Probability Theory

Naotaka Kajino

WS 2012/2013, Universität Bielefeld

Preface

This is a lecture note for the lecture course "Probability Theory" in the University of Bielefeld (240111, WS 2012/2013).

Several theorems and exercises are adopted from an unpublished lecture note [6] on measure theory by Professor Jun Kigami in Kyoto University, and some other problems are borrowed from an unpublished lecture note by Professor Grigor'yan in the University of Bielefeld.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to express his deepest gratitude toward Professor Kigami and Professor Grigor'yan for their permission to quote their unpublished notes in this lecture note.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

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.

Next, let us throw this dice *infinitely many* times and let X_n be the n-th outcome. From our intuition we naturally expect that

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mathbb{E}[X], \tag{0.1}$$

where $\mathbb{E}[X]$ is the "expectation" ("expected value") or "mean" of the outcome of a trial, given by

$$\mathbb{E}[X] = \sum_{k=1}^{6} k \cdot \mathbb{P}[X = k] = \frac{1 + \cdots + 6}{6} = \frac{7}{2}. \tag{0.2}$$

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

The convergence as in (0.1) is called the *law of large numbers*. This "law" is usually taken for granted, but *why should it be true at all?* At this moment this fact is just an experimental observation, but with a mathematically rigorous formulation of the notions of "probability" and "random variable" we can in fact prove (0.1) as a mathematical theorem!

The purpose of this lecture course is to give such a rigorous formulation of "probability" and prove various probabilistic phenomena like (0.1) as mathematical theorems.

How to formulate "probability" rigorously?

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.
- 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 $\{A_i\}_{i=1}^n \subset \mathcal{F}$[4] 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[5] 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.

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 $i, j \in \{1, \ldots, n\}$ 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 $\mathbb{P}\big[\bigcup_{n=1}^\infty A_n\big] = \sum_{n=1}^\infty \mathbb{P}[A_n]$.

Countable additivity plays significant roles in the proofs of various limit theorems like (0.1) where an *infinite* sequence of random variables should be inevitably involved. 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.

[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] $\{A_i\}_{i=1}^n \subset \mathcal{F}$ means that $\{A_i\}_{i=1}^n$ is a family of elements of $\mathcal{F}$ indexed by $i \in \{1, \ldots, n\}$, or in other words, $A_i \in \mathcal{F}$ for each $i \in \{1, \ldots, n\}$. The notation "$\subset$" is used here since $\{A_i\}_{i=1}^n$ can be considered as a subfamily of $\mathcal{F}$, although it may happen that $A_i = A_j$ for some $i \neq j$.

[5] 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.

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* $\mathbb{P}$:

$$\mathbb{E}[X] = \int_\Omega X d\mathbb{P}. \tag{0.3}$$

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 $\{X_n\}_{n=1}^\infty$,

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X], \qquad \text{that is,} \qquad \lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}\Big[\lim_{n\to\infty} X_n\Big]. \tag{0.4}$$

(0.4) 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 $\mathbb{N} = \{1, 2, 3, \dots\}$.

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. By $\{x_\lambda\}_{\lambda\in\Lambda} \subset X$, where Λ is another set, we mean that $\{x_\lambda\}_{\lambda\in\Lambda}$ is a family of elements of X indexed by $\lambda \in \Lambda$, or in other words, $x_\lambda \in X$ for each $\lambda \in \Lambda$. 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 verify the following facts:

$$\textit{If } n \in \mathbb{N} \textit{ and } \{X_i\}_{i=1}^n \textit{ are countable sets, then } X_1 \times \cdots \times X_n \textit{ is countable.} \tag{0.5}$$

$$\textit{If } A_n \textit{ is a countable set for each } n \in \mathbb{N}, \textit{ then } \textstyle\bigcup_{n=1}^\infty A_n \textit{ is countable.} \tag{0.6}$$

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 distinct elements which are also 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

$$(a, b) := \{x \in [-\infty, \infty] \mid a < x < b\}, \quad [a, b] := \{x \in [-\infty, \infty] \mid a \leq x \leq b\},$$
$$(a, b] := \{x \in [-\infty, \infty] \mid a < x \leq b\}, \quad [a, b) := \{x \in [-\infty, \infty] \mid a \leq x < b\}.$$

(2) We say that a sequence $\{a_n\}_{n=1}^\infty \subset [-\infty, \infty]$ *converges to* ∞ (resp. *to* $-\infty$)[6], 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 $a_n > b$ (resp. $a_n < b$) 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: we write $\lim_{n\to\infty} a_n = a$ if and only if for any $\varepsilon \in (0, \infty)$ there exists $N \in \mathbb{N}$ such that $a_n \in (a - \varepsilon, a + \varepsilon)$ for any $n \geq N$.

Below we state basic definitions and facts concerning $[-\infty, \infty]$.

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

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 in $[-\infty, \infty]$ exist.*[7]

Proposition 0.3. *Let $\{a_n\}_{n=1}^{\infty} \subset [-\infty, \infty]$.*
(1) *If $a_n \leq a_{n+1}$ for any $n \in \mathbb{N}$, then* $\lim_{n\to\infty} a_n = \sup_{n\geq 1} a_n$.
(2) *If $a_n \geq a_{n+1}$ for any $n \in \mathbb{N}$, then* $\lim_{n\to\infty} a_n = \inf_{n\geq 1} a_n$.

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

$$\limsup_{n\to\infty} a_n := \inf_{n\geq 1}\left(\sup_{k\geq n} a_k\right), \qquad \liminf_{n\to\infty} a_n := \sup_{n\geq 1}\left(\inf_{k\geq n} a_k\right). \tag{0.7}$$

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 by Proposition 0.3,

$$\lim_{n\to\infty}\left(\sup_{k\geq n} a_k\right) = \limsup_{n\to\infty} a_n, \qquad \lim_{n\to\infty}\left(\inf_{k\geq n} a_k\right) = \liminf_{n\to\infty} a_n. \tag{0.8}$$

It also holds that

$$\liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n. \tag{0.9}$$

Indeed, $\inf_{k\geq m} a_k \leq a_{\max\{m,n\}} \leq \sup_{k\geq n} a_k$ for any $m, n \in \mathbb{N}$, and taking the infimum of the right-hand side in n shows that $\inf_{k\geq m} a_k \leq \limsup_{n\to\infty} a_n$ for any $m \in \mathbb{N}$. Then taking the supremum of the left-hand side in m shows (0.9).

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

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

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

[7] 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.

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 have the following useful proposition.

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

$$a + 0 = 0 + a = a, \quad a + b = b + a, \quad (a + b) + c = a + (b + c),$$
$$a \cdot 1 = 1 \cdot a = a, \quad ab = ba, \quad (ab)c = a(bc),$$
$$a(b + c) = ab + ac, \quad (a + b)c = ac + bc.$$

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

$$\lim_{n\to\infty}(a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n, \tag{0.10}$$

$$\lim_{n\to\infty} a_n b_n = \Big(\lim_{n\to\infty} a_n\Big)\Big(\lim_{n\to\infty} b_n\Big). \tag{0.11}$$

Remark 0.8. *It also holds that* $a \cdot 1 = 1 \cdot a = a$, $ab = ba$ *and* $(ab)c = a(bc)$ *for any* $a, b, c \in [-\infty, \infty]$. Indeed, these equalities are all immediate from Definition 0.6.

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

$$\sum_{n=1}^{\infty} a_n := \lim_{n\to\infty} \sum_{i=1}^{n} a_i = \sup_{n\in\mathbb{N}} \sum_{i=1}^{n} a_i = \sup_{A \subset \mathbb{N}:\ \text{finite}} \sum_{n\in A} a_n.^{8} \tag{0.12}$$

The equality $\lim_{n\to\infty} \sum_{i=1}^{n} a_i = \sup_{n\in\mathbb{N}} \sum_{i=1}^{n} a_i$ follows by Proposition 0.3-(1). For the third equality of (0.12), $\sum_{i=1}^{k} a_i = \sum_{i\in\{1,\ldots,k\}} a_i \leq \sup_{A \subset \mathbb{N}:\ \text{finite}} \sum_{n\in A} a_n$ for any $k \in \mathbb{N}$ and hence $\sup_{n\in\mathbb{N}} \sum_{i=1}^{n} a_i \leq \sup_{A \subset \mathbb{N}:\ \text{finite}} \sum_{n\in A} a_n$. For the converse inequality, let $A \subset \mathbb{N}$ be non-empty finite and set $k := \max A$. Then $\sum_{n\in A} a_n \leq \sum_{i=1}^{k} a_i \leq \sup_{n\in\mathbb{N}} \sum_{i=1}^{n} a_i$, and hence $\sup_{A \subset \mathbb{N}:\ \text{finite}} \sum_{n\in A} a_n \leq \sup_{n\in\mathbb{N}} \sum_{i=1}^{n} a_i$. Thus the equalities in (0.12) follows.

Note that, by the last equality in (0.12), 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*

$$\sum_{n=1}^{\infty}\sum_{k=1}^{\infty} a_{n,k} = \sum_{k=1}^{\infty}\sum_{n=1}^{\infty} a_{n,k} = \sum_{\ell=1}^{\infty} a_{n_\ell,k_\ell} = \sup_{\substack{A\subset\mathbb{N}\times\mathbb{N}:\\ \text{finite}}} \sum_{(n,k)\in A} a_{n,k} =: \sum_{n,k=1}^{\infty} a_{n,k}. \tag{0.13}$$

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

[8] 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 $x, y \in \mathbb{R}^d$, $x = (x_1, \ldots, x_d)$, $y = (y_1, \ldots, y_d)$,

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

Also for $x \in \mathbb{R}^d$ and $r \in (0, \infty)$ 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 $\mathbb{R}^d$.*

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 $\mathbb{R}^k$.*

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.*

Exercises

Problem 0.1. (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 0.2. Let $\{a_n\}_{n=1}^{\infty}, \{b_n\}_{n=1}^{\infty} \subset [-\infty, \infty]$.
(1) Suppose $a_n \leq b_n$ for any $n \in \mathbb{N}$. Prove that

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

(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

$$\limsup_{n\to\infty}(a_n + b_n) \leq \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \tag{0.14}$$

and that the equality holds in (0.14) 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 (0.14).

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* $A \setminus B \in \mathcal{M}$.

Definition 1.3 (Measures)**.** (1) Let $(X, \mathcal{M})$ be a measurable space. 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 $i, j \in \{1, \dots, n\}$ with $i \neq j$, then $\mu(A_1 \cup \dots \cup A_n) = \mu(A_1) + \dots + \mu(A_n)$.*
(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 $\lim_{n\to\infty} \mu(A_n) = \mu\bigl(\bigcup_{n=1}^{\infty} A_n\bigr)$.*
(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 $\lim_{n\to\infty} \mu(A_n) = \mu\bigl(\bigcap_{n=1}^{\infty} A_n\bigr)$.*

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 probability 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 $\mathcal{A}$.

[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.

$\sigma_X(\mathcal{A})$ 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.

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.

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

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 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* $[a_1, b_1] \times \cdots \times [a_d, b_d]$,

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

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*

$$\mathcal{F}_d := \big\{[a_1, b_1] \times \cdots \times [a_d, b_d] \bigm| a_k, b_k \in \mathbb{R}, a_k \le b_k \text{ for } 1 \le k \le d\big\} \cup \{\emptyset\}, \tag{1.6}$$
$$\mathcal{F}_d^{\mathbb{Q}} := \big\{[a_1, b_1] \times \cdots \times [a_d, b_d] \bigm| a_k, b_k \in \mathbb{Q}, a_k \le b_k \text{ for } 1 \le k \le d\big\} \cup \{\emptyset\}. \tag{1.7}$$

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

The following lemma is sometimes useful.

Lemma 1.10. *Let X be a set and let $Y \subset X$. For $\mathcal{A} \subset 2^X$, 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) *If $\mathcal{A} \subset 2^X$, then $\sigma_Y(\mathcal{A}|_Y) = \sigma_X(\mathcal{A})|_Y$.*

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.

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

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

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

[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.

an application of Lemma 1.10 shows that

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

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 $A_n \times \{0,1\}^{\mathbb{N}\setminus\{1,\dots,n\}}$ with $A_n \subset \{0,1\}^n$, should be measurable. Therefore an easy choice is to consider the following σ-algebra $\mathcal{F}$:

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

$\mathcal{F}$ 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]

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

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

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* $a \in \mathbb{R}$*).*

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.*

[3]The number p corresponds to the probability of heads at each flip.
[4]Here $0^0 := 1$.
[5]that is, provided neither "$\infty + (-\infty)$" nor "$-\infty + \infty$" appears in the sum $f(x) + g(x)$

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

$$\left(\sup_{n\geq 1} f_n\right)(x) := \sup_{n\geq 1}\bigl(f_n(x)\bigr), \qquad \left(\limsup_{n\to\infty} f_n\right)(x) := \limsup_{n\to\infty}\bigl(f_n(x)\bigr),$$
$$\left(\inf_{n\geq 1} f_n\right)(x) := \inf_{n\geq 1}\bigl(f_n(x)\bigr), \qquad \left(\liminf_{n\to\infty} f_n\right)(x) := \liminf_{n\to\infty}\bigl(f_n(x)\bigr).$$

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.*

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

$$\mathbf{1}_E(x) := \begin{cases} 1 & \text{if } x \in E, \\ 0 & \text{if } x \notin E. \end{cases} \tag{1.13}$$

$\mathbf{1}_E$ 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 $E \in \mathcal{M}$.

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

$$s = \sum_{i=1}^n a_i \mathbf{1}_{A_i} \qquad \text{for some } n \in \mathbb{N},\ \{a_i\}_{i=1}^n \subset \mathbb{R} \text{ and } \{A_i\}_{i=1}^n \subset \mathcal{M}. \tag{1.14}$$

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

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

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 $(X, \mathcal{M}, \mu)$.

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

1.3.1 Integration of 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

$$\int_X s d\mu := \sum_{a \in s(X)} a\mu\big(s^{-1}(a)\big). \tag{1.15}$$

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

$$\int_X (\alpha s + \beta t) d\mu = \alpha \int_X s d\mu + \beta \int_X t d\mu. \tag{1.16}$$

Note that $\mathbf{1}_E$ is $\mathcal{M}$-simple and $\int_X \mathbf{1}_E d\mu = \mu(E)$ for any $E \in \mathcal{M}$. Therefore Lemma 1.21 in particular implies that for $n \in \mathbb{N}$, $\{a_i\}_{i=1}^n \subset [0, \infty)$ and $\{A_i\}_{i=1}^n \subset \mathcal{M}$,

$$\int_X \left(\sum_{i=1}^n a_i \mathbf{1}_{A_i} \right) d\mu = \sum_{i=1}^n a_i \mu(A_i). \tag{1.17}$$

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

$$\int_X f d\mu := \sup \left\{ \int_X s d\mu \;\middle|\; s : X \to \mathbb{R},\, s \text{ is } \mathcal{M}\text{-simple and } 0 \leq s \leq f \text{ on } X \right\}. \tag{1.18}$$

Note that (1.18) is consistent with (1.15) for non-negative $\mathcal{M}$-simple functions; indeed, the supremum in (1.18) is attained by f if $f : X \to [0, \infty]$ is itself $\mathcal{M}$-simple, since we see from Lemma 1.21 that $\int_X s d\mu \leq \int_X s d\mu + \int_X (t - s) d\mu = \int_X t d\mu$ for $\mathcal{M}$-simple functions $s, t : X \to [0, \infty)$ with $s \leq t$ on X.

The following lemma is immediate from (1.18).

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*

$$\lim_{n \to \infty} \int_X f_n d\mu = \int_X f d\mu. \tag{1.19}$$

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

$$\int_X (\alpha f + \beta g) d\mu = \alpha \int_X f d\mu + \beta \int_X g d\mu. \tag{1.20}$$

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

$$\int_X \left(\sum_{n=1}^{\infty} f_n\right) d\mu = \sum_{n=1}^{\infty} \int_X f_n d\mu. \tag{1.21}$$

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*

$$\int_X \left(\liminf_{n\to\infty} f_n\right) d\mu \leq \liminf_{n\to\infty} \int_X f_n d\mu. \tag{1.22}$$

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

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

$$f^+(x) := \max\{f(x), 0\} \qquad \text{and} \qquad f^-(x) := -\min\{f(x), 0\}, \tag{1.23}$$

so that $f = f^+ - f^-$ and $|f| = f^+ + f^-$ (recall that 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)**.** (1) For an $\mathcal{M}$-measurable function $f : X \to [-\infty,\infty]$, we say that f *admits the μ-integral* or *the μ-integral of f exists* (or simply *$\int_X f d\mu$ exists*) if and only if

$$\min\left\{\int_X f^+ d\mu, \int_X f^- d\mu\right\} < \infty, \tag{1.24}$$

and in this case its *μ-integral* $\int_X f d\mu$ is defined by

$$\int_X f d\mu := \int_X f^+ d\mu - \int_X f^- d\mu. \tag{1.25}$$

Moreover, f is called *μ-integrable* if and only if $\int_X |f| d\mu < \infty$. Finally, we set

$$\mathcal{L}^1(X,\mathcal{M},\mu) := \{f : X \to \mathbb{R} \mid f \text{ is } \mathcal{M}\text{-measurable and } \mu\text{-integrable}\}, \tag{1.26}$$

which will be simply written as $\mathcal{L}^1(X,\mu)$ or $\mathcal{L}^1(\mu)$ when no confusion can occur.
(2) Let $A \in \mathcal{M}$. For an $\mathcal{M}$-measurable function $f : X \to [-\infty,\infty]$, we say that f *admits the μ-integral on A* or *the μ-integral of f on A exists* (or simply *$\int_A f d\mu$ exists*) if and only if $\int_X f \mathbf{1}_A d\mu$ exists, and in this case its *μ-integral $\int_X f d\mu$ on A* is defined by $\int_A f d\mu := \int_X f \mathbf{1}_A d\mu$. Moreover, f is called *μ-integrable on A* if and only if $f\mathbf{1}_A$ is μ-integrable.

Note that (1.25) is consistent with (1.18) for non-negative functions, since $f^+ = f$ and $f^- = 0$ for $\mathcal{M}$-measurable $f : X \to [0,\infty]$. Note also that for $A \in \mathcal{M}$, f is μ-integrable on A if and only if $\int_A f d\mu$ exists 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.

$$\int_A f(x)d\mu(x) := \int_A f(x)\mu(dx) := \int_A f d\mu. \tag{1.27}$$

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

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 $0 \leq s \leq |f|\mathbf{1}_A$ on X. Then for any $a \in s(X) \setminus \{0\}$, $s^{-1}(a) \subset A$ and hence $\mu\big(s^{-1}(a)\big) = 0$. Thus $\int_X s d\mu = 0$ for any such s and therefore $\int_X |f|\mathbf{1}_A d\mu = 0$.
(2) Set $A := f^{-1}(\infty) \cup f^{-1}(-\infty)$ and let $n \in \mathbb{N}$. Then $|f| \geq |f|\mathbf{1}_A \geq n\mathbf{1}_A$ on X and hence $n\mu(A) = \int_X n\mathbf{1}_A d\mu \leq \int_X |f| d\mu < \infty$. Thus $0 \leq \mu(A) \leq n^{-1} \int_X |f| d\mu$, and letting $n \to \infty$ yields $\mu(A) = 0$. $\square$

Proposition 1.31. (1) *If $f, g : X \to [-\infty, \infty]$ are $\mathcal{M}$-measurable, $f \leq g$ on X and $\int_X f d\mu, \int_X g d\mu$ exist, then*

$$\int_X f d\mu \leq \int_X g d\mu. \tag{1.28}$$

In particular, if $f : X \to [-\infty, \infty]$ is $\mathcal{M}$-measurable and $\int_X f d\mu$ exists, then

$$\left|\int_X f d\mu\right| \leq \int_X |f| d\mu. \tag{1.29}$$

(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*

$$\int_X (\alpha f + \beta g) d\mu = \alpha \int_X f d\mu + \beta \int_X g d\mu. \tag{1.30}$$

The following proposition says that sets of μ-measure zero are in fact negligible as long as μ-integrals are concerned. Note that we have $\{x \in X \mid f(x) \neq g(x)\} \in \mathcal{M}$ for $\mathcal{M}$-measurable functions $f, g : X \to [-\infty, \infty]$; see Problem 1.15-(1).

Proposition 1.32. *Let $f, g : X \to [-\infty, \infty]$ be $\mathcal{M}$-measurable and suppose that $\mu\big(\{x \in X \mid f(x) \neq g(x)\}\big) = 0$. Then for any $A \in \mathcal{M}$, $\int_A f d\mu$ exists if and only if $\int_A g d\mu$ exists, and in this case*

$$\int_A f d\mu = \int_A g d\mu. \tag{1.31}$$

The following convergence theorem often plays fundamental roles in analysis.

[7] The first and second notations in (1.27) have exactly the same meaning, but for certain reasons the second notation is often preferred in the context of probability theory.

Theorem 1.33 (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 $|f_n(x)| \le g(x)$ 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

$$\lim_{n\to\infty} \int_X f_n d\mu = \int_X f d\mu. \tag{1.32}$$

Note that $\sum_{n=1}^{\infty} a_n = \int_{\mathbb{N}} a_n d\#(n)$ for any $\{a_n\}_{n=1}^{\infty} \subset [0, \infty]$ by Problem 1.19, where # denotes the counting measure on $\mathbb{N}$ defined in Example 1.5-(1), so that *all the results established so far in this section are applicable to such series $\sum_{n=1}^{\infty} a_n$.*

Example 1.34. As an application of the dominated convergence theorem (Theorem 1.33), for $\alpha, \beta \in \mathbb{R}$ with $\alpha + \beta > 2$ let us verify the limit

$$\lim_{N\to\infty} \sum_{n=1}^{\infty} \frac{N}{n^\alpha + N^2 n^\beta} = 0. \tag{1.33}$$

For any $n \in \mathbb{N}$, we have

$$\frac{N}{n^\alpha + N^2 n^\beta} = \frac{1}{N} \frac{1}{n^\alpha N^{-2} + n^\beta} \xrightarrow{N\to\infty} 0 \cdot \frac{1}{n^\beta} = 0, \tag{1.34}$$

$$0 < \frac{N}{n^\alpha + N^2 n^\beta} = \left(\frac{n^\alpha}{N} + N n^\beta\right)^{-1} \le \frac{1}{2n^{(\alpha+\beta)/2}}, \tag{1.35}$$

where we used $a + b \ge 2\sqrt{ab}$, $a, b \in [0, \infty)$[8], for the inequality in (1.35). Now since

$$\left(\int_{\mathbb{N}} \frac{1}{2n^{(\alpha+\beta)/2}} d\#(n) =\right) \sum_{n=1}^{\infty} \frac{1}{2n^{(\alpha+\beta)/2}} < \infty \tag{1.36}$$

by $\alpha + \beta > 2$, the dominated convergence theorem (Theorem 1.33) together with (1.34), (1.35) and (1.36) implies that

$$\lim_{N\to\infty} \sum_{n=1}^{\infty} \frac{N}{n^\alpha + N^2 n^\beta} = \sum_{n=1}^{\infty} 0 = 0$$

(in other words, $\lim_{N\to\infty} \int_{\mathbb{N}} \frac{N}{n^\alpha + N^2 n^\beta} d\#(n) = \int_{\mathbb{N}} 0 d\#(n) = 0$), proving (1.33).

Note that (1.33) also holds if $\beta > 1$ instead of $\alpha + \beta > 2$, since $\sum_{n=1}^{\infty} n^{-\beta} < \infty$ and hence

$$0 < \sum_{n=1}^{\infty} \frac{N}{n^\alpha + N^2 n^\beta} < \frac{1}{N} \sum_{n=1}^{\infty} \frac{1}{n^\beta} \xrightarrow{N\to\infty} 0 \cdot \sum_{n=1}^{\infty} \frac{1}{n^\beta} = 0.$$

[8]This inequality is valid since $a + b - 2\sqrt{ab} = (\sqrt{a} - \sqrt{b})^2 \ge 0$.

1.3.3 Sets of measure zero and completion of measure spaces

In the above proof of Theorem 1.33, 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 1.35 (Almost everywhere, a.e.)**.** Let $\mathbf{P}(x)$ be a statement 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 "$f(x) = 0$" or "$f(x) = g(x)$" for given functions $f, g : X \to [-\infty, \infty]$, 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 by the following proposition.

Proposition 1.36. (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 $f, g : X \to [-\infty, \infty]$ are $\mathcal{M}$-measurable, μ-integrable and satisfy $\int_A f d\mu = \int_A g d\mu$ for any $A \in \mathcal{M}$, then $f = g$ μ-a.e.*

Recall Proposition 1.32, which asserts that for any two $\mathcal{M}$-measurable functions f, g with $f = g$ μ-a.e., the μ-integrals $\int_A f d\mu$ and $\int_A g d\mu$ are *always* the same. In other words, *sets of zero μ-measure can be neglected as long as μ-integrals are concerned.* By taking this fact into consideration, we can slightly weaken the assumptions of the results in this section by allowing exceptional sets of μ-measure zero.

For example, Theorem 1.33 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 $N := \bigcup_{n=0}^{\infty} N_n$ satisfies $\mu(N) = 0$ by Problem 1.10, we obtain (1.32) by applying the original Theorem 1.33 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 $A := \{x \in X \mid \limsup_{n\to\infty} f_n(x) = \liminf_{n\to\infty} f_n(x)\}$ (recall that $A \in \mathcal{M}$ by Problem 1.15-(1)), 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 $\limsup_{n\to\infty} f_n$ is $\mathcal{M}$-measurable, if we extend f outside A by defining $f := h$ on A^c, where $h : X \to [-\infty, \infty]$ is an arbitrary $\mathcal{M}$-measurable function, then f is $\mathcal{M}$-measurable (see Problem 1.15-(2)) and $\int_X f d\mu$ is defined. Furthermore Proposition 1.32 together with $\mu(A^c) = 0$ assures that this integral $\int_X f d\mu$ is independent of a particular choice of the extension $h|_{A^c}$ of f to 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.

Since we may neglect sets of μ-measure zero as long as μ-integrals are concerned, it sounds quite natural that *any* subset of a set $N \in \mathcal{M}$ of μ-measure zero should also be of μ-measure zero. As a matter of fact, this is not always the case for a general measure space $(X, \mathcal{M}, \mu)$ since such N may include non-measurable sets, but we can still *define* the μ-measure of any subset of such N to be 0, so that μ is extended to a *measure* defined on a larger σ-algebra, as follows.

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

$$\overline{\mathcal{M}}^{\mu} := \{A \subset X \mid B \subset A \subset C \textit{ for some } B, C \in \mathcal{M} \textit{ with } \mu(C \setminus B) = 0\}. \quad (1.37)$$

Then $\overline{\mathcal{M}}^{\mu}$ is a σ-algebra in X satisfying $\mathcal{M} \subset \overline{\mathcal{M}}^{\mu}$, 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*. Note that, as shown in the proof of this theorem below, *if $A \in \overline{M}^{\mu}$ and $B, C \in \mathcal{M}$ satisfy $B \subset A \subset C$ and $\mu(C \setminus B) = 0$, then $\overline{\mu}(A) = \mu(B) = \mu(C)$.*

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 (1.37) 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 $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$ (Example 1.8) and the Bernoulli measure $\mathbb{P}_p$ on $\mathcal{F}$ (Example 1.12) are not complete.

1.3.4 Integration of 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.*

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, if $f : X \to \mathbb{C}$ is $\mathcal{M}$-measurable then $|f|$ is $\mathcal{M}$-measurable by virtue of Problem 1.16.

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

$$\int_X f d\mu := \int_X \mathrm{Re}(f) d\mu + i \int_X \mathrm{Im}(f) d\mu. \tag{1.38}$$

We also set

$$\mathcal{L}^1(X, \mathcal{M}, \mu, \mathbb{C}) := \{ f : X \to \mathbb{C} \mid f \text{ is } \mathcal{M}\text{-measurable and } \mu\text{-integrable} \}, \tag{1.39}$$

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.
(2) Let $A \in \mathcal{M}$. An $\mathcal{M}$-measurable function $f : X \to \mathbb{C}$ is called *μ-integrable on A* if and only if $f \mathbf{1}_A$ is μ-integrable, and in this case its *μ-integral* $\int_A f d\mu$ *on A* is defined by $\int_A f d\mu := \int_X f \mathbf{1}_A d\mu$.

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

$$\left| \int_X f d\mu \right| \leq \int_X |f| d\mu. \tag{1.40}$$

(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*

$$\int_X (\alpha f + \beta g) d\mu = \alpha \int_X f d\mu + \beta \int_X g d\mu. \tag{1.41}$$

1.4 Some Basic Consequences

In this section, we present some consequences of the integration theory developed so far in this chapter. In the proofs 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*

$$\nu(A) := \int_A f d\mu. \tag{1.42}$$

Then ν is a measure on $(X, \mathcal{M})$. Moreover, if $g : X \to [-\infty, \infty]$ is $\mathcal{M}$-measurable, then $\int_X g d\nu$ exists if and only if $\int_X g f d\mu$ exists, and in this case

$$\int_X g d\nu = \int_X g f d\mu. \tag{1.43}$$

The measure ν is denoted by $f \cdot \mu$, and (1.43) is often abbreviated as $d\nu = f d\mu$.

Remark 1.44. Note that the measure $\nu = f \cdot \mu$ satisfies $\nu(A) = 0$ for any $A \in \mathcal{M}$ with $\mu(A) = 0$ by Proposition 1.30-(1). 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 a measure ν on $(X, \mathcal{M})$ of this form under certain mild assumptions on μ and ν. 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 $\int_S f d(\mu \circ \varphi^{-1})$ exists if and only if $\int_X (f \circ \varphi) d\mu$ exists, and in this case*

$$\int_S f d(\mu \circ \varphi^{-1}) = \int_X (f \circ \varphi) d\mu. \tag{1.44}$$

The measure $\mu \circ \varphi^{-1}$ is called the *image measure of μ by φ*. An application of the dominated convergence theorem (Theorem 1.33) 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 $|(\partial f/\partial t)(x, t)| \leq g(x)$ 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*

$$\frac{d}{dt} \int_X f(x, t) d\mu(x) = \int_X \frac{\partial f}{\partial t}(x, t) d\mu(x). \tag{1.45}$$

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 by Problem 1.20-(1), 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*

$$\int_X fg d\mu \leq \left(\int_X f^p d\mu\right)^{1/p} \left(\int_X g^q d\mu\right)^{1/q}. \tag{1.46}$$

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

$$\|f\|_{L^p(X,\mu)} := \left(\int_X |f|^p d\mu\right)^{1/p}, \tag{1.47}$$

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

$$\mathcal{L}^p(X, \mathcal{M}, \mu) := \{f : X \to \mathbb{R} \mid f \text{ is } \mathcal{M}\text{-measurable and } \|f\|_{L^p(X,\mu)} < \infty\}, \tag{1.48}$$

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

Note that (1.48) is consistent with (1.26). We easily see that $\mathcal{L}^p(\mu)$ is a $\mathbb{R}$-vector space[9] 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 1.29, 1.30 and 1.31 and Exercise 1.35 below for other important facts concerning $\mathcal{L}^p(\mu)$.

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 $t \in [0,1]$,

$$\varphi\big((1-t)x + ty\big) \le (1-t)\varphi(x) + t\varphi(y), \tag{1.49}$$

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

$$\frac{\varphi(z)-\varphi(x)}{z-x} \le \frac{\varphi(y)-\varphi(z)}{y-z}. \tag{1.50}$$

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.*

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 (1.49) 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 $\int_X f d\mu \in (a,b)$, $\int_X (\varphi\circ f)^- d\mu < \infty$ and*

$$\varphi\left(\int_X f d\mu\right) \le \int_X (\varphi\circ f) d\mu. \tag{1.51}$$

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 $\Lambda_0 \subset \Lambda$.

[9] That is, $\alpha f + \beta g \in \mathcal{L}^p(\mu)$ for any $f, g \in \mathcal{L}^p(\mu)$ and any $\alpha, \beta \in \mathbb{R}$.

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 $\mathbb{P}[A \cap B \cap C] \neq \mathbb{P}[A]\mathbb{P}[B]\mathbb{P}[C]$.

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.

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

$$\mathbb{P}[A \mid B] := \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}. \tag{1.52}$$

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,

$$\mathbb{P}[\Omega_n \mid A] = \frac{\mathbb{P}[A \mid \Omega_n]\mathbb{P}[\Omega_n]}{\sum_{k=1}^N \mathbb{P}[A \mid \Omega_k]\mathbb{P}[\Omega_k]}. \tag{1.53}$$

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*:

$$\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n \sum_{1\le i_1<\cdots<i_k\le n} (-1)^{k-1} \mu\left(\bigcap_{\ell=1}^k A_{i_\ell}\right). \tag{1.54}$$

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

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n). \tag{1.55}$$

Problem 1.11. 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

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k, \qquad \liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k, \tag{1.56}$$

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

$$\begin{aligned}\limsup_{n\to\infty} A_n &= \{x \in X \mid x \in A_n \text{ for infinitely many } n \in \mathbb{N}\},\\ \liminf_{n\to\infty} A_n &= \{x \in X \mid x \in A_n \text{ for sufficiently large } n \in \mathbb{N}\}.\end{aligned} \tag{1.57}$$

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

$$\mu\left(\limsup_{n\to\infty} A_n\right) = \mu\left(\left(\liminf_{n\to\infty} A_n^c\right)^c\right) = 0. \tag{1.58}$$

Problem 1.12. 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 $\lim_{n\to\infty} \#A_n \neq \#\bigl(\bigcap_{n=1}^{\infty} A_n\bigr)$.

Problem 1.12 shows that the conclusion of Proposition 1.4-(4) is not necessarily valid if the assumption "$\mu(A_1) < \infty$" is dropped.

Problem 1.13. 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 1.14. 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.

Problem 1.15. (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

$$f(x) := \begin{cases} \lim_{n\to\infty} f_n(x) & \text{if the limit } \lim_{n\to\infty} f_n(x) \text{ exists in } \mathbb{R},\\ h(x) & \text{otherwise,}\end{cases} \tag{1.59}$$

$$g(x) := \begin{cases} \lim_{n\to\infty} f_n(x) & \text{if the limit } \lim_{n\to\infty} f_n(x) \text{ exists in } [-\infty, \infty],\\ h(x) & \text{otherwise.}\end{cases} \tag{1.60}$$

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

Problem 1.16. Let $(S, \mathcal{B})$ be a measurable space, let $\varphi : X \to S$ be $\mathcal{M}/\mathcal{B}$-measurable (see Definition 1.45) and let $f : S \to [-\infty, \infty]$ be $\mathcal{B}$-measurable. Prove that $f \circ \varphi : X \to [-\infty, \infty]$ is $\mathcal{M}$-measurable.

Problem 1.17. (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 $f = (f_1, \dots, f_d) : X \to \mathbb{R}^d$, where $f_i : X \to \mathbb{R}$ for each $i \in \{1, \dots, d\}$. Prove that f is $\mathcal{M}/\mathcal{B}(\mathbb{R}^d)$-measurable if and only if f_i is $\mathcal{M}$-measurable for any $i \in \{1, \dots, d\}$.

Exercise 1.18. 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

$$f^\varepsilon(x) := \sup_{y \in B_S(x,\varepsilon)} f(y) \quad \text{and} \quad f_\varepsilon(x) := \inf_{y \in B_S(x,\varepsilon)} f(y). \tag{1.61}$$

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

$$\overline{f}(x) := \limsup_{S \ni y \to x} f(y) \quad \text{and} \quad \underline{f}(x) := \liminf_{S \ni y \to x} f(y) \tag{1.62}$$

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 1.19. 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 $\varphi : [0, \infty] \to [0, \infty]$ be non-decreasing and let $f : X \to [0, \infty]$ be $\mathcal{M}$-measurable. Prove the following assertions.
(1) $\varphi \circ f$ is $\mathcal{M}$-measurable.
(2) (Chebyshev's inequality) For any $a \in [0, \infty]$ with $\varphi(a) \in (0, \infty)$,

$$\mu\bigl(\{x \in X \mid f(x) \geq a\}\bigr) \leq \frac{1}{\varphi(a)} \int_X (\varphi \circ f) d\mu. \tag{1.63}$$

Problem 1.21. Let $f_n : X \to [-\infty, \infty]$ be $\mathcal{M}$-measurable for each $n \in \mathbb{N}$ and suppose that $\sum_{n=1}^\infty \int_X |f_n| d\mu < \infty$. Prove that $\lim_{n \to \infty} f_n(x) = 0$ for μ-a.e. $x \in X$.

Problem 1.22. Find the limits as $N \to \infty$ of the following series:

$$(1)\ \sum_{n=1}^\infty 2^{-n}\left(1 + \frac{\sin(2^N n)}{N}\right)^{-1} \qquad (2)\ \sum_{n=1}^\infty \frac{1}{n(n+N)} \qquad (3)\ \sum_{n=1}^\infty \left(1 + \frac{n}{N}\right)^{-N}$$

Problem 1.23. 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. For each $n \in \mathbb{N}$, define $f_n : [a, b] \to \mathbb{R}$ by

$$f_n := \sum_{k=1}^n f\left(a + \frac{k}{n}(b-a)\right) \mathbf{1}_{\left(a + \frac{k-1}{n}(b-a),\, a + \frac{k}{n}(b-a)\right]} + f(a)\mathbf{1}_{\{a\}}. \tag{1.64}$$

(i) Prove that $\lim_{n\to\infty} f_n(x) = f(x)$ for any $x \in [a,b]$.
(ii) By considering $\lim_{n\to\infty} \int_{[a,b]} f_n d\mathrm{m}_1$, prove that

$$\int_{[a,b]} f d\mathrm{m}_1 = \int_a^b f(x)dx, \tag{1.65}$$

where the integral in the right-hand side denotes the Riemann integral on $[a,b]$.
(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 $\lim_{b\to\infty} \int_a^b |f(x)|dx < \infty$,[10] and in that case

$$\int_{[a,\infty)} f d\mathrm{m}_1 = \lim_{b\to\infty} \int_a^b f(x)dx. \tag{1.66}$$

By Problem 1.23-(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 1.23-(3) says that the same is true also for a continuous function on an unbounded interval *provided the improper Riemann integral* $\lim_{b\to\infty} \int_a^b f(x)dx$ *is absolutely convergent*. Here the assumption of the absolute convergence is necessary; see Problem 2.14 in this connection.

Problem 1.24. Find the limits as $n \to \infty$ of the following integrals:

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

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 $\lim_{n\to\infty} \int_X |f - f_n| d\mu = 0$.

Problem 1.26 ([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 $\log\infty := \infty^\alpha := \infty$)

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

Exercise 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.

[10] 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 1.28. Let $p \in (0, \infty)$ and let $f \in \mathcal{L}^p(\mu)$. Prove that

$$\lim_{n\to\infty} \int_X \left| f - f\mathbf{1}_{\{|f|\le n\}} \right|^p d\mu = 0. \tag{1.67}$$

Problem 1.29. Let $p, q \in (0, \infty)$, $p < q$, and let $f : X \to [0, \infty]$ be $\mathcal{M}$-measurable. Prove that

$$\left(\int_X f^p d\mu \right)^{1/p} \le \left(\int_X f^q d\mu \right)^{1/q} \mu(X)^{(q-p)/pq}. \tag{1.68}$$

By Problem 1.29, 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 1.30 (Minkowski's inequality)**.** Let $p \in [1, \infty)$ and let $f, g : X \to [0, \infty]$ be $\mathcal{M}$-measurable. Prove that

$$\left(\int_X (f + g)^p d\mu \right)^{1/p} \le \left(\int_X f^p d\mu \right)^{1/p} + \left(\int_X g^p d\mu \right)^{1/p}. \tag{1.69}$$

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 $\varepsilon \in (0, \infty)$,

$$\lim_{n\to\infty} \mu\bigl(\{x \in X \mid |f_n(x) - f(x)| \ge \varepsilon\}\bigr) = 0. \tag{1.70}$$

Problem 1.31. 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. $x \in X$.

Problem 1.32. 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 $\int_X f\mathbf{1}_A d\mu$ exists if and only if $\int_A f|_A d(\mu|_A)$ exists, and in this case

$$\left(\int_A f d\mu := \right) \int_X f\mathbf{1}_A d\mu = \int_A f|_A \, d(\mu|_A). \tag{1.71}$$

According to Problem 1.32, $\int_A f d\mu$ could alternatively be defined as the integral of $f|_A$ with respect to $\mu|_A = \mu|_{\mathcal{M}|_A}$, the *restriction of μ to A*.

Exercise 1.33. 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 $\int_X f d\mu$ exists if and only if $\int_X f d(\mu|_{\mathcal{N}})$ exists (note that $\mu|_{\mathcal{N}}$ is a measure on $(X, \mathcal{N})$), and in this case

$$\int_X f d\mu = \int_X f d(\mu|_{\mathcal{N}}). \tag{1.72}$$

Exercise 1.34. 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 $\mu(A) < \delta$.

Exercise 1.35. Assume that $(X, \mathcal{M}, \mu)$ is σ-finite (see Definition 2.25). Let $p \in (1, \infty)$, $q := p/(p-1)$, and let $f : X \to [0, \infty]$ be $\mathcal{M}$-measurable. Prove that

$$\|f\|_{L^p} = \sup\left\{\int_X fg d\mu \;\middle|\; g : X \to [0, \infty],\ g \text{ is } \mathcal{M}\text{-measurable and } \|g\|_{L^q} \leq 1\right\}. \tag{1.73}$$

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, some uniqueness theorems for measures are 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 $\delta_X(\mathcal{A}) \subset \sigma_X(\mathcal{A})$.

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

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*

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

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.*

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

Theorem 2.5 (Uniqueness of probability measures)**.** *Let X be a set, let $\mathcal{A} \subset 2^X$ be a π-system and let $\nu : \mathcal{A} \to [0,1]$. Then a probability measure μ on $\sigma(\mathcal{A})$ with $\mu|_{\mathcal{A}} = \nu$, if exists, is unique, i.e. if μ_1, μ_2 are probability measures on $\sigma(\mathcal{A})$ with $\mu_1|_{\mathcal{A}} = \mu_2|_{\mathcal{A}} = \nu$, then $\mu_1 = \mu_2$.*

For instance, Theorem 2.5 can be used to prove the uniqueness of the Bernoulli measure $\mathbb{P}_p$ of probability p stated in Example 1.12; see Problem 2.2.

With exactly the same idea and a more complicated calculation using the inclusion-exclusion formula (Problem 1.9), we can also prove the following more general uniqueness theorem applicable to non-probability measures.

Theorem 2.6 (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 a measure μ on $\sigma(\mathcal{A})$ with $\mu|_{\mathcal{A}} = \nu$, if exists, is unique, i.e. if μ_1, μ_2 are measures on $\sigma(\mathcal{A})$ with $\mu_1|_{\mathcal{A}} = \mu_2|_{\mathcal{A}} = \nu$, then $\mu_1 = \mu_2$.

Example 2.7. Let $d \in \mathbb{N}$, let $\mathcal{F}_d$ be as in (1.6), and define $\nu : \mathcal{F}_d \to [0,\infty)$ by

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

Then $\mathcal{F}_d$ is clearly a π-system and (2.3) is satisfied with $X_n := [-n,n]^d$. Thus by Theorem 2.6, 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 2.8 (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

$$\mu(A) := \inf\left\{ \sum_{n=1}^{\infty} \nu(A_n) \;\middle|\; \{A_n\}_{n=1}^{\infty} \subset \mathcal{A},\ A \subset \bigcup_{n=1}^{\infty} A_n \right\} \quad (\inf \emptyset := \infty) \tag{2.4}$$

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 2.8. We need the following definition and theorem, which are also fundamental in measure theory.

Definition 2.9 (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

$$\mathcal{M}(\nu) := \{A \subset X \mid \nu(E) = \nu(E \cap A) + \nu(E \setminus A) \text{ for any } E \subset X\}. \tag{2.5}$$

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 $E = (E \cap A) \cup (E \setminus A) \cup \emptyset \cup \emptyset \cup \ldots$, 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 2.10 (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 $\mathcal{M}(\nu)$.*

We also need the following easy lemma.

Lemma 2.11. *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*

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

is an outer measure on X.

The proof of Lemma 2.11 is left to the reader as an exercise (Problem 2.4).

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 2.8. 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 2.12. A function $F : \mathbb{R} \to \mathbb{R}$ is called *right-continuous* if and only if[1]

$$\lim_{y \downarrow x} F(y) = F(x) \qquad \text{for any } x \in \mathbb{R}. \tag{2.7}$$

Proposition 2.13. *Let μ be a Borel measure on $\mathbb{R}$ such that $\mu\bigl((-n, n]\bigr) < \infty$ for any $n \in \mathbb{N}$. Define $F : \mathbb{R} \to \mathbb{R}$ by*

$$F(x) := \begin{cases} \mu\bigl((0, x]\bigr) & \text{if } x \in (0, \infty), \\ 0 & \text{if } x = 0, \\ -\mu\bigl((x, 0]\bigr) & \text{if } x \in (-\infty, 0]. \end{cases} \tag{2.8}$$

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

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

Theorem 2.14. *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 $\mu_F\bigl((a, b]\bigr) = F(b) - F(a)$ for any $a, b \in \mathbb{R}$ with $a < b$.*

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

Corollary 2.15 (Lebesgue measure on $\mathcal{B}(\mathbb{R})$)**.** *There exists a unique Borel measure* m_1 *on $\mathbb{R}$ such that* $\mathrm{m}_1\bigl([a, b]\bigr) = b - a$ *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* $\mathbb{R}$. The case of probability measures is of particular importance.

Definition 2.16 (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 $F_\mu(x) := \mu\bigl((-\infty, x]\bigr)$ is called the *distribution function of μ*.

[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 $y \in (x - \delta, x)$).

Similarly to Proposition 2.13, 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 2.14, 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.*[2]

Corollary 2.17. *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 $\lim_{x\to-\infty} F(x) = 0$.*

According to Corollary 2.17 and the argument after Definition 2.16, $\mu \mapsto F_\mu$ 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 2.17 can be generalized to Borel probability measures on $\mathbb{R}^d$, as described below in this subsection.

Definition 2.18 (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

$$F_\mu(x_1, \dots, x_d) := \mu\big((-\infty, x_1] \times \cdots \times (-\infty, x_d]\big) \tag{2.9}$$

is called the *distribution function of μ.*

Proposition 2.19. *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 $(x_1, \dots, x_d) \in \mathbb{R}^d$ and any $(h_1, \dots, h_d) \in [0, \infty)^d$,*

$$\begin{aligned} &\mu\big((x_1 - h_1, x_1] \times \cdots \times (x_d - h_d, x_d]\big) \\ &\qquad = \sum_{(\alpha_1,\dots,\alpha_d)\in\{0,1\}^d} (-1)^{\sum_{i=1}^d \alpha_i} F_\mu(x_1 - \alpha_1 h_1, \dots, x_d - \alpha_d h_d) \geq 0, \end{aligned} \tag{2.10}$$

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

$$\lim_{\substack{(y_1,\dots,y_d)\to x \\ y_i \geq x_i,\, i\in\{1,\dots,d\}}} F_\mu(y_1, \dots, y_d) = F_\mu(x). \tag{2.11}$$

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

[2] That is, if ν is a Borel probability measure on $\mathbb{R}$ whose distribution function is F_μ, then $\nu = \mu$.
[3] That is, if ν is a Borel probability measure on $\mathbb{R}^d$ whose distribution function is F_μ, then $\nu = \mu$.

The proof of Proposition 2.19 is left to the reader as an exercise (Problem 2.8).

Theorem 2.20. *Let $d \in \mathbb{N}$, and let $F : \mathbb{R}^d \to \mathbb{R}$ satisfy the following conditions:*

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

$$\sum_{(\alpha_1,\dots,\alpha_d)\in\{0,1\}^d} (-1)^{\sum_{i=1}^d \alpha_i} F(x_1 - \alpha_1 h_1, \dots, x_d - \alpha_d h_d) \geq 0. \tag{2.12}$$

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

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

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

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 2.21. *Let $d \in \mathbb{N}$, and let μ be a Borel measure on $\mathbb{R}^d$ with $\mu\big([-n, n]^d\big) < \infty$ for any $n \in \mathbb{N}$. Then for any $A \in \mathcal{B}(\mathbb{R}^d)$,*

$$\mu(A) = \inf\{\mu(U) \mid A \subset U \subset \mathbb{R}^d, U \text{ is open in } \mathbb{R}^d\} \tag{2.13}$$

$$= \sup\{\mu(K) \mid K \subset A, K \text{ is compact}\}. \tag{2.14}$$

Note that Theorem 2.21 is applicable to the Lebesgue measure m_d on $\mathbb{R}^d$, since m_d satisfies $\mathrm{m}_d\big([-n, n]^d\big) = (2n)^d < \infty$ for any $n \in \mathbb{N}$.

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*

$$\int_{[0,1]^2} f(z)dz = \int_0^1 \left(\int_0^1 f(x, y)dx\right)dy = \int_0^1 \left(\int_0^1 f(x, y)dy\right)dx. \tag{2.15}$$

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 2.22 (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

$$\mathcal{M}_1 \times \cdots \times \mathcal{M}_n := \big\{A_1 \times \cdots \times A_n \bigm| A_i \in \mathcal{M}_i \text{ for } i \in \{1, \dots, n\}\big\}, \tag{2.16}$$

$$\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n := \sigma_{X_1 \times \cdots \times X_n}(\mathcal{M}_1 \times \cdots \times \mathcal{M}_n) \, (= \mathcal{M}_1 \text{ if } n = 1). \tag{2.17}$$

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

Proposition 2.23. *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*

$$(\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n) \otimes (\mathcal{M}_{n+1} \otimes \cdots \otimes \mathcal{M}_{n+k}) = \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_{n+k}. \tag{2.18}$$

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

Proposition 2.24. (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).*

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

Definition 2.25. 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

$$X = \bigcup_{n=1}^\infty X_n \qquad \text{and} \qquad \mu(X_n) < \infty \quad \text{for any } n \in \mathbb{N}. \tag{2.19}$$

Note that, by considering $\{\bigcup_{i=1}^n X_i\}_{n=1}^\infty$ instead of $\{X_n\}_{n=1}^\infty$, *in* (2.19) *we may assume without loss of generality that $X_n \subset X_{n+1}$ for any $n \in \mathbb{N}$.*

Theorem 2.26 (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 $A_i \in \mathcal{M}_i$, $i \in \{1, \dots, n\}$,*

$$\mu(A_1 \times \cdots \times A_n) = \mu_1(A_1) \cdots \mu_n(A_n). \tag{2.20}$$

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* $\{\mu_i\}_{i=1}^n$.

Corollary 2.27. *Let $n, k \in \mathbb{N}$, and for each $i \in \{1, \dots, n+k\}$ let $(X_i, \mathcal{M}_i, \mu_i)$ be a σ-finite measure space. Then*

$$(\mu_1 \times \cdots \times \mu_n) \times (\mu_{n+1} \times \cdots \times \mu_{n+k}) = \mu_1 \times \cdots \times \mu_{n+k}. \tag{2.21}$$

Theorem 2.26 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 2.15 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 2.28 (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 $a_i \leq b_i$, $i \in \{1, \dots, d\}$,*

$$\mathrm{m}_d\big([a_1, b_1] \times \cdots \times [a_d, b_d]\big) = (b_1 - a_1) \cdots (b_d - a_d). \tag{2.22}$$

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* $\mathbb{R}^d$. 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, $i \in \{1, \ldots, n\}$. This is established in Theorem 2.30 below, which requires some preparations concerning measurability of functions. Note that, in view of Proposition 2.23 and Corollary 2.27, it suffices to consider the case of $n = 2$.

Proposition 2.29. *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* $x \in X$.

Theorem 2.30 (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*

$$\int_{X \times Y} f d(\mu \times \nu) = \int_X \Big(\int_Y f(x, y) d\nu(y) \Big) d\mu(x) = \int_Y \Big(\int_X f(x, y) d\mu(x) \Big) d\nu(y). \tag{2.23}$$

(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* (2.23) *holds.*

Remark 2.31. (1) In the situation of Theorem 2.30-(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 2.30-(1). The first assertion of Theorem 2.30-(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 2.30-(2) is easily verified also for $\mathbb{C}$*-valued* $\mathcal{M} \otimes \mathcal{N}$-measurable f.

The assumption of σ-finiteness of μ and ν and the integrability assumption in (2) are indeed necessary in Theorem 2.30; 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 2.30.

2.5 Fubini's Theorem for Completed Product Measures

In the last section we have proved Fubini's theorem (Theorem 2.30). 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

2.29). As a consequence, we cannot apply Theorem 2.30 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 2.32. *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 σ-finite measure space. Then it holds that*

$$\overline{\mu_1 \times \cdots \times \mu_n} = \overline{\overline{\mu_1} \times \cdots \times \overline{\mu_n}}. \tag{2.24}$$

Corollary 2.33. *Let $n, k \in \mathbb{N}$. Then $\overline{\mathrm{m}_{n+k}} = \overline{\overline{\mathrm{m}_n} \times \overline{\mathrm{m}_k}}$.*

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

Theorem 2.34 (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*

$$\int_{X \times Y} f d(\overline{\mu \times \nu}) = \int_X \Big(\int_Y f(x, y) d\nu(y) \Big) d\mu(x) = \int_Y \Big(\int_X f(x, y) d\mu(x) \Big) d\nu(y). \tag{2.25}$$

(2) *Suppose that any one of $\int_{X \times Y} |f| d(\overline{\mu \times \nu})$, $\int_X \big(\int_Y |f(x, y)| d\nu(y) \big) d\mu(x)$ 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 $\overline{\mu \times \nu}$-integrable, and* (2.25) *holds.*

Remark 2.35. (1) In the situation of Theorem 2.34-(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 2.34-(0) and the completeness of $(X, \mathcal{M}, \mu)$. Similarly to Remark 2.31-(1), the first assertion of Theorem 2.34-(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 2.31 apply to Theorem 2.34-(2).

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 2.36. *Let $d \in \mathbb{N}$, let $a_i, b_i \in \mathbb{R}$, $a_i < b_i$ for each $i \in \{1, \ldots, d\}$ 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 $f \in \mathcal{L}^1\big(I, \overline{\mathcal{B}(I)}^{\mathrm{m}_d}, \overline{\mathrm{m}_d}\big)$ and*

$$\int_I f d\overline{\mathrm{m}_d} = \int_I f(x)dx, \tag{2.26}$$

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

Remark 2.37. *In Theorem* 2.36*, 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 2.36, 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 $\int_A f(x)dx$:

$$\int_A f dx := \int_A f(x)dx := \int_A f d\overline{\mathrm{m}_d}. \tag{2.27}$$

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

$$\int_a^b f dx := \int_a^b f(x)dx := \int_{(a,b)} f d\overline{\mathrm{m}_1}. \tag{2.28}$$

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 2.38. 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 2.39. *Let $d \in \mathbb{N}$.*
(1) *If $\alpha \in \mathbb{R}^d$, then*

$$\mathrm{m}_d(A + \alpha) = \mathrm{m}_d(A) \tag{2.29}$$

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 $A \in \mathcal{B}(\mathbb{R}^d)$,*

$$\mathrm{m}_d(T(A)) = |\det T| \mathrm{m}_d(A). \tag{2.30}$$

Remark 2.40. (1) Note that $A + \alpha, T(A) \in \mathcal{B}(\mathbb{R}^d)$ in the situation of Theorem 2.39; indeed, since T^{-1} is continuous, it is $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^d)$-measurable by Lemma 1.17 and

Problem 1.17-(2) 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

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

for some $\ell \in \{1, \ldots, d\}$ 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 2.28 and Fubini's theorem (Theorem 2.30-(1)), which implies the claim.

In view of the image measure theorem (Theorem 1.46), Theorem 2.39 yields the following change-of-variables formula.

Corollary 2.41 (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* $\int_{\mathbb{R}^d} f(y)dy$ *exists if and only if* $\int_{\mathbb{R}^d} f(Tx+\alpha)dx$ *exists, and in this case*

$$\int_{\mathbb{R}^d} f(y)dy = \int_{\mathbb{R}^d} f(Tx + \alpha)|\det T|dx. \tag{2.31}$$

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

Definition 2.42. Let $d \in \mathbb{N}$, let U be an open subset of $\mathbb{R}^d$ and let $\varphi : U \to \mathbb{R}^d$, $\varphi = (\varphi_1, \ldots, \varphi_d)$.
(1) φ is called *continuously differentiable*, or simply C^1, if and only if φ is continuous, all its partial derivatives $\partial\varphi_i/\partial x_j$, $i, j \in \{1, \ldots, d\}$, 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 $D\varphi(x) := \bigl((\partial\varphi_i/\partial x_j)(x)\bigr)_{i,j=1}^d$.
(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 2.43 (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* $\int_{\varphi(U)} f(y)dy$ *exists if and only if* $\int_U f(\varphi(x))|\det D\varphi(x)|dx$ *exists, and in this case*

$$\int_{\varphi(U)} f(y)dy = \int_U f(\varphi(x))|\det D\varphi(x)|dx. \tag{2.32}$$

The proof of Theorem 2.43 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 2.43 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 $\bigcup_{n=1}^{\infty} A_n \in \mathcal{D}$.

Problem 2.2. Let $\Omega := \{0,1\}^{\mathbb{N}} = \{(\omega_n)_{n=1}^{\infty} \mid \omega_n \in \{0,1\}\}$, let $\mathcal{F}$ be the σ-algebra in Ω defined by (1.11) and let $p \in [0,1]$. Prove the uniqueness of the Bernoulli measure $\mathbb{P}_p$ on $(\Omega, \mathcal{F})$ of probability p stated in Example 1.12.

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

$$\mathcal{M}(\mathcal{A}) := \mathcal{M}_X(\mathcal{A}) := \bigcap_{\mathcal{M}\text{: monotone class in } X,\, \mathcal{A} \subset \mathcal{M}} \mathcal{M} \tag{2.33}$$

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

$$\mathcal{M}(\mathcal{A}) = \sigma(\mathcal{A}). \tag{2.34}$$

Problem 2.4. Prove Lemma 2.11.

Problem 2.5 ([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) $\mu\bigl([a,b]\bigr) = F(b) - F(a-)$.
(2) $\mu\bigl([a,b)\bigr) = F(b-) - F(a-)$.
(3) $\mu\bigl((a,b)\bigr) = F(b-) - F(a)$.
(4) $\mu(\{a\}) = F(a) - F(a-)$. (Thus $\mu(\{a\}) = 0$ if and only if F is continuous at a.)

Problem 2.6. 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 2.5.

Problem 2.7 ([4, Exercise 7.18])**.** Define $F : \mathbb{R} \to \mathbb{R}$ by

$$F := \sum_{n=1}^{\infty} \frac{1}{2^n} \mathbf{1}_{[n^{-1},\infty)}. \tag{2.35}$$

(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):

$$\begin{array}{llllll} \text{(i)} & \mu\big([1,\infty)\big) & \text{(ii)} & \mu\big([1/10,\infty)\big) & \text{(iii)} & \mu(\{0\}) \\ \text{(iv)} & \mu\big([0,1/2)\big) & \text{(v)} & \mu\big((-\infty,0)\big) & \text{(vi)} & \mu\big((0,\infty)\big) \end{array}$$

Problem 2.8. Prove Proposition 2.19.

Exercise 2.9. Let $d \in \mathbb{N}$ and let μ be a Borel probability measure on $\mathbb{R}^d$. Define

$$C_{\mu,i} := \big\{a \in \mathbb{R} \bigm| \mu\big(H_i(a)\big) = 0\big\}, \quad \text{where } H_i(a) := \{(x_1, \ldots, x_d) \in \mathbb{R}^d \mid x_i = a\}, \tag{2.36}$$

for each $i \in \{1, \ldots, d\}$ 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 $i \in \{1, \ldots, d\}$.
(2) The distribution function $F_\mu : \mathbb{R}^d \to [0, 1]$ of μ is continuous at x for any $x \in C_\mu$.

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. For each $i \in \{1, \ldots, n\}$ define $F_i : X_1 \times \cdots \times X_n \to [-\infty, \infty]$ by $F_i(x_1, \ldots, x_n) := f_i(x_i)$, and define $F : X_1 \times \cdots \times X_n \to [-\infty, \infty]$ by $F(x_1, \ldots, x_n) := f_1(x_1) \cdots f_n(x_n)$. Prove the following statements:
(1) F_i is $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$-measurable for any $i \in \{1, \ldots, n\}$.
(2) F is $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$-measurable.
(3) If f_i is μ_i-integrable for any $i \in \{1, \ldots, n\}$, then F is $\mu_1 \times \cdots \times \mu_n$-integrable and

$$\int_{X_1 \times \cdots \times X_n} F d(\mu_1 \times \cdots \times \mu_n) = \int_{X_1} f_1 d\mu_1 \cdots \int_{X_n} f_n d\mu_n. \tag{2.37}$$

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 $S_f := \{(x, t) \in X \times \mathbb{R} \mid 0 \leq t < f(x)\}$.
(1) Prove that $S_f \in \mathcal{M} \otimes \mathcal{B}(\mathbb{R})$ and that $[0, \infty) \ni t \mapsto \mu\big(\{x \in X \mid f(x) > t\}\big) \in [0, \infty]$ is Borel measurable.
(2) Prove that $\int_X f d\mu = \mu \times \mathrm{m}_1(S_f)$ and that for any $p \in (0, \infty)$,

$$\int_X f^p d\mu = p \int_0^\infty t^{p-1} \mu\big(\{x \in X \mid f(x) > t\}\big) dt. \tag{2.38}$$

(3) Prove that $\mathrm{m}_2\big(\{x \in \mathbb{R}^2 \mid |x| < r\}\big) = \pi r^2$ 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

$$\int_0^1 \left(\int_{[0,1]} \mathbf{1}_{\Delta_{[0,1]}}(x, y) d\#(y) \right) dx = 1 \neq 0 = \int_{[0,1]} \left(\int_0^1 \mathbf{1}_{\Delta_{[0,1]}}(x, y) dx \right) d\#(y). \tag{2.39}$$

(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

$$f(x, y) := \sum_{n=1}^{\infty} \bigl(g_n(x) - g_{n+1}(x)\bigr) g_n(y). \tag{2.40}$$

Prove the following statements:
(i) f is continuous and $\int_0^1 \bigl(\int_0^1 |f(x, y)|dx\bigr)dy = \infty$.
(ii) For any $x, y \in [0, 1)$, $f(x, \cdot), f(\cdot, y) \in \mathcal{L}^1\bigl([0, 1), \mathrm{m}_1\bigr)$, $\int_0^1 f(x, z)dz = g_1(x)$ and $\int_0^1 f(z, y)dz = 0$. In particular,

$$\int_0^1 \left(\int_0^1 f(x, y)dy \right) dx = 1 \neq 0 = \int_0^1 \left(\int_0^1 f(x, y)dx \right) dy. \tag{2.41}$$

Problem 2.14. (1) Prove that

$$\int_0^{\infty} \left| \frac{\sin x}{x} \right| dx = \infty. \tag{2.42}$$

(2) Use $x^{-1} = \int_0^{\infty} e^{-xt} dt$, $x \in (0, \infty)$, to prove that

$$\lim_{A\to\infty} \int_0^A \frac{\sin x}{x} dx = \int_0^{\infty} \frac{1 - \cos x}{x^2} dx = \int_0^{\infty} \left(\frac{\sin x}{x} \right)^2 dx = \frac{\pi}{2}. \tag{2.43}$$

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 $i \in \{1, \dots, d\}$.*

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.*

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 Lemma 1.17 and Problem 1.17-(2).

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* (*mean*) or *the expectation* (*mean*) *of X exists* (or simply $\mathbb{E}[X]$ *exists*) if and only if the $\mathbb{P}$-integral $\int_\Omega X(\omega)\mathbb{P}(d\omega)$ of X exists, and in this case its *expectation* (*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 the mean of X exists 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, $\mathcal{L}^q(\mathbb{P}) \subset \mathcal{L}^p(\mathbb{P})$.*

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.

The following theorem is just a paraphrase 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* $\mathbb{E}[f(X)]$ *exists if and only if* $\int_S f(x)\mathbb{P}_X(dx)$ *exists, 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) $\mathbb{E}[X]$ *exists if and only if* $\int_{\mathbb{R}} x\mathbb{P}_X(dx)$ *exists, 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*

$$\operatorname{var}(X) = \int_{\mathbb{R}} \left(x - \int_{\mathbb{R}} y \mathbb{P}_X(dy) \right)^2 \mathbb{P}_X(dx) = \int_{\mathbb{R}} x^2 \mathbb{P}_X(dx) - \left(\int_{\mathbb{R}} x \mathbb{P}_X(dx) \right)^2. \quad (3.8)$$

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,

$$\mu(A) = \int_A \rho(x) dx \qquad \text{for any } A \in \mathcal{B}(\mathbb{R}^d). \quad (3.9)$$

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.*

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* $\mathbb{E}[f(X)]$ *exists if and only if* $\int_{\mathbb{R}^d} f(x)\rho(x)dx$ *exists, and in this case*

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f(x)\rho(x)dx. \quad (3.10)$$

Corollary 3.15. *Let* X *be a real random variable with a density* ρ*.*
(1) $\mathbb{E}[X]$ *exists if and only if* $\int_{-\infty}^{\infty} x\rho(x)dx$ *exists, and in this case*

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\rho(x)dx. \quad (3.11)$$

(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*

$$\begin{aligned} \operatorname{var}(X) &= \int_{-\infty}^{\infty} \left(x - \int_{-\infty}^{\infty} y\rho(y)dy \right)^2 \rho(x)dx \\ &= \int_{-\infty}^{\infty} x^2\rho(x)dx - \left(\int_{-\infty}^{\infty} x\rho(x)dx \right)^2. \end{aligned} \quad (3.12)$$

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 $\varphi(X)$.* Note that by Proposition 3.3, if $X = (X_1, \dots, X_n)$ is an n-dimensional random variable and $Y = (Y_1, \dots, Y_k)$ is a k-dimensional random variable, then $(X, Y) = (X_1, \dots, X_n, Y_1, \dots, Y_k)$ 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*

$$\rho_X(x) := \int_{\mathbb{R}^k} \rho(x, y)dy \quad \text{and} \quad \rho_Y(y) := \int_{\mathbb{R}^n} \rho(x, y)dx. \tag{3.14}$$

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 $\{0, \dots, n\}$ given by

$$B(n, p)(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k \in \{0, \dots, n\}, \tag{3.15}$$

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

$$N(m,v)(dx) = \frac{1}{\sqrt{2\pi v}} \exp\biggl(-\frac{(x-m)^2}{2v}\biggr) dx \tag{3.23}$$

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 2.41 with $y = \sqrt{v}x + m$ yields

$$\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi v}} \exp\biggl(-\frac{(y-m)^2}{2v}\biggr) dy = \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp\biggl(-\frac{x^2}{2}\biggr) dx,$$

and by the first half of Fubini's theorem (Theorem 2.30-(1)) and Theorem 2.43 with the polar coordinates $(0,\infty) \times (0,2\pi) \ni (r,\theta) \mapsto (r\cos\theta, r\sin\theta) \in \mathbb{R}^2 \setminus \bigl([0,\infty) \times \{0\}\bigr)$,

$$\biggl(\int_{-\infty}^\infty e^{-x^2/2} dx\biggr)^2 = \int_{\mathbb{R}^2} e^{-|z|^2/2} dz = \int_0^\infty \biggl(\int_0^{2\pi} e^{-r^2/2} r\, d\theta\biggr) dr = 2\pi.$$

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 $v = 0$, then $X = m$ a.s., hence $\mathbb{E}[X] = m$ and $\mathrm{var}(X) = 0 = v$. Suppose $v > 0$. Then Theorem 3.14 and Corollary 2.41 yield

$$\begin{aligned} \mathbb{E}[(X-m)^2] &= \int_{-\infty}^\infty \frac{(x-m)^2}{\sqrt{2\pi v}} \exp\biggl(-\frac{(x-m)^2}{2v}\biggr) dx = v \int_{-\infty}^\infty y^2 \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy \\ &= v \lim_{n\to\infty} \int_{-n}^n y^2 \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy \\ &= v \lim_{n\to\infty} \biggl(\biggl[-y\frac{e^{-y^2/2}}{\sqrt{2\pi}}\biggr]_{-n}^n + \int_{-n}^n \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy\biggr) \\ &= v \int_{-\infty}^\infty \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy = v < \infty, \end{aligned} \tag{3.24}$$

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, $\mathbb{E}[X^2] < \infty$,

[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 $n \in \mathbb{N}$.

hence $\mathbb{E}[|X|] < \infty$, and then by Corollary 3.15-(1) and Corollary 2.41 we have

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi v}} \exp\biggl(-\frac{(x-m)^2}{2v}\biggr) dx = \int_{-\infty}^{\infty} \bigl(m + \sqrt{v} y\bigr) \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy = m,$$

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 $i \in \{1, \dots, n\}$ 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 $A_i \in \mathcal{B}_i$, $i \in \{1, \dots, n\}$,

$$\mathbb{P}[X_1 \in A_1, \dots, X_n \in A_n] = \mathbb{P}[X_1 \in A_1] \cdots \mathbb{P}[X_n \in A_n]. \tag{3.26}$$

According to Problem 2.10, in the situation of Definition 3.26, $(X_1, \dots, X_n)$ 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 $\mathbb{P}_{(X_1,\dots,X_n)}$ 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 $i \in \{1, \dots, n\}$ 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*

$$\mathbb{P}_{(X_1,\dots,X_n)} = \mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_n}. \tag{3.27}$$

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.*

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 $\rho(x_1, \dots, x_n) = \rho_1(x_1) \cdots \rho_n(x_n)$.*

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 $\varphi^{-1}(x, y) = \bigl(\frac{x+y}{2}, \frac{x-y}{2}\bigr)$, $D(\varphi^{-1})(x, y) = \frac{1}{2}\bigl(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\bigr)$, $|\det D(\varphi^{-1})(x, y)| = 1/2$ and therefore by Theorem 3.16, $(X + Y, X - Y)$ has a density given by

$$\rho\biggl(\frac{x+y}{2}, \frac{x-y}{2}\biggr) \cdot \frac{1}{2} = \frac{1}{4\pi} e^{-(x^2+y^2)/4} = \frac{1}{\sqrt{4\pi}} e^{-x^2/4} \cdot \frac{1}{\sqrt{4\pi}} e^{-y^2/4}.$$

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 $i \in \{1, \dots, n\}$ 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 $1 \le i_1 < \cdots < i_k \le n$, $\{X_{i_\ell}\}_{\ell=1}^k$ is independent.*
(2) *For each $i \in \{1, \dots, n\}$, 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 $Y := (X_1, \dots, X_k)$ and $Z := (X_{k+1}, \dots, X_n)$. Then $\{Y, Z\}$ is independent.*

Note that $\mathrm{var}(X) = \mathbb{E}\bigl[\bigl(X - \mathbb{E}[X]\bigr)^2\bigr]$ for any $X \in \mathcal{L}^1(\mathbb{P})$ *even if* $\mathbb{E}[X^2] = \infty$.

Proposition 3.32. *Let $n \in \mathbb{N}$ and let $\{X_i\}_{i=1}^n \subset \mathcal{L}^1(\mathbb{P})$ be independent. Then*

$$X_1 \cdots X_n \in \mathcal{L}^1(\mathbb{P}), \qquad \mathbb{E}[X_1 \cdots X_n] = \mathbb{E}[X_1] \cdots \mathbb{E}[X_n], \tag{3.28}$$

$$\mathrm{var}\biggl(\sum_{i=1}^n X_i\biggr) = \sum_{i=1}^n \mathrm{var}(X_i). \tag{3.29}$$

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

$$\mathbb{E}[X] = \mathbb{E}[X^2] = 0 \cdot (1 - p) + 1 \cdot p = p,$$
$$\mathrm{var}(X) = \mathbb{E}[X^2] - \bigl(\mathbb{E}[X]\bigr)^2 = p(1 - p).$$

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:

$$\mathbb{E}[S] = \sum_{i=1}^n \mathbb{E}[X_i] = np, \qquad \mathrm{var}(S) = \sum_{i=1}^n \mathrm{var}(X_i) = np(1-p). \tag{3.30}$$

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 $n \in \mathbb{N}$, $\mathbb{R}^{dn} \ni (x_1, \dots, x_n) \mapsto \mathbf{1}_A(x_1 + \dots + x_n)$ 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 $x \in \mathbb{R}^d$.*

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 $A \in \mathcal{B}(\mathbb{R}^d)$,

$$(\mu * \nu)(A) := \int_{\mathbb{R}^{2d}} \mathbf{1}_A(x+y)(\mu \times \nu)(dxdy) = \int_{\mathbb{R}^d} \nu(A-x)\mu(dx). \tag{3.31}$$

The second equality in (3.31) follows by Lemma 3.34 and Fubini's theorem (Theorem 2.30-(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 $\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y$.*

Proposition 3.37. *Let $d \in \mathbb{N}$ and let λ, μ, ν be laws on $\mathbb{R}^d$. Then*

$$\mu * \nu = \nu * \mu \qquad \text{and} \qquad (\mu * \nu) * \lambda = \mu * (\nu * \lambda). \tag{3.32}$$

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 $h(x) = \int_{\mathbb{R}^d} \rho(x-y)g(y)dy$.*

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 1.11, for $\{A_n\}_{n=1}^{\infty} \subset 2^{\Omega}$ we set

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k, \qquad \liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k, \tag{1.56}$$

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 $\sum_{n=1}^{\infty} p_n = \infty$.*

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

$$\sigma\big(\{X_\lambda\}_{\lambda\in\Lambda}\big) := \sigma_\Omega\big(\{\{X_\lambda \in A_\lambda\} \mid \lambda \in \Lambda,\ A_\lambda \in \mathcal{B}_\lambda\}\big) \tag{3.33}$$

so that $\sigma\big(\{X_\lambda\}_{\lambda\in\Lambda}\big) \subset \mathcal{F}$. We call $\sigma\big(\{X_\lambda\}_{\lambda\in\Lambda}\big)$ the *σ-algebra generated by $\{X_\lambda\}_{\lambda\in\Lambda}$*.

By definition, $\sigma\big(\{X_\lambda\}_{\lambda\in\Lambda}\big)$ 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

$$\sigma_{\infty}\big(\{X_n\}_{n=1}^{\infty}\big) := \bigcap_{n=1}^{\infty} \sigma\big(\{X_k\}_{k=n}^{\infty}\big), \tag{3.34}$$

so that $\sigma_{\infty}\big(\{X_n\}_{n=1}^{\infty}\big)$ is a σ-algebra in Ω and $\sigma_{\infty}\big(\{X_n\}_{n=1}^{\infty}\big) \subset \mathcal{F}$. $\sigma_{\infty}\big(\{X_n\}_{n=1}^{\infty}\big)$ is called the *tail σ-algebra of* $\{X_n\}_{n=1}^{\infty}$, and each $A \in \sigma_{\infty}\big(\{X_n\}_{n=1}^{\infty}\big)$ 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 $A \in \sigma_{\infty}\big(\{X_n\}_{n=1}^{\infty}\big)$, $\mathbb{P}[A]$ 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 $c \in [-\infty, \infty]$.*

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:

$$\limsup_{n\to\infty} X_n, \quad \liminf_{n\to\infty} X_n, \quad \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{3.35}$$

Indeed, let $N \in \mathbb{N}$. Then for any $n \in \mathbb{N}$ with $n \geq N$, $\sup_{k\geq n} X_k$ is $\sigma\big(\{X_k\}_{k=N}^{\infty}\big)$-measurable, and hence so is $\limsup_{n\to\infty} X_n = \lim_{n\to\infty} \sup_{k\geq n} X_k$. 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 (3.35) are constant a.s.

3.4 Convergence of Random Variables

In the next section, we consider convergence of the form

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_n = m \tag{3.36}$$

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, $\mathbb{P}[\lim_{n\to\infty} X_n = X] = 1$. (Note here that $\{\lim_{n\to\infty} X_n = X\} = \bigcap_{n=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty} \{|X_k - X| < 1/n\} \in \mathcal{F}$.)
(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

$$\lim_{n\to\infty} \mathbb{P}[|X_n - X| \geq \varepsilon] = 0 \qquad \text{for any } \varepsilon \in (0, \infty) \tag{3.37}$$

(that is, X_n converges to X in $\mathbb{P}$-measure; recall the definition before Problem 1.31).
(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] $f : \mathbb{R}^d \to \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)], \tag{3.38}$$

or equivalently (by virtue of Theorem 3.10),

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} f(x)\mathbb{P}_{X_n}(dx) = \int_{\mathbb{R}^d} f(x)\mathbb{P}_X(dx). \tag{3.39}$$

(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

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0. \tag{3.40}$$

(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

$$X_n \xrightarrow{\text{BC}} X$$

---

[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 $\sup_{x\in S} |f(x)| < \infty$.

if and only if

$$\sum_{n=1}^{\infty} \mathbb{P}[|X_n - X| \geq \varepsilon] < \infty \qquad \text{for any } \varepsilon \in (0,\infty). \tag{3.41}$$

Remark 3.50. Note that *only the laws of X and X_n, $n \in \mathbb{N}$, are involved in the definition* (3.39) *of* $X_n \xrightarrow{\mathcal{L}} X$. In particular, in defining $X_n \xrightarrow{\mathcal{L}} X$, it is enough to assume that X_n is defined on a (not necessarily common) probability space $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ for each $n \in \mathbb{N} \cup \{0\}$, where $X_0 := X$.

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:*

$$\left(X_n \xrightarrow{\mathrm{BC}} X\right) \Longrightarrow \left(X_n \xrightarrow{\mathrm{a.s.}} X\right) \Longrightarrow \left(X_n \xrightarrow{\mathrm{P}} X\right) \Longrightarrow \left(X_n \xrightarrow{\mathcal{L}} X\right) \tag{3.42}$$

$$\left(X_n \xrightarrow{L^p} X\right) \Longrightarrow \left(X_n \xrightarrow{\mathrm{P}} X\right) \tag{3.43}$$

Theorem 3.52. *Let $d \in \mathbb{N}$, and let $X, \{X_n\}_{n=1}^{\infty}$ be d-dimensional random variables. Then $X_n \xrightarrow{\mathrm{P}} X$ 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 $X_{n(k(\ell))} \xrightarrow{\mathrm{a.s.}} X$.*

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 $X_n \xrightarrow{\mathrm{a.s.}} X$ then $f(X_n) \xrightarrow{\mathrm{a.s.}} f(X)$.*
(2) *If $X_n \xrightarrow{\mathrm{P}} X$ then $f(X_n) \xrightarrow{\mathrm{P}} f(X)$.*
(3) *If $X_n \xrightarrow{\mathcal{L}} X$ then $f(X_n) \xrightarrow{\mathcal{L}} f(X)$.*

Example 3.54. Let us show that the converses of the implications in Theorem 3.51 are **not** true in general: for $p \in (0,\infty)$,

$$\left(X_n \xrightarrow{\mathcal{L}} X\right) \not\Rightarrow \left(X_n \xrightarrow{\mathrm{P}} X\right) \not\Rightarrow \left(X_n \xrightarrow{\mathrm{a.s.}} X\right) \not\Rightarrow \left(X_n \xrightarrow{\mathrm{BC}} X\right)$$

$$\left(X_n \xrightarrow{\mathrm{P}} X\right) \not\Rightarrow \left(X_n \xrightarrow{L^p} X\right)$$

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 $n \in \mathbb{N}$,

$$X_n := \sum_{k=1}^{n} \frac{k}{n} \mathbf{1}_{(\frac{k-1}{n}, \frac{k}{n}]}.$$

Then for any $\omega \in [0,1]$, clearly $|X_n(\omega) - X(\omega)| < 1/n$ and hence $\lim_{n\to\infty} X_n(\omega) = X(\omega)$. In particular, as random variables on $\big([0,1], \mathcal{B}, \mathrm{m}_1\big)$, $X_n \xrightarrow{\mathrm{a.s.}} X$ and hence $X_n \xrightarrow{\mathcal{L}} X$ by Theorem 3.51. On the other hand, for $n \geq 2$,

$$|X_n - (1 - X)| = |1 - 2X + X - X_n| \geq |1 - 2X| - |X_n - X|$$

$$> |1-2X| - \frac{1}{n} \geq |1-2X| - \frac{1}{2},$$

and therefore if $X \leq 1/8$ then $|X_n - (1-X)| \geq 1 - 2X - 1/2 = 1/4$. Thus $\mathbb{P}[|X_n - (1-X)| \geq 1/4] \geq \mathbb{P}[X \leq 1/8] = 1/8$, so that $X_n \xrightarrow{\mathrm{P}} 1 - X$ does not hold.
(2) Let $\{I_n\}_{n=1}^\infty$ be the sequence of intervals given by $I_n := \left[\frac{n-2^k}{2^k}, \frac{n-2^k+1}{2^k}\right]$, $k := \max\{\ell \in \mathbb{N} \cup \{0\} \mid 2^\ell \leq n\}$, that is,

$$I_1, I_2, I_3, \ldots, := [0,1], [0, \tfrac{1}{2}], [\tfrac{1}{2}, 1], [0, \tfrac{1}{4}], [\tfrac{1}{4}, \tfrac{1}{2}], [\tfrac{1}{2}, \tfrac{3}{4}], [\tfrac{3}{4}, 1], [0, \tfrac{1}{8}] \ldots,$$

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 $\varepsilon \in (0,1]$, $\mathrm{m}_1(|X_n| \geq \varepsilon) = \mathrm{m}_1(I_n) < 2/n \to 0$ 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 $\limsup_{n\to\infty} X_n(\omega) = 1$. 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 $\varepsilon \in (0,1]$, $\mathrm{m}_1(|X_n| \geq \varepsilon) = \mathrm{m}_1(I_n) = 1/n$ and hence $\sum_{n=1}^\infty \mathrm{m}_1(|X_n| \geq \varepsilon) = \infty$. Therefore $X_n \xrightarrow{\mathrm{BC}} 0$ does not hold. Moreover, $\mathrm{m}_1(|X_n| \geq \varepsilon) = 1/n \to 0$ 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 (3.36) 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,*

$$\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{\mathrm{P}} m. \tag{3.44}$$

In particular, the weak law of large numbers holds for any i.i.d. $\{X_n\}_{n=1}^{\infty} \subset \mathcal{L}^2(\mathbb{P})$.

By using a similar method, we can 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*

$$\sup_{x \in [a,b]} |f(x) - P(x)| < \varepsilon. \tag{3.45}$$

Theorem 3.58 (Strong 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* strong law of large numbers *holds, that is,*

$$\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{\text{a.s.}} m. \tag{3.46}$$

In particular, the strong law of large numbers holds for any i.i.d. $\{X_n\}_{n=1}^{\infty} \subset \mathcal{L}^2(\mathbb{P})$.

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 $k \in \{1, \ldots, 6\}$. Then for any $k \in \{1, \ldots, 6\}$, Theorem 3.58 yields

$$\frac{\#\{i \in \{1, \ldots, n\} \mid X_i = k\}}{n} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{k\}}(X_i) \xrightarrow{\text{a.s.}} \frac{1}{6}$$

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 $\big([0, 1), \mathcal{B}, \mathrm{m}_1\big)$, $\mathcal{B} := \mathcal{B}\big([0, 1)\big)$. For each $\omega \in [0, 1)$, let

$$\omega = 0.\omega_1 \omega_2 \omega_3 \ldots$$

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 $k \in \{0, \ldots, 9\}$, X_n is a $\{0, \ldots, 9\}$-valued random variable on $\big([0,1), \mathcal{B}, \mathrm{m}_1\big)$ and

$$\mathrm{m}_1(X_n = k) = 10^{n-1}10^{-n} = \frac{1}{10} \qquad \text{for any } k \in \{0, \ldots, 9\}. \tag{3.47}$$

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

$$A_k := \left\{\omega \in [0,1) \;\middle|\; \lim_{n\to\infty} \frac{\#\{i \in \{1, \ldots, n\} \mid \omega_i = k\}}{n} = \frac{1}{10}\right\}. \tag{3.48}$$

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

$$\mathrm{m}_1(A_k) = \mathrm{m}_1\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{k\}}(X_i) = \frac{1}{10}\right) = 1, \quad \text{whence} \quad \mathrm{m}_1\left(\bigcap_{k=0}^9 A_k\right) = 1. \tag{3.49}$$

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

$$A_{p,k} := \left\{\omega \in [0,1) \;\middle|\; \lim_{n\to\infty} \frac{\#\{i \in \{1, \ldots, n\} \mid \omega_{p,i} = k\}}{n} = \frac{1}{p}\right\} \tag{3.50}$$

for $p \geq 2$ and $k \in \{0, \ldots, p-1\}$, then similarly to (3.49) we have $\mathrm{m}_1(A_{p,k}) = 1$, and hence we conclude that

$$\mathrm{m}_1\left(\bigcap_{p=2}^\infty \bigcap_{k=0}^{p-1} A_{p,k}\right) = 1. \tag{3.51}$$

In fact, for i.i.d. real random variables $\{X_n\}_{n=1}^\infty$, the strong law of large numbers still holds as long as $X_1 \in \mathcal{L}^1(\mathbb{P})$, and this is necessary and sufficient for the validity of the strong law of large numbers, as follows.

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 $\big\{\frac{1}{n}\sum_{i=1}^n X_i\big\}_{n=1}^\infty$ 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*

$$\mathbb{E}[X] \leq \sum_{n=0}^\infty \mathbb{P}[X > n] \leq \mathbb{E}[X] + 1. \tag{3.52}$$

In particular, $\mathbb{E}[X] < \infty$ if and only if $\sum_{n=0}^\infty \mathbb{P}[X > n] < \infty$.

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,*

$$\limsup_{n\to\infty} \frac{\sum_{i=1}^n X_i}{\sqrt{2n\log\log n}} = 1 \quad \text{and} \quad \liminf_{n\to\infty} \frac{\sum_{i=1}^n X_i}{\sqrt{2n\log\log n}} = -1. \tag{3.53}$$

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

$$\prod_{n=1}^\infty \mathcal{F}_n := \left\{ A_1 \times \cdots \times A_n \times \prod_{i=n+1}^\infty \Omega_i \;\middle|\; n \in \mathbb{N},\, A_i \in \mathcal{F}_i \text{ for } i \in \{1,\ldots,n\} \right\}, \tag{3.54}$$

$$\bigotimes_{n=1}^\infty \mathcal{F}_n := \sigma_\Omega\left(\prod_{n=1}^\infty \mathcal{F}_n \right). \tag{3.55}$$

$\bigotimes_{n=1}^\infty \mathcal{F}_n$ 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* $A_n \in \mathcal{F}_n$*,* $n \in \{1,\ldots,k\}$*,*

$$\mathbb{P}\left[A_1 \times \cdots \times A_k \times \prod_{n=k+1}^\infty \Omega_n \right] = \mathbb{P}_1[A_1]\cdots\mathbb{P}_k[A_k]. \tag{3.56}$$

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* $\{\mathbb{P}_n\}_{n=1}^\infty$.

Example 3.66 (Bernuolli measures)**.** Let $\Omega := \{0,1\}^{\mathbb{N}} = \left\{ (\omega_n)_{n=1}^\infty \mid \omega_n \in \{0,1\} \right\}$. In Example 1.12, we have introduced a σ-algebra $\mathcal{F}$ in Ω given by

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

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 has been already verified in Problem 2.2.

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 $A \in \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_k$, $A \times \prod_{n=k+1}^{\infty} \Omega_n \in \bigotimes_{n=1}^{\infty} \mathcal{F}_n$.*

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

$$\mathbb{P}[X > s+t \mid X > s] = \mathbb{P}[X > t] \qquad \text{for any } s,t \in [0,\infty) \tag{3.57}$$

(recall (1.52) for the definition of conditional probabilities).

(3.57) 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

$$\rho_Y(x) = \frac{1}{x\sqrt{2\pi v}} \exp\left(-\frac{(\log x - m)^2}{2v}\right)\mathbf{1}_{(0,\infty)}(x). \tag{3.58}$$

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

$$\rho_Z(x) = \frac{1}{\sqrt{2\pi x}} e^{-x/2}\mathbf{1}_{(0,\infty)}(x). \tag{3.59}$$

The law of Z is called the *chi square distribution with one degree of freedom* and denoted as χ_1^2. (In fact, (3.59) and (3.21) easily imply that $\chi_1^2 = \mathrm{Gamma}(1/2, 1/2)$.)

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 $X \sim \mathrm{Unif}(0, \pi/2)$ and set $Y := \sin X$. Find the following quantities:

(i) a density of Y (ii) $\mathbb{E}[Y]$ (iii) $\mathrm{var}(Y)$

Exercise 3.11. Define $\rho : \mathbb{R}^2 \to [0,\infty)$ by

$$\rho(x,y) := \frac{1}{2}(x+y)e^{-x-y}\mathbf{1}_{(0,\infty)^2}(x,y). \tag{3.60}$$

(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 $X+Y$		

In Exercise 3.11-(2), you will see that $\mathrm{cov}(X,Y) \neq 0$, which together with (3.28) 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 $\alpha X \sim N(\alpha m, \alpha^2 v)$.

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 $X + Y \sim N(m_1 + m_2, v_1 + v_2)$.

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 $X \sim N(m, v)$.

Problem 3.15. Let $\{X_n\}_{n=1}^\infty$ be real random variables. Prove the following statements:
(1) $\{\lim_{n\to\infty} X_n \text{ exists in } \mathbb{R}\}$ 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 $\sigma_\infty\big(\{X_n\}_{n=1}^\infty\big)$-measurable.

Exercise 3.16. Let $d \in \mathbb{N}$, and let $\{X_n\}_{n=1}^\infty$ be d-dimensional random variables. Prove that $\left\{\lim_{n\to\infty} X_n \text{ exists in } \mathbb{R}^d\right\}$ 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 $X \sim \mathrm{Po}(\lambda)$.

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 $\mu(dx) = F'_\mu(x)\mathbf{1}_{(a,b)}(x)dx$.

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 $Z := X/Y$.

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)	$\mathbb{E}[\max\{X, Y\} \cdot \min\{X, Y\}]$

Problem 3.22. Let $X, Y, \{X_n\}_{n=1}^\infty, \{Y_n\}_{n=1}^\infty$ be real random variables such that

$$X_n \xrightarrow{\mathrm{P}} X \quad \text{and} \quad Y_n \xrightarrow{\mathrm{P}} Y. \tag{3.61}$$

(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 $X_n Y_n \xrightarrow{\mathrm{P}} XY$.

Problem 3.23. Let $X, Y, \{X_n\}_{n=1}^\infty, \{Y_n\}_{n=1}^\infty$ be real random variables such that

$$\frac{1}{n}\sum_{k=1}^n X_k \xrightarrow{\mathrm{P}} X \quad \text{and} \quad \frac{1}{n}\sum_{k=1}^n Y_k \xrightarrow{\mathrm{P}} Y. \tag{3.62}$$

Define $\{Z_n\}_{n=1}^\infty$ by $Z_{2n-1} := X_n$ and $Z_{2n} := Y_n$. Prove that

$$\frac{1}{n}\sum_{k=1}^n Z_k \xrightarrow{\mathrm{P}} \frac{X + Y}{2}. \tag{3.63}$$

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 $X_n \xrightarrow{\mathrm{P}} x$.

Exercise 3.25. Let $X, \{X_n\}_{n=1}^{\infty}$ be real random variables with $X_n \xrightarrow{\mathrm{P}} X$ and suppose $X \neq 0$ a.s. Prove that $X_n^{-1}\mathbf{1}_{\{X_n \neq 0\}} \xrightarrow{\mathrm{P}} X^{-1}$.

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

$$(Y_1 \cdots Y_n)^{1/n} \xrightarrow{\text{a.s.}} \exp\bigl(\mathbb{E}[X_1]\bigr). \tag{3.64}$$

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

$$\frac{1}{n}\sum_{k=1}^{n} X_k^N \xrightarrow{\text{a.s.}} \mathbb{E}[X_1^N]. \tag{3.65}$$

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

$$\frac{\sum_{k=1}^{n} X_k}{\sum_{k=1}^{n} X_k^2} \xrightarrow{\text{a.s.}} \frac{m}{m^2 + v}. \tag{3.66}$$

Problem 3.30. Let $\{X_n\}_{n=1}^{\infty} \subset \mathcal{L}^2(\mathbb{P})$ be i.i.d. Prove that

$$\frac{1}{n}\sum_{k=1}^{n} \bigl(X_k - \mathbb{E}[X_1]\bigr)^2 \xrightarrow{\text{a.s.}} \operatorname{var}(X_1). \tag{3.67}$$

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*

$$f_n(x) := \min\{1, n \cdot \mathrm{dist}(x, F)\}, \quad \textit{where} \quad \mathrm{dist}(x, F) := \inf_{y\in F} |x - y|. \tag{4.10}$$

Then $\{f_n\}_{n=1}^{\infty} \subset C_b(\mathbb{R}^d)$, and for any $x \in \mathbb{R}^d$, $0 \le f_n(x) \le f_{n+1}(x) \le 1$ for any $n \in \mathbb{N}$ and $\lim_{n\to\infty} f_n(x) = \mathbf{1}_{\mathbb{R}^d \setminus F}(x)$.

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 $[-N, N]^d$.

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 $\mu = \nu$.*

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 $\mu_n \xrightarrow{\mathcal{L}} \mu$.*

[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 4.9. *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 $\mu_n \xrightarrow{\mathcal{L}} \mu$.*

Recall that, for $\mu \in \mathcal{P}(\mathbb{R}^d)$, F_μ denotes its distribution function given in Definition 2.18 (and in Definition 2.16 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 $N \in \mathbb{N}$.*

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 $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, y)$.*

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 $\mu_{n(k)} \xrightarrow{\mathcal{L}} \mu$.*

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 $\overline{e^{i\theta}} = e^{-i\theta}$, $\big|e^{i\theta}\big| = 1$ 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

$$\varphi_\mu(t) := \int_{\mathbb{R}^d} e^{i\langle t,x\rangle}\mu(dx) \qquad \text{for each } t \in \mathbb{R}^d, \tag{4.11}$$

where the integral is always defined since $x \mapsto e^{i\langle t,x\rangle}$ is continuous and $\big|e^{i\langle t,x\rangle}\big| = 1$.

(2) For a d-dimensional random variable X, we define its *characteristic function* $\varphi_X : \mathbb{R}^d \to \mathbb{C}$ by

$$\varphi_X(t) := \mathbb{E}\big[e^{i\langle t,X\rangle}\big] = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle}\mathbb{P}_X(dx) \qquad \text{for each } t \in \mathbb{R}^d, \tag{4.12}$$

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* $\{t_k\}_{k=1}^n \subset \mathbb{R}^d$,

$$\sum_{k,\ell=1}^{n} \varphi_\mu(t_k - t_\ell) z_k \overline{z_\ell} \ge 0. \tag{4.13}$$

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*

$$\frac{\partial^k \varphi_X}{\partial t_{j_1} \dots \partial t_{j_k}}(t) = i^k \mathbb{E}\big[X_{j_1} \cdots X_{j_k} e^{i\langle t,X\rangle}\big] \qquad \textit{for any } t = (t_1, \dots, t_d) \in \mathbb{R}^d. \tag{4.14}$$

In particular, for any $k \in \{1, \dots, n\}$ and $\{j_\ell\}_{\ell=1}^k \subset \{1, \dots, d\}$,

$$\mathbb{E}[X_{j_1} \cdots X_{j_k}] = (-i)^k \frac{\partial^k \varphi_X}{\partial t_{j_1} \dots \partial t_{j_k}}(0). \tag{4.15}$$

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*

$$f(h) = f(0) + f'(0)h + \frac{1}{2} f''(0)h^2 + o(h^2) \qquad \text{as } h \to 0, \tag{4.16}$$

$$f''(0) = \lim_{h \downarrow 0} \frac{f(h) + f(-h) - 2f(0)}{h^2}. \tag{4.17}$$

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*

$$\varphi_{X_1+\cdots+X_n}(t) = \varphi_{X_1}(t) \cdots \varphi_{X_n}(t) \qquad \text{for any } t \in \mathbb{R}^d. \tag{4.18}$$

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

$$\varphi_{X_1+\cdots+X_n}(t) = \varphi_\mu(t)^n \qquad \text{for any } t \in \mathbb{R}^d. \tag{4.19}$$

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*

$$\varphi_{TX+m}(t) = e^{i\langle t,m\rangle} \varphi_X(T^* t) \qquad \text{for any } t \in \mathbb{R}^k. \tag{4.20}$$

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

$$\varphi_X(t) = \bigl(1 + p(e^{it} - 1)\bigr)^n. \tag{4.21}$$

(2) If X has the Poisson distribution $\mathrm{Po}(\lambda)$, $\lambda \in (0, \infty)$, then

$$\varphi_X(t) = \exp\bigl(\lambda(e^{it} - 1)\bigr). \tag{4.22}$$

(3) If X has the geometric distribution $\mathrm{Geom}(\alpha)$, $\alpha \in [0, 1)$, then

$$\varphi_X(t) = \frac{1 - \alpha}{1 - \alpha e^{it}}. \tag{4.23}$$

(4) If X has the uniform distribution $\mathrm{Unif}(-a, a)$ on $(-a, a)$, $a \in (0, \infty)$, then

$$\varphi_X(t) = \frac{\sin at}{at}. \tag{4.24}$$

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 $t \in \mathbb{R}$,

$$\varphi_X(t) = \frac{\alpha}{\alpha - it}. \tag{4.25}$$

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 1.33) 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 $t \in \mathbb{R}$,

$$\varphi_X(t) = \exp\big(itm - t^2 v/2\big). \tag{4.26}$$

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 2.41 we have

$$\begin{aligned}\varphi_X(t) &= \int_{-\infty}^\infty e^{itx} \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(x-m)^2}{2v}\right) dx = e^{itm} \int_{-\infty}^\infty e^{it\sqrt{v}y} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy \\ &= e^{itm}\left(\int_{-\infty}^\infty \cos\big(t\sqrt{v}y\big)\frac{e^{-y^2/2}}{\sqrt{2\pi}} dy + i\int_{-\infty}^\infty \sin\big(t\sqrt{v}y\big)\frac{e^{-y^2/2}}{\sqrt{2\pi}} dy\right) \\ &= e^{itm}\varphi\big(\sqrt{v}t\big), \quad \text{where} \quad \varphi(t) := \int_{-\infty}^\infty \cos(tx)\frac{e^{-x^2/2}}{\sqrt{2\pi}} dx. \end{aligned} \tag{4.27}$$

Thus it suffices to calculate $\varphi(t)$ defined in (4.27). 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 1.33) twice, we obtain

$$\begin{aligned}\varphi'(t) &= -\int_{-\infty}^\infty x\sin(tx)\frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \lim_{n\to\infty}\int_{-n}^n \sin(tx)\frac{-xe^{-x^2/2}}{\sqrt{2\pi}} dx \\ &= \lim_{n\to\infty}\left(\left[\sin(tx)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\right]_{-n}^n - \int_{-n}^n t\cos(tx)\frac{e^{-x^2/2}}{\sqrt{2\pi}} dx\right) \\ &= -t\int_{-\infty}^\infty \cos(tx)\frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = -t\varphi(t).\end{aligned}$$

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 $\varphi_X(t) = e^{itm}\varphi\big(\sqrt{v}t\big) = e^{itm - t^2 v/2}$.

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 $t \in \mathbb{R}$,

$$\varphi_X(t) = \frac{\beta^\alpha}{(\beta - it)^\alpha}, \tag{4.28}$$

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 (4.28) is similar to the case of normal distributions in Example 4.22. Recalling Example 3.23, we have

$$\varphi_X(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{itx} x^{\alpha-1} e^{-\beta x}\, dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\beta - it)x}\, dx. \tag{4.29}$$

Since $\bigl|\frac{\partial}{\partial t}\bigl(x^{\alpha-1} e^{-(\beta-it)x}\bigr)\bigr| = x^\alpha e^{-\beta x}$ and $\int_0^\infty x^\alpha e^{-\beta x}\, dx < \infty$, by using Theorem 1.47 and the dominated convergence theorem (Theorem 1.33) we obtain

$$\begin{aligned}
\varphi_X'(t) &= \frac{i\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^\alpha e^{-(\beta-it)x}\, dx = \frac{i\beta^\alpha}{\Gamma(\alpha)} \lim_{n\to\infty} \int_0^n x^\alpha e^{-(\beta-it)x}\, dx \\
&= \frac{i\beta^\alpha}{\Gamma(\alpha)} \lim_{n\to\infty} \left(\left[\frac{-1}{\beta - it} x^\alpha e^{-(\beta-it)x} \right]_0^n + \frac{\alpha}{\beta - it} \int_0^n x^{\alpha-1} e^{-(\beta-it)x}\, dx \right) \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{i\alpha}{\beta - it} \int_0^\infty x^{\alpha-1} e^{-(\beta-it)x}\, dx = \frac{i\alpha}{\beta - it} \varphi_X(t).
\end{aligned}$$

Therefore $\frac{d}{dt}\bigl((\beta - it)^\alpha \varphi_X(t)\bigr) = (\beta - it)^{\alpha-1}\bigl((\beta - it)\varphi_X'(t) - i\alpha\varphi_X(t)\bigr) = 0$ and hence $(\beta - it)^\alpha \varphi_X(t) = (\beta - i0)^\alpha \varphi_X(0) = \beta^\alpha$ for any $t \in (0, \infty)$, which shows (4.29).

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 $t \in \mathbb{R}$,

$$\varphi_X(t) = \exp\bigl(itm - \alpha|t|\bigr). \tag{4.30}$$

Recalling Example 3.25 and using Corollary 2.41 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 (4.30) for $m = 0$ and $\alpha = 1$, that is, for any $t \in \mathbb{R}$,

$$\varphi_{\mathrm{Cauchy}(0,1)}(t) = \frac{1}{\pi} \int_{-\infty}^\infty e^{itx} \frac{1}{1 + x^2}\, dx = e^{-|t|}. \tag{4.31}$$

(4.31) 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 (4.31) 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*

$$\rho_\mu^{(v)}(x) := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi_\mu(t) \exp\bigl(-i\langle t, x\rangle - |t|^2 v/2\bigr) dt. \tag{4.32}$$

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 $\mu * N(0, n^{-2} I_d) \xrightarrow{\mathcal{L}} \mu$.*

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 $t_k \in \mathbb{R}^{d_k}$, $k \in \{1, \ldots, n\}$,*

$$\varphi_{(X_1,\ldots,X_n)}(t_1, \ldots, t_n) = \varphi_{X_1}(t_1) \cdots \varphi_{X_n}(t_n). \tag{4.33}$$

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, uniformly continuous on $\mathbb{R}^d$ and given by*

$$\rho(x) := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi_\mu(t) e^{-i\langle t, x\rangle} dt. \tag{4.34}$$

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

$$\varphi_{\mathrm{Cauchy}(0,1)}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} e^{itx} \frac{1}{1 + x^2} dx = e^{-|t|} \tag{4.31}$$

for any $t \in \mathbb{R}$, which completes the proof of (4.30) in Example 4.24. Indeed, Let $\mu \in \mathcal{P}(\mathbb{R})$ be the *Laplace distribution* (or *double exponential distribution*), defined by

$$\mu(dx) := \frac{1}{2} e^{-|x|} dx. \tag{4.35}$$

The reader will see in Problem 4.6 that for any $t \in \mathbb{R}$,

$$\varphi_\mu(t) = \frac{1}{1 + t^2}. \tag{4.36}$$

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 $\bigl|e^{-|x|} - \varphi_{\mathrm{Cauchy}(0,1)}(-x)\bigr|$ 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 (4.31).

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 $t \in \mathbb{R}^d$.*

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 $k \in \{1, \dots, d\}$, where $e_k := \big(\mathbf{1}_{\{k\}}(\ell)\big)_{\ell=1}^{d} \in \mathbb{R}^d$, 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 Unif$(-n, n)$ on $[-n, n]$. Then by (4.24) 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 $\varepsilon \in (0, \infty)$,*

$$\mu\big(\{x \in \mathbb{R} \mid |x| \geq 1/\varepsilon\}\big) \leq \frac{8}{\varepsilon} \int_0^{\varepsilon} \Big(1 - \mathrm{Re}\big(\varphi_\mu(t)\big)\Big) dt. \tag{4.37}$$

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 $t \in \mathbb{R}^d$.*

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 $\{t_k\}_{k=1}^{n} \subset \mathbb{R}^d$,*

$$\sum_{k,\ell=1}^{n} \varphi(t_k - t_\ell) z_k \overline{z_\ell} \geq 0. \tag{4.38}$$

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 $t \in \mathbb{R}^d$.*

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*

$$\lim_{n\to\infty} (1 + z_n/n)^n = e^z. \tag{4.39}$$

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 $t \in \mathbb{R}$,*

$$\lim_{n\to\infty} \mathbb{E}\left[\exp\left(it\frac{S_n - nm}{\sqrt{n}}\right)\right] = \exp\bigl(-t^2 v/2\bigr). \tag{4.40}$$

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 $X = (X_1, \dots, X_d)$ 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 $X = (X_1, \dots, X_d)$ is clearly *symmetric*, i.e. $v_{jk} = v_{kj}$ for any $j, k \in \{1, \dots, d\}$. Moreover, it is *non-negative definite*, i.e.

$$\langle a, Va\rangle = \sum_{j,k=1}^{d} v_{jk} a_j a_k \geq 0 \qquad \text{for any } a = (a_1, \dots, a_d) \in \mathbb{R}^d, \tag{4.41}$$

where $Va \in \mathbb{R}^d$ is the matrix product of V and a with a regarded as a *column* vector. Indeed,

$$\begin{aligned}\langle a, Va\rangle = \sum_{j,k=1}^{d} \mathrm{cov}(X_j, X_k) a_j a_k &= \mathbb{E}\left[\sum_{j,k=1}^{d} \bigl(X_j - \mathbb{E}[X_j]\bigr)\bigl(X_k - \mathbb{E}[X_k]\bigr) a_j a_k\right] \\ &= \mathbb{E}\left[\left(\sum_{k=1}^{d} a_k\bigl(X_k - \mathbb{E}[X_k]\bigr)\right)^2\right] \geq 0.\end{aligned}$$

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*

$$\varphi_\mu(t) = \exp\bigl(i\langle t, m\rangle - \langle t, Vt\rangle/2\bigr) \qquad \text{for any } t \in \mathbb{R}^d. \tag{4.42}$$

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

$$\varphi_{N(0,vI_d)}(t) = \varphi_{X_1}(t_1) \cdots \varphi_{X_d}(t_d) = e^{-t_1^2 v/2} \cdots e^{-t_d^2 v/2} = e^{-\langle t, vI_d t\rangle/2}, \tag{4.43}$$

where $\{X_k\}_{k=1}^d$ are i.i.d. d-dimensional random variables with $X_1 \sim N(0, v)$.

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*

$$N(m, V)(dx) = \frac{1}{(2\pi)^{d/2}\sqrt{\det V}} \exp\biggl(-\frac{\langle x - m, V^{-1}(x - m)\rangle}{2}\biggr) dx. \tag{4.44}$$

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*

$$\mathcal{L}\biggl(\frac{S_n - nm}{\sqrt{n}}\biggr) \xrightarrow{\mathcal{L}} N(0, V). \tag{4.45}$$

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) $\mathcal{L}\biggl(\dfrac{S_n - n}{\sqrt{n}}\biggr) \xrightarrow{\mathcal{L}} N(0, 1)$.

(2) $\mathbb{P}[S_n \le n] = e^{-n} \displaystyle\sum_{k=0}^n \frac{n^k}{k!}$ for any $n \in \mathbb{N}$.

(3) $\displaystyle\lim_{n\to\infty} e^{-n} \sum_{k=0}^n \frac{n^k}{k!} = \frac{1}{2}$.

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

$$X_n \xrightarrow{\mathcal{L}} X \quad \text{and} \quad Y_n \xrightarrow{\mathrm{P}} y. \tag{4.46}$$

(1) Prove that $X_n + Y_n \xrightarrow{\mathcal{L}} X + y$ and that $X_n Y_n \xrightarrow{\mathcal{L}} yX$.
(2) Suppose $y \neq 0$. Prove that

$$\frac{X_n}{Y_n}\mathbf{1}_{\{Y_n \neq 0\}} \xrightarrow{\mathcal{L}} \frac{X}{y}. \tag{4.47}$$

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 $n \in \mathbb{N}$,

$$\sqrt{S_n} - \sqrt{n} = \frac{S_n - n}{\sqrt{n}} \frac{1}{1 + \sqrt{S_n/n}}. \tag{4.48}$$

(2) Prove that

$$\mathcal{L}\big(\sqrt{S_n} - \sqrt{n}\big) \xrightarrow{\mathcal{L}} N(0, v/4). \tag{4.49}$$

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

$$\mathcal{L}\left(\frac{\sum_{k=1}^{n} X_k}{\sqrt{\sum_{k=1}^{n} X_k^2}} \mathbf{1}_{\{\sum_{k=1}^{n} X_k^2 \neq 0\}} \right) \xrightarrow{\mathcal{L}} N(0, 1). \tag{4.50}$$

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

$$\mu(dx) := \frac{1}{2} e^{-|x|} dx. \tag{4.35}$$

(μ is also called the *double exponential distribution*.) Prove that for any $t \in \mathbb{R}$,

$$\varphi_\mu(t) = \frac{1}{1 + t^2}. \tag{4.36}$$

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 $X + Y \sim N(m_1 + m_2, v_1 + v_2)$.
(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 $k \in \{1, \dots, n\}$. 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 $X \sim N(m, v)$.

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 $k \in \{1,\dots,n\}$. 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 $k \in \{1,\dots,n\}$. 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 $k \in \{1,\dots,n\}$. 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 $n \in \mathbb{N}$.

Problem 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 Problem 4.12 yields the characteristic function (4.26) 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 $\mathcal{L}(-X) = \mathcal{L}(X)$.

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 $\mathrm{Po}(\lambda)$ ($\lambda \in (0, \infty)$)	$\mu(\{n\}) = e^{-\lambda} \frac{\lambda^n}{n!}$	$\varphi_\mu(t) = \exp\big(\lambda(e^{it} - 1)\big)$
Geometric $\mathrm{Geom}(\alpha)$ ($\alpha \in [0, 1)$)	$\mu(\{n\}) = (1-\alpha)\alpha^n$	$\varphi_\mu(t) = \frac{1-\alpha}{1-\alpha e^{it}}$
Uniform $\mathrm{Unif}(a, b)$ ($a, b \in \mathbb{R}$, $a < b$)	$\mu(dx) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x)dx$	$\varphi_\mu(t) = \frac{e^{itb}-e^{ita}}{it(b-a)}$
Exponential $\mathrm{Exp}(\alpha)$ ($\alpha \in (0, \infty)$)	$\mu(dx) = \alpha e^{-\alpha x}\mathbf{1}_{(0,\infty)}(x)dx$	$\varphi_\mu(t) = \frac{\alpha}{\alpha - it}$
Gamma $\mathrm{Gamma}(\alpha, \beta)$ ($\alpha, \beta \in (0, \infty)$)	$\mu(dx) =$ $\frac{\beta^\alpha}{\Gamma(\alpha)}\lvert x\rvert^{\alpha-1}e^{-\beta x}\mathbf{1}_{(0,\infty)}(x)dx$	$\varphi_\mu(t) = \frac{\beta^\alpha}{(\beta - it)^\alpha}$
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 $\mathrm{Cauchy}(m, \alpha)$ ($m \in \mathbb{R}$, $\alpha \in (0, \infty)$)	$\mu(dx) = \frac{1}{\pi}\frac{\alpha}{\alpha^2 + (x-m)^2}dx$	$\varphi_\mu(t) = \exp\big(itm - \alpha\lvert t\rvert\big)$

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.