int_X n \mathbf{1}_A d\mu \le \int_X |f| d\mu < \infty$. Thus $0 \le \mu(A) \le n^{-1} \int_X |f| d\mu$, and letting $n \to \infty$ yields $\mu(A) = 0$. □

[7] The first and second notations in (1.29) have exactly the same meaning, but for certain reasons the second notation is often preferred in the context of probability theory.

Proposition 1.31. (1) *If $f, g : X \to [-\infty, \infty]$ are $\mathcal{M}$-measurable, admit the μ-integrals and satisfy $f \leq g$ on X, then*

$$\int_X f d\mu \leq \int_X g d\mu. \tag{1.31}$$

In particular, if $f : X \to [-\infty, \infty]$ is $\mathcal{M}$-measurable and admits the μ-integral, then

$$\left| \int_X f d\mu \right| \leq \int_X |f| d\mu. \tag{1.32}$$

(2) *If $f, g \in \mathcal{L}^1(\mu)$ and $\alpha, \beta \in \mathbb{R}$, then $\alpha f + \beta g \in \mathcal{L}^1(\mu)$ and*

$$\int_X (\alpha f + \beta g) d\mu = \alpha \int_X f d\mu + \beta \int_X g d\mu. \tag{1.33}$$

Proof. (1) $f \leq g$ implies $f^+ \leq g^+$ and $f^- \geq g^-$, and hence by Lemma 1.23 we have $\int_X f^+ d\mu \leq \int_X g^+ d\mu$ and $\int_X f^- d\mu \geq \int_X g^- d\mu$. (1.31) is immediate from these two inequalities, and (1.32) follows from (1.31) and $-|f| \leq f \leq |f|$.
(2) It is easy to see that $\alpha f \in \mathcal{L}^1(\mu)$ and that $\int_X \alpha f d\mu = \alpha \int_X f d\mu$; indeed, this is immediate from Proposition 1.25 for $\alpha \in [0, \infty)$, and the case of $\alpha = -1$ follows from $(-f)^+ = f^-$ and $(-f)^- = f^+$. Thus it remains to prove the assertions for $\alpha = \beta = 1$. $f + g \in \mathcal{L}^1(\mu)$ follows by $|f + g| \leq |f| + |g|$, Lemma 1.23 and Proposition 1.25. From $(f + g)^+ - (f + g)^- = f + g = f^+ - f^- + g^+ - g^-$ we have $(f + g)^+ + f^- + g^- = (f + g)^- + f^+ + g^+$, and Proposition 1.25 yields

$$\int_X (f + g)^+ d\mu + \int_X f^- d\mu + \int_X g^- d\mu = \int_X (f + g)^- d\mu + \int_X f^+ d\mu + \int_X g^+ d\mu$$

where all the integrals are finite. Therefore $\int_X (f + g) d\mu = \int_X f d\mu + \int_X g d\mu$. □

The following convergence theorem often plays fundamental roles in analysis.

Theorem 1.32 (Lebesgue's dominated convergence theorem, DCT)**.** *Let $f_n : X \to [-\infty, \infty]$ be $\mathcal{M}$-measurable for each $n \in \mathbb{N}$. Suppose the following two conditions:*

(L1) *The limit $f(x) := \lim_{n\to\infty} f_n(x)$ exists in $[-\infty, \infty]$ for any $x \in X$.*

(L2) *There exists an $\mathcal{M}$-measurable, μ-integrable function $g : X \to [0, \infty]$ such that $|f_n(x)| \leq g(x)$ for any $x \in X$ and any $n \in \mathbb{N}$.*

Then $f : X \to [-\infty, \infty]$ is $\mathcal{M}$-measurable and μ-integrable, and

$$\lim_{n\to\infty} \int_X f_n d\mu = \int_X f d\mu. \tag{1.34}$$

Proof. The $\mathcal{M}$-measurability of f follows from Proposition 1.16. Since $|f_n| \leq g$ we have $|f| \leq g$, and Lemma 1.23 yields $\int_X |f| d\mu \leq \int_X g d\mu < \infty$. Let $A := g^{-1}(\infty)$, so that $\mu(A) = 0$ by Proposition 1.30-(3). Then

$$\int_X f_n d\mu - \int_X f d\mu = \int_{A^c} f_n d\mu - \int_{A^c} f d\mu = \int_X (f_n - f) \mathbf{1}_{A^c} d\mu \tag{1.35}$$

by Proposition 1.30-(1),(2) and Proposition 1.31-(2) (note that f_n and f are $\mathbb{R}$-valued on A^c). Since $2g - |f_n - f| \geq 0$ on A^c, Theorem 1.27 and Proposition 1.31-(2) yield

$$\begin{aligned}\int_X 2g\mathbf{1}_{A^c}\,d\mu &= \int_X \liminf_{n\to\infty}\Big((2g-|f_n-f|)\mathbf{1}_{A^c}\Big)d\mu \\ &\leq \liminf_{n\to\infty}\int_X (2g-|f_n-f|)\mathbf{1}_{A^c}\,d\mu \\ &= \liminf_{n\to\infty}\Big(\int_X 2g\mathbf{1}_{A^c} - \int_X |f_n-f|\mathbf{1}_{A^c}\,d\mu\Big) \\ &= \int_X 2g\mathbf{1}_{A^c}\,d\mu - \limsup_{n\to\infty}\int_X |f_n-f|\mathbf{1}_{A^c}\,d\mu.\end{aligned} \tag{1.36}$$

Since each term in (1.36) is finite, we may subtract $\int_X 2g\mathbf{1}_{A^c}\,d\mu$ to obtain

$$\limsup_{n\to\infty}\int_X |f_n-f|\mathbf{1}_{A^c}\,d\mu \leq 0, \quad \text{i.e.} \quad \lim_{n\to\infty}\int_X |f_n-f|\mathbf{1}_{A^c}\,d\mu = 0. \tag{1.37}$$

Now (1.34) follows from (1.32), (1.35) and (1.37). □

1.3.3 Sets of measure zero and completion of measure spaces

In the above proof of Theorem 1.32, we already utilized the fact that the set $g^{-1}(\infty)$ is *"negligible"* since it is of μ-measure zero. There are a lot of situations in measure theory where it is necessary to neglect sets of measure zero appropriately, and here is an important definition used in those situations.

Definition 1.33 (Almost everywhere, a.e.)**.** Let $\mathbf{P}(x)$ be a proposition on x for each $x \in X$, and let $A \in \mathcal{M}$. Then we say that $\mathbf{P}$ *holds μ-almost everywhere on A*, or $\mathbf{P}$ *holds μ-a.e. on A* for short, if and only if there exists $N \in \mathcal{M}$ with $\mu(N) = 0$ such that $\mathbf{P}(x)$ holds for any $x \in A \setminus N$. For $A = X$, we simply say $\mathbf{P}$ *holds μ-a.e.* instead of saying $\mathbf{P}$ *holds μ-a.e. on X*.

For example, $\mathbf{P}(x)$ can be "$f(x) = 0$" for a given function $f : X \to \mathbb{R}$, or can be "the limit $\lim_{n\to\infty} f_n(x)$ exists in $\mathbb{R}$" for a given sequence $\{f_n\}_{n=1}^{\infty}$ of functions on X.

Measure theoretic assumptions naturally imply μ-a.e. assertions, as illustrated in the following proposition.

Proposition 1.34. (1) *If $f : X \to [0,\infty]$ is $\mathcal{M}$-measurable and $\int_X f\,d\mu = 0$, then $f = 0$ μ-a.e.*
(2) *If $f : X \to [-\infty,\infty]$ is $\mathcal{M}$-measurable, μ-integrable and satisfies $\int_A f\,d\mu = 0$ for any $A \in \mathcal{M}$, then $f = 0$ μ-a.e.*

Proof. (1) Let $n \in \mathbb{N}$ and set $A_n := f^{-1}\big([n^{-1},\infty]\big)$. Then $0 = \int_X f\,d\mu \geq \int_{A_n} f\,d\mu \geq n^{-1}\mu(A_n)$ and hence $\mu(A_n) = 0$. Letting $n \to \infty$ yields $\mu\big(f^{-1}\big((0,\infty]\big)\big) = 0$.
(2) Let $A := f^{-1}\big([0,\infty]\big)$. Then since $f^+\mathbf{1}_A = f^+$, $f^-\mathbf{1}_{A^c} = f^-$ and $f^-\mathbf{1}_A = f^+\mathbf{1}_{A^c} = 0$, we have $0 = \int_A f\,d\mu = \int_X f^+d\mu$ and $0 = -\int_{A^c} f\,d\mu = \int_X f^-d\mu$. Thus $f^+ = f^- = 0$ μ-a.e. by (1) and hence $f = 0$ μ-a.e. □

The following proposition says that sets of μ-measure zero are in fact negligible as long as μ-integrals are concerned.

Proposition 1.35. *Let $f, g : X \to [-\infty, \infty]$ be $\mathcal{M}$-measurable and satisfy $f = g$ μ-a.e. Then for any $A \in \mathcal{M}$, f admits the μ-integral on A if and only if so does g, and in that case we have*

$$\int_A f d\mu = \int_A g d\mu. \tag{1.38}$$

Proof. Choose $N \in \mathcal{M}$ with $\mu(N) = 0$ so that $f(x) = g(x)$ for any $x \in X \setminus N$. Then since $\mu(A \cap N) = 0$, Proposition 1.25 and Proposition 1.30-(1) imply

$$\begin{aligned}\int_X f^{\pm}\mathbf{1}_A d\mu &= \int_X f^{\pm}\mathbf{1}_{A\setminus N} d\mu + \int_X f^{\pm}\mathbf{1}_{A\cap N} d\mu = \int_X g^{\pm}\mathbf{1}_{A\setminus N} d\mu + 0\\ &= \int_X g^{\pm}\mathbf{1}_{A\setminus N} d\mu + \int_X g^{\pm}\mathbf{1}_{A\cap N} d\mu = \int_X g^{\pm}\mathbf{1}_A d\mu,\end{aligned}$$

from which the assertions are immediate. □

Remark 1.36. In the above proof, we can take $\{x \in X \mid f(x) \neq g(x)\}$ as the set N; in fact, it is not difficult to verify that $\{x \in X \mid f(x) \neq g(x)\} \in \mathcal{M}$ (see Problem 1.16).

By virtue of Proposition 1.35, we can slightly weaken the assumptions of the results in this section by allowing exceptional sets of μ-measure zero. For example, Theorem 1.32 is still valid if *"for any $x \in X$"* in the conditions (L1) and (L2) are replaced by *"for μ-a.e. $x \in X$"*; indeed, if $N_n \in \mathcal{M}$ with $\mu(N_n) = 0$, $n \in \mathbb{N} \cup \{0\}$, are chosen so that

(L1)′ the limit $f(x) := \lim_{n\to\infty} f_n(x)$ exists in $[-\infty, \infty]$ for any $x \in X \setminus N_0$, and

(L2)′ $|f_n(x)| \leq g(x)$ for any $x \in X \setminus N_n$ for each $n \in \mathbb{N}$,

then since $N := \bigcup_{n=0}^{\infty} N_n$ is of μ-measure zero by Problem 1.10, we obtain (1.34) by applying the original Theorem 1.32 to $\{g_n\}_{n=1}^{\infty}$ defined by

$$g_n(x) := \begin{cases} f_n(x) & \text{if } x \in X \setminus N,\\ 0 & \text{if } x \in N.\end{cases}$$

Note here that the limit function f *is defined only μ-almost everywhere*, only on the set $A := \{x \in X \mid \limsup_{n\to\infty} f_n(x) = \liminf_{n\to\infty} f_n(x)\} \in \mathcal{M}$, but still *its μ-integral $\int_X f d\mu$ is uniquely defined.* Indeed, since $f = \limsup_{n\to\infty} f_n$ on A and it is $\mathcal{M}|_A$-measurable, if we set $f := h$ on A^c, where $h : A^c \to [-\infty, \infty]$ is an arbitrary $\mathcal{M}|_{A^c}$-measurable function, then f is $\mathcal{M}$-measurable, and Proposition 1.35 together with $\mu(A^c) = 0$ assures that the integral $\int_X f d\mu$ is independent of $h = f|_{A^c}$.

Such a situation is quite common in measure theory and probability theory: once an $\mathcal{M}|_{X\setminus N}$-measurable function $f : X \setminus N \to [-\infty, \infty]$ is defined outside a set $N \in \mathcal{M}$ with $\mu(N) = 0$, we define $\int_X f d\mu$ as the μ-integral of any $\mathcal{M}$-measurable extension of f to X, and we often do NOT specify the values on N.

For $f : X \setminus N \to [-\infty, \infty]$ as above, it is natural to regard *any* extension of f to X as measurable since the way of extension does not affect the integral. This measurability is, however, not necessarily true since N may contain a subset which does not belong to $\mathcal{M}$. In fact, we can achieve this measurability by enlarging the σ-algebra $\mathcal{M}$ in the following way.

Theorem 1.37 (Completion of a measure space)**.** *We define*

$$\overline{\mathcal{M}}^{\mu} := \{A \subset X \mid B \subset A \subset C \text{ for some } B, C \in \mathcal{M} \text{ with } \mu(C \setminus B) = 0\}, \tag{1.39}$$

Then $\overline{\mathcal{M}}^{\mu}$ is a σ-algebra in X larger than $\mathcal{M}$, and μ is uniquely extended to a measure $\overline{\mu}$ on $\overline{\mathcal{M}}^{\mu}$. $\overline{\mathcal{M}}^{\mu}$ is called the μ-completion of $\mathcal{M}$*, and $\overline{\mu}$ is called the* completion of μ.

Proof. If $A \in \mathcal{M}$ then setting $B := A =: C$ shows $A \in \overline{\mathcal{M}}^{\mu}$. Thus $\mathcal{M} \subset \overline{\mathcal{M}}^{\mu}$, and in particular $\emptyset \in \overline{\mathcal{M}}^{\mu}$. If $A \in \overline{\mathcal{M}}^{\mu}$ and $B, C \in \mathcal{M}$ are as in (1.39), then $C^c \subset A^c \subset B^c$, $\mu(B^c \setminus C^c) = \mu(C \setminus B) = 0$ and hence $A^c \in \overline{\mathcal{M}}^{\mu}$. If $\{A_n\}_{n=1}^{\infty} \subset \overline{\mathcal{M}}^{\mu}$ and $B_n, C_n \in \mathcal{M}$ are as in (1.39) for A_n for each $n \in \mathbb{N}$, then $\bigcup_{n=1}^{\infty} B_n \subset \bigcup_{n=1}^{\infty} A_n \subset \bigcup_{n=1}^{\infty} C_n$,

$$\bigcup_{n=1}^{\infty} C_n \setminus \bigcup_{n=1}^{\infty} B_n \subset \bigcup_{n=1}^{\infty} (C_n \setminus B_n)$$

and hence by Problem 1.10,

$$\mu\left(\bigcup_{n=1}^{\infty} C_n \setminus \bigcup_{n=1}^{\infty} B_n\right) \leq \mu\left(\bigcup_{n=1}^{\infty} (C_n \setminus B_n)\right) \leq \sum_{n=1}^{\infty} \mu(C_n \setminus B_n) = 0.$$

Thus $\bigcup_{n=1}^{\infty} A_n \in \overline{\mathcal{M}}^{\mu}$, and $\overline{\mathcal{M}}^{\mu}$ is a σ-algebra in X.

We would like to define $\overline{\mu} : \overline{\mathcal{M}}^{\mu} \to [0, \infty]$ by $\overline{\mu}(A) := \mu(B) = \mu(C \setminus B) + \mu(B) = \mu(C)$, where $B, C \in \mathcal{M}$ are as in (1.39). If $B_1, C_1 \in \mathcal{M}$ also satisfy $B_1 \subset A \subset C_1$ and $\mu(C_1 \setminus B_1) = 0$, then since $B \cup B_1 \subset A \subset C \cap C_1$ we have $\mu(B) \leq \mu(C_1) = \mu(B_1)$, $\mu(B_1) \leq \mu(C) = \mu(B)$ and hence $\mu(B) = \mu(B_1)$. Thus $\overline{\mu}(A) := \mu(B) = \mu(C)$ is independent of a particular choice of $B, C \in \mathcal{M}$ as in (1.39), and it defines a function $\overline{\mu} : \overline{\mathcal{M}}^{\mu} \to [0, \infty]$. For $A \in \mathcal{M}$, we may take $B := A =: C$ and therefore $\overline{\mu}(A) = \mu(A)$. In particular, $\overline{\mu}(\emptyset) = 0$. The countable additivity of $\overline{\mu}$ is immediate from that of μ, and the uniqueness of such an extension is also clear. □

Definition 1.38. We call μ, or $(X, \mathcal{M}, \mu)$, *complete* if and only if $A \in \mathcal{M}$ whenever $A \subset N$ for some $N \in \mathcal{M}$ with $\mu(N) = 0$.

By the construction, the completion $\overline{\mu}$ of μ is actually complete, which and (1.39) easily imply that $(X, \mathcal{M}, \mu)$ *is complete if and only if* $\overline{\mathcal{M}}^{\mu} = \mathcal{M}$. On the other hand, it is known that the Lebesgue measure m_d on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ (Example 1.8) and the Bernoulli measure $\mathbb{P}_p$ on $\mathcal{F}$ (Example 1.12) are not complete.

1.3.4 Complex functions

In this course, we usually consider $\mathbb{R}$-valued or $[-\infty, \infty]$-valued functions, but we will need integration of complex functions later in Chapter 4. Here we collect some basic definitions and facts concerning integration of complex functions.

Let i denote the imaginary unit. As usual, $\mathbb{C} = \{x + iy \mid x, y \in \mathbb{R}\}$ is naturally identified with $\mathbb{R}^2$, so that $\mathbb{C}$ is equipped with the metric structure inherited from $\mathbb{R}^2$.

Definition 1.39. $f : X \to \mathbb{C}$ is called $\mathcal{M}$-*measurable* if and only if $f^{-1}(A) \in \mathcal{M}$ for any $A \in \mathcal{B}(\mathbb{C})$.

Proposition 1.40. *$f : X \to \mathbb{C}$ is $\mathcal{M}$-measurable if and only if its real part* $\mathrm{Re}(f)$ *and imaginary part* $\mathrm{Im}(f)$ *are both $\mathbb{R}$-valued $\mathcal{M}$-measurable functions.*

Proof. It is clear that f is $\mathcal{M}$-measurable if and only if $\tilde{f} := \big(\mathrm{Re}(f), \mathrm{Im}(f)\big) : X \to \mathbb{R}^2$ satisfies $\tilde{f}^{-1}(A) \in \mathcal{M}$ for any $A \in \mathcal{B}(\mathbb{R}^2)$. By Problem 1.15, this is equivalent to the condition that $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are $\mathcal{M}$-measurable. □

Note that $|f|$ is $\mathcal{M}$-measurable if $f : X \to \mathbb{C}$ is $\mathcal{M}$-measurable, since the function $\mathbb{C} \ni z \mapsto |z| \in \mathbb{R}$ is continuous and hence $\mathcal{B}(\mathbb{C})$-measurable by Lemma 1.17.

Definition 1.41 (Integration of complex functions)**.** Let $f : X \to \mathbb{C}$ be $\mathcal{M}$-measurable and let $A \in \mathcal{M}$. f is called *μ-integrable on A* if and only if $\int_X |f|\mathbf{1}_A d\mu < \infty$, or equivalently, $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are μ-integrable on A, and in this case its *μ-integral $\int_A f d\mu$ on A* is defined by

$$\int_A f d\mu := \int_A \mathrm{Re}(f) d\mu + i \int_A \mathrm{Im}(f) d\mu. \tag{1.40}$$

When $A = X$, the part *"on X"* will be omitted from these phrases. Finally, we set

$$\mathcal{L}^1(X, \mathcal{M}, \mu, \mathbb{C}) := \{f : X \to \mathbb{C} \mid f \text{ is } \mathcal{M}\text{-measurable and } \mu\text{-integrable}\}, \tag{1.41}$$

which will be simply written as $\mathcal{L}^1(X, \mu, \mathbb{C})$ or $\mathcal{L}^1(\mu, \mathbb{C})$ when no confusion can occur.

Proposition 1.42. (1) *If $f \in \mathcal{L}^1(\mu, \mathbb{C})$, then*

$$\left| \int_X f d\mu \right| \le \int_X |f| d\mu. \tag{1.42}$$

(2) *If $f, g \in \mathcal{L}^1(\mu, \mathbb{C})$ and $\alpha, \beta \in \mathbb{C}$, then $\alpha f + \beta g \in \mathcal{L}^1(\mu, \mathbb{C})$ and*

$$\int_X (\alpha f + \beta g) d\mu = \alpha \int_X f d\mu + \beta \int_X g d\mu. \tag{1.43}$$

Proof. (2) Using Proposition 1.31-(2), we have

$$\begin{aligned}\int_X i f d\mu &= \int_X \big(-\mathrm{Im}(f) + i\,\mathrm{Re}(f)\big) d\mu = \int_X \big(-\mathrm{Im}(f)\big) d\mu + i \int_X \mathrm{Re}(f) d\mu \\ &= -\int_X \mathrm{Im}(f) d\mu + i \int_X \mathrm{Re}(f) d\mu = i \int_X f d\mu,\end{aligned}$$

which together with Proposition 1.31-(2) easily implies the assertion.
(1) Choose $\alpha \in \mathbb{C}$ with $|\alpha| = 1$ so that $\left|\int_X f d\mu\right| = \alpha \int_X f d\mu$. Then

$$\left| \int_X f d\mu \right| = \alpha \int_X f d\mu = \int_X \alpha f d\mu = \int_X \mathrm{Re}(\alpha f) d\mu \le \int_X |f| d\mu,$$

where the third equality is due to $\int_X \alpha f d\mu \in \mathbb{R}$ and the inequality follows from $\mathrm{Re}(\alpha f) \le |\alpha f| = |f|$ and (1.31). □

1.4 Some basic consequences

In this section, we present some consequences of the integration theory developed so far in this chapter. In the proof of the first two theorems, we will utilize monotone approximation of a measurable function by simple functions (Proposition 1.19) and the monotone convergence theorem (Theorem 1.24) in a typical way.

Throughout this section, $(X, \mathcal{M}, \mu)$ denotes a given measure space.

Theorem 1.43. *Let $f : X \to [0, \infty]$ be $\mathcal{M}$-measurable and define $\nu : \mathcal{M} \to [0, \infty]$ by*

$$\nu(A) := \int_A f d\mu. \tag{1.44}$$

Then ν is a measure on $(X, \mathcal{M})$. Moreover, if $g : X \to [-\infty, \infty]$ is $\mathcal{M}$-measurable, then g admits the ν-integral if and only if gf admits the μ-integral, and in that case

$$\int_X g d\nu = \int_X g f d\mu. \tag{1.45}$$

The measure ν is denoted by $f \cdot \mu$, and (1.45) is sometimes written as $d\nu = f d\mu$.

Proof. $\nu(\emptyset) = \int_X 0 d\mu = 0$. If $\{A_n\}_{n=1}^{\infty} \subset \mathcal{M}$ and $A_i \cap A_j = \emptyset$ for $i, j \in \mathbb{N}$ with $i \neq j$, then $f \mathbf{1}_{\bigcup_{n=1}^{\infty} A_n} = \sum_{n=1}^{\infty} f \mathbf{1}_{A_n}$ and therefore Proposition 1.26 implies that

$$\nu\left(\bigcup_{n=1}^{\infty} A_n\right) = \int_X \sum_{n=1}^{\infty} f \mathbf{1}_{A_n} d\mu = \sum_{n=1}^{\infty} \int_X f \mathbf{1}_{A_n} d\mu = \sum_{n=1}^{\infty} \nu(A_n).$$

Thus ν is a measure on $(X, \mathcal{M})$. For the other assertions, since $(gf)^{\pm} = g^{\pm} f$, it suffices to show (1.45) for $g : X \to [0, \infty]$. For an $\mathcal{M}$-simple function $s : X \to [0, \infty)$, Proposition 1.25 yields

$$\begin{aligned} \int_X s d\nu &= \sum_{a \in s(X)} a \nu\bigl(s^{-1}(a)\bigr) = \sum_{a \in s(X)} a \int_X f \mathbf{1}_{s^{-1}(a)} d\mu \\ &= \int_X \sum_{a \in s(X)} a f \mathbf{1}_{s^{-1}(a)} d\mu = \int_X f \sum_{a \in s(X)} a \mathbf{1}_{s^{-1}(a)} = \int_X s f d\mu. \end{aligned} \tag{1.46}$$

Therefore by choosing a sequence $\{s_n\}_{n=1}^{\infty}$ of $\mathcal{M}$-simple functions monotonically increasing to g, as in Proposition 1.19, we have $\int_X s_n d\nu = \int_X s_n f d\mu$, and letting $n \to \infty$ results in $\int_X g d\nu = \int_X g f d\mu$ by virtue of Theorem 1.24. □

Remark 1.44. Note that the measure $\nu = f \cdot \mu$ satisfies $\nu(A) = 0$ for any $A \in \mathcal{M}$ with $\mu(A) = 0$. A measure on $(X, \mathcal{M})$ with this property is called *absolutely continuous with respect to* μ, and it is known that this property completely characterizes measures on $(X, \mathcal{M})$ of the form $f \cdot \mu$ for some $f \in \mathcal{L}^1(\mu)$. This fact is very fundamental in measure theory and probability theory and known as the *Radon-Nikodym theorem*, but we do not treat this theorem in this course. See [7, Chapter 6] and [1, Sections 5.5 and 5.6] for details of the Radon-Nikodym theorem.

Definition 1.45. Let $(S, \mathcal{B})$ be a measurable space. A map $\varphi : X \to S$ is called $\mathcal{M}/\mathcal{B}$-*measurable* if and only if $\varphi^{-1}(A) \in \mathcal{M}$ for any $A \in \mathcal{B}$.

The following result is a fundamental tool in probability theory.

Theorem 1.46 (Image measure theorem)**.** *Let $(S, \mathcal{B})$ be a measurable space and let $\varphi : X \to S$ be $\mathcal{M}/\mathcal{B}$-measurable. Then the function $\mu \circ \varphi^{-1} : \mathcal{B} \to [0, \infty]$ defined by $(\mu \circ \varphi^{-1})(A) := \mu\bigl(\varphi^{-1}(A)\bigr)$ is a measure on $(S, \mathcal{B})$. Moreover, if $f : S \to [-\infty, \infty]$ is $\mathcal{B}$-measurable, then f admits the $(\mu \circ \varphi^{-1})$-integral if and only if $f \circ \varphi$ admits the μ-integral, and in that case*

$$\int_S f d(\mu \circ \varphi^{-1}) = \int_X (f \circ \varphi) d\mu. \tag{1.47}$$

The measure $\mu \circ \varphi^{-1}$ is called the *image measure of μ by φ*.

Proof. It is immediate to see that $\mu \circ \varphi^{-1}$ is a measure on $(S, \mathcal{B})$. Note that the $\mathcal{B}$-measurability of f yields the $\mathcal{M}$-measurability of $f \circ \varphi$. If $f = \mathbf{1}_A$ for some $A \in \mathcal{B}$, then $\int_S \mathbf{1}_A d(\mu \circ \varphi^{-1}) = \mu\bigl(\varphi^{-1}(A)\bigr) = \int_X \mathbf{1}_{\varphi^{-1}(A)} d\mu = \int_X (\mathbf{1}_A \circ \varphi) d\mu$, which and Proposition 1.25 together imply (1.47) for $\mathcal{M}$-simple $f : X \to [0, \infty)$ similarly to (1.46). Then by using Theorem 1.24, we obtain (1.47) for $[0, \infty]$-valued f and hence for $[-\infty, \infty]$-valued f as well, noting that $(f \circ \varphi)^{\pm} = f^{\pm} \circ \varphi$. □

An application of the dominated convergence theorem (Theorem 1.32) gives rise to the following theorem.

Theorem 1.47. *Let $a, b \in [-\infty, \infty]$, $a < b$ and let $f : X \times (a, b) \to \mathbb{R}$ be such that $f(\cdot, t) \in \mathcal{L}^1(\mu)$ for any $t \in (a, b)$ and $f(x, \cdot) : (a, b) \to \mathbb{R}$ is differentiable for any $x \in X$. Suppose there exists an $\mathcal{M}$-measurable μ-integrable function $g : X \to [0, \infty]$ such that $|(\partial f/\partial t)(x, t)| \le g(x)$ for any $(x, t) \in X \times (a, b)$. Then $\int_X f(x, \cdot) d\mu(x) : (a, b) \to \mathbb{R}$ is differentiable, and for any $t \in (a, b)$, $(\partial f/\partial t)(\cdot, t) \in \mathcal{L}^1(\mu)$ and*

$$\frac{d}{dt} \int_X f(x, t) d\mu(x) = \int_X \frac{\partial f}{\partial t}(x, t) d\mu(x). \tag{1.48}$$

Proof. Let $t \in (a, b)$. By the definition of $\partial f/\partial t$,

$$\frac{\partial f}{\partial t}(x, t) = \lim_{n \to \infty} \frac{f(x, t + 1/n) - f(x, t)}{1/n}, \qquad x \in X,$$

which is $\mathcal{M}$-measurable in $x \in X$ by Proposition 1.16 since $f(\cdot, t + 1/n)$ and $f(\cdot, t)$ are $\mathcal{M}$-measurable. Then it follows from $|(\partial f/\partial t)(\cdot, t)| \le g$ that $(\partial f/\partial t)(\cdot, t) \in \mathcal{L}^1(\mu)$.

Let $\{h_n\}_{n=1}^{\infty} \subset \mathbb{R} \setminus \{0\}$ satisfy $t + h_n \in (a, b)$ and $\lim_{n \to \infty} h_n = 0$. Then

$$\frac{1}{h_n}\left(\int_X f(x, t + h_n) d\mu(x) - \int_X f(x, t) d\mu(x)\right) = \int_X \frac{f(x, t + h_n) - f(x, t)}{h_n} d\mu(x), \tag{1.49}$$

where the integrand in the right-hand side satisfies

$$\left|\frac{f(x, t + h_n) - f(x, t)}{h_n}\right| = \left|\frac{\partial f}{\partial t}\bigl(x, t + \theta(x, t, h_n) h_n\bigr)\right| \le g(x), \qquad x \in X, \tag{1.50}$$

for some $\theta(x,t,h_n) \in (0,1)$, by the mean value theorem. By (1.50), the dominated convergence theorem (Theorem 1.32) applies to the right-hand side of (1.49) to imply that the left-hand side of (1.49) converges to $\int_X (\partial f/\partial t)(x,t)d\mu(x)$. This proves that $\int_X f(x,\cdot)d\mu(x)$ is differentiable at t and that (1.48) holds, since $\{h_n\}_{n=1}^\infty \subset \mathbb{R} \setminus \{0\}$ is an arbitrary sequence satisfying $t + h_n \in (a,b)$ and $\lim_{n\to\infty} h_n = 0$. □

Next we present two frequently used inequalities. For $p \in (0,\infty)$, we naturally extend the power function $[0,\infty) \ni x \mapsto x^p$ to $[0,\infty]$ by setting $\infty^p := \infty$. Note that, if $f : X \to [0,\infty]$ is $\mathcal{M}$-measurable then so is f^p for any $p \in (0,\infty)$.

Theorem 1.48 (Hölder's inequality). *Let $p \in (1,\infty)$ and set $q := p/(p-1)$, so that $p^{-1} + q^{-1} = 1$.* (q is called the *conjugate exponent* of p.) *Let $f, g : X \to [0,\infty]$ be $\mathcal{M}$-measurable. Then*

$$\int_X fg\,d\mu \le \left(\int_X f^p d\mu\right)^{1/p} \left(\int_X g^q d\mu\right)^{1/q}. \tag{1.51}$$

Proof. Let $A := \left(\int_X f^p d\mu\right)^{1/p}$ and $B := \left(\int_X g^q d\mu\right)^{1/q}$. (1.51) is trivial if $AB = \infty$. If either A or B is equal to 0, then either $f = 0$ μ-a.e. or $g = 0$ μ-a.e. by Proposition 1.34-(1) and hence $\int_X fg\,d\mu = 0 = AB$ by Proposition 1.30-(1).

Thus we may assume that $A, B \in (0,\infty)$. A standard one-dimensional differential calculus together with $p^{-1} + q^{-1} = 1$ easily shows that

$$xy \le \frac{x^p}{p} + \frac{y^q}{q} \qquad \text{for any } x, y \in [0,\infty]. \tag{1.52}$$

By applying (1.52) to $A^{-1}f$ and $B^{-1}g$, we obtain

$$\frac{1}{AB}\int_X fg\,d\mu = \int_X \frac{f}{A}\cdot\frac{g}{B}d\mu \le \frac{1}{p}\int_X \frac{f^p}{A^p}d\mu + \frac{1}{q}\int_X \frac{g^q}{B^q}d\mu = \frac{1}{p} + \frac{1}{q} = 1$$

by virtue of Lemma 1.23 and Proposition 1.25, completing the proof. □

Definition 1.49. Let $p \in (0,\infty)$. For an $\mathcal{M}$-measurable function $f : X \to [-\infty,\infty]$, we define

$$\|f\|_{L^p(X,\mu)} := \left(\int_X |f|^p d\mu\right)^{1/p}, \tag{1.53}$$

which will be simply denoted as $\|f\|_{L^p(\mu)}$ or $\|f\|_{L^p}$ when no confusion can occur. Moreover, we also define

$$\mathcal{L}^p(X,\mathcal{M},\mu) := \{f : X \to \mathbb{R} \mid f \text{ is } \mathcal{M}\text{-measurable and } \|f\|_{L^p(X,\mu)} < \infty\}, \tag{1.54}$$

which will be simply written as $\mathcal{L}^p(X,\mu)$ or $\mathcal{L}^p(\mu)$ when no confusion can occur.

Note that (1.54) is consistent with (1.28). We easily see that $\mathcal{L}^p(\mu)$ is a vector space over $\mathbb{R}$ for each $p \in (0,\infty)$, since $(a+b)^p \le (2\max\{a,b\})^p \le 2^p(a^p + b^p)$ for $a, b \in [0,\infty]$. According to Theorem 1.48, for $p \in (1,\infty)$, $q = p/(p-1)$, $f \in \mathcal{L}^p(\mu)$ and $g \in \mathcal{L}^q(\mu)$ we have $fg \in \mathcal{L}^1(\mu)$ and $\|fg\|_{L^1} \le \|f\|_{L^p}\|g\|_{L^q}$. See Problems 1.31 and 1.32 below for other important facts concerning $\mathcal{L}^p(\mu)$, $p \in [1,\infty)$.

To state and prove another inequality, we need the following definition and lemma.

Definition 1.50 (Convex functions)**.** Let $a, b \in [-\infty, \infty]$, $a < b$ and let $\varphi : (a, b) \to \mathbb{R}$. Then φ is called *convex* if and only if for any $x, y \in (a, b)$ and any $t \in [0, 1]$,

$$\varphi\big((1-t)x + ty\big) \leq (1-t)\varphi(x) + t\varphi(y), \tag{1.55}$$

or equivalently, for any $x, y, z \in (a, b)$ with $x < z < y$,

$$\frac{\varphi(z) - \varphi(x)}{z - x} \leq \frac{\varphi(y) - \varphi(z)}{y - z}. \tag{1.56}$$

For example, φ is convex if φ is differentiable on (a, b) and φ' is non-decreasing, by virtue of the mean value theorem in one-dimensional calculus.

Lemma 1.51. *Let $a, b \in [-\infty, \infty]$, $a < b$. If $\varphi : (a, b) \to \mathbb{R}$ is convex, then it is continuous.*

Proof. Let $x, y \in (a, b)$, $x < y$ and choose $c, d \in (a, b)$ so that $c < x < y < d$. Set $s := \frac{x-c}{y-c}$ and $t := \frac{y-x}{d-x}$, so that $s, t \in (0, 1)$, $x = (1-s)c + sy$ and $y = (1-t)x + td$. From the convexity of φ we see that

$$\frac{\varphi(x) - (1-s)\varphi(c)}{s} \leq \varphi(y) \leq (1-t)\varphi(x) + t\varphi(d). \tag{1.57}$$

Since $s \to 1$ and $t \to 0$ as $y \downarrow x$ or as $x \uparrow y$, it follows from (1.57) that $\lim_{y \downarrow x} \varphi(y) = \varphi(x)$ and $\lim_{x \uparrow y} \varphi(x) = \varphi(y)$. This establishes the continuity of φ, as $x, y \in (a, b)$ are arbitrary. □

Remark 1.52. Note that Lemma 1.51 is based on the assumption that the domain of φ is an *open* interval. In fact, if we define $\varphi : [0, 1] \to \mathbb{R}$ by $\varphi(x) := 0$ for $x \in [0, 1)$ and $\varphi(1) := 1$, then φ satisfies (1.55) for any $x, y, t \in [0, 1]$ but it is not continuous.

Theorem 1.53 (Jensen's inequality)**.** *Assume that μ is a probability measure, that is, $\mu(X) = 1$. Let $a, b \in [-\infty, \infty]$, $a < b$ and let $\varphi : (a, b) \to \mathbb{R}$ be convex. If $f : X \to (a, b)$ and $f \in \mathcal{L}^1(\mu)$, then $\varphi \circ f$ admits the μ-integral and*

$$\varphi\left(\int_X f d\mu\right) \leq \int_X (\varphi \circ f) d\mu. \tag{1.58}$$

Proof. Note that $\varphi \circ f$ is $\mathcal{M}$-measurable by virtue of Lemmas 1.17 and 1.51 and $\mathcal{B}\big((a, b)\big) \subset \mathcal{B}(\mathbb{R})$. Let $z := \int_X f d\mu$. Then $z \in (a, b)$ by $\mu(X) = 1$, $a < f < b$, Proposition 1.31-(2) and Proposition 1.34-(1). Let $\gamma := \sup_{x \in (a,z)} \frac{\varphi(z) - \varphi(x)}{z - x}$. Then $\gamma \in \mathbb{R}$ and $\gamma \leq \frac{\varphi(y) - \varphi(z)}{y - z}$ for any $y \in (z, b)$ by (1.56), and hence

$$\varphi(x) \geq \varphi(z) + \gamma(x - z) \qquad \text{for any } x \in (a, b). \tag{1.59}$$

Thus $\varphi \circ f \geq \varphi(z) + \gamma(f - z)$, which implies $\int_X (\varphi \circ f)^- d\mu < \infty$, and taking the μ-integrals of both sides results in (1.58) in view of (1.31) and Proposition 1.31-(2). □

Exercises

Problem 1.1. Let $X := \{1, 2, 3\}$. Provide all σ-algebras in X.

Problem 1.2. For a set X and $A \subset X$, prove that $\{\emptyset, A, A^c, X\}$ is a σ-algebra in X.

The notion of independence is very important in probability theory. The following definitions, problems and exercises provide some basics about independence of events.

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
(1) A pair $\{A, B\}$ of events $A, B \in \mathcal{F}$ is called *independent* if and only if $\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$.
(2) A (possibly infinite) family $\{A_\lambda\}_{\lambda\in\Lambda} \subset \mathcal{F}$ of events is called *independent* if and only if it holds that $\mathbb{P}\big[\bigcap_{\lambda\in\Lambda_0} A_\lambda\big] = \prod_{\lambda\in\Lambda_0} \mathbb{P}[A_\lambda]$ for any non-empty finite $\Lambda_0 \subset \Lambda$.

Problem 1.3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
(1) Let $A, B \in \mathcal{F}$. Prove that if $\{A, B\}$ is independent then $\{A^c, B\}$, $\{A, B^c\}$ and $\{A^c, B^c\}$ are also independent.
(2) Let $\{A_\lambda\}_{\lambda\in\Lambda} \subset \mathcal{F}$ be a (possibly infinite) family of events. Prove that $\{A_\lambda\}_{\lambda\in\Lambda}$ is independent if and only if $\mathbb{P}\big[\bigcap_{\lambda\in\Lambda_0} B_\lambda\big] = \prod_{\lambda\in\Lambda_0} \mathbb{P}[B_\lambda]$ for any non-empty finite $\Lambda_0 \subset \Lambda$ and any $B_\lambda \in \{\emptyset, A_\lambda, A_\lambda^c, \Omega\}$, $\lambda \in \Lambda_0$.

Problem 1.4. Give an example of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and events $A, B, C \in \mathcal{F}$ such that the pairs $\{A, B\}$, $\{B, C\}$ and $\{A, C\}$ are independent but $\mathbb{P}[A \cap B \cap C] \neq \mathbb{P}[A]\mathbb{P}[B]\mathbb{P}[C]$. (Consider $\Omega := \{1, 2, 3, 4\}$ and $\mathbb{P}[A] := \#A/4$, $A \subset \Omega$.)

Exercise 1.5. Give an example of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and events $A, B, C \in \mathcal{F}$ such that $\{A, B\}$ and $\{B, C\}$ are independent, $\mathbb{P}[A \cap B \cap C] = \mathbb{P}[A]\mathbb{P}[B]\mathbb{P}[C]$ but $\{A, C\}$ is not independent. (Consider $\Omega := \{1, \dots, 16\}$ and $\mathbb{P}[A] := \#A/16$, $A \subset \Omega$.)

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $B \in \mathcal{F}$ satisfy $\mathbb{P}[B] > 0$. For each $A \in \mathcal{F}$, We define the *conditional probability* $\mathbb{P}[A \mid B]$ *of* A *given* B by

$$\mathbb{P}[A \mid B] := \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}. \tag{1.60}$$

Problem 1.6. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $B \in \mathcal{F}$ satisfy $\mathbb{P}[B] > 0$.
(1) Let $A \in \mathcal{F}$. Prove that $\{A, B\}$ is independent if and only if $\mathbb{P}[A \mid B] = \mathbb{P}[A]$.
(2) Prove that the set function $\mathcal{F} \ni A \mapsto \mathbb{P}[A \mid B]$ is a probability measure on $(\Omega, \mathcal{F})$. This probability measure is called the *conditional probability measure given* B.

Problem 1.7. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{\Omega_n\}_{n=1}^N \subset \mathcal{F}$, where $N \in \mathbb{N} \cup \{\infty\}$, satisfy $\mathbb{P}[\Omega_n] > 0$ for any n, $\Omega_i \cap \Omega_j = \emptyset$ for any i, j with $i \neq j$ and $\bigcup_{n=1}^N \Omega_n = \Omega$. Also let $A \in \mathcal{F}$. Prove the following statements:
(1) $\mathbb{P}[A] = \sum_{n=1}^N \mathbb{P}[A \mid \Omega_n]\mathbb{P}[\Omega_n]$.
(2) (Bayes' theorem) If $\mathbb{P}[A] > 0$, then for each n,

$$\mathbb{P}[\Omega_n \mid A] = \frac{\mathbb{P}[A \mid \Omega_n]\mathbb{P}[\Omega_n]}{\sum_{k=1}^N \mathbb{P}[A \mid \Omega_k]\mathbb{P}[\Omega_k]}. \tag{1.61}$$

Exercise 1.8. Suppose people have a certain disease with probability 0.001. Doctors use a test to detect the disease, and suppose that the test gives a positive result on a patient with the disease with probability 0.99 and on a patient without it with probability 0.004. Evaluate the probability that one has this disease under the condition that
(1) the result of the test was positive.
(2) the result of the test was negative.

In the problems and the exercises below, $(X, \mathcal{M}, \mu)$ denotes a given measure space.

Problem 1.9. Let $n \in \mathbb{N}$ and let $\{A_i\}_{i=1}^n \subset \mathcal{M}$ satisfy $\mu\big(\bigcup_{i=1}^n A_i\big) < \infty$. Prove the following *inclusion-exclusion formula*:

$$\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n \sum_{1\le i_1<\cdots<i_k\le n} (-1)^{k-1}\mu\left(\bigcap_{\ell=1}^k A_{i_\ell}\right). \tag{1.62}$$

Problem 1.10. Prove the following *countable subadditivity* of μ: for $\{A_n\}_{n=1}^\infty \subset \mathcal{M}$,

$$\mu\left(\bigcup_{n=1}^\infty A_n\right) \le \sum_{n=1}^\infty \mu(A_n). \tag{1.63}$$

(Set $B_1 := A_1$ and $B_n := A_n \setminus \bigcup_{i=1}^{n-1} A_i$, $n \ge 2$, and show that $\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty B_n$.)

Problem 1.11. Let # be the counting measure on $\mathbb{N}$ (recall Example 1.5-(1)). Provide an example of $\{A_n\}_{n=1}^\infty \subset 2^{\mathbb{N}}$ such that $A_n \supset A_{n+1}$ for any $n \in \mathbb{N}$ but $\lim_{n\to\infty} \#A_n \neq \#\big(\bigcap_{n=1}^\infty A_n\big)$.

Problem 1.11 shows that the conclusion of Proposition 1.4-(4) is not necessarily valid if the assumption "$\mu(A_1) < \infty$" is dropped.

Problem 1.12. Let Y be a set and define $\mathcal{N} := \{A \subset Y \mid \text{either } A \text{ or } A^c \text{ is countable}\}$ and $\mathcal{N}_0 := \{A \subset Y \mid \text{either } A \text{ or } A^c \text{ is finite}\}$. Prove that $\mathcal{N}$ is a σ-algebra in Y and that $\sigma(\mathcal{N}_0) = \mathcal{N}$.

Problem 1.13. Let $\{A_n\}_{n=1}^\infty \subset 2^X$ and define $\limsup_{n\to\infty} A_n$ and $\liminf_{n\to\infty} A_n$ by

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^\infty \bigcup_{k=n}^\infty A_k, \qquad \liminf_{n\to\infty} A_n := \bigcup_{n=1}^\infty \bigcap_{k=n}^\infty A_k, \tag{1.64}$$

so that they belong to $\mathcal{M}$ if $\{A_n\}_{n=1}^\infty \subset \mathcal{M}$. Prove the following assertions.
(1) $\big(\limsup_{n\to\infty} A_n\big)^c = \liminf_{n\to\infty} A_n^c$ and

$$\begin{aligned}\limsup_{n\to\infty} A_n &= \{x \in X \mid x \in A_n \text{ for infinitely many } n \in \mathbb{N}\},\\ \liminf_{n\to\infty} A_n &= \{x \in X \mid x \in A_n \text{ for sufficiently large } n \in \mathbb{N}\}.\end{aligned} \tag{1.65}$$

(2) (First Borel-Cantelli lemma) If $\{A_n\}_{n=1}^\infty \subset \mathcal{M}$ and $\sum_{n=1}^\infty \mu(A_n) < \infty$, then

$$\mu\left(\limsup_{n\to\infty} A_n\right) = \mu\left(\left(\liminf_{n\to\infty} A_n^c\right)^c\right) = 0. \tag{1.66}$$

(Noting $\limsup_{n\to\infty} A_n \subset \bigcup_{n=k}^\infty A_n$, use the countable subadditivity (1.63) of μ.)

Problem 1.14. Assume $\mu(X) < \infty$. Let Λ be a set and let $\{A_\lambda\}_{\lambda\in\Lambda} \subset \mathcal{M}$ be such that $A_{\lambda_1} \cap A_{\lambda_2} = \emptyset$ for any $\lambda_1, \lambda_2 \in \Lambda$ with $\lambda_1 \neq \lambda_2$. Prove that $\{\lambda \in \Lambda \mid \mu(A_\lambda) > 0\}$ is a countable set. (Show that $\{\lambda \in \Lambda \mid \mu(A_\lambda) \geq 1/n\}$ is finite for any $n \in \mathbb{N}$.)

Problem 1.15. (1) Let S be a set, let $\mathcal{A} \subset 2^S$ and let $f : X \to S$. Prove that f is $\mathcal{M}/\sigma_S(\mathcal{A})$-measurable (see Definition 1.45) if and only if $f^{-1}(A) \in \mathcal{M}$ for any $A \in \mathcal{A}$.
(2) Let $d \in \mathbb{N}$ and let $f = (f_1, \ldots, f_d) : X \to \mathbb{R}^d$, where $f_i : X \to \mathbb{R}$ for each $i \in \{1, \ldots, d\}$. Prove that f is $\mathcal{M}/\mathcal{B}(\mathbb{R}^d)$-measurable if and only if f_i is $\mathcal{M}$-measurable for any $i \in \{1, \ldots, d\}$.

Problem 1.16. (1) Let $f, g : X \to [-\infty, \infty]$ be $\mathcal{M}$-measurable. Prove that the following sets belong to $\mathcal{M}$:

$$\{x \in X \mid f(x) < g(x)\}, \quad \{x \in X \mid f(x) = g(x)\}, \quad \{x \in X \mid f(x) > g(x)\}.$$

(2) Let $f_n : X \to [-\infty, \infty]$ be $\mathcal{M}$-measurable for each $n \in \mathbb{N}$ and let $h : X \to [-\infty, \infty]$ be $\mathcal{M}$-measurable. Define $f, g : X \to [-\infty, \infty]$ by

$$f(x) := \begin{cases} \lim_{n\to\infty} f_n(x) & \text{if the limit } \lim_{n\to\infty} f_n(x) \text{ exists in } \mathbb{R}, \\ h(x) & \text{otherwise,} \end{cases} \tag{1.67}$$

$$g(x) := \begin{cases} \lim_{n\to\infty} f_n(x) & \text{if the limit } \lim_{n\to\infty} f_n(x) \text{ exists in } [-\infty, \infty], \\ h(x) & \text{otherwise.} \end{cases} \tag{1.68}$$

Prove that the functions f and g are $\mathcal{M}$-measurable.

Exercise 1.17. Let $d \in \mathbb{N}$, let $S \subset \mathbb{R}^d$ and let $f : S \to [-\infty, \infty]$.
(1) Let $\varepsilon \in (0, \infty)$ and define $f^\varepsilon, f_\varepsilon : S \to [-\infty, \infty]$ by

$$f^\varepsilon(x) := \sup_{y\in B_S(x,\varepsilon)} f(y) \quad \text{and} \quad f_\varepsilon(x) := \inf_{y\in B_S(x,\varepsilon)} f(y). \tag{1.69}$$

Prove that f^ε and f_ε are Borel measurable. (Show that $(f^\varepsilon)^{-1}\big((a, \infty]\big)$ is open in S.)
(2) Prove that the functions $\overline{f}, \underline{f} : S \to [-\infty, \infty]$ defined by

$$\overline{f}(x) := \limsup_{S\ni y\to x} f(y) \quad \text{and} \quad \underline{f}(x) := \liminf_{S\ni y\to x} f(y) \tag{1.70}$$

are Borel measurable.
(3) Prove that $\{x \in S \mid \lim_{S\ni y\to x} f(y) = f(x)\}$ is a Borel set of S.

Problem 1.18 (Chebyshev's inequality)**.** Let $\varphi : [0, \infty] \to [0, \infty]$ be non-decreasing and let $f : X \to [0, \infty]$ be $\mathcal{M}$-measurable. Prove that $\varphi \circ f$ is $\mathcal{M}$-measurable and that

$$\mu\big(\{x \in X \mid f(x) \geq a\}\big) \leq \frac{1}{\varphi(a)} \int_X (\varphi \circ f) d\mu \tag{1.71}$$

for any $a \in [0, \infty]$ with $\varphi(a) \in (0, \infty)$.

Problem 1.19. Let X be a countable set and let μ be a measure on $(X, 2^X)$.
(1) Prove that any function $f : X \to [-\infty, \infty]$ on X is 2^X-measurable.
(2) Let $f : X \to [0, \infty]$. Prove that $\int_X f d\mu = \sum_{x \in X} f(x)\mu(\{x\})$.

Problem 1.20. Let $A \in \mathcal{M}$, and define a measure $\mu|_A$ on $\mathcal{M}|_A = \{B \cap A \mid B \in \mathcal{M}\}$ by $\mu|_A := \mu|_{\mathcal{M}|_A}$ (note that $\mathcal{M}|_A \subset \mathcal{M}$). Let $f : X \to [-\infty, \infty]$ be $\mathcal{M}$-measurable. Prove that f admits the μ-integral on A if and only if $f|_A$ admits the $\mu|_A$-integral on A, and in that case

$$\int_A f d\mu = \int_A f|_A \, d(\mu|_A). \tag{1.72}$$

(It suffices to prove (1.72) when f is non-negative. Show first for $\mathcal{M}$-simple functions and then use Proposition 1.19 and Theorem 1.24 for general non-negative f.)

According to Problem 1.20, $\int_A f d\mu$ could alternatively be defined as the integral of f with respect to $\mu|_A = \mu|_{\mathcal{M}|_A}$, the *restriction of* μ *to* A.

Problem 1.21. Let $\mathcal{N}$ be a σ-algebra in X such that $\mathcal{N} \subset \mathcal{M}$, and let $f : X \to [-\infty, \infty]$ be $\mathcal{N}$-measurable. Prove that f admits the μ-integral if and only if it admits the $\mu|_{\mathcal{N}}$-integral (note that $\mu|_{\mathcal{N}}$ is a measure on $(X, \mathcal{N})$), and in that case

$$\int_X f d\mu = \int_X f d(\mu|_{\mathcal{N}}). \tag{1.73}$$

Problem 1.22. Let m_1 be the Lebesgue measure on $\mathcal{B}(\mathbb{R})$ introduced in Example 1.8.
(1) Prove that $\mathrm{m}_1(\{a\}) = 0$ for any $a \in \mathbb{R}$.
(2) Let $a, b \in \mathbb{R}$, $a < b$, and let $f : [a, b] \to \mathbb{R}$ be continuous. Prove that

$$\int_{[a,b]} f d\mathrm{m}_1 = \int_a^b f(x)dx, \tag{1.74}$$

where the integral in the right-hand side denotes the Riemann integral on $[a, b]$. (The right hand side is the limit of some Riemann sums. Regarding each Riemann sum as the integral of a simple function, use the dominated convergence theorem.)
(3) Let $a \in \mathbb{R}$ and let $f : [a, \infty) \to \mathbb{R}$ be continuous. Prove that f is m_1-integrable on $[a, \infty)$ if and only if $\lim_{b\to\infty} \int_a^b |f(x)|dx < \infty$,[8] and in that case

$$\int_{[a,\infty)} f d\mathrm{m}_1 = \lim_{b\to\infty} \int_a^b f(x)dx. \tag{1.75}$$

By Problem 1.22-(2), for a continuous function on a bounded closed interval, its integral with respect to the Lebesgue measure m_1 coincides with its Riemann integral. In fact, this fact can be generalized to any Riemann integrable function f on a bounded closed interval of any dimension. See Section 2.6 below for details.

On the other hand, Problem 1.22-(3) says that the same is true also for a continuous function on an unbounded interval *provided the improper Riemann integral is absolutely convergent*. Here the assumption of the absolute convergence is necessary; see Problem 2.14 in this connection.

[8]Note that the limit $\lim_{b\to\infty} \int_a^b |f(x)|dx$ always exists in $[0, \infty]$, since $\int_a^b |f(x)|dx$ is non-decreasing in $b \in (a, \infty)$.

Problem 1.23. Find the limits as $n \to \infty$ of the following integrals:

$$(1) \int_0^\infty \frac{1}{1+x^n} dx \quad (2) \int_0^\infty \frac{\sin e^x}{1+nx^2} dx \quad (3) \int_0^1 \frac{n \cos x}{1+n^2 x^{3/2}} dx$$

Problem 1.24 ([7, Chapter 1, Exercise 9])**.** Let $\alpha \in (0,\infty)$, let $f : X \to [0,\infty]$ be $\mathcal{M}$-measurable and suppose $\int_X f d\mu \in (0,\infty)$. Find the limit (with $\log \infty := \infty^\alpha := \infty$)

$$\lim_{n\to\infty} \int_X n \log\bigl(1 + (f/n)^\alpha\bigr) d\mu.$$

(The integrands are dominated by αf if $\alpha \geq 1$, and otherwise Fatou's lemma applies.)

Exercise 1.25 ([1, Section 4.3, Problem 1])**.** Let $f \in \mathcal{L}^1(\mu)$ and $\{f_n\}_{n=1}^\infty \subset \mathcal{L}^1(\mu)$. Suppose that $f_n \geq 0$ on X for any $n \in \mathbb{N}$, that $\lim_{n\to\infty} f_n(x) = f(x)$ for any $x \in X$, and that $\lim_{n\to\infty} \int_X f_n d\mu = \int_X f d\mu$. Prove that $\lim_{n\to\infty} \int_X |f_n - f| d\mu = 0$.

Problem 1.26. Let $f_n : X \to [-\infty,\infty]$ be $\mathcal{M}$-measurable for each $n \in \mathbb{N}$ and suppose $\sum_{n=1}^\infty \int_X |f_n| d\mu < \infty$. Prove that $\lim_{n\to\infty} f_n(x) = 0$ for μ-a.e. $x \in X$.

Problem 1.27. Let $f : X \to [-\infty,\infty]$. Prove that the following three conditions are equivalent:
(1) f is $\overline{\mathcal{M}}^\mu$-measurable.
(2) There exist $\mathcal{M}$-measurable functions $f_1, f_2 : X \to [-\infty,\infty]$ such that $f_1 \leq f \leq f_2$ on X and $f_1 = f_2$ μ-a.e.
(3) There exists a $\mathcal{M}$-measurable function $f_0 : X \to [-\infty,\infty]$ such that $f_0 = f$ μ-a.e.
((1) $\Rightarrow$ (2): this is easy if f is $\overline{\mathcal{M}}^\mu$-simple. For a general $\overline{\mathcal{M}}^\mu$-measurable function f, take non-decreasing sequences of $\overline{\mathcal{M}}^\mu$-simple functions converging to $f^\pm$.)

Exercise 1.28. Let $f : X \to [0,\infty]$ be $\mathcal{M}$-measurable and μ-integrable. Prove that, for any $\varepsilon \in (0,\infty)$ there exists $\delta \in (0,\infty)$ such that $\int_A f d\mu < \varepsilon$ for any $A \in \mathcal{M}$ with $\mu(A) < \delta$. (Proof by contradiction. Problem 1.13-(2) can be used.)

Problem 1.29. Let $p \in (0,\infty)$ and let $f \in \mathcal{L}^p(\mu)$. Prove that

$$\lim_{n\to\infty} \int_X \bigl|f - f \mathbf{1}_{\{|f|\leq n\}}\bigr|^p d\mu = 0. \tag{1.76}$$

Problem 1.30. Let $p, q \in (0,\infty)$, $p < q$, and let $f : X \to [0,\infty]$ be $\mathcal{M}$-measurable. Prove that

$$\left(\int_X f^p d\mu\right)^{1/p} \leq \left(\int_X f^q d\mu\right)^{1/q} \mu(X)^{(q-p)/pq}. \tag{1.77}$$

By Problem 1.30, if $\mu(X) < \infty$, then $\mathcal{L}^q(X,\mu) \subset \mathcal{L}^p(X,\mu)$ for any $p, q \in (0,\infty)$ with $p < q$.

Problem 1.31 (Minkowski's inequality)**.** Let $p \in [1,\infty)$ and let $f, g : X \to [0,\infty]$ be $\mathcal{M}$-measurable. Use Hölder's inequality to prove that

$$\left(\int_X (f+g)^p d\mu\right)^{1/p} \leq \left(\int_X f^p d\mu\right)^{1/p} + \left(\int_X g^p d\mu\right)^{1/p}. \tag{1.78}$$

Problem 1.32. Assume that $(X, \mathcal{M}, \mu)$ is σ-finite (see Definition 2.24). Let $p \in (1, \infty)$, $q := p/(p-1)$, and let $f : X \to [0, \infty]$ be $\mathcal{M}$-measurable. Prove that

$$\|f\|_{L^p} = \sup\left\{ \int_X fg\,d\mu \;\middle|\; g : X \to [0, \infty],\ g \text{ is } \mathcal{M}\text{-measurable and } \|g\|_{L^q} \le 1 \right\}. \tag{1.79}$$

(Let $g := (h/\|h\|_{L^p})^{p-1}$ for a suitable h with $\|h\|_{L^p} \in (0, \infty)$. Treat the case of $\|f\|_{L^p} < \infty$ and that of $\|f\|_{L^p} = \infty$ separately.)

For the next problem, we need the following definition.

Definition. Let $f : X \to \mathbb{R}$ and $f_n : X \to \mathbb{R}$, $n \in \mathbb{N}$, be $\mathcal{M}$-measurable. We say that $\{f_n\}_{n=1}^{\infty}$ *converges in μ-measure to f* if and only if for any $\varepsilon \in (0, \infty)$,

$$\lim_{n\to\infty} \mu\big(\{x \in X \mid |f_n(x) - f(x)| \ge \varepsilon\}\big) = 0. \tag{1.80}$$

Problem 1.33. Let $f : X \to \mathbb{R}$ and $f_n : X \to \mathbb{R}$, $n \in \mathbb{N}$, be $\mathcal{M}$-measurable.
(1) Let $p \in (0, \infty)$ and suppose $\lim_{n\to\infty} \|f_n - f\|_{L^p(\mu)} = 0$. Prove that $\{f_n\}_{n=1}^{\infty}$ converges in μ-measure to f.
(2) Suppose that $\{f_n\}_{n=1}^{\infty}$ converges in μ-measure to f. Prove that there exists a strictly increasing sequence $\{n_k\}_{k=1}^{\infty} \subset \mathbb{N}$ such that $\lim_{k\to\infty} f_{n_k}(x) = f(x)$ for μ-a.e. $x \in X$. (Choose $n_k \in \mathbb{N}$ so that $\mu\big(\{x \in X \mid |f_{n_k}(x) - f(x)| \ge 2^{-k}\}\big) \le 2^{-k}$ and use Problem 1.13-(2).)

Chapter 2

Construction and Uniqueness of Measures

In this chapter, we provide general criteria for existence and uniqueness of measures and apply them to some important examples. In the latter part of this chapter, we will also discuss products of measures and integration of functions in two variables.

2.1 Uniqueness of Measures: Dynkin System Theorem

The purpose of this section is to state and prove the *Dynkin system theorem*, which is a fundamental tool in probability theory. This theorem enables us to establish various equalities and measurability properties among measures and integrals. As an easy application, a uniqueness theorem for measures is also proved at the last of this section.

Definition 2.1 (π-systems and Dynkin systems)**.** Let X be a set and let $\mathcal{A}, \mathcal{D} \subset 2^X$.
(1) $\mathcal{A}$ is called a *π-system* if and only if $A \cap B \in \mathcal{A}$ for any $A, B \in \mathcal{A}$.
(2) $\mathcal{D}$ is called a *Dynkin system in X* if and only if the following conditions are satisfied:

(D1) $X \in \mathcal{D}$.

(D2) If $A, B \in \mathcal{D}$ and $A \subset B$, then $B \setminus A \in \mathcal{D}$.

(D3) If $\{A_n\}_{n=1}^{\infty} \subset \mathcal{D}$ and $A_n \subset A_{n+1}$ for any $n \in \mathbb{N}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{D}$.

Proposition 2.2. *Let X be a set.*
(1) *Let Λ be a non-empty set and suppose that $\mathcal{D}_\lambda$ is a Dynkin system in X for each $\lambda \in \Lambda$. Then $\bigcap_{\lambda\in\Lambda} \mathcal{D}_\lambda$ is a Dynkin system in X.*
(2) *Let $\mathcal{A} \subset 2^X$ and set*

$$\delta_X(\mathcal{A}) := \bigcap_{\mathcal{D}\text{: Dynkin system in } X,\, \mathcal{A} \subset \mathcal{D}} \mathcal{D}. \tag{2.1}$$

Then $\delta_X(\mathcal{A})$ is the smallest Dynkin system in X that includes $\mathcal{A}$, and $\delta_X(\mathcal{A}) \subset \sigma_X(\mathcal{A})$.

$\mathcal{S}_X(\mathcal{A})$ in (2.1) is called the Dynkin system in X *generated by* $\mathcal{A}$, and it is simply denoted as $\mathcal{S}(\mathcal{A})$ when no confusion can occur.

Proof. $\mathcal{S}_X(\mathcal{A}) \subset \sigma_X(\mathcal{A})$ holds since a σ-algebra in X is a Dynkin system in X. The other assertions are proved in exactly the same way as Proposition 1.7. □

Here is the statement of the Dynkin system theorem.

Theorem 2.3 (Dynkin system theorem)**.** *Let X be a set and let $\mathcal{A} \subset 2^X$ be a π-system. Then*

$$\mathcal{S}(\mathcal{A}) = \sigma(\mathcal{A}). \tag{2.2}$$

We need the following lemma.

Lemma 2.4. *Let X be a set and let $\mathcal{D} \subset 2^X$ be a Dynkin system in X. If $\mathcal{D}$ is a π-system, then it is a σ-algebra in X.*

Proof. $X \in \mathcal{D}$ by (D1), and therefore $\emptyset = X \setminus X \in \mathcal{D}$ and $A^c = X \setminus A \in \mathcal{D}$ for any $A \in \mathcal{D}$ by (D2). If $A, B \in \mathcal{D}$, then $A^c, B^c \in \mathcal{D}$, $A^c \cap B^c \in \mathcal{D}$ by the assumption that $\mathcal{D}$ is a π-system, and hence $A \cup B = (A^c \cap B^c)^c \in \mathcal{D}$. Now let $\{A_n\}_{n=1}^{\infty} \subset \mathcal{D}$. If we set $B_n := \bigcup_{i=1}^{n} A_i$ for $n \in \mathbb{N}$, then $B_n \subset B_{n+1}$, $B_n \in \mathcal{D}$ by the previous argument, and therefore $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n \in \mathcal{D}$ by (D3). Thus $\mathcal{D}$ is a σ-algebra in X. □

Proof of Theorem 2.3. Once we show that $\mathcal{S}(\mathcal{A})$ is a σ-algebra in X, we obtain $\sigma(\mathcal{A}) \subset \mathcal{S}(\mathcal{A})$ and (2.2) follows. By Lemma 2.4, it suffices to show that $\mathcal{S}(\mathcal{A})$ is a π-system.

Let $Y \in \mathcal{S}(\mathcal{A})$ and set $\mathcal{D}_Y := \{A \subset X \mid A \cap Y \in \mathcal{S}(\mathcal{A})\}$. Then $\mathcal{D}_Y$ is a Dynkin system in X. Indeed, $X \cap Y = Y \in \mathcal{S}(\mathcal{A})$ and hence $X \in \mathcal{D}_Y$. If $A, B \in \mathcal{D}_Y$ and $A \subset B$, then $(B \setminus A) \cap Y = (B \cap Y) \setminus (A \cap Y) \in \mathcal{S}(\mathcal{A})$ since $A \cap Y, B \cap Y \in \mathcal{S}(\mathcal{A})$ and $A \cap Y \subset B \cap Y$. If $\{A_n\}_{n=1}^{\infty} \subset \mathcal{D}_Y$ and $A_n \subset A_{n+1}$ for any $n \in \mathbb{N}$, then $A_n \cap Y \in \mathcal{S}(\mathcal{A})$, $A_n \cap Y \subset A_{n+1} \cap Y$ and hence $Y \cap \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty}(A_n \cap Y) \in \mathcal{S}(\mathcal{A})$.

Since $\mathcal{A}$ is a π-system, if $Y \in \mathcal{A}$ then $A \cap Y \in \mathcal{A} \subset \mathcal{S}(\mathcal{A})$ for any $A \in \mathcal{A}$ and hence $\mathcal{A} \subset \mathcal{D}_Y$. Thus $\mathcal{S}(\mathcal{A}) \subset \mathcal{D}_Y$ for any $Y \in \mathcal{A}$, which means that $\mathcal{A} \subset \mathcal{D}_Y$ for any $Y \in \mathcal{S}(\mathcal{A})$. Thus $\mathcal{S}(\mathcal{A}) \subset \mathcal{D}_Y$ for any $Y \in \mathcal{S}(\mathcal{A})$, that is, $\mathcal{S}(\mathcal{A})$ is a π-system. □

Now we present a uniqueness theorem for measures, whose proof illustrates when and how to use the Dynkin system theorem (Theorem 2.3).

Theorem 2.5 (Uniqueness of measures)**.** *Let X be a set, let $\mathcal{A} \subset 2^X$ be a π-system and let $\nu : \mathcal{A} \to [0, \infty]$. Suppose that there exists $\{X_n\}_{n=1}^{\infty} \subset \mathcal{A}$ such that*

$$X = \bigcup_{n=1}^{\infty} X_n \qquad \text{and} \qquad \nu(X_n) < \infty \quad \text{for any } n \in \mathbb{N}. \tag{2.3}$$

Then there exists at most one measure μ on $\sigma(\mathcal{A})$ such that $\mu|_{\mathcal{A}} = \nu$.

Proof. Suppose we have two measures μ_1 and μ_2 on $\sigma(\mathcal{A})$ such that $\mu_1|_{\mathcal{A}} = \mu_2|_{\mathcal{A}} = \nu$. Let $Y \in \mathcal{A}$ satisfy $\mu_1(Y) < \infty$, and define

$$\mathcal{D} := \{A \in \sigma(\mathcal{A}) \mid \mu_1(A \cap Y) = \mu_2(A \cap Y)\}.$$

Then $\mathcal{A} \subset \mathcal{D}$, since for $A \in \mathcal{A}$ we have $A \cap Y \in \mathcal{A}$ and hence $\mu_1(A \cap Y) = \nu(A \cap Y) = \mu_2(A \cap Y)$. We claim that $\mathcal{D}$ is a Dynkin system in X. Indeed, $\mu_i(X \cap Y) = \mu_i(Y) = \nu(Y)$ for $i = 1, 2$ and hence $X \in \mathcal{D}$. If $A, B \in \mathcal{D}$ and $A \subset B$, then by $\mu_1(Y) = \nu(Y) = \mu_2(Y) < \infty$,

$$\begin{aligned}\mu_1\bigl((B \setminus A) \cap Y\bigr) &= \mu_1\bigl((B \cap Y) \setminus (A \cap Y)\bigr) = \mu_1(B \cap Y) - \mu_1(A \cap Y)\\ &= \mu_2(B \cap Y) - \mu_2(A \cap Y) = \mu_2\bigl((B \cap Y) \setminus (A \cap Y)\bigr)\\ &= \mu_2\bigl((B \setminus A) \cap Y\bigr)\end{aligned}$$

and hence $B \setminus A \in \mathcal{D}$. If $\{A_n\}_{n=1}^{\infty} \subset \mathcal{D}$ and $A_n \subset A_{n+1}$ for any $n \in \mathbb{N}$, then

$$\mu_1\left(Y \cap \bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu_1(A_n \cap Y) = \lim_{n\to\infty} \mu_2(A_n \cap Y) = \mu_2\left(Y \cap \bigcup_{n=1}^{\infty} A_n\right)$$

by Proposition 1.4-(3) and hence $\bigcup_{n=1}^{\infty} A_n \in \mathcal{D}$. Thus $\mathcal{D}$ is a Dynkin system in X.

Now using Theorem 2.3, we obtain $\sigma(\mathcal{A}) = \delta(\mathcal{A}) \subset \mathcal{D}$, that is, for $Y \in \mathcal{A}$ with $\mu_1(Y) < \infty$,

$$\mu_1(A \cap Y) = \mu_2(A \cap Y) \qquad \text{for any } A \in \sigma(\mathcal{A}). \tag{2.4}$$

Finally, let $A \in \sigma(\mathcal{A})$ and $n \in \mathbb{N}$. Then the inclusion-exclusion formula (Problem 1.9) yields

$$\mu_i\left(A \cap \bigcup_{j=1}^{n} X_j\right) = \sum_{k=1}^{n} \sum_{1\le j_1<\cdots<j_k\le n} (-1)^{k-1}\mu_i\left(A \cap \bigcap_{\ell=1}^{k} X_{j_\ell}\right) \tag{2.5}$$

for $i = 1, 2$, where $\mu_1\bigl(\bigcap_{\ell=1}^{k} X_{j_\ell}\bigr) \le \mu_1(X_{j_1}) = \nu(X_{j_1}) < \infty$, and $\bigcap_{\ell=1}^{k} X_{j_\ell} \in \mathcal{A}$ since $\mathcal{A}$ is a π-system. Thus (2.4) applies to the right-hand side of (2.5) to imply that $\mu_1\bigl(A \cap \bigcup_{j=1}^{n} X_j\bigr) = \mu_2\bigl(A \cap \bigcup_{j=1}^{n} X_j\bigr)$, and letting $n \to \infty$ results in $\mu_1(A) = \mu_2(A)$ by virtue of Proposition 1.4-(3) and $\bigcup_{n=1}^{\infty} X_n = X$. □

Example 2.6. Let $d \in \mathbb{N}$, let $\mathcal{F}_d$ be as in (1.6), and define $\nu : \mathcal{F}_d \to [0, \infty)$ by

$$\nu\bigl([a_1, b_1] \times \cdots \times [a_d, b_d]\bigr) := (b_1 - a_1)\cdots(b_d - a_d), \qquad \nu(\emptyset) := 0.$$

Then $\mathcal{F}_d$ is clearly a π-system and (2.3) is satisfied with $X_n := [-n, n]^d$. Thus by Theorem 2.5, a measure on $\sigma(\mathcal{F}_d) = \mathcal{B}(\mathbb{R}^d)$ extending ν is unique. This is nothing but the uniqueness of the Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ stated in Example 1.8.

2.2 Construction of Measures

The following theorem is our criterion for construction of measures, which is due to Jun Kigami in Kyoto University and has been borrowed from his unpublished lecture note [6]. We use this theorem in the next section to construct measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Theorem 2.7 (Kigami [6, Theorem 1.4.3]). *Let X be a set, let $\mathcal{A} \subset 2^X$ be a π-system and let $\nu : \mathcal{A} \to [0, \infty]$. Suppose that the following three conditions are satisfied:*

(C1) *$\emptyset \in \mathcal{A}$ and $\nu(\emptyset) = 0$.*

(C2) *If $A \in \mathcal{A}$, $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ and $A \subset \bigcup_{n=1}^{\infty} A_n$, then $\nu(A) \leq \sum_{n=1}^{\infty} \nu(A_n)$.*

(C3) *For any $A, B \in \mathcal{A}$, there exist $n \in \mathbb{N}$ and $\{A_i\}_{i=1}^{n} \subset \mathcal{A}$ such that $A \setminus B \subset \bigcup_{i=1}^{n} A_i$ and $\nu(A) \geq \nu(A \cap B) + \sum_{i=1}^{n} \nu(A_i)$.*

Then the set function $\mu : \sigma(\mathcal{A}) \to [0, \infty]$ defined by

$$\mu(A) := \inf\left\{ \sum_{n=1}^{\infty} \nu(A_n) \;\middle|\; \{A_n\}_{n=1}^{\infty} \subset \mathcal{A},\ A \subset \bigcup_{n=1}^{\infty} A_n \right\} \qquad (\inf \emptyset := \infty) \tag{2.6}$$

is a measure on $\sigma(\mathcal{A})$ such that $\mu|_{\mathcal{A}} = \nu$.

The rest of this section is devoted to the proof of Theorem 2.7. We need the following definition and theorem, which are also fundamental in measure theory.

Definition 2.8 (Outer measures)**.** Let X be a set. A set function $\nu : 2^X \to [0, \infty]$ is called an *outer measure on X* if and only if it has the following properties:

(O1) $\nu(\emptyset) = 0$.

(O2) If $A \subset B \subset X$, then $\nu(A) \leq \nu(B)$.

(O3) If $\{A_n\}_{n=1}^{\infty} \subset 2^X$, then $\nu\big(\bigcup_{n=1}^{\infty} A_n\big) \leq \sum_{n=1}^{\infty} \nu(A_n)$. (countable subadditivity)

Moreover, for an outer measure ν on X, we define $\mathcal{M}(\nu) \subset 2^X$ by

$$\mathcal{M}(\nu) := \{A \subset X \mid \nu(E) = \nu(E \cap A) + \nu(E \setminus A) \text{ for any } E \subset X\}. \tag{2.7}$$

Each $A \in \mathcal{M}(\nu)$ is called *ν-measurable*.

Note that an outer measure ν on a set X satisfies $\nu(E) \leq \nu(E \cap A) + \nu(E \setminus A)$ for any $A, E \subset X$ by (O1), (O3) and $E = (E \cap A) \cup (E \setminus A) \cup \emptyset \cup \emptyset \cup \dots$, and hence that $A \subset X$ belongs to $\mathcal{M}(\nu)$ if and only if $\nu(E) \geq \nu(E \cap A) + \nu(E \setminus A)$ for any $E \subset X$.

Theorem 2.9 (Carathéodory's theorem)**.** *Let X be a set and let ν be an outer measure on X. Then $\mathcal{M}(\nu)$ is a σ-algebra in X and $\nu|_{\mathcal{M}(\nu)}$ is a complete measure on $\mathcal{M}(\nu)$.*

Proof. Let $E \subset X$. $\emptyset \in \mathcal{M}(\nu)$ since $\nu(E) = 0 + \nu(E) = \nu(E \cap \emptyset) + \nu(E \setminus \emptyset)$. If $A \in \mathcal{M}(\nu)$, then $\nu(E) = \nu(E \cap A) + \nu(E \setminus A) = \nu(E \setminus A^c) + \nu(E \cap A^c)$ and hence $A^c \in \mathcal{M}(\nu)$. If $A, B \in \mathcal{M}(\nu)$ then $A \cup B \in \mathcal{M}(\nu)$, because

$$\begin{aligned}
&\nu\big(E \cap (A \cup B)\big) + \nu\big(E \setminus (A \cup B)\big) \\
&= \nu\big(E \cap (A \cup B) \cap A\big) + \nu\big(E \cap (A \cup B) \setminus A\big) + \nu(E \cap A^c \cap B^c) \\
&= \nu(E \cap A) + \nu\big((E \setminus A) \cap B)\big) + \nu\big((E \setminus A) \setminus B)\big) \\
&= \nu(E \cap A) + \nu(E \setminus A) = \nu(E).
\end{aligned}$$

Let $\{A_n\}_{n=1}^{\infty} \subset \mathcal{M}(\nu)$. Set $B_1 := A_1$ and $B_n := A_n \setminus \bigcup_{i=1}^{n-1} A_i$ for $n \geq 2$. Then for any $n \in \mathbb{N}$, we have $\nu\big(E \cap \bigcup_{i=1}^{n+1} A_i\big) = \nu(E \cap B_{n+1}) + \nu\big(E \cap \bigcup_{i=1}^{n} A_i\big)$ by $\bigcup_{i=1}^{n} A_i \in \mathcal{M}(\nu)$, and hence inductively

$$\nu\left(E \cap \bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \nu(E \cap B_i). \tag{2.8}$$

Using $\bigcup_{i=1}^{n} A_i \in \mathcal{M}(\nu)$ and (2.8), we obtain

$$\nu(E) = \nu\left(E \cap \bigcup_{i=1}^{n} A_i\right) + \nu\left(E \setminus \bigcup_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} \nu(E \cap B_i) + \nu\left(E \setminus \bigcup_{i=1}^{\infty} A_i\right), \quad (2.9)$$

where the inequality follows by $\bigcup_{i=1}^{n} A_i \subset \bigcup_{i=1}^{\infty} A_i$ and the condition (O2). Now by letting $n \to \infty$ in (2.9) and using the condition (O3) and $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$, we get

$$\nu(E) \geq \sum_{n=1}^{\infty} \nu(E \cap B_n) + \nu\left(E \setminus \bigcup_{n=1}^{\infty} A_n\right) \geq \nu\left(E \cap \bigcup_{n=1}^{\infty} A_n\right) + \nu\left(E \setminus \bigcup_{n=1}^{\infty} A_n\right), \quad (2.10)$$

where actually the equalities hold by the remark after Definition 2.8. Thus $\bigcup_{n=1}^{\infty} A_n \in \mathcal{M}(\nu)$, proving that $\mathcal{M}(\nu)$ is a σ-algebra in X. If in addition $A_i \cap A_j = \emptyset$ for any $i, j \in \mathbb{N}$ with $i \neq j$, then $B_n = A_n$ in the above argument, and hence the equalities in (2.10) with $E := \bigcup_{n=1}^{\infty} A_n$ yield $\nu\big(\bigcup_{n=1}^{\infty} A_n\big) = \sum_{n=1}^{\infty} \nu(A_n)$. Thus $\nu|_{\mathcal{M}(\nu)}$ is a measure on $\mathcal{M}(\nu)$.

Finally, if $A \subset X$ and $\nu(A) = 0$, then $\nu(E \cap A) + \nu(E \setminus A) \leq \nu(A) + \nu(E) = \nu(E)$ for any $E \subset X$ by the condition (O2) and hence $A \in \mathcal{M}(\nu)$. In particular, if $A \subset N$ for some $N \in \mathcal{M}(\nu)$ with $\nu(N) = 0$, then $\nu(A) = 0$ by (O2) and hence $A \in \mathcal{M}(\nu)$, proving that $\nu|_{\mathcal{M}(\nu)}$ is a complete measure. □

We also need the following easy lemma.

Lemma 2.10. *Let X be a set, let $\mathcal{A} \subset 2^X$ and let $\nu : \mathcal{A} \to [0, \infty]$. Suppose $\emptyset \in \mathcal{A}$ and $\nu(\emptyset) = 0$. Then the set function $\nu_* : 2^X \to [0, \infty]$ defined by*

$$\nu_*(A) := \inf\left\{ \sum_{n=1}^{\infty} \nu(A_n) \;\middle|\; \{A_n\}_{n=1}^{\infty} \subset \mathcal{A},\ A \subset \bigcup_{n=1}^{\infty} A_n \right\} \qquad (\inf \emptyset := \infty) \quad (2.11)$$

is an outer measure on X.

The proof of Lemma 2.10 is left to the reader as an exercise (Problem 2.3).

Proof of Theorem 2.7. Define $\nu_* : 2^X \to [0, \infty]$ by (2.11), so that $\mu = \nu_*|_{\sigma(\mathcal{A})}$. ν_* is an outer measure on X by (C1) and Lemma 2.10. For $A \in \mathcal{A}$, (C2) yields $\nu(A) \leq \nu_*(A)$, and we obtain $\nu_*(A) \leq \nu(A)$ by choosing $A_1 := A$ and $A_n := \emptyset$ for $n \geq 2$ in (2.11). Thus $\mu|_{\mathcal{A}} = \nu_*|_{\mathcal{A}} = \nu$.

To show that $\mathcal{A} \subset \mathcal{M}(\nu_*)$, let $A \in \mathcal{A}$ and $E \subset X$. It suffices to show that $\nu_*(E) \geq \nu_*(E \cap A) + \nu_*(E \setminus A)$, for which we may assume $\nu_*(E) < \infty$. Let $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ satisfy $E \subset \bigcup_{n=1}^{\infty} A_n$. Then $E \cap A \subset \bigcup_{n=1}^{\infty}(A_n \cap A)$, and $\{A_n \cap A\}_{n=1}^{\infty} \subset \mathcal{A}$ since $\mathcal{A}$ is a π-system. Moreover, $E \setminus A \subset \bigcup_{n=1}^{\infty}(A_n \setminus A)$ and, by (C3), for each $n \in \mathbb{N}$ there exist $m_n \in \mathbb{N}$ and $\{B_{n,k}\}_{k=1}^{m_n} \subset \mathcal{A}$ such that $A_n \setminus A \subset \bigcup_{k=1}^{m_n} B_{n,k}$ and $\nu(A_n) \geq \nu(A_n \cap A) + \sum_{k=1}^{m_n} \nu(B_{n,k})$. Thus $\{B_{n,k}\}_{n\in\mathbb{N}, 1\leq k\leq m_n} \subset \mathcal{A}$ and $E \setminus A \subset \bigcup_{n=1}^{\infty} \bigcup_{k=1}^{m_n} B_{n,k}$, and it follows that

$$\sum_{n=1}^{\infty} \nu(A_n) \geq \sum_{n=1}^{\infty} \nu(A_n \cap A) + \sum_{n=1}^{\infty} \sum_{k=1}^{m_n} \nu(B_{n,k}) \geq \nu_*(E \cap A) + \nu_*(E \setminus A).$$

Hence $\nu_*(E) \geq \nu_*(E \cap A) + \nu_*(E \setminus A)$, proving $A \in \mathcal{M}(\nu_*)$ and $\mathcal{A} \subset \mathcal{M}(\nu_*)$. Now since $\mathcal{M}(\nu_*)$ is a σ-algebra in X and $\nu_*|_{\mathcal{M}(\nu_*)}$ is a measure on $\mathcal{M}(\nu_*)$ by Theorem 2.9, $\mathcal{A} \subset \mathcal{M}(\nu_*)$ yields $\sigma(\mathcal{A}) \subset \mathcal{M}(\nu_*)$, and $\mu = \nu_*|_{\sigma(\mathcal{A})} = (\nu_*|_{\mathcal{M}(\nu_*)})|_{\sigma(\mathcal{A})}$ is a measure on $\sigma(\mathcal{A})$. □

2.3 Borel Measures on $\mathbb{R}^d$ and Distribution Functions

In this section, we construct *Borel measures on* $\mathbb{R}^d$ (i.e. measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$) by using Theorem 2.7. At the last of this section, we will also present a useful result concerning approximation of measures by open sets and compact sets.

2.3.1 Borel measures on $\mathbb{R}$: Lebesgue-Stieltjes measures

This subsection is devoted to the construction of Borel measures on $\mathbb{R}$ from right-continuous non-decreasing functions on $\mathbb{R}$. In particular, we prove the existence of the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ stated in Example 1.8.

Definition 2.11. A function $F : \mathbb{R} \to \mathbb{R}$ is called *right-continuous* if and only if[1]

$$\lim_{y \downarrow x} F(y) = F(x) \qquad \text{for any } x \in \mathbb{R}. \tag{2.12}$$

Proposition 2.12. *Let μ be a Borel measure on $\mathbb{R}$ such that $\mu\big((-n, n]\big) < \infty$ for any $n \in \mathbb{N}$. Define $F : \mathbb{R} \to \mathbb{R}$ by*

$$F(x) := \begin{cases} \mu\big((0, x]\big) & \text{if } x \in (0, \infty), \\ 0 & \text{if } x = 0, \\ -\mu\big((x, 0]\big) & \text{if } x \in (-\infty, 0]. \end{cases} \tag{2.13}$$

Then F is right-continuous, non-decreasing and satisfies $\mu\big((a, b]\big) = F(b) - F(a)$ for any $a, b \in \mathbb{R}$ with $a < b$.

Proof. For $a, b \in \mathbb{R}$ with $a < b$, $\mu\big((a, b]\big) = F(b) - F(a)$ easily follows since $\mu\big((x, y]\big) = \mu\big((x, z]\big) + \mu\big((z, y]\big)$ for any $x, y, z \in \mathbb{R}$ with $x < z < y$ by Proposition 1.4-(1). In particular, $F(b) - F(a) \geq 0$, that is, F is non-decreasing.

Let $x \in \mathbb{R}$ and let $\{x_n\}_{n=1}^{\infty} \subset (x, \infty)$ be a non-increasing sequence converging to x. Then Proposition 1.4-(4) yields

$$|F(x_n) - F(x)| = \mu\big((x, x_n]\big) \xrightarrow{n \to \infty} \mu\left(\bigcap_{n=1}^{\infty} (x, x_n] \right) = \mu(\emptyset) = 0,$$

which means that $\lim_{y \downarrow x} F(y) = F(x)$. □

Conversely, any right-continuous non-decreasing function on $\mathbb{R}$ gives rise to exactly one Borel measure on $\mathbb{R}$, as follows.

Theorem 2.13. *Let $F : \mathbb{R} \to \mathbb{R}$ be right-continuous and non-decreasing. Then there exists a unique Borel measure μ_F on $\mathbb{R}$ such that $\mu_F\big((a, b]\big) = F(b) - F(a)$ for any $a, b \in \mathbb{R}$ with $a < b$.*

μ_F is called the *Lebesgue-Stieltjes measure associated with F*.

[1]For $a \in \mathbb{R}$, $\lim_{y \downarrow x} F(y) = a$ (resp. $\lim_{y \uparrow x} F(y) = a$) means that for any $\varepsilon \in (0, \infty)$ there exists $\delta \in (0, \infty)$ such that $|F(y) - a| < \varepsilon$ for any $y \in (x, x + \delta)$ (resp. for any $y \in (x - \delta, x)$).

Proof. Let $\mathcal{A} := \{(a,b] \mid a,b \in \mathbb{R}, a < b\} \cup \{\emptyset\}$ and define $\nu : \mathcal{A} \to [0,\infty)$ by $\nu(\emptyset) := 0$ and $\nu\big((a,b]\big) := F(b) - F(a)$. Clearly $\mathcal{A}$ is a π-system, $\mathcal{A} \subset \mathcal{B}(\mathbb{R})$ and hence $\sigma(\mathcal{A}) \subset \mathcal{B}(\mathbb{R})$. Since $[a,b] = \bigcap_{n=1}^{\infty}(a - 1/n, b] \in \sigma(\mathcal{A})$ for any $a,b \in \mathbb{R}$ with $a \leq b$, $\mathcal{B}(\mathbb{R}) = \sigma\big(\{[a,b] \mid a,b \in \mathbb{R}, a \leq b\} \cup \{\emptyset\}\big) \subset \sigma(\mathcal{A})$ by Proposition 1.9. Thus $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$. The condition (2.3) is satisfied with $X_n = (-n,n]$, and therefore Theorem 2.5 implies the uniqueness of a measure on $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$ extending ν, that is, the uniqueness of μ_F.

Thus it remains to verify the conditions (C2) and (C3) of Theorem 2.7. For any $a,b,c,d \in \mathbb{R}$ with $a < b$ and $c < d$, we have

$$(a,b] \cap (c,d] = \begin{cases} \emptyset \\ (a,d] \\ (a,b] \\ (c,d] \\ (c,b] \end{cases} \quad \text{and} \quad (a,b] \setminus (c,d] = \begin{cases} (a,b] & \text{if } d \leq a \text{ or } b \leq c, \\ (d,b] & \text{if } c \leq a < d < b, \\ \emptyset & \text{if } c \leq a < b \leq d, \\ (a,c] \cup (d,b] & \text{if } a < c < d < b, \\ (a,c] & \text{if } a < c < b \leq d, \end{cases} \tag{2.14}$$

from which (C3) easily follows.

For (C2), let $a,b \in \mathbb{R}$, $a < b$. We first show that for any $n \in \mathbb{N}$ and $a_i, b_i \in \mathbb{R}$ with $a_i < b_i$, $i \in \{1,\ldots,n\}$,

$$\nu\big((a,b]\big) \leq \sum_{i=1}^{n} \nu\big((a_i,b_i]\big) \quad \text{whenever} \quad (a,b] \subset \bigcup_{i=1}^{n}(a_i,b_i]. \tag{2.15}$$

The proof is by induction in n. (2.15) is clear if $n = 1$. Suppose (2.15) is valid for $n \in \mathbb{N}$. Let $a_i, b_i \in \mathbb{R}$, $a_i < b_i$, $i \in \{1,\ldots,n+1\}$, and suppose $(a,b] \subset \bigcup_{i=1}^{n+1}(a_i,b_i]$. Then $b \in (a_j,b_j]$ for some $j \in \{1,\ldots,n+1\}$. If $a_j \leq a$ then $(a,b] \subset (a_j,b_j]$ and hence $\nu\big((a,b]\big) \leq \sum_{i=1}^{n+1} \nu\big((a_i,b_i]\big)$. If $a < a_j$, then $(a,a_j] \subset \bigcup_{1 \leq i \leq n+1, i \neq j}(a_i,b_i]$, and the induction hypothesis together with $b \leq b_j$ yields

$$\begin{aligned} \nu\big((a,b]\big) &= F(a_j) - F(a) + F(b) - F(a_j) = \nu\big((a,a_j]\big) + \nu\big((a_j,b]\big) \\ &\leq \sum_{1 \leq i \leq n+1, i \neq j} \nu\big((a_i,b_i]\big) + \nu\big((a_j,b_j]\big) \leq \sum_{i=1}^{n+1} \nu\big((a_i,b_i]\big), \end{aligned}$$

completing the induction procedure. Now let $a_n, b_n \in \mathbb{R}$, $n \in \mathbb{N}$, be such that $a_n < b_n$ and $(a,b] \subset \bigcup_{n=1}^{\infty}(a_n,b_n]$. Let $\varepsilon \in (0,\infty)$. By the right-continuity of F, we can choose $\{\delta_n\}_{n=1}^{\infty} \subset (0,\infty)$ and $\delta \in (0, b-a)$ so that $F(b_n + \delta_n) - F(b_n) \leq 2^{-n}\varepsilon$ for any $n \in \mathbb{N}$ and $F(a + \delta) - F(a) \leq \varepsilon$. Since $[a+\delta, b] \subset \bigcup_{n=1}^{\infty}(a_n, b_n + \delta_n)$, the compactness of $[a+\delta,b]$ yields a finite set $I \subset \mathbb{N}$ such that $(a+\delta,b] \subset [a+\delta,b] \subset \bigcup_{n\in I}(a_n, b_n+\delta_n) \subset \bigcup_{n \in I}(a_n, b_n + \delta_n]$, and then by (2.15),

$$\begin{aligned} \nu\big((a,b]\big) &= F(b) - F(a) \leq \nu\big((a+\delta,b]\big) + \varepsilon \leq \sum_{n \in I} \nu\big((a_n, b_n+\delta_n]\big) + \varepsilon \\ &\leq \sum_{n=1}^{\infty}\Big(\nu\big((a_n,b_n]\big) + 2^{-n}\varepsilon\Big) + \varepsilon = \sum_{n=1}^{\infty} \nu\big((a_n,b_n]\big) + 2\varepsilon. \end{aligned}$$

Letting $\varepsilon \downarrow 0$ results in (C2), completing the proof. □

Corollary 2.14 (Lebesgue measure on $\mathcal{B}(\mathbb{R})$)**.** *There exists a unique Borel measure* m_1 *on* $\mathbb{R}$ *such that* $\mathrm{m}_1\big([a,b]\big) = b - a$ *for any* $a, b \in \mathbb{R}$ *with* $a \leq b$.

As already mentioned in Example 1.8, m_1 is called the *Lebesgue measure on* $\mathbb{R}$.

Proof. Define $F : \mathbb{R} \to \mathbb{R}$ by $F(x) := x$, which is continuous and non-decreasing. Set $\mathrm{m}_1 := \mu_F$. Then Proposition 1.4-(4) implies that for any $a, b \in \mathbb{R}$ with $a \leq b$,

$$\mathrm{m}_1\big([a,b]\big) = \lim_{n\to\infty} \mathrm{m}_1\big((a - 1/n, b]\big) = \lim_{n\to\infty} (b - a + 1/n) = b - a.$$

Thus m_1 is a Borel measure on $\mathbb{R}$ with the desired property. The uniqueness of such a measure easily follows from Theorem 2.5, as described in Example 2.6. □

The case of probability measures is of particular importance.

Definition 2.15 (Distribution functions)**.** Let μ be a *Borel probability measure on* $\mathbb{R}$ (i.e. a probability measure on $\mathcal{B}(\mathbb{R})$). Then the function $F_\mu : \mathbb{R} \to [0,1]$ defined by $F_\mu(x) := \mu\big((-\infty, x]\big)$ is called the *distribution function of* μ.

Similarly to Proposition 2.12, F_μ is right-continuous, non-decreasing and satisfies $\mu\big((a,b]\big) = F_\mu(b) - F_\mu(a)$ for any $a, b \in \mathbb{R}$ with $a < b$. By Theorem 2.13, μ is equal to μ_{F_μ}, the Lebesgue-Stieltjes measure associated with F_μ, and in particular μ is uniquely determined by its distribution function F_μ.

Corollary 2.16. *A function* $F : \mathbb{R} \to \mathbb{R}$ *is the distribution function of a (unique) Borel probability measure on* $\mathbb{R}$ *if and only if* F *is right-continuous, non-decreasing and satisfies* $\lim_{x\to\infty} F(x) = 1$ *and* $\lim_{x\to-\infty} F(x) = 0$.

Proof. Suppose $F(x) = \mu\big((-\infty, x]\big)$, $x \in \mathbb{R}$, for a Borel probability measure μ on $\mathbb{R}$. Then F is right-continuous and non-decreasing, and Proposition 1.4-(3),(4) yield

$$\begin{aligned} 1 = \mu(\mathbb{R}) &= \lim_{n\to\infty} \mu\big((-\infty, n]\big) = \lim_{n\to\infty} F(n) = \lim_{x\to\infty} F(x), \\ 0 = \mu(\emptyset) &= \lim_{n\to\infty} \mu\big((-\infty, -n]\big) = \lim_{n\to\infty} F(-n) = \lim_{x\to-\infty} F(x). \end{aligned}$$

Conversely, suppose F satisfies these conditions, and let $\mu := \mu_F$ be the Lebesgue-Stieltjes measure associated with F. Then Proposition 1.4-(3) yields

$$\begin{aligned} \mu\big((-\infty, x]\big) &= \lim_{n\to\infty} \mu\big((-n, x]\big) = \lim_{n\to\infty} (F(x) - F(-n)) = F(x) \quad \text{for any } x \in \mathbb{R}, \\ \mu(\mathbb{R}) &= \lim_{n\to\infty} \mu\big((-\infty, n]\big) = \lim_{n\to\infty} F(n) = 1. \end{aligned}$$

Thus μ is a Borel probability measure on $\mathbb{R}$ and F is its distribution function. □

According to Corollary 2.16 and the argument after Definition 2.15, $\mu \mapsto F_\mu$ gives a bijection from the set of Borel probability measures on $\mathbb{R}$ to the set

$$\left\{ F : \mathbb{R} \to \mathbb{R} \;\middle|\; \begin{array}{l} F \text{ is right continuous, non-decreasing and satisfies} \\ \lim_{x\to\infty} F(x) = 1 \text{ and } \lim_{x\to-\infty} F(x) = 0 \end{array} \right\},$$

and its inverse map is given by $F \mapsto \mu_F$. Through this bijection, a Borel probability measure on $\mathbb{R}$ is often identified with its distribution function.

2.3.2 Borel probability measures on $\mathbb{R}^d$ and distribution functions

Corollary 2.16 can be generalized to Borel probability measures on $\mathbb{R}^d$, as described below in this subsection.

Definition 2.17 (Distribution functions on $\mathbb{R}^d$)**.** Let $d \in \mathbb{N}$ and let μ be a Borel probability measure on $\mathbb{R}^d$. Then the function $F_\mu : \mathbb{R}^d \to [0,1]$ defined by

$$F_\mu(x_1, \ldots, x_d) := \mu\big((-\infty, x_1] \times \cdots \times (-\infty, x_d]\big) \tag{2.16}$$

is called the *distribution function of* μ.

Proposition 2.18. *Let $d \in \mathbb{N}$, let μ be a Borel probability measure on $\mathbb{R}^d$ and let F_μ be the distribution function of μ.*
(1) *For any $(x_1, \ldots, x_d) \in \mathbb{R}^d$ and any $(h_1, \ldots, h_d) \in [0, \infty)^d$,*

$$\begin{aligned}\mu\big((x_1 - h_1, x_1] \times \cdots \times (x_d - h_d, x_d]\big) \\ = \sum_{(\alpha_1,\ldots,\alpha_d)\in\{0,1\}^d} (-1)^{\sum_{i=1}^d \alpha_i} F_\mu(x_1 - \alpha_1 h_1, \ldots, x_d - \alpha_d h_d) \geq 0,\end{aligned} \tag{2.17}$$

where $(a, a] := \emptyset$ for $a \in \mathbb{R}$.
(2) *For any $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$,*

$$\lim_{\substack{(y_1,\ldots,y_d)\to x \\ y_i \geq x_i,\, i\in\{1,\ldots,d\}}} F_\mu(y_1, \ldots, y_d) = F_\mu(x). \tag{2.18}$$

(3) $\lim_{x\to\infty} F_\mu(x, \ldots, x) = 1$, *and* $\lim_{x_i\to-\infty} F_\mu(x_1, \ldots, x_i, \ldots, x_d) = 0$ *for any $i \in \{1, \ldots, d\}$ and any $x_j \in \mathbb{R}$, $j \in \{1, \ldots, d\} \setminus \{i\}$*
(4) *μ is uniquely determined by its distribution function F_μ.*

The proof of Proposition 2.18 is left to the reader as an exercise (Problem 2.7).

Theorem 2.19. *Let $d \in \mathbb{N}$, and let $F : \mathbb{R}^d \to \mathbb{R}$ satisfy the following conditions:*

(F1) *For any $(x_1, \ldots, x_d) \in \mathbb{R}^d$ and any $(h_1, \ldots, h_d) \in (0, \infty)^d$,*

$$\sum_{(\alpha_1,\ldots,\alpha_d)\in\{0,1\}^d} (-1)^{\sum_{i=1}^d \alpha_i} F(x_1 - \alpha_1 h_1, \ldots, x_d - \alpha_d h_d) \geq 0. \tag{2.19}$$

(F2) $\lim_{h\downarrow 0} F(x_1 + h, \ldots, x_d + h) = F(x_1, \ldots, x_d)$ *for any $(x_1, \ldots, x_d) \in \mathbb{R}^d$.*

(F3) $\lim_{x\to\infty} F(x, \ldots, x) = 1$, *and* $\lim_{x_i\to-\infty} F(x_1, \ldots, x_i, \ldots, x_d) = 0$ *for any $i \in \{1, \ldots, d\}$ and any $x_j \in \mathbb{R}$, $j \in \{1, \ldots, d\} \setminus \{i\}$.*

Then F is the distribution function of a (unique) Borel probability measure on $\mathbb{R}^d$.

Proof. We have already seen the uniqueness of such μ in Proposition 2.18-(4). To use Theorem 2.7 to construct μ, let $\mathcal{A} := \big\{(a_1, b_1] \times \cdots \times (a_d, b_d] \bigm| a_i, b_i \in \mathbb{R}, a_i < b_i, i \in \{1, \ldots, d\}\big\} \cup \{\emptyset\}$. Clearly $\mathcal{A}$ is a π-system, and by using Proposition 1.9 we easily see that $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R}^d)$. We define $\nu : \mathcal{A} \to [0,\infty)$ by setting $\nu\big((x_1-h_1]\times\cdots\times(x_d-h_d,x_d]\big)$ to be the left-hand side of (2.19) for each $(x_1,\ldots,x_d)\in\mathbb{R}^d$ and $(h_1,\ldots,h_d)\in(0,\infty)^d$, and $\nu(\emptyset):=0$. Let us verify the conditions (C2) and (C3) of Theorem 2.7 for $\mathcal{A}$ and ν. Note that

$$F(x-he_i)\le F(x)\qquad \text{for any } x\in\mathbb{R}^d,\ h\in[0,\infty) \text{ and } i\in\{1,\ldots,d\}, \tag{2.20}$$

where $e_i := \big(\mathbf{1}_{\{i\}}(j)\big)_{j=1}^d \in \mathbb{R}^d$; indeed, letting $h_j\to-\infty$ in (2.19) for $j\in\{1,\ldots,d\}\setminus\{i\}$ yields (2.20) by virtue of (F3). Then by (F2) and (2.20), we obtain (2.18) with F in place of F_μ. Note also that $\nu\big((x_1-h_1]\times\cdots\times(x_d-h_d,x_d]\big)$ is equal to the left-hand side of (2.19), not only for $(h_1,\ldots,h_d)\in(0,\infty)^d$ but also for $(h_1,\ldots,h_d)\in[0,\infty)^d$, since the left-hand side of (2.19) is easily seen to be 0 if $h_i=0$ for some $i\in\{1,\ldots,d\}$.

For the rest of the proof, we use the following notations: for $a,b\in\mathbb{R}^d$, $a=(a_1,\ldots,a_d)$, $b=(b_1,\ldots,b_d)$, we write $a\le b$ if and only if $a_i\le b_i$ for each $i\in\{1,\ldots,d\}$, and $a<b$ if and only if $a_i<b_i$ for each $i\in\{1,\ldots,d\}$. We set $(a,b]:=(a_1,b_1]\times\cdots\times(a_d,b_d]$ and $[a,b]:=[a_1,b_1]\times\cdots\times[a_d,b_d]$ if $a\le b$, and $(a,b):=(a_1,b_1)\times\cdots\times(a_d,b_d)$ if $a<b$.

We first prove (C3). Let $x=(x_1,\ldots,x_d)\in\mathbb{R}^d$ and $h^k=(h^k_1,\ldots,h^k_d)\in[0,\infty)^d$, $k=1,2,3$. We claim that for any $G:\mathbb{R}^d\to\mathbb{R}$,

$$\begin{aligned}&\sum_{(\alpha_1,\ldots,\alpha_d)\in\{0,1\}^d}(-1)^{\sum_{i=1}^d\alpha_i}G\Big(x_1-\alpha_1\sum_{k=1}^3h^k_1,\ldots,x_d-\alpha_d\sum_{k=1}^3h^k_d\Big)\\&=\sum_{\substack{(\alpha_1,\ldots,\alpha_d)\in\{0,1\}^d\\(\beta_1,\ldots,\beta_d)\in\{1,2,3\}^d}}(-1)^{\sum_{i=1}^d\alpha_i}G\Big(x_1-\sum_{k=1}^{\beta_1-1}h^k_1-\alpha_1h^{\beta_1}_1,\ldots,x_d-\sum_{k=1}^{\beta_d-1}h^k_d-\alpha_dh^{\beta_d}_d\Big),\end{aligned}\tag{2.21}$$

which we prove by induction in d. (2.21) is immediate if $d=1$. Let $d\ge2$ and suppose (2.21) is valid with $d-1$ in place of d. Then by the induction hypothesis and (2.21) for $d=1$, the right-hand side of (2.21) is equal to

$$\begin{aligned}&\sum_{\substack{(\alpha_1,\ldots,\alpha_d)\in\{0,1\}^d\\\beta\in\{1,2,3\}}}(-1)^{\sum_{i=1}^d\alpha_i}G\Big(x_1-\sum_{k=1}^{\beta-1}h^k_1-\alpha_1h^{\beta}_1,x_2-\alpha_2\sum_{k=1}^3h^k_2,\ldots,x_d-\alpha_d\sum_{k=1}^3h^k_d\Big)\\&=\sum_{(\alpha_1,\ldots,\alpha_d)\in\{0,1\}^d}(-1)^{\sum_{i=1}^d\alpha_i}G\Big(x_1-\alpha_1\sum_{k=1}^3h^k_1,\ldots,x_d-\alpha_d\sum_{k=1}^3h^k_d\Big),\end{aligned}$$

completing the induction procedure of the proof of (2.21). For $G=F$, (2.21) means that

$$\nu\bigg(\Big(x-\sum_{k=1}^3h^k,x\Big]\bigg)=\sum_{(\beta_1,\ldots,\beta_d)\in\{1,2,3\}^d}\nu\bigg(\prod_{i=1}^d\Big(x_i-\sum_{k=1}^{\beta_i}h^k_i,x_i-\sum_{k=1}^{\beta_i-1}h^k_i\Big]\bigg).\tag{2.22}$$

In relation to (2.22), it also holds that

$$\Big(x-\sum_{k=1}^3h^k,x\Big]=\bigcup_{(\beta_1,\ldots,\beta_d)\in\{1,2,3\}^d}\prod_{i=1}^d\Big(x_i-\sum_{k=1}^{\beta_i}h^k_i,x_i-\sum_{k=1}^{\beta_i-1}h^k_i\Big],\tag{2.23}$$

where the right-hand side is a disjoint union. Now if $A,B\in\mathcal{A}$, then we can choose $x\in\mathbb{R}^d$ and $h^k\in[0,\infty)^d$, $k=1,2,3$, so that $A=\big(x-\sum_{k=1}^3h^k,x\big]$ and $A\cap B=(x-h^1-h^2,x-h^1]$, and then (C3) is immediate from (2.23) and (2.22).

It remains to prove (C2). Let $a, b \in \mathbb{R}^d$, $a < b$. We first show that for any $n \in \mathbb{N}$ and $a^i, b^i \in \mathbb{R}^d$ with $a^i \leq b^i$, $i \in \{1, \ldots, n\}$,

$$\nu\big((a,b]\big) \leq \sum_{i=1}^{n} \nu\big((a^i, b^i]\big) \quad \text{whenever} \quad (a,b] \subset \bigcup_{i=1}^{n} (a^i, b^i]. \tag{2.24}$$

The proof is by induction in n. (2.24) is clear from (C3) if $n = 1$. Let $n \in \mathbb{N}$ and suppose (2.24) is valid for n. Let $a^i, b^i \in \mathbb{R}^d$, $a^i \leq b^i$, $i \in \{1, \ldots, n+1\}$, and suppose $(a,b] \subset \bigcup_{i=1}^{n+1}(a^i, b^i]$. We can choose $v^k \in [0,\infty)^d$, $k = 1,2,3$, so that $(a,b] = \big(b - \sum_{k=1}^{3} v^k, b\big]$ and $(a,b] \cap (a^{n+1}, b^{n+1}] = (b - v^1 - v^2, b - v^1]$. For $\beta = (\beta_1, \ldots, \beta_d) \in \{1,2,3\}^d$ we set $I_\beta := \prod_{i=1}^{d}\big(b_i - \sum_{k=1}^{\beta_i} v_i^k, b_i - \sum_{k=1}^{\beta_i - 1} v_i^k\big]$, where $b = (b_1, \ldots, b_d)$. By (2.23) and (2.22),

$$(a,b] = \bigcup_{\beta \in \{1,2,3\}^d} I_\beta \quad \text{(disjoint union)}, \qquad \nu\big((a,b]\big) = \sum_{\beta \in \{1,2,3\}^d} \nu(I_\beta). \tag{2.25}$$

Let $\beta \in \{1,2,3\}^d$, $\beta \neq (2, \ldots, 2)$. Since $(a,b] \cap (a^{n+1}, b^{n+1}] = I_{(2,\ldots,2)}$ we have $I_\beta \cap (a^{n+1}, b^{n+1}] = \emptyset$ and hence $I_\beta \subset \bigcup_{i=1}^{n}(a^i, b^i]$, i.e. $I_\beta \subset \bigcup_{i=1}^{n}\big(I_\beta \cap (a^i, b^i]\big)$. Then the induction hypothesis shows $\nu(I_\beta) \leq \sum_{i=1}^{n} \nu\big(I_\beta \cap (a^i, b^i]\big) = \sum_{i=1}^{n+1} \nu\big(I_\beta \cap (a^i, b^i]\big)$ for $\beta \in \{1,2,3\}^d$, $\beta \neq (2, \ldots, 2)$, and therefore

$$\nu(I_\beta) \leq \sum_{i=1}^{n+1} \nu\big(I_\beta \cap (a^i, b^i]\big), \qquad \beta \in \{1,2,3\}^d. \tag{2.26}$$

On the other hand, for each $i \in \{1, \ldots, n+1\}$, $(a,b] \cap (a^i, b^i] = \bigcup_{\beta \in \{1,2,3\}^d}\big(I_\beta \cap (a^i, b^i]\big)$, and it is easy to see that there exist $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and $h^k = (h_1^k, \ldots, h_d^k) \in [0,\infty)^d$, $k = 1,2,3$, such that for any $\beta = (\beta_1, \ldots, \beta_d) \in \{1,2,3\}^d$,

$$(a,b] \cap (a^i, b^i] = \left(x - \sum_{k=1}^{3} h^k, x\right] \quad \text{and} \quad I_\beta \cap (a^i, b^i] = \prod_{i=1}^{d}\left(x_i - \sum_{k=1}^{\beta_i} h_i^k, x_i - \sum_{k=1}^{\beta_i - 1} h_i^k\right],$$

which together with (2.22) and (C3) yields

$$\sum_{\beta \in \{1,2,3\}^d} \nu\big(I_\beta \cap (a^i, b^i]\big) = \nu\big((a,b] \cap (a^i, b^i]\big) \leq \nu\big((a^i, b^i]\big), \quad i \in \{1, \ldots, n+1\}. \tag{2.27}$$

By combining (2.25), (2.26) and (2.27), we obtain

$$\nu\big((a,b]\big) = \sum_{\beta \in \{1,2,3\}^d} \nu(I_\beta) \leq \sum_{i=1}^{n+1} \sum_{\beta \in \{1,2,3\}^d} \big(I_\beta \cap (a^i, b^i]\big) \leq \sum_{i=1}^{n+1} \nu\big((a^i, b^i]\big),$$

completing the induction procedure and the proof of (2.24).

Now let $a, b \in \mathbb{R}^d$, $a < b$, and let $a^n, b^n \in \mathbb{R}^d$, $n \in \mathbb{N}$, be such that $a^n \leq b^n$ and $(a,b] \subset \bigcup_{n=1}^{\infty}(a^n, b^n]$. Set $\mathbf{1} := (1, \ldots, 1) \in \mathbb{R}^d$ and let $\varepsilon \in (0,\infty)$. By (2.18) for F, we can choose $\{\delta_n\}_{n=1}^{\infty} \subset (0,\infty)$ and $\delta \in (0,\infty)$ so that $\nu\big((a^n, b^n + \delta_n \mathbf{1}]\big) - \nu\big((a^n, b^n]\big) \leq 2^{-n}\varepsilon$ for any $n \in \mathbb{N}$, $a + \delta\mathbf{1} < b$ and $\nu\big((a,b]\big) - \nu\big((a + \delta\mathbf{1}, b]\big) \leq \varepsilon$. Since $[a + \delta\mathbf{1}, b] \subset \bigcup_{n=1}^{\infty}(a^n, b^n + \delta_n \mathbf{1})$, the compactness of $[a + \delta, b]$ yields a finite set $I \subset \mathbb{N}$ such that $(a + \delta\mathbf{1}, b] \subset [a + \delta\mathbf{1}, b] \subset \bigcup_{n \in I}(a^n, b^n + \delta_n \mathbf{1}) \subset \bigcup_{n \in I}(a^n, b^n + \delta_n \mathbf{1}]$, and then by (2.24),

$$\nu\big((a,b]\big) \leq \nu\big((a + \delta\mathbf{1}, b]\big) + \varepsilon \leq \sum_{n \in I} \nu\big((a^n, b^n + \delta_n \mathbf{1}]\big) + \varepsilon$$

$$\leq \sum_{n=1}^{\infty} \Bigl(\nu\bigl((a^n, b^n]\bigr) + 2^{-n}\varepsilon\Bigr) + \varepsilon = \sum_{n=1}^{\infty} \nu\bigl((a^n, b^n]\bigr) + 2\varepsilon.$$

Letting $\varepsilon \downarrow 0$ results in (C2), and by Theorem 2.7 there exists a measure μ on $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R}^d)$ such that $\mu|_{\mathcal{A}} = \nu$. Finally, by (F3), letting $h_i \to \infty$ for $i \in \{1, \dots, d\}$ successively in (2.19) yields $F(x_1, \dots, x_d) = \mu\bigl((-\infty, x_1] \times \cdots \times (-\infty, x_d]\bigr)$ for any $(x_1, \dots, x_d) \in \mathbb{R}^d$, and $\mu(\mathbb{R}^d) = \lim_{x\to\infty} \mu\bigl((-\infty, x]^d\bigr) = \lim_{x\to\infty} F(x, \dots, x) = 1$. □

2.3.3 Topology and Borel measures on $\mathbb{R}^d$

The purpose of this subsection is to prove the following theorem, which asserts that the measure of a Borel set can be approximated from above by open sets and from below by compact sets.

Theorem 2.20. *Let $d \in \mathbb{N}$, and let μ be a Borel measure on $\mathbb{R}^d$ such that $\mu(\mathbb{R}^d) < \infty$. Then for any $A \in \mathcal{B}(\mathbb{R}^d)$,*

$$\mu(A) = \inf\{\mu(U) \mid A \subset U \subset \mathbb{R}^d, U \text{ is open in } \mathbb{R}^d\} \tag{2.28}$$

$$= \sup\{\mu(K) \mid K \subset A, K \text{ is compact}\}. \tag{2.29}$$

Proof. Define

$$\mathcal{A} := \left\{ A \subset \mathbb{R}^d \;\middle|\; \begin{array}{l} \text{for any } \varepsilon \in (0,\infty) \text{ there exist an open subset } U \text{ of } \mathbb{R}^d \text{ and a} \\ \text{closed subset } F \text{ of } \mathbb{R}^d \text{ such that } F \subset A \subset U \text{ and } \mu(U \setminus F) < \varepsilon \end{array} \right\}. \tag{2.30}$$

We prove that $\mathcal{A}$ is a σ-algebra in $\mathbb{R}^d$ containing all closed subsets of $\mathbb{R}^d$. Let $F \subset \mathbb{R}^d$ be closed in $\mathbb{R}^d$ and let $U_n := \bigcup_{x\in F} B_d(x, 1/n)$ for $n \in \mathbb{N}$. Then U_n is open in $\mathbb{R}^d$, $F \subset U_{n+1} \subset U_n$, and $F = \bigcap_{n=1}^{\infty} U_n$ since F is closed in $\mathbb{R}^d$. Therefore $\mu(F) = \lim_{n\to\infty} \mu(U_n)$ by Proposition 1.4-(4) and $\mu(\mathbb{R}^d) < \infty$, and hence $F \in \mathcal{A}$.

We have $\emptyset \in \mathcal{A}$ since $\emptyset$ is closed in $\mathbb{R}^d$. If $A \in \mathcal{A}$, then for $\varepsilon \in (0, \infty)$ and $U, F \subset \mathbb{R}^d$ as in (2.30), $U^c \subset A^c \subset F^c$ and $\mu(F^c \setminus U^c) < \varepsilon$ since $F^c \setminus U^c = U \setminus F$, and hence $A^c \in \mathcal{A}$. Let $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ and $\varepsilon \in (0, \infty)$, and for each $n \in \mathbb{N}$ choose an open subset U_n of $\mathbb{R}^d$ and a closed subset F_n of $\mathbb{R}^d$ so that $F_n \subset A_n \subset U_n$ and $\mu(U_n \setminus F_n) < 2^{-n}\varepsilon$. Then $\bigcup_{n=1}^{\infty} F_n \subset \bigcup_{n=1}^{\infty} A_n \subset \bigcup_{n=1}^{\infty} U_n$ and

$$\mu\left(\bigcup_{n=1}^{\infty} U_n \setminus \bigcup_{n=1}^{\infty} F_n\right) \leq \mu\left(\bigcup_{n=1}^{\infty} (U_n \setminus F_n)\right) \leq \sum_{n=1}^{\infty} \mu(U_n \setminus F_n) < \varepsilon.$$

Since $\mu\bigl(\bigcup_{n=1}^{\infty} F_n\bigr) = \lim_{n\to\infty} \mu\bigl(\bigcup_{i=1}^{n} F_i\bigr)$ by Proposition 1.4-(3), we can take $k \in \mathbb{N}$ such that $\mu\bigl(\bigcup_{n=1}^{\infty} F_n\bigr) \leq \mu\bigl(\bigcup_{n=1}^{k} F_n\bigr) + \varepsilon$, and then $\mu\bigl(\bigcup_{n=1}^{\infty} U_n \setminus \bigcup_{n=1}^{k} F_n\bigr) < 2\varepsilon$, where $\bigcup_{n=1}^{\infty} U_n$ is open and $\bigcup_{n=1}^{k} F_n$ is closed in $\mathbb{R}^d$. Therefore $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ and hence $\mathcal{A}$ is a σ-algebra in $\mathbb{R}^d$.

Thus we conclude that $\mathcal{B}(\mathbb{R}^d) \subset \mathcal{A}$, and this means (2.28) and (2.29) with "closed in $\mathbb{R}^d$" in place of "compact", from which (2.29) also follows since any closed subset F of $\mathbb{R}^d$ satisfies $\mu(F) = \lim_{n\to\infty} \mu\bigl(F \cap [-n, n]^d\bigr)$ with $F \cap [-n, n]^d$ compact. □

The equalities (2.28) and (2.29) are true also for a certain class of infinite Borel measures on $\mathbb{R}^d$, e.g. for the Lebesgue measure. See Exercise 2.9 in this connection.

2.4 Product Measures and Fubini's Theorem

Recall the following basic fact for Riemann integrals: *Let* $f : [0,1]^2 \to \mathbb{R}$ *be bounded and Riemann integrable on* $[0,1]^2$. *If* $f(x,\cdot)$ *and* $f(\cdot,y)$ *are Riemann integrable on* $[0,1]$ *for any* $x, y \in [0,1]$, *then so are* $\int_0^1 f(\cdot,y)dy$ *and* $\int_0^1 f(x,\cdot)dx$, *and*

$$\int_{[0,1]^2} f(z)dz = \int_0^1 \left(\int_0^1 f(x,y)dx \right) dy = \int_0^1 \left(\int_0^1 f(x,y)dy \right) dx. \tag{2.31}$$

The aim of this section is to establish the counterpart of this fact in the framework of measure theory, for which we need the notions of the product of σ-algebras and that of measures. We start with the definition of the product of σ-algebras.

Definition 2.21 (Product σ-algebras)**.** Let $n \in \mathbb{N}$, and for each $i \in \{1,\dots,n\}$ let $(X_i, \mathcal{M}_i)$ be a measurable space. We define $\mathcal{M}_1 \times \cdots \times \mathcal{M}_n \subset 2^{X_1 \times \cdots \times X_n}$ and a σ-algebra $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ in $X_1 \times \cdots \times X_n$ by

$$\mathcal{M}_1 \times \cdots \times \mathcal{M}_n := \big\{ A_1 \times \cdots \times A_n \bigm| A_i \in \mathcal{M}_i \text{ for } i \in \{1,\dots,n\} \big\}, \tag{2.32}$$

$$\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n := \sigma_{X_1 \times \cdots \times X_n}(\mathcal{M}_1 \times \cdots \times \mathcal{M}_n) \, (= \mathcal{M}_1 \text{ if } n = 1). \tag{2.33}$$

$\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ is called the *product* σ*-algebra of* $\{\mathcal{M}_i\}_{i=1}^n$.

Proposition 2.22. *Let* $n, k \in \mathbb{N}$, *and for each* $i \in \{1,\dots,n+k\}$ *let* $(X_i, \mathcal{M}_i)$ *be a measurable space. Then*

$$(\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n) \otimes (\mathcal{M}_{n+1} \otimes \cdots \otimes \mathcal{M}_{n+k}) = \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_{n+k}. \tag{2.34}$$

Proof. Set $Y_1 := X_1 \times \cdots \times X_n$, $Y_2 := X_{n+1} \times \cdots \times X_{n+k}$, $\mathcal{N}_1 := \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ and $\mathcal{N}_2 := \mathcal{M}_{n+1} \otimes \cdots \otimes \mathcal{M}_{n+k}$. Clearly $\mathcal{N}_1 \times \mathcal{N}_2 \supset \mathcal{M}_1 \times \cdots \times \mathcal{M}_{n+k}$ and hence $\mathcal{N}_1 \otimes \mathcal{N}_2 \supset \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_{n+k}$.

For the converse inclusion, define $\mathcal{A} := \{A \subset Y_1 \mid A \times Y_2 \in \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_{n+k}\}$. Clearly $\mathcal{M}_1 \times \cdots \times \mathcal{M}_n \subset \mathcal{A}$. We claim that $\mathcal{A}$ is a σ-algebra in Y_1. Indeed, $\emptyset \in \mathcal{A}$ by $\emptyset \times Y_2 = \emptyset \in \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_{n+k}$, and if $A \in \mathcal{A}$ then $Y_1 \setminus A \in \mathcal{A}$ since $(Y_1 \setminus A) \times Y_2 = Y_1 \times Y_2 \setminus A \times Y_2 \in \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_{n+k}$ If $\{A_i\}_{i=1}^\infty \subset \mathcal{A}$, then $\big(\bigcup_{i=1}^\infty A_i\big) \times Y_2 = \bigcup_{i=1}^\infty (A_i \times Y_2) \in \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_{n+k}$ and hence $\bigcup_{i=1}^\infty A_i \in \mathcal{A}$. Thus we conclude that $\mathcal{N}_1 = \sigma_{Y_1}(\mathcal{M}_1 \times \cdots \times \mathcal{M}_n) \subset \mathcal{A}$, i.e. $A \times Y_2 \in \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_{n+k}$ for any $A \in \mathcal{N}_1$. It follows in exactly the same way that $Y_1 \times B \in \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_{n+k}$ for any $B \in \mathcal{N}_2$. Thus $A \times B = (A \times Y_2) \cap (Y_1 \times B) \in \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_{n+k}$ for any $A \in \mathcal{N}_1$ and $B \in \mathcal{N}_2$, that is, $\mathcal{N}_1 \times \mathcal{N}_2 \subset \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_{n+k}$, and hence $\mathcal{N}_1 \otimes \mathcal{N}_2 \subset \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_{n+k}$. □

The following proposition provides an important example of product σ-algebras.

Proposition 2.23. (1) *Let* $n, k \in \mathbb{N}$. *Then* $\mathcal{B}(\mathbb{R}^{n+k}) = \mathcal{B}(\mathbb{R}^n) \otimes \mathcal{B}(\mathbb{R}^k)$.
(2) *Let* $d \in \mathbb{N}$. *Then* $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R})^{\otimes d} := \mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R})$ *(*d*-fold product).*

Proof. (1) Let $\mathcal{F}_{n+k}$ be as in (1.6) with $d = n + k$. Then clearly $\mathcal{F}_{n+k} \subset \mathcal{B}(\mathbb{R}^n) \times \mathcal{B}(\mathbb{R}^k)$, and therefore $\mathcal{B}(\mathbb{R}^{n+k}) = \sigma(\mathcal{F}_{n+k}) \subset \mathcal{B}(\mathbb{R}^n) \otimes \mathcal{B}(\mathbb{R}^k)$. On the other hand, $\{A \subset \mathbb{R}^n \mid A \times \mathbb{R}^k \in \mathcal{B}(\mathbb{R}^{n+k})\}$ is easily shown to be a σ-algebra in $\mathbb{R}^n$ in the same

way as the above proof of Proposition 2.22, and it contains all open subsets of $\mathbb{R}^n$ and hence all Borel sets of $\mathbb{R}^n$ as well. Thus for $A \in \mathcal{B}(\mathbb{R}^n)$ and $B \in \mathcal{B}(\mathbb{R}^k)$, we have $A \times \mathbb{R}^k \in \mathcal{B}(\mathbb{R}^{n+k})$, similarly $\mathbb{R}^n \times B \in \mathcal{B}(\mathbb{R}^{n+k})$ as well, and hence $A \times B = (A \times \mathbb{R}^k) \cap (\mathbb{R}^n \times B) \in \mathcal{B}(\mathbb{R}^{n+k})$. Therefore $\mathcal{B}(\mathbb{R}^n) \times \mathcal{B}(\mathbb{R}^k) \subset \mathcal{B}(\mathbb{R}^{n+k})$, which yields $\mathcal{B}(\mathbb{R}^n) \otimes \mathcal{B}(\mathbb{R}^k) \subset \mathcal{B}(\mathbb{R}^{n+k})$.
(2) This easily follows by an induction in d using (1) and Proposition 2.22. □

Next we prove the existence and the uniqueness of the product of measures. We need the following definition for the uniqueness statement.

Definition 2.24. Let $(X, \mathcal{M}, \mu)$ be a measure space. Then μ (or $(X, \mathcal{M}, \mu)$) is called *σ-finite* if and only if there exists $\{X_n\}_{n=1}^{\infty} \subset \mathcal{M}$ such that

$$X = \bigcup_{n=1}^{\infty} X_n \qquad \text{and} \qquad \mu(X_n) < \infty \quad \text{for any } n \in \mathbb{N}. \tag{2.35}$$

Note that, by considering $\{\bigcup_{i=1}^{n} X_i\}_{n=1}^{\infty}$ instead of $\{X_n\}_{n=1}^{\infty}$, *in* (2.35) *we may assume without loss of generality that* $X_n \subset X_{n+1}$ *for any* $n \in \mathbb{N}$.

Theorem 2.25 (Product measures)**.** *Let* $n \in \mathbb{N}$, $n \geq 2$, *and for each* $i \in \{1, \dots, n\}$ *let* $(X_i, \mathcal{M}_i, \mu_i)$ *be a measure space. Then there exists a measure* μ *on* $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ *such that for any* $A_i \in \mathcal{M}_i$, $i \in \{1, \dots, n\}$,

$$\mu(A_1 \times \cdots \times A_n) = \mu_1(A_1) \cdots \mu_n(A_n). \tag{2.36}$$

If $(X_i, \mathcal{M}_i, \mu_i)$ *is σ-finite for each* $i \in \{1, \dots, n\}$ *in addition, then such a measure* μ *on* $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ *is unique and σ-finite, and it is denoted as* $\mu_1 \times \cdots \times \mu_n$.

In the latter case, $\mu_1 \times \cdots \times \mu_n$ is called the *product measure of* $\{\mu_i\}_{i=1}^{n}$.

Proof. Define $\nu : \mathcal{M}_1 \times \cdots \times \mathcal{M}_n \to [0, \infty]$ by $\nu(A_1 \times \cdots \times A_n) := \mu_1(A_1) \cdots \mu_n(A_n)$ for $A_n \in \mathcal{M}_i$, $i \in \{1, \dots, n\}$, so that $\nu(\emptyset) = 0$, regardless of how $\emptyset$ is written as an element of $\mathcal{M}_1 \times \cdots \times \mathcal{M}_n$. Since $\mathcal{M}_1 \times \cdots \times \mathcal{M}_n$ is a π-system and generates $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$, Theorem 2.5 applied to $\mathcal{M}_1 \times \cdots \times \mathcal{M}_n$ and ν immediately shows the uniqueness assertion; here the σ-finiteness of $(X_i, \mathcal{M}_i, \mu_i)$ for $i \in \{1, \dots, n\}$ assures the condition (2.3), and then clearly the measure extending ν has to be σ-finite.

By virtue of Proposition 2.22, the existence of μ for general n easily follows from that for $n = 2$ and an induction in n. Thus it suffices to prove the existence of μ when $n = 2$. To apply Theorem 2.7, we need to verify its conditions (C2) and (C3) for $\mathcal{M}_1 \times \mathcal{M}_2$ and ν. For the proof of (C3), let $A_i, B_i \in \mathcal{M}_i$, $i = 1, 2$. Then

$$A_1 \times A_2 \setminus B_1 \times B_2 = \big((A_1 \setminus B_1) \times A_2\big) \cup \big((A_1 \cap B_1) \times (A_2 \setminus B_2)\big)$$

and

$$\begin{aligned}
&\nu\big((A_1 \times A_2) \cap (B_1 \times B_2)\big) + \nu\big((A_1 \setminus B_1) \times A_2\big) + \nu\big((A_1 \cap B_1) \times (A_2 \setminus B_2)\big) \\
&= \mu_1(A_1 \cap B_1)\mu_2(A_2 \cap B_2) + \mu_1(A_1 \cap B_1)\mu_2(A_2 \setminus B_2) + \mu_1(A_1 \setminus B_1)\mu_2(A_2) \\
&= \mu_1(A_1 \cap B_1)\mu_2(A_2) + \mu_1(A_1 \setminus B_1)\mu_2(A_2) = \mu_1(A_1)\mu_2(A_2) = \nu(A_1 \times A_2),
\end{aligned}$$

proving (C3). Next for the proof of (C2), let $A \in \mathcal{M}_1$, $B \in \mathcal{M}_2$, $\{A_i\}_{i=1}^{\infty} \subset \mathcal{M}_1$ and $\{B_i\}_{i=1}^{\infty} \subset \mathcal{M}_2$ be such that $A \times B \subset \bigcup_{i=1}^{\infty}(A_i \times B_i)$. Then $\mathbf{1}_A(x)\mathbf{1}_B(y) \leq \sum_{i=1}^{\infty} \mathbf{1}_{A_i}(x)\mathbf{1}_{B_i}(y)$ for any $(x, y) \in X_1 \times X_2$, and therefore by using Propositions 1.25 and 1.26, we obtain

$$\begin{aligned}\mu_2(B)\mathbf{1}_A(x) &= \int_{X_2} \mathbf{1}_A(x)\mathbf{1}_B(y)d\mu_2(y) \leq \int_{X_2} \biggl(\sum_{i=1}^{\infty} \mathbf{1}_{A_i}(x)\mathbf{1}_{B_i}(y)\biggr) d\mu_2(y) \\ &= \sum_{i=1}^{\infty} \int_{X_2} \mathbf{1}_{A_i}(x)\mathbf{1}_{B_i}(y)d\mu_2(y) = \sum_{i=1}^{\infty} \mu_2(B_i)\mathbf{1}_{A_i}(x)\end{aligned}$$

for each $x \in X_1$, and hence

$$\begin{aligned}\nu(A \times B) = \mu_2(B)\mu_1(A) &= \int_{X_1} \mu_2(B)\mathbf{1}_A(x)d\mu_1(x) \\ &\leq \int_{X_1} \biggl(\sum_{i=1}^{\infty} \mu_2(B_i)\mathbf{1}_{A_i}(x)\biggr) d\mu_1(x) = \sum_{i=1}^{\infty} \int_{X_1} \mu_2(B_i)\mathbf{1}_{A_i}(x)d\mu_1(x) \\ &= \sum_{i=1}^{\infty} \mu_2(B_i)\mu_1(A_i) = \sum_{i=1}^{\infty} \nu(A_i \times B_i),\end{aligned}$$

proving (C2). Now Theorem 2.7 applied to $X_1 \times X_2$, $\mathcal{M}_1 \times \mathcal{M}_2$ and ν shows that ν is extended to a measure μ on $\sigma(\mathcal{M}_1 \times \mathcal{M}_2) = \mathcal{M}_1 \otimes \mathcal{M}_2$. □

Corollary 2.26. *Let $n, k \in \mathbb{N}$, and for each $i \in \{1, \ldots, n+k\}$ let $(X_i, \mathcal{M}_i, \mu_i)$ be a σ-finite measure space. Then*

$$(\mu_1 \times \cdots \times \mu_n) \times (\mu_{n+1} \times \cdots \times \mu_{n+k}) = \mu_1 \times \cdots \times \mu_{n+k}. \tag{2.37}$$

Proof. According to Proposition 2.22, the two measure in (2.37) are defined on the same σ-algebra $(\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n) \otimes (\mathcal{M}_{n+1} \otimes \cdots \otimes \mathcal{M}_{n+k}) = \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_{n+k}$, and they clearly coincide on $\mathcal{M}_1 \times \cdots \times \mathcal{M}_{n+k}$. Now (2.37) follows from the uniqueness of $\mu_1 \times \cdots \times \mu_{n+k}$ stated in Theorem 2.25. □

Theorem 2.25 gives rise to the existence of the Lebesgue measure on $\mathbb{R}^d$, $d \geq 2$. Note that the Lebesgue measure m_1 on $\mathbb{R}$ constructed in Corollary 2.14 is σ-finite and hence that its product $\mathrm{m}_1 \times \cdots \times \mathrm{m}_1$ (d-fold product) is defined and σ-finite.

Corollary 2.27 (Lebesgue measure on $\mathcal{B}(\mathbb{R}^d)$). *Let $d \in \mathbb{N}$ and define $\mathrm{m}_d := \mathrm{m}_1^d := \mathrm{m}_1 \times \cdots \times \mathrm{m}_1$ (d-fold product). Then m_d is the unique Borel measure on $\mathbb{R}^d$ such that for any $a_i, b_i \in \mathbb{R}$ with $a_i \leq b_i$, $i \in \{1, \ldots, d\}$,*

$$\mathrm{m}_d\bigl([a_1, b_1] \times \cdots \times [a_d, b_d]\bigr) = (b_1 - a_1) \cdots (b_d - a_d). \tag{2.38}$$

Moreover, $\mathrm{m}_{n+k} = \mathrm{m}_n \times \mathrm{m}_k$ for any $n, k \in \mathbb{N}$.

As already mentioned in Example 1.8, m_d is called the *Lebesgue measure on* $\mathbb{R}^d$.

Proof. m_d is a measure on $\mathcal{B}(\mathbb{R})^{\otimes d} = \mathcal{B}(\mathbb{R}^d)$ by Proposition 2.23-(2), and it clearly satisfies (2.38). The uniqueness of such m_d is already verified in Example 2.6, and the last assertion is nothing but (2.37) with $\mu_i = \mathrm{m}_1$, $i \in \{1, \dots, n+k\}$. □

We would like to write down integrals with respect to $\mu_1 \times \cdots \times \mu_n$ as iterated integrals with respect to μ_i, $i \in \{1, \dots, n\}$. This is established in Theorem 2.29 below, which requires some preparations concerning measurability of functions. Note that, in view of Proposition 2.22 and Corollary 2.26, it suffices to consider the case of $n = 2$.

Proposition 2.28. *Let $(X, \mathcal{M}), (Y, \mathcal{N})$ be measurable spaces and let $f : X \times Y \to [-\infty, \infty]$ be $\mathcal{M} \otimes \mathcal{N}$-measurable. Then $f(\cdot, y) : X \to [-\infty, \infty]$ is $\mathcal{M}$-measurable for any $y \in Y$, and $f(x, \cdot) : Y \to [-\infty, \infty]$ is $\mathcal{N}$-measurable for any $x \in X$.*

Proof. Let $y \in Y$ and define $\mathcal{A}_y := \{A \subset X \times Y \mid \mathbf{1}_A(\cdot, y) \text{ is } \mathcal{M}\text{-measurable}\}$. Then $\mathcal{M} \times \mathcal{N} \subset \mathcal{A}_y$ since $\mathbf{1}_{A\times B}(\cdot, y) = \mathbf{1}_B(y)\mathbf{1}_A$ for $A \in \mathcal{M}$ and $B \in \mathcal{N}$, and $\mathcal{A}_y$ is a σ-algebra in $X \times Y$; indeed, $\mathbf{1}_\emptyset(\cdot, y) = 0$ yields $\emptyset \in \mathcal{A}_y$, and $\mathbf{1}_{X\times Y\setminus A}(\cdot, y) = 1 - \mathbf{1}_A(\cdot, y)$ shows $X \times Y \setminus A \in \mathcal{A}_y$ for $A \in \mathcal{A}_y$. If $\{A_n\}_{n=1}^\infty \subset \mathcal{A}_y$, then $\mathbf{1}_{\bigcap_{n=1}^\infty (X\times Y\setminus A_n)}(\cdot, y) = \lim_{n\to\infty}(\mathbf{1}_{X\times Y\setminus A_1} \cdots \mathbf{1}_{X\times Y\setminus A_n})(\cdot, y)$ shows $\bigcap_{n=1}^\infty (X \times Y \setminus A_n) \in \mathcal{A}_y$ and hence $\bigcup_{n=1}^\infty A_n = X \times Y \setminus \bigcap_{n=1}^\infty (X \times Y \setminus A_n) \in \mathcal{A}_y$. It follows that $\mathcal{M} \otimes \mathcal{N} \subset \mathcal{A}_y$.

Now let A be either a Borel set of $\mathbb{R}$ or any one of $\{\infty\}$ and $\{-\infty\}$. Then we have $\mathbf{1}_{(f(\cdot,y))^{-1}(A)} = \mathbf{1}_{f^{-1}(A)}(\cdot, y)$, which is $\mathcal{M}$-measurable since $f^{-1}(A) \in \mathcal{M} \otimes \mathcal{N}$. Thus $\big(f(\cdot, y)\big)^{-1}(A) \in \mathcal{M}$, that is, $f(\cdot, y)$ is $\mathcal{M}$-measurable. The $\mathcal{N}$-measurability of $f(x, \cdot)$ is proved in exactly the same way. □

Theorem 2.29 (Fubini's theorem)**.** *Let $(X, \mathcal{M}, \mu), (Y, \mathcal{N}, \nu)$ be σ-finite measure spaces and let $f : X \times Y \to [-\infty, \infty]$ be $\mathcal{M} \otimes \mathcal{N}$-measurable.*
(1) *If $f \geq 0$ on $X \times Y$, then $\int_Y f(\cdot, y)d\nu(y) : X \to [0, \infty]$ is $\mathcal{M}$-measurable, $\int_X f(x, \cdot)d\mu(x) : Y \to [0, \infty]$ is $\mathcal{N}$-measurable, and*

$$\int_{X\times Y} f d(\mu \times \nu) = \int_X \left(\int_Y f(x, y)d\nu(y) \right) d\mu(x) = \int_Y \left(\int_X f(x, y)d\mu(x) \right) d\nu(y). \tag{2.39}$$

(2) *Suppose that any one of $\int_{X\times Y} |f| d(\mu \times \nu)$, $\int_X \big(\int_Y |f(x, y)| d\nu(y)\big) d\mu(x)$ and $\int_Y \big(\int_X |f(x, y)| d\mu(x)\big) d\nu(y)$ is finite. Then $f(x, \cdot)$ is ν-integrable for μ-a.e. $x \in X$ with $\int_Y f(\cdot, y)d\nu(y)$ $\mathcal{M}$-measurable and μ-integrable, $f(\cdot, y)$ is μ-integrable for ν-a.e. $y \in Y$ with $\int_X f(x, \cdot)d\mu(x)$ $\mathcal{N}$-measurable and ν-integrable, f is $\mu \times \nu$-integrable, and* (2.39) *holds.*

Remark 2.30. (1) In the situation of Theorem 2.29-(2), the function $\int_Y f(\cdot, y)d\nu(y)$ is defined only off $M := \{x \in X \mid \int_Y |f(x, y)| d\nu(y) = \infty\}$, which belongs to $\mathcal{M}$ by Theorem 2.29-(1). The first assertion of Theorem 2.29-(2) means that $\mu(M) = 0$ and that the function $\int_Y f(\cdot, y)d\nu(y)$ on $X \setminus M$ is $\mathcal{M}|_{X\setminus M}$-measurable and μ-integrable. The same remark of course applies to $\int_X f(x, \cdot)d\mu(x)$ as well.
(2) Theorem 2.29-(2) is easily verified also for $\mathbb{C}$-*valued* $\mathcal{M} \otimes \mathcal{N}$-measurable f.

Proof of Theorem 2.29. (1) Choose $\{X_n\}_{n=1}^\infty \subset \mathcal{M}$ and $\{Y_n\}_{n=1}^\infty \subset \mathcal{N}$ so that $X = \bigcup_{n=1}^\infty X_n$, $Y = \bigcup_{n=1}^\infty Y_n$, and for any $n \in \mathbb{N}$, $X_n \subset X_{n+1}$, $Y_n \subset Y_{n+1}$, $\mu(X_n) < \infty$

and $\nu(Y_n) < \infty$. Let $n \in \mathbb{N}$ and define

$$\mathcal{A}_n := \{A \in \mathcal{M} \otimes \mathcal{N} \mid \text{all the conclusions of (1) are valid for } f = \mathbf{1}_{A\cap(X_n\times Y_n)}\}.$$

Then by virtue of $\mu(X_n) < \infty$, $\nu(Y_n) < \infty$ and the monotone convergence theorem (Theorem 1.24), we easily see that $\mathcal{A}_n$ is a Dynkin system in $X \times Y$ including $\mathcal{M} \times \mathcal{N}$, and the Dynkin system theorem (Theorem 2.3) yields $\mathcal{M}\otimes\mathcal{N} = \delta(\mathcal{M}\times\mathcal{N}) \subset \mathcal{A}_n$. Thus for $A \in \mathcal{M} \otimes \mathcal{N}$, the conclusions of (1) are valid with $f = \mathbf{1}_{A\cap(X_n\times Y_n)}$ for any $n \in \mathbb{N}$, and letting $n \to \infty$ yields those with $f = \mathbf{1}_A$ by virtue of the monotone convergence theorem (Theorem 1.24), since $\mathbf{1}_{A\cap(X_n\times Y_n)}(x, y)$ is non-decreasing in n and converges to $\mathbf{1}_A(x, y)$ for any $(x, y) \in X \times Y$.

Now for $\mathcal{M}\otimes\mathcal{N}$-measurable $f : X\times Y \to [0,\infty]$, let $\{s_n\}_{n=1}^{\infty}$ be a non-decreasing sequence of non-negative $\mathcal{M} \otimes \mathcal{N}$-simple functions converging to f, as in Proposition 1.19. Then by the previous paragraph and Proposition 1.25, the conclusions of (1) is valid with s_n in place of f, and letting $n \to \infty$ results in (1), again by Theorem 1.24.

(2) $\int_X \big(\int_Y |f(x, y)|d\nu(y)\big)d\mu(x) = \int_Y \big(\int_X |f(x, y)|d\mu(x)\big)d\nu(y) < \infty$ by (2.39) with $|f|$ in place of f, and Proposition 1.30-(3) yields $\mu(M) = \nu(N) = 0$, where

$$M := \big\{x \in X \bigm| \textstyle\int_Y|f(x, y)|d\nu(y) = \infty\big\} \in \mathcal{M},$$
$$N := \big\{y \in Y \bigm| \textstyle\int_X|f(x, y)|d\mu(x) = \infty\big\} \in \mathcal{N}.$$

$\int_Y f(\cdot, y)d\nu(y) = \int_Y f^+(\cdot, y)d\nu(y) - \int_Y f^-(\cdot, y)d\nu(y)$ is defined on $X \setminus M$, and it is $\mathcal{M}|_{X\setminus M}$-measurable and μ-integrable since $\int_Y f^{\pm}(\cdot, y)d\nu(y)$ are $\mathcal{M}$-measurable and μ-integrable by (1). Similarly $\int_X f(x,\cdot)d\mu(x)$ is defined on $Y \setminus N$, $\mathcal{N}|_{Y\setminus N}$-measurable and ν-integrable. Finally, (2.39) applied to $f^{\pm}$ yields (2.39) for f. □

The assumption of σ-finiteness of μ and ν and the integrability assumption in (2) are indeed necessary in Theorem 2.29; see Exercise 2.13 for concrete counterexamples. The assumption of $\mathcal{M} \otimes \mathcal{N}$-measurability of f is much more subtle and there is no easy counterexample that shows its necessity, but the reader should always keep this measurability assumption in mind when using Theorem 2.29.

2.5 Fubini's Theorem for Completed Product Measures

In the last section we have proved Fubini's theorem (Theorem 2.29). In fact, however, it is still insufficient when we consider *complete measures*, e.g. the completion $\overline{\mathrm{m}_d}$ of the Lebesgue measure on $\mathcal{B}(\mathbb{R}^d)$. A simple reason for this is that the product measure $\mu \times \nu$ of two σ-finite measures μ on $(X, \mathcal{M})$ and ν on $(Y, \mathcal{N})$ is usually not complete even if μ and ν are complete; indeed, if $N \in \mathcal{N}$, $N \neq \emptyset$, $\nu(N) = 0$ and $A \subset X$, $A \notin \mathcal{M}$, then $A\times N \subset X\times N \in \mathcal{M}\otimes\mathcal{N}$ and $(\mu\times\nu)(X\times N) = 0$, but $A\times N \notin \mathcal{M}\otimes\mathcal{N}$ since $\mathbf{1}_{A\times N}(\cdot, y) = \mathbf{1}_N(y)\mathbf{1}_A$ is not $\mathcal{M}$-measurable for $y \in N$ (recall Proposition 2.28). As a consequence, we cannot apply Theorem 2.29 directly to $\overline{\mathrm{m}_d}$-integrals of $\overline{\mathcal{B}(\mathbb{R}^d)}^{\mathrm{m}_d}$-measurable functions.

The purpose of this section is to overcome this difficulty by extending Fubini's theorem to the case of the *completion of the product measure*. We first prove a theorem which asserts a certain uniqueness of the completion of a product measure.

Theorem 2.31. *Let $n \in \mathbb{N}$, $n \geq 2$, and for each $i \in \{1, \ldots, n\}$ let $(X_i, \mathcal{M}_i, \mu_i)$ be a σ-finite measure space. Then it holds that*

$$\overline{\mu_1 \times \cdots \times \mu_n} = \overline{\overline{\mu_1} \times \cdots \times \overline{\mu_n}}. \tag{2.40}$$

Proof. We prove that the domains of the two measures in (2.40) coincide, that is,

$$\overline{\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n}^{\mu_1 \times \cdots \times \mu_n} = \overline{\overline{\mathcal{M}_1}^{\mu_1} \otimes \cdots \otimes \overline{\mathcal{M}_n}^{\mu_n}}^{\overline{\mu_1} \times \cdots \times \overline{\mu_n}}. \tag{2.41}$$

Since $\overline{\mu_1} \times \cdots \times \overline{\mu_n}|_{\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n} = \mu_1 \times \cdots \times \mu_n$ by the uniqueness of $\mu_1 \times \cdots \times \mu_n$, (2.41) means that the two measures in (2.40) are both extensions of $\mu_1 \times \cdots \times \mu_n$ to $\overline{\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n}^{\mu_1 \times \cdots \times \mu_n}$, and therefore they are equal by the uniqueness assertion of Theorem 1.37.

Thus it suffices to show (2.41). The inclusion "$\subset$" easily follows from $\mu_1 \times \cdots \times \mu_n = \overline{\mu_1} \times \cdots \times \overline{\mu_n}|_{\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n}$. For the converse inclusion, let us first prove that

$$\overline{\mathcal{M}_1}^{\mu_1} \otimes \cdots \otimes \overline{\mathcal{M}_n}^{\mu_n} \subset \overline{\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n}^{\mu_1 \times \cdots \times \mu_n}. \tag{2.42}$$

For $i \in \{1, \ldots, n\}$, let $A_i \in \overline{\mathcal{M}_i}^{\mu_i}$, and choose $B_i, C_i \in \mathcal{M}_i$ so that $B_i \subset A_i \subset C_i$ and $\mu_i(C_i \setminus B_i) = 0$. Then $B_1 \times \cdots \times B_n \subset A_1 \times \cdots \times A_n \subset C_1 \times \cdots \times C_n$ and

$$\begin{aligned} &\mu_1 \times \cdots \times \mu_n\big(C_1 \times \cdots \times C_n \setminus B_1 \times \cdots \times B_n\big) \\ &\qquad = \mu_1 \times \cdots \times \mu_n\left(\bigcup_{i=1}^{n} \Big(C_1 \times \cdots \times (C_i \setminus B_i) \times \cdots \times C_n\Big)\right) = 0, \end{aligned}$$

where we used (1.63). Thus $A_1 \times \cdots \times A_n \in \overline{\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n}^{\mu_1 \times \cdots \times \mu_n}$, which implies (2.42).

Now let A belong to the right-hand side of (2.41), and choose $B, C \in \overline{\mathcal{M}_1}^{\mu_1} \otimes \cdots \otimes \overline{\mathcal{M}_n}^{\mu_n}$ so that $B \subset A \subset C$ and $\overline{\mu_1} \times \cdots \times \overline{\mu_n}(C \setminus B) = 0$. Then by (2.42) there exist $B_1, B_2, C_1, C_2 \in \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ such that $B_1 \subset B \subset B_2$, $C_1 \subset C \subset C_2$, $\mu_1 \times \cdots \times \mu_n(B_2 \setminus B_1) = 0$ and $\mu_1 \times \cdots \times \mu_n(C_2 \setminus C_1) = 0$. Then $B_1 \subset A \subset C_2$ and $C_2 \setminus B_1 \subset (B_2 \setminus B_1) \cup (C \setminus B) \cup (C_2 \setminus C_1)$, which together with $\mu_1 \times \cdots \times \mu_n = \overline{\mu_1} \times \cdots \times \overline{\mu_n}|_{\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n}$ implies that

$$\mu_1 \times \cdots \times \mu_n(C_2 \setminus B_1) \leq \overline{\mu_1} \times \cdots \times \overline{\mu_n}\big((B_2 \setminus B_1) \cup (C \setminus B) \cup (C_2 \setminus C_1)\big) = 0.$$

Thus $A \in \overline{\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n}^{\mu_1 \times \cdots \times \mu_n}$, proving (2.41). □

Corollary 2.32. *Let $n, k \in \mathbb{N}$. Then $\overline{\mathrm{m}_{n+k}} = \overline{\overline{\mathrm{m}_n} \times \overline{\mathrm{m}_k}}$.*

Proof. This is immediate by $\mathrm{m}_{n+k} = \mathrm{m}_n \times \mathrm{m}_k$ (Corollary 2.27) and Theorem 2.31. □

Now we state and prove Fubini's theorem for the completion of a product measure.

Theorem 2.33 (Fubini's theorem for completion)**.** *Let $(X, \mathcal{M}, \mu), (Y, \mathcal{N}, \nu)$ be complete σ-finite measure spaces and $f : X \times Y \to [-\infty, \infty]$ be $\overline{\mathcal{M} \otimes \mathcal{N}}^{\mu \times \nu}$-measurable.*
(0) $f(\cdot, y) : X \to [-\infty, \infty]$ is $\mathcal{M}$-measurable for ν-a.e. $y \in Y$ and $f(x, \cdot) : Y \to [-\infty, \infty]$ is $\mathcal{N}$-measurable for μ-a.e. $x \in X$.
(1) If $f \geq 0$ on $X \times Y$, then $\int_Y f(\cdot, y) d\nu(y)$ is defined μ-a.e. on X and $\mathcal{M}$-measurable, $\int_X f(x, \cdot) d\mu(x)$ is defined ν-a.e. on Y and $\mathcal{N}$-measurable, and

$$\int_{X \times Y} f d(\overline{\mu \times \nu}) = \int_X \left(\int_Y f(x, y) d\nu(y)\right) d\mu(x) = \int_Y \left(\int_X f(x, y) d\mu(x)\right) d\nu(y). \tag{2.43}$$

(2) *Suppose that any one of* $\int_{X\times Y}|f|d(\overline{\mu\times\nu})$, $\int_X\bigl(\int_Y|f(x,y)|d\nu(y)\bigr)d\mu(x)$ *and* $\int_Y\bigl(\int_X|f(x,y)|d\mu(x)\bigr)d\nu(y)$ *is finite. Then* $f(x,\cdot)$ *is* ν*-integrable for* μ*-a.e.* $x\in X$ *with* $\int_Y f(\cdot,y)d\nu(y)$ $\mathcal{M}$*-measurable and* μ*-integrable,* $f(\cdot,y)$ *is* μ*-integrable for* ν*-a.e.* $y\in Y$ *with* $\int_X f(x,\cdot)d\mu(x)$ $\mathcal{N}$*-measurable and* ν*-integrable,* f *is* $\overline{\mu\times\nu}$*-integrable, and* (2.43) *holds.*

Remark 2.34. (1) In the situation of Theorem 2.33-(1), $\int_Y f(\cdot,y)d\nu(y)$ is defined only off $M:=\{x\in X\mid f(x,\cdot)\text{ is not }\mathcal{N}\text{-measurable}\}$, which belongs to $\mathcal{M}$ by Theorem 2.33-(0) and the completeness of $(X,\mathcal{M},\mu)$. Similarly to Remark 2.30-(1), the first assertion of Theorem 2.33-(1) means that the function $\int_Y f(\cdot,y)d\nu(y)$ on $X\setminus M$ is $\mathcal{M}|_{X\setminus M}$-measurable. The same remark of course applies to $\int_X f(x,\cdot)d\mu(x)$ as well.
(2) The same remarks as those in Remark 2.30 apply to Theorem 2.33-(2).

Proof of Theorem 2.33. By Problem 1.27, there exist $\mathcal{M}\otimes\mathcal{N}$-measurable functions $f_1,f_2:X\times Y\to[-\infty,\infty]$ such that $f_1\le f\le f_2$ on $X\times Y$ and the set $N:=\{z\in X\times Y\mid f_1(z)<f_2(z)\}\in\mathcal{M}\otimes\mathcal{N}$ has $\mu\times\nu$-measure 0. Then $0=\mu\times\nu(N)=\int_X\bigl(\int_Y\mathbf{1}_N(x,y)d\nu(y)\bigr)d\mu(x)$ by Theorem 2.29-(1) and hence $\nu\bigl(\{y\in Y\mid(x,y)\in N\}\bigr)=\int_Y\mathbf{1}_N(x,y)d\nu(y)=0$ for μ-a.e. $x\in X$ by Proposition 1.34. This means that $\nu\bigl(\{y\in Y\mid f_1(x,y)<f_2(x,y)\}\bigr)=0$ for μ-a.e. $x\in X$, that is,

$$f_1(x,\cdot)=f(x,\cdot)=f_2(x,\cdot)\quad\nu\text{-a.e.}\qquad\text{for }\mu\text{-a.e. }x\in X. \tag{2.44}$$

In exactly the same way, we also obtain

$$f_1(\cdot,y)=f(\cdot,y)=f_2(\cdot,y)\quad\mu\text{-a.e.}\qquad\text{for }\nu\text{-a.e. }y\in Y. \tag{2.45}$$

Now by virtue of (2.44), (2.45) and $f_1=f=f_2$ $\mu\times\nu$-a.e., the assertions of Theorem 2.33 are all immediate from Proposition 2.28, Theorem 2.29 and the completeness of μ and ν. □

2.6 Riemann Integrals and Lebesgue Integrals

The purpose of this section is to prove the following theorem, which asserts that Riemann integrals on bounded closed intervals are just special cases of integrals with respect to (the completion of) the Lebesgue measure. Recall that a function $f:X\to\mathbb{C}$ on a set X is called *bounded* if and only if $\sup_{x\in X}|f(x)|<\infty$.

Theorem 2.35. *Let* $d\in\mathbb{N}$, *let* $a_i,b_i\in\mathbb{R}$, $a_i<b_i$ *for each* $i\in\{1,\dots,d\}$ *and set* $I:=[a_1,b_1]\times\cdots\times[a_d,b_d]$. *Let* $f:I\to\mathbb{R}$ *be bounded and Riemann integrable on* I. *Then* $f\in\mathcal{L}^1\bigl(I,\overline{\mathcal{B}(I)}^{\mathrm{m}_d},\overline{\mathrm{m}_d}\bigr)$ *and*

$$\int_I f\,d\overline{\mathrm{m}_d}=\int_I f(x)dx, \tag{2.46}$$

where the integral in the right-hand side denotes the Riemann integral on I.

Proof. For each $n\in\mathbb{N}$, we define Borel measurable functions $g_n,h_n:I\to\mathbb{R}$ by

$$g_n:=\sum_{i_1,\dots,i_d=1}^{2^n}\Bigl(\inf_{x\in I_n(i_1,\dots,i_d)}f(x)\Bigr)\mathbf{1}_{J_n(i_1,\dots,i_d)}+\Bigl(\inf_{x\in I}f(x)\Bigr)\mathbf{1}_{I\setminus J},$$

$$h_n := \sum_{i_1,\ldots,i_d=1}^{2^n} \Bigl(\sup_{x \in I_n(i_1,\ldots,i_d)} f(x) \Bigr) \mathbf{1}_{J_n(i_1,\ldots,i_d)} + \Bigl(\sup_{x \in I} f(x) \Bigr) \mathbf{1}_{I \setminus J},$$

where $J := (a_1, b_1] \times \cdots \times (a_d, b_d]$,

$$I_n(i_1, \ldots, i_d) := \prod_{k=1}^{d} \Bigl[a_k + \frac{i_k - 1}{2^n}(b_k - a_k), a_k + \frac{i_k}{2^n}(b_k - a_k) \Bigr],$$
$$J_n(i_1, \ldots, i_d) := \prod_{k=1}^{d} \Bigl(a_k + \frac{i_k - 1}{2^n}(b_k - a_k), a_k + \frac{i_k}{2^n}(b_k - a_k) \Bigr].$$

Then

$$\inf_{x \in I} f(x) \le g_n \le g_{n+1} \le f \le h_{n+1} \le h_n \le \sup_{x \in I} f(x) \quad \text{on } I,$$

so that $g := \lim_{n\to\infty} g_n$ and $h := \lim_{n\to\infty} h_n$ are defined, Borel measurable and satisfy $\inf_{x\in I} f(x) \le g \le f \le h \le \sup_{x \in I} f(x)$ on I. In particular, we have $g, h \in \mathcal{L}^1(I, \mathcal{B}(I), \mathrm{m}_d)$ by $\mathrm{m}_d(I) < \infty$.

Since $\mathrm{m}_d\bigl(I_n(i_1, \ldots, i_d)\bigr) = \mathrm{m}_d\bigl(J_n(i_1, \ldots, i_d)\bigr)$ and $\mathrm{m}_d(I \setminus J) = 0$, we see that

$$\int_I g_n d\mathrm{m}_d = \sum_{i_1,\ldots,i_d=1}^{2^n} \Bigl(\inf_{x \in I_n(i_1,\ldots,i_d)} f(x) \Bigr) \mathrm{m}_d\bigl(I_n(i_1, \ldots, i_d)\bigr),$$
$$\int_I h_n d\mathrm{m}_d = \sum_{i_1,\ldots,i_d=1}^{2^n} \Bigl(\sup_{x \in I_n(i_1,\ldots,i_d)} f(x) \Bigr) \mathrm{m}_d\bigl(I_n(i_1, \ldots, i_d)\bigr),$$

which both converge to $\int_I f(x)dx$ by the Riemann integrability of f on I. On the other hand, since $\inf_{x\in I} f(x) \le g_n \le h_n \le \sup_{x\in I} f(x)$ on I and $\mathrm{m}_d(I) < \infty$, the dominated convergence theorem (Theorem 1.32) yields

$$\int_I g d\mathrm{m}_d = \lim_{n\to\infty} \int_I g_n d\mathrm{m}_d = \int_I f(x)dx = \lim_{n\to\infty} \int_I h_n d\mathrm{m}_d = \int_I h d\mathrm{m}_d. \quad (2.47)$$

Thus $\int_I (h - g) d\mathrm{m}_d = 0$, which and $h - g \ge 0$ imply $g = h$ m_d-a.e. on I in view of Proposition 1.34-(1). Finally, since $g \le f \le h$ on I and $g = f = h$ m_d-a.e. on I, f is $\overline{\mathcal{B}(I)}^{\mathrm{m}_d}$-measurable by Problem 1.27, and f is $\overline{\mathrm{m}_d}$-integrable on I and $\int_I f d\overline{\mathrm{m}_d} = \int_I g d\mathrm{m}_d = \int_I f(x)dx$ by Proposition 1.35 and (2.47). □

Remark 2.36. *In Theorem* 2.35*, we cannot conclude that f is Borel measurable.* In fact, there exists a Riemann integrable function on I which is NOT Borel measurable.

Notation. In view of Theorem 2.35, an integral $\int_A f d\overline{\mathrm{m}_d}$ with respect to (the completion of) the Lebesgue measure $\overline{\mathrm{m}_d}$ is also denoted as $\int_A f dx$ or $\int_A f(x)dx$:

$$\int_A f dx := \int_A f(x)dx := \int_A f d\overline{\mathrm{m}_d}. \quad (2.48)$$

If $d = 1$ and $A = (a, b)$, $a, b \in [-\infty, \infty]$, $a < b$, then we write

$$\int_a^b f dx := \int_a^b f(x) dx := \int_{(a,b)} f d\overline{\mathrm{m}_1}. \tag{2.49}$$

In short, *an integral on a subset A of $\mathbb{R}^d$ written as $\int_A f dx$ or $\int_A f(x) dx$ will always mean one with respect to (the completion of) the Lebesgue measure $\overline{\mathrm{m}_d}$.*

Remark 2.37. Let $d \in \mathbb{N}$. Elements of $\overline{\mathcal{B}(\mathbb{R}^d)}^{\mathrm{m}_d}$ are called *Lebesgue measurable sets of $\mathbb{R}^d$* and $\overline{\mathcal{B}(\mathbb{R}^d)}^{\mathrm{m}_d}$-measurable functions are called *Lebesgue measurable*. $\overline{\mathcal{B}(\mathbb{R}^d)}^{\mathrm{m}_d}$ is called the *Lebesgue σ-algebra of $\mathbb{R}^d$* or the *σ-algebra of Lebesgue measurable sets of $\mathbb{R}^d$*.

2.7 Change-of-Variables Formula

At the last of this chapter, we prove the invariance of the Lebesgue measure m_d under parallel translations and invertible linear transformations and present the change-of-variables formulas for m_d.

Theorem 2.38. *Let $d \in \mathbb{N}$.*
(1) *If $\alpha \in \mathbb{R}^d$, then*

$$\mathrm{m}_d(A + \alpha) = \mathrm{m}_d(A) \tag{2.50}$$

for any $A \in \mathcal{B}(\mathbb{R}^d)$, where $A + \alpha := \{x + \alpha \mid x \in A\}$.
(2) *If $T : \mathbb{R}^d \to \mathbb{R}^d$ is linear and invertible, then for any $A \in \mathcal{B}(\mathbb{R}^d)$,*

$$\mathrm{m}_d(T(A)) = |\det T| \mathrm{m}_d(A). \tag{2.51}$$

Remark 2.39. (1) Note that $A + \alpha, T(A) \in \mathcal{B}(\mathbb{R}^d)$ in the situation of Theorem 2.38; indeed, since T^{-1} is continuous, it is $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^d)$-measurable by Problem 1.15-(1) (see also Lemma 1.17) and hence $T(A) = (T^{-1})^{-1}(A) \in \mathcal{B}(\mathbb{R}^d)$. The same argument works for $A + \alpha$ as well.
(2) *If $T : \mathbb{R}^d \to \mathbb{R}^d$ is linear and NOT invertible, then $T(A) \in \overline{\mathcal{B}(\mathbb{R}^d)}^{\mathrm{m}_d}$ and $\overline{\mathrm{m}_d}(T(A)) = 0$ for any $A \in \mathcal{B}(\mathbb{R}^d)$.* Indeed, $T(\mathbb{R}^d)$ is contained in a $(d-1)$-dimensional subspace H, which can be written as

$$H = \left\{(x_1, \dots, x_d) \in \mathbb{R}^d \;\middle|\; x_\ell = \textstyle\sum_{1 \le k \le d, k \ne \ell} a_k x_k \right\}$$

for some $\ell \in \{1, \dots, d\}$ and $a_k \in \mathbb{R}$, $k \neq \ell$. Therefore $H \in \mathcal{B}(\mathbb{R}^d)$ and $\mathrm{m}_d(H) = 0$ by Corollary 2.27 and Fubini's theorem (Theorem 2.29-(1)), which implies the claim.

Proof of Theorem 2.38. (1) Set $\mu(A) := \mathrm{m}_d(A + \alpha)$ for $A \in \mathcal{B}(\mathbb{R}^d)$. Then μ is clearly a Borel measure on $\mathbb{R}^d$ and satisfies $\mu\big([a_1, b_1] \times \cdots \times [a_d, b_d]\big) = (b_1 - a_1) \cdots (b_d - a_d)$ for any $a_i, b_i \in \mathbb{R}$ with $a_i \le b_i$, $i \in \{1, \dots, d\}$. Therefore $\mu = \mathrm{m}_d$ by the uniqueness of the Lebesgue measure m_d on $\mathbb{R}^d$ stated in Corollary 2.27.
(2) Set $\mu_T(A) := \mathrm{m}_d(T(A))$ for $A \in \mathcal{B}(\mathbb{R}^d)$, so that μ_T is a Borel measure on $\mathbb{R}^d$. If $d = 1$, then $T(x) = \beta x$ for some $\beta \in \mathbb{R} \setminus \{0\}$ and hence $\mu_T\big([a, b]\big) = |\beta|(b - a) =$

$|\det T|(b-a)$ for any $a, b \in \mathbb{R}$ with $a \leq b$. Therefore $|\det T|^{-1}\mu_T = \mathrm{m}_1$ by the uniqueness of the Lebesgue measure m_1 on $\mathbb{R}$ stated in Corollary 2.14.

Thus we may assume that $d \geq 2$. Let $(e_1, \ldots, e_d)$ be the standard basis of $\mathbb{R}^d$, that is, $e_i = \big(\mathbf{1}_{\{i\}}(k)\big)_{k=1}^d \in \mathbb{R}^d$, $i \in \{1, \ldots, d\}$. Recall that T can be written as $T = T_1 \cdots T_N$ for some $N \in \mathbb{N}$ and T_i, where each T_i is of one of the following three types:

(i) $(T_i e_1, \ldots, T_i e_d)$ is a permutation of $(e_1, \ldots, e_d)$.

(ii) $T_i e_1 = \beta e_1$ and $T_i e_k = e_k$, $k \in \{2, \ldots, d\}$, for some $\beta \in \mathbb{R} \setminus \{0\}$.

(iii) $T_i e_1 = e_1 + e_2$ and $T_i e_k = e_k$, $k \in \{2, \ldots, d\}$.

Since $\det T = (\det T_1) \cdots (\det T_N)$, it suffices to prove $\mu_T = |\det T|\mathrm{m}_d$ when T itself is of one of the above three types. For this purpose, let $a_i, b_i \in \mathbb{R}$ satisfy $a_i \leq b_i$ for $i \in \{1, \ldots, d\}$ and set $I := [a_1, b_1] \times \cdots \times [a_d, b_d]$. If T is of the type (i), then $|\det T| = 1$ and $\mu_T(I) = (b_1 - a_1) \cdots (b_d - a_d) = |\det T|\mathrm{m}_d(I)$. If T of the type (ii), then $|\det T| = |\beta|$ and $\mu_T(I) = |\beta|(b_1 - a_1) \cdots (b_d - a_d) = |\det T|\mathrm{m}_d(I)$.

Now suppose T is of the type (iii). Then $T(I) = J \times [a_3, b_3] \times \cdots \times [a_d, b_d]$ with $J := \{(s, s+t) \mid s \in [a_1, b_1],\ t \in [a_2, b_2]\} \subset \mathbb{R}^2$, which is a parallelogram formed by the vectors $(b_1 - a_1, b_1 - a_1)$ and $(0, b_2 - a_2)$. Therefore it is immediate that $\mathrm{m}_2(J) = (b_1 - a_1)(b_2 - a_2)$, from which and $\det T = 1$ we see that

$$\mu_T(I) = \mathrm{m}_d(T(I)) = \mathrm{m}_2(J)(b_3 - a_3) \cdots (b_d - a_d) = \mathrm{m}_d(I) = |\det T|\mathrm{m}_d(I).$$

Thus if T is of the type (i), (ii) or (iii) as above, then $|\det T|^{-1}\mu_T = \mathrm{m}_d$ by the uniqueness of m_d stated in Corollary 2.14, completing the proof. □

In view of the image measure theorem (Theorem 1.46), Theorem 2.38 yields the following change-of-variables formula.

Corollary 2.40 (Change-of-variables formula: linear version). *Let $d \in \mathbb{N}$, $\alpha \in \mathbb{R}^d$ and let $T : \mathbb{R}^d \to \mathbb{R}^d$ be linear and invertible. Let $f : \mathbb{R}^d \to [-\infty, \infty]$ be Borel measurable (i.e. $\mathcal{B}(\mathbb{R}^d)$-measurable). Then f admits the* m_d*-integral if and only if so does the function $\mathbb{R}^d \ni x \mapsto f(Tx + \alpha)$, and in that case*

$$\int_{\mathbb{R}^d} f(y)dy = \int_{\mathbb{R}^d} f(Tx + \alpha)|\det T|dx. \tag{2.52}$$

Proof. Define $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ by $\varphi(x) := Tx + \alpha$, so that φ is continuous and bijective. By Theorem 2.38, for any $A \in \mathcal{B}(\mathbb{R}^d)$ we have

$$\mathrm{m}_d(\varphi(A)) = \mathrm{m}_d(T(A) + \alpha) = \mathrm{m}_d(T(A)) = |\det T|\mathrm{m}_d(A)$$

and hence

$$\mathrm{m}_d(A) = \mathrm{m}_d\big(\varphi(\varphi^{-1}(A))\big) = |\det T|\mathrm{m}_d\big(\varphi^{-1}(A)\big) \tag{2.53}$$

since $\varphi^{-1}(A) \in \mathcal{B}(\mathbb{R}^d)$ by Problem 1.15-(1) and the continuity of φ, similarly to Remark 2.39-(1). Now (2.53) means $\mathrm{m}_d = |\det T|\mathrm{m}_d \circ \varphi^{-1}$, which and the image measure theorem (Theorem 1.46) immediately show the assertion. □

In fact, we have a much more general change-of-variables formula for the Lebesgue measure. Recall the following notions from multivariable calculus.

Definition 2.41. Let $d \in \mathbb{N}$, let U be an open subset of $\mathbb{R}^d$ and let $\varphi : U \to \mathbb{R}^d$, $\varphi = (\varphi_1, \dots, \varphi_d)$.
(1) φ is called *continuously differentiable*, or simply C^1, if and only if φ is continuous, all its partial derivatives $\partial\varphi_i/\partial x_j$, $i, j \in \{1, \dots, d\}$, exist at any point of U and they are continuous on U. If φ is C^1, then for $x \in U$, its *derivative* (or *Jacobian matrix*) at x is defined as the matrix $D\varphi(x) := \big((\partial\varphi_i/\partial x_j)(x)\big)_{i,j=1}^d$.
(2) φ is called a C^1*-embedding* if and only if φ is C^1 and injective and $D\varphi(x)$ is invertible for any $x \in U$.

Note also the following fact, which follows by the inverse mapping theorem: *if $\varphi : U \to \mathbb{R}^d$ is a C^1-embedding defined on an open subset U of $\mathbb{R}^d$, then its image $\varphi(U)$ is open in $\mathbb{R}^d$ and the inverse $\varphi^{-1} : \varphi(U) \to U$ is also a C^1-embedding.*

Theorem 2.42 (Change-of-variables formula: general version)**.** *Let $d \in \mathbb{N}$, let U be an open subset of $\mathbb{R}^d$ and let $\varphi : U \to \mathbb{R}^d$ be a C^1-embedding. Let $f : \varphi(U) \to [-\infty, \infty]$ be Borel measurable (i.e. $\mathcal{B}(\varphi(U))$-measurable). Then f admits the* m_d*-integral on $\varphi(U)$ if and only if $(f \circ \varphi)|\det D\varphi|$ admits the* m_d*-integral on U, and in that case*

$$\int_{\varphi(U)} f(y)dy = \int_U f(\varphi(x))|\det D\varphi(x)|dx. \tag{2.54}$$

The proof of Theorem 2.42 requires various preparations and is too long to be given here. We refer the interested readers to the proof in Rudin's book [7, Definition 7.22 – Theorem 7.26]. (In fact, the change-of-variables formula [7, Theorem 7.26] in his book is proved under much weaker assumptions than those of Theorem 2.42 above.)

Exercises

Problem 2.1. Let X be a set and let $\mathcal{D} \subset 2^X$. Prove that $\mathcal{D}$ is a Dynkin system in X if and only if $\mathcal{D}$ satisfies the conditions (D1) and (D2) of Definition 2.1-(2) and the following condition (D3)$'$:

(D3)$'$ If $\{A_n\}_{n=1}^\infty \subset \mathcal{D}$ and $A_i \cap A_j = \emptyset$ for any $i, j \in \mathbb{N}$ with $i \neq j$, then $\bigcup_{n=1}^\infty A_n \in \mathcal{D}$.

The next exercise requires the following definition.

Definition. Let X be a set and let $\mathcal{A}, \mathcal{M} \subset 2^X$.
(1) $\mathcal{A}$ is called an *algebra in X* if and only if it possesses the following properties:

(A1) $\emptyset \in \mathcal{A}$.

(A2) If $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$, where $A^c := X \setminus A$.

(A3) If $n \in \mathbb{N}$ and $\{A_i\}_{i=1}^n \subset \mathcal{A}$ then $\bigcup_{i=1}^n A_i \in \mathcal{A}$.

(2) $\mathcal{M}$ is called a *monotone class in* X if and only if it satisfies the following conditions:

(M1) If $\{A_n\}_{n=1}^{\infty} \subset \mathcal{M}$ and $A_n \subset A_{n+1}$ for any $n \in \mathbb{N}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{M}$.

(M2) If $\{A_n\}_{n=1}^{\infty} \subset \mathcal{M}$ and $A_n \supset A_{n+1}$ for any $n \in \mathbb{N}$, then $\bigcap_{n=1}^{\infty} A_n \in \mathcal{M}$.

Exercise 2.2. Let X be a set and let $\mathcal{A} \subset 2^X$.
(1) Prove that

$$\mathcal{M}(\mathcal{A}) := \mathcal{M}_X(\mathcal{A}) := \bigcap_{\mathcal{M}\text{: monotone class in } X,\ \mathcal{A} \subset \mathcal{M}} \mathcal{M} \tag{2.55}$$

is the smallest monotone class in X that includes $\mathcal{A}$, and that $\mathcal{M}(\mathcal{A}) \subset \sigma(\mathcal{A})$.
(2) (Monotone class theorem) Suppose $\mathcal{A}$ is an algebra in X. Prove that

$$\mathcal{M}(\mathcal{A}) = \sigma(\mathcal{A}). \tag{2.56}$$

Problem 2.3. Prove Lemma 2.10.

Problem 2.4 ([4, Corollary 7.1])**.** Let μ be a Borel probability measure on $\mathbb{R}$ and let F be its distribution function. Recalling that F is non-decreasing, we define $F(x-) := \lim_{y \uparrow x} F(y)$ for each $x \in \mathbb{R}$. Let $a, b \in \mathbb{R}$, $a < b$. Prove the following equalities:
(1) $\mu([a,b]) = F(b) - F(a-)$.
(2) $\mu([a,b)) = F(b-) - F(a-)$.
(3) $\mu((a,b)) = F(b-) - F(a)$.
(4) $\mu(\{a\}) = F(a) - F(a-)$. (Thus $\mu(\{a\}) = 0$ if and only if F is continuous at a.)

Problem 2.5. Let F be the distribution function of a Borel probability measure on $\mathbb{R}$. Prove that the set $\{x \in \mathbb{R} \mid F(x) \neq F(x-)\}$ is countable, where $F(x-)$ is as in Problem 2.4. (Noting Problem 2.4-(4), use Problem 1.14.)

Problem 2.6 ([4, Exercise 7.18])**.** Define $F : \mathbb{R} \to \mathbb{R}$ by

$$F := \sum_{n=1}^{\infty} \frac{1}{2^n} \mathbf{1}_{[n^{-1},\infty)}. \tag{2.57}$$

(1) Prove that F is the distribution function of a Borel probability measure μ on $\mathbb{R}$.
(2) Let μ be as in (1). Calculate the following values (i)–(vi):

(i)	$\mu([1,\infty))$	(ii)	$\mu([1/10,\infty))$	(iii)	$\mu(\{0\})$
(iv)	$\mu([0,1/2))$	(v)	$\mu((-\infty,0))$	(vi)	$\mu((0,\infty))$

Problem 2.7. Prove Proposition 2.18. (For (1), use the inclusion-exclusion formula (1.62).)

Problem 2.8. Let $d \in \mathbb{N}$ and let μ be a Borel probability measure on $\mathbb{R}^d$. Define

$$C_{\mu,i} := \{a \in \mathbb{R} \mid \mu(H_i(a)) = 0\}, \quad \text{where } H_i(a) := \{(x_1,\dots,x_d) \in \mathbb{R}^d \mid x_i = a\}, \tag{2.58}$$

for each $i \in \{1,\dots,d\}$ and $C_\mu := C_{\mu,1} \times \cdots \times C_{\mu,d}$. Prove the following statements:
(1) $\mathbb{R} \setminus C_{\mu,i}$ is a countable set for any $i \in \{1,\dots,d\}$. (Use Problem 1.14.)
(2) The distribution function $F_\mu : \mathbb{R}^d \to [0,1]$ of μ is continuous at x for any $x \in C_\mu$.

Exercise 2.9. Let $d \in \mathbb{N}$. Prove that for any $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned} \mathrm{m}_d(A) &= \inf\{\mathrm{m}_d(U) \mid A \subset U \subset \mathbb{R}^d,\, U \text{ is open in } \mathbb{R}^d\} &(2.59)\\ &= \sup\{\mathrm{m}_d(K) \mid K \subset A,\, K \text{ is compact}\}. &(2.60)\end{aligned}$$

Problem 2.10. Let $(X, \mathcal{M})$ be a measurable space. Let $n \in \mathbb{N}$, and for each $i \in \{1, \ldots, n\}$, let $(S_i, \mathcal{B}_i)$ be a measurable space and let $f_i : X \to S_i$. Prove that the map $f = (f_1, \ldots, f_d) : X \to S_1 \times \cdots \times S_n$ is $\mathcal{M}/\mathcal{B}_1 \otimes \cdots \otimes \mathcal{B}_n$-measurable if and only if f_i is $\mathcal{M}/\mathcal{B}_i$-measurable for any $i \in \{1, \ldots, n\}$.

Problem 2.11. Let $n \in \mathbb{N}$. For each $i \in \{1, \ldots, n\}$, let $(X_i, \mathcal{M}_i, \mu_i)$ be a σ-finite measure space and let $f_i : X_i \to [-\infty, \infty]$ be $\mathcal{M}_i$-measurable. Define $f_1 \otimes \cdots \otimes f_n : X_1 \times \cdots \times X_n \to [-\infty, \infty]$ by $(f_1 \otimes \cdots \otimes f_n)(x_1, \ldots, x_n) := f_1(x_1) \cdots f_n(x_n)$. Prove the following statements:
(1) $f_1 \otimes \cdots \otimes f_n$ is $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$-measurable.
(2) If f_i is μ_i-integrable for each $i \in \{1, \ldots, n\}$, then $f_1 \otimes \cdots \otimes f_n$ is $\mu_1 \times \cdots \times \mu_n$-integrable and

$$\int_{X_1 \times \cdots \times X_n} f_1 \otimes \cdots \otimes f_n d(\mu_1 \times \cdots \times \mu_n) = \int_{X_1} f_1 d\mu_1 \cdots \int_{X_n} f_n d\mu_n. \tag{2.61}$$

(Induction in n. Use Proposition 2.22 and Corollary 2.26 to apply Theorem 2.29-(2).)

Problem 2.12. Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space, let $f : X \to [0, \infty]$ be $\mathcal{M}$-measurable and set $S_f := \{(x, t) \in X \times \mathbb{R} \mid 0 \le t < f(x)\}$.
(1) Prove that $S_f \in \mathcal{M} \otimes \mathcal{B}(\mathbb{R})$ and that $[0, \infty) \ni t \mapsto \mu\bigl(\{x \in X \mid f(x) > t\}\bigr) \in [0, \infty]$ is Borel measurable.
(2) Prove that $\int_X f d\mu = \mu \times \mathrm{m}_1(S_f)$ and that for any $p \in (0, \infty)$,

$$\int_X f^p d\mu = p \int_0^\infty t^{p-1} \mu\bigl(\{x \in X \mid f(x) > t\}\bigr) dt. \tag{2.62}$$

(3) Prove that $\mathrm{m}_2\bigl(\{x \in \mathbb{R}^2 \mid |x| < r\}\bigr) = \pi r^2$ for any $r \in (0, \infty)$.

Exercise 2.13 ([7, Counterexamples 8.9])**.** (1) Let # denote the counting measure on $[0, 1]$ and set $\Delta_{[0,1]} := \{(x, y) \in [0, 1]^2 \mid x = y\}$, which is closed in $\mathbb{R}^2$. Prove that

$$\int_0^1 \biggl(\int_{[0,1]} \mathbf{1}_{\Delta_{[0,1]}}(x, y) d\#(y) \biggr) dx = 1 \neq 0 = \int_{[0,1]} \biggl(\int_0^1 \mathbf{1}_{\Delta_{[0,1]}}(x, y) dx \biggr) d\#(y). \tag{2.63}$$

(2) Let $\{\delta_n\}_{n=0}^\infty \subset [0, 1)$ be such that $\delta_0 = 0$, $\delta_{n-1} < \delta_n$ for any $n \in \mathbb{N}$ and $\lim_{n\to\infty} \delta_n = 1$. Also for each $n \in \mathbb{N}$, let $g_n : [0, 1) \to \mathbb{R}$ be a continuous function such that $g_n|_{[0,1)\setminus(\delta_{n-1}, \delta_n)} = 0$ and $\int_0^1 g_n(x) dx = 1$. Define $f : [0, 1)^2 \to \mathbb{R}$ by

$$f(x, y) := \sum_{n=1}^\infty \bigl(g_n(x) - g_{n+1}(x)\bigr) g_n(y). \tag{2.64}$$

Prove the following statements:

(i) f is continuous and $\int_0^1 \big(\int_0^1 |f(x,y)|dx\big)dy = \infty$.
(ii) For any $x, y \in [0,1)$, $f(x,\cdot), f(\cdot,y) \in \mathcal{L}^1\big([0,1), \mathrm{m}_1\big)$, $\int_0^1 f(x,z)dz = g_1(x)$ and $\int_0^1 f(z,y)dz = 0$. In particular,

$$\int_0^1 \left(\int_0^1 f(x,y)dy\right)dx = 1 \neq 0 = \int_0^1 \left(\int_0^1 f(x,y)dx\right)dy. \tag{2.65}$$

Problem 2.14. (1) Prove that

$$\int_0^\infty \left|\frac{\sin x}{x}\right| dx = \infty. \tag{2.66}$$

(2) Use $x^{-1} = \int_0^\infty e^{-xt}dt$, $x \in (0,\infty)$, to prove that

$$\lim_{A\to\infty} \int_0^A \frac{\sin x}{x} dx = \int_0^\infty \frac{1-\cos x}{x^2} dx = \int_0^\infty \left(\frac{\sin x}{x}\right)^2 dx = \frac{\pi}{2}. \tag{2.67}$$

Part II

Probability Theory

Chapter 3

Random Variables and Independence

On the basis of measure theoretic tools developed so far, from this chapter on we present various limit theorems in probability theory. First in this chapter, we introduce various notions concerning random variables including *independence* of random variables, which is one of the most important notions in probability theory, and state the *laws of large numbers* for sequences of independent real random variables. In Section 3.6, we also prove the existence and the uniqueness of the product of an infinite sequence of probability measures, which assures the existence of infinite sequences of independent random variables.

3.1 Random Variables and their Probability Laws

In this section, we give the precise definition of random variables and state basic facts for them, which are more or less immediate from the results of the preceding chapters.

Throughout this section, we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$; recall from Definition 1.3-(2) that a *probability space* is the triple $(\Omega, \mathcal{F}, \mathbb{P})$ of a set Ω, a σ-algebra $\mathcal{F}$ in Ω and a probability measure $\mathbb{P}$ on $\mathcal{F}$. We begin with some probabilistic terminology.

Definition 3.1. (1) The set Ω is called the *sample space of* $(\Omega, \mathcal{F}, \mathbb{P})$.
(2) Each $A \in \mathcal{F}$ is called an *event*. For an event $A \in \mathcal{F}$, $\mathbb{P}[A]$ is called its *probability*.
(3) We use the phrase *"almost surely"* (or *"a.s."* for short) as a synonym for "$\mathbb{P}$-almost everywhere". When an explicit reference to the probability measure $\mathbb{P}$ is necessary, we also say *"$\mathbb{P}$-almost surely"* (or *"$\mathbb{P}$-a.s."* for short).

Definition 3.2 (Random variables)**.** (1) Let $(S, \mathcal{B})$ be a measurable space. An $\mathcal{F}/\mathcal{B}$-measurable map $X : \Omega \to S$ (recall Definition 1.45) is called an $(S, \mathcal{B})$-*valued random variable*, or a *random variable taking values in* $(S, \mathcal{B})$, or simply an S-*valued random variable* when $\mathcal{B}$ is clear from the context.

(2) When $d \in \mathbb{N}$ and $S \subset \mathbb{R}^d$, we *always* equip S with its Borel σ-algebra $\mathcal{B}(S)$ unless otherwise stated, and an $(S, \mathcal{B}(S))$-valued random variable is simply called an S-valued random variable.
(3) An $\mathbb{R}$-valued random variable is called a *real random variable*. For $d \in \mathbb{N}$, an $\mathbb{R}^d$-valued random variable is called a *d-dimensional random variable*.

Note that a real random variable is nothing but an $\mathbb{R}$-valued $\mathcal{F}$-measurable function on Ω.

Proposition 3.3. *Let $d \in \mathbb{N}$ and let $X = (X_1, \dots, X_d) : \Omega \to \mathbb{R}^d$, where $X_i : \Omega \to \mathbb{R}$ for each $i \in \{1, \dots, d\}$. Then X is a d-dimensional random variable if and only if X_i is a real random variable for any $i \in \{1, \dots, d\}$.*

Proof. By Problem 1.15-(2) (or by Problem 2.10 together with Proposition 2.23-(2)), X is $\mathcal{F}/\mathcal{B}(\mathbb{R}^d)$-measurable if and only if X_i is $\mathcal{F}$-measurable for any $i \in \{1, \dots, d\}$, which is the asserted equivalence. □

Proposition 3.4. *Let $(S, \mathcal{B})$ be a measurable space and let X be an $(S, \mathcal{B})$-valued random variable. If $(E, \mathcal{E})$ is a measurable space and $f : S \to E$ is $\mathcal{B}/\mathcal{E}$-measurable, then $f(X)$ $(:= f \circ X)$ is an $(E, \mathcal{E})$-valued random variable. In particular, if $f : S \to \mathbb{R}$ is $\mathcal{B}$-measurable, then $f(X)$ is a real random variable.*

Proof. For any $A \in \mathcal{E}$, $f^{-1}(A) \in \mathcal{B}$ and hence $(f(X))^{-1}(A) = X^{-1}(f^{-1}(A)) \in \mathcal{F}$. Thus $f(X)$ is $\mathcal{F}/\mathcal{E}$-measurable, that is, it is an $(E, \mathcal{E})$-valued random variable.
The latter assertion is nothing but the case where $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. □

In particular, for $d, k \in \mathbb{N}$, *if X is a d-dimensional random variable and $f : \mathbb{R}^d \to \mathbb{R}^k$ is continuous, then $f(X)$ is a k-dimensional random variable*, since such f is $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^k)$-measurable by Problem 1.15-(1) (see also Lemma 1.17).

Definition 3.5 (Expectation (mean))**.** Let X be a real random variable (or more generally, a $[-\infty, \infty]$-valued $\mathcal{F}$-measurable function on Ω). We say that X *admits the expectation* (or it *admits the mean*) if and only if X admits the $\mathbb{P}$-integral, and in this case its *expectation* (or *mean*) $\mathbb{E}[X]$ is defined by

$$\mathbb{E}[X] := \int_\Omega X(\omega)\mathbb{P}(d\omega). \tag{3.1}$$

X is called *integrable* if and only if it is $\mathbb{P}$-integrable, or equivalently, if and only if X admits the mean and $\mathbb{E}[X] \in \mathbb{R}$.

Recall the definition of $\mathcal{L}^p(\mathbb{P}) = \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ for $p \in (0, \infty)$ (Definition 1.49):

$$\mathcal{L}^p(\mathbb{P}) := \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P}) := \{X \mid X \text{ is a real random variable}, \mathbb{E}[|X|^p] < \infty\}. \tag{3.2}$$

Proposition 3.6. *Let X be a real random variable.*
(1) *If X is* almost surely bounded*, that is, $|X| \le M$ a.s. for some $M \in [0, \infty)$, then $X \in \mathcal{L}^p(\mathbb{P})$ for any $p \in (0, \infty)$.*
(2) *Let $p, q \in (0, \infty)$, $p < q$. Then $\|X\|_{L^p} \le \|X\|_{L^q}$. In particular, $\mathcal{L}^q(\mathbb{P}) \subset \mathcal{L}^p(\mathbb{P})$.*

Proof. (1) Since $|X|^p = \min\{|X|^p, M^p\}$ a.s., by Proposition 1.35, Lemma 1.23 and $\mathbb{P}[\Omega] = 1$ we have $\mathbb{E}[|X|^p] = \mathbb{E}[\min\{|X|^p, M^p\}] \leq \mathbb{E}[M^p] = M^p < \infty$.
(2) This is immediate from Problem 1.30 (or more directly, this is an easy consequence of Hölder's inequality, Theorem 1.48). □

By Proposition 3.6-(2), if $X \in \mathcal{L}^2(\mathbb{P})$ then $X \in \mathcal{L}^1(\mathbb{P})$ and hence $\mathbb{E}[X]$ is defined and finite. Note also that, by Hölder's inequality, if $X, Y \in \mathcal{L}^2(\mathbb{P})$ then $XY \in \mathcal{L}^1(\mathbb{P})$.

Definition 3.7 (Variance and covariance)**.** (1) Let X be a real random variable. We define the *variance of* X, denoted as $\mathrm{var}(X)$ or $\sigma^2(X)$, by

$$\mathrm{var}(X) := \begin{cases} \mathbb{E}\big[\big(X - \mathbb{E}[X]\big)^2\big] = \mathbb{E}[X^2] - \big(\mathbb{E}[X]\big)^2 & \text{if } \mathbb{E}[X^2] < \infty, \\ \infty & \text{if } \mathbb{E}[X^2] = \infty. \end{cases} \tag{3.3}$$

Then $\sigma(X) := \sqrt{\mathrm{var}(X)}$ is called the *standard deviation of* X.
(2) For $X, Y \in \mathcal{L}^2(\mathbb{P})$, we define their *covariance* $\mathrm{cov}(X, Y)$ by

$$\mathrm{cov}(X, Y) := \mathbb{E}\big[\big(X - \mathbb{E}[X]\big)\big(Y - \mathbb{E}[Y]\big)\big] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \tag{3.4}$$

Mean and (co-)variance are the most fundamental quantities in probability theory. In fact, they naturally appear in the statements of limit theorems for random variables presented in the rest of this course.

The following definition is quite fundamental in the development of probability theory. Recall Theorem 1.46 for the notion of image measures.

Definition 3.8 (Law of a random variable)**.** Let $(S, \mathcal{B})$ be a measurable space and let X be an $(S, \mathcal{B})$-valued random variable. The *law* (or *distribution*) $\mathbb{P}_X$ *of* X is defined as the image measure $\mathbb{P} \circ X^{-1}$ of $\mathbb{P}$ by X, that is, $\mathbb{P}_X$ is a measure on $(S, \mathcal{B})$ given by

$$\mathbb{P}_X(A) := \mathbb{P} \circ X^{-1}(A) := \mathbb{P}[X^{-1}(A)] = \mathbb{P}[X \in A], \qquad A \in \mathcal{B}. \tag{3.5}$$

$\mathbb{P}_X$ is in fact a probability measure on $(S, \mathcal{B})$ since $\mathbb{P}_X(S) = \mathbb{P}[X^{-1}(S)] = \mathbb{P}[\Omega] = 1$. $\mathbb{P}_X$ is also referred to as the *probability law of* X or the *probability distribution of* X.

Notation. Let $(S, \mathcal{B})$ be a measurable space and X an $(S, \mathcal{B})$-valued random variable.
(1) As already used in (3.5), for $A \in \mathcal{B}$, the event $X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}$ is abbreviated as $\{X \in A\}$ and its probability is simply written as $\mathbb{P}[X \in A]$.
(2) The law $\mathbb{P}_X$ of a random variable X is also denoted as $\mathcal{L}(X)$. (This notation is used especially when no explicit reference is made to the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where the random variable X is defined.)
(3) For a probability measure μ on $(S, \mathcal{B})$, we write $X \sim \mu$ if and only if $\mathcal{L}(X) = \mu$.

The following proposition asserts that any probability measure on any measurable space is the law of a random variable on some probability space.

Proposition 3.9. *Let $(S, \mathcal{B})$ be a measurable space and let μ be a probability measure on $(S, \mathcal{B})$. Then the map $X : S \to S$ defined by $X(x) := x$ is an $(S, \mathcal{B})$-valued random variable on the probability space $(S, \mathcal{B}, \mu)$ whose law is μ.*

Proof. This is immediate from the fact that $X^{-1}(A) = A$ for any $A \subset S$. $\square$

The following theorem is just a rephrase of the latter half of Theorem 1.46.

Theorem 3.10. *Let $(S, \mathcal{B})$ be a measurable space, let X be an $(S, \mathcal{B})$-valued random variable and let $f : S \to \mathbb{R}$ be $\mathcal{B}$-measurable. Then $f(X)$ admits the mean if and only if f admits the $\mathbb{P}_X$-integral, and in this case*

$$\mathbb{E}[f(X)] = \int_S f(x)\mathbb{P}_X(dx). \tag{3.6}$$

Corollary 3.11. *Let X be a real random variable.*
(1) *X admits the mean if and only if the function $\mathbb{R} \ni x \mapsto x$ admits the $\mathbb{P}_X$-integral, and in this case*

$$\mathbb{E}[X] = \int_{\mathbb{R}} x\mathbb{P}_X(dx). \tag{3.7}$$

(2) $\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2\mathbb{P}_X(dx)$. *Moreover, if $\int_{\mathbb{R}} x^2\mathbb{P}_X(dx) < \infty$ then*

$$\mathrm{var}(X) = \int_{\mathbb{R}}\left(x-\int_{\mathbb{R}} y\mathbb{P}_X(dy)\right)^2\mathbb{P}_X(dx) = \int_{\mathbb{R}} x^2\mathbb{P}_X(dx)-\left(\int_{\mathbb{R}} x\mathbb{P}_X(dx)\right)^2. \tag{3.8}$$

Proof. (1) This is nothing but Theorem 3.10 with $(S, \mathcal{B}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $f(x) = x$.
(2) $\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2\mathbb{P}_X(dx)$ follows by applying Theorem 3.10 with $f(x) = x^2$. If these integrals are finite, then $X \in \mathcal{L}^1(\mathbb{P})$, so that $\mathbb{E}[X] = \int_{\mathbb{R}} x\mathbb{P}_X(dx) \in \mathbb{R}$ by (1), and Theorem 3.10 with $f(x) = \big(x - \mathbb{E}[X]\big)^2$ yields (3.8). $\square$

Definition 3.12. Let $d \in \mathbb{N}$, and let μ be a Borel probability measure on $\mathbb{R}^d$. A Borel measurable function $\rho : \mathbb{R}^d \to [0, \infty]$ is called a *density of μ* if and only if $\mu = \rho \cdot \mathrm{m}_d$ (recall Theorem 1.43), that is,

$$\mu(A) = \int_A \rho(x)dx \qquad \text{for any } A \in \mathcal{B}(\mathbb{R}^d). \tag{3.9}$$

The relation $\mu = \rho \cdot \mathrm{m}_d$ is also written as $\mu(dx) = \rho(x)dx$. If the law $\mathbb{P}_X$ of a d-dimensional random variable X has a density ρ, it is referred to as a *density of X*.

A density ρ of a Borel probability measure on $\mathbb{R}^d$ clearly satisfies $\int_{\mathbb{R}^d} \rho(x)dx = 1$. Conversely by Theorem 1.43, *any Borel measurable function $\rho : \mathbb{R}^d \to [0, \infty]$ with $\int_{\mathbb{R}^d} \rho(x)dx = 1$ defines a Borel probability measure $\rho \cdot \mathrm{m}_d$ on $\mathbb{R}^d$ (with a density ρ).*

Proposition 3.13. *Let $d \in \mathbb{N}$ and let μ be a Borel probability measure on $\mathbb{R}^d$ with a density ρ. If $h : \mathbb{R}^d \to [0, \infty]$ is Borel measurable, then h is a density of μ if and only if $h = \rho$ m_d-a.e.*

Proof. If $h = \rho$ m_d-a.e., then $\int_A h(x)dx = \int_A \rho(x)dx = \mu(A)$ for any $A \in \mathcal{B}(\mathbb{R}^d)$ by Proposition 1.35 and hence h is a density of μ. Conversely suppose h is a density of μ. Then $h = h\mathbf{1}_{\{h<\infty\}}$ m_d-a.e. and $\rho = \rho\mathbf{1}_{\{\rho<\infty\}}$ m_d-a.e. by Proposition 1.30-(3), $h\mathbf{1}_{\{h<\infty\}}, \rho\mathbf{1}_{\{\rho<\infty\}} \in \mathcal{L}^1(\mathrm{m}_d)$, and Proposition 1.35 implies that for any $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\int_A \big(h(x)\mathbf{1}_{\{h<\infty\}} - \rho(x)\mathbf{1}_{\{\rho<\infty\}}\big)dx = \int_A h(x)dx - \int_A \rho(x)dx = \mu(A) - \mu(A) = 0.$$

By Proposition 1.34-(2), we conclude that $h = h\mathbf{1}_{\{h<\infty\}} = \rho\mathbf{1}_{\{\rho<\infty\}} = \rho$ m_d-a.e. $\square$

Not every Borel probability measure on $\mathbb{R}^d$ has a density (see Problem 3.1), but many important probability measures are defined by determining densities ρ on $\mathbb{R}^d$, as we will see in the next section.

For a random variable with a density, Theorem 3.10 and Corollary 3.11 take the following form by virtue of Theorem 1.43.

Theorem 3.14. *Let $d \in \mathbb{N}$, let X be a d-dimensional random variable with a density ρ, and let $f : \mathbb{R}^d \to \mathbb{R}$ be Borel measurable. Then $f(X)$ admits the mean if and only if $f\rho$ admits the m_d-integral, and in this case*

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f(x)\rho(x)dx. \tag{3.10}$$

Corollary 3.15. *Let X be a real random variable with a density ρ.*
(1) *X admits the mean if and only if the function $\mathbb{R} \ni x \mapsto x\rho(x)$ admits the m_1-integral, and in this case*

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\rho(x)dx. \tag{3.11}$$

(2) $\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2\rho(x)dx$. *Moreover, if $\int_{-\infty}^{\infty} x^2\rho(x)dx < \infty$ then*

$$\begin{aligned} \mathrm{var}(X) &= \int_{-\infty}^{\infty}\left(x - \int_{-\infty}^{\infty} y\rho(y)dy\right)^2 \rho(x)dx \\ &= \int_{-\infty}^{\infty} x^2\rho(x)dx - \left(\int_{-\infty}^{\infty} x\rho(x)dx\right)^2. \end{aligned} \tag{3.12}$$

Theorem 3.16. *Let $d \in \mathbb{N}$, let U be an open subset of $\mathbb{R}^d$ and let X be a d-dimensional random variable with a density ρ_X and such that $X \in U$ a.s. Let $\varphi : U \to \mathbb{R}^d$ be a C^1-embedding and let $\psi := \varphi^{-1} : \varphi(U) \to U$. Then $Y := \begin{cases} \varphi(X) & \text{on } \{X \in U\} \\ 0 & \text{on } \{X \notin U\} \end{cases}$ is a d-dimensional random variable with a density ρ_Y given by*

$$\rho_Y = (\rho_X \circ \psi)|\det D\psi|\mathbf{1}_{\varphi(U)}. \tag{3.13}$$

Since $X \in U$ a.s. and hence $Y = \varphi(X)$ a.s., *in what follows the random variable Y in Theorem* 3.16 *will be simply denoted as $\varphi(X)$.*

Proof. Since φ is continuous, it is $\mathcal{B}(U)/\mathcal{B}(\mathbb{R}^d)$-measurable. Therefore if we define $\widetilde{\varphi} : \mathbb{R}^d \to \mathbb{R}^d$ by $\widetilde{\varphi}(x) := \varphi(x)$ for $x \in U$ and $\widetilde{\varphi}(x) := 0$ for $x \in \mathbb{R}^d \setminus U$, then $\widetilde{\varphi}$ is $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^d)$-measurable and hence $\widetilde{\varphi}(X) = Y$ is $\mathcal{F}/\mathcal{B}(\mathbb{R}^d)$-measurable. Now $\psi : \varphi(U) \to U$ is a surjective C^1-embedding, and therefore Theorem 2.42 together with $\mathbb{P}[X \in U] = 1$ implies that for any $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned} \mathbb{P}_Y(A) &= \mathbb{P}[\widetilde{\varphi}(X) \in A] = \mathbb{P}[X \in \widetilde{\varphi}^{-1}(A) \cap U] = \mathbb{P}_X\big(\widetilde{\varphi}^{-1}(A) \cap U\big) \\ &= \int_U \mathbf{1}_A\big(\widetilde{\varphi}(x)\big)\rho_X(x)dx = \int_{\varphi(U)} \mathbf{1}_A(y)\rho_X(\psi(y))|\det D\psi(y)|dy \\ &= \int_A \rho_X(\psi(y))|\det D\psi(y)|\mathbf{1}_{\varphi(U)}(y)dy. \end{aligned}$$

Thus $\rho_Y := (\rho_X \circ \psi)|\det D\psi|\mathbf{1}_{\varphi(U)}$ is a density of Y. □

Note that by Proposition 3.3, if $X = (X_1, \ldots, X_n)$ is an n-dimensional random variable and $Y = (Y_1, \ldots, Y_k)$ is a k-dimensional random variable, then $(X, Y) = (X_1, \ldots, X_n, Y_1, \ldots, Y_k)$ is an $(n+k)$-dimensional random variable. In this situation, the law of (X, Y) is often called the *joint law* (or *joint distribution*) *of* X, Y.

Proposition 3.17. *Let $n, k \in \mathbb{N}$, let X be an n-dimensional random variable and Y a k-dimensional random variable. If the $(n+k)$-dimensional random variable (X, Y) has a density ρ, then X and Y have densities ρ_X and ρ_Y, respectively, given by*

$$\rho_X(x) = \int_{\mathbb{R}^k} \rho(x, y) dy \quad \text{and} \quad \rho_Y(y) = \int_{\mathbb{R}^n} \rho(x, y) dx. \tag{3.14}$$

Proof. The first half of Fubini's theorem (Theorem 2.29-(1)) implies that for any $A \in \mathcal{B}(\mathbb{R}^n)$,

$$\mathbb{P}_X(A) = \mathbb{P}[(X, Y) \in A \times \mathbb{R}^k] = \int_{A \times \mathbb{R}^k} \rho(z) d\mathrm{m}_{n+k}(z) = \int_A \left(\int_{\mathbb{R}^k} \rho(x, y) dy \right) dx,$$

which means $\mathbb{P}_X(dx) = \left(\int_{\mathbb{R}^k} \rho(x, y) dy \right) dx$. Exactly the same argument also shows $\mathbb{P}_Y(dy) = \left(\int_{\mathbb{R}^n} \rho(x, y) dx \right) dy$. □

3.2 Basic Examples of Probability Distributions

In this section, we collect several important examples of Borel probability measures on $\mathbb{R}$ and illustrate usages of the tools presented in the last section by concrete calculations of means and variances of random variables.

Convention. (1) In accordance with the terminology in Definition 3.8, for $d \in \mathbb{N}$ and $S \subset \mathbb{R}^d$, a Borel probability measure on S is often referred to as a *law* (*probability law*) on S or a *distribution* (*probability distribution*) on S.
(2) A random variable X with a known probability distribution will be referred to with the name of that distribution. For example, an *exponential random variable* is a random variable whose law is an exponential distribution.

3.2.1 Probability distributions on integers

We start with examples of probability measures on (subsets of) $\mathbb{N} \cup \{0\}$. Note that, if $S \subset \mathbb{R}$ is a countable set then $\mathcal{B}(S) = 2^S$, since $\{x\}$ is a closed set in S and hence belongs to $\mathcal{B}(S)$ for each $x \in S$.

Example 3.18 (Binomial distribution)**.** Let $n \in \mathbb{N}$ and $p \in [0, 1]$. The *binomial distribution of size n and probability p* is the probability measure $B(n, p)$ on $\{0, \ldots, n\}$ given by

$$B(n, p)(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k \in \{0, \ldots, n\}, \tag{3.15}$$

where $\binom{n}{k} := \frac{n!}{k!(n-k)!}$ and $0^0 := 1$. ((3.15) is nothing but the probability of having heads exactly k times from n flips of a coin which shows heads with probability p; see Example 3.33 below.)

Recall the following equality (the binomial theorem): for any $x, y \in \mathbb{C}$,

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}. \tag{3.16}$$

(3.16) with $x = p$ and $y = 1 - p$ shows $\sum_{k=0}^{n} B(n, p)(\{k\}) = 1$, which means that $B(n, p)$ is actually a probability measure on $\{0, \dots, n\}$.

Example 3.19 (Poisson distribution)**.** Let $\lambda \in (0, \infty)$. The *Poisson distribution of parameter* λ is the probability measure $\mathrm{Po}(\lambda)$ on $\mathbb{N} \cup \{0\}$ given by

$$\mathrm{Po}(\lambda)(\{n\}) = e^{-\lambda} \frac{\lambda^n}{n!}, \qquad n \in \mathbb{N} \cup \{0\}. \tag{3.17}$$

Example 3.20 (Geometric distribution)**.** Let $\alpha \in [0, 1)$. The *geometric distribution of parameter* α is the probability measure $\mathrm{Geom}(\alpha)$ on $\mathbb{N} \cup \{0\}$ given by (with $0^0 := 1$)

$$\mathrm{Geom}(\alpha)(\{n\}) = (1-\alpha)\alpha^n, \qquad n \in \mathbb{N} \cup \{0\}. \tag{3.18}$$

It is clear that $\mathrm{Po}(\lambda)$ and $\mathrm{Geom}(\alpha)$ are probability measures on $\mathbb{N} \cup \{0\}$. Calculation of mean and variance for random variables with these distributions is left to the readers as an exercise (Problem 3.2). Note that an $\mathbb{N} \cup \{0\}$-valued random variable X can be naturally regarded as a real random variable, and then the law $\mathcal{L}(X)$ of X is regarded as a law on $\mathbb{R}$. In particular, $B(n, p)$, $\mathrm{Po}(\lambda)$ and $\mathrm{Geom}(\alpha)$ are regarded as laws on $\mathbb{R}$.

3.2.2 Probability distributions on $\mathbb{R}$

Next we give examples of probability distributions on $\mathbb{R}$.

Example 3.21 (Uniform distribution)**.** Let $a, b \in \mathbb{R}$, $a < b$. The *uniform distribution on* $[a, b]$ is the probability distribution $\mathrm{Unif}(a, b)$ on $\mathbb{R}$ given by

$$\mathrm{Unif}(a, b)(dx) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x) dx. \tag{3.19}$$

Example 3.22 (Exponential distribution)**.** Let $\alpha \in (0, \infty)$. The *exponential distribution of parameter* α is the probability distribution $\mathrm{Exp}(\alpha)$ on $\mathbb{R}$ given by

$$\mathrm{Exp}(\alpha)(dx) = \alpha e^{-\alpha x} \mathbf{1}_{(0,\infty)}(x) dx. \tag{3.20}$$

The exponential distributions are characterized by their *"memoryless property"*; see Problem 3.4 and Exercise 3.5.

Example 3.23 (Gamma distribution)**.** Let $\alpha, \beta \in (0, \infty)$. The *gamma distribution with parameters* α, β is the probability distribution $\mathrm{Gamma}(\alpha, \beta)$ on $\mathbb{R}$ given by

$$\mathrm{Gamma}(\alpha, \beta)(dx) = \frac{\beta^\alpha}{\Gamma(\alpha)} |x|^{\alpha-1} e^{-\beta x} \mathbf{1}_{(0,\infty)}(x) dx, \tag{3.21}$$

where Γ denotes the *gamma function*[1] $\Gamma : (0, \infty) \to (0, \infty)$ defined by

$$\Gamma(s) := \int_0^\infty x^{s-1} e^{-x} dx. \tag{3.22}$$

It is again clear that $\mathrm{Unif}(a, b)$, $\mathrm{Exp}(\alpha)$ and $\mathrm{Gamma}(\alpha, \beta)$ are probability distributions on $\mathbb{R}$. Calculation of mean and variance for random variables with these distributions is left to the readers as an exercise (Problem 3.3).

Example 3.24 (Normal distribution)**.** Let $m \in \mathbb{R}$ and $v \in [0, \infty)$. The *normal* (or *Gaussian*) *distribution with mean* m *and variance* v is the probability distribution $N(m, v)$ on $\mathbb{R}$ given by $N(m, 0) = \delta_m$ (the unit mass at m) if $v = 0$ and

$$N(m, v)(dx) = \frac{1}{\sqrt{2\pi v}} \exp\Big(-\frac{(x-m)^2}{2v}\Big) dx \tag{3.23}$$

if $v > 0$. In particular, $N(0, 1)$ is called the *standard normal distribution*.

The following calculations show that (3.23) actually defines a probability distribution on $\mathbb{R}$: Corollary 2.40 with $y = \sqrt{v}x + m$ yields

$$\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi v}} \exp\Big(-\frac{(y-m)^2}{2v}\Big) dy = \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big) dx,$$

and by the first half of Fubini's theorem (Theorem 2.29-(1)) and Theorem 2.42 with the polar coordinates $(0, \infty) \times (0, 2\pi) \ni (r, \theta) \mapsto (r\cos\theta, r\sin\theta) \in \mathbb{R}^2 \setminus \big([0, \infty) \times \{0\}\big)$,

$$\Big(\int_{-\infty}^\infty e^{-x^2/2} dx\Big)^2 = \int_{\mathbb{R}^2} e^{-|z|^2/2} dz = \int_0^\infty \Big(\int_0^{2\pi} e^{-r^2/2} r d\theta\Big) dr = 2\pi.$$

As suggested in the name of $N(m, v)$, a real random variable X with $X \sim N(m, v)$ has mean m and variance v. Indeed, if $v > 0$, Theorem 3.14 and Corollary 2.40 yield

$$\begin{aligned} \mathbb{E}[(X-m)^2] &= \int_{-\infty}^\infty \frac{(x-m)^2}{\sqrt{2\pi v}} \exp\Big(-\frac{(x-m)^2}{2v}\Big) dx = v \int_{-\infty}^\infty y^2 \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy \\ &= v \lim_{n\to\infty} \int_{-n}^n y^2 \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy \\ &= v \lim_{n\to\infty} \Big(\Big[-y \frac{e^{-y^2/2}}{\sqrt{2\pi}}\Big]_{-n}^n + \int_{-n}^n \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy\Big) \\ &= v \int_{-\infty}^\infty \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy = v < \infty, \end{aligned} \tag{3.24}$$

where we used integration by parts in the third line and the monotone convergence theorem (Theorem 1.24) in the second and fourth lines. In particular, $\mathbb{E}[X^2] < \infty$, hence $\mathbb{E}[|X|] < \infty$, and then by Corollary 3.15-(1) and Corollary 2.40 we have

$$\mathbb{E}[X] = \int_{-\infty}^\infty x \frac{1}{\sqrt{2\pi v}} \exp\Big(-\frac{(x-m)^2}{2v}\Big) dx = \int_{-\infty}^\infty (m + \sqrt{v}y) \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy = m,$$

[1] It easily follows by integration by parts that $\Gamma(x+1) = x\Gamma(x)$ for any $x \in (0, \infty)$, which and $\Gamma(1) = 1$ imply that $\Gamma(n) = (n-1)!$ for any $n \in \mathbb{N}$.

which and (3.24) show $\mathrm{var}(X) = v$.

The following example presents a probability distribution on $\mathbb{R}$ with which a random variable does **not** admit the mean.

Example 3.25 (Cauchy distribution)**.** Let $m \in \mathbb{R}$ and $\alpha \in (0, \infty)$. The *Cauchy distribution with parameters* m, α is the probability distribution $\mathrm{Cauchy}(m, \alpha)$ on $\mathbb{R}$ given by

$$\mathrm{Cauchy}(m, \alpha)(dx) = \frac{1}{\pi} \frac{\alpha}{\alpha^2 + (x - m)^2} dx. \tag{3.25}$$

This indeed defines a probability distribution on $\mathbb{R}$ since

$$\int_{-\infty}^{\infty} \frac{\alpha}{\alpha^2 + (x - m)^2} dx = \lim_{n \to \infty} \int_{m-n}^{m+n} \frac{\alpha}{\alpha^2 + (x - m)^2} dx = \lim_{n \to \infty} 2 \arctan \frac{n}{\alpha} = \pi$$

by the monotone convergence theorem (Theorem 1.24) and $(\arctan x)' = (1 + x^2)^{-1}$. It is easy to see that a Cauchy random variable does not admit the mean (Problem 3.8).

3.3 Independence of Random Variables

Throughout this section, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a probability space, and random variables are always assumed to be defined on $(\Omega, \mathcal{F}, \mathbb{P})$ unless otherwise stated.

Definition 3.26 (Independence)**.** Let $n \in \mathbb{N}$, and for each $i \in \{1, \ldots, n\}$ let $(S_i, \mathcal{B}_i)$ be a measurable space and X_i an $(S_i, \mathcal{B}_i)$-valued random variable. We call $\{X_i\}_{i=1}^n$ *independent* if and only if for any $A_i \in \mathcal{B}_i$, $i \in \{1, \ldots, n\}$,

$$\mathbb{P}[X_1 \in A_1, \ldots, X_n \in A_n] = \mathbb{P}[X_1 \in A_1] \cdots \mathbb{P}[X_n \in A_n]. \tag{3.26}$$

According to Problem 2.10, in the situation of Definition 3.26, $(X_1, \ldots, X_n)$ is an $(S_1 \times \cdots \times S_n, \mathcal{B}_1 \otimes \cdots \otimes \mathcal{B}_n)$-valued random variable and hence its law $\mathbb{P}_{(X_1,\ldots,X_n)}$ is defined as a probability measure on $(S_1 \times \cdots \times S_n, \mathcal{B}_1 \otimes \cdots \otimes \mathcal{B}_n)$.

Theorem 3.27. *Let $n \in \mathbb{N}$, and for each $i \in \{1, \ldots, n\}$ let $(S_i, \mathcal{B}_i)$ be a measurable space and X_i an $(S_i, \mathcal{B}_i)$-valued random variable. Then $\{X_i\}_{i=1}^n$ is independent if and only if*

$$\mathbb{P}_{(X_1,\ldots,X_n)} = \mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_n}. \tag{3.27}$$

Proof. Note that for $A_i \in \mathcal{B}_i$, $i \in \{1, \ldots, n\}$, we have

$$\mathbb{P}[X_1 \in A_1, \ldots, X_n \in A_n] = \mathbb{P}_{(X_1,\ldots,X_n)}(A_1 \times \cdots \times A_n) \tag{3.28}$$

and

$$\begin{aligned} \mathbb{P}[X_1 \in A_1] \cdots \mathbb{P}[X_n \in A_n] &= \mathbb{P}_{X_1}(A_1) \cdots \mathbb{P}_{X_n}(A_n) \\ &= \mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_n}(A_1 \times \cdots \times A_n). \end{aligned} \tag{3.29}$$

Thus if (3.27) holds then (3.28) and (3.29) are equal and hence $\{X_i\}_{i=1}^n$ is independent. Conversely if $\{X_i\}_{i=1}^n$ is independent, then by (3.28) and (3.29), $\mathbb{P}_{(X_1,\ldots,X_n)}$ coincides with $\mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_n}$ on $\mathcal{B}_1 \times \cdots \times \mathcal{B}_n$, and hence on $\sigma(\mathcal{B}_1 \times \cdots \times \mathcal{B}_n) = \mathcal{B}_1 \otimes \cdots \otimes \mathcal{B}_n$ as well by Theorem 2.5. □

Theorem 3.28. *Let $n \in \mathbb{N}$, and for each $i \in \{1, \dots, n\}$ let $(S_i, \mathcal{B}_i)$ be a measurable space and μ_i a probability measure on $(S_i, \mathcal{B}_i)$. Then there exist a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ and an $(S_i, \mathcal{B}_i)$-valued random variable X_i on $(\Omega', \mathcal{F}', \mathbb{P}')$ with $X_i \sim \mu_i$ for $i \in \{1, \dots, n\}$, such that $\{X_i\}_{i=1}^n$ is independent.*

Proof. Let $\Omega' := S_1 \times \cdots \times S_n$, $\mathcal{F}' := \mathcal{B}_1 \otimes \cdots \otimes \mathcal{B}_n$ and $\mathbb{P}' := \mu_1 \times \cdots \times \mu_n$, and for each $i \in \{1, \dots, n\}$ define $X_i : \Omega' \to S_i$ by $X_i(x_1, \dots, x_n) := x_i$. Then for each $i \in \{1, \dots, n\}$, X_i is clearly $\mathcal{F}'/\mathcal{B}_i$-measurable, and for any $A_i \in \mathcal{B}_i$,

$$\mathbb{P}'_{X_i}(A_i) = \mathbb{P}'[X_i \in A_i] = \mu_1 \times \cdots \times \mu_n(S_1 \times \cdots \times A_i \times \cdots \times S_n) = \mu_i(A_i),$$

i.e. $X_i \sim \mu_i$. Moreover, $X := (X_1, \dots, X_n)$ is the identity map on $\Omega' = S_1 \times \cdots \times S_n$ (i.e. $X(x_1, \dots, x_n) = (x_1, \dots, x_n)$ for any $(x_1, \dots, x_n) \in \Omega'$) and therefore $\mathbb{P}'_X = \mathbb{P}' \circ X^{-1} = \mathbb{P}' = \mu_1 \times \cdots \times \mu_n = \mathbb{P}'_{X_1} \times \cdots \times \mathbb{P}'_{X_n}$. Thus $\{X_i\}_{i=1}^n$ is independent. $\square$

Theorem 3.29. *Let $n \in \mathbb{N}$. For each $i \in \{1, \dots, n\}$, let $d_i \in \mathbb{N}$, let X_i be a d_i-dimensional random variable and let $\rho_i : \mathbb{R}^{d_i} \to [0, \infty]$ be Borel measurable and satisfy $\int_{\mathbb{R}^{d_i}} \rho_i(x)dx = 1$. Then the following conditions are equivalent to each other:*
(1) *$\{X_i\}_{i=1}^n$ is independent and X_i has a density ρ_i for any $i \in \{1, \dots, n\}$.*
(2) *$(X_1, \dots, X_n)$ has a density ρ given by $\rho(x_1, \dots, x_n) = \rho_1(x_1) \cdots \rho_n(x_n)$.*

Proof. (1) $\Rightarrow$ (2): For any $A_i \in \mathcal{B}(\mathbb{R}^{d_i})$, $i \in \{1, \dots, n\}$, the independence yields

$$\begin{aligned}
&\mathbb{P}_{(X_1,\dots,X_n)}(A_1 \times \cdots \times A_n) = \mathbb{P}[X_1 \in A_1, \dots, X_n \in A_n] \\
&= \mathbb{P}[X_1 \in A_1] \cdots \mathbb{P}[X_n \in A_n] = \int_{A_1} \rho_1(x_1)dx_1 \cdots \int_{A_n} \rho_n(x_n)dx_n \\
&\text{(by Problem 2.11-(2))} = \int_{A_1 \times \cdots \times A_n} \rho_1(x_1) \cdots \rho_n(x_n) d(x_1, \dots, x_n),
\end{aligned}$$

which and Theorem 2.5 imply $\mathbb{P}_{(X_1,\dots,X_n)}(dx) = \rho_1(x_1) \cdots \rho_n(x_n) d(x_1, \dots, x_n)$.
(2) $\Rightarrow$ (1): For any $A_i \in \mathcal{B}(\mathbb{R}^{d_i})$, $i \in \{1, \dots, n\}$, by Problem 2.11-(2) we have

$$\begin{aligned}
&\mathbb{P}[X_1 \in A_1, \dots, X_n \in A_n] = \mathbb{P}_{(X_1,\dots,X_n)}(A_1 \times \cdots \times A_n) \qquad (3.30) \\
&= \int_{A_1 \times \cdots \times A_n} \rho_1(x_1) \cdots \rho_n(x_n) d(x_1, \dots, x_n) = \int_{A_1} \rho_1(x_1)dx_1 \cdots \int_{A_n} \rho_n(x_n)dx_n.
\end{aligned}$$

For each $i \in \{1, \dots, n\}$, setting $A_k := \mathbb{R}^{d_k}$ for $k \in \{1, \dots, n\} \setminus \{i\}$ in (3.30) shows $\mathbb{P}_{X_i}(dx) = \rho_i(x)dx$, and then (3.30) implies the independence of $\{X_i\}_{i=1}^n$. $\square$

Example 3.30. Let X, Y be independent real random variables with $X \sim N(0, 1)$ and $Y \sim N(0, 1)$. We calculate densities of $X + Y$ and $X - Y$ by using Theorems 3.16 and 3.29. By Theorem 3.29, (X, Y) has a density ρ given by

$$\rho(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}.$$

Define $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ by $\varphi(x, y) := (x + y, x - y)$. Then $\varphi^{-1}(x, y) = \left(\frac{x+y}{2}, \frac{x-y}{2}\right)$, $D(\varphi^{-1})(x, y) = \frac{1}{2}\left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$, $|\det D(\varphi^{-1})(x, y)| = 1/2$ and therefore by Theorem 3.16,

$(X + Y, X - Y)$ has a density given by

$$\rho\left(\frac{x+y}{2}, \frac{x-y}{2}\right) \cdot \frac{1}{2} = \frac{1}{4\pi} e^{-(x^2+y^2)/4} = \frac{1}{\sqrt{4\pi}} e^{-x^2/4} \cdot \frac{1}{\sqrt{4\pi}} e^{-y^2/4}.$$

Thus by Theorem 3.29, $\{X+Y, X-Y\}$ is independent, $X+Y \sim N(0,2)$ and $X-Y \sim N(0,2)$.

Proposition 3.31. *Let $n \in \mathbb{N}$, and for each $i \in \{1, \ldots, n\}$ let $(S_i, \mathcal{B}_i)$ be a measurable space and X_i an $(S_i, \mathcal{B}_i)$-valued random variable. Suppose $\{X_i\}_{i=1}^n$ is independent.*
(1) *For any $1 \leq i_1 < \cdots < i_k \leq n$, $\{X_{i_\ell}\}_{\ell=1}^k$ is independent.*
(2) *For each $i \in \{1, \ldots, n\}$, let $(E_i, \mathcal{E}_i)$ be a measurable space and let $f_i : S_i \to E_i$ be $\mathcal{B}_i/\mathcal{E}_i$-measurable. Then $\{f_i(X_i)\}_{i=1}^n$ is independent.*
(3) *Let $k \in \mathbb{N}$, $k < n$ and set $Y := (X_1, \ldots, X_k)$ and $Z := (X_{k+1}, \ldots, X_n)$. Then $\{Y, Z\}$ is independent.*

Proof. (1) Let $A_{i_\ell} \in \mathcal{B}_{i_\ell}$, $\ell \in \{1, \ldots, k\}$ and set $A_i := S_i$, $i \in \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\}$. Then since $\mathbb{P}[X_i \in A_i] = 1$ for $i \in \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\}$,

$$\begin{aligned}&\mathbb{P}[X_{i_1} \in A_{i_1}, \ldots, X_{i_k} \in A_{i_k}] = \mathbb{P}[X_1 \in A_1, \ldots, X_n \in A_n] \\ &= \mathbb{P}[X_1 \in A_1] \cdots \mathbb{P}[X_n \in A_n] = \mathbb{P}[X_{i_1} \in A_{i_1}] \cdots \mathbb{P}[X_{i_k} \in A_{i_k}].\end{aligned}$$

(2) Let $A_i \in \mathcal{E}_i$, $i \in \{1, \ldots, n\}$. Then since $f_i^{-1}(A_i) \in \mathcal{B}_i$,

$$\begin{aligned}&\mathbb{P}[f_1(X_1) \in A_1, \ldots, f_n(X_n) \in A_n] = \mathbb{P}[X_1 \in f_1^{-1}(A_1), \ldots, X_n \in f_n^{-1}(A_n)] \\ &= \mathbb{P}[X_1 \in f_1^{-1}(A_1)] \cdots \mathbb{P}[X_n \in f_n^{-1}(A_n)] = \mathbb{P}[f_1(X_1) \in A_1] \cdots \mathbb{P}[f_n(X_n) \in A_n].\end{aligned}$$

(3) By Theorem 3.27, Corollary 2.26 and (1) above,

$$\mathbb{P}_{(Y,Z)} = \mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_n} = (\mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_k}) \times (\mathbb{P}_{X_{k+1}} \times \cdots \times \mathbb{P}_{X_n}) = \mathbb{P}_Y \times \mathbb{P}_Z,$$

and hence $\{Y, Z\}$ is independent by Theorem 3.27. □

Proposition 3.32. *Let $n \in \mathbb{N}$ and let $\{X_i\}_{i=1}^n \subset \mathcal{L}^1(\mathbb{P})$ be independent. Then*

$$X_1 \cdots X_n \in \mathcal{L}^1(\mathbb{P}), \qquad \mathbb{E}[X_1 \cdots X_n] = \mathbb{E}[X_1] \cdots \mathbb{E}[X_n], \tag{3.31}$$

$$\operatorname{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \operatorname{var}(X_i). \tag{3.32}$$

Proof. By Problem 2.11-(2), $\mathbb{R}^n \ni (x_1, \ldots, x_n) \mapsto x_1 \cdots x_n$ is integrable with respect to $\mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_n} = \mathbb{P}_{(X_1,\ldots,X_n)}$ and

$$\int_{\mathbb{R}^n} x_1 \cdots x_n \mathbb{P}_{(X_1,\ldots,X_n)}(dx_1 \ldots dx_n) = \int_{\mathbb{R}} x_1 \mathbb{P}_{X_1}(dx_1) \ldots \int_{\mathbb{R}} x_n \mathbb{P}_{X_n}(dx_n),$$

which and Theorem 3.10 yields (3.31).

For (3.32), let $Y_i := X_i - \mathbb{E}[X_i] \in \mathcal{L}^1(\mathbb{P})$, so that $\{Y_i\}_{i=1}^n$ is independent, $\mathbb{E}[Y_i] = 0$ and $\mathrm{var}(X_i) = \mathbb{E}[Y_i^2]$ for $i \in \{1, \dots, n\}$, and $\mathrm{var}\big(\sum_{i=1}^n X_i\big) = \mathbb{E}\big[\big(\sum_{i=1}^n Y_i\big)^2\big]$. Then

$$\left(\sum_{i=1}^n Y_i\right)^2 = \sum_{i=1}^n Y_i^2 + 2 \sum_{1 \le i < j \le n} Y_i Y_j, \tag{3.33}$$

and $\sum_{1 \le i < j \le n} Y_i Y_j \in \mathcal{L}^1(\mathbb{P})$ and $\mathbb{E}\big[\sum_{1 \le i < j \le n} Y_i Y_j\big] = 0$ by (3.31). It follows from these facts that $\mathrm{var}\big(\sum_{i=1}^n X_i\big) = \mathbb{E}\big[\big(\sum_{i=1}^n Y_i\big)^2\big] < \infty$ if and only if $\sum_{i=1}^n \mathrm{var}(X_i) = \sum_{i=1}^n \mathbb{E}[Y_i^2] < \infty$, and in this case taking the mean of (3.33) shows (3.32). □

Example 3.33. Let $p \in [0, 1]$. A *Bernoulli random variable of probability* p is a $\{0, 1\}$-valued random variable X with $\mathbb{P}[X = 1] = p$ and $\mathbb{P}[X = 0] = 1 - p$. For such X we have

$$\mathbb{E}[X] = \mathbb{E}[X^2] = 0 \cdot (1 - p) + 1 \cdot p = p,$$
$$\mathrm{var}(X) = \mathbb{E}[X^2] - \big(\mathbb{E}[X]\big)^2 = p(1 - p).$$

Now let $n \in \mathbb{N}$ and let $\{X_i\}_{i=1}^n$ be independent Bernoulli random variables of probability p, which exist by Theorem 3.28. Then $S := \sum_{i=1}^n X_i$ is a binomial random variable of size n and probability p; indeed, for $k \in \{0, 1, \dots, n\}$,

$$\mathbb{P}[S = k] = \sum_{\substack{(\alpha_1, \dots, \alpha_n) \in \{0,1\}^n \\ \sum_{i=1}^n \alpha_i = k}} \mathbb{P}[(X_1, \dots, X_n) = (\alpha_1, \dots, \alpha_n)] = \binom{n}{k} p^k (1 - p)^{n-k}.$$

These facts together with Proposition 3.32 allow us to calculate easily the mean and the variance of a binomial random variable of size n and probability p, as follows:

$$\mathbb{E}[S] = \sum_{i=1}^n \mathbb{E}[X_i] = np, \qquad \mathrm{var}(S) = \sum_{i=1}^n \mathrm{var}(X_i) = np(1 - p). \tag{3.34}$$

We need a lemma for the next definition.

Lemma 3.34. *Let $d \in \mathbb{N}$, $A \in \mathcal{B}(\mathbb{R}^d)$ and let ν be a law on $\mathbb{R}^d$.*
(1) *For $n \in \mathbb{N}$, $\mathbb{R}^{dn} \ni (x_1, \dots, x_n) \mapsto \mathbf{1}_A(x_1 + \cdots + x_n)$ is Borel measurable.*
(2) *Let $x \in \mathbb{R}^d$ and set $A - x := \{z - x \mid z \in A\}$. Then $\nu(A - x) = \int_{\mathbb{R}^d} \mathbf{1}_A(x + y)\nu(dy)$ and it is a Borel measurable function in $x \in \mathbb{R}^d$.*

Proof. (1) Since $\mathbb{R}^{dn} \ni (x_1, \dots, x_n) \mapsto x_1 + \cdots + x_n \in \mathbb{R}^d$ is continuous, it is $\mathcal{B}(\mathbb{R}^{dn})/\mathcal{B}(\mathbb{R}^d)$-measurable and hence $\mathbb{R}^{dn} \ni (x_1, \dots, x_n) \mapsto \mathbf{1}_A(x_1 + \cdots + x_n)$ is $\mathcal{B}(\mathbb{R}^{dn})$-measurable.
(2) $\mathbf{1}_A(x + y) = \mathbf{1}_{A-x}(y)$ yields $\int_{\mathbb{R}^d} \mathbf{1}_A(x + y)\nu(dy) = \nu(A - x)$, and then its Borel measurability in $x \in \mathbb{R}^d$ follows by Fubini's theorem (Theorem 2.29-(1)). □

Definition 3.35 (Convolution)**.** Let $d \in \mathbb{N}$. For probability laws μ, ν on $\mathbb{R}^d$, their *convolution* $\mu * \nu$ is defined as the law on $\mathbb{R}^d$ given by, for each $A \in \mathcal{B}(\mathbb{R}^d)$,

$$(\mu * \nu)(A) := \int_{\mathbb{R}^{2d}} \mathbf{1}_A(x+y)(\mu \times \nu)(dxdy) = \int_{\mathbb{R}^d} \nu(A-x)\mu(dx). \qquad (3.35)$$

The second equality in (3.35) follows by Lemma 3.34 and Fubini's theorem (Theorem 2.29-(1)). Clearly $(\mu * \nu)(\mathbb{R}^d) = 1$, and Proposition 1.26 easily shows that $\mu * \nu$ is a Borel measure on $\mathbb{R}^d$. Thus $\mu * \nu$ is indeed a law on $\mathbb{R}^d$.

Proposition 3.36. *Let $d \in \mathbb{N}$ and let X, Y be independent d-dimensional random variables. Then $\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y$.*

Proof. For $A \in \mathcal{B}(\mathbb{R}^d)$, Theorem 3.10 applied to $\mathbb{R}^{2d} \ni (x,y) \mapsto \mathbf{1}_A(x+y)$ yields

$$\begin{aligned}\mathbb{P}_{X+Y}(A) = \mathbb{P}[X+Y \in A] = \mathbb{E}[\mathbf{1}_A(X+Y)] &= \int_{\mathbb{R}^{2d}} \mathbf{1}_A(x+y)\mathbb{P}_{(X,Y)}(dxdy) \\ = \int_{\mathbb{R}^{2d}} \mathbf{1}_A(x+y)(\mathbb{P}_X \times \mathbb{P}_Y)(dxdy) &= (\mathbb{P}_X * \mathbb{P}_Y)(A),\end{aligned}$$

proving $\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y$. □

Proposition 3.37. *Let $d \in \mathbb{N}$ and let λ, μ, ν be laws on $\mathbb{R}^d$. Then*

$$\mu * \nu = \nu * \mu \quad \text{and} \quad (\mu * \nu) * \lambda = \mu * (\nu * \lambda). \qquad (3.36)$$

Proof. By Fubini's theorem (Theorem 2.29-(1)), the definition (3.35) of $\mu * \nu$ is independent of the order of μ, ν and hence $\mu * \nu = \nu * \mu$. Furthermore for $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned}\big(\mu * (\nu * \lambda)\big)(A) &= \int_{\mathbb{R}^d} (\nu * \lambda)(A-x)\mu(dx) \\ &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \lambda\big((A-x)-y\big)\nu(dy) \right) \mu(dx) \\ &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \mathbf{1}_A(x+y+z)\lambda(dz) \right) \nu(dy) \right) \mu(dx),\end{aligned}$$

which is independent of the order of μ, ν, λ by Fubini's theorem (Theorem 2.29-(1)). Therefore $(\mu * \nu) * \lambda = \mu * (\nu * \lambda)$. □

Proposition 3.38. *Let $d \in \mathbb{N}$ and let μ, ν be laws on $\mathbb{R}^d$. If ν has a density ρ, then $\mu * \nu$ has a density h given by $h(x) := \int_{\mathbb{R}^d} \rho(x-y)\mu(dy)$. If μ also has a density g, then $h(x) = \int_{\mathbb{R}^d} \rho(x-y)g(y)dy$.*

Proof. For $A \in \mathcal{B}(\mathbb{R}^d)$, from Theorem 1.43, Corollary 2.40 and Fubini's theorem (Theorem 2.29-(1)) we see that

$$\begin{aligned}&(\mu * \nu)(A) \\ &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \mathbf{1}_A(x+y)\nu(dx) \right) \mu(dy) = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \mathbf{1}_A(x+y)\rho(x)dx \right) \mu(dy)\end{aligned}$$

$$= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \mathbf{1}_A(x) \rho(x-y) dx \right) \mu(dy) = \int_A \left(\int_{\mathbb{R}^d} \rho(x-y) \mu(dy) \right) dx = \int_A h dx.$$

Note here that the Borel measurability of h is also a consequence of Fubini's theorem (Theorem 2.29-(1)). The latter assertion is immediate by Theorem 1.43. □

So far we have considered independence for finitely many random variables only. Next we define the independence of an *infinite* sequence of random variables.

Definition 3.39. For each $n \in \mathbb{N}$, let $(S_n, \mathcal{B}_n)$ be a measurable space and let X_n be an $(S_n, \mathcal{B}_n)$-valued random variable. We call $\{X_n\}_{n=1}^{\infty}$ *independent* if and only if $\{X_i\}_{i=1}^{n}$ is independent for any $n \in \mathbb{N}$.

Then in accordance with Theorem 3.28, we have the following existence theorem for independent sequences of random variables, whose proof will be provided later in Section 3.6.

Theorem 3.40. *For each $n \in \mathbb{N}$ let $(S_n, \mathcal{B}_n)$ be a measurable space and let μ_n be a probability measure on $(S_n, \mathcal{B}_n)$. Then there exist a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ and an $(S_n, \mathcal{B}_n)$-valued random variable X_n on $(\Omega', \mathcal{F}', \mathbb{P}')$ with $X_n \sim \mu_n$ for $n \in \mathbb{N}$, such that $\{X_n\}_{n=1}^{\infty}$ is independent.*

The following theorem is frequently used in probability theory. Recall that, as in Problem 1.13, for $\{A_n\}_{n=1}^{\infty} \subset 2^{\Omega}$ we set

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k, \qquad \liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k, \tag{1.64}$$

so that $\limsup_{n\to\infty} A_n, \liminf_{n\to\infty} A_n \in \mathcal{F}$ if $\{A_n\}_{n=1}^{\infty} \subset \mathcal{F}$.

Theorem 3.41 (Borel-Cantelli lemma)**.** *Let $\{A_n\}_{n=1}^{\infty} \subset \mathcal{F}$.*
(1) *If $\sum_{n=1}^{\infty} \mathbb{P}[A_n] < \infty$, then $\mathbb{P}[\liminf_{n\to\infty} A_n^c] = 1$.*
(2) *If $\{A_n\}_{n=1}^{\infty}$ is independent and $\sum_{n=1}^{\infty} \mathbb{P}[A_n] = \infty$, then $\mathbb{P}[\limsup_{n\to\infty} A_n] = 1$.*

Remark 3.42. Recall that the notion of independence of *events* has been treated in Problem 1.3 and the definition before it. In fact, *$\{A_n\}_{n=1}^{\infty} \subset \mathcal{F}$ is independent if and only if $\{\mathbf{1}_{A_n}\}_{n=1}^{\infty}$ is independent.* This equivalence easily follows from Problem 1.3-(2) and the fact that for $A \in \mathcal{F}$ and $B \in \mathcal{B}(\mathbb{R})$,

$$\{\mathbf{1}_A \in B\} = \begin{cases} \emptyset & \text{if } 0 \notin B \text{ and } 1 \notin B, \\ A & \text{if } 0 \notin B \text{ and } 1 \in B, \\ A^c & \text{if } 0 \in B \text{ and } 1 \notin B, \\ \Omega & \text{if } 0 \in B \text{ and } 1 \in B. \end{cases}$$

We need the following easy lemma.

Lemma 3.43. *Let $\{p_n\}_{n=1}^{\infty} \subset [0,1)$. Then $\lim_{n\to\infty}(1-p_1)\cdots(1-p_n) = 0$ if and only if $\sum_{n=1}^{\infty} p_n = \infty$.*

Proof. Since $1-x \le e^{-x}$ for $x \in [0,1]$, for $n \in \mathbb{N}$ we have $0 \le (1-p_1)\cdots(1-p_n) \le \exp\bigl(-\sum_{i=1}^{n} p_i\bigr)$ and hence $\sum_{n=1}^{\infty} p_n = \infty$ implies $\lim_{n\to\infty}(1-p_1)\cdots(1-p_n) = 0$.

For the converse, suppose $\lim_{n\to\infty}(1-p_1)\cdots(1-p_n) = 0$. If $p_n > 1/2$ for infinitely many $n \in \mathbb{N}$ then clearly $\sum_{n=1}^{\infty} p_n = \infty$, and therefore we may assume that there exists $N \in \mathbb{N}$ such that $p_n \le 1/2$ for any $n \ge N$. Note that $1-x \ge e^{-2x}$ for $x \in [0,1/2]$; indeed, $g(x) := 1-x-e^{-2x}$ increases on $[0,(\log 2)/2]$ and decreases on $[(\log 2)/2, 1/2]$ since $g'(x) = 2e^{-2x}-1$, and thus $g(x) \ge \min\{g(0), g(1/2)\} = 0$ for $x \in [0,1/2]$. Now for $n \in \mathbb{N}, n \ge N$, $\exp\bigl(-2\sum_{k=N}^{n} p_k\bigr) \le (1-p_N)\cdots(1-p_n)$ and hence $2\sum_{k=N}^{n} p_k \ge -\log\bigl((1-p_N)\cdots(1-p_n)\bigr)$, which together with the assumption $\lim_{n\to\infty}(1-p_1)\cdots(1-p_n) = 0$ shows $\sum_{n=1}^{\infty} p_n = \infty$. □

Proof of Theorem 3.41. (1) By Problem 1.13, $\bigl(\limsup_{n\to\infty} A_n\bigr)^c = \liminf_{n\to\infty} A_n^c$ and $\mathbb{P}[\liminf_{n\to\infty} A_n^c] = 1 - \mathbb{P}[\limsup_{n\to\infty} A_n] = 1-0 = 1$.
(2) For $n \in \mathbb{N}$, by the independence of $\{A_n\}_{n=1}^{\infty}$, $\sum_{n=1}^{\infty} \mathbb{P}[A_n] = \infty$ and Lemma 3.43,

$$\mathbb{P}\left[\left(\bigcup_{k=n}^{\infty} A_k\right)^c\right] = \mathbb{P}\left[\bigcap_{k=n}^{\infty} A_k^c\right] = \lim_{\ell\to\infty} \mathbb{P}\left[\bigcap_{k=n}^{\ell} A_k^c\right] = \lim_{\ell\to\infty} \prod_{k=n}^{\ell}\bigl(1-\mathbb{P}[A_k]\bigr) = 0$$

and hence $\mathbb{P}\bigl[\bigcup_{k=n}^{\infty} A_k\bigr] = 1$. Letting $n \to \infty$, we obtain $\mathbb{P}[\limsup_{n\to\infty} A_n] = 1$. □

We conclude this subsection with another important consequence of independence. We need some definitions.

Definition 3.44. Let $\{X_\lambda\}_{\lambda\in\Lambda}$ be a family of random variables with X_λ taking values in a measurable space $(S_\lambda, \mathcal{B}_\lambda)$ for each $\lambda \in \Lambda$. We define

$$\sigma\bigl(\{X_\lambda\}_{\lambda\in\Lambda}\bigr) := \sigma_\Omega\bigl(\{\{X_\lambda \in A_\lambda\} \mid \lambda \in \Lambda,\ A_\lambda \in \mathcal{B}_\lambda\}\bigr) \tag{3.37}$$

so that $\sigma\bigl(\{X_\lambda\}_{\lambda\in\Lambda}\bigr) \subset \mathcal{F}$. We call $\sigma\bigl(\{X_\lambda\}_{\lambda\in\Lambda}\bigr)$ the *σ-algebra generated by* $\{X_\lambda\}_{\lambda\in\Lambda}$.

By definition, $\sigma\bigl(\{X_\lambda\}_{\lambda\in\Lambda}\bigr)$ is the smallest σ-algebra in Ω with respect to which X_λ is measurable for any $\lambda \in \Lambda$.

Definition 3.45 (Tail σ-algebra)**.** Let $\{X_n\}_{n=1}^{\infty}$ be random variables with X_n taking values in a measurable space $(S_n, \mathcal{B}_n)$ for each $n \in \mathbb{N}$. We define

$$\sigma_\infty\bigl(\{X_n\}_{n=1}^{\infty}\bigr) := \bigcap_{n=1}^{\infty} \sigma\bigl(\{X_k\}_{k=n}^{\infty}\bigr), \tag{3.38}$$

so that $\sigma_\infty\bigl(\{X_n\}_{n=1}^{\infty}\bigr)$ is a σ-algebra in Ω and $\sigma_\infty\bigl(\{X_n\}_{n=1}^{\infty}\bigr) \subset \mathcal{F}$. $\sigma_\infty\bigl(\{X_n\}_{n=1}^{\infty}\bigr)$ is called the *tail σ-algebra of* $\{X_n\}_{n=1}^{\infty}$, and each $A \in \sigma_\infty\bigl(\{X_n\}_{n=1}^{\infty}\bigr)$ is called a *tail event for* $\{X_n\}_{n=1}^{\infty}$.

Theorem 3.46 (Kolmogorov's 0-1 law)**.** *Let $\{X_n\}_{n=1}^{\infty}$ be random variables with X_n taking values in a measurable space $(S_n, \mathcal{B}_n)$ for each $n \in \mathbb{N}$. If $\{X_n\}_{n=1}^{\infty}$ is independent, then for any $A \in \sigma_\infty\bigl(\{X_n\}_{n=1}^{\infty}\bigr)$, $\mathbb{P}[A]$ is either* 0 *or* 1.

Proof. Let $n \in \mathbb{N}$ and set

$$\mathcal{A}_n := \big\{\{X_1 \in A_1, \dots, X_n \in A_n\} \bigm| A_i \in \mathcal{B}_i,\ i \in \{1, \dots, n\}\big\},$$
$$\mathcal{D}_n := \big\{\{X_{n+1} \in A_{n+1}, \dots, X_{n+k} \in A_{n+k}\} \bigm| k \in \mathbb{N}, A_{n+i} \in \mathcal{B}_{n+i}, i \in \{1, \dots, k\}\big\},$$

so that $\mathcal{D}_n$ is a π-system containing Ω and $\sigma_\Omega(\mathcal{D}_n) = \sigma\big(\{X_i\}_{i=n+1}^\infty\big)$.

First we prove that for any $A \in \mathcal{A}_n$ and $B \in \sigma\big(\{X_i\}_{i=n+1}^\infty\big)$,

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]. \tag{3.39}$$

The independence of $\{X_i\}_{i=1}^\infty$ implies that (3.39) holds for $A \in \mathcal{A}_n$ and $B \in \mathcal{D}_n$. If we fix $A \in \mathcal{A}_n$, then $\mathbb{P}[A \cap (\cdot)]$ and $\mathbb{P}[A]\mathbb{P}$ are both $[0,1]$-valued measures on $\mathcal{F}$ and they coincide on $\mathcal{D}_n$, so that they are equal on $\sigma_\Omega(\mathcal{D}_n) = \sigma\big(\{X_i\}_{i=n+1}^\infty\big)$ by Theorem 2.5. Thus (3.39) holds for any $A \in \mathcal{A}_n$ and $B \in \sigma\big(\{X_i\}_{i=n+1}^\infty\big)$.

Now let $B \in \sigma_\infty\big(\{X_n\}_{n=1}^\infty\big)$. If $A \in \mathcal{D}_1$, then (3.39) holds since $A \in \mathcal{A}_n$ for some $n \in \mathbb{N}$ and $B \in \sigma\big(\{X_i\}_{i=n+1}^\infty\big)$. This means that the $[0,1]$-valued measures $\mathbb{P}[(\cdot) \cap B]$ and $\mathbb{P}[B]\mathbb{P}$ coincide on $\mathcal{D}_1$, and hence Theorem 2.5 again implies that they are equal on $\sigma_\Omega(\mathcal{D}_1) = \sigma\big(\{X_i\}_{i=2}^\infty\big)$, i.e. $\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$ for any $A \in \sigma\big(\{X_i\}_{i=2}^\infty\big)$. Since $B \in \sigma\big(\{X_i\}_{i=2}^\infty\big)$, we may let $A := B$ here and obtain $\mathbb{P}[B] = \mathbb{P}[B]^2$. Thus $\mathbb{P}[B]$ is either 0 or 1. ☐

Corollary 3.47. *Let $\{X_n\}_{n=1}^\infty$ be random variables with X_n taking values in a measurable space $(S_n, \mathcal{B}_n)$ for each $n \in \mathbb{N}$. If $\{X_n\}_{n=1}^\infty$ is independent and $Z : \Omega \to [-\infty, \infty]$ is $\sigma_\infty\big(\{X_n\}_{n=1}^\infty\big)$-measurable, then $Z = c$ a.s. for some $c \in [-\infty, \infty]$.*

Proof. Let $F(x) := \mathbb{P}[Z \le x]$ for $x \in \mathbb{R}$. Since $\{Z \le x\}$ is a tail event for $\{X_n\}_{n=1}^\infty$, $F(x)$ is either 0 or 1 for each $x \in \mathbb{R}$ by Theorem 3.46. If $F(x) = 0$ for any $x \in \mathbb{R}$, then $\mathbb{P}[Z < \infty] = \lim_{n\to\infty} \mathbb{P}[Z \le n] = 0$ and $Z = \infty$ a.s. If $F(x) = 1$ for any $x \in \mathbb{R}$, then $\mathbb{P}[Z = -\infty] = \lim_{n\to\infty} \mathbb{P}[Z \le -n] = 1$ and $Z = -\infty$ a.s.

Thus we may assume that $F(a) = 0$ and $F(b) = 1$ for some $a, b \in \mathbb{R}$. Then since F is non-decreasing, $a < b$, $F(x) = 0$ for any $x \in (-\infty, a]$ and $F(x) = 1$ for any $x \in [b, \infty)$. Now let $c := \sup\{x \in \mathbb{R} \mid F(x) = 0\}$, so that $c \in [a, b]$. Then $F(x) = 1$ for any $x \in (c, \infty)$, $F(x) = 0$ for any $x \in (-\infty, c)$, and hence $\mathbb{P}[Z \le c] = F(c) = \lim_{n\to\infty} F(c+1/n) = 1, \mathbb{P}[Z < c] = \lim_{n\to\infty} F(c-1/n) = 0$. Thus $\mathbb{P}[Z = c] = \mathbb{P}[Z \le c] - \mathbb{P}[Z < c] = 1$. ☐

Example 3.48. Let $\{X_n\}_{n=1}^\infty$ be real random variables. Then the following $[-\infty, \infty]$-valued random variables are all $\sigma_\infty\big(\{X_n\}_{n=1}^\infty\big)$-measurable:

$$\limsup_{n\to\infty} X_n, \quad \liminf_{n\to\infty} X_n, \quad \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i, \quad \liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i. \tag{3.40}$$

Indeed, let $N \in \mathbb{N}$. Then for any $n \in \mathbb{N}$ with $n \ge N$, $\sup_{k\ge n} X_k$ is $\sigma\big(\{X_k\}_{k=N}^\infty\big)$-measurable, and hence so is $\limsup_{n\to\infty} X_n = \lim_{n\to\infty} \sup_{k\ge n} X_k$. Moreover, by $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{N-1} X_i = 0$, we have

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i = \limsup_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{N-1} X_i + \frac{1}{n}\sum_{i=N}^n X_i\right) = 0 + \limsup_{n\to\infty} \frac{1}{n}\sum_{i=N}^\infty X_i,$$

which is $\sigma\big(\{X_k\}_{k=N}^{\infty}\big)$-measurable. Since $N \in \mathbb{N}$ is arbitrary, $\limsup_{n\to\infty} X_n$ and $\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i$ are $\bigcap_{N=1}^{\infty} \sigma\big(\{X_k\}_{k=N}^{\infty}\big) = \sigma_{\infty}\big(\{X_n\}_{n=1}^{\infty}\big)$-measurable. The same proof applies to $\liminf_{n\to\infty} X_n$ and $\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i$ as well.

Therefore by Corollary 3.47, if $\{X_n\}_{n=1}^{\infty}$ is independent, then the random variables in (3.40) are constant a.s.

3.4 Convergence of Random Variables

In the next section, we consider convergence of the form

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_n = m \tag{3.41}$$

for independent real random variables $\{X_n\}_{n=1}^{\infty}$ such that $\mathbb{E}[|X_n|] < \infty$ and $\mathbb{E}[X_n] = m$ for any $n \in \mathbb{N}$. Such a convergence is called a *law of large numbers*. In probability theory, however, there are several ways of *"convergence"* of random variables, depending on how one measures the size of the difference between each random variable and the limit. The purpose of this section is to introduce various notions of convergence of random variables and study relations between them.

Again throughout this section, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a probability space, and random variables are always assumed to be defined on $(\Omega, \mathcal{F}, \mathbb{P})$ unless otherwise stated.

Definition 3.49. Let $d \in \mathbb{N}$, and let $X, \{X_n\}_{n=1}^{\infty}$ be d-dimensional random variables.
(1) We say that $\{X_n\}_{n=1}^{\infty}$ converges to X *almost surely* and write

$$X_n \xrightarrow{\text{a.s.}} X$$

if and only if $\lim_{n\to\infty} X_n = X$ a.s., that is, $\mathbb{P}[\lim_{n\to\infty} X_n = X] = 1$.
(2) We say that $\{X_n\}_{n=1}^{\infty}$ converges to X *in probability* and write

$$X_n \xrightarrow{\text{P}} X$$

if and only if

$$\lim_{n\to\infty} \mathbb{P}[|X_n - X| \geq \varepsilon] = 0 \qquad \text{for any } \varepsilon \in (0, \infty) \tag{3.42}$$

(that is, X_n converges to X in $\mathbb{P}$-measure; recall the definition before Problem 1.33).
(3) We say that $\{X_n\}_{n=1}^{\infty}$ converges to X *in law* (or *in distribution*) and write[2]

$$X_n \xrightarrow{\mathcal{L}} X$$

if and only if, for any bounded continuous function[3] $f : \mathbb{R}^d \to \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)], \tag{3.43}$$

[2] $X_n \xrightarrow{\mathcal{L}} X$ is also written as $X_n \xrightarrow{\mathcal{D}} X$, but we do not use this latter notation in this course.
[3] Recall that a function $f : S \to \mathbb{C}$ on a set S is called *bounded* if and only if $\sup_{x\in S} |f(x)| < \infty$.

or equivalently (by virtue of Theorem 3.10),

$$\lim_{n\to\infty}\int_{\mathbb{R}^d} f(x)\mathbb{P}_{X_n}(dx) = \int_{\mathbb{R}^d} f(x)\mathbb{P}_X(dx). \tag{3.44}$$

(4) Let $p \in (0,\infty)$. We say that $\{X_n\}_{n=1}^{\infty}$ converges to X *in* L^p and write

$$X_n \xrightarrow{L^p} X$$

if and only if

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0. \tag{3.45}$$

(5) (Adopted from Grigor'yan [3, Section 5.6]) We say that $\{X_n\}_{n=1}^{\infty}$ converges to X *in the Borel-Cantelli sense* and write

$$X_n \xrightarrow{\mathrm{BC}} X$$

if and only if

$$\sum_{n=1}^{\infty} \mathbb{P}[|X_n - X| \geq \varepsilon] < \infty \qquad \text{for any } \varepsilon \in (0,\infty). \tag{3.46}$$

Remark 3.50. Note that *only the laws of* X *and* X_n, $n \in \mathbb{N}$, *are involved in the definition* (3.44) *of* $X_n \xrightarrow{\mathcal{L}} X$. In particular, in defining $X_n \xrightarrow{\mathcal{L}} X$, we do **not** have to assume that X and $\{X_n\}_{n=1}^{\infty}$ are defined on the same probability space,

Theorem 3.51. *Let* $d \in \mathbb{N}$, *and let* $X, \{X_n\}_{n=1}^{\infty}$ *be* d*-dimensional random variables. Let* $p \in (0,\infty)$. *Then we have the following four implications:*

$$\left(X_n \xrightarrow{\mathrm{BC}} X\right) \Longrightarrow \left(X_n \xrightarrow{\mathrm{a.s.}} X\right) \Longrightarrow \left(X_n \xrightarrow{\mathrm{P}} X\right) \Longrightarrow \left(X_n \xrightarrow{\mathcal{L}} X\right) \tag{3.47}$$

$$\left(X_n \xrightarrow{L^p} X\right) \Longrightarrow \left(X_n \xrightarrow{\mathrm{P}} X\right) \tag{3.48}$$

Proof. The implication (3.48) has been already seen in Problem 1.33-(1).

Assume $X_n \xrightarrow{\mathrm{BC}} X$. Then the first Borel-Cantelli lemma (Theorem 3.41-(1)) shows that $\mathbb{P}[\liminf_{n\to\infty}\{|X_n - X| < 1/k\}] = 1$ for any $k \in \mathbb{N}$ and hence that

$$\mathbb{P}[\Omega_0] = 1, \quad \text{where} \quad \Omega_0 := \bigcap_{k=1}^{\infty} \liminf_{n\to\infty}\bigl\{|X_n - X| < 1/k\bigr\}.$$

Let $\omega \in \Omega_0$ and $\varepsilon \in (0,\infty)$. Choose $k \in \mathbb{N}$ so that $1/k < \varepsilon$. Since ω belongs to $\liminf_{n\to\infty}\{|X_n - X| < 1/k\}$, there exists $N_k(\omega) \in \mathbb{N}$ such that

$$|X_n(\omega) - X(\omega)| < 1/k < \varepsilon \qquad \text{for any } n \geq N_k(\omega),$$

which means $\lim_{n\to\infty} X_n(\omega) = X(\omega)$. Thus $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ for any $\omega \in \Omega_0$ and hence $X_n \xrightarrow{\mathrm{a.s.}} X$ by $\mathbb{P}[\Omega_0] = 1$.

Next suppose $X_n \xrightarrow{\text{a.s.}} X$ and let $\varepsilon \in (0,\infty)$. Then since $\mathbf{1}_{\{|X_n - X| \geq \varepsilon\}} \xrightarrow{\text{a.s.}} 0$, $0 \leq \mathbf{1}_{\{|X_n - X| \geq \varepsilon\}} \leq 1$ and $\mathbb{E}[1] = 1 < \infty$, the dominated convergence theorem (Theorem 1.32) yields

$$\mathbb{P}[|X_n - X| \geq \varepsilon] = \mathbb{E}\big[\mathbf{1}_{\{|X_n - X| \geq \varepsilon\}}\big] \xrightarrow{n\to\infty} 0.$$

Thus $X_n \xrightarrow{\text{P}} X$.

Finally, assume $X_n \xrightarrow{\text{P}} X$ and let $f : \mathbb{R}^d \to \mathbb{R}$ be bounded and continuous. Suppose $\mathbb{E}[f(X_n)]$ does not converge to $\mathbb{E}[f(X)]$. Then there exist $\varepsilon \in (0,\infty)$ and a strictly increasing sequence $\{n(k)\}_{k=1}^{\infty} \subset \mathbb{N}$ such that

$$\big|\mathbb{E}[f(X_{n(k)})] - \mathbb{E}[f(X)]\big| \geq \varepsilon \qquad \text{for any } k \in \mathbb{N}. \tag{3.49}$$

On the other hand, $X_n \xrightarrow{\text{P}} X$ yields $X_{n(k)} \xrightarrow{\text{P}} X$, and by Problem 1.33-(2) there exists a further strictly increasing sequence $\{k(\ell)\}_{\ell=1}^{\infty} \subset \mathbb{N}$ such that $X_{n(k(\ell))} \xrightarrow{\text{a.s.}} X$. Then $f(X_{n(k(\ell))}) \xrightarrow{\text{a.s.}} f(X)$ by the continuity of f, $|f(X_n)| \leq \sup_{x\in\mathbb{R}^d} |f(x)| < \infty$, and therefore an application of the dominated convergence theorem (Theorem 1.32) yields $\lim_{\ell\to\infty} \mathbb{E}[f(X_{n(k(\ell))})] = \mathbb{E}[f(X)]$, which contradicts (3.49). Thus we obtain $\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$ and hence $X_n \xrightarrow{\mathcal{L}} X$. □

Theorem 3.52. *Let $d \in \mathbb{N}$, and let $X, \{X_n\}_{n=1}^{\infty}$ be d-dimensional random variables. Then $X_n \xrightarrow{\text{P}} X$ if and only if for any strictly increasing sequence $\{n(k)\}_{k=1}^{\infty} \subset \mathbb{N}$ there exists a further strictly increasing sequence $\{k(\ell)\}_{\ell=1}^{\infty} \subset \mathbb{N}$ such that $X_{n(k(\ell))} \xrightarrow{\text{a.s.}} X$.*

Proof. If $X_n \xrightarrow{\text{P}} X$ and $\{n(k)\}_{k=1}^{\infty} \subset \mathbb{N}$ is strictly increasing, then $X_{n(k)} \xrightarrow{\text{P}} X$, and Problem 1.33-(2) implies that $X_{n(k(\ell))} \xrightarrow{\text{a.s.}} X$ for some strictly increasing $\{k(\ell)\}_{\ell=1}^{\infty} \subset \mathbb{N}$. Conversely if $X_n \xrightarrow{\text{P}} X$ does not hold, then for some $\varepsilon \in (0,\infty)$, $\mathbb{P}[|X_n - X| \geq \varepsilon]$ does not converge to 0 and hence there exist $\delta \in (0,\infty)$ and a strictly increasing sequence $\{n(k)\}_{k=1}^{\infty} \subset \mathbb{N}$ such that $\mathbb{P}[|X_{n(k)} - X| \geq \varepsilon] \geq \delta$ for any $k \in \mathbb{N}$. Then for any strictly increasing $\{k(\ell)\}_{\ell=1}^{\infty}$, $X_{n(k(\ell))} \xrightarrow{\text{P}} X$ does not hold and hence neither does $X_{n(k(\ell))} \xrightarrow{\text{a.s.}} X$ by (3.47) of Theorem 3.51. □

Corollary 3.53. *Let $d, k \in \mathbb{N}$, let $X, \{X_n\}_{n=1}^{\infty}$ be d-dimensional random variables and let $f : \mathbb{R}^d \to \mathbb{R}^k$ be continuous.*
(1) If $X_n \xrightarrow{\text{a.s.}} X$ then $f(X_n) \xrightarrow{\text{a.s.}} f(X)$.
(2) If $X_n \xrightarrow{\text{P}} X$ then $f(X_n) \xrightarrow{\text{P}} f(X)$.
(3) If $X_n \xrightarrow{\mathcal{L}} X$ then $f(X_n) \xrightarrow{\mathcal{L}} f(X)$.

Proof. (1) is obvious. (2) follows from Theorem 3.52, since the latter condition of Theorem 3.52 for $\{X_n\}_{n=1}^{\infty}$ implies that for $\{f(X_n)\}_{n=1}^{\infty}$.
(3) If $g : \mathbb{R}^k \to \mathbb{R}$ is bounded and continuous, then $g \circ f : \mathbb{R}^d \to \mathbb{R}$ is also bounded and continuous, and therefore

$$\mathbb{E}[g(f(X_n))] = \mathbb{E}[(g \circ f)(X_n)] \xrightarrow{n\to\infty} \mathbb{E}[(g \circ f)(X)] = \mathbb{E}[g(f(X))]$$

by $X_n \xrightarrow{\mathcal{L}} X$. Thus $f(X_n) \xrightarrow{\mathcal{L}} f(X)$. □

Example 3.54. Let us show that the converses of the implications in Theorem 3.51 are **not** true in general: for $p \in (0, \infty)$,

$$\left(X_n \xrightarrow{\mathcal{L}} X\right) \not\Rightarrow \left(X_n \xrightarrow{\mathrm{P}} X\right) \not\Rightarrow \left(X_n \xrightarrow{\text{a.s.}} X\right) \not\Rightarrow \left(X_n \xrightarrow{\mathrm{BC}} X\right)$$
$$\left(X_n \xrightarrow{\mathrm{P}} X\right) \not\Rightarrow \left(X_n \xrightarrow{L^p} X\right)$$

For this purpose, we consider the probability space $\big([0,1], \mathcal{B}, \mathrm{m}_1\big)$, $\mathcal{B} := \mathcal{B}\big([0,1]\big)$.
(1) Define $X(\omega) := \omega$, $\omega \in [0,1]$, and for $n \in \mathbb{N}$,

$$X_n := \sum_{k=1}^{n} \frac{k}{n} \mathbf{1}_{\left(\frac{k-1}{n}, \frac{k}{n}\right]}.$$

Then clearly $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ for any $\omega \in [0,1]$. In particular, as random variables on $\big([0,1], \mathcal{B}, \mathrm{m}_1\big)$, $X_n \xrightarrow{\text{a.s.}} X$ and hence $X_n \xrightarrow{\mathcal{L}} X$ by Theorem 3.51. On the other hand, for any $a \in [0,1]$, $\mathrm{m}_1\big(1-X \le a\big) = \mathrm{m}_1(X \ge 1-a) = \mathrm{m}_1\big([1-a, 1]\big) = a$, and hence $\mathcal{L}(1-X) = \mathcal{L}(X)$. Thus $X_n \xrightarrow{\mathcal{L}} 1-X$, but subsequences of $\{X_n(\omega)\}_{n=1}^{\infty}$ can converge to $1 - X(\omega)$ only if $1 - X(\omega) = X(\omega)$, i.e. $\omega = 1/2$, and hence $X_n \xrightarrow{\mathrm{P}} 1 - X$ does not hold.
(2) Let $\{I_n\}_{n=1}^{\infty}$ be the sequence of intervals given by $I_n := \left[\frac{n-2^k}{2^k}, \frac{n-2^k+1}{2^k}\right]$, $k := \max\{\ell \in \mathbb{N} \cup \{0\} \mid 2^\ell \le n\}$, that is,

$$I_1, I_2, I_3, \ldots, := [0,1], \left[0, \tfrac{1}{2}\right], \left[\tfrac{1}{2}, 1\right], \left[0, \tfrac{1}{4}\right], \left[\tfrac{1}{4}, \tfrac{1}{2}\right], \left[\tfrac{1}{2}, \tfrac{3}{4}\right], \left[\tfrac{3}{4}, 1\right], \left[0, \tfrac{1}{8}\right] \ldots,$$

and define real random variables $\{X_n\}_{n=1}^{\infty}$ on $\big([0,1], \mathcal{B}, \mathrm{m}_1\big)$ by $X_n := \mathbf{1}_{I_n}$. Then for $\varepsilon \in (0,1]$, $\mathrm{m}_1(|X_n| \ge \varepsilon) = \mathrm{m}_1(I_n) \to 0$ as $n \to \infty$, and hence $X_n \xrightarrow{\mathrm{P}} 0$. On the other hand, for each $\omega \in [0,1]$, $X_n(\omega) = \mathbf{1}_{I_n}(\omega) = 1$ for infinitely many $n \in \mathbb{N}$, and hence $\{X_n(\omega)\}_{n=1}^{\infty}$ does not converge to 0. Thus $X_n \xrightarrow{\text{a.s.}} 0$ does not hold.
(3) For $n \in \mathbb{N}$ let $I_n := [0, 1/n]$ and define a real random variable X_n on $\big([0,1], \mathcal{B}, \mathrm{m}_1\big)$ by $X_n := n^{1/p}\mathbf{1}_{I_n}$. Then $\lim_{n\to\infty} X_n(\omega) = 0$ for $\omega \in (0,1]$, so that $X_n \xrightarrow{\text{a.s.}} 0$, but for $\varepsilon \in (0,1]$, $\mathrm{m}_1(|X_n| \ge \varepsilon) = \mathrm{m}_1(I_n) = 1/n$ and hence $\sum_{n=1}^{\infty} \mathrm{m}_1(|X_n| \ge \varepsilon) = \infty$. Therefore $X_n \xrightarrow{\mathrm{BC}} 0$ does not hold. Moreover, $\mathrm{m}_1(|X_n| \ge \varepsilon) = 1/n \to 0$ for $\varepsilon \in (0,1]$ and hence $X_n \xrightarrow{\mathrm{P}} 0$, but $\int_{[0,1]} |X_n|^p d\mathrm{m}_1 = n\mathrm{m}_1(I_n) = 1$ for any $n \in \mathbb{N}$, so that $X_n \xrightarrow{L^p} 0$ does not hold.

3.5 Laws of Large Numbers

Once again throughout this section, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a probability space, and random variables are always assumed to be defined on $(\Omega, \mathcal{F}, \mathbb{P})$ unless otherwise stated.

As described at the beginning of the last section, in this section we prove laws of large numbers, which assert convergence of the form (3.41) for independent real random variables $\{X_n\}_{n=1}^{\infty}$ with $\mathbb{E}[X_n] = m$, $n \in \mathbb{N}$. The most important case is that of *independent and identically distributed* random variables, which appear quite often in probability theory:

Definition 3.55 (Independent and identically distributed, i.i.d.)**.** Let $(S, \mathcal{B})$ be a measurable space and let $\{X_n\}_{n=1}^{N}$ be $(S, \mathcal{B})$-valued random variables, where $N \in \mathbb{N} \cup \{\infty\}$. $\{X_n\}_{n=1}^{N}$ is called *independent and identically distributed*, or *i.i.d.* for short, if and only if it is independent and $\mathcal{L}(X_n) = \mathcal{L}(X_1)$ for any $n \in \mathbb{N}$, $n \leq N$.

Note that by Theorem 3.40, *for any measurable space* $(S, \mathcal{B})$ *and any probability measure* μ *on* $(S, \mathcal{B})$*, there exist a probability space* $(\Omega', \mathcal{F}', \mathbb{P}')$ *and i.i.d.* $(S, \mathcal{B})$*-valued random variables* $\{X_n\}_{n=1}^{\infty}$ *on* $(\Omega', \mathcal{F}', \mathbb{P}')$ *with* $X_1 \sim \mu$.

Theorem 3.56 (Weak law of large numbers)**.** *Let* $m \in \mathbb{R}$*, and let* $\{X_n\}_{n=1}^{\infty} \subset \mathcal{L}^2(\mathbb{P})$ *be independent and satisfy* $\mathbb{E}[X_n] = m$ *for any* $n \in \mathbb{N}$ *and* $\sup_{n\in\mathbb{N}} \mathrm{var}(X_n) < \infty$*. Then the* weak law of large numbers *holds, that is,*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathrm{P}} m. \tag{3.50}$$

In particular, the weak law of large numbers holds for any i.i.d. $\{X_n\}_{n=1}^{\infty} \subset \mathcal{L}^2(\mathbb{P})$.

Proof. Let $\varepsilon \in (0, \infty)$, $n \in \mathbb{N}$ and set $S_n := \sum_{i=1}^{n} X_i$. Then by Chebyshev's inequality (Problem 1.18 with $\varphi(x) = x^2$) and (3.32) of Proposition 3.32,

$$\begin{aligned}\mathbb{P}\left[\left|\frac{S_n}{n} - m\right| \geq \varepsilon\right] &\leq \frac{1}{\varepsilon^2}\mathbb{E}\left[\left|\frac{S_n}{n} - m\right|^2\right] = \frac{1}{\varepsilon^2 n^2}\mathrm{var}(S_n) = \frac{1}{\varepsilon^2 n^2}\sum_{i=1}^{n}\mathrm{var}(X_i)\\ &\leq \frac{\sup_{k\in\mathbb{N}}\mathrm{var}(X_k)}{\varepsilon^2 n} \xrightarrow{n\to\infty} 0. \end{aligned} \tag{3.51}$$

Thus $S_n/n \xrightarrow{\mathrm{P}} m$. □

The estimate (3.51) can be used to prove the following well-known result from calculus.

Theorem 3.57 (Weierstrass approximation theorem)**.** *Let* $a, b \in \mathbb{R}$*,* $a < b$*, and let* $f : [a, b] \to \mathbb{R}$ *be continuous. Then for any* $\varepsilon \in (0, \infty)$*, there exists a polinomial* $P(x) = \sum_{k=0}^{n} a_k x^k$*, where* $n \in \mathbb{N} \cup \{0\}$ *and* $\{a_k\}_{k=0}^{n} \subset \mathbb{R}$*, such that*

$$\sup_{x\in[a,b]} |f(x) - P(x)| < \varepsilon. \tag{3.52}$$

Proof. It suffices to consider the case of $[a, b] = [0, 1]$. As in Example 3.33, let $p \in [0, 1]$, $n \in \mathbb{N}$, and let $\{X_i\}_{i=1}^{n}$ be independent Bernoulli random variables of probability p, which exist by Theorem 3.28. Then $S_n := \sum_{i=1}^n X_i$ is a binomial random variable of size n and probability p, and

$$\mathbb{E}[f(S_n/n)] = \sum_{k=0}^{n} f(k/n)\mathbb{P}[S_n = k] = \sum_{k=0}^{n} f(k/n)\binom{n}{k} p^k(1-p)^{n-k} =: B_n(p). \tag{3.53}$$

The polinomial $B_n(p)$ is called the *Bernstein polynomial for f of degree n*. We claim that

$$\lim_{n\to\infty} \sup_{p\in[0,1]} |f(p) - B_n(p)| = 0, \tag{3.54}$$

which immediately implies the assertion. Let $\varepsilon \in (0,\infty)$. Recall that f is *uniformly continuous on* $[0,1]$: there exists $\delta \in (0,\infty)$ such that for any $x, y \in [0,1]$,

$$|x - y| < \delta \Longrightarrow |f(x) - f(y)| < \varepsilon. \tag{3.55}$$

Note also that $M := \sup_{x\in[0,1]} |f(x)| < \infty$. By using (3.53), (3.55) and (3.51), for $n \in \mathbb{N}$ we obtain

$$\begin{aligned}
|f(p) - B_n(p)| &= \bigl|\mathbb{E}[f(p) - f(S_n/n)]\bigr| \le \mathbb{E}\bigl[\bigl|f(p) - f(S_n/n)\bigr|\bigr] \\
&= \mathbb{E}\bigl[\bigl|f(p) - f(S_n/n)\bigr|\mathbf{1}_{\{|S_n/n-p|<\delta\}}\bigr] + \mathbb{E}\bigl[\bigl|f(p) - f(S_n/n)\bigr|\mathbf{1}_{\{|S_n/n-p|\ge\delta\}}\bigr] \\
&\le \varepsilon + 2M\mathbb{P}\left[\left|\frac{S_n}{n} - p\right| \ge \delta\right] \le \varepsilon + 2M\frac{p(1-p)}{\delta^2 n} \le \varepsilon + \frac{M}{\delta^2 n},
\end{aligned}$$

which implies that $\sup_{p\in[0,1]} |f(p) - B_n(p)| \le 2\varepsilon$ for any $n \in \mathbb{N}$ with $n \ge M/(\delta^2\varepsilon)$. Thus (3.54) is proved. □

Theorem 3.58 (Strong law of large numbers)**.** *If $\{X_n\}_{n=1}^\infty \subset \mathcal{L}^2(\mathbb{P})$ is i.i.d., then the* strong law of large numbers *holds, that is,*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{a.s.}} \mathbb{E}[X_1]. \tag{3.56}$$

Proof. If we define $f, g : \mathbb{R} \to \mathbb{R}$ by $f(x) := \max\{x, 0\}$ and $g(x) := -\min\{x, 0\}$, then $X_n^+ = f(X_n)$ and $X_n^- = g(X_n)$. Therefore by Proposition 3.31-(2), $\{X_n^+\}_{n=1}^\infty$ and $\{X_n^-\}_{n=1}^\infty$ are both independent, and it is clear that they are identically distributed and belong to $\mathcal{L}^2(\mathbb{P})$. It is also immediate that (3.56) follows from $\frac{1}{n}\sum_{i=1}^n X_i^+ \xrightarrow{\text{a.s.}} \mathbb{E}[X_1^+]$ and $\frac{1}{n}\sum_{i=1}^n X_i^- \xrightarrow{\text{a.s.}} \mathbb{E}[X_1^-]$. Thus by considering $\{X_n^+\}_{n=1}^\infty$ and $\{X_n^-\}_{n=1}^\infty$ instead of $\{X_n\}_{n=1}^\infty$, we assume without loss of generality that $X_n \ge 0$ for any $n \in \mathbb{N}$.

Set $S_n := \sum_{i=1}^n X_i$ for $n \in \mathbb{N}$. Then for any $\varepsilon \in (0,\infty)$, (3.51) yields

$$\sum_{k=1}^{\infty} \mathbb{P}\left[\left|\frac{S_{k^2}}{k^2} - \mathbb{E}[X_1]\right| \ge \varepsilon\right] \le \sum_{k=1}^{\infty} \frac{\mathrm{var}(X_1)}{\varepsilon^2 k^2} < \infty$$

and hence $S_{k^2}/k^2 \xrightarrow{\text{BC}} \mathbb{E}[X_1]$. Then $S_{k^2}/k^2 \xrightarrow{\text{a.s.}} \mathbb{E}[X_1]$ by Theorem 3.51.

Let $n \in \mathbb{N}$ and set $k(n) := \max\{k \in \mathbb{N} \mid k^2 \leq n\}$, so that $k(n)^2 \leq n < (k(n)+1)^2$. Then $k(n) > \sqrt{n} - 1$, hence $\lim_{n\to\infty} k(n) \to \infty$, and

$$\begin{aligned}\left(\frac{k(n)}{k(n)+1}\right)^2 \frac{S_{k(n)^2}}{k(n)^2} = \frac{S_{k(n)^2}}{(k(n)+1)^2} \leq \frac{S_n}{n} \leq \frac{S_{(k(n)+1)^2}}{k(n)^2} \\ = \left(\frac{k(n)+1}{k(n)}\right)^2 \frac{S_{(k(n)+1)^2}}{(k(n)+1)^2}.\end{aligned} \tag{3.57}$$

Since $\lim_{n\to\infty} k(n) = \infty$, $\lim_{n\to\infty}(k(n)+1)/k(n) = \lim_{n\to\infty}\bigl(1 + 1/k(n)\bigr) = 1$ and $S_{k^2}/k^2 \xrightarrow{\text{a.s.}} \mathbb{E}[X_1]$, letting $n \to \infty$ in (3.57) results in $S_n/n \xrightarrow{\text{a.s.}} \mathbb{E}[X_1]$. □

Example 3.59. (1) Let $p \in [0,1]$ and let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. Bernoulli random variables of probability p. Then since $\mathbb{E}[X_1] = \mathbb{E}[X_1^2] = p$, Theorem 3.58 yields

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{a.s.}} p,$$

This result fits our intuition that, if we flip a coin and see the outcome (heads or tails) very many times, then the number of heads divided by the total number of trials should give an approximation of the probability for the coin to show heads.
(2) Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. $\{1,2,3,4,5,6\}$-valued random variables with $\mathbb{P}[X_1 = k] = 1/6$ for any $k \in \{1,\dots,6\}$. Then for any $k \in \{1,\dots,6\}$, Theorem 3.58 yields

$$\frac{\#\{i \in \{1,\dots,n\} \mid X_i = k\}}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{k\}}(X_i) \xrightarrow{\text{a.s.}} \frac{1}{6}$$

since $\{\mathbf{1}_{\{k\}}(X_n)\}_{n=1}^{\infty}$ is i.i.d. Bernoulli random variables of probability $1/6$. This result again fits our intuition that, if we throw very many times a dice whose all sides are equally likely to appear, then all sides should appear approximately the same number of times.

Example 3.60. Consider the probability space $\bigl([0,1), \mathcal{B}, \mathrm{m}_1\bigr)$, $\mathcal{B} := \mathcal{B}\bigl([0,1)\bigr)$. For each $\omega \in [0,1)$, let

$$\omega = 0.\omega_1\omega_2\omega_3\dots$$

be the usual decimal expansion of ω, where we choose the finite decimal expansion if exists. Then let $X_n(\omega) := \omega_n$ for $n \in \mathbb{N}$. Since

$$\{X_n = k\} = \bigcup_{j=0}^{10^{n-1}-1} \left[\frac{j}{10^{n-1}} + \frac{k}{10^n}, \frac{j}{10^{n-1}} + \frac{k+1}{10^n}\right)$$

for each $k \in \{0,\dots,9\}$, X_n is a $\{0,\dots,9\}$-valued random variable on $\bigl([0,1), \mathcal{B}, \mathrm{m}_1\bigr)$ and

$$\mathrm{m}_1(X_n = k) = 10^{n-1}10^{-n} = \frac{1}{10} \qquad \text{for any } k \in \{0,\dots,9\}. \tag{3.58}$$

Moreover, for each $\{k_i\}_{i=1}^n \subset \{0, \ldots, 9\}$,

$$\mathrm{m}_1(X_1 = k_1, \ldots, X_n = k_n) = \mathrm{m}_1\left(\left[\sum_{i=1}^n \frac{k_i}{10^i}, \sum_{i=1}^n \frac{k_i}{10^i} + \frac{1}{10^n}\right)\right) = \frac{1}{10^n},$$

from which it immediately follows that $\{X_n\}_{n=1}^\infty$ is independent. Thus $\{X_n\}_{n=1}^\infty$ is i.i.d.

Now for $k \in \{0, \ldots, 9\}$, let

$$A_k := \left\{\omega \in [0,1) \;\middle|\; \lim_{n\to\infty} \frac{\#\{i \in \{1, \ldots, n\} \mid \omega_i = k\}}{n} = \frac{1}{10}\right\}. \tag{3.59}$$

Then since $\#\{i \in \{1, \ldots, n\} \mid \omega_i = k\} = \sum_{i=1}^n \mathbf{1}_{\{k\}}(X_i(\omega))$ and $\{\mathbf{1}_{\{k\}}(X_n)\}_{n=1}^\infty$ is i.i.d. Bernoulli random variables of probability $1/10$, Theorem 3.58 implies that

$$\mathrm{m}_1(A_k) = \mathrm{m}_1\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{k\}}(X_i) = \frac{1}{10}\right) = 1, \quad \text{whence} \quad \mathrm{m}_1\left(\bigcap_{k=0}^9 A_k\right) = 1. \tag{3.60}$$

The same argument applies to the *p-ary expansion* $\omega =_p 0.\omega_{p,1}\omega_{p,2}\omega_{p,3}\ldots$ of $\omega \in [0,1)$ for any $p \in \mathbb{N}$, $p \geq 2$, by replacing 10 by p. Thus if we set

$$A_{p,k} := \left\{\omega \in [0,1) \;\middle|\; \lim_{n\to\infty} \frac{\#\{i \in \{1, \ldots, n\} \mid \omega_{p,i} = k\}}{n} = \frac{1}{p}\right\} \tag{3.61}$$

for $p \geq 2$ and $k \in \{0, \ldots, p-1\}$, then similarly to (3.60) we have $\mathrm{m}_1(A_{p,k}) = 1$, and hence we conclude that

$$\mathrm{m}_1\left(\bigcap_{p=2}^\infty \bigcap_{k=0}^{p-1} A_{p,k}\right) = 1. \tag{3.62}$$

In fact, we can prove the following stronger version of the strong law of large numbers with a quite similar idea to, but by more complicated arguments than, the above proof of Theorem 3.58.

Theorem 3.61 (Strong law of large numbers)**.** *Let $\{X_n\}_{n=1}^\infty$ be i.i.d. real random variables. If $\mathbb{E}[|X_1|] < \infty$, then the strong law of large numbers holds: $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{\text{a.s.}} \mathbb{E}[X_1]$. If $\mathbb{E}[|X_1|] = \infty$, then almost surely $\left\{\frac{1}{n}\sum_{i=1}^n X_i\right\}_{n=1}^\infty$ does not converge in $\mathbb{R}$.*

We follow [1, Proof of Theorem 8.3.5] for the following proof of Theorem 3.61. We need the following lemma.

Lemma 3.62. *Let X be a non-negative real random variable. Then*

$$\mathbb{E}[X] \leq \sum_{n=0}^\infty \mathbb{P}[X > n] \leq \mathbb{E}[X] + 1. \tag{3.63}$$

In particular, $\mathbb{E}[X] < \infty$ if and only if $\sum_{n=0}^\infty \mathbb{P}[X > n] < \infty$.

Proof. Let $A_k := \{k < X \le k+1\}$, $k \in \mathbb{N} \cup \{0\}$. Then since $\{X > n\} = \bigcup_{k=n}^{\infty} A_k$,

$$\sum_{n=0}^{\infty} \mathbb{P}[X > n] = \sum_{n=0}^{\infty}\sum_{k=n}^{\infty} \mathbb{P}[A_k] = \sum_{k=0}^{\infty}\sum_{n=0}^{k} \mathbb{P}[A_k] = \sum_{k=0}^{\infty}(k+1)\mathbb{P}[A_k]. \tag{3.64}$$

Let $Y := \sum_{k=0}^{\infty} k\mathbf{1}_{A_k}$, so that $Y \le X \le Y + \mathbf{1}_{\{X>0\}}$ and $\mathbb{E}[X] \le \mathbb{E}[Y] + \mathbb{P}[X > 0] \le \mathbb{E}[X]+1$. Now since

$$\mathbb{E}[Y] + \mathbb{P}[X > 0] = \sum_{k=0}^{\infty} k\mathbb{P}[A_k] + \sum_{k=0}^{\infty} \mathbb{P}[A_k] = \sum_{n=0}^{\infty} \mathbb{P}[X > n]$$

by Proposition 1.26 and (3.64), we obtain (3.63). □

Proof of Theorem 3.61. Set $S_n := \sum_{j=1}^{n} X_j$ for $n \in \mathbb{N}$. Let $\omega \in \Omega$ and suppose for some $c \in \mathbb{R}$ we have $\lim_{n\to\infty} S_n(\omega)/n = c$. Then as $n \to \infty$, $S_{n-1}(\omega)/n = (1-1/n)S_{n-1}(\omega)/(n-1) \to c$ and hence $X_n(\omega)/n = \big(S_n(\omega) - S_{n-1}(\omega)\big)/n \to 0$. On the other hand, if $\mathbb{E}[|X_1|] = \infty$, then $\big\{\{X_n > n\}\big\}_{n=1}^{\infty}$ is independent and $\sum_{n=1}^{\infty} \mathbb{P}[|X_n| > n] = \sum_{n=1}^{\infty} \mathbb{P}[|X_1| > n] = \infty$ by Lemma 3.62. Therefore by the second Borel-Cantelli lemma (Theorem 3.41-(2)), for $\mathbb{P}$-a.e. $\omega \in \Omega$, $X_n(\omega) > n$ for infinitely many $n \in \mathbb{N}$. For such $\omega \in \Omega$, $\{X_n(\omega)/n\}_{n=1}^{\infty}$ does not converge to 0 and hence $\{S_n(\omega)/n\}_{n=1}^{\infty}$ cannot converge in $\mathbb{R}$.

For the converse, suppose $\mathbb{E}[|X_1|] < \infty$. As in the proof of Theorem 3.58, by considering $\{X_n^+\}_{n=1}^{\infty}$ and $\{X_n^-\}_{n=1}^{\infty}$ instead of $\{X_n\}_{n=1}^{\infty}$, we assume without loss of generality that $X_n \ge 0$ for any $n \in \mathbb{N}$. For $x \in \mathbb{R}$, we write $\lfloor x \rfloor := \max\{k \in \mathbb{N} \mid k \le x\}$, so that $\lfloor x \rfloor \le x < \lfloor x \rfloor + 1$. Note that $\lfloor x \rfloor \ge x/2$ for $x \in [1,\infty)$, since $1 \le \lfloor x \rfloor \le x < \lfloor x \rfloor + 1 \le 2\lfloor x \rfloor$.

Let $\alpha \in (1,\infty)$ and $\varepsilon \in (0,\infty)$. For $n \in \mathbb{N}$, we set $k(n) := \lfloor \alpha^n \rfloor$ and define

$$Y_n := X_n \mathbf{1}_{\{X_n \le n\}} = X_n \mathbf{1}_{[0,n]}(X_n) \qquad \text{and} \qquad T_n := \sum_{j=1}^{n} Y_j,$$

so that $k(n)^{-2} \le 4\alpha^{-2n}$, and $\{Y_n\}_{n=1}^{\infty}$ is independent by Proposition 3.31-(2). We first show that

$$\frac{T_{k(n)} - \mathbb{E}[T_{k(n)}]}{k(n)} \xrightarrow{\text{BC}} 0. \tag{3.65}$$

By Chebyshev's inequality (Problem 1.18 with $\varphi(x) = x^2$) and (3.32) of Proposition 3.32,

$$\begin{aligned}\Sigma_{\alpha,\varepsilon} &:= \sum_{n=1}^{\infty} \mathbb{P}\Big[\big|T_{k(n)} - \mathbb{E}[T_{k(n)}]\big| \ge \varepsilon k(n)\Big] \le \sum_{n=1}^{\infty} \frac{\operatorname{var}\big(T_{k(n)}\big)}{\big(\varepsilon k(n)\big)^2} \\ &= \varepsilon^{-2} \sum_{n=1}^{\infty}\sum_{j=1}^{k(n)} k(n)^{-2} \operatorname{var}(Y_j) = \varepsilon^{-2} \sum_{j=1}^{\infty} \operatorname{var}(Y_j) \sum_{n=1}^{\infty} k(n)^{-2} \mathbf{1}_{[j,\infty)}\big(k(n)\big).\end{aligned}$$

We would like to show $\Sigma_{\alpha,\varepsilon} < \infty$. Let μ be the law of X_1; recall that μ is the law of X_n for any $n \in \mathbb{N}$. Since $\operatorname{var}(Y_j) \le \mathbb{E}[Y_j^2] = \int_{(0,j]} x^2 \mu(dx)$ and

$$\sum_{n=1}^{\infty} k(n)^{-2} \mathbf{1}_{[j,\infty)}\big(k(n)\big) \le 4 \sum_{n=1}^{\infty} \alpha^{-2n} \mathbf{1}_{[j,\infty)}(\alpha^n) \le \frac{4}{1-\alpha^{-2}} j^{-2},$$

setting $c_{\alpha,\varepsilon} := 4\varepsilon^{-2}(1-\alpha^{-2})$, we obtain $\Sigma_{\alpha,\varepsilon} < \infty$, i.e. (3.65), as follows:

$$\Sigma_{\alpha,\varepsilon} \le c_{\alpha,\varepsilon} \sum_{j=1}^{\infty} j^{-2} \int_{(0,j]} x^2 \mu(dx) = c_{\alpha,\varepsilon} \sum_{j=1}^{\infty}\sum_{k=1}^{j} j^{-2} \int_{(k-1,k]} x^2 \mu(dx)$$

$$
\begin{aligned}
&= c_{\alpha,\varepsilon} \sum_{k=1}^{\infty} \int_{(k-1,k]} x^2 \mu(dx) \sum_{j=k}^{\infty} j^{-2} \\
&= c_{\alpha,\varepsilon} \sum_{k=1}^{\infty} \int_{(k-1,k]} x^2 \mu(dx) \int_{k}^{\infty} \big(\lfloor x \rfloor\big)^{-2} dx \\
&\leq c_{\alpha,\varepsilon} \sum_{k=1}^{\infty} \int_{(k-1,k]} x^2 \mu(dx) \int_{k}^{\infty} \frac{4}{x^2} dx \\
&= c_{\alpha,\varepsilon} \sum_{k=1}^{\infty} \int_{(k-1,k]} x^2 \mu(dx) \cdot \frac{4}{k} = 4c_{\alpha,\varepsilon} \sum_{k=1}^{\infty} \int_{(k-1,k]} \frac{x^2}{k} \mu(dx) \\
&\leq 4c_{\alpha,\varepsilon} \sum_{k=1}^{\infty} \int_{(k-1,k]} x \mu(dx) = 4c_{\alpha,\varepsilon} \int_{(0,\infty)} x\mu(dx) = 4c_{\alpha,\varepsilon} \mathbb{E}[X_1] < \infty.
\end{aligned}
$$

(3.65) and Theorem 3.51 imply that $\big(T_{k(n)} - \mathbb{E}[T_{k(n)}]\big)/k(n) \xrightarrow{\text{a.s.}} 0$. Since

$$
\mathbb{E}[Y_n] = \int_{[0,n]} x\mu(dx) \xrightarrow{n\to\infty} \int_{[0,\infty)} x\mu(dx) = \mathbb{E}[X_1]
$$

by the monotone convergence theorem (Theorem 1.24), we easily see that $\mathbb{E}[T_{k(n)}]/k(n) = \sum_{j=1}^{k(n)} \mathbb{E}[Y_j]/k(n) \to \mathbb{E}[X_1]$, and hence $T_{k(n)}/k(n) \xrightarrow{\text{a.s.}} \mathbb{E}[X_1]$. On the other hand, by Lemma 3.62,

$$
\sum_{j=1}^{\infty} \mathbb{P}[X_j \neq Y_j] = \sum_{j=1}^{\infty} \mathbb{P}[X_j > j] = \sum_{j=1}^{\infty} \mathbb{P}[X_1 > j] \leq \mathbb{E}[X_1] + 1 < \infty,
$$

and therefore the first Borel-Cantelli lemma (Theorem 3.41-(1)) implies that, for $\mathbb{P}$-a.e. $\omega \in \Omega$, there exists $\ell(\omega) \in \mathbb{N}$ such that $X_j(\omega) = Y_j(\omega)$ for any $j \geq \ell(\omega)$. For such ω, we have

$$
\frac{S_{k(n)}(\omega) - T_{k(n)}(\omega)}{k(n)} = \frac{1}{k(n)} \sum_{j=1}^{\ell(\omega)} \big(X_j(\omega) - Y_j(\omega)\big) \xrightarrow{n\to\infty} 0, \quad \text{where } \ S_n := \sum_{j=1}^{n} X_j.
$$

Hence $S_{k(n)}/k(n) \xrightarrow{\text{a.s.}} \mathbb{E}[X_1]$. Clearly $\lim_{n\to\infty} k(n+1)/k(n) = \alpha$, and therefore $1 < k(n+1)/k(n) < \alpha^2$ for any $n \geq N$ for some $N \in \mathbb{N}$. Now let $j \in \mathbb{N}$ satisfy $j > k(N)$ and let $n_j \in \mathbb{N}$ be such that $k(n_j) < j \leq k(n_j + 1)$. Then $n_j \geq N$ and

$$
\alpha^{-2} \frac{S_{k(n_j)}}{k(n_j)} \leq \frac{S_{k(n_j)}}{k(n_j+1)} \leq \frac{S_j}{j} \leq \frac{S_{k(n_j+1)}}{k(n_j)} \leq \alpha^2 \frac{S_{k(n_j+1)}}{k(n_j+1)}. \tag{3.66}
$$

Since $S_{k(n)}/k(n) \xrightarrow{\text{a.s.}} \mathbb{E}[X_1]$, letting $j \to \infty$ in (3.66) yields

$$
\alpha^{-2}\mathbb{E}[X_1] \leq \liminf_{n\to\infty} \frac{S_n}{n} \leq \limsup_{n\to\infty} \frac{S_n}{n} \leq \alpha^2 \mathbb{E}[X_1] \qquad \text{a.s.}
$$

Finally, since $\alpha \in (1,\infty)$ is arbitrary, we obtain

$$
\frac{\mathbb{E}[X_1]}{(1+j^{-2})^2} \leq \liminf_{n\to\infty} \frac{S_n}{n} \leq \limsup_{n\to\infty} \frac{S_n}{n} \leq (1+j^{-2})^2 \mathbb{E}[X_1] \quad \text{ for any } j \in \mathbb{N}, \text{ a.s.}
$$

and letting $j \to \infty$ shows $S_n/n \xrightarrow{\text{a.s.}} \mathbb{E}[X_1]$. ☐

Note that in the situation of Theorem 3.58 or Theorem 3.61, if $\mathbb{E}[X_1] = 0$ then $n^{-1}\sum_{i=1}^{n} X_i \xrightarrow{\text{a.s.}} 0$. In view of this, it is natural to expect $n^{-\alpha}\sum_{i=1}^{n} X_i \xrightarrow{\text{a.s.}} 0$ even for $\alpha < 1$. In fact, this is true for $\alpha \in (1/2, 1)$ under certain mild assumptions, and indeed the following much stronger result is valid.

Theorem 3.63 (Law of iterated logarithm)**.** *If* $\{X_n\}_{n=1}^{\infty} \subset \mathcal{L}^2(\mathbb{P})$ *is i.i.d.,* $\mathbb{E}[X_1] = 0$ *and* $\mathbb{E}[X_1^2] = 1$*, then the* law of iterated logarithm *holds, that is, almost surely,*

$$\limsup_{n\to\infty} \frac{\sum_{i=1}^{n} X_i}{\sqrt{2n\log\log n}} = 1 \quad \textit{and} \quad \liminf_{n\to\infty} \frac{\sum_{i=1}^{n} X_i}{\sqrt{2n\log\log n}} = -1. \tag{3.67}$$

The proof of Theorem 3.63 is lengthy and difficult and is not given in this lecture note. A proof of Theorem 3.63 is found in Dudley [1, Section 12.5], but the reader will have to learn quite a lot to follow the proof there.

3.6 Infinite Product of Probability Spaces

The purpose of this section is to give a proof of Theorem 3.40. Similarly to Theorem 3.28, this amounts to construct the *product of an infinite sequence of probability spaces*.

Definition 3.64 (Infinite product σ-algebras)**.** Let $(\Omega_n, \mathcal{F}_n)$ be a measurable space for each $n \in \mathbb{N}$ and set $\Omega := \prod_{n=1}^{\infty} \Omega_n$. We define $\prod_{n=1}^{\infty} \mathcal{F}_n \subset 2^{\Omega}$ and a σ-algebra $\bigotimes_{n=1}^{\infty} \mathcal{F}_n$ in Ω by

$$\prod_{n=1}^{\infty} \mathcal{F}_n := \left\{ A_1 \times \cdots \times A_n \times \prod_{i=n+1}^{\infty} \Omega_i \,\middle|\, n \in \mathbb{N},\ A_i \in \mathcal{F}_i \text{ for } i \in \{1, \ldots, n\} \right\}, \tag{3.68}$$

$$\bigotimes_{n=1}^{\infty} \mathcal{F}_n := \sigma_{\Omega}\left(\prod_{n=1}^{\infty} \mathcal{F}_n \right). \tag{3.69}$$

$\bigotimes_{n=1}^{\infty} \mathcal{F}_n$ is called the *product σ-algebra of* $\{\mathcal{F}_n\}_{n=1}^{\infty}$.

Theorem 3.65 (Infinite product probability measures)**.** *Let* $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ *be a probability space for each* $n \in \mathbb{N}$*. Then there exists a unique probability measure* $\mathbb{P}$ *on* $\left(\prod_{n=1}^{\infty} \Omega_n, \bigotimes_{n=1}^{\infty} \mathcal{F}_n\right)$ *such that for any* $k \in \mathbb{N}$ *and any* $A_n \in \mathcal{F}_n$*,* $n \in \{1, \ldots, k\}$*,*

$$\mathbb{P}\left[A_1 \times \cdots \times A_k \times \prod_{n=k+1}^{\infty} \Omega_n \right] = \mathbb{P}_1[A_1] \cdots \mathbb{P}_k[A_k]. \tag{3.70}$$

The probability measure $\mathbb{P}$ in Theorem 3.65 is denoted by $\prod_{n=1}^{\infty} \mathbb{P}_n$ and called the *product probability measure of* $\{\mathbb{P}_n\}_{n=1}^{\infty}$.

Proof of Theorem 3.40. Set $\Omega' := \prod_{n=1}^{\infty} S_n$, $\mathcal{F}' := \bigotimes_{n=1}^{\infty} \mathcal{B}_n$ and $\mathbb{P}' := \prod_{n=1}^{\infty} \mu_n$. For each $n \in \mathbb{N}$, define $X_n : \Omega' \to S_n$ by $X_n\big((x_k)_{k=1}^{\infty}\big) := x_n$ (the projection onto the n-th component). Then for $A_n \in \mathcal{B}_n$, $\{X_n \in A_n\} = S_1 \times \cdots \times S_{n-1} \times A_n \times \prod_{i=n+1}^{\infty} S_i$, which belongs to $\mathcal{F}'$ and has $\mathbb{P}'$-probability $\mu_1[S_1] \cdots \mu_{n-1}[S_{n-1}]\mu_n[A_n] = \mu_n[A_n]$.

Thus X_n is $\mathcal{F}'/\mathcal{B}_n$-measurable and $X_n \sim \mu_n$. Moreover, for $k \in \mathbb{N}$ and $A_n \in \mathcal{B}_n$, $n \in \{1, \dots, k\}$,

$$\mathbb{P}'[X_1 \in A_1, \dots, X_k \in A_k] = \mathbb{P}'\left[A_1 \times \cdots \times A_k \times \prod_{n=k+1}^{\infty} S_n\right]$$
$$= \mu_1(A_1) \cdots \mu_k(A_k) = \mathbb{P}'[X_1 \in A_1] \cdots \mathbb{P}'[X_k \in A_k].$$

Hence $\{X_n\}_{n=1}^k$ is independent, and so is $\{X_n\}_{n=1}^\infty$ since $k \in \mathbb{N}$ is arbitrary. □

Example 3.66 (Bernuolli measures)**.** Let $\Omega := \{0,1\}^{\mathbb{N}} = \{(\omega_n)_{n=1}^\infty \mid \omega_n \in \{0,1\}\}$. In Example 1.12, we have introduced a σ-algebra $\mathcal{F}$ in Ω given by

$$\mathcal{F} := \sigma\Big(\big\{A_n \times \{0,1\}^{\mathbb{N}\setminus\{1,\dots,n\}} \bigm| n \in \mathbb{N},\ A_n \subset \{0,1\}^n\big\}\Big). \tag{1.11}$$

which is nothing but the product σ-algebra $\bigotimes_{n=1}^\infty 2^{\{0,1\}}$ of countable copies of $2^{\{0,1\}}$. Let $p \in [0,1]$ and set $\mathbb{P}_p := \prod_{n=1}^\infty B(1,p)$, where $B(1,p)$ is as in Example 3.18 (note that $B(1,p)$ is nothing but the law of a Bernoulli random variable of probability p). Then

$$\mathbb{P}_p\big[\{(\omega_i)_{i=1}^n\} \times \{0,1\}^{\mathbb{N}\setminus\{1,\dots,n\}}\big] = \prod_{i=1}^n p^{\omega_i}(1-p)^{1-\omega_i} \tag{1.12}$$

for any $n \in \mathbb{N}$ and any $(\omega_i)_{i=1}^n \in \{0,1\}^n$, which shows the existence of the Bernoulli measure on $\{0,1\}^{\mathbb{N}}$ of probability p stated in Example 1.12. Its uniqueness follows by applying Theorem 2.5 to the π-system $\{\{\omega\} \times \{0,1\}^{\mathbb{N}\setminus\{1,\dots,n\}} \mid n \in \mathbb{N},\ \omega \in \{0,1\}^n\}$.

The rest of this section is devoted to the proof of Theorem 3.65. We need the following proposition.

Proposition 3.67. *Let $(\Omega_n, \mathcal{F}_n)$ be a measurable space for each $n \in \mathbb{N}$. Then for any $k \in \mathbb{N}$ and any $A \in \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_k$, $A \times \prod_{n=k+1}^\infty \Omega_n \in \bigotimes_{n=1}^\infty \mathcal{F}_n$.*

Proof. Let $k \in \mathbb{N}$ and define $\pi_k : \prod_{n=1}^\infty \Omega_n \to \Omega_1 \times \cdots \times \Omega_k$ by $\pi\big((\omega_n)_{n=1}^\infty\big) := (\omega_1, \dots, \omega_k)$ (the projection onto the first k components). Then for any $A_n \in \mathcal{F}_n$, $n \in \{1, \dots, k\}$,

$$\pi_k^{-1}\big(A_1 \times \cdots \times A_k\big) = \big(A_1 \times \cdots \times A_k\big) \times \prod_{n=k+1}^\infty \Omega_n \in \prod_{n=1}^\infty \mathcal{F}_n \subset \bigotimes_{n=1}^\infty \mathcal{F}_n,$$

that is, $\pi_k^{-1}(A) \in \bigotimes_{n=1}^\infty \mathcal{F}_n$ for any $A \in \mathcal{F}_1 \times \cdots \times \mathcal{F}_k$. Then since $\sigma_{\Omega_1 \times \cdots \times \Omega_k}(\mathcal{F}_1 \times \cdots \times \mathcal{F}_k) = \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_k$, Problem 1.15-(1) implies that π_k is $\bigotimes_{n=1}^\infty \mathcal{F}_n / \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_k$-measurable, i.e. $A \times \prod_{n=k+1}^\infty \Omega_n = \pi_k^{-1}(A) \in \bigotimes_{n=1}^\infty \mathcal{F}_n$ for any $A \in \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_k$. □

Proof of Theorem 3.65. Set $\Omega := \prod_{n=1}^\infty \Omega_n$ and $\mathcal{F} := \bigotimes_{n=1}^\infty \mathcal{F}_n$. Since $\prod_{n=1}^\infty \mathcal{F}_n$ is a π-system, $\Omega \in \prod_{n=1}^\infty \mathcal{F}_n$ and $\mathcal{F} = \sigma_\Omega\big(\prod_{n=1}^\infty \mathcal{F}_n\big)$, Theorem 2.5 implies that, if two probability measures $\mathbb{P}, \mathbb{P}'$ on $(\Omega, \mathcal{F})$ coincide on $\prod_{n=1}^\infty \mathcal{F}_n$ then $\mathbb{P} = \mathbb{P}'$, which shows the uniqueness of $\mathbb{P}$ asserted in Theorem 3.65.

To apply Theorem 2.7 to prove the existence of a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ with the desired properties, we define $\mathcal{A} \subset 2^\Omega$ and $\nu : \mathcal{A} \to [0,1]$ by

$$\mathcal{A} := \left\{A \times \prod_{n=k+1}^\infty \Omega_n \,\middle|\, k \in \mathbb{N},\ A \in \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_k\right\}, \tag{3.71}$$

$$\nu(C) := \mathbb{P}_1 \times \cdots \times \mathbb{P}_k[A] \quad \text{for } C = A \times \prod_{n=k+1}^{\infty} \Omega_n, k \in \mathbb{N}, A \in \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_k. \quad (3.72)$$

Note that the definition of $\nu(C)$ in (3.72) is independent of how C is written in the form $C = A \times \prod_{n=k+1}^{\infty} \Omega_n$, so that the definition (3.72) of $\nu : \mathcal{A} \to [0, 1]$ makes sense; indeed, suppose $C \in \mathcal{A}$ is written as $C = A \times \prod_{n=k+1}^{\infty} \Omega_n = B \times \prod_{n=\ell+1}^{\infty} \Omega_n$ for $k, \ell \in \mathbb{N}$, $A \in \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_k$ and $B \in \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_\ell$. We must show that $\mathbb{P}_1 \times \cdots \times \mathbb{P}_k[A] = \mathbb{P}_1 \times \cdots \times \mathbb{P}_\ell[B]$. Without loss of generality we may assume $k \leq \ell$. Taking the projection of C onto the first ℓ components shows that $B = A \times \prod_{k<n\leq\ell} \Omega_n$, and then by Proposition 2.26,

$$\begin{aligned}\mathbb{P}_1 \times \cdots \times \mathbb{P}_\ell[B] &= \big(\mathbb{P}_1 \times \cdots \times \mathbb{P}_k\big) \times \big(\mathbb{P}_{k+1} \times \cdots \times \mathbb{P}_\ell\big)\left[A \times \prod_{n=k+1}^{\ell} \Omega_n\right] \\ &= \mathbb{P}_1 \times \cdots \times \mathbb{P}_k[A] \cdot \mathbb{P}_{k+1} \times \cdots \times \mathbb{P}_\ell\left[\prod_{n=k+1}^{\ell} \Omega_n\right] = \mathbb{P}_1 \times \cdots \times \mathbb{P}_k[A].\end{aligned}$$

By virtue of Proposition 3.67, we have $\prod_{n=1}^{\infty} \mathcal{F}_n \subset \mathcal{A} \subset \mathcal{F}$ and hence $\sigma_\Omega(\mathcal{A}) = \mathcal{F}$. Clearly $\emptyset, \Omega \in \mathcal{A}$, $\nu(\emptyset) = \mathbb{P}_1[\emptyset] = 0$ and $\nu(\Omega) = \mathbb{P}_1[\Omega_1] = 1$. Let $C, D \in \mathcal{A}$ and write $C = A \times \prod_{n=k+1}^{\infty} \Omega_n$, $D = B \times \prod_{n=\ell+1}^{\infty} \Omega_n$ for some $k, \ell \in \mathbb{N}$, $A \in \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_k$ and $B \in \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_\ell$. Note that here k can be replaced by any $j \in \mathbb{N}$ with $j > k$, since for such j we have $C = A \times \prod_{n=k+1}^{\infty} \Omega_n = (A \times \Omega_{k+1} \times \cdots \times \Omega_j) \times \prod_{n=j+1}^{\infty} \Omega_n$ and $A \times \Omega_{k+1} \times \cdots \times \Omega_j \in \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_j$ by Proposition 2.22. Therefore by replacing both k and ℓ by $\max\{k, \ell\}$, we may assume without loss of generality that $k = \ell$. Then we see that $C \cup D = (A \cup B) \times \prod_{n=k+1}^{\infty} \Omega_n \in \mathcal{A}$, $C \cap D = (A \cap B) \times \prod_{n=k+1}^{\infty} \Omega_n \in \mathcal{A}$, $C \setminus D = (A \setminus B) \times \prod_{n=k+1}^{\infty} \Omega_n \in \mathcal{A}$ and

$$\nu(C \cap D) + \nu(C \setminus D) = \mathbb{P}_1 \times \cdots \times \mathbb{P}_k[A \cap B] + \mathbb{P}_1 \times \cdots \times \mathbb{P}_k[A \setminus B] = \mathbb{P}_1 \times \cdots \times \mathbb{P}_k[A] = \nu(C).$$

Thus we have verified that for any $C, D \in \mathcal{A}$,

$$C \cup D, C \cap D, C \setminus D \in \mathcal{A} \quad \text{and} \quad \nu(C) = \nu(C \cap D) + \nu(C \setminus D), \quad (3.73)$$

which implies that $\mathcal{A}$ is a π-system and that $\mathcal{A}$ and ν satisfy the condition (C3) of Theorem 2.7. The condition (C1) of Theorem 2.7 has been already verified. Hence if we prove that $\mathcal{A}$ and ν satisfy the condition (C2) of Theorem 2.7 as well, then Theorem 2.7 yields a probability measure $\mathbb{P}$ on $\sigma_\Omega(\mathcal{A}) = \mathcal{F}$ such that $\mathbb{P}|_{\mathcal{A}} = \nu$, which clearly satisfies (3.70) by the definition (3.72) of ν.

Thus it remains to verify the condition (C2) of Theorem 2.7 for $\mathcal{A}$ and ν. We prove that

$$\text{if } \{A_n\}_{n=1}^{\infty} \subset \mathcal{A}, A_n \supset A_{n+1} \text{ for any } n \in \mathbb{N} \text{ and } \bigcap_{n=1}^{\infty} A_n = \emptyset, \text{ then } \lim_{n\to\infty} \nu(A_n) = 0, \quad (3.74)$$

which easily implies the condition (C2) of Theorem 2.7 as follows: let $A \in \mathcal{A}$, $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ and $A \subset \bigcup_{n=1}^{\infty} A_n$. Set $B_k := A \setminus \bigcup_{n=1}^{k} A_n$ for $k \in \mathbb{N}$, so that $\{B_k\}_{k=1}^{\infty} \subset \mathcal{A}$ by (3.73), $B_k \supset B_{k+1}$ for $k \in \mathbb{N}$, and $\bigcap_{k=1}^{\infty} B_k = \emptyset$ by $A \subset \bigcup_{n=1}^{\infty} A_n$. Let $k \in \mathbb{N}$. (3.73) yields

$$\begin{aligned}\nu\left(\bigcup_{n=1}^{k+1} A_n\right) &= \nu\left(A_{k+1} \cap \bigcup_{n=1}^{k+1} A_n\right) + \nu\left(\left(\bigcup_{n=1}^{k+1} A_n\right) \setminus A_{k+1}\right) \\ &= \nu(A_{k+1}) + \nu\left(\left(\bigcup_{n=1}^{k} A_n\right) \setminus A_{k+1}\right) \leq \nu(A_{k+1}) + \nu\left(\bigcup_{n=1}^{k} A_n\right)\end{aligned}$$

and hence inductively $\nu\big(\bigcup_{n=1}^{k} A_n\big) \le \sum_{n=1}^{k} \nu(A_n)$. Then again by (3.73),

$$\nu(A) = \nu\left(A \cap \bigcup_{n=1}^{k} A_n\right) + \nu\left(A \setminus \bigcup_{n=1}^{k} A_n\right) \le \nu\left(\bigcup_{n=1}^{k} A_n\right) + \nu(B_k) \le \sum_{n=1}^{k} \nu(A_n) + \nu(B_k),$$

and letting $k \to \infty$ results in $\nu(A) \le \sum_{n=1}^{\infty} \nu(A_n)$ since $\lim_{k\to\infty} \nu(B_k) = 0$ by (3.74).

Thus it suffices to show (3.74). We closely follow [1, Proof of Theorem 8.2.2] for the argument below. We need some preparation. Let $k \in \mathbb{N}$ and $\Omega^{(k)} := \prod_{n=k+1}^{\infty} \Omega_n$. Define $\mathcal{A}^{(k)} \subset 2^{\Omega^{(k)}}$ and $\nu^{(k)} : \mathcal{A}^{(k)} \to [0,1]$ by the right-hand sides of (3.71) and (3.72), respectively, with $(\Omega_{k+n}, \mathcal{F}_{k+n}, \mathbb{P}_{k+n})$ in place of $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$. For each $C \subset \Omega$ and $\omega_n \in \Omega_n$, $n \in \{1, \dots, k\}$, we set

$$C^{(k)}(\omega_1, \dots, \omega_k) := \Big\{(\omega_n)_{n=k+1}^{\infty} \in \Omega^{(k)} \;\Big|\; (\omega_n)_{n=1}^{\infty} \in C\Big\}. \tag{3.75}$$

We claim that, if $C \in \mathcal{A}$, then $C^{(k)}(\omega_1, \dots, \omega_k) \in \mathcal{A}^{(k)}$ for any $(\omega_1, \dots, \omega_k) \in \Omega_1 \times \cdots \times \Omega_k$ and

$$\nu(C) = \int_{\Omega_1 \times \cdots \times \Omega_k} \nu^{(k)}\big(C^{(k)}(\omega_1, \dots, \omega_k)\big) \mathbb{P}_1 \times \cdots \times \mathbb{P}_k(d\omega_1 \dots d\omega_k). \tag{3.76}$$

Indeed, we can choose $\ell \in \mathbb{N}$ with $\ell > k$ and $A \in \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_\ell$ so that $C = A \times \prod_{n=\ell+1}^{\infty} \Omega_n$. Setting $A(\omega_1, \dots, \omega_k) := \big\{(\omega_{k+1}, \dots, \omega_\ell) \in \Omega_{k+1} \times \cdots \times \Omega_\ell \;\big|\; (\omega_1, \dots, \omega_\ell) \in A\big\}$, we have

$$C^{(k)}(\omega_1, \dots, \omega_k) = A(\omega_1, \dots, \omega_k) \times \prod_{n=\ell+1}^{\infty} \Omega_n \tag{3.77}$$

and $\mathbf{1}_{A(\omega_1,\dots,\omega_k)}(\omega_{k+1}, \dots, \omega_\ell) = \mathbf{1}_A(\omega_1, \dots, \omega_\ell)$, which is a $\mathcal{F}_{k+1} \otimes \cdots \otimes \mathcal{F}_\ell$-measurable function of $(\omega_{k+1}, \dots, \omega_\ell) \in \Omega_{k+1} \times \cdots \times \Omega_\ell$ by Propositions 2.22 and 2.28. It follows that $A(\omega_1, \dots, \omega_k) \in \mathcal{F}_{k+1} \otimes \cdots \otimes \mathcal{F}_\ell$ and hence that $C^{(k)}(\omega_1, \dots, \omega_k) \in \mathcal{A}^{(k)}$ in view of (3.77). Furthermore Fubini's theorem (Theorem 2.29-(1)) together with Corollary 2.26 yields

$$\begin{aligned}
&\nu(C) = \mathbb{P}_1 \times \cdots \times \mathbb{P}_\ell[A] \\
&= \int_{\Omega_1 \times \cdots \times \Omega_k} \left(\int_{\Omega_{k+1} \times \cdots \times \Omega_\ell} \mathbf{1}_{A(\omega_1,\dots,\omega_k)} d\big(\mathbb{P}_{k+1} \times \cdots \times \mathbb{P}_\ell\big)\right) \mathbb{P}_1 \times \cdots \times \mathbb{P}_k(d\omega_1 \dots d\omega_k) \\
&= \int_{\Omega_1 \times \cdots \times \Omega_k} \mathbb{P}_{k+1} \times \cdots \times \mathbb{P}_\ell\big[A(\omega_1, \dots, \omega_k)\big] \mathbb{P}_1 \times \cdots \times \mathbb{P}_k(d\omega_1 \dots d\omega_k) \\
&= \int_{\Omega_1 \times \cdots \times \Omega_k} \nu^{(k)}\big(C^{(k)}(\omega_1, \dots, \omega_k)\big) \mathbb{P}_1 \times \cdots \times \mathbb{P}_k(d\omega_1 \dots d\omega_k),
\end{aligned}$$

proving (3.76).

Now let $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ satisfy $A_n \supset A_{n+1}$ for any $n \in \mathbb{N}$, so that $\nu(A_n) \ge \nu(A_{n+1})$ by (3.73) and hence $\varepsilon := \lim_{n\to\infty} \nu(A_n) \in [0,1]$ exists. For the proof of (3.74), we deduce $\bigcap_{n=1}^{\infty} A_n \neq \emptyset$ by supposing $\varepsilon > 0$. For $n, k \in \mathbb{N}$ and $(\omega_1, \dots, \omega_k) \in \Omega_1 \times \cdots \times \Omega_k$, let $A_n^{(k)}(\omega_1, \dots, \omega_k)$ be as in (3.75) with $C := A_n$. We prove that

$$\text{there exists } \omega_1 \in \Omega_1 \text{ such that} \quad \nu^{(1)}\big(A_n^{(1)}(\omega_1)\big) \ge \frac{\varepsilon}{2} \quad \text{for any } n \in \mathbb{N}. \tag{3.78}$$

Let $n \in \mathbb{N}$ and $F_n := \{\omega_1 \in \Omega_1 \mid \nu^{(1)}(A_n^{(1)}(\omega_1)) \geq \varepsilon/2\}$. Since $A_n \supset A_{n+1}$, we have $A_n^{(1)}(\omega_1) \supset A_{n+1}^{(1)}(\omega_1)$ for each $\omega_1 \in \Omega_1$ and hence $F_n \supset F_{n+1}$. Moreover, (3.76) yields

$$\begin{aligned}\varepsilon \leq \nu(A_n) &= \int_{\Omega_1} \nu^{(1)}(A_n^{(1)}(\omega_1))\mathbb{P}_1(d\omega_1)\\ &= \int_{F_n} \nu^{(1)}(A_n^{(1)}(\omega_1))\mathbb{P}_1(d\omega_1) + \int_{\Omega_1\setminus F_n} \nu^{(1)}(A_n^{(1)}(\omega_1))\mathbb{P}_1(d\omega_1) \leq \mathbb{P}_1[F_n] + \frac{\varepsilon}{2}\end{aligned}$$

and hence $\mathbb{P}_1[F_n] \geq \varepsilon/2$. Letting $n \to \infty$, we obtain $\mathbb{P}_1\bigl[\bigcap_{n=1}^\infty F_n\bigr] \geq \varepsilon/2 > 0$. In particular, $\bigcap_{n=1}^\infty F_n \neq \emptyset$, and we can choose $\omega_1 \in \bigcap_{n=1}^\infty F_n$. This ω_1 satisfies the condition of (3.78).

Next let $k \in \mathbb{N}$ and suppose that we already have $(\omega_1, \ldots, \omega_k) \in \Omega_1 \times \cdots \times \Omega_k$ such that $\nu^{(k)}(A_n^{(k)}(\omega_1, \ldots, \omega_k)) \geq \varepsilon/2^k$ for any $n \in \mathbb{N}$. Then $\{A_n^{(k)}(\omega_1, \ldots, \omega_k)\}_{n=1}^\infty \subset \mathcal{A}^{(k)}$ by the claim in the previous paragraph, and $A_n^{(k)}(\omega_1, \ldots, \omega_k) \supset A_{n+1}^{(k)}(\omega_1, \ldots, \omega_k)$ for any $n \in \mathbb{N}$ by $A_n \supset A_{n+1}$. If we replace A_n with $A_n^{(k)}(\omega_1, \ldots, \omega_k)$ and $(\Omega_j, \mathcal{F}_j, \mathbb{P}_j)$ with $(\Omega_{j+k}, \mathcal{F}_{j+k}, \mathbb{P}_{j+k})$ for each $j \in \mathbb{N}$, then (3.78) is still applicable, and hence there exists $\omega_{k+1} \in \Omega_{k+1}$ such that for any $n \in \mathbb{N}$,

$$\begin{aligned}&\nu^{(k+1)}\Bigl(A_n^{(k+1)}(\omega_1, \ldots, \omega_{k+1})\Bigr)\\ &= \nu^{(k+1)}\Bigl(\Bigl\{(\omega_j)_{j=k+2}^\infty \in \Omega^{(k+1)} \Bigm| (\omega_j)_{j=k+1}^\infty \in A_n^{(k)}(\omega_1, \ldots, \omega_k)\Bigr\}\Bigr) \geq \frac{\varepsilon}{2^{k+1}}.\end{aligned}$$

Thus by induction in k, we conclude that there exists $(\omega_j)_{j=1}^\infty \in \Omega$ such that for any $n, k \in \mathbb{N}$, $\nu^{(k)}(A_n^{(k)}(\omega_1, \ldots, \omega_k)) \geq \varepsilon/2^k$ and hence $A_n^{(k)}(\omega_1, \ldots, \omega_k) \neq \emptyset$. Now let $n \in \mathbb{N}$ and choose $k \in \mathbb{N}$ and $B_n \in \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_k$ so that $A_n = B_n \times \prod_{j=k+1}^\infty \Omega_j$. Then $A_n^{(k)}(\omega_1, \ldots, \omega_k) \neq \emptyset$ implies $(\omega_1, \ldots, \omega_k) \in B_n$ and hence $(\omega_j)_{j=1}^\infty \in B_n \times \prod_{j=k+1}^\infty = A_n$. Since $n \in \mathbb{N}$ is arbitrary, $(\omega_j)_{j=1}^\infty \in \bigcap_{n=1}^\infty A_n$ and therefore $\bigcap_{n=1}^\infty A_n \neq \emptyset$, proving (3.74). □

Exercises

In the problems and the exercises below, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a probability space and all random variables are assumed to be defined on $(\Omega, \mathcal{F}, \mathbb{P})$.

Problem 3.1. Let $d \in \mathbb{N}$ and let $x \in \mathbb{R}^d$. Prove that the unit mass δ_x at x defined by $\delta_x(A) := \mathbf{1}_A(x)$, $A \in \mathcal{B}(\mathbb{R}^d)$ (recall Example 1.5-(2)), does not have a density.

Problem 3.2. Calculate $\mathbb{E}[X]$ and $\mathrm{var}(X)$ for a real random variable X with
(1) the binomial distribution $B(n, p)$, $n \in \mathbb{N}$, $p \in [0, 1]$.
(2) the Poisson distribution $\mathrm{Po}(\lambda)$, $\lambda \in (0, \infty)$.
(3) the geometric distribution $\mathrm{Geom}(\alpha)$, $\alpha \in [0, 1)$.

Problem 3.3. Calculate $\mathbb{E}[X]$ and $\mathrm{var}(X)$ for a real random variable X with
(1) the uniform distribution $\mathrm{Unif}(a, b)$, $a, b \in \mathbb{R}$, $a < b$.
(2) the exponential distribution $\mathrm{Exp}(\alpha)$, $\alpha \in (0, \infty)$.
(3) the gamma distribution $\mathrm{Gamma}(\alpha, \beta)$, $\alpha, \beta \in (0, \infty)$.

Problem 3.4. Let X be an exponential random variable. Prove that

$$\mathbb{P}[X > s + t \mid X > s] = \mathbb{P}[X > t] \qquad \text{for any } s, t \in [0, \infty) \tag{3.79}$$

(recall (1.60) for the definition of conditional probabilities).

(3.79) is known as the "memoryless property" of exponential random variables. Due to this property, exponential random variables are often used as *"random alarm clocks with no memory"*.

Exercise 3.5. Let X be a real random variable such that $\mathbb{P}[X > 0] > 0$, and suppose $\mathbb{P}[X > s + t \mid X > s] = \mathbb{P}[X > t]$ for any $s, t \in (0, \infty)$ with $\mathbb{P}[X > s] > 0$. Define $h : \mathbb{R} \to [0, 1]$ by $h(t) := \mathbb{P}[X > t]$. Prove the following statements:
(1) h is right-continuous and $h(s + t) = h(s)h(t)$ for any $s, t \in [0, \infty)$.
(2) There exists $\alpha \in (0, \infty)$ such that $h(t) = e^{-\alpha t}$ for any $t \in [0, \infty)$.
(3) X is an exponential random variable of parameter α.

Problem 3.6. Let X be a normal random variable with mean m and variance $v \in (0, \infty)$. Prove that the real random variable $Y := e^X$ has a density ρ_Y given by

$$\rho_Y(x) = \frac{1}{x\sqrt{2\pi v}} \exp\left(-\frac{(\log x - m)^2}{2v}\right)\mathbf{1}_{(0,\infty)}(x). \tag{3.80}$$

The law of Y is called the *lognormal distribution with parameters m, v.*

Problem 3.7. Let X be a normal random variable with mean 0 and variance 1. Prove that the real random variable $Z := X^2$ has a density ρ_Z given by

$$\rho_Z(x) = \frac{1}{\sqrt{2\pi x}} e^{-x/2}\mathbf{1}_{(0,\infty)}(x). \tag{3.81}$$

The law of Z is called the *chi square distribution with one degree of freedom* and denoted as χ_1^2. (In fact, (3.81) and (3.21) easily imply that $\chi_1^2 = \text{Gamma}(1/2, 1/2)$.)

Problem 3.8. Let $m \in \mathbb{R}$, $\alpha \in (0, \infty)$ and let X be a Cauchy random variable with parameters m, α. Prove that X does not admit the mean, i.e. $\mathbb{E}[X^+] = \mathbb{E}[X^-] = \infty$.

Problem 3.9. Let X, Y be independent geometric random variables of parameter $1/2$. Let $k \in \mathbb{N} \cup \{0\}$. Calculate the following probabilities:

$$\text{(i)} \quad \mathbb{P}[\min\{X, Y\} \le k] \quad \text{(ii)} \quad \mathbb{P}[X < Y] \quad \text{(iii)} \quad \mathbb{P}[X = Y]$$

Problem 3.10. Let X be a real random variable with $X \sim \text{Unif}(0, \pi/2)$ and set $Y := \sin X$. Find the following quantities:

$$\text{(i)} \quad \text{a density of } Y \quad \text{(ii)} \quad \mathbb{E}[Y] \quad \text{(iii)} \quad \text{var}(Y)$$

Exercise 3.11. Define $\rho : \mathbb{R}^2 \to [0, \infty)$ by

$$\rho(x, y) := \frac{1}{2}(x + y)e^{-x-y}\mathbf{1}_{(0,\infty)^2}(x, y). \tag{3.82}$$

(1) Prove that $\int_{\mathbb{R}^2} \rho(z)dz = 1$, so that $\mu := \rho \cdot \mathrm{m}_2$ is a probability law on $\mathbb{R}^2$.
(2) Let X, Y be real random variables with $(X, Y) \sim \mu$. Find the following quantities:

(i)	a density of X	(ii)	$\mathbb{E}[X]$	(iii)	$\mathrm{var}(X)$
(iv)	$\mathrm{cov}(X, Y)$	(v)	a density of $X + Y$		

((iv): $\mathbb{E}[XY] = \int_{\mathbb{R}^2} xy\rho(x, y)d\mathrm{m}_2(x, y)$ by Theorem 3.14. (v): find a density of $(X + Y, X - Y)$ in the same way as Example 3.30 and then use Proposition 3.17.)

In Exercise 3.11-(2), you will see that $\mathrm{cov}(X, Y) \neq 0$, which together with (3.31) in Proposition 3.32 implies that $\{X, Y\}$ is not independent.

Problem 3.12. Let X be a real random variable with $X \sim N(m, v)$. Let $\alpha \in \mathbb{R}$. Prove that $\alpha X \sim N(\alpha m, \alpha^2 v)$. (Note that a special treatment is required if $v = 0$ or $\alpha = 0$.)

Problem 3.13. Let X, Y be independent real random variables with $X \sim N(m_1, v_1)$ and $Y \sim N(m_2, v_2)$. Prove that $X + Y \sim N(m_1 + m_2, v_1 + v_2)$. (Use Propositions 3.36 and 3.38. Note again that a special treatment is required if $v_1 = 0$ or $v_2 = 0$.)

Exercise 3.14. Let $n \in \mathbb{N}$, and let $\{X_i\}_{i=1}^n$ be independent real random variables with $X_i \sim N(m_i, v_i)$ for any $i \in \{1, \dots, n\}$. Set $X := \sum_{i=1}^n X_i$, $m := \sum_{i=1}^n m_i$ and $v := \sum_{i=1}^n v_i$. Prove that $X \sim N(m, v)$. (Induction in n. Use Proposition 3.31.)

Problem 3.15. Let $\{X_n\}_{n=1}^\infty$ be real random variables. Prove the following statements:
(1) $\{\lim_{n\to\infty} X_n$ exists in $\mathbb{R}\}$ is a tail event for $\{X_n\}_{n=1}^\infty$.
(2) If $\{a_n\}_{n=1}^\infty \subset \mathbb{R}$ satisfies $\lim_{n\to\infty} a_n = 0$, then $\limsup_{n\to\infty} a_n \sum_{i=1}^n X_i$ and $\liminf_{n\to\infty} a_n \sum_{i=1}^n X_i$ are $\sigma_\infty(\{X_n\}_{n=1}^\infty)$-measurable. (Imitate Example 3.48.)

Exercise 3.16. Let $d \in \mathbb{N}$, and let $\{X_n\}_{n=1}^\infty$ be d-dimensional random variables. Prove that $\{\lim_{n\to\infty} X_n \text{ exists in } \mathbb{R}^d\}$ is a tail event for $\{X_n\}_{n=1}^\infty$.

Problem 3.17. Let X, Y be independent real random variables with $X \sim \mathrm{Po}(\lambda_1)$ and $Y \sim \mathrm{Po}(\lambda_2)$. Prove that $X + Y \sim \mathrm{Po}(\lambda_1 + \lambda_2)$.

Exercise 3.18. Let $n \in \mathbb{N}$, and let $\{X_i\}_{i=1}^n$ be independent real random variables with $X_i \sim \mathrm{Po}(\lambda_i)$ for any $i \in \{1, \dots, n\}$. Set $X := \sum_{i=1}^n X_i$ and $\lambda := \sum_{i=1}^n \lambda_i$. Prove that $X \sim \mathrm{Po}(\lambda)$. (Induction in n. Similarly to Exercise 3.14, use Proposition 3.31.)

Problem 3.19. Let $a, b \in [-\infty, \infty]$, $a < b$ and let μ be a law on $\mathbb{R}$. Prove that, if the distribution function F_μ of μ is C^1 on (a, b), $\lim_{x\uparrow b} F_\mu(x) = 1$ and $\lim_{x\downarrow a} F_\mu(x) = 0$, then $\mu(dx) = F'_\mu(x)\mathbf{1}_{(a,b)}(x)dx$. (Show $\int_{-\infty}^x F'_\mu(y)\mathbf{1}_{(a,b)}(y)dy = F_\mu(x)$, $x \in \mathbb{R}$.)

Problem 3.20. Let X, Y be independent real random variables with $X \sim \mathrm{Exp}(1)$ and $Y \sim \mathrm{Exp}(1)$. Find a density of the random variable $Z := X/Y$. (Calculate $F_Z(t) := \mathbb{P}[Z \leq t]$, differentiate F_Z and use Problem 3.19.)

Problem 3.21. Let X, Y be independent real random variables with $X \sim \mathrm{Unif}(0, 1)$ and $Y \sim \mathrm{Unif}(0, 1)$. Find the following quantities:

(i)	a density of $X + Y$	(ii)	a density of XY	(iii)	a density of X^2
(iv)	$\mathbb{E}[\max\{X, Y\}]$	(v)	$\mathbb{E}[\min\{X, Y\}]$	(vi)	$\mathbb{E}[\max\{X, Y\} \cdot \min\{X, Y\}]$

((i): Use Propositions 3.36 and 3.38. (ii), (iii): Use Problem 3.19. (iv), (v), (vi): Apply Theorem 3.10 to the random variable (X, Y) and use the independence of X, Y.)

Problem 3.22. Let $X, Y, \{X_n\}_{n=1}^\infty, \{Y_n\}_{n=1}^\infty$ be real random variables such that

$$X_n \xrightarrow{\mathrm{P}} X \quad \text{and} \quad Y_n \xrightarrow{\mathrm{P}} Y. \tag{3.83}$$

(1) Prove that $(X_n, Y_n) \xrightarrow{\mathrm{P}} (X, Y)$.
(2) Prove that $X_n + Y_n \xrightarrow{\mathrm{P}} X + Y$ and that $X_n Y_n \xrightarrow{\mathrm{P}} XY$. (By (1), Corollary 3.53-(2) applies to (X, Y) and $\{(X_n, Y_n)\}_{n=1}^\infty$.)

Problem 3.23. Let $X, Y, \{X_n\}_{n=1}^\infty, \{Y_n\}_{n=1}^\infty$ be real random variables such that

$$\frac{1}{n}\sum_{k=1}^n X_k \xrightarrow{\mathrm{P}} X \quad \text{and} \quad \frac{1}{n}\sum_{k=1}^n Y_k \xrightarrow{\mathrm{P}} Y. \tag{3.84}$$

Define $\{Z_n\}_{n=1}^\infty$ by $Z_{2n-1} := X_n$ and $Z_{2n} := Y_n$. Prove that

$$\frac{1}{n}\sum_{k=1}^n Z_k \xrightarrow{\mathrm{P}} \frac{X+Y}{2}. \tag{3.85}$$

(Use Problem 3.22-(2).)

Exercise 3.24. Let $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ and let $\{X_n\}_{n=1}^\infty$ be d-dimensional random variables with $X_n \xrightarrow{\mathcal{L}} x$. Prove that $X_n \xrightarrow{\mathrm{P}} x$. ($\mathbb{P}[|X_n - x| \geq \varepsilon] = \mathbb{P}[\min\{2\varepsilon, |X_n - x|\} \geq \varepsilon]$ for $\varepsilon \in (0, \infty)$. Apply Chebyshev's inequality (Problem 1.18) with $\varphi(x) = x$ and then use $X_n \xrightarrow{\mathcal{L}} x$, noting that $\mathbb{R}^d \ni y \mapsto \min\{2\varepsilon, |y - x|\}$ is bounded continuous.)

Exercise 3.25. Let $X, \{X_n\}_{n=1}^\infty$ be real random variable with $X_n \xrightarrow{\mathrm{P}} X$ and suppose $X \neq 0$ a.s. Prove that $X_n^{-1}\mathbf{1}_{\{X_n \neq 0\}} \xrightarrow{\mathrm{P}} X^{-1}$. (Use Theorem 3.52, similarly to the proof of Corollary 3.53-(2).)

Problem 3.26. Let $(S, \mathcal{B})$ be a measurable space and let $\{X_n\}_{n=1}^\infty$ be i.i.d. $(S, \mathcal{B})$-valued random variables. Let $(E, \mathcal{E})$ be a measurable space and let $f : S \to E$ be $\mathcal{B}/\mathcal{E}$-measurable. Prove that $\{f(X_n)\}_{n=1}^\infty$ is i.i.d. $(E, \mathcal{E})$-valued random variables.

Problem 3.27. Let $\{X_n\}_{n=1}^\infty \subset \mathcal{L}^1(\mathbb{P})$ be i.i.d. and set $Y_n := e^{X_n}$ for each $n \in \mathbb{N}$. Prove that

$$(Y_1 \cdots Y_n)^{1/n} \xrightarrow{\text{a.s.}} \exp\bigl(\mathbb{E}[X_1]\bigr). \tag{3.86}$$

($(Y_1 \cdots Y_n)^{1/n} = \exp\bigl(\frac{1}{n}\sum_{k=1}^n X_k\bigr)$, to which Theorem 3.61 applies.)

Problem 3.28. Let $N \in \mathbb{N}$ and let $\{X_n\}_{n=1}^\infty \subset \mathcal{L}^N(\mathbb{P})$ be i.i.d. Prove that

$$\frac{1}{n}\sum_{k=1}^n X_k^N \xrightarrow{\text{a.s.}} \mathbb{E}[X_1^N]. \tag{3.87}$$

(Apply Theorem 3.61 to $\{X_n^N\}_{n=1}^\infty$, which is i.i.d. by Problem 3.26.)

Problem 3.29. Let $m \in \mathbb{R}$, $v \in (0, \infty)$ and let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. with $X_1 \sim N(m, v)$. Prove that

$$\frac{\sum_{k=1}^{n} X_k}{\sum_{k=1}^{n} X_k^2} \xrightarrow{\text{a.s.}} \frac{m}{m^2 + v}. \tag{3.88}$$

(Divide both the numerator and the denominator by n and apply Theorem 3.61.)

Problem 3.30. Let $\{X_n\}_{n=1}^{\infty} \subset \mathcal{L}^2(\mathbb{P})$ be i.i.d. Prove that

$$\frac{1}{n} \sum_{k=1}^{n} \big(X_k - \mathbb{E}[X_1]\big)^2 \xrightarrow{\text{a.s.}} \operatorname{var}(X_1). \tag{3.89}$$

Chapter 4

Convergence of Laws and Central Limit Theorem

In Definition 3.49, we have defined the notion of convergence *in law* (or convergence *in distribution*) of random variables, along with various other forms of convergence of random variables. The aim of this section is to develop further theory of convergence in law of random variables. Our principal goal is to state and prove the *central limit theorem*. Its precise statement is first described in Section 4.1 in the case of i.i.d. real random variables and then Sections 4.1 and 4.2 are devoted to preparing important tools for the proof of the central limit theorem. The key notions of this chapter are:

- *convergence of laws* on $\mathbb{R}^d$ (Section 4.1)
- *characteristic functions* of laws on $\mathbb{R}^d$ (Section 4.2)

Using the theories developed in Sections 4.1 and 4.2, in Section 4.3 we state and prove the central limit theorem for i.i.d. d-dimensional random variables, which involves *d-dimensional normal distributions*. Some details on d-dimensional normal distributions are also presented in Section 4.3.

Throughout this chapter, we fix $d \in \mathbb{N}$ and a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and random variables are always assumed to be defined on $(\Omega, \mathcal{F}, \mathbb{P})$ unless otherwise stated.

4.1 Convergence of Laws

We start with some notations which will be frequently used in this chapter. Recall that "law" is a synonym for "Borel probability measure" and that a function $f : S \to \mathbb{C}$ on a set S is called *bounded* if and only if $\sup_{x \in S} |f(x)| < \infty$.

Definition 4.1. For $S \subset \mathbb{R}^d$, we define

$$\mathcal{P}(S) := \{\mu \mid \mu \text{ is a law on } S\}, \tag{4.1}$$

$$C_b(S) := \{f \mid f : S \to \mathbb{R},\ f \text{ is bounded and continuous}\}. \tag{4.2}$$

The following definition is at the center of consideration in this chapter.

Definition 4.2 (Convergence of laws)**.** Let $S \subset \mathbb{R}^d$, $\mu \in \mathcal{P}(S)$ and $\{\mu_n\}_{n=1}^{\infty} \subset \mathcal{P}(S)$. We say that $\{\mu_n\}_{n=1}^{\infty}$ *converges weakly to* μ, or simply $\{\mu_n\}_{n=1}^{\infty}$ *converges to* μ, and write $\mu_n \xrightarrow{\mathcal{L}} \mu$, if and only if

$$\lim_{n\to\infty} \int_S f d\mu_n = \int_S f d\mu \qquad \text{for any } f \in C_b(S). \tag{4.3}$$

This convergence is called *weak convergence of laws* or simply *convergence of laws*.

According to Definition 3.49-(3), for d-dimensional random variables $X, \{X_n\}_{n=1}^{\infty}$,

$$X_n \xrightarrow{\mathcal{L}} X \qquad \text{if and only if} \qquad \mathcal{L}(X_n) \xrightarrow{\mathcal{L}} \mathcal{L}(X); \tag{4.4}$$

recall that $\mathcal{L}(Y)$ denotes the law of a random variable Y.

In the situation of Definition 4.2, one could consider other ways of convergence of laws, e.g.

$$\lim_{n\to\infty} \mu_n(A) = \mu(A) \qquad \text{for any } A \in \mathcal{B}(S). \tag{4.5}$$

The convergence in the sense of (4.5), however, is actually a stronger requirement than $\mu_n \xrightarrow{\mathcal{L}} \mu$, which will be verified in Theorem 4.10. The following example illustrates the situation.

Example 4.3. For each $x \in \mathbb{R}^d$ let δ_x denote the unit mass at x given by $\delta_x(A) := \mathbf{1}_A(x)$, $A \in \mathcal{B}(\mathbb{R}^d)$ (recall Example 1.5-(2)). Let $x \in \mathbb{R}^d$ and $\{x_n\}_{n=1}^{\infty} \subset \mathbb{R}^d$. Then $\lim_{n\to\infty} x_n = x$ if and only if $\delta_{x_n} \xrightarrow{\mathcal{L}} \delta_x$; indeed, if $\lim_{n\to\infty} x_n = x$ then for any $f \in C_b(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} f(y)\delta_{x_n}(dy) = f(x_n) \xrightarrow{n\to\infty} f(x) = \int_{\mathbb{R}^d} f(y)\delta_x(dy)$$

and hence $\delta_{x_n} \xrightarrow{\mathcal{L}} \delta_x$, and conversely if $\delta_{x_n} \xrightarrow{\mathcal{L}} \delta_x$ then $\lim_{n\to\infty} x_n = x$ since

$$\min\{1, |x_n - x|\} = \int_{\mathbb{R}^d} \min\{1, |y-x|\}\delta_{x_n}(dy) \xrightarrow{n\to\infty} \int_{\mathbb{R}^d} \min\{1, |y-x|\}\delta_x(dy) = 0.$$

On the other hand, $\{\delta_{x_n}\}_{n=1}^{\infty}$ converges to δ_x in the sense of (4.5), i.e.

$$\lim_{n\to\infty} \delta_{x_n}(A) = \delta_x(A) \qquad \text{for any } A \in \mathcal{B}(\mathbb{R}^d) \tag{4.6}$$

if and only if there exists $k \in \mathbb{N}$ *such that* $x_n = x$ *for any* $n \in \mathbb{N}$ *with* $n \geq k$. Indeed, "if" part is clear, and conversely if (4.6) holds, then $\lim_{n\to\infty} \delta_{x_n}(\{x\}) = 1$, hence there exists $k \in \mathbb{N}$ such that $\delta_{x_n}(\{x\}) > 0$ for any $n \geq k$, and thus $x_n = x$ for any $n \geq k$.

The principal aim of this chapter is to prove the following *central limit theorem*, which occupies a central position in modern probability theory as its name suggests.

Theorem 4.4 (Central limit theorem)**.** *Let* $\{X_n\}_{n=1}^{\infty} \subset \mathcal{L}^2(\mathbb{P})$ *be i.i.d. Set* $m := \mathbb{E}[X_1]$, $v := \mathrm{var}(X_1)$ *and* $S_n := \sum_{k=1}^{n} X_k$ *for each* $n \in \mathbb{N}$.
(1) *It holds that*

$$\mathcal{L}\left(\frac{S_n - nm}{\sqrt{n}}\right) \xrightarrow{\mathcal{L}} N(0, v). \tag{4.7}$$

(2) *If* $v > 0$, *then for any* $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{S_n - nm}{\sqrt{n}} \le x\right] = \lim_{n\to\infty} \mathbb{P}\left[\frac{S_n - nm}{\sqrt{n}} < x\right] = \int_{-\infty}^{x} \frac{e^{-y^2/(2v)}}{\sqrt{2\pi v}} dy. \tag{4.8}$$

Note that in the situation of Theorem 4.4, if $v > 0$ then by Theorem 3.63 we have almost surely

$$\limsup_{n\to\infty} \frac{S_n - nm}{\sqrt{2n \log\log n}} = \sqrt{v} \quad \text{and} \quad \liminf_{n\to\infty} \frac{S_n - nm}{\sqrt{2n \log\log n}} = -\sqrt{v}. \tag{4.9}$$

Thus roughly speaking, almost surely $(S_n - nm)/\sqrt{n}$ oscillates between $\sqrt{2v \log\log n}$ and $-\sqrt{2v \log\log n}$ as $n \to \infty$, and the amplitude $\sqrt{2v \log\log n}$ of the oscillation grows only very slowly. Then one might expect $(S_n - nm)/\sqrt{n}$ to converge in some sense as $n \to \infty$. Theorem 4.4 asserts that $(S_n - nm)/\sqrt{n}$ does converge in law and that the limit distribution is *always* the normal distribution $N(0, v)$, as long as the i.i.d. real random variables $\{X_n\}_{n=1}^{\infty}$ have finite variance v.[1] In this sense, the normal distributions can be considered as the most fundamental probability laws on $\mathbb{R}$.

Now we present basic facts concerning convergence of laws.

Lemma 4.5. *Let* F *be a non-empty closed subset of* $\mathbb{R}^d$, *and for each* $n \in \mathbb{N}$ *define* $f_n : \mathbb{R}^d \to [0, 1]$ *by*

$$f_n(x) := \min\{1, n\mathrm{dist}(x, F)\}, \quad \text{where} \quad \mathrm{dist}(x, F) := \inf_{y\in F} |x - y|. \tag{4.10}$$

Then $\{f_n\}_{n=1}^{\infty} \subset C_b(\mathbb{R}^d)$ *and* $\lim_{n\to\infty} f_n(x) = \mathbf{1}_{\mathbb{R}^d\setminus F}(x)$ *for any* $x \in \mathbb{R}^d$.

Proof. $\mathrm{dist}(x, F) = 0$ for $x \in F$, and if $x \in \mathbb{R}^d \setminus F$ then $\mathrm{dist}(x, F) > 0$ since $\mathbb{R}^d \setminus F$ is open in $\mathbb{R}^d$ and hence $B_d(x, \varepsilon) \subset \mathbb{R}^d \setminus F$ for some $\varepsilon \in (0, \infty)$. (Recall that $B_d(x, \varepsilon) = \{y \in \mathbb{R}^d \mid |y - x| < \varepsilon\}$.) Therefore $\lim_{n\to\infty} f_n(x) = \mathbf{1}_{\mathbb{R}^d\setminus F}(x)$ for any $x \in \mathbb{R}^d$. Moreover, we have

$$|\mathrm{dist}(x, F) - \mathrm{dist}(y, F)| \le |x - y| \qquad \text{for any } x, y \in \mathbb{R}^d \tag{4.11}$$

and hence $\{f_n\}_{n=1}^{\infty} \subset C_b(\mathbb{R}^d)$. Indeed, for $x, y \in \mathbb{R}^d$ and $z \in F$, $|y - z| \ge |x - z| - |x - y| \ge \mathrm{dist}(x, F) - |x - y|$ and taking the infimum over $z \in F$ yields $\mathrm{dist}(y, F) \ge \mathrm{dist}(x, F) - |x - y|$, i.e. $\mathrm{dist}(x, F) - \mathrm{dist}(y, F) \le |x - y|$. Interchanging the role of x and y shows $\mathrm{dist}(x, F) - \mathrm{dist}(y, F) \ge -|x - y|$, and (4.11) follows. □

[1]Honestly, the explanation of the appearance of $\sqrt{n}$ here is a little cheating, since Theorem 4.4 can be proved much more elementarily, and theoretically it should come earlier, than Theorem 3.63.

Proposition 4.6. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$. Suppose $\int_{\mathbb{R}^d} f d\mu = \int_{\mathbb{R}^d} f d\nu$ for any continuous function $f : \mathbb{R}^d \to \mathbb{R}$ satisfying $f|_{\mathbb{R}^d \setminus [-N,N]^d} = 0$ for some $N \in \mathbb{N}$. Then $\mu = \nu$.*

Note that *if $f : \mathbb{R}^d \to \mathbb{R}$ is as in Proposition* 4.6 *then* $f \in C_b(\mathbb{R}^d)$, since $\sup_{x \in \mathbb{R}^d} |f(x)| = \sup_{x \in [-N,N]^d} |f(x)| < \infty$ by the compactness of $[-N, N]^d$.

Proof. Let U be an open subset of $\mathbb{R}^d$. Let $N \in \mathbb{N}$ and set $F := \mathbb{R}^d \setminus \bigl(U \cap (-N, N)^d\bigr)$. Let $\{f_n\}_{n=1}^{\infty} \subset C_b(\mathbb{R}^d)$ be as in Lemma 4.5. Since $0 \le f_n \le f_{n+1} \le \mathbf{1}_{U \cap (-N,N)^d}$ on $\mathbb{R}^d$ for any $n \in \mathbb{N}$, the monotone convergence theorem together with the assumption on μ and ν yields

$$\mu\bigl(U \cap (-N, N)^d\bigr) = \lim_{n\to\infty} \int_{\mathbb{R}^d} f_n d\mu = \lim_{n\to\infty} \int_{\mathbb{R}^d} f_n d\nu = \nu\bigl(U \cap (-N, N)^d\bigr)$$

and letting $N \to \infty$ results in $\mu(U) = \nu(U)$. Thus $\mu(U) = \nu(U)$ for any open subset U of $\mathbb{R}^d$, and hence $\mu = \nu$ by Theorem 2.5 since $\{U \subset \mathbb{R}^d \mid U \text{ is open in } \mathbb{R}^d\}$ is a π-system and $\mathcal{B}(\mathbb{R}^d) = \sigma\bigl(\{U \subset \mathbb{R}^d \mid U \text{ is open in } \mathbb{R}^d\}\bigr)$. □

Corollary 4.7. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and $\{\mu_n\}_{n=1}^{\infty} \subset \mathcal{P}(\mathbb{R}^d)$. If $\mu_n \xrightarrow{\mathcal{L}} \mu$ and $\mu_n \xrightarrow{\mathcal{L}} \nu$, then $\mu = \nu$.*

Proof. Since $\mu_n \xrightarrow{\mathcal{L}} \mu$ and $\mu_n \xrightarrow{\mathcal{L}} \nu$, for any $f \in C_b(\mathbb{R}^d)$ we have

$$\int_{\mathbb{R}^d} f d\mu = \lim_{n\to\infty} \int_{\mathbb{R}^d} f d\mu_n = \int_{\mathbb{R}^d} f d\nu$$

and hence $\mu = \nu$ by Proposition 4.6. □

Proposition 4.8. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\{\mu_n\}_{n=1}^{\infty} \subset \mathcal{P}(\mathbb{R}^d)$. Suppose that for any strictly increasing sequence $\{n(k)\}_{k=1}^{\infty} \subset \mathbb{N}$ there exists a further strictly increasing sequence $\{k(\ell)\}_{\ell=1}^{\infty} \subset \mathbb{N}$ such that $\mu_{n(k(\ell))} \xrightarrow{\mathcal{L}} \mu$. Then $\mu_n \xrightarrow{\mathcal{L}} \mu$.*

Proof. Suppose $\mu_n \xrightarrow{\mathcal{L}} \mu$ does not hold. Then for some $f \in C_b(\mathbb{R}^d)$, $\int_{\mathbb{R}^d} f d\mu_n$ does not converge to $\int_{\mathbb{R}^d} f d\mu$ and hence there exist $\varepsilon \in (0, \infty)$ and a strictly increasing sequence $\{n(k)\}_{k=1}^{\infty} \subset \mathbb{N}$ such that $\bigl|\int_{\mathbb{R}^d} f d\mu_{n(k)} - \int_{\mathbb{R}^d} f d\mu\bigr| \ge \varepsilon$ for any $k \in \mathbb{N}$. Then for any strictly increasing sequence $\{k(\ell)\}_{\ell=1}^{\infty} \subset \mathbb{N}$, $\int_{\mathbb{R}^d} f d\mu_{n(k(\ell))}$ cannot converge to $\int_{\mathbb{R}^d} f d\mu$ and hence $\mu_{n(k(\ell))} \xrightarrow{\mathcal{L}} \mu$ does not hold, contradicting the assumption. □

Proposition 4.9. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ have a density ρ and let $\mu_n \in \mathcal{P}(\mathbb{R}^d)$ have a density ρ_n for each $n \in \mathbb{N}$. If $\lim_{n\to\infty} \rho_n(x) = \rho(x)$ for m_d-a.e. $x \in \mathbb{R}^d$, then $\mu_n \xrightarrow{\mathcal{L}} \mu$.*

Proof. Let $f : \mathbb{R}^d \to \mathbb{R}$ be bounded Borel measurable and set $M := \sup_{x \in \mathbb{R}^d} |f(x)|$. Then since $f + M \ge 0$, Fatou's lemma (Theorem 1.27) yields

$$\liminf_{n\to\infty} \int_{\mathbb{R}^d} f d\mu_n + M = \liminf_{n\to\infty} \int_{\mathbb{R}^d} (f + M)\rho_n dx \ge \int_{\mathbb{R}^d} \liminf_{n\to\infty}\bigl((f + M)\rho_n\bigr) dx$$

$$= \int_{\mathbb{R}^d} (f + M)\rho dx = \int_{\mathbb{R}^d} f d\mu + M.$$

Thus $\int_{\mathbb{R}^d} f d\mu \leq \liminf_{n\to\infty} \int_{\mathbb{R}^d} f d\mu_n$, and by replacing f with $-f$ we also obtain $\limsup_{n\to\infty} \mathbb{R}^d f d\mu_n \leq \int_{\mathbb{R}^d} f d\mu$. Therefore $\int_{\mathbb{R}^d} f d\mu = \limsup_{n\to\infty} \int_{\mathbb{R}^d} f d\mu_n = \liminf_{n\to\infty} \int_{\mathbb{R}^d} f d\mu_n$ and hence $\lim_{n\to\infty} \int_{\mathbb{R}^d} f d\mu_n = \int_{\mathbb{R}^d} f d\mu$. This is in particular true for any $f \in C_b(\mathbb{R}^d)$, that is, $\mu_n \xrightarrow{\mathcal{L}} \mu$. □

Recall that, for $\mu \in \mathcal{P}(\mathbb{R}^d)$, F_μ denotes its distribution function given in Definition 2.17 (and in Definition 2.15 for $d = 1$).

Theorem 4.10. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\{\mu_n\}_{n=1}^\infty \subset \mathcal{P}(\mathbb{R}^d)$. Then the following conditions are equivalent:*
(1) $\mu_n \xrightarrow{\mathcal{L}} \mu$.
(2) $\liminf_{n\to\infty} \mu_n(U) \geq \mu(U)$ *for any open subset U of $\mathbb{R}^d$.*
(3) $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$ *for any closed subset F of $\mathbb{R}^d$.*
(4) $\lim_{n\to\infty} F_{\mu_n}(x) = F_\mu(x)$ *for any $x \in \mathbb{R}^d$ at which F_μ is continuous.*
(5) $\lim_{n\to\infty} \int_{\mathbb{R}^d} f d\mu_n = \int_{\mathbb{R}^d} f d\mu$ *for any continuous function $f : \mathbb{R}^d \to \mathbb{R}$ such that $f|_{\mathbb{R}^d\setminus[-N,N]^d} = 0$ for some $N \in \mathbb{N}$.*

Proof. (1) ⇒ (2): Let U be an open subset of $\mathbb{R}^d$ and set $F := \mathbb{R}^d \setminus U$. The assertion is clear if $U = \mathbb{R}^d$, and hence we may assume $U \neq \mathbb{R}^d$. Let $\{f_n\}_{n=1}^\infty \subset C_b(\mathbb{R}^d)$ be as in Lemma 4.5. Then for $k \in \mathbb{N}$, $f_k \leq f_{k+1} \leq \mathbf{1}_U$ on $\mathbb{R}^d$ and hence

$$\liminf_{n\to\infty} \mu_n(U) \geq \liminf_{n\to\infty} \int_{\mathbb{R}^d} f_k d\mu_n = \lim_{n\to\infty} \int_{\mathbb{R}^d} f_k d\mu_n = \int_{\mathbb{R}^d} f_k d\mu \xrightarrow{k\to\infty} \mu(U),$$

where we used the monotone convergence theorem (Theorem 1.24) at the last part.
(2) ⇒ (3): Let F be a closed subset of $\mathbb{R}^d$. Then since $U := \mathbb{R}^d \setminus F$ is open in $\mathbb{R}^d$,

$$\limsup_{n\to\infty} \mu_n(F) = \limsup_{n\to\infty}\bigl(1 - \mu_n(U)\bigr) = 1 - \liminf_{n\to\infty} \mu_n(U) \leq 1 - \mu(U) = \mu(F).$$

(3) ⇒ (2): Let U be an open subset of $\mathbb{R}^d$. Then since $F := \mathbb{R}^d \setminus U$ is closed in $\mathbb{R}^d$,

$$\liminf_{n\to\infty} \mu_n(U) = \liminf_{n\to\infty}\bigl(1 - \mu_n(F)\bigr) = 1 - \limsup_{n\to\infty} \mu_n(F) \geq 1 - \mu(F) = \mu(U).$$

(2), (3) ⇒ (4): Let $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and suppose F_μ is continuous at x. Set $I := (-\infty, x_1] \times \cdots \times (-\infty, x_d]$ and $J := (-\infty, x_1) \times \cdots \times (-\infty, x_d)$. Then

$$\begin{aligned}\mu(J) &= \lim_{n\to\infty} \mu\bigl((-\infty, x_1 - 1/n] \times \cdots \times (-\infty, x_d - 1/n]\bigr)\\ &= \lim_{n\to\infty} F_\mu((x_1 - 1/n, \ldots, x_d - 1/n) = F_\mu(x) = \mu(I).\end{aligned}$$

Since I is closed in $\mathbb{R}^d$, J is open in $\mathbb{R}^d$ and $J \subset I$, by virtue of (2) and (3) we obtain

$$\limsup_{n\to\infty} F_{\mu_n}(x) = \limsup_{n\to\infty} \mu_n(I) \leq \mu(I) = F_\mu(x) = \mu(J) \leq \liminf_{n\to\infty} \mu_n(J)$$

$$\leq \liminf_{n\to\infty} \mu_n(I) = \liminf_{n\to\infty} F_{\mu_n}(x) \leq \limsup_{n\to\infty} F_{\mu_n}(x).$$

Thus $F_\mu(x) = \limsup_{n\to\infty} F_{\mu_n}(x) = \liminf_{n\to\infty} F_{\mu_n}(x) = \lim_{n\to\infty} F_{\mu_n}(x)$.
(5) $\Rightarrow$ (1): Let $k \in \mathbb{N}$ and define a continuous function $g_k : \mathbb{R}^d \to [0,1]$ by $g_k(x) := (\min\{k - |x|, 1\})^+$, so that $g_k(x) = 1$ if $|x| \leq k-1$ and $g_k(x) = 0$ if $|x| \geq k$. Let $f \in C_b(\mathbb{R}^d)$ and set $M := \sup_{x\in\mathbb{R}^d} |f(x)|$. Then since $f + M \geq 0$ on $\mathbb{R}^d$ and $\big((f+M)g_k\big)|_{\mathbb{R}^d\setminus[-k,k]^d} = 0$, by using (5) we obtain

$$\begin{aligned}\liminf_{n\to\infty} \int_{\mathbb{R}^d} f d\mu_n + M &= \liminf_{n\to\infty} \int_{\mathbb{R}^d} (f+M) d\mu_n \geq \liminf_{n\to\infty} \int_{\mathbb{R}^d} (f+M) g_k d\mu_n \\ &= \lim_{n\to\infty} \int_{\mathbb{R}^d} (f+M) g_k d\mu_n = \int_{\mathbb{R}^d} (f+M) g_k d\mu. \quad (4.12)\end{aligned}$$

By virtue of the monotone convergence theorem (Theorem 1.24), letting $k \to \infty$ in (4.12) yields $\liminf_{n\to\infty} \int_{\mathbb{R}^d} f d\mu_n + M \geq \int_{\mathbb{R}^d} (f+M) d\mu = \int_{\mathbb{R}^d} f d\mu + M$. Thus $\int_{\mathbb{R}^d} f d\mu \leq \liminf_{n\to\infty} \int_{\mathbb{R}^d} f d\mu_n$, and by replacing f with $-f$ we also obtain $\limsup_{n\to\infty} \int_{\mathbb{R}^d} f d\mu_n \leq \int_{\mathbb{R}^d} f d\mu$. Hence $\int_{\mathbb{R}^d} f d\mu = \limsup_{n\to\infty} \int_{\mathbb{R}^d} f d\mu_n = \liminf_{n\to\infty} \int_{\mathbb{R}^d} f d\mu_n$, that is, $\lim_{n\to\infty} \int_{\mathbb{R}^d} f d\mu_n = \int_{\mathbb{R}^d} f d\mu$. Thus $\mu_n \xrightarrow{\mathcal{L}} \mu$.
(4) $\Rightarrow$ (5): For simplicity we assume $d = 1$; the proof for general d is provided after the proof for $d = 1$. Let $C_\mu := \{a \in \mathbb{R} \mid \mu(\{a\}) = 0\}$. By Problem 2.4-(4) and Problem 2.5, F_μ is continuous at any $x \in C_\mu$ and $\mathbb{R} \setminus C_\mu$ is a countable set, so that C_μ *is dense in* $\mathbb{R}$, i.e. $C_\mu \cap U \neq \emptyset$ for any non-empty open subset U of $\mathbb{R}$.

Let $f : \mathbb{R} \to \mathbb{R}$ be continuous, let $N \in \mathbb{N}$ and suppose $f|_{\mathbb{R}\setminus[-N+1,N-1]} = 0$. Let $\varepsilon \in (0,\infty)$. Since $[-N,N]$ is compact, f is uniformly continuous on $[-N,N]$ and hence there exists $\delta \in (0,1)$ such that $|f(x) - f(y)| < \varepsilon$ for any $x, y \in [-N,N]$ with $|x-y| < \delta$. In fact, more strongly

$$|f(x) - f(y)| < \varepsilon \qquad \text{for any } x, y \in \mathbb{R} \text{ with } |x-y| < \delta. \tag{4.13}$$

Indeed, let $x, y \in \mathbb{R}$ satisfy $|x-y| < \delta$. If $x, y \in \mathbb{R} \setminus [-N,N]$ then $|f(x) - f(y)| = 0$, and if $x \in [-N,N]$ and $y \in \mathbb{R} \setminus [-N,N]$ then by $|x-y| < \delta < 1$ we have $x \in \mathbb{R} \setminus [-N+1, N-1]$ and hence $|f(x) - f(y)| = 0$.

Since C_μ is dense in $\mathbb{R}$, we can choose $x_k \in C_\mu \cap \big(k\frac{\delta}{2}, (k+1)\frac{\delta}{2}\big)$ for each $k \in \mathbb{Z}$, so that $x_k < x_{k+1} < x_k + \delta$. We define $g : \mathbb{R} \to \mathbb{R}$ by

$$g := \sum_{k=-\infty}^{\infty} f(x_k) \mathbf{1}_{(x_{k-1}, x_k]}, \tag{4.14}$$

which is actually a finite sum since $f|_{\mathbb{R}\setminus[-N,N]} = 0$. We claim that

$$|f(x) - g(x)| < \varepsilon \qquad \text{for any } x \in \mathbb{R}, \tag{4.15}$$

$$\lim_{n\to\infty} \int_{\mathbb{R}} g d\mu_n = \int_{\mathbb{R}} g d\mu. \tag{4.16}$$

Indeed, if $x \in \mathbb{R}$ then $x \in (x_{k-1}, x_k]$ for a unique $k \in \mathbb{Z}^d$, so that $|x - x_k| < \delta$ and hence $|f(x) - g(x)| = |f(x) - f(x_k)| < \varepsilon$ by virtue of (4.13), proving (4.15). For

(4.16), let $k \in \mathbb{Z}$. By $x_{k-1}, x_k \in C_\mu$, F_μ is continuous at x_{k-1} and x_k and therefore the assumption (4) yields

$$\begin{aligned}\mu_n\bigl((x_{k-1}, x_k]\bigr) &= F_{\mu_n}(x_k) - F_{\mu_n}(x_{k-1}) \\ &\xrightarrow[n\to\infty]{(4)} F_\mu(x_k) - F_\mu(x_{k-1}) = \mu\bigl((x_{k-1}, x_k]\bigr),\end{aligned}$$

which immediately implies (4.16) since (4.14) is a finite sum.

Now by (4.16), we can choose $\ell \in \mathbb{N}$ so that $\left|\int_{\mathbb{R}^d} g d\mu_n - \int_{\mathbb{R}^d} g d\mu\right| < \varepsilon$ for any $n \in \mathbb{N}$ with $n \geq \ell$. For such n, by (4.15),

$$\begin{aligned}&\left|\int_{\mathbb{R}} f d\mu_n - \int_{\mathbb{R}} f d\mu\right| \\ &= \left|\int_{\mathbb{R}} (f-g) d\mu_n + \int_{\mathbb{R}} g d\mu_n - \int_{\mathbb{R}} g d\mu + \int_{\mathbb{R}} (g-f) d\mu\right| \qquad (4.17) \\ &\leq \int_{\mathbb{R}} |f-g| d\mu_n + \left|\int_{\mathbb{R}} g d\mu_n - \int_{\mathbb{R}} g d\mu\right| + \int_{\mathbb{R}} |g-f| d\mu < 3\varepsilon,\end{aligned}$$

proving $\lim_{n\to\infty} \int_{\mathbb{R}^d} f d\mu_n = \int_{\mathbb{R}^d} f d\mu$.

(4) $\Rightarrow$ (5) for general d: Following Problem 2.8, we set

$$C_{\mu,i} := \bigl\{a \in \mathbb{R} \bigm| \mu\bigl(H_i(a)\bigr) = 0\bigr\}, \quad \text{where } H_i(a) := \{(x_1, \ldots, x_d) \in \mathbb{R}^d \mid x_i = a\},$$

for each $i \in \{1, \ldots, d\}$ and $C_\mu := C_{\mu,1} \times \cdots \times C_{\mu,d}$. Then since $H_i(a) \cap H_i(b) = \emptyset$ for $a, b \in \mathbb{R}$ with $a \neq b$, it follows from Problem 1.14 that $\mathbb{R} \setminus C_{\mu,i}$ is a countable set, so that $C_{\mu,i}$ is dense in $\mathbb{R}$.

Let $f : \mathbb{R}^d \to \mathbb{R}$ be continuous, let $N \in \mathbb{N}$ and suppose $f|_{\mathbb{R}^d \setminus [-N+1, N-1]^d} = 0$. Let $\varepsilon \in (0, \infty)$. Since $[-N, N]^d$ is compact, f is uniformly continuous on $[-N, N]^d$, and similarly to (4.13) we see that there exists $\delta \in (0, 1)$ such that

$$|f(x) - f(y)| < \varepsilon \qquad \text{for any } x, y \in \mathbb{R}^d \text{ with } |x - y| < \delta. \tag{4.18}$$

Set $\delta_0 := \delta/\sqrt{4d}$ and let $i \in \{1, \ldots, d\}$. Since $C_{\mu,i}$ is dense in $\mathbb{R}$, we can choose $x_{i,k} \in C_{\mu,i} \cap \bigl(k\delta_0, (k+1)\delta_0\bigr)$ for each $k \in \mathbb{Z}$, so that $x_{i,k} < x_{i,k+1} < x_{i,k} + 2\delta_0$. We define $g : \mathbb{R}^d \to \mathbb{R}$ by

$$g := \sum_{k(1),\ldots,k(d)=-\infty}^{\infty} f\bigl(x_{1,k(1)}, \ldots, x_{d,k(d)}\bigr) \mathbf{1}_{(x_{1,k(1)-1}, x_{1,k(1)}] \times \cdots \times (x_{d,k(d)-1}, x_{d,k(d)}]}, \tag{4.19}$$

which is actually a finite sum since $f|_{\mathbb{R}^d \setminus [-N,N]^d} = 0$. We claim that

$$|f(x) - g(x)| < \varepsilon \qquad \text{for any } x \in \mathbb{R}^d, \tag{4.20}$$

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} g d\mu_n = \int_{\mathbb{R}^d} g d\mu. \tag{4.21}$$

Indeed, if $x \in \mathbb{R}^d$ then $x \in (x_{1,k(1)-1}, x_{1,k(1)}] \times \cdots \times (x_{d,k(d)-1}, x_{d,k(d)}]$ for a unique $\bigl(k(1), \ldots, k(d)\bigr) \in \mathbb{Z}^d$, so that $\bigl|x - \bigl(x_{1,k(1)}, \ldots, x_{d,k(d)}\bigr)\bigr| < \sqrt{d(2\delta_0)^2} = \delta$ and hence $|f(x) - g(x)| = \bigl|f(x) - f\bigl(x_{1,k(1)}, \ldots, x_{d,k(d)}\bigr)\bigr| < \varepsilon$ by virtue of (4.18), proving (4.20). For

(4.21), let $\big(k(1), \ldots, k(d)\big) \in \mathbb{Z}^d$ and set $h_i := x_{i,k(i)} - x_{i,k(i)-1}$ for $i \in \{1, \ldots, d\}$. Then for any $(\alpha_1, \ldots, \alpha_d) \in \{0,1\}^d$, $x_{i,k(i)} - \alpha_i h_i$ is either $x_{i,k(i)-1}$ or $x_{i,k(i)}$, which belongs to $C_{\mu,i}$, and hence we have $\big(x_{1,k(1)} - \alpha_1 h_1, \ldots, x_{d,k(d)} - \alpha_d h_d\big) \in C_\mu$. Since F_μ is continuous at any $x \in C_\mu$ by Problem 2.8-(2), using Proposition 2.18-(1) and the assumption (4), we obtain

$$
\begin{aligned}
&\mu_n\big((x_{1,k(1)-1}, x_{1,k(1)}] \times \cdots \times (x_{d,k(d)-1}, x_{d,k(d)}]\big) \\
&= \mu_n\big((x_{1,k(1)} - h_1, x_{1,k(1)}] \times \cdots \times (x_{d,k(d)} - h_d, x_{d,k(d)}]\big) \\
&= \sum_{(\alpha_1,\ldots,\alpha_d)\in\{0,1\}^d} (-1)^{\sum_{i=1}^d \alpha_i} F_{\mu_n}\big(x_{1,k(1)} - \alpha_1 h_1, \ldots, x_{d,k(d)} - \alpha_d h_d\big) \\
&\xrightarrow[n\to\infty]{(4)} \sum_{(\alpha_1,\ldots,\alpha_d)\in\{0,1\}^d} (-1)^{\sum_{i=1}^d \alpha_i} F_{\mu}\big(x_{1,k(1)} - \alpha_1 h_1, \ldots, x_{d,k(d)} - \alpha_d h_d\big) \\
&= \mu\big((x_{1,k(1)} - h_1, x_{1,k(1)}] \times \cdots \times (x_{d,k(d)} - h_d, x_{d,k(d)}]\big) \\
&= \mu\big((x_{1,k(1)-1}, x_{1,k(1)}] \times \cdots \times \big(x_{d,k(d)-1}, x_{d,k(d)}]\big),
\end{aligned}
$$

which immediately implies (4.21) since (4.19) is a finite sum.

Now by (4.21), we can choose $\ell \in \mathbb{N}$ so that $\left|\int_{\mathbb{R}^d} g\,d\mu_n - \int_{\mathbb{R}^d} g\,d\mu\right| < \varepsilon$ for any $n \in \mathbb{N}$ with $n \geq \ell$, and exactly the same calculation as in (4.17) shows that $\left|\int_{\mathbb{R}^d} f\,d\mu_n - \int_{\mathbb{R}^d} f\,d\mu\right| < 3\varepsilon$ for any $n \in \mathbb{N}$ with $n \geq \ell$, proving $\lim_{n\to\infty} \int_{\mathbb{R}^d} f\,d\mu_n = \int_{\mathbb{R}^d} f\,d\mu$. □

Proposition 4.11. *Let $X, \{X_n\}_{n=1}^\infty$ be d-dimensional random variables with $X_n \xrightarrow{\mathcal{L}} X$. Let $k \in \mathbb{N}$, $y \in \mathbb{R}^k$ and let $\{Y_n\}_{n=1}^\infty$ be k-dimensional random variables with $Y_n \xrightarrow{\mathrm{P}} y$. Then $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, y)$.*

Proof. Let $f : \mathbb{R}^{d+k} \to \mathbb{R}$ be continuous and satisfy $f|_{\mathbb{R}^{d+k}\setminus[-N,N]^{d+k}} = 0$ for some $N \in \mathbb{N}$. We show $\lim_{n\to\infty} \mathbb{E}[f(X_n, Y_n)] = \mathbb{E}[f(X, y)]$, which yields $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, y)$ by virtue of Theorem 4.10. Let $\varepsilon \in (0, \infty)$. Similarly to (4.13), f is uniformly continuous on $\mathbb{R}^{d+k}$ and hence there exists $\delta \in (0, \infty)$ such that

$$|f(x) - f(z)| < \varepsilon \qquad \text{for any } x, z \in \mathbb{R}^{d+k} \text{ with } |x - z| < \delta. \tag{4.22}$$

Then by $Y_n \xrightarrow{\mathrm{P}} y$ there exists $j \in \mathbb{N}$ such that

$$\mathbb{P}[|Y_n - y| \geq \delta] < \frac{\varepsilon}{M+1} \qquad \text{for any } n \in \mathbb{N} \text{ with } n \geq j, \tag{4.23}$$

where $M := \sup_{x\in\mathbb{R}^{d+k}} |f(x)|$. Moreover, since $f(\cdot, y) : \mathbb{R}^d \to \mathbb{R}$, $x \mapsto f(x, y)$, is bounded and continuous, by $X_n \xrightarrow{\mathcal{L}} X$ there exists $\ell \in \mathbb{N}$ such that

$$\big|\mathbb{E}[f(X_n, y)] - \mathbb{E}[f(X, y)]\big| < \varepsilon \qquad \text{for any } n \in \mathbb{N} \text{ with } n \geq \ell. \tag{4.24}$$

Now for any $n \in \mathbb{N}$ with $n \geq \max\{j, \ell\}$, by using (4.24), (4.22) and (4.23) we see that

$$
\begin{aligned}
&\big|\mathbb{E}[f(X_n, Y_n)] - \mathbb{E}[f(X, y)]\big| \\
&\leq \big|\mathbb{E}[f(X_n, Y_n) - f(X_n, y)]\big| + \big|\mathbb{E}[f(X_n, y)] - \mathbb{E}[f(X, y)]\big| \\
&\leq \mathbb{E}\big[|f(X_n, Y_n) - f(X_n, y)|\mathbf{1}_{\{|Y_n - y|\geq\delta\}} + |f(X_n, Y_n) - f(X_n, y)|\mathbf{1}_{\{|Y_n - y|<\delta\}}\big] + \varepsilon
\end{aligned}
$$

$$\leq 2M\mathbb{P}[|Y_n - y| \geq \delta] + \varepsilon + \varepsilon < 4\varepsilon,$$

proving $\lim_{n\to\infty} \mathbb{E}[f(X_n, Y_n)] = \mathbb{E}[f(X, y)]$ and hence $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, y)$. □

Theorem 4.12. *Let $\mathcal{K} \subset \mathcal{P}(\mathbb{R}^d)$. Then the following conditions are equivalent:*
(1) $\mathcal{K}$ *is* tight, *that is,* $\lim_{N\to\infty} \sup_{\mu\in\mathcal{K}} \mu\big(\mathbb{R}^d \setminus [-N, N]^d\big) = 0$.
(2) *For any $\{\mu_n\}_{n=1}^\infty \subset \mathcal{K}$, there exist $\mu \in \mathcal{P}(\mathbb{R}^d)$ and a strictly increasing sequence $\{n(k)\}_{k=1}^\infty \subset \mathbb{N}$ such that $\mu_{n(k)} \xrightarrow{\mathcal{L}} \mu$.*

Proof. (2) ⇒ (1): Suppose $\varepsilon := \lim_{N\to\infty} \sup_{\mu\in\mathcal{K}} \mu\big(\mathbb{R}^d \setminus [-N, N]^d\big) > 0$. Then for any $n \in \mathbb{N}$ there exists $\mu_n \in \mathcal{K}$ such that $\mu_n\big(\mathbb{R}^d \setminus [-n, n]^d\big) > \varepsilon/2$. By (2), there exist $\mu \in \mathcal{P}(\mathbb{R}^d)$ and a strictly increasing sequence $\{n(k)\}_{k=1}^\infty \subset \mathbb{N}$ such that $\mu_{n(k)} \xrightarrow{\mathcal{L}} \mu$. Let $N \in \mathbb{N}$. Since $\mathbb{R}^d \setminus (-N, N)^d$ is closed in $\mathbb{R}^d$, Theorem 4.10 yields

$$\begin{aligned}\frac{\varepsilon}{2} &\leq \limsup_{k\to\infty} \mu_{n(k)}\big(\mathbb{R}^d \setminus [-n(k), n(k)]^d\big)\\ &\leq \limsup_{k\to\infty} \mu_{n(k)}\big(\mathbb{R}^d \setminus (-N, N)^d\big) \leq \mu\big(\mathbb{R}^d \setminus (-N, N)^d\big)\end{aligned}$$

and letting $N \to \infty$ results in $\varepsilon/2 \leq 0$, a contradiction.
(1) ⇒ (2): For simplicity we assume $d = 1$; the proof for general d is provided after the proof for $d = 1$. Set $F_n := F_{\mu_n}$ for $n \in \mathbb{N}$ and let $\{q_k\}_{k=1}^\infty$ be an enumeration of $\mathbb{Q}$. Since $\{F_n(q_1)\}_{n=1}^\infty \subset [0, 1]$, by the Bolzano-Weierstrass theorem there exists a strictly increasing sequence $\{n(1, \ell)\}_{\ell=1}^\infty \subset \mathbb{N}$ such that the limit $\lim_{\ell\to\infty} F_{n(1,\ell)}(q_1) =: \rho(q_1)$ exists. Inductively for $k \in \mathbb{N}$ and a strictly increasing sequence $\{n(k, \ell)\}_{\ell=1}^\infty \subset \mathbb{N}$, by using the Bolzano-Weierstrass theorem we choose a subsequence $\{n(k + 1, \ell)\}_{\ell=1}^\infty$ of $\{n(k, \ell)\}_{\ell=1}^\infty$ so that the limit $\lim_{\ell\to\infty} F_{n(k+1,\ell)}(q_{k+1}) =: \rho(q_{k+1})$ exists. Thus inductively we can choose a strictly increasing sequence $\{n(k, \ell)\}_{\ell=1}^\infty \subset \mathbb{N}$ for each $k \in \mathbb{N}$, so that for any $k \in \mathbb{N}$, the limit $\lim_{\ell\to\infty} F_{n(k,\ell)}(q_k) =: \rho(q_k)$ exists and $\{n(k + 1, \ell)\}_{\ell=1}^\infty$ is a subsequence of $\{n(k, \ell)\}_{\ell=1}^\infty$. Let $n(\ell) := n(\ell, \ell)$ for $\ell \in \mathbb{N}$. Then $\{n(\ell)\}_{\ell=1}^\infty$ is strictly increasing, and for any $k \in \mathbb{N}$, $\{n(\ell)\}_{\ell=k}^\infty$ is a subsequence of $\{n(k, \ell)\}_{\ell=1}^\infty$. Thus $\lim_{\ell\to\infty} F_{n(\ell)}(q_k) = \rho(q_k)$ for any $k \in \mathbb{N}$, or in other words, $\lim_{\ell\to\infty} F_{n(\ell)}(q) = \rho(q)$ for any $q \in \mathbb{Q}$.

We define $F : \mathbb{R} \to [0, 1]$ by

$$F(x) := \inf_{q\in(x,\infty)\cap\mathbb{Q}} \rho(q), \tag{4.25}$$

so that F is non-decreasing since $(y, \infty) \cap \mathbb{Q} \subset (x, \infty) \cap \mathbb{Q}$ for any $x, y \in \mathbb{R}$ with $x \leq y$. We claim that F is right-continuous. Indeed, let $x \in \mathbb{R}$ and $\varepsilon \in (0, \infty)$. By the definition (4.25) of F we can take $q \in (x, \infty) \cap \mathbb{Q}$ such that $F(x) \leq \rho(q) < F(x) + \varepsilon$. If $y \in (x, q)$, then $q \in (y, \infty) \cap \mathbb{Q}$ and hence $F(x) \leq F(y) \leq \rho(q) < F(x) + \varepsilon$. Thus $\lim_{y\downarrow x} F(y) = F(x)$.

Next we prove $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$. Let $\varepsilon \in (0, \infty)$. By the assumption (1), there exists $N \in \mathbb{N}$ such that $\sup_{\ell\in\mathbb{N}} \mu_{n(\ell)}\big(\mathbb{R} \setminus [-N, N]\big) < \varepsilon$.

If $x \in (-\infty, -N)$, then we can choose $q \in (x, -N) \cap \mathbb{Q}$, and we have $(-\infty, q] \subset \mathbb{R} \setminus [-N, N]$ and

$$0 \leq F(x) \leq \rho(q) = \lim_{\ell \to \infty} F_{n(\ell)}(q) = \lim_{\ell \to \infty} \mu_{n(\ell)}\big((-\infty, q]\big) \leq \varepsilon,$$

proving $\lim_{x \to -\infty} F(x) = 0$. On the other hand, if $x \in [N, \infty)$ then for any $q \in (x, \infty) \cap \mathbb{Q}$, we have $[-N, N] \subset (-\infty, q]$ and

$$\rho(q) = \lim_{\ell \to \infty} F_{n(\ell)}(q) = \lim_{\ell \to \infty} \mu_{n(\ell)}\big((-\infty, q]\big) \geq 1 - \varepsilon.$$

Taking the infimum over $q \in (x, \infty) \cap \mathbb{Q}$, we conclude that $F(x) \geq 1 - \varepsilon$ for any $x \in [N, \infty)$. Thus $\lim_{x \to \infty} F(x) = 1$.

Since F is right-continuous, non-decreasing and satisfies $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$, $F = F_\mu$ for some $\mu \in \mathcal{P}(\mathbb{R})$ by Corollary 2.16 (recall that F_μ denotes the distribution function of μ). Now let $x \in \mathbb{R}$ and suppose F is continuous at x. We prove $\lim_{\ell \to \infty} F_{n(\ell)}(x) = F(x)$, which and Theorem 4.10 together imply $\mu_{n(\ell)} \xrightarrow{\mathcal{L}} \mu$. Let $p, q, r \in \mathbb{Q}$ be such that $p < q < x < r$. Then

$$\begin{aligned} F(p) \leq \rho(q) = \lim_{\ell \to \infty} F_{n(\ell)}(q) &\leq \liminf_{\ell \to \infty} F_{n(\ell)}(x) \\ &\leq \limsup_{\ell \to \infty} F_{n(\ell)}(x) \leq \lim_{\ell \to \infty} F_{n(\ell)}(r) = \rho(r), \end{aligned}$$

from which we obtain $F(x) = \limsup_{\ell \to \infty} F_{n(\ell)}(x) = \liminf_{\ell \to \infty} F_{n(\ell)}(x)$ by taking the infimum over $r \in (x, \infty) \cap \mathbb{Q}$ and using the continuity of F at x to let $p \uparrow x$. Thus $\lim_{\ell \to \infty} F_{n(\ell)}(x) = F(x)$.

(1) $\Rightarrow$ (2) for general d: Set $F_n := F_{\mu_n}$ for $n \in \mathbb{N}$. Since $\mathbb{Q}^d$ is a countable set, in exactly the same way as the above proof for $d = 1$, there exists a strictly increasing sequence $\{n(\ell)\}_{\ell=1}^{\infty} \subset \mathbb{N}$ such that the limit $\lim_{\ell \to \infty} F_{n(\ell)}(q) =: \rho(q)$ exists for any $q \in \mathbb{Q}^d$.

For each $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, set $J_x := (x_1, \infty) \times \cdots \times (x_d, \infty)$ and define

$$F(x) := \inf_{q \in J_x \cap \mathbb{Q}^d} \rho(q), \tag{4.26}$$

so that $F : \mathbb{R}^d \to [0, 1]$ is a function on $\mathbb{R}^d$. We claim that for any $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$,

$$F(x) = \lim_{J_x \cap \mathbb{Q}^d \ni q \to x} \rho(q) = \lim_{J_x \ni y \to x} F(y). \tag{4.27}$$

Indeed, let $\varepsilon \in (0, \infty)$. By the definition (4.26) of F we can take $q = (q_1, \ldots, q_n) \in J_x \cap \mathbb{Q}^d$ such that $F(x) \leq \rho(q) < F(x) + \varepsilon$. Note that $x_k < q_k$ for any $k \in \{1, \ldots, d\}$. If $y = (y_1, \ldots, y_d) \in J_x$ and $|y - x| < \min\{q_1 - x_1, \ldots, q_d - x_d\}$, then $y_k < q_k$ for any $k \in \{1, \ldots, d\}$, so that $q \in J_y \cap \mathbb{Q}^d$ and hence $F(x) \leq F(y) \leq \rho(q) < F(x) + \varepsilon$, where $F(x) \leq F(y)$ follows by $J_y \cap \mathbb{Q}^d \subset J_x \cap \mathbb{Q}^d$. Thus $\lim_{J_x \ni y \to x} F(y) = F(x)$. Moreover, if $r = (r_1, \ldots, r_d) \in J_x \cap \mathbb{Q}^d$ and $|r - x| < \min\{q_1 - x_1, \ldots, q_d - x_d\}$, then since $r_k < q_k$ for any $k \in \{1, \ldots, d\}$, we have $F_{n(\ell)}(r) \leq F_{n(\ell)}(q)$ for any $\ell \in \mathbb{N}$ and hence

$$F(x) \leq \rho(r) = \lim_{\ell \to \infty} F_{n(\ell)}(r) \leq \lim_{\ell \to \infty} F_{n(\ell)}(q) = \rho(q) < F(x) + \varepsilon,$$

proving $F(x) = \lim_{J_x \cap \mathbb{Q}^d \ni r \to x} \rho(r)$. Therefore (4.27) follows.

Next we show that F satisfies the conditions (F1), (F2) and (F3) in Theorem 2.19, so that $F = F_\mu$ for some $\mu \in \mathcal{P}(\mathbb{R}^d)$ by Theorem 2.19. (F2) follows from the latter equality in (4.27). We set $\alpha h := (\alpha_1 h_1, \dots, \alpha_d h_d)$ for $h = (h_1, \dots, h_d) \in (0,\infty)^d$, $\alpha = (\alpha_1, \dots, \alpha_d) \in \{0,1\}^d$. Let $x \in \mathbb{R}^d$, $q \in J_x \cap \mathbb{Q}^d$ $h = (h_1, \dots, h_d) \in (0,\infty)^d$, and let $\eta = (\eta_1, \dots, \eta_d) \in (0,\infty)^d \cap \mathbb{Q}^d$ be such that $\eta_k \leq h_k$ for any $k \in \{1, \dots, d\}$. By Proposition 2.18-(1),

$$\begin{aligned} &\sum_{(\alpha_1,\dots,\alpha_d)\in\{0,1\}^d} (-1)^{\sum_{i=1}^d \alpha_i} \rho(q - \alpha\eta) \\ &= \lim_{\ell\to\infty} \sum_{(\alpha_1,\dots,\alpha_d)\in\{0,1\}^d} (-1)^{\sum_{i=1}^d \alpha_i} F_{n(\ell)}(q - \alpha\eta) \geq 0, \end{aligned} \tag{4.28}$$

and since $q - \alpha\eta \in J_{x-\alpha h} \cap \mathbb{Q}^d$ for any $\alpha \in \{0,1\}^d$, letting $q \to x$ and $\eta \to h$ at the same time in (4.28) yields $\sum_{(\alpha_1,\dots,\alpha_d)\in\{0,1\}^d} (-1)^{\sum_{i=1}^d \alpha_i} F(x - \alpha h) \geq 0$ by virtue of (4.27). Thus (F1) follows. Next for (F3), let $\varepsilon \in (0,\infty)$. By the assumption (1), there exists $N \in \mathbb{N}$ such that $\sup_{\ell\in\mathbb{N}} \mu_{n(\ell)}\bigl(\mathbb{R}^d \setminus [-N,N]^d\bigr) < \varepsilon$. Let $k \in \{1, \dots, d\}$ and $x = (x_1, \dots, x_d) \in \mathbb{R}^d$. If $x_k < -N$, then we can choose $q = (q_1, \dots, q_d) \in J_x \cap \mathbb{Q}^d$ so that $q_k < -N$, and we have $(-\infty, q_1] \times \cdots \times (-\infty, q_d] \subset \mathbb{R}^d \setminus [-N,N]^d$ and

$$0 \leq F(x) \leq \rho(q) = \lim_{\ell\to\infty} F_{n(\ell)}(q) = \lim_{\ell\to\infty} \mu_{n(\ell)}\bigl((-\infty, q_1] \times \cdots \times (-\infty, q_d]\bigr) \leq \varepsilon,$$

proving $\lim_{x_k\to-\infty} F(x_1, \dots, x_k, \dots, x_d) = 0$. On the other hand, if $x \in [N,\infty)^d$ then for any $q = (q_1, \dots, q_d) \in J_x \cap \mathbb{Q}^d$, $[-N,N]^d \subset (-\infty, q_1] \times \cdots \times (-\infty, q_d]$ and

$$\rho(q) = \lim_{\ell\to\infty} F_{n(\ell)}(q) = \lim_{\ell\to\infty} \mu_{n(\ell)}\bigl((-\infty, q_1] \times \cdots \times (-\infty, q_d]\bigr) \geq 1 - \varepsilon.$$

Taking the infimum over $q \in J_x \cap \mathbb{Q}^d$ yields $F(x) \geq 1 - \varepsilon$. In particular, $F(x, \dots, x) \geq 1 - \varepsilon$ for any $x \in [N,\infty)$, hence $\lim_{x\to\infty} F(x, \dots, x) = 1$ and (F3) follows.

Now let $x \in \mathbb{R}^d$ and suppose F is continuous at x. We prove $\lim_{\ell\to\infty} F_{n(\ell)}(x) = F(x)$, which and Theorem 4.10 together imply $\mu_{n(\ell)} \xrightarrow{\mathcal{L}} \mu$ for $\mu \in \mathcal{P}(\mathbb{R}^d)$ satisfying $F = F_\mu$. Let $p, q, r \in \mathbb{Q}^d$ be such that $r \in J_x$, $x \in J_q$ and $q \in J_p$. Then

$$\begin{aligned} \limsup_{\ell\to\infty} F_{n(\ell)}(x) &\leq \lim_{\ell\to\infty} F_{n(\ell)}(r) = \rho(r) \xrightarrow[(4.27)]{r\to x} F(x), \\ \liminf_{\ell\to\infty} F_{n(\ell)}(x) &\geq \lim_{\ell\to\infty} F_{n(\ell)}(q) = \rho(q) \geq F(p) \xrightarrow{p\to x} F(x) \end{aligned}$$

by the continuity of F at x. Thus $F(x) = \limsup_{\ell\to\infty} F_{n(\ell)}(x) = \liminf_{\ell\to\infty} F_{n(\ell)}(x)$ and hence $\lim_{\ell\to\infty} F_{n(\ell)}(x) = F(x)$. □

4.2 Characteristic Functions

This section is devoted to preparing for the proof of the central limit theorem given in the next section. The key tool for the proof is characteristic functions of laws on $\mathbb{R}^d$, defined as follows. Recall that $\langle x, y\rangle$ denotes the usual inner product of $x, y \in \mathbb{R}^d$.

Convention. (1) From this section on, the symbol i always denotes the imaginary unit.
(2) For $z \in \mathbb{C}$, $\overline{z}$ denotes its *complex conjugate*, i.e. $\overline{z} := \mathrm{Re}(z) - i\,\mathrm{Im}(z)$. Note that $|z|^2 = z\overline{z}$.
(3) A $(\mathbb{C}, \mathcal{B}(\mathbb{C}))$-valued random variable is simply called a *complex random variable*.

Recall that for any $\theta \in \mathbb{R}$, $e^{i\theta} = \cos\theta + i\sin\theta$, hence $\overline{e^{i\theta}} = e^{-i\theta}$, $\left|e^{i\theta}\right| = 1$ and $\frac{d}{d\theta}e^{i\theta} = ie^{i\theta}$.

Definition 4.13. (1) For $\mu \in \mathcal{P}(\mathbb{R}^d)$, we define its *characteristic function* $\varphi_\mu : \mathbb{R}^d \to \mathbb{C}$ by

$$\varphi_\mu(t) := \int_{\mathbb{R}^d} e^{i\langle t,x\rangle}\mu(dx) \qquad \text{for each } t \in \mathbb{R}^d, \tag{4.29}$$

where the integral is always defined since $x \mapsto e^{i\langle t,x\rangle}$ is continuous and $\left|e^{i\langle t,x\rangle}\right| = 1$.
(2) For a d-dimensional random variable X, we define its *characteristic function* $\varphi_X : \mathbb{R}^d \to \mathbb{C}$ by

$$\varphi_X(t) := \mathbb{E}\bigl[e^{i\langle t,X\rangle}\bigr] = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle}\mathbb{P}_X(dx) \qquad \text{for each } t \in \mathbb{R}^d, \tag{4.30}$$

i.e. φ_X is defined as the characteristic function $\varphi_{\mathcal{L}(X)}$ of the law $\mathcal{L}(X) = \mathbb{P}_X$ of X.

Proposition 4.14. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$. Then its characteristic function φ_μ possesses the following properties:*

(φ1) $\varphi_\mu(0) = 1$.

(φ2) $|\varphi_\mu(t)| \le 1$ *and* $\varphi_\mu(-t) = \overline{\varphi_\mu(t)}$ *for any* $t \in \mathbb{R}^d$.

(φ3) φ_μ *is uniformly continuous on* $\mathbb{R}^d$.

(φ4) (Non-negative definiteness) *For any* $n \in \mathbb{N}$, $\{z_k\}_{k=1}^n \subset \mathbb{C}$ *and* $\{t_k\}_{k=1}^n \subset \mathbb{R}^d$,

$$\sum_{k,\ell=1}^{n} \varphi_\mu(t_k - t_\ell) z_k \overline{z_\ell} \ge 0. \tag{4.31}$$

Proof. (φ1) $\varphi_\mu(0) = \mu(\mathbb{R}^d) = 1$.
(φ2) $|\varphi_\mu(t)| \le \int_{\mathbb{R}^d}\left|e^{i\langle t,x\rangle}\right|\mu(dx) = 1$ by Proposition 1.42-(1) and $\left|e^{i\langle t,x\rangle}\right| = 1$. The latter assertion follows by $e^{i\langle -t,x\rangle} = \overline{e^{i\langle t,x\rangle}}$.
(φ3) For any $\{h_n\}_{n=1}^\infty \subset \mathbb{R}^d$ with $\lim_{n\to\infty} h_n = 0$, by $\left|e^{i\langle h_n,x\rangle}\right| \le 1$ the dominated convergence theorem (Theorem 1.32) applies to yield $\lim_{n\to\infty}\int_{\mathbb{R}^d} e^{i\langle h_n,x\rangle}\mu(dx) = 1$, which means $\lim_{h\to 0}\varphi_\mu(h) = \lim_{h\to 0}\int_{\mathbb{R}^d} e^{i\langle h,x\rangle}\mu(dx) = 1$ since $\{h_n\}_{n=1}^\infty \subset \mathbb{R}^d$ is an arbitrary sequence with $\lim_{n\to\infty} h_n = 0$.

Now let $t, h \in \mathbb{R}^d$. Then since $\left|e^{i\langle h,x\rangle} - 1\right|^2 = 2 - 2\,\mathrm{Re}\bigl(e^{i\langle h,x\rangle}\bigr)$, by using Hölder's inequality (Theorem 1.48) we obtain

$$\begin{aligned}
|\varphi_\mu(t+h) - \varphi(t)| &\le \int_{\mathbb{R}^d}\left|e^{i\langle t+h,x\rangle} - e^{i\langle t,x\rangle}\right|\mu(dx) = \int_{\mathbb{R}^d}\left|e^{i\langle h,x\rangle} - 1\right|\mu(dx)\\
&\le \left(\int_{\mathbb{R}^d}\left|e^{i\langle h,x\rangle} - 1\right|^2\mu(dx)\right)^{1/2}\mu(\mathbb{R}^d)^{1/2}\\
&= \left(\int_{\mathbb{R}^d}\Bigl(2 - 2\,\mathrm{Re}\bigl(e^{i\langle h,x\rangle}\bigr)\Bigr)\mu(dx)\right)^{1/2} = \sqrt{2 - 2\,\mathrm{Re}\bigl(\varphi_\mu(h)\bigr)},
\end{aligned}$$

which tends to 0 as $h \to 0$, *independently of* $t \in \mathbb{R}^d$. Thus φ_μ is uniformly continuous.
(φ4) By the definition (4.29) of φ_μ,

$$\begin{aligned}\sum_{k,\ell=1}^{n} \varphi_\mu(t_k - t_\ell) z_k \overline{z_\ell} &= \int_{\mathbb{R}^d} \sum_{k,\ell=1}^{n} z_k \overline{z_\ell} e^{i\langle t_k - t_\ell, x\rangle} \mu(dx) \\ &= \int_{\mathbb{R}^d} \sum_{k,\ell=1}^{n} z_k e^{i\langle t_k, x\rangle} \overline{z_\ell e^{i\langle t_\ell, x\rangle}} \mu(dx) \\ &= \int_{\mathbb{R}^d} \sum_{k=1}^{n} z_k e^{i\langle t_k, x\rangle} \overline{\sum_{\ell=1}^{n} z_\ell e^{i\langle t_\ell, x\rangle}} \mu(dx) \\ &= \int_{\mathbb{R}^d} \left| \sum_{k=1}^{n} z_k e^{i\langle t_k, x\rangle} \right|^2 \mu(dx) \geq 0,\end{aligned}$$

proving (φ4). □

Various properties of laws and random variables are reflected in their characteristic functions. The integrability of a random variables is closely related with smoothness of its characteristic function in the following way, which also provides a method of calculating mean and (co-)variance via characteristic functions.

Theorem 4.15. *Let $X = (X_1, \dots, X_d)$ be a d-dimensional random variable and let φ_X be its characteristic function. Let $n \in \mathbb{N}$ and suppose $\mathbb{E}[|X|^n] < \infty$. Then for any $k \in \{1, \dots, n\}$ and $\{j_\ell\}_{\ell=1}^k \subset \{1, \dots, d\}$, the partial derivative $\partial^k \varphi_X / \partial t_{j_1} \dots \partial t_{j_k}$ exists on $\mathbb{R}^d$, is continuous and*

$$\frac{\partial^k \varphi_X}{\partial t_{j_1} \dots \partial t_{j_k}}(t) = i^k \mathbb{E}\big[X_{j_1} \cdots X_{j_k} e^{i\langle t, X\rangle}\big] \qquad \text{for any } t = (t_1, \dots, t_d) \in \mathbb{R}^d. \tag{4.32}$$

In particular, for any $k \in \{1, \dots, n\}$ and $\{j_\ell\}_{\ell=1}^k \subset \{1, \dots, d\}$,

$$\mathbb{E}[X_{j_1} \cdots X_{j_k}] = (-i)^k \frac{\partial^k \varphi_X}{\partial t_{j_1} \dots \partial t_{j_k}}(0). \tag{4.33}$$

Proof. Since $\big|X_{j_1} \cdots X_{j_k} e^{i\langle t, X\rangle}\big| \leq |X|^k$ and $\mathbb{E}[|X|^k]^{1/k} \leq \mathbb{E}[|X|^n]^{1/n} < \infty$ by Proposition 3.6-(2), the mean $\mathbb{E}\big[X_{j_1} \cdots X_{j_k} e^{i\langle t, X\rangle}\big]$ in (4.32) is defined for any $t \in \mathbb{R}^d$. Moreover, the dominated convergence theorem (Theorem 1.32) implies that for any $t \in \mathbb{R}^d$ and $\{t_n\}_{n=1}^\infty \subset \mathbb{R}^d$ with $\lim_{n\to\infty} t_n = t$, $\lim_{n\to\infty} \mathbb{E}\big[X_{j_1} \cdots X_{j_k} e^{i\langle t_n, X\rangle}\big] = \mathbb{E}\big[X_{j_1} \cdots X_{j_k} e^{i\langle t, X\rangle}\big]$, that is, the right-hand side of (4.32) is continuous in $t \in \mathbb{R}^d$.

The proof of (4.32) is by induction in k. Let $k \in \{0, \dots, n-1\}$ and suppose that the assertion is valid for k (we suppose nothing if $k = 0$). Let $\{j_\ell\}_{\ell=1}^{k+1} \subset \{1, \dots, d\}$. Since $\mathbb{E}[|X|^{k+1}] < \infty$ and for any $t \in \mathbb{R}^d$ we have $\big|X_{j_1} \cdots X_{j_{k+1}} e^{i\langle t, X\rangle}\big| \leq |X|^{k+1}$ and

$$\frac{\partial}{\partial t_{j_1}}\big(X_{j_2} \cdots X_{j_{k+1}} e^{i\langle t, X\rangle}\big)(t) = i X_{j_1} \cdots X_{j_{k+1}} e^{i\langle t, X\rangle},$$

the induction hypothesis and Theorem 1.47 together imply that for any $t \in \mathbb{R}^d$,

$$\begin{aligned}\frac{\partial}{\partial t_{j_1}}\left(\frac{\partial^k \varphi_X}{\partial t_{j_2}\ldots\partial t_{j_{k+1}}}\right)(t) &= \frac{\partial}{\partial t_{j_1}}\Big(i^k \mathbb{E}\big[X_{j_2}\cdots X_{j_{k+1}} e^{i\langle t,X\rangle}\big]\Big)(t)\\ &= i^{k+1}\mathbb{E}\big[X_{j_1}\cdots X_{j_{k+1}} e^{i\langle t,X\rangle}\big],\end{aligned}$$

which is (4.32). Finally, setting $t := 0$ in (4.32) yields (4.33). □

The following proposition is a partial converse of Theorem 4.15.

Theorem 4.16. *Let X be a real random variable and let $n \in \mathbb{N}$. If the characteristic function φ_X of X has the $(2n-1)$-th derivative $\varphi_X^{(2n-1)}$ on $(-a, a)$ for some $a \in (0,\infty)$ and has the $2n$-th derivative $\varphi_X^{(2n)}(0)$ at 0, then $\mathbb{E}[X^{2n}] < \infty$.*

We need the following easy fact from calculus for the proof of Theorem 4.16. Recall that "$f(x) = g(x) + o\big(h(x)\big)$ as $x \to a$" means $\lim_{x\to a} \frac{f(x)-g(x)}{h(x)} = 0$.

Lemma 4.17. *Let $a \in (0,\infty)$ and let $f : (-a, a) \to \mathbb{C}$ be differentiable. If $f''(0)$ exists, then*

$$f(h) = f(0) + f'(0)h + \frac{1}{2} f''(0)h^2 + o(h^2) \qquad \text{as } h \to 0, \tag{4.34}$$

$$f''(0) = \lim_{h\downarrow 0} \frac{f(h) + f(-h) - 2f(0)}{h^2}. \tag{4.35}$$

Proof. Define $g : (-a, a) \to \mathbb{R}$ by $g(x) := f(x) - f(0) - f'(0)x - f''(0)x^2/2$, so that $g'(x) = f'(x) - f'(0) - f''(0)x$ and $g''(0) = 0$. Let $\varepsilon \in (0,\infty)$. Since $g'(0) = g''(0) = 0$, there exists $\delta \in (0, a)$ such that $|g'(h)/h| \le \varepsilon$ for any $h \in (-\delta, \delta) \setminus \{0\}$. If $h \in (-\delta, \delta)\setminus\{0\}$, then by $g(0) = 0$ and the mean value theorem there exists $\theta_h \in (0, 1)$ such that $g(h) = g(h) - g(0) = hg'(\theta_h h)$, and hence

$$\left| f(h) - f(0) - f'(0)h - \frac{1}{2} f''(0)h^2 \right| = \left|\frac{g(h)}{h^2}\right| = \left|\frac{g'(\theta_h h)}{h}\right| \le \left|\frac{\varepsilon\theta_h h}{h}\right| < \varepsilon,$$

which shows (4.34). Then $f(h) + f(-h) - 2f(0) = f''(0)h^2 + o(h^2)$ as $h \to 0$ by (4.34) and hence (4.35) follows. □

Proof of Theorem 4.16. Let $k \in \{0,\ldots,n-1\}$ and suppose $\mathbb{E}[X^{2k}] < \infty$. Then $\varphi_X^{(2k)}(t) = (-1)^k \mathbb{E}[X^{2k} e^{itX}]$ for $t \in \mathbb{R}$ by Theorem 4.15, and hence by (4.35),

$$\begin{aligned}\big|\varphi_X^{(2k+2)}(0)\big| &= \lim_{j\to\infty}\left|\frac{\varphi_X^{(2k)}(1/j) + \varphi_X^{(2k)}(-1/j) - 2\varphi_X^{(2k)}(0)}{1/j^2}\right|\\ &= \lim_{j\to\infty} \mathbb{E}\left[X^{2k}\left(\frac{\sin(Xj^{-2}/2)}{j^{-2}/2}\right)^2\right] \ge \mathbb{E}[X^{2k+2}],\end{aligned}$$

where the inequality follows by Fatou's lemma (Theorem 1.27) and $\lim_{h\downarrow 0} \frac{\sin Xh}{h} = X$. Thus $\mathbb{E}[X^{2k}] < \infty$ implies $\mathbb{E}[X^{2k+2}] < \infty$ for each $k \in \{0,\ldots,n-1\}$, and hence $\mathbb{E}[X^{2n}] < \infty$ by induction in k. □

Multiplication of characteristic functions corresponds to sum of independent random variables in the following sense.

Proposition 4.18. *Let $n \in \mathbb{N}$ and let $\{X_k\}_{k=1}^n$ be independent d-dimensional random variables. Then*

$$\varphi_{X_1+\cdots+X_n}(t) = \varphi_{X_1}(t)\cdots\varphi_{X_n}(t) \qquad \text{for any } t \in \mathbb{R}^d. \tag{4.36}$$

In particular, if $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\{X_k\}_{k=1}^n$ is i.i.d. with $X_1 \sim \mu$, then

$$\varphi_{X_1+\cdots+X_n}(t) = \varphi_\mu(t)^n \qquad \text{for any } t \in \mathbb{R}^d. \tag{4.37}$$

Proof. Let $t \in \mathbb{R}^d$. By Theorem 3.10, the independence of $\{X_k\}_{k=1}^n$ and Problem 2.11 (for $\mathbb{C}$-valued measurable functions, which can be verified in exactly the same way by using Fubini's theorem (Theorem 2.29)), we obtain

$$\begin{aligned}\varphi_{X_1+\cdots+X_n}(t) &= \mathbb{E}\big[e^{i\langle t, X_1+\cdots+X_n\rangle}\big] = \int_{\mathbb{R}^{dn}} e^{i\langle t, x_1+\cdots+x_n\rangle}\mathbb{P}_{(X_1,\ldots,X_n)}(dx_1\ldots dx_n)\\ &= \int_{\mathbb{R}^{dn}} e^{i\langle t,x_1\rangle}\ldots e^{i\langle t,x_n\rangle}\mathbb{P}_{X_1}\times\cdots\times\mathbb{P}_{X_n}(dx_1\ldots dx_n)\\ &= \int_{\mathbb{R}^d} e^{i\langle t,x_1\rangle}\mathbb{P}_{x_1}(dx_1)\cdots\int_{\mathbb{R}^d} e^{i\langle t,x_n\rangle}\mathbb{P}_{x_n}(dx_n) = \varphi_{X_1}(t)\cdots\varphi_{X_n}(t),\end{aligned}$$

proving (4.36). (4.37) is immediate from (4.36) and $\varphi_{X_k} = \varphi_\mu$, $k \in \{1,\ldots,n\}$. □

Proposition 4.19. *Let X be a d-dimensional random variable. Let $k \in \mathbb{N}$, let $m \in \mathbb{R}^k$, let $T : \mathbb{R}^d \to \mathbb{R}^k$ be linear and let $T^* : \mathbb{R}^k \to \mathbb{R}^d$ be the adjoint (i.e. transpose) of T. Then*

$$\varphi_{TX+m}(t) = e^{i\langle t,m\rangle}\varphi_X(T^*t) \qquad \text{for any } t \in \mathbb{R}^k. \tag{4.38}$$

Proof. Let $t \in \mathbb{R}^k$. Recalling that $\langle t, Tx\rangle = \langle T^*t, x\rangle$ for $x \in \mathbb{R}^d$, we see that

$$\varphi_{TX+m}(t) = \mathbb{E}\big[e^{i\langle t,TX+m\rangle}\big] = e^{i\langle t,m\rangle}\mathbb{E}\big[e^{i\langle T^*t,X\rangle}\big] = e^{i\langle t,m\rangle}\varphi_X(T^*t),$$

proving (4.38). □

Next we present concrete examples of the characteristic functions Recall Section 3.2 for the definitions of probability distributions on $\mathbb{R}$ mentioned below.

Example 4.20. Let X be a real random variable and let $t \in \mathbb{R}$.
(1) If X has the binomial distribution $B(n, p)$, $n \in \mathbb{N}$, $p \in [0, 1]$, then

$$\varphi_X(t) = \big(1 + p(e^{it} - 1)\big)^n. \tag{4.39}$$

(2) If X has the Poisson distribution $\mathrm{Po}(\lambda)$, $\lambda \in (0, \infty)$, then

$$\varphi_X(t) = \exp\big(\lambda(e^{it} - 1)\big). \tag{4.40}$$

(3) If X has the geometric distribution $\mathrm{Geom}(\alpha)$, $\alpha \in [0, 1)$, then

$$\varphi_X(t) = \frac{1-\alpha}{1-\alpha e^{it}}. \tag{4.41}$$

(4) If X has the uniform distribution $\mathrm{Unif}(-a,a)$ on $(-a,a)$, $a \in (0,\infty)$, then

$$\varphi_X(t) = \frac{\sin at}{at}. \tag{4.42}$$

It is left to the reader as an exercise to verify these equalities (Exercise 4.5).

Example 4.21. Let $\alpha \in (0,\infty)$ and let X be a real random variable with $X \sim \mathrm{Exp}(\alpha)$. Then for any $t \in \mathbb{R}$,

$$\varphi_X(t) = \frac{\alpha}{\alpha - it}. \tag{4.43}$$

Indeed, since $\left|\alpha e^{(-\alpha+it)x}\right| = \alpha e^{-\alpha x}$, $\int_0^\infty \alpha e^{-\alpha x}dx = 1 < \infty$, and $\left(e^{(-\alpha+it)x}\right)' = (-\alpha+it)e^{(-\alpha+it)x}$, by using the dominated convergence theorem (Theorem 1.32) we see that

$$\begin{aligned}\varphi_X(t) &= \mathbb{E}\left[e^{itX}\right] = \int_0^\infty e^{itx}\alpha e^{-\alpha x}dx = \int_0^\infty \alpha e^{(-\alpha+it)x}dx \\ &= \lim_{n\to\infty}\int_0^n \alpha e^{(-\alpha+it)x}dx = \lim_{n\to\infty}\left[\frac{\alpha}{-\alpha+it}e^{(-\alpha+it)x}\right]_0^n = \frac{\alpha}{\alpha-it}.\end{aligned}$$

Example 4.22. Let $m \in \mathbb{R}$, $v \in [0,\infty)$ and let X be a real random variable with $X \sim N(m,v)$. Then for any $t \in \mathbb{R}$,

$$\varphi_X(t) = \exp\bigl(itm - t^2 v/2\bigr). \tag{4.44}$$

Indeed, if $v = 0$ then $X \sim \delta_m$ and hence $\varphi_X(t) = \int_{\mathbb{R}} e^{itx}\delta_m(dx) = e^{itm}$ for any $t \in \mathbb{R}$. Next assume $v > 0$ and let $t \in \mathbb{R}$. Since $\sin(-x) = -\sin x$ for $x \in \mathbb{R}$, by using Corollary 2.40 we have

$$\begin{aligned}\varphi_X(t) &= \int_{-\infty}^\infty e^{itx}\frac{1}{\sqrt{2\pi v}}\exp\left(-\frac{(x-m)^2}{2v}\right)dx = e^{itm}\int_{-\infty}^\infty e^{it\sqrt{v}y}\frac{e^{-y^2/2}}{\sqrt{2\pi}}dy \\ &= e^{itm}\left(\int_{-\infty}^\infty \cos\bigl(t\sqrt{v}y\bigr)\frac{e^{-y^2/2}}{\sqrt{2\pi}}dy + i\int_{-\infty}^\infty \sin\bigl(t\sqrt{v}y\bigr)\frac{e^{-y^2/2}}{\sqrt{2\pi}}dy\right) \\ &= e^{itm}\varphi\bigl(\sqrt{v}t\bigr), \quad \text{where} \quad \varphi(t) := \int_{-\infty}^\infty \cos(tx)\frac{e^{-x^2/2}}{\sqrt{2\pi}}dx.\end{aligned} \tag{4.45}$$

Thus it suffices to calculate $\varphi(t)$ defined in (4.45). Since

$$\left|\frac{\partial}{\partial t}\left(\cos(tx)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\right)(t)\right| = \left|x\sin(tx)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\right| \le |x|\frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

and $\int_{-\infty}^\infty |x|e^{-x^2/2}dx < \infty$, by using Theorem 1.47 and the dominated convergence theorem (Theorem 1.32) twice, we obtain

$$\varphi'(t) = -\int_{-\infty}^\infty x\sin(tx)\frac{e^{-x^2/2}}{\sqrt{2\pi}}dx = \lim_{n\to\infty}\int_{-n}^n \sin(tx)\frac{-xe^{-x^2/2}}{\sqrt{2\pi}}dx$$

$$= \lim_{n\to\infty}\left(\left[\sin(tx)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\right]_{-n}^{n} - \int_{-n}^{n} t\cos(tx)\frac{e^{-x^2/2}}{\sqrt{2\pi}}dx\right)$$
$$= -t\int_{-\infty}^{\infty}\cos(tx)\frac{e^{-x^2/2}}{\sqrt{2\pi}}dx = -t\varphi(t).$$

Therefore $\frac{d}{dt}\big(e^{t^2/2}\varphi(t)\big) = e^{t^2/2}\big(\varphi'(t)+t\varphi(t)\big) = 0$ and hence $e^{t^2/2}\varphi(t) = e^0\varphi(0) = 1$ for any $t \in \mathbb{R}$. Thus $\varphi(t) = e^{-t^2/2}$, so that $\varphi_X(t) = e^{itm}\varphi\big(\sqrt{v}t\big) = e^{itm-t^2v/2}$.

Example 4.23. Let $\alpha, \beta \in (0,\infty)$ and let X be a real random variable with $X \sim \mathrm{Gamma}(\alpha,\beta)$. Then for any $t \in \mathbb{R}$,

$$\varphi_X(t) = \frac{\beta^\alpha}{(\beta - it)^\alpha}, \tag{4.46}$$

where $z^\gamma := e^{\gamma\log z}$ for $\gamma \in \mathbb{C}$ and $z \in \mathbb{C}\setminus(-\infty,0]$, with $(-\infty,0] \subset \mathbb{R}$ regarded as a subset of $\mathbb{C}$ and $\log : \mathbb{C}\setminus(-\infty,0] \to \{|\operatorname{Im}(z)| < \pi\}$ denoting the inverse map of the C^1-embedding $\exp : \{|\operatorname{Im}(z)| < \pi\} \to \mathbb{C}\setminus(-\infty,0]$.

The proof of (4.46) is similar to the case of normal distributions in Example 4.22. Recalling Example 3.23, we have

$$\varphi_X(t) = \frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty e^{itx}x^{\alpha-1}e^{-\beta x}dx = \frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty x^{\alpha-1}e^{-(\beta-it)x}dx. \tag{4.47}$$

Since $\big|\frac{\partial}{\partial t}\big(x^{\alpha-1}e^{-(\beta-it)x}\big)\big| = x^\alpha e^{-\beta x}$ and $\int_0^\infty x^\alpha e^{-\beta x}dx < \infty$, by using Theorem 1.47 and the dominated convergence theorem (Theorem 1.32) we obtain

$$\varphi_X'(t) = \frac{i\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty x^\alpha e^{-(\beta-it)x}dx = \frac{i\beta^\alpha}{\Gamma(\alpha)}\lim_{n\to\infty}\int_0^n x^\alpha e^{-(\beta-it)x}dx$$
$$= \frac{i\beta^\alpha}{\Gamma(\alpha)}\lim_{n\to\infty}\left(\left[\frac{-1}{\beta-it}x^\alpha e^{-(\beta-it)x}\right]_0^n + \frac{\alpha}{\beta-it}\int_0^n x^{\alpha-1}e^{-(\beta-it)x}dx\right)$$
$$= \frac{\beta^\alpha}{\Gamma(\alpha)}\frac{i\alpha}{\beta-it}\int_0^\infty x^{\alpha-1}e^{-(\beta-it)x}dx = \frac{i\alpha}{\beta-it}\varphi_X(t).$$

Therefore $\frac{d}{dt}\big((\beta-it)^\alpha\varphi_X(t)\big) = (\beta-it)^{\alpha-1}\big((\beta-it)\varphi_X'(t) - i\alpha\varphi_X(t)\big) = 0$ and hence $(\beta-it)^\alpha\varphi_X(t) = (\beta-i0)^\alpha\varphi_X(0) = \beta^\alpha$ for any $t \in (0,\infty)$, which shows (4.47).

Example 4.24. Let $m \in \mathbb{R}$, $\alpha \in (0,\infty)$ and let X be a real random variable with $X \sim \mathrm{Cauchy}(m,\alpha)$. Then for any $t \in \mathbb{R}$,

$$\varphi_X(t) = \exp\big(itm - \alpha|t|\big). \tag{4.48}$$

Recalling Example 3.25 and using Corollary 2.40 with $x := m + \alpha y$, we have

$$\varphi_X(t) = \frac{1}{\pi}\int_{-\infty}^\infty e^{itx}\frac{\alpha}{\alpha^2+(x-m)^2}dx = e^{itm}\frac{1}{\pi}\int_{-\infty}^\infty e^{i(\alpha t)y}\frac{1}{1+y^2}dy,$$

and hence it suffices to show (4.48) for $m = 0$ and $\alpha = 1$, that is, for any $t \in \mathbb{R}$,

$$\varphi_{\mathrm{Cauchy}(0,1)}(t) = \frac{1}{\pi}\int_{-\infty}^\infty e^{itx}\frac{1}{1+x^2}dx = e^{-|t|}. \tag{4.49}$$

(4.49) could be proved in a similar way to Examples 4.22 and 4.23 but the calculation would be rather involved. We give a proof of (4.49) later as an application of the Fourier inversion formula (Theorem 4.29).

In fact, a law on $\mathbb{R}^d$ is uniquely determined by its characteristic function, as stated in the following theorem. In the rest of this section, we closely follow [1, Sections 9.5 and 9.8].

Theorem 4.25 (Uniqueness theorem). *If $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and $\varphi_\mu = \varphi_\nu$, then $\mu = \nu$.*

We need to prepare two lemmas. Let I_d denote the $d \times d$ identity matrix.

Lemma 4.26. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$, let $v \in (0, \infty)$ and set $N(0, vI_d) := N(0, v)^d := N(0, v) \times \cdots \times N(0, v)$ (d-fold product). Then $\mu * N(0, vI_d)$ has a density $\rho_\mu^{(v)}$ which is $[0, (2\pi v)^{-d/2}]$-valued and given by*

$$\rho_\mu^{(v)}(x) := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi_\mu(t) \exp\bigl(-i\langle t, x\rangle - |t|^2 v/2\bigr) dt. \tag{4.50}$$

Proof. If $\{X_k\}_{k=1}^d$ are i.i.d. real random variables with $X_1 \sim N(0, v)$, which exist by Theorem 3.28, then the law of $(X_1, \dots, X_d)$ is $N(0, vI_d)$ and, by Theorem 3.29, has a density ρ_v given by

$$\begin{aligned} \rho_v(x) &= \frac{e^{-x_1^2/(2v)}}{\sqrt{2\pi v}} \cdots \frac{e^{-x_d^2/(2v)}}{\sqrt{2\pi v}} \left(= \frac{e^{-|x|^2/(2v)}}{(2\pi v)^{d/2}} \right) \\ &= \frac{1}{(2\pi v)^{d/2}} \int_{-\infty}^{\infty} e^{ix_1 t_1} \frac{e^{-t_1^2 v/2}}{\sqrt{2\pi/v}} dt_1 \cdots \int_{-\infty}^{\infty} e^{ix_d t_d} \frac{e^{-t_d^2 v/2}}{\sqrt{2\pi/v}} dt_d \\ &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{i\langle t, x\rangle - |t|^2 v/2} dt \quad \text{for each } x = (x_1, \dots, x_d) \in \mathbb{R}^d. \end{aligned} \tag{4.51}$$

Here the second equality is due to $e^{-x^2/(2v)} = (2\pi/v)^{-1/2} \int_{-\infty}^{\infty} e^{ixt} e^{-t^2 v/2} dt$ which follows by (4.44) with $m = 0$ and $1/v$ in place of v, and for the third equality we used Problem 2.11-(2). Therefore by Proposition 3.38, $\mu * N(0, vI_d)$ has a density $\rho_\mu^{(v)}$ given by $\rho_\mu^{(v)}(x) := \int_{\mathbb{R}^d} \rho_v(x - y)\mu(dy)$, so that $\rho_\mu^{(v)}(x) \in [0, (2\pi v)^{-d/2}]$ for any $x \in \mathbb{R}$ by (4.51), and

$$\begin{aligned} \rho_\mu^{(v)}(x) &= \int_{\mathbb{R}^d} \rho_v(x - y)\mu(dy) \\ &= \int_{\mathbb{R}^d} \rho_v(y - x)\mu(dy) \\ &= \int_{\mathbb{R}^d} \left(\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{i\langle t, y - x\rangle - |t|^2 v/2} dt \right) \mu(dy) \\ &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} e^{i\langle t, y - x\rangle - |t|^2 v/2} \mu(dy) \right) dt \\ &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \left(e^{-i\langle t, x\rangle - |t|^2 v/2} \int_{\mathbb{R}^d} e^{i\langle t, y\rangle} \mu(dy) \right) dt \end{aligned}$$

$$= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle t,x\rangle - |t|^2 v/2} \varphi_\mu(t)dt,$$

where the second equality is due to $\rho_v(-x) = \rho_v(x)$ and we used Fubini's theorem (Theorem 2.29-(2)) for the fourth equality. Thus the proof is complete. □

Lemma 4.27. *For each $v \in (0,\infty)$ let $N(0, vI_d) := N(0,v)^d$, as in Lemma 4.26. If $\mu \in \mathcal{P}(\mathbb{R}^d)$, then $\mu * N(0, n^{-2}I_d) \xrightarrow{\mathcal{L}} \mu$.*

Proof. Let X, Y be independent d-dimensional random variables with $X \sim \mu$ and $Y \sim N(0, I_d)$, which exist by Theorem 3.28. Then for each $n \in \mathbb{N}$, $\{X, n^{-1}Y\}$ is independent by Proposition 3.31-(2), and Theorem 3.16 and (4.51) easily imply $n^{-1}Y \sim N(0, n^{-2}I_d)$. Therefore $X + n^{-1}Y \sim \mu * N(0, n^{-2}I_d)$ by Proposition 3.36, and since $X + n^{-1}Y \xrightarrow{\text{a.s.}} X$ we have $X + n^{-1}Y \xrightarrow{\mathcal{L}} X$ by (3.47) of Theorem 3.51, that is, $\mu * N(0, n^{-2}I_d) = \mathcal{L}(X + n^{-1}Y) \xrightarrow{\mathcal{L}} \mathcal{L}(X) = \mu$. □

Proof of Theorem 4.25. If $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and $\varphi_\mu = \varphi_\nu$, then $\mu * N(0, n^{-2}I_d) = \nu * N(0, n^{-2}I_d)$ $(=: \lambda_n)$ for any $n \in \mathbb{N}$ since they have the same density by Lemma 4.26, and $\lambda_n \xrightarrow{\mathcal{L}} \mu$ and $\lambda_n \xrightarrow{\mathcal{L}} \nu$ by Lemma 4.27. Hence $\mu = \nu$ by Corollary 4.7. □

Corollary 4.28. *Let $n \in \mathbb{N}$, and let $d_k \in \mathbb{N}$ and let X_k be a d_k-dimensional random variable for each $k \in \{1, \ldots, n\}$. Then $\{X_k\}_{k=1}^n$ is independent if and only if for any $t_k \in \mathbb{R}^{d_k}$, $k \in \{1, \ldots, n\}$,*

$$\varphi_{(X_1,\ldots,X_n)}(t_1, \ldots, t_n) = \varphi_{X_1}(t_1) \cdots \varphi_{X_n}(t_n). \tag{4.52}$$

Proof. Set $\mu := \mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_n}$. By Problem 2.11 (for $\mathbb{C}$-valued measurable functions), for any $t_k \in \mathbb{R}^{d_k}$, $k \in \{1, \ldots, n\}$, we have

$$\begin{aligned}
\varphi_\mu(t_1, \ldots, t_n) &= \int_{\mathbb{R}^{\sum_{k=1}^n d_k}} e^{i\langle (t_1,\ldots,t_n), x\rangle} \mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_n}(dx) \\
&= \int_{\mathbb{R}^{d_1}} e^{i\langle t_1, x_1\rangle} \mathbb{P}_{X_1}(dx_1) \cdots \int_{\mathbb{R}^{d_n}} e^{i\langle t_n, x_n\rangle} \mathbb{P}_{X_n}(dx_n) \\
&= \varphi_{X_1}(t_1) \cdots \varphi_{X_n}(t_n).
\end{aligned}$$

Thus (4.52) holds for any $t_k \in \mathbb{R}^{d_k}$, $k \in \{1, \ldots, n\}$, if and only if $\varphi_{(X_1,\ldots,X_n)} = \varphi_\mu$. On the other hand, Theorem 4.25 implies that $\varphi_{(X_1,\ldots,X_n)} = \varphi_\mu$ if and only if $\mathbb{P}_{(X_1,\ldots,X_n)} = \mu = \mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_n}$, that is, $\{X_k\}_{k=1}^n$ is independent. Hence the assertion follows. □

By using Lemmas 4.26 and 4.27, we can also prove the following theorem.

Theorem 4.29 (Fourier inversion formula)**.** *If $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\int_{\mathbb{R}^d} |\varphi_\mu(t)|dt < \infty$, then μ has a density ρ which is $[0,\infty)$-valued, continuous and given by*

$$\rho(x) := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi_\mu(t) e^{-i\langle t,x\rangle} dt. \tag{4.53}$$

Proof. Since $\left|\varphi_\mu(t)e^{-i\langle t,x\rangle}\right| = |\varphi_\mu(t)|$ and $\int_{\mathbb{R}^d} |\varphi_\mu(t)|dt < \infty$, the continuity of ρ given by (4.53) easily follows by the dominated convergence theorem (Theorem 1.32) in exactly the same way as the proof of (φ3) in Proposition 4.14. Set $\rho_n := \rho_\mu^{(n^{-2})}$ for $n \in \mathbb{N}$, where $\rho_\mu^{(v)}$ is as in (4.50) of Lemma 4.26. Then for $n \in \mathbb{N}$ and for any $x \in \mathbb{R}^d$,

$$\begin{aligned}|\rho(x) - \rho_n(x)| &= \left|\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \varphi_\mu(t)e^{-i\langle t,x\rangle}\bigl(1 - e^{-|t|^2n^{-2}/2}\bigr)dt\right| \\ &\le \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} |\varphi_\mu(t)|\bigl(1 - e^{-|t|^2n^{-2}/2}\bigr)dt =: a_n.\end{aligned} \tag{4.54}$$

Since $\lim_{n\to\infty} a_n = 0$ by $\int_{\mathbb{R}^d} |\varphi_\mu(t)|dt < \infty$ and the dominated convergence theorem (Theorem 1.32), (4.54) yields $\rho(x) = \lim_{n\to\infty} \rho_n(x) \ge 0$ for any $x \in \mathbb{R}^d$. Moreover, if $f : \mathbb{R}^d \to \mathbb{R}$ is continuous and satisfies $f|_{\mathbb{R}^d\setminus[-N,N]^d} = 0$ for some $N \in \mathbb{N}$, then noting that $f\rho$ is m_d-integrable by $\sup_{x\in\mathbb{R}^d} |\rho(x)| \le (2\pi)^{-d}\int_{\mathbb{R}^d} |\varphi_\mu(t)|dt$, from Lemmas 4.27, 4.26 and (4.54) we see that

$$\begin{aligned}&\left|\int_{\mathbb{R}^d} f(x)\rho(x)dx - \int_{\mathbb{R}^d} f(x)\mu(dx)\right| \\ &= \lim_{n\to\infty}\left|\int_{\mathbb{R}^d} f(x)\rho(x)dx - \int_{\mathbb{R}^d} f(x)\bigl(\mu * N(0, n^{-2}I_d)\bigr)(dx)\right| \\ &= \lim_{n\to\infty}\left|\int_{\mathbb{R}^d} f(x)\rho(x)dx - \int_{\mathbb{R}^d} f(x)\rho_n(x)dx\right| \le \limsup_{n\to\infty} a_n \int_{\mathbb{R}^d} |f(x)|dx = 0,\end{aligned}$$

so that $\int_{\mathbb{R}^d} f(x)\rho(x)dx = \int_{\mathbb{R}^d} f(x)\mu(dx)$. In particular, for any $k \in \mathbb{N}$ we have $\int_{\mathbb{R}^d} (\min\{k - |x|, 1\})^+\rho(x)dx = \int_{\mathbb{R}^d} (\min\{k - |x|, 1\})^+\mu(dx)$, and letting $k \to \infty$ yields $\int_{\mathbb{R}^d} \rho(x)dx = \mu(\mathbb{R}^d) = 1$ by the monotone convergence theorem (Theorem 1.24). It follows that μ and $\nu(dx) := \rho(x)dx$ are laws on $\mathbb{R}^d$ satisfying the assumption of Proposition 4.6, and hence $\mu = \nu$, i.e. $\mu(dx) = \rho(x)dx$. □

Example 4.30. As an application of the Fourier inversion formula (Theorem 4.29), we show that the characteristic function of Cauchy(0, 1) is given by

$$\varphi_{\mathrm{Cauchy}(0,1)}(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} e^{itx}\frac{1}{1+x^2}dx = e^{-|t|} \tag{4.49}$$

for any $t \in \mathbb{R}$, which completes the proof of (4.48) in Example 4.24. Indeed, Let $\mu \in \mathcal{P}(\mathbb{R})$ be the *Laplace distribution* (or *double exponential distribution*), defined by

$$\mu(dx) := \frac{1}{2}e^{-|x|}dx. \tag{4.55}$$

The reader will see in Problem 4.6 that for any $t \in \mathbb{R}$,

$$\varphi_\mu(t) = \frac{1}{1+t^2}. \tag{4.56}$$

Then $\int_{-\infty}^{\infty} |\varphi_\mu(t)|dt = \pi < \infty$ and hence Theorem 4.29 applies to μ to imply that μ has a $[0,\infty)$-valued density given by $\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1}{1+t^2}e^{-itx}dt$, which is nothing but

$\varphi_{\mathrm{Cauchy}(0,1)}(-x)/2$. Since μ has another density $e^{-|x|}/2$, Proposition 3.13 shows that $e^{-|x|} = \varphi_{\mathrm{Cauchy}(0,1)}(-x)$ for m_1-a.e. $x \in \mathbb{R}$, but if $\left|e^{-|x|} - \varphi_{\mathrm{Cauchy}(0,1)}(-x)\right|$ were not 0 at some point $x_0 \in \mathbb{R}$ then by its continuity it would be strictly positive on $(x_0 - \delta, x_0 + \delta)$ for some $\delta \in (0, \infty)$, contradicting the fact that it is 0 m_1-a.e. Thus $e^{-|-x|} = e^{-|x|} = \varphi_{\mathrm{Cauchy}(0,1)}(-x)$ for any $x \in \mathbb{R}$, proving (4.49).

At the last of this section, we study the relation between convergence of laws and pointwise convergence of characteristic functions. First, the definition of convergence of laws easily implies the following lemma.

Lemma 4.31. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$, let $\{\mu_n\}_{n=1}^{\infty} \subset \mathcal{P}(\mathbb{R}^d)$ and suppose $\mu_n \xrightarrow{\mathcal{L}} \mu$. Then* $\lim_{n\to\infty} \varphi_{\mu_n}(t) = \varphi_{\mu}(t)$ *for any $t \in \mathbb{R}^d$.*

Proof. For each $t \in \mathbb{R}^d$, $\mathbb{R}^d \ni x \mapsto \cos\langle t, x\rangle$ and $\mathbb{R}^d \ni x \mapsto \sin\langle t, x\rangle$ are bounded continuous functions on $\mathbb{R}^d$ and hence $\mu_n \xrightarrow{\mathcal{L}} \mu$ yields

$$\begin{aligned}\varphi_{\mu_n}(t) &= \int_{\mathbb{R}^d} \cos\langle t, x\rangle \mu_n(dx) + i \int_{\mathbb{R}^d} \sin\langle t, x\rangle \mu_n(dx) \\ &\xrightarrow{n\to\infty} \int_{\mathbb{R}^d} \cos\langle t, x\rangle \mu(dx) + i \int_{\mathbb{R}^d} \sin\langle t, x\rangle \mu(dx) = \varphi_{\mu}(t),\end{aligned}$$

completing the proof. □

An important feature of characteristic functions is that the converse implication of Lemma 4.31 is also true. In fact, we can prove an even stronger assertion, as follows.

Theorem 4.32 (Lévy's continuity theorem)**.** *Let $\{\mu_n\}_{n=1}^{\infty} \subset \mathcal{P}(\mathbb{R}^d)$ and suppose that the limit* $\lim_{n\to\infty} \varphi_{\mu_n}(t) =: \varphi(t)$ *exists for any $t \in \mathbb{R}^d$. If φ is* continuous at 0 along each coordinate axis*, i.e.* $\lim_{t\to 0} \varphi(te_k) = \varphi(0)$ *for any $k \in \{1, \ldots, d\}$, where $e_k := \left(\mathbf{1}_{\{k\}}(\ell)\right)_{\ell=1}^{d} \in \mathbb{R}^d$, then there exists $\mu \in \mathcal{P}(\mathbb{R}^d)$ such that $\varphi = \varphi_{\mu}$ and $\mu_n \xrightarrow{\mathcal{L}} \mu$.*

In Theorem 4.32, the assumption of the continuity of φ at 0 can NOT be dropped, as shown in the following example.

Example 4.33. For each $n \in \mathbb{N}$, let μ_n be the uniform distribution $\mathrm{Unif}(-n, n)$ on $[-n, n]$. Then by (4.42) of Example 4.20, we see that for any $t \in \mathbb{R}$,

$$\varphi_{\mu_n}(t) = \frac{\sin nt}{nt} \xrightarrow{n\to\infty} \mathbf{1}_{\{0\}}(t).$$

Thus the limit $\lim_{n\to\infty} \varphi_{\mu_n}(t)$ exists for any $t \in \mathbb{R}$ but the limit function $\mathbf{1}_{\{0\}}$ is not continuous at 0 and hence cannot be the characteristic function of a law on $\mathbb{R}$.

The proof of Theorem 4.32 requires the following lemma.

Lemma 4.34 (Truncation inequality)**.** *Let $\mu \in \mathcal{P}(\mathbb{R})$. Then for any $\varepsilon \in (0, \infty)$,*

$$\mu\bigl(\{x \in \mathbb{R} \mid |x| \geq 1/\varepsilon\}\bigr) \leq \frac{8}{\varepsilon} \int_0^{\varepsilon} \Bigl(1 - \mathrm{Re}\bigl(\varphi_{\mu}(t)\bigr)\Bigr) dt. \tag{4.57}$$

Proof. Recall that $|(\sin t)/t| \leq 1$ for any $t \in \mathbb{R}$, where $(\sin 0)/0 := 1$. Note that $(\sin t)/t \leq 7/8$ if $|t| \geq 1$; indeed, it suffices to verify this for $t \in [1, \infty)$ since $\bigl(\sin(-t)\bigr)/(-t) = (\sin t)/t$, and $(\sin t)/t \leq 1/t \leq 2/\pi < 7/8$ for $t \in [\pi/2, \infty)$. For $[1, \pi/2)$ we have $\frac{d}{dt}\bigl((\sin t)/t\bigr) = \cos t(t - \tan t)/t^2 \leq 0$ and hence $(\sin t)/t \leq \sin 1 \leq \sin(\pi/3) = \sqrt{3}/2 < 7/8$.

Now for any $\varepsilon \in (0, \infty)$, by using Fubini's theorem (Theorem 2.29-(1)) we see that

$$\begin{aligned}
&\frac{1}{\varepsilon}\int_0^{\varepsilon}\Bigl(1 - \mathrm{Re}\bigl(\varphi_\mu(t)\bigr)\Bigr)dt = \frac{1}{\varepsilon}\int_0^{\varepsilon}\biggl(\int_{\mathbb{R}}\bigl(1 - \cos(tx)\bigr)\mu(dx)\biggr)dt \\
&= \frac{1}{\varepsilon}\int_{\mathbb{R}}\biggl(\int_0^{\varepsilon}\bigl(1 - \cos(tx)\bigr)dt\biggr)\mu(dx) = \int_{\mathbb{R}}\biggl(1 - \frac{\sin(\varepsilon x)}{\varepsilon x}\biggr)\mu(dx) \\
&\geq \int_{\{|x| \geq 1/\varepsilon\}}\biggl(1 - \frac{\sin(\varepsilon x)}{\varepsilon x}\biggr)\mu(dx) \geq \frac{1}{8}\mu\bigl(\{x \in \mathbb{R} \mid |x| \geq 1/\varepsilon\}\bigr),
\end{aligned}$$

proving (4.57). □

Proof of Theorem 4.32. We first prove that $\{\mu_n\}_{n=1}^{\infty}$ is *tight* (recall Theorem 4.12), i.e.

$$\lim_{N \to \infty}\sup_{n \in \mathbb{N}}\mu_n\bigl(\mathbb{R}^d \setminus [-N, N]^d\bigr) = 0. \tag{4.58}$$

Let $k \in \{1, \ldots, d\}$ and define $X_k : \mathbb{R}^d \to \mathbb{R}$ by $X_k(x_1, \ldots, x_d) := x_k$. Then for any $\nu \in \mathcal{P}(\mathbb{R}^d)$, X_k can be considered as a real random variable on $\bigl(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \nu\bigr)$, and then $\varphi_\nu(te_k) = \int_{\mathbb{R}^d} e^{itX_k(x)}\nu(dx)$ is its characteristic function and hence Lemma 4.34 implies that for any $\varepsilon \in (0, \infty)$,

$$\nu(|X_k| \geq 1/\varepsilon) \leq \frac{8}{\varepsilon}\int_0^{\varepsilon}\Bigl(1 - \mathrm{Re}\bigl(\varphi_\nu(te_k)\bigr)\Bigr)dt. \tag{4.59}$$

Let $\varepsilon \in (0, \infty)$. By $\lim_{t \to 0}\varphi(te_k) = \varphi(0)$ and $\varphi(0) = \lim_{n \to \infty}\varphi_{\mu_n}(0) = 1$, we can choose $\delta \in (0, \infty)$ so that $|1 - \varphi(te_k)| \leq \varepsilon/(9d)$ for any $t \in [-\delta, \delta]$. Then by (4.59), $0 \leq 1 - \mathrm{Re}\bigl(\varphi_{\mu_n}(te_k)\bigr) \leq 2$ and the dominated convergence theorem (Theorem 1.32),

$$\begin{aligned}
&\mu_n(|X_k| \geq 1/\delta) \leq \frac{8}{\delta}\int_0^{\delta}\Bigl(1 - \mathrm{Re}\bigl(\varphi_{\mu_n}(te_k)\bigr)\Bigr)dt \\
&\xrightarrow{n \to \infty} \frac{8}{\delta}\int_0^{\delta}\Bigl(1 - \mathrm{Re}\bigl(\varphi(te_k)\bigr)\Bigr)dt \leq \frac{8}{\delta}\int_0^{\delta}|1 - \varphi(te_k)|dt \leq \frac{8\varepsilon}{9d} < \frac{\varepsilon}{d},
\end{aligned}$$

and therefore we can choose $\ell \in \mathbb{N}$ so that for any $n \in \mathbb{N}$ with $n \geq \ell$,

$$\mu_n(|X_k| \geq 1/\delta) \leq \frac{8}{\delta}\int_0^{\delta}\Bigl(1 - \mathrm{Re}\bigl(\varphi_{\mu_n}(te_k)\bigr)\Bigr)dt < \frac{\varepsilon}{d}. \tag{4.60}$$

For each $n \in \{1, \ldots, \ell - 1\}$, $\lim_{j \to \infty}\mu_n(|X_k| \geq j) = 0$ and hence $\mu_n(|X_k| \geq j_n) < \varepsilon/d$ for some $j_n \in \mathbb{N}$, which and (4.60) imply that $\mu_n(|X_k| \geq M_k) < \varepsilon/d$ for any

$n \in \mathbb{N}$, where $M_k := \max\{1/\delta, j_1, \dots, j_{\ell-1}\}$. Now choosing such $M_k \in (0,\infty)$ for each $k \in \{1,\dots,d\}$, for $N \in \mathbb{N}$ with $N \geq \max\{M_1,\dots,M_d\}$ we see that

$$\mu_n\big(\mathbb{R}^d \setminus [-N,N]^d\big) = \mu_n\left(\bigcup_{k=1}^{d}\{|X_k| > N\}\right) \leq \sum_{k=1}^{d} \mu_n(|X_k| \geq M_k) < \varepsilon$$

for any $n \in \mathbb{N}$, so that $\sup_{n\in\mathbb{N}} \mu_n\big(\mathbb{R}^d \setminus [-N,N]^d\big) \leq \varepsilon$, proving (4.58).

Now since $\{\mu_n\}_{n=1}^{\infty}$ is tight, by Theorem 4.12 there exist $\mu \in \mathcal{P}(\mathbb{R}^d)$ and a strictly increasing sequence $\{m(j)\}_{j=1}^{\infty} \subset \mathbb{N}$ such that $\mu_{m(j)} \xrightarrow{\mathcal{L}} \mu$, and then $\varphi_\mu(t) = \lim_{j\to\infty} \varphi_{\mu_{m(j)}}(t) = \varphi(t)$ for any $t \in \mathbb{R}^d$ by Lemma 4.31. Moreover, for any strictly increasing sequence $\{n(k)\}_{k=1}^{\infty} \subset \mathbb{N}$, Theorem 4.12 also implies that there exist $\nu \in \mathcal{P}(\mathbb{R}^d)$ and a strictly increasing sequence $\{k(\ell)\}_{\ell=1}^{\infty} \subset \mathbb{N}$ such that $\mu_{n(k(\ell))} \xrightarrow{\mathcal{L}} \nu$, but then $\varphi_\nu(t) = \lim_{\ell\to\infty} \varphi_{\mu_{n(k(\ell))}}(t) = \varphi(t) = \varphi_\mu(t)$ for any $t \in \mathbb{R}^d$ by Lemma 4.31 and hence $\nu = \mu$ by Theorem 4.25, so that $\mu_{n(k(\ell))} \xrightarrow{\mathcal{L}} \mu$. Thus μ and $\{\mu_n\}_{n=1}^{\infty}$ satisfy the condition of Proposition 4.8 and hence $\mu_n \xrightarrow{\mathcal{L}} \mu$. □

Corollary 4.35. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\{\mu_n\}_{n=1}^{\infty} \subset \mathcal{P}(\mathbb{R}^d)$. Then $\mu_n \xrightarrow{\mathcal{L}} \mu$ if and only if $\lim_{n\to\infty} \varphi_{\mu_n}(t) = \varphi_\mu(t)$ for any $t \in \mathbb{R}^d$.*

Proof. "only if" part has been already verified in Lemma 4.31. Conversely suppose $\lim_{n\to\infty} \varphi_{\mu_n}(t) = \varphi_\mu(t)$ for any $t \in \mathbb{R}^d$. Since φ_μ is continuous, by Lévy's continuity theorem (Theorem 4.32) there exists $\nu \in \mathcal{P}(\mathbb{R}^d)$ such that $\varphi_\mu = \varphi_\nu$ and $\mu_n \xrightarrow{\mathcal{L}} \nu$. Then $\mu = \nu$ by Theorem 4.25 and hence $\mu_n \xrightarrow{\mathcal{L}} \mu$. □

As an application of Lévy's continuity theorem (Theorem 4.32), at the last of this section we prove the following theorem, which asserts that the properties (φ1), (φ3) and (φ4) in Proposition 4.14 characterize characteristic functions of laws on $\mathbb{R}^d$.

Theorem 4.36 (Bochner's theorem)**.** *Let $\varphi : \mathbb{R}^d \to \mathbb{C}$ be continuous and satisfy $\varphi(0) = 1$. Suppose that φ is* non-negative definite*, that is, for any $n \in \mathbb{N}$, $\{z_k\}_{k=1}^{n} \subset \mathbb{C}$ and $\{t_k\}_{k=1}^{n} \subset \mathbb{R}^d$,*

$$\sum_{k,\ell=1}^{n} \varphi(t_k - t_\ell) z_k \overline{z_\ell} \geq 0. \tag{4.61}$$

Then there exists $\mu \in \mathcal{P}(\mathbb{R}^d)$ such that $\varphi = \varphi_\mu$.

We need the following lemma for the proof of Theorem 4.36.

Lemma 4.37. *Let $\varphi : \mathbb{R}^d \to \mathbb{C}$ be non-negative definite. Then $\varphi(0) \in [0,\infty)$, and $|\varphi(t)| \leq \varphi(0)$ and $\varphi(-t) = \overline{\varphi(t)}$ for any $t \in \mathbb{R}^d$.*

Proof. (4.61) with $n = 1$, $z_1 = 1$ and $t_1 = 0$ shows $\varphi(0) \geq 0$. Let $t \in \mathbb{R}^d$ and $z \in \mathbb{C}$. Then setting $n = 2$, $z_1 = -z$, $z_2 = 1$, $t_1 = t$ and $t_2 = 0$ in (4.61), we obtain

$$(1 + |z|^2)\varphi(0) - \varphi(t)z - \varphi(-t)\overline{z} \geq 0. \tag{4.62}$$

Set $u := \varphi(t) + \varphi(-t)$ and $v := -i\varphi(t) + i\varphi(-t)$, so that $u, v \in \mathbb{R}$ by (4.62) with $z = 1, i$. Solving these equalities in $\varphi(t), \varphi(-t)$ yields $\varphi(t) = (u + iv)/2$ and $\varphi(-t) = (u - iv)/2$ and therefore $\varphi(-t) = \overline{\varphi(t)}$. Finally, if $\varphi(t) = 0$ then $|\varphi(t)| \leq \varphi(0)$, and otherwise (4.62) with $z := \overline{\varphi(t)}/|\varphi(t)|$ and $\varphi(-t) = \overline{\varphi(t)}$ together show that

$$2\varphi(0) \geq \varphi(t)\frac{\overline{\varphi(t)}}{|\varphi(t)|} + \overline{\varphi(t)}\frac{\varphi(t)}{|\varphi(t)|} = 2|\varphi(t)|,$$

proving $|\varphi(t)| \leq \varphi(0)$. □

Proof of Theorem 4.36. Let $g : \mathbb{R}^d \to \mathbb{C}$ be continuous and satisfy $\int_{\mathbb{R}^d} |g(x)|dx < \infty$. Then $\mathbb{R}^{2d} \ni (t, s) \mapsto \varphi(t - s)g(t)\overline{g(s)}$ is continuous and therefore for each $N \in \mathbb{N}$, Theorem 2.35 and (4.61) imply that

$$\begin{aligned}\int_{[-N,N]^{2d}} &\varphi(t - s)g(t)\overline{g(s)}dtds \\ &= \lim_{n\to\infty} \sum_{\alpha,\beta\in\mathbb{N}^d\cap(-nN,nN]^d} \varphi\left(\frac{\alpha}{n} - \frac{\beta}{n}\right)g\left(\frac{\alpha}{n}\right)\overline{g\left(\frac{\beta}{n}\right)}\left(\frac{1}{n}\right)^{2d} \geq 0. \quad (4.63)\end{aligned}$$

Lemma 4.37 and $\varphi(0) = 1$ imply $\left|\varphi(t - s)g(t)\overline{g(s)}\right| \leq |g(t)||g(s)|$, and by Problem 2.11-(2) we have $\int_{\mathbb{R}^{2d}} |g(t)||g(s)|dtds = \left(\int_{\mathbb{R}^d} |g(t)|dt\right)^2 < \infty$. Therefore the dominated convergence theorem (Theorem 1.32) applies to the left-hand side of (4.63) and yields

$$\int_{\mathbb{R}^{2d}} \varphi(t - s)g(t)\overline{g(s)}dtds = \lim_{N\to\infty} \int_{[-N,N]^{2d}} \varphi(t - s)g(t)\overline{g(s)}dtds \geq 0. \quad (4.64)$$

In particular, for $n \in \mathbb{N}$ and $x \in \mathbb{R}^d$, by setting $g(t) := (\pi^3 n)^{-d/4} \exp\left(-i\langle t, x\rangle - 2|t|^2/n\right)$ in (4.64) and using Corollary 2.40 and Fubini's theorem (Theorem 2.29-(2)), we obtain

$$\begin{aligned}0 &\leq \int_{\mathbb{R}^{2d}} \varphi(t - s)g(t)\overline{g(s)}dtds = \int_{\mathbb{R}^{2d}} \varphi(t)g(t + s)\overline{g(s)}dtds \\ &= \int_{\mathbb{R}^d} \left(\varphi(t) \int_{\mathbb{R}^d} g(t + s)\overline{g(s)}ds\right)dt \\ &= \int_{\mathbb{R}^d} \left(\varphi(t)(\pi^3 n)^{-d/2}e^{-i\langle t,x\rangle} \int_{\mathbb{R}^d} \exp\left(-\frac{2}{n}\left(|t + s|^2 + |s|^2\right)\right)ds\right)dt \\ &= \int_{\mathbb{R}^d} \left(\varphi(t)(\pi^3 n)^{-d/2}e^{-i\langle t,x\rangle-|t|^2/n} \int_{\mathbb{R}^d} \exp\left(-\frac{4}{n}\left|s + \frac{t}{2}\right|^2\right)ds\right)dt \\ &= \int_{\mathbb{R}^d} \varphi(t)(\pi^3 n)^{-d/2}e^{-i\langle t,x\rangle-|t|^2/n}(\pi n/4)^{d/2}dt \\ &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi(t)e^{-i\langle t,x\rangle-|t|^2/n}dt =: \rho_n(x). \quad (4.65)\end{aligned}$$

Note that, similarly to (φ3) in Proposition 4.14, the function $\rho_n : \mathbb{R}^d \to [0, \infty)$ defined by (4.65) is continuous by virtue of the dominated convergence theorem (Theorem 1.32).

Next we prove that $\int_{\mathbb{R}^d} \rho_n(x)dx = 1$. Using Fubini's theorem (Theorem 2.29-(2)), we see that for any $(s_1, \ldots, s_d) \in (0, \infty)^d$,

$$h_n(s_1, \ldots, s_d) := \int_{[-s_1,s_1]\times\cdots\times[-s_d,s_d]} \rho_n(x)dx \quad (4.66)$$

$$= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \left(\varphi(t) e^{-|t|^2/n} \int_{[-s_1,s_1]\times\cdots\times[-s_d,s_d]} e^{-i\langle t,x\rangle} dx \right) dt$$
$$= \frac{1}{\pi^d} \int_{\mathbb{R}^d} \varphi(t_1,\ldots,t_d) e^{-|(t_1,\ldots,t_d)|^2/n} \frac{\sin(s_1 t_1)}{t_1} \cdots \frac{\sin(s_d t_d)}{t_d} d(t_1,\ldots,t_d).$$

Note that $h_n : (0,\infty)^d \to [0,\infty)$ defined by (4.66) is again continuous by the dominated convergence theorem (Theorem 1.32). For any $s \in (0,\infty)$, $\{h_n(Ns)\}_{N=1}^{\infty}$ is clearly non-decreasing, and $\int_{\mathbb{R}^d} \rho_n(x)dx = \lim_{N\to\infty} h_n(Ns)$ by the monotone convergence theorem (Theorem 1.24). Hence by using the monotone convergence theorem, Fubini's theorem (Theorem 2.29), Corollary 2.40 and the dominated convergence theorem (Theorem 1.32), we obtain

$$\int_{\mathbb{R}^d} \rho_n(x)dx = \lim_{N\to\infty} \int_{(0,1)^d} h_n(Ns)ds$$
$$= \lim_{N\to\infty} \frac{1}{\pi^d} \int_{\mathbb{R}^d} \varphi(t_1,\ldots,t_d) e^{-|(t_1,\ldots,t_d)|^2/n} \frac{1-\cos(Nt_1)}{Nt_1^2} \cdots \frac{1-\cos(Nt_d)}{Nt_d^2} d(t_1,\ldots,t_d)$$
$$= \lim_{N\to\infty} \frac{1}{\pi^d} \int_{\mathbb{R}^d} \varphi\left(\frac{t_1}{N},\ldots,\frac{t_d}{N}\right) e^{-|(t_1,\ldots,t_d)|^2/(nN^2)} \frac{1-\cos t_1}{t_1^2} \cdots \frac{1-\cos t_d}{t_d^2} d(t_1,\ldots,t_d)$$
$$= \frac{1}{\pi^d} \int_{\mathbb{R}^d} \varphi(0) \frac{1-\cos t_1}{t_1^2} \cdots \frac{1-\cos t_d}{t_d^2} d(t_1,\ldots,t_d) = \frac{1}{\pi^d} \left(\int_{-\infty}^{\infty} \frac{1-\cos t}{t^2} dt \right)^d = 1,$$

where the last equality is due to Problem 2.14-(2).

Let $\mu_n(dx) := \rho_n(x)dx$, so that $\mu_n \in \mathcal{P}(\mathbb{R}^d)$ by the previous paragraph. We claim that for any $t \in \mathbb{R}^d$,

$$\varphi_{\mu_n}(t) = \int_{\mathbb{R}^d} \rho_n(x) e^{i\langle t,x\rangle} dx = \varphi(t) e^{-|t|^2/n}. \tag{4.67}$$

Indeed, let $u = (u_1,\ldots,u_d) \in \mathbb{R}^d$. Similarly to (4.66), using Fubini's theorem (Theorem 2.29-(2)), for $(s_1,\ldots,s_d) \in (0,\infty)^d$ we have

$$f_n(s_1,\ldots,s_d) := \int_{[-s_1,s_1]\times\cdots\times[-s_d,s_d]} \rho_n(x) e^{i\langle u,x\rangle} dx \tag{4.68}$$
$$= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \left(\varphi(t) e^{-|t|^2/n} \int_{[-s_1,s_1]\times\cdots\times[-s_d,s_d]} e^{i\langle u-t,x\rangle} dx \right) dt$$
$$= \frac{1}{\pi^d} \int_{\mathbb{R}^d} \varphi(t_1,\ldots,t_d) e^{-|(t_1,\ldots,t_d)|^2/n} \prod_{k=1}^{d} \frac{\sin\big(s_k(u_k-t_k)\big)}{u_k-t_k} d(t_1,\ldots,t_d).$$

By virtue of the dominated convergence theorem (Theorem 1.32), $f_n : (0,\infty)^d \to \mathbb{C}$ defined by (4.68) is continuous, and for any $s \in (0,\infty)^d$, $|f_n(s)| \le \int_{\mathbb{R}^d} \rho_n(x)dx = 1$ and $\varphi_{\mu_n}(u) = \int_{\mathbb{R}^d} \rho_n(x) e^{i\langle u,x\rangle} dx = \lim_{N\to\infty} f_n(Ns)$. Then again by the dominated convergence theorem and by using Fubini's theorem (Theorem 2.29) and Corollary 2.40, we obtain

$$\varphi_{\mu_n}(u) = \lim_{N\to\infty} \int_{(0,1)^d} f_n(Ns)ds$$
$$= \lim_{N\to\infty} \frac{1}{\pi^d} \int_{\mathbb{R}^d} \varphi(t_1,\ldots,t_d) e^{-|(t_1,\ldots,t_d)|^2/n} \prod_{k=1}^{d} \frac{1-\cos\big(N(u_k-t_k)\big)}{N(u_k-t_k)^2} d(t_1,\ldots,t_d)$$
$$= \lim_{N\to\infty} \frac{1}{\pi^d} \int_{\mathbb{R}^d} \varphi\left(u_1+\frac{t_1}{N},\ldots,u_d+\frac{t_d}{N}\right) \prod_{k=1}^{d} e^{-(u_k+t_k/N)^2/n} \frac{1-\cos t_k}{t_k^2} d(t_1,\ldots,t_d)$$

$$= \frac{1}{\pi^d} \int_{\mathbb{R}^d} \varphi(u) e^{-|u|^2/n} \prod_{k=1}^{d} \frac{1-\cos t_k}{t_k^2} d(t_1, \dots, t_d)$$
$$= \frac{1}{\pi^d} \varphi(u) e^{-|u|^2/n} \left(\int_{-\infty}^{\infty} \frac{1-\cos t}{t^2} dt \right)^d = \varphi(u) e^{-|u|^2/n},$$

proving (4.67), where the last equality is due to Problem 2.14-(2).

Now by (4.67), we have $\lim_{n\to\infty} \varphi_{\mu_n}(t) = \varphi(t)$ for any $t \in \mathbb{R}^d$, and φ is continuous by the assumption. Hence it follows by Lévy's continuity theorem (Theorem 4.32) that $\varphi = \varphi_\mu$ for some $\mu \in \mathcal{P}(\mathbb{R}^d)$. □

4.3 Central Limit Theorem

Based on the properties of characteristic functions established in the previous section, now we prove the central limit theorem (Theorem 4.4).

Lemma 4.38. *Let $z \in \mathbb{C}$ and $\{z_n\}_{n=1}^{\infty} \subset \mathbb{C}$. If $\lim_{n\to\infty} z_n = z$, then*

$$\lim_{n\to\infty} (1 + z_n/n)^n = e^z. \tag{4.69}$$

Proof. Since $\lim_{n\to\infty} z_n = z$, we can choose $N \in \mathbb{N}$ so that $|z - z_n| \le 1$ for any $n \in \mathbb{N}$ with $n \ge N$. Let $n \in \mathbb{N}$ satisfy $n \ge N$. Then

$$(1 + z_n/n)^n = \sum_{k=0}^{n} \binom{n}{k} \frac{z_n^k}{n^k} = \sum_{k=0}^{n} \frac{n!}{n^k (n-k)!} \frac{z_n^k}{k!}. \tag{4.70}$$

Set $f_n(k) := \frac{n!}{n^k(n-k)!} \frac{z_n^k}{k!} \mathbf{1}_{\{0,\dots,n\}}(k)$ for $k \in \mathbb{N} \cup \{0\}$. Then $\lim_{n\to\infty} f_n(k) = z^k/k!$, $|f_n(k)| \le |z_n|^k/k! \le (|z|+1)^k/k!$, and (4.70) means $(1 + z_n/n)^n = \int_{\mathbb{N}\cup\{0\}} f_n d\#$, where $\#$ denotes the counting measure on $\mathbb{N} \cup \{0\}$. Since $\int_{\mathbb{N}\cup\{0\}} (|z|+1)^k/k! d\#(k) = e^{|z|+1} < \infty$, the dominated convergence theorem (Theorem 1.32) applied to $\{f_n\}_{n=N}^{\infty}$ yields

$$(1 + z_n/n)^n = \int_{\mathbb{N}\cup\{0\}} f_n d\# \xrightarrow{n\to\infty} \int_{\mathbb{N}\cup\{0\}} \frac{z^k}{k!} d\#(k) = e^z,$$

proving (4.69). □

Lemma 4.39. *Let $\{X_n\}_{n=1}^{\infty} \subset \mathcal{L}^2(\mathbb{P})$ be i.i.d. Set $m := \mathbb{E}[X_1]$, $v := \mathrm{var}(X_1)$ and $S_n := \sum_{k=1}^{n} X_k$ for each $n \in \mathbb{N}$. Then for any $t \in \mathbb{R}$,*

$$\lim_{n\to\infty} \mathbb{E}\left[\exp\left(it \frac{S_n - nm}{\sqrt{n}}\right)\right] = \exp\bigl(-t^2 v/2\bigr). \tag{4.71}$$

Proof. Set $f := \varphi_{X_1 - m}$. By Theorem 4.15, f' and f'' are continuous on $\mathbb{R}$, $f(0) = 1$, $f'(0) = i\mathbb{E}[X_1 - m] = 0$ and $f''(0) = -\mathbb{E}[(X_1 - m)^2] = -v$. (4.71) is clear if $t = 0$. Let $t \in \mathbb{R} \setminus \{0\}$. Since $\{X_n - m\}_{n=1}^{\infty}$ is i.i.d., by using Proposition 4.18 and (4.34) of Lemma 4.17 we see that

$$\mathbb{E}\left[\exp\left(it \frac{S_n - nm}{\sqrt{n}}\right)\right] = \varphi_{S_n - nm}\bigl(t/\sqrt{n}\bigr) = \varphi_{X_1 - m}\bigl(t/\sqrt{n}\bigr) \cdots \varphi_{X_n - m}\bigl(t/\sqrt{n}\bigr)$$

$$= f\big(t/\sqrt{n}\big)^n = \left(1 - \frac{v}{2}\frac{t^2}{n} + o\Big(\frac{t^2}{n}\Big)\right)^n = \left(1 + \frac{1}{n}\Big(-\frac{t^2 v}{2} + t^2\frac{o(t^2/n)}{t^2/n}\Big)\right)^n$$

as $n \to \infty$, and therefore it converges to $e^{-t^2 v/2}$ as $n \to \infty$ by Lemma 4.38. □

Proof of Theorem 4.4. (1) By Lemma 4.39, for any $t \in \mathbb{R}$, the characteristic function $\mathbb{E}\big[e^{it(S_n - nm)/\sqrt{n}}\big]$ of $\mathcal{L}\big(\frac{S_n-nm}{\sqrt{n}}\big)$ converges to $e^{-t^2 v/2}$, which is the characteristic function of $N(0,v)$ by Example 4.22. Thus $\mathcal{L}\big(\frac{S_n-nm}{\sqrt{n}}\big) \xrightarrow{\mathcal{L}} N(0,v)$ by Corollary 4.35.
(2) Set $\mu_n := \mathcal{L}\big(\frac{S_n-nm}{\sqrt{n}}\big)$ for each $n \in \mathbb{N}$ and let $x \in \mathbb{R}$. Since $(-\infty, x)$ is open in $\mathbb{R}$, $(-\infty, x]$ is closed in $\mathbb{R}$ and $v > 0$, from $\mathcal{L}\big(\frac{S_n-nm}{\sqrt{n}}\big) \xrightarrow{\mathcal{L}} N(0,v)$ and Theorem 4.10-(2),(3) we see that

$$\int_{-\infty}^{x} \frac{e^{-y^2/(2v)}}{\sqrt{2\pi v}}\,dy \le \liminf_{n\to\infty} \mu_n\big((-\infty,x)\big) \le \begin{matrix} \limsup_{n\to\infty} \mu_n\big((-\infty,x)\big) & \text{or} \\ \liminf_{n\to\infty} \mu_n\big((-\infty,x]\big) & \end{matrix}$$
$$\le \limsup_{n\to\infty} \mu_n\big((-\infty,x]\big) \le \int_{-\infty}^{x} \frac{e^{-y^2/(2v)}}{\sqrt{2\pi v}}\,dy,$$

which is valid for either choice of the third term. Thus we get $\lim_{n\to\infty} \mu_n\big((-\infty,x]\big) = \lim_{n\to\infty} \mu_n\big((-\infty,x)\big) = (2\pi v)^{-1/2}\int_{-\infty}^{x} e^{-y^2/2v}\,dy$, which was to be proved. □

In fact, Theorem 4.4-(1) is generalized to i.i.d. *d-dimensional* random variables. We need some preparations to state and prove that generalization.

Definition 4.40 (Covariance matrix)**.** Let $X = (X_1, \ldots, X_d)$ be a d-dimensional random variable with $\mathbb{E}[|X|^2] < \infty$. Then the real $d \times d$ matrix $V = (v_{jk})_{j,k=1}^d$ given by $v_{jk} := \mathrm{cov}(X_j, X_k)$ is called the *covariance matrix of* X.

The covariance matrix $V = (v_{jk})_{j,k=1}^d$ of such a d-dimensional random variable $X = (X_1, \ldots, X_d)$ is clearly *symmetric*, i.e. $v_{jk} = v_{kj}$ for any $j,k \in \{1,\ldots,d\}$. Moreover, it is *non-negative definite*, i.e.

$$\langle a, Va\rangle = \sum_{j,k=1}^{d} v_{jk} a_j a_k \ge 0 \qquad \text{for any } a = (a_1,\ldots,a_d) \in \mathbb{R}^d, \tag{4.72}$$

where $Va \in \mathbb{R}^d$ is the matrix product of V and a with a regarded as a *column* vector. Indeed,

$$\langle a, Va\rangle = \sum_{j,k=1}^{d} \mathrm{cov}(X_j, X_k) a_j a_k = \mathbb{E}\left[\sum_{j,k=1}^{d} \big(X_j - \mathbb{E}[X_j]\big)\big(X_k - \mathbb{E}[X_k]\big) a_j a_k\right]$$
$$= \mathbb{E}\left[\left(\sum_{k=1}^{d} a_k\big(X_k - \mathbb{E}[X_k]\big)\right)^2\right] \ge 0.$$

Theorem 4.41 (Normal distribution on $\mathbb{R}^d$). *Let $m \in \mathbb{R}^d$ and let V be a non-negative definite real symmetric $d \times d$ matrix. Then there exists a unique $\mu \in \mathcal{P}(\mathbb{R}^d)$ such that*

$$\varphi_\mu(t) = \exp\bigl(i\langle t, m\rangle - \langle t, Vt\rangle/2\bigr) \qquad \text{for any } t \in \mathbb{R}^d. \tag{4.73}$$

Moreover, if X is a d-dimensional random variable with $X \sim \mu$, then $\mathbb{E}[|X|^2] < \infty$, $\mathbb{E}[X] = m$, and the covariance matrix of X is V.

The law μ in Theorem 4.41 is denoted by $N(m, V)$ and called the *d-dimensional normal* (or *Gaussian*) *distribution with mean m and covariance matrix V*. Note that the notation $N(m, V)$ is consistent with $N(0, vI_d) = N(0, v)^d$ introduced for $v \in [0, \infty)$ in Lemma 4.26 since, by Corollary 4.28, for any $t = (t_1, \dots, t_d) \in \mathbb{R}^d$ we have

$$\varphi_{N(0,vI_d)}(t) = \varphi_{X_1}(t_1) \cdots \varphi_{X_d}(t_d) = e^{-t_1^2 v/2} \cdots e^{-t_d^2 v/2} = e^{-\langle t, vI_d t\rangle/2}, \tag{4.74}$$

where $\{X_k\}_{k=1}^d$ are i.i.d. d-dimensional random variables with $X_1 \sim N(0, v)$.

Proof of Theorem 4.41. Recall the following basic fact from linear algebra: since V is a real symmetric $d \times d$ matrix, there exist a real *orthogonal* $d \times d$ matrix U (i.e. a real $d \times d$ matrix satisfying $U^*U = UU^* = I_d$) and $(\lambda_1, \dots, \lambda_d) \in \mathbb{R}^d$ such that

$$U^*VU = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_d \end{pmatrix} =: D(\lambda_1, \dots, \lambda_d). \tag{4.75}$$

Then for each $k \in \{1, \dots, d\}$, $\lambda_k = \langle e_k, U^*VUe_k\rangle = \langle Ue_k, VUe_k\rangle \geq 0$, where $e_k := \bigl(\mathbf{1}_{\{k\}}(\ell)\bigr)_{\ell=1}^d \in \mathbb{R}^d$, and hence we can define $W := UD\bigl(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d}\bigr)U^*$, so that $W^* = W$ and $W^2 = UD(\lambda_1, \dots, \lambda_d)U^* = UU^*VUU^* = V$.

Now let Y be a d-dimensional random variable with $Y \sim N(0, I_d)$. Then by using Proposition 4.19 and (4.74) we see that for any $t \in \mathbb{R}^d$,

$$\varphi_{WY+m}(t) = e^{i\langle t, m\rangle}\varphi_Y(W^*t) = e^{i\langle t, m\rangle}e^{-\langle Wt, Wt\rangle/2} = e^{i\langle t, m\rangle - \langle t, Vt\rangle/2},$$

and therefore $\mu := \mathcal{L}(WY + m)$ satisfies (4.73). Moreover, if $\nu \in \mathcal{P}(\mathbb{R}^d)$ also satisfies (4.73), then $\varphi_\mu = \varphi_\nu$ and hence $\mu = \nu$ by Theorem 4.25, proving the uniqueness of μ.

Finally, for the latter assertion let $X = (X_1, \dots, X_d)$ be a d-dimensional random variable with $X \sim \mu$. For $k \in \{1, \dots, d\}$, since $\mathbb{R} \ni t \mapsto \varphi_\mu(te_k)$ is the characteristic function of X_k and has continuous second derivative, $\mathbb{E}[X_k^2] < \infty$ by Theorem 4.16, so that $\mathbb{E}[|X|^2] = \sum_{\ell=1}^d \mathbb{E}[X_\ell^2] < \infty$. Writing $m = (m_1, \dots, m_d)$ and $V = (v_{jk})_{j,k=1}^d$, for any $j, k \in \{1, \dots, d\}$ and $t = (t_1, \dots, t_d) \in \mathbb{R}^d$, by (4.33) of Theorem 4.15 we get

$$\frac{\partial \varphi_\mu}{\partial t_k}(t) = \left(im_k - \sum_{\ell=1}^d v_{k\ell}t_\ell\right)\varphi_\mu(t), \qquad \mathbb{E}[X_k] = -i\frac{\partial \varphi_\mu}{\partial t_k}(0) = m_k,$$

$$\operatorname{cov}(X_j, X_k) = \mathbb{E}[X_jX_k] - \mathbb{E}[X_j]\mathbb{E}[X_k] = -\frac{\partial^2 \varphi_\mu}{\partial t_j \partial t_k}(0) - m_jm_k = v_{jk},$$

completing the proof. $\square$

Proposition 4.42. *Let $m \in \mathbb{R}^d$ and let V be a non-negative definite real symmetric $d \times d$ matrix. Then $N(m, V)$ has a density if and only if V is invertible, and in this case*

$$N(m,V)(dx) = \frac{1}{(2\pi)^{d/2}\sqrt{\det V}} \exp\biggl(-\frac{\langle x-m, V^{-1}(x-m)\rangle}{2}\biggr)dx. \tag{4.76}$$

Proof. Let W be as in the above proof of Theorem 4.41, so that $W^* = W$ and $W^2 = V$. Let Y be a d-dimensional random variable with $Y \sim N(0, I_d)$, so that $N(m, V) = \mathcal{L}(WY + m)$ by above proof of Theorem 4.41. Recall that Y has a density ρ_1 given by $\rho_1(x) = (2\pi)^{-d/2}e^{-|x|^2/2}$ ((4.51) with $v = 1$).

Define $f : \mathbb{R}^d \to \mathbb{R}^d$ by $f(y) := Wy + m$. Note that $(\det W)^2 = \det(W^2) = \det V$. If V is invertible, then $\det W \neq 0$, so that W is invertible and f is a C^1-embedding with $f^{-1}(x) = W^{-1}(x - m)$. Therefore by Theorem 3.16, $N(m, V) = \mathcal{L}(f(Y))$ has a density ρ given by

$$\begin{aligned}
\rho(x) = \rho_1\bigl(f^{-1}(x)\bigr)\bigl|\det f^{-1}(x)\bigr| &= \frac{1}{(2\pi)^{d/2}} \exp\Biggl(-\frac{\bigl|W^{-1}(x-m)\bigr|^2}{2}\Biggr)\bigl|\det\bigl(W^{-1}\bigr)\bigr| \\
&= \frac{1}{(2\pi)^{d/2}|\det W|} \exp\biggl(-\frac{\langle W^{-1}(x-m), W^{-1}(x-m)\rangle}{2}\biggr) \\
&= \frac{1}{(2\pi)^{d/2}\sqrt{\det V}} \exp\biggl(-\frac{\langle x-m, V^{-1}(x-m)\rangle}{2}\biggr),
\end{aligned}$$

proving (4.76), where we used $\det\bigl(W^{-1}\bigr) = (\det W)^{-1}$ for the equality in the second line and $(W^{-1})^* = W^{-1}$ for the equality in the third line. ($(W^{-1})^* = W^{-1}$ holds since $W^*(W^{-1})^* = (W^{-1}W)^* = I_d^* = I_d$ and hence $(W^{-1})^* = (W^*)^{-1} = W^{-1}$.)

Finally, suppose V is not invertible, so that W is not invertible by $W^2 = V$. Then we have $\mathrm{m}_d(W(\mathbb{R}^d)) = 0$ by Remark 2.39-(2) and therefore $\mathrm{m}_d(f(\mathbb{R}^d)) = \mathrm{m}_d(W(\mathbb{R}^d) + m) = \mathrm{m}_d(W(\mathbb{R}^d)) = 0$ by Theorem 2.38-(1).[2] Hence if $N(m, V) = \mathcal{L}(f(Y))$ has a density ρ, then $\mathcal{L}(f(Y))(f(\mathbb{R}^d)) = \int_{f(\mathbb{R}^d)} \rho(x)dx = 0$, which contradicts $\mathcal{L}(f(Y))(f(\mathbb{R}^d)) = \mathbb{P}[f(Y) \in f(\mathbb{R}^d)] = 1$. Thus $N(m, V)$ does not have a density, and the proof is complete. □

Now we establish the central limit theorem for i.i.d. d-dimensional random variables.

Theorem 4.43 (Central limit theorem)**.** *Let $\{X_n\}_{n=1}^\infty$ be i.i.d. d-dimensional random variables with $\mathbb{E}[|X_1|^2] < \infty$. Set $m := \mathbb{E}[X_1](\in \mathbb{R}^d)$, let V be the covariance matrix of X_1 and set $S_n := \sum_{k=1}^n X_k$ for each $n \in \mathbb{N}$. Then*

$$\mathcal{L}\biggl(\frac{S_n - nm}{\sqrt{n}}\biggr) \xrightarrow{\mathcal{L}} N(0, V). \tag{4.77}$$

[2]Note that $W(\mathbb{R}^d)$ and $f(\mathbb{R}^d) = W(\mathbb{R}^d) + m$ are closed (and hence Borel) subsets of $\mathbb{R}^d$, since $W(\mathbb{R}^d)$ is a linear subspace of $\mathbb{R}^d$.

Proof. Let $t \in \mathbb{R}^d$. Then $|\langle t, X_1\rangle|^2 \le |t|^2|X_1|^2$, so that $\{\langle t, X_n\rangle\}_{n=1}^{\infty} \subset \mathcal{L}^2(\mathbb{P})$, and they are clearly i.i.d., where the independence follows by Proposition 3.31-(2). Some easy calculations show $\mathbb{E}[\langle t, X_1\rangle] = \langle t, m\rangle$ and $\mathrm{var}(\langle t, X_1\rangle) = \langle t, Vt\rangle$, and therefore by Lemma 4.39,

$$\mathbb{E}\left[\exp\left(i\left\langle t, \frac{S_n - nm}{\sqrt{n}}\right\rangle\right)\right] = \mathbb{E}\left[\exp\left(i\frac{\sum_{k=1}^{n}\langle t, X_k\rangle - n\langle t, m\rangle}{\sqrt{n}}\right)\right]$$
$$\xrightarrow{n\to\infty} \exp\bigl(-\langle t, Vt\rangle/2\bigr) = \varphi_{N(0,V)}(t).$$

Thus the characteristic function $\mathbb{E}\bigl[e^{i\langle t,(S_n-nm)/\sqrt{n}\rangle}\bigr]$ of $\mathcal{L}\bigl(\frac{S_n-nm}{\sqrt{n}}\bigr)$ converges to the characteristic function $\varphi_{N(0,V)}(t)$ of $N(0,V)$ for any $t \in \mathbb{R}^d$, and it follows from Corollary 4.35 that $\mathcal{L}\bigl(\frac{S_n-nm}{\sqrt{n}}\bigr) \xrightarrow{\mathcal{L}} N(0,V)$. □

Exercises

In the problems and the exercises below, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a probability space and all random variables are assumed to be defined on $(\Omega, \mathcal{F}, \mathbb{P})$.

Problem 4.1. Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. real random variables with $X_1 \sim \mathrm{Po}(1)$, and set $S_n := \sum_{k=1}^{n} X_k$ for each $n \in \mathbb{N}$. Prove the following statements:

(1) $\mathcal{L}\left(\dfrac{S_n - n}{\sqrt{n}}\right) \xrightarrow{\mathcal{L}} N(0,1)$. (Simply apply Theorem 4.4-(1).)

(2) $\mathbb{P}[S_n \le n] = e^{-n}\displaystyle\sum_{k=0}^{n}\frac{n^k}{k!}$ for any $n \in \mathbb{N}$. (Use Exercise 3.18.)

(3) $\displaystyle\lim_{n\to\infty} e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!} = \frac{1}{2}$. (Theorem 4.4-(2) applies by (2) above.)

Problem 4.2. Let $y \in \mathbb{R}$ and let $X, \{X_n\}_{n=1}^{\infty}, \{Y_n\}_{n=1}^{\infty}$ be real random variables such that

$$X_n \xrightarrow{\mathcal{L}} X \quad \text{and} \quad Y_n \xrightarrow{\mathrm{P}} y. \tag{4.78}$$

(1) Prove that $X_n + Y_n \xrightarrow{\mathcal{L}} X + y$ and that $X_n Y_n \xrightarrow{\mathcal{L}} yX$. (Since $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, y)$ by Proposition 4.11, Corollary 3.53-(3) applies to (X, y) and $\{(X_n, Y_n)\}_{n=1}^{\infty}$.)
(2) Suppose $y \neq 0$. Prove that

$$\frac{X_n}{Y_n}\mathbf{1}_{\{Y_n \neq 0\}} \xrightarrow{\mathcal{L}} \frac{X}{y}. \tag{4.79}$$

(Use Exercise 3.25 to apply the latter assertion of (1).)

Remark. Note that in the statements of Problem 4.2, the random variable X is involved *only in terms of its law* $\mathcal{L}(X)$ since the laws of $X + y, yX, X/y$ are determined solely by $\mathcal{L}(X)$ and y. In particular, the statements of Problem 4.2 are valid even if X is replaced by another real random variable X_0 with $\mathcal{L}(X_0) = \mathcal{L}(X)$ which is defined on a *different* probability space.

Exercise 4.3 ([2, Exercise 3.4.4]). Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. $[0,\infty)$-valued random variables with $\mathbb{E}[X_1]=1$ and $v := \mathrm{var}(X_1) < \infty$. Set $S_n := \sum_{k=1}^{n} X_k$ for each $n \in \mathbb{N}$.
(1) Prove that for any $n \in \mathbb{N}$,

$$\sqrt{S_n} - \sqrt{n} = \frac{S_n - n}{\sqrt{n}} \frac{1}{1+\sqrt{S_n/n}}. \tag{4.80}$$

(2) Prove that

$$\mathcal{L}\big(\sqrt{S_n} - \sqrt{n}\big) \xrightarrow{\mathcal{L}} N(0, v/4). \tag{4.81}$$

((4.81) can be rephrased as "$\sqrt{S_n} - \sqrt{n} \xrightarrow{\mathcal{L}} Z/2$" for a real random variable Z with $Z \sim N(0,v)$, and Theorem 4.4-(1) can be also rephrased in the same way. Apply this version of Theorem 4.4-(1) to $(S_n - n)/\sqrt{n}$ and then use (4.80) and the latter part of Problem 4.2-(1), noting that it is irrelevant on which probability space Z is defined.)

Problem 4.4 ([2, Exercise 3.4.5]). Let $\{X_n\}_{n=1}^{\infty} \subset \mathcal{L}^2(\mathbb{P})$ be i.i.d. with $\mathbb{E}[X_1]=0$ and $v := \mathrm{var}(X_1) > 0$. Prove that

$$\mathcal{L}\left(\frac{\sum_{k=1}^{n} X_k}{\sqrt{\sum_{k=1}^{n} X_k^2}} \mathbf{1}_{\{\sum_{k=1}^{n} X_k^2 \neq 0\}} \right) \xrightarrow{\mathcal{L}} N(0,1). \tag{4.82}$$

($\big(\sum_{k=1}^{n} X_k\big)/\sqrt{\sum_{k=1}^{n} X_k^2} = \big(\frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k\big)/\sqrt{\frac{1}{n}\sum_{k=1}^{n} X_k^2}$ on $\big\{\sum_{k=1}^{n} X_k^2 \neq 0\big\}$. Similarly to Exercise 4.3-(2), use Theorem 4.4-(1) and Problem 4.2-(2).)

Exercise 4.5. Verify the assertions (1), (2), (3) and (4) of Example 4.20.

Problem 4.6. Let $\mu \in \mathcal{P}(\mathbb{R})$ be the *Laplace distribution*, that is, the law on $\mathbb{R}$ given by

$$\mu(dx) := \frac{1}{2} e^{-|x|} dx. \tag{4.55}$$

(μ is also called the *double exponential distribution*.) Prove that for any $t \in \mathbb{R}$,

$$\varphi_\mu(t) = \frac{1}{1+t^2}. \tag{4.56}$$

(The result for $\mathrm{Exp}(\alpha)$ in Example 4.21 can be used with $\alpha = 1$.)

For Problem 4.7 and Exercises 4.8, 4.9 and 4.10 below, recall Proposition 4.18 and Examples 4.20, 4.22, 4.23 and 4.24. Note also the following immediate corollary of Theorem 4.25:

Corollary. *Let $d \in \mathbb{N}$, $\mu \in \mathcal{P}(\mathbb{R}^d)$ and let X be a d-dimensional random variable. If $\varphi_X = \varphi_\mu$ then $X \sim \mu$.*

Problem 4.7. (1) (Problem 3.13) Let X, Y be independent real random variables with $X \sim N(m_1, v_1)$ and $Y \sim N(m_2, v_2)$. Prove that $X+Y \sim N(m_1+m_2, v_1+v_2)$. (Use Proposition 4.18 and (4.44) of Example 4.22 to show that $\varphi_{X+Y} = \varphi_{N(m_1+m_2, v_1+v_2)}$.)
(2) (Exercise 3.14) Let $n \in \mathbb{N}$, and let $\{X_k\}_{k=1}^{n}$ be independent real random variables with $X_k \sim N(m_k, v_k)$ for any $k \in \{1,\ldots,n\}$. Set $X := \sum_{k=1}^{n} X_k$, $m := \sum_{k=1}^{n} m_k$ and $v := \sum_{k=1}^{n} v_k$. Prove that $X \sim N(m,v)$. (Similarly to (1), verify $\varphi_X = \varphi_{N(m,v)}$.)

Recall that Problem 4.7 already appeared as Problem 3.13 and Exercise 3.14, where some tedious calculations on density functions were necessary. Here the same assertions can be verified rather easily by virtue of Proposition 4.18 and Theorem 4.25. The same argument applies to Poisson, gamma and Cauchy random variables, as follows.

Exercise 4.8 (Exercise 3.18)**.** Let $n \in \mathbb{N}$, and let $\{X_k\}_{k=1}^n$ be independent real random variables with $X_k \sim \mathrm{Po}(\lambda_k)$ for any $k \in \{1, \ldots, n\}$. Set $X := \sum_{k=1}^n X_k$ and $\lambda := \sum_{k=1}^n \lambda_k$. Prove that $X \sim \mathrm{Po}(\lambda)$.

Exercise 4.9. Let $n \in \mathbb{N}$, $\beta \in (0, \infty)$ and let $\{X_k\}_{k=1}^n$ be independent real random variables with $X_k \sim \mathrm{Gamma}(\alpha_k, \beta)$ for any $k \in \{1, \ldots, n\}$. Set $X := \sum_{k=1}^n X_k$ and $\alpha := \sum_{k=1}^n \alpha_k$. Prove that $X \sim \mathrm{Gamma}(\alpha, \beta)$.

Exercise 4.10. Let $n \in \mathbb{N}$, and let $\{X_k\}_{k=1}^n$ be independent real random variables with $X_k \sim \mathrm{Cauchy}(m_k, \alpha_k)$ for any $k \in \{1, \ldots, n\}$. Set $X := \sum_{k=1}^n X_k$, $m := \sum_{k=1}^n m_k$ and $\alpha := \sum_{k=1}^n \alpha_k$. Prove that $X \sim \mathrm{Cauchy}(m, \alpha)$.

Problem 4.11. Let X be a real random variable with $X \sim N(0, 1)$. Calculate $\mathbb{E}[X^n]$ for any $n \in \mathbb{N}$. (Use the Taylor series expansion of $\varphi_X(t) = e^{-t^2/2}$ to apply (4.33) of Theorem 4.15.)

Exercise 4.12. Let $m \in \mathbb{R}$, $v \in [0, \infty)$ and let X be a real random variable with $X \sim N(m, v)$. Prove that $\mathbb{E}\big[e^{sX}\big] = \exp\big(sm + s^2 v/2\big)$ for any $s \in \mathbb{R}$.

Remark. Formally, replacing s by it in Exercise 4.12 yields the characteristic function (4.44) of $N(m, v)$ in Example 4.22, but some task is required to justify this reasoning.

Problem 4.13. Let $d \in \mathbb{N}$ and let X be a d-dimensional random variable.
(1) Prove that $\varphi_{-X}(t) = \overline{\varphi_X(t)}$ for any $t \in \mathbb{R}^d$.
(2) Prove that φ_X is real-valued (i.e. $\varphi_X(t) \in \mathbb{R}$ for any $t \in \mathbb{R}^d$) if and only if $\mathcal{L}(-X) = \mathcal{L}(X)$. (Use (1) and Theorem 4.25. Recall that for $z \in \mathbb{C}$, $z \in \mathbb{R}$ if and only if $\overline{z} = z$.)

Appendix: Examples of Probability Distributions

Name of distribution	Density or weights	Characteristic function
binomial $B(n,p)$ ($n \in \mathbb{N}$, $p \in [0,1]$)	$\mu(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}$	$\varphi_\mu(t) = \big(1 + p(e^{it} - 1)\big)^n$
Poisson Po(λ) ($\lambda \in (0,\infty)$)	$\mu(\{n\}) = e^{-\lambda} \frac{\lambda^n}{n!}$	$\varphi_\mu(t) = \exp\big(\lambda(e^{it} - 1)\big)$
Geometric Geom(α) ($\alpha \in [0,1)$)	$\mu(\{n\}) = (1-\alpha)\alpha^n$	$\varphi_\mu(t) = \frac{1-\alpha}{1-\alpha e^{it}}$
Uniform Unif(a,b) ($a,b \in \mathbb{R}$, $a < b$)	$\mu(dx) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x)dx$	$\varphi_\mu(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}$
Exponential Exp(α) ($\alpha \in (0,\infty)$)	$\mu(dx) = \alpha e^{-\alpha x} \mathbf{1}_{(0,\infty)}(x)dx$	$\varphi_\mu(t) = \frac{\alpha}{\alpha - it}$
Gamma Gamma(α, β) ($\alpha, \beta \in (0,\infty)$)	$\mu(dx) =$ $\frac{\beta^\alpha}{\Gamma(\alpha)} \lvert x\rvert^{\alpha-1} e^{-\beta x} \mathbf{1}_{(0,\infty)}(x)dx$	$\varphi_\mu(t) = \frac{\beta^\alpha}{(\beta - it)^\alpha}$
Normal $N(m,v)$ ($m \in \mathbb{R}$, $v \in [0,\infty)$)	$\mu = \delta_m$ if $v = 0$, otherwise $\mu(dx) = \frac{1}{\sqrt{2\pi v}} \exp\big(-\frac{(x-m)^2}{2v}\big)dx$	$\varphi_\mu(t) = \exp\big(itm - t^2 v/2\big)$
Cauchy Cauchy(m, α) ($m \in \mathbb{R}$, $\alpha \in (0,\infty)$)	$\mu(dx) = \frac{1}{\pi} \frac{\alpha}{\alpha^2 + (x-m)^2} dx$	$\varphi_\mu(t) = \exp\big(itm - \alpha \lvert t\rvert\big)$

Bibliography

[1] R. M. Dudley, *Real Analysis and Probability*, Cambridge studies in advanced math. **74**, Cambridge University Press, Cambridge, 2002.

[2] R. Durrett, *Probability: Theory and Examples*, 4th ed., Cambridge University Press, Cambridge, 2010.

[3] A. Grigor'yan, *Probability Theory*, an unpublished lecture note in Universität Bielefeld, 2010.

[4] J. Jacod and P. Protter, *Probability Essentials*, 2nd ed. (corrected second printing), Universitext, Springer-Verlag, New York - Berlin - Heidelberg, 2004.

[5] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed., Graduate Texts in Math. **113**, Springer-Verlag, New York - Berlin - Heidelberg, 1991.

[6] J. Kigami, *Analysis I*, an unpublished lecture note in Kyoto University, July 3, 2009 (in Japanese).

[7] W. Rudin, *Real and Complex Analysis*, 3rd ed., McGraw-Hill, New York, 1987.