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Several theorems and exercises are adopted from an unpublished lecture note [6] on measure theory by Professor Jun Kigami in Kyoto University, and some other problems are borrowed from an unpublished lecture note by Professor Grigor'yan in the University of Bielefeld.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to express his deepest gratitude toward Professor Kigami and Professor Grigor'yan for their permission to quote their unpublished notes in this lecture note.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 $\mathbb{P}\bigl[\bigcup_{n=1}^\infty A_n\bigr] = \sum_{n=1}^\infty \mathbb{P}[A_n]$.

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:

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

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

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 $a_n \le a_{n+1}$ for any $n \in \mathbb{N}$, then $\lim_{n\to\infty} a_n = \sup_{n\ge 1} a_n$.*
(2) *If $a_n \ge a_{n+1}$ for any $n \in \mathbb{N}$, then $\lim_{n\to\infty} a_n = \inf_{n\ge 1} a_n$.*

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

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

Since the set $\{a_k \mid k \ge n\}$ is decreasing in n, $\sup_{k\ge n} a_k$ is non-increasing in n and $\inf_{k\ge n} a_k$ is non-decreasing in n, so that by Proposition 0.3,

$$\lim_{n\to\infty}\Bigl(\sup_{k\ge n} a_k\Bigr) = \limsup_{n\to\infty} a_n, \qquad \lim_{n\to\infty}\Bigl(\inf_{k\ge n} a_k\Bigr) = \liminf_{n\to\infty} a_n. \tag{0.8}$$

It also holds that

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

Indeed, $\inf_{k\ge m} a_k \le a_{\max\{m,n\}} \le \sup_{k\ge n} a_k$ for any $m, n \in \mathbb{N}$, and taking the infimum of the right-hand side in n shows that $\inf_{k\ge m} a_k \le \limsup_{n\to\infty} a_n$ 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

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

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 $a \le M$ for any $a \in A$ and $M \le b$ whenever $b \in [-\infty, \infty]$ satisfies $a \le b$ 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), $\sum_{i=1}^{k} a_i = \sum_{i\in\{1,\dots,k\}} a_i \leq \sup_{A \subset \mathbb{N}:\text{ finite}} \sum_{n\in A} a_n$ 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 $i, j \in \{1, \ldots, n\}$ with $i \neq j$, then $\mu(A_1 \cup \cdots \cup A_n) = \mu(A_1) + \cdots + \mu(A_n)$.*
(2) *If $A, B \in \mathcal{M}$ and $A \subset B$ then $\mu(A) \leq \mu(B)$.*
(3) *If $\{A_n\}_{n=1}^{\infty} \subset \mathcal{M}$ satisfies $A_n \subset A_{n+1}$ for any $n \in \mathbb{N}$, then $\lim_{n\to\infty} \mu(A_n) = \mu\bigl(\bigcup_{n=1}^{\infty} A_n\bigr)$.*
(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.

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

Then each $A \in \mathcal{B}(\mathbb{R}^d)$ is called a *Borel set of* $\mathbb{R}^d$. In fact, as stated in the following proposition, $\mathcal{B}(\mathbb{R}^d)$ is generated by d-dimensional intervals. As we will see in the course of this lecture, $\mathcal{B}(\mathbb{R}^d)$ is the right σ-algebra to be considered when dealing with measures on $\mathbb{R}^d$ and $\mathbb{R}^d$-valued functions.

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

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

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

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

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

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

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

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

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.

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

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 $\mathcal{F}$:

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

$\mathcal{F}$ 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 $\sup_{n\ge1} f_n$, $\inf_{n\ge1} f_n$, $\limsup_{n\to\infty} f_n$ and $\liminf_{n\to\infty} f_n$ on X by

$$\Big(\sup_{n\ge1} f_n\Big)(x) := \sup_{n\ge1}\big(f_n(x)\big), \qquad \Big(\limsup_{n\to\infty} f_n\Big)(x) := \limsup_{n\to\infty}\big(f_n(x)\big),$$
$$\Big(\inf_{n\ge1} f_n\Big)(x) := \inf_{n\ge1}\big(f_n(x)\big), \qquad \Big(\liminf_{n\to\infty} f_n\Big)(x) := \liminf_{n\to\infty}\big(f_n(x)\big).$$

Proposition 1.16. *Let $f_n : X \to [-\infty,\infty]$ be $\mathcal{M}$-measurable for each $n \in \mathbb{N}$. Then $\sup_{n\ge1} f_n$, $\inf_{n\ge1} f_n$, $\limsup_{n\to\infty} f_n$ 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) $0 \le s_n(x) \le s_{n+1}(x)$ *for any $n \in \mathbb{N}$,*

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

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

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

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 $\int_X s d\mu \le \int_X s d\mu + \int_X (t - s) d\mu = \int_X t d\mu$ for $\mathcal{M}$-simple functions $s, t : X \to [0, \infty)$ with $s \le t$ on X.

The following lemma is immediate from (1.18).

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

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 $f_n(x) \le f_{n+1}(x)$ 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*

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

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 $\mu\bigl(f^{-1}(\infty) \cup f^{-1}(-\infty)\bigr) = 0$.*

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

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 $\mu\bigl(\{x \in X \mid f(x) \neq g(x)\}\bigr) = 0$. 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)′ $|f_n(x)| \le g(x)$ 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*

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

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 $|(\partial f/\partial t)(x, t)| \le g(x)$ for any $(x, t) \in X \times (a, b)$. Then $\int_X f(x, \cdot) d\mu(x) : (a, b) \to \mathbb{R}$ is differentiable, and for any $t \in (a, b)$, $(\partial f/\partial t)(\cdot, t) \in \mathcal{L}^1(\mu)$ and*

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

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*

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

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*:

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

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 $\varphi(a) \in (0, \infty)$,

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

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:

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

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

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

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

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

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

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

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

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

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 $\nu\bigl(\bigcup_{n=1}^{\infty} A_n\bigr) \leq \sum_{n=1}^{\infty} \nu(A_n)$. (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 $\mu\big((-n, n]\big) < \infty$ for any $n \in \mathbb{N}$. Define $F : \mathbb{R} \to \mathbb{R}$ by*

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

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

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

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

Corollary 2.15 (Lebesgue measure on $\mathcal{B}(\mathbb{R})$). *There exists a unique Borel measure* m_1 *on $\mathbb{R}$ such that* $\mathrm{m}_1\big([a, b]\big) = b - a$ *for any $a, b \in \mathbb{R}$ with $a \leq b$.*

As already mentioned in Example 1.8, m_1 is called the *Lebesgue measure on* $\mathbb{R}$. 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 $F_\mu(x) := \mu\big((-\infty, x]\big)$ 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

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

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

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

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

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

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

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

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

$\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ 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 $i \in \{1, \ldots, n+k\}$ 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 $i \in \{1, \ldots, n\}$ 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 $A_i \in \mathcal{M}_i$, $i \in \{1, \ldots, n\}$,*

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

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

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

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 $a_i \leq b_i$, $i \in \{1, \ldots, d\}$,*

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

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*

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

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

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

(2) *Suppose that any one of $\int_{X \times Y} |f| d(\overline{\mu \times \nu})$, $\int_X \bigl(\int_Y |f(x, y)| d\nu(y) \bigr) d\mu(x)$ and $\int_Y \bigl(\int_X |f(x, y)| d\mu(x) \bigr) d\nu(y)$ is finite. Then $f(x, \cdot)$ is ν-integrable for μ-a.e. $x \in X$ with $\int_Y f(\cdot, y) d\nu(y)$ $\mathcal{M}$-measurable and μ-integrable, $f(\cdot, y)$ is μ-integrable for ν-a.e. $y \in Y$ with $\int_X f(x, \cdot) d\mu(x)$ $\mathcal{N}$-measurable and ν-integrable, f is $\overline{\mu \times \nu}$-integrable, and* (2.25) *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 $i \in \{1, \dots, d\}$ and set $I := [a_1, b_1] \times \cdots \times [a_d, b_d]$. Let $f : I \to \mathbb{R}$ be bounded and Riemann integrable on I. Then $f \in \mathcal{L}^1\big(I, \overline{\mathcal{B}(I)}^{\mathrm{m}_d}, \overline{\mathrm{m}_d}\big)$ 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 $a_k \in \mathbb{R}$, $k \ne \ell$. 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 $D\varphi(x) := \big((\partial\varphi_i/\partial x_j)(x)\big)_{i,j=1}^d$.
(2) φ is called a C^1*-embedding* if and only if φ is C^1 and injective and $D\varphi(x)$ is invertible for any $x \in U$.

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

Theorem 2.43 (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 $\Omega := \{0,1\}^{\mathbb{N}} = \big\{(\omega_n)_{n=1}^{\infty} \mid \omega_n \in \{0,1\}\big\}$, 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) $\mu\big([a,b]\big) = F(b) - F(a-)$.
(2) $\mu\big([a,b)\big) = F(b-) - F(a-)$.
(3) $\mu\big((a,b)\big) = F(b-) - F(a)$.
(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

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

for each $i \in \{1,\dots,d\}$ and $C_\mu := C_{\mu,1} \times \cdots \times C_{\mu,d}$. Prove the following statements:
(1) $\mathbb{R} \setminus C_{\mu,i}$ is a countable set for any $i \in \{1,\dots,d\}$.
(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 $i \in \{1,\dots,n\}$, let $(S_i,\mathcal{B}_i)$ be a measurable space and let $f_i : X \to S_i$. Prove that the map $f = (f_1,\dots,f_d) : X \to S_1 \times \cdots \times S_n$ 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 $i \in \{1,\dots,n\}$.

Problem 2.11. Let $n \in \mathbb{N}$. For each $i \in \{1,\dots,n\}$, 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 $i \in \{1,\dots,n\}$ define $F_i : X_1 \times \cdots \times X_n \to [-\infty,\infty]$ by $F_i(x_1,\dots,x_n) := f_i(x_i)$, and define $F : X_1 \times \cdots \times X_n \to [-\infty,\infty]$ by $F(x_1,\dots,x_n) := f_1(x_1)\cdots f_n(x_n)$. Prove the following statements:
(1) F_i is $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$-measurable for any $i \in \{1,\dots,n\}$.
(2) F is $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$-measurable.
(3) If f_i is μ_i-integrable for any $i \in \{1,\dots,n\}$, 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 $S_f := \{(x,t) \in X \times \mathbb{R} \mid 0 \le t < f(x)\}$.
(1) Prove that $S_f \in \mathcal{M} \otimes \mathcal{B}(\mathbb{R})$ and that $[0,\infty) \ni t \mapsto \mu\big(\{x \in X \mid f(x) > t\}\big) \in [0,\infty]$ 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

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