An Introduction to Probability Theory

Christel Geiss and Stefan Geiss
Department of Mathematics and Statistics
University of Jyväskylä (Finland)

These lecture notes are combined from [9] and [10] and are still work in progress.

October 31, 2019

Contents

Introduction

This book presents a self-contained introduction to probability based on measure and integration theory, without assuming knowledge of the latter beforehand. It evolved from courses given at the *University of Jyväskylä* (Finland) and the *University of Innsbruck* (Austria) with the aim to provide the necessary knowledge and tools that are required for further studies in stochastic process theory or stochastic analysis.

Probability theory can be understood as a mathematical model for the intuitive notion of *uncertainty*. Some parts of probability theory belong to pure mathematics, other parts to applied mathematics. Starting with the applied side, without probability theory all the stochastic models in physics, biology, and economics would either not have been developed or would not be rigorous. In order to be rigorous, a solid mathematical foundation is needed to assist us with various tools, like for example set theory, functional analysis, complex analysis, and the theory of special functions. On the other hand, probability is used in many branches of pure mathematics itself, even in branches where one does not expect this, like in convex geometry.

The modern period of probability theory is connected with names like S.N. Bernstein (1880-1968), E. Borel (1871-1956), and A.N. Kolmogorov (1903-1987). In particular, in 1933 A.N. Kolmogorov [14] published his modern approach of Probability Theory, including the notion of a measurable space and a probability space. This book will start from this notion to continue with random variables and basic parts of integration theory. Then we investigate the various types of convergence and conclude with the powerful theory of characteristic functions (known from real analysis as FOURIER-transforms). Having this tools, limit theorems will be studied, like the fundamental LAW OF LARGE NUMBERS or the CENTRAL LIMIT THEOREM.

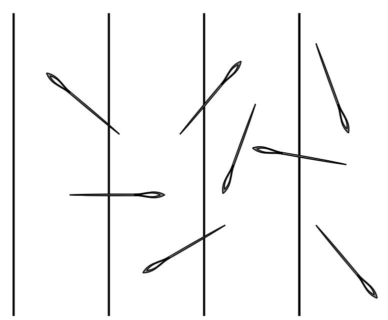
The book follows an axiomatic approach and is intended for students from mathematics, but also for other students who need more mathematical background for their further studies. We do not assume the knowledge of measure- and integration theory, instead we will introduce the necessary parts. In this way, we offer a somewhat demanding but rewarding access to rigorous probability theory.

Historical information about mathematicians can be found in the *MacTutor History of Mathematics Archive* under

www-history.mcs.st-andrews.ac.uk/history/index.html,

and this source has been also used throughout this book.

An example. Let us start with a classical example from 1733, Buffon's needle experiment [1] (see [5] and [2] for the historical background): We take needles of a fixed length $L > 0$ and a plain area with parallel lines, where the distance of two lines is the length L of the needles. We throw one needle after the other on the plane area, without caring about the direction and position of the needles, and count how many needles intersect one line or touch two lines. If we did throw n needles and counted f_n cases where a needle intersects one line or touches two lines, then we will observe for large n that $f_n/n \sim 2/\pi$, or more formally that



$$\lim_{n\to\infty} \frac{f_n}{n} = \frac{2}{\pi}.$$

So we obtain by a simple experiment the number π by $\pi = \lim_{n\to\infty} \frac{2n}{f_n}$. Besides this surprising fact, the basic question is how to model this experiment mathematically and how to confirm within this model the above observation. Intuitively, the randomness comes from two sources: the position of one end point of the needle and the angle against the parallel lines. It turns

[1] Georges Louis Leclerc Comte de Buffon, 07/09/1707 (Montbard, Côte d'Or, France) - 16/04/1788 (Paris, France)

out that we need to consider only one stripe so that we have the state space $\Omega := [0, L] \times [0, 2\pi)$ and the parameters $\omega = (l, \alpha)$ model the position of the needle in one stripe. Then we consider the map

$$f : \Omega = [0, L] \times [0, 2\pi) \to \{0, 1\}$$

such that $f(\omega) := 1$ if the needle intersects one line or touches two lines, and $f(\omega) := 0$ otherwise. Now we proceed as follows:

STEP 1: We model the randomness of our experiment by considering a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ will be the BOREL-σ-algebra on $[0, L] \times [0, 2\pi)$ and $\mathbb{P}$ be the normalized LEBESGUE measure. The knowledge about this will be provided in Chapters 1 and 2.

STEP 2: We interprete the map f as a measurable map or random variable. These concepts are introduced in Chapter 3.

STEP 3: We model the repeated throwing of a needle by taking *independent copies* $f_1, f_2, \ldots : \overline{\Omega} \to \mathbb{R}$ of f, where $(\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathbb{P}})$ is an appropriate 'larger' probability space. The concept of independence is introduced in Chapter 4.

STEP 4: In the final step we use the *Strong Law of Large Numbers* discussed in Chapter 8 and obtain that

$$\lim_{n\to\infty} \frac{f_1 + \cdots + f_n}{n} = \mathbb{E} f_1$$

in an almost sure sense. The various types of convergence are treated in Chapter 6, the *expected value* $\mathbb{E} f_1$ in Chapter 5. It turns out that $\mathbb{E} f_1 = \frac{2}{\pi}$ which confirms the experiment.

Some notation. We close with some notation used in the sequel.

set of all subsets of Ω	2^Ω	$=$	$\{A : A \subseteq \Omega\}$
empty set	$\emptyset$	$=$	set, without any element
For $A, B \subseteq \Omega$:			
intersection	$A \cap B$	$=$	$\{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$
union	$A \cup B$	$=$	$\{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\}$
set-theoretical minus	$A \backslash B$	$=$	$\{\omega \in \Omega : \omega \in A \text{ and } \omega \notin B\}$
complement	A^c	$=$	$\{\omega \in \Omega : \omega \notin A\}$
symmetric difference	$A \Delta B$	$=$	$(A \cup B) \backslash (A \cap B)$
real numbers	$\mathbb{R}$		
rational numbers	$\mathbb{Q}$		
natural numbers	$\mathbb{N}$	$=$	$\{1, 2, 3, \ldots\}$
natural numbers with 0	$\mathbb{N}_0$	$=$	$\{0, 1, 2, \ldots\}$
indicator-function	$1\!\!\mathrm{I}_A(x)$	$=$	$\begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$
extended real line	$[0, \infty]$	$=$	$[0, \infty) \cup \{\infty\}$
minimum	$\alpha \wedge \beta$	$=$	$\min\{\alpha, \beta\}$
cardinality of a set	$\#A$	$=$	number of elements in A

The sections that are marked with a * are intended for extended reading.

Chapter 1

Measure spaces and probability spaces

In this chapter we introduce measure spaces and, as a special case the probability space, the fundamental notion of probability theory. A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ consists of three components.

The **elementary events** or **states** ω which are collected in a non-empty set Ω.

Example 1.0.1. (1) If we roll a die, then the possible outcomes are the numbers between 1 and 6. We write

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

(2) If we flip a coin, then we have either "heads" or "tails" on top, so we put

$$\Omega = \{H, T\}.$$

If we have two coins, then

$$\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$$

is the set of all possible outcomes.

(3) The choice $\Omega = [0, \infty)$ could model the lifetime of a bulb.

A σ-**algebra** $\mathcal{F}$, which is a system of *observable* subsets or *events* $A \subseteq \Omega$, is the second component.

The third component is a **measure** $\mathbb{P}$, which assigns a **probability** to all $A \in \mathcal{F}$. This probability is a number $\mathbb{P}(A) \in [0,1]$ which measures the likelihood that a random experiment has an outcome ω which belongs to A. Let us illustrate this with the example of rolling a die. Here we choose $\mathcal{F}$ to contain all subsets of $\Omega = \{1,2,3,4,5,6\}$ and $\mathbb{P}(A) = \frac{\#A}{6}$. If $A = \{2,4,6\}$, then '$\omega \in A$' means 'the outcome of rolling the die is an even number.' We have

$$\mathbb{P}(A) = \frac{\#\{2,4,6\}}{6} = \frac{1}{2}.$$

We proceed now as follows: First we define σ-algebras. For this one does not need any measure. Afterwards we introduce measures and probability measures

1.1 σ-algebras

The σ-algebra is a basic tool in measure and probability theory. It serves as the domain of definition of measures. The notion of a σ-algebra is crucial to introduce the fundamental Lebesgue measure on the real line or Gaussian measures, without which many parts of mathematics could not live.

Definition 1.1.1 (σ-ALGEBRA, ALGEBRA, MEASURABLE SPACE)**.** Let Ω be a non-empty set. A system $\mathcal{F}$ of subsets $A \subseteq \Omega$ is called σ-**algebra** on Ω if

(1) $\emptyset, \Omega \in \mathcal{F}$,

(2) $A \in \mathcal{F}$ implies that $A^c = \Omega \backslash A \in \mathcal{F}$,

(3) $A_1, A_2, ... \in \mathcal{F}$ implies that $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

The pair $(\Omega, \mathcal{F})$, where $\mathcal{F}$ is a σ-algebra on Ω, is called **measurable space**. If one replaces (3) by

(3') $A, B \in \mathcal{F}$ implies that $A \cup B \in \mathcal{F}$,

then $\mathcal{F}$ is called an **algebra**.

Remark 1.1.2. (1) In probability theory and statistics the sets $A \in \mathcal{F}$ are called **events**. One says that an event A occurs if $\omega \in A$ and that it does not occur if $\omega \notin A$. The idea behind this is the intuition that one usually does not know in which particular $\omega \in \Omega$ a system is, but given an event $A \in \mathcal{F}$, one can decide at least whether $\omega \in A$ or $\omega \notin A$.

To illustrate this, consider the example of throwing of two coins and choose

$$A = \{(H,H),(T,T)\}.$$

If one is told that the outcome ω of throwing two coins is 'both coins show the same', then one would know that $\omega \in A$, but we can not deduce from the given information whether we have $\omega = (H,H)$ or $\omega = (T,T)$.

This way of thinking is in accordance with the properties of a σ-algebra: For example, if $\omega \in A$ can be decided, then $\omega \notin A$ can be decided, and if $\omega \in A_1$, $\omega \in A_2$, ...can be decided, then $\omega \in \bigcup_{i=1}^{\infty} A_i$ can be decided as well.

(2) One of the main aspects in Definition 1.1.1 is that we use an *infinite* union in (3), but do *not* allow uncountably many sets, only at most *countably many* sets.

(3) By 2^Ω we denote the so-called *power set*, the set which contains all subsets of Ω. The power set 2^Ω is always a σ-algebra. Non-trivial σ-algebras $\mathcal{F}$ on Ω in the sense that $\mathcal{F} \subsetneq 2^\Omega$ are needed for several reasons. For example, we will see that various important measures cannot be defined on 2^Ω, but only on smaller σ-algebras. Furthermore, the concept of σ-algebras is used to describe in stochastic modeling the notion of *information* or by a sequence of σ-algebras, a flow of information.

(4) Sometimes, the terms σ-field and field are used instead of σ-algebra and algebra, respectively.

The next proposition follows directly from Definition 1.1.1:

Proposition 1.1.3. (1) *Every σ-algebra is an algebra.*

(2) *Given a measurable space $(\Omega, \mathcal{F})$ and $A, B, A_1, A_2, \dots \in \mathcal{F}$, then*

$$\bigcap_{i=1}^{\infty} A_i \in \mathcal{F} \quad \textit{and} \quad A \backslash B \in \mathcal{F}.$$

Proof. (1) In Definition 1.1.1 we have that (3) implies (3'). This can be seen by taking $A_1 = A$ and $A_2 = A_3 = \cdots = B$. (2) is an exercise (Exercise 1). □

We continue with some examples.

Example 1.1.4 (The smallest and the largest σ-algebra)**.** For any σ-algebra $\mathcal{F}$ on Ω one has

$$\{\emptyset, \Omega\} \subseteq \mathcal{F} \subseteq 2^\Omega,$$

where one checks that 2^Ω and $\{\Omega, \emptyset\}$ are σ-algebras as well. Consequently, $\{\Omega, \emptyset\}$ and 2^Ω are the smallest and largest σ-algebras on Ω. Something 'in between' one can obtain by taking a set $A \subseteq \Omega$ not being empty nor being Ω itself. Then

$$\mathcal{F} = \{\Omega, \emptyset, A, A^c\}$$

is a σ-algebra not equal to $\{\Omega, \emptyset\}$ and 2^Ω.

Some more basic examples are the following:

Example 1.1.5 (Rolling Dice)**.** (1) Assume first the model for one die, i.e. $\Omega := \{1, ..., 6\}$ and $\mathcal{F} := 2^\Omega$. Then, for example, the event "the die shows an even number" is described by $A = \{2, 4, 6\}$.

(2) Assume now that the model with two dice, i.e. $\Omega := \{(a, b) : a, b = 1, ..., 6\}$. We want to describe the event "the sum of the two dice equals four", i.e. $A = \{(1, 3), (2, 2), (3, 1)\}$. There are at least two options to take an appropriate σ-algebra: Either the largest one $\mathcal{F}_1 := 2^\Omega$. But a smaller natural one would be

$$\mathcal{F}_2 := \{A \subseteq \Omega : A \text{ is the empty set or a union of } A_2, ..., A_{12}\},$$

where $A_m := \{(k, l) : k + l = m\}$. The σ-algebra $\mathcal{F}_2$ would fit the following model: One player throws two dice but only tells the other player the sum of the outcomes, not the values of the particular dice.

Example 1.1.6 (Heads and Tails)**.** (1) Assume a model for two coins, i.e. $\Omega := \{(H, H), (H, T), (T, H), (T, T)\}$ and $\mathcal{F} := 2^\Omega$. "Exactly one of two coins shows heads" is modelled by

$$A = \{(H, T), (T, H)\}.$$

(2) Assume now that we have a game with infinitely many trials. As set of elementary events we take the set of all possible outcomes, i.e. $\Omega := \{(a_1, a_2, a_3, ...) : a_k \in \{T, H\}\}$. Our game is such hat player I wins, if there are 10 heads more than tails for the first time, and player IIwins if there are 10 tails more than heads for the first time.

If Ω has finitely many elements, $\Omega = \{\omega_1, ..., \omega_n\}$, then any algebra $\mathcal{F}$ on Ω is automatically a σ-algebra. However, in general this is not the case as shown by the next example:

Example 1.1.7 (ALGEBRA, WHICH IS NOT A σ-ALGEBRA)**.** Let $\mathcal{A}$ be the system of subsets $A \subseteq \mathbb{R}$ such that A can be written as

$$A = (a_1, b_1] \cup (a_2, b_2] \cup \cdots \cup (a_n, b_n]$$

or

$$A = (a_1, b_1] \cup (a_2, b_2] \cup \cdots \cup (a_n, \infty)$$

where $-\infty \leqslant a_1 \leqslant b_1 \leqslant \cdots \leqslant a_n \leqslant b_n < \infty$ with the convention that $(a, a] = \emptyset$. Then $\mathcal{A}$ is an algebra, but not a σ-algebra (Exercise 2).

1.2 Measures

1.3 Definition of measures and first properties

Before we introduce the notion of a measure, we define what we mean by a partition of Ω.

Definition 1.3.1 (PARTITION)**.** Let Ω be a non-empty set and let I be a non-empty index-set. A system $(\Omega_i)_{i \in I}$ with $\Omega_i \subseteq \Omega$ is called partition of Ω if

(1) $\Omega_i \cap \Omega_j = \emptyset$ for $i \neq j$,

(2) $\Omega = \bigcup_{i \in I} \Omega_i$.

Definition 1.3.2 (MEASURE SPACE, PROBABILITY SPACE)**.** Let $(\Omega, \mathcal{F})$ be a measurable space.

(1) A map $\mu : \mathcal{F} \to [0, \infty]$ is called **measure** if

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

  (b) for all $A_1, A_2, ... \in \mathcal{F}$ with $A_i \cap A_j = \emptyset$ for $i \neq j$ one has

$$\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mu(A_i). \tag{1.1}$$

The triplet $(\Omega, \mathcal{F}, \mu)$ is called **measure space**.

(2) A measure space $(\Omega, \mathcal{F}, \mu)$ or measure μ is called **finite** if $\mu(\Omega) < \infty$.

(3) A measure space $(\Omega, \mathcal{F}, \mu)$ or measure μ is called σ**-finite** provided that there exists a partition $(\Omega_k)_{k=1}^{\infty} \subseteq \mathcal{F}$ of Ω such that for all $k = 1, 2, ...$ it holds $\mu(\Omega_k) < \infty$.

(4) A finite measure μ on $(\Omega, \mathcal{F})$ is called **probability measure** if $\mu(\Omega) = 1$. The triplet $(\Omega, \mathcal{F}, \mu)$ is called **probability space**.

Remark 1.3.3. (1) If we define a measure μ as above, but *without* the assumption $\mu(\emptyset) = 0$, and if we assume that there is at least one set $A \in \mathcal{F}$ with $\mu(A) < \infty$, then necessarily it holds that $\mu(\emptyset) = 0$: Indeed, we have the disjoint union $A = A \cup \bigcup_{i=1}^{\infty} \emptyset$ so that the σ-additivity (1.1) implies

$$\infty > \mu(A) = \mu(A) + \sum_{i=1}^{\infty} \mu(\emptyset)$$

and $\mu(\emptyset) = 0$. The only situation, where $\mu(\emptyset) > 0$ would be possible, consists in $\mu(A) \equiv \infty$ for all $A \in \mathcal{F}$, so that we would have $\mu(\emptyset) = \infty$ as well.

(2) In case that μ is a probability measure we will usually write $\mathbb{P}$ instead of μ.

Example 1.3.4. (1) We assume the model of a die, i.e. $\Omega = \{1, ..., 6\}$ and $\mathcal{F} = 2^{\Omega}$. Assuming that all outcomes for rolling a die are equally likely, leads to

$$\mathbb{P}(\{\omega\}) := \frac{1}{6}.$$

Then, for example, $\mathbb{P}(\{2, 4, 6\}) = \frac{1}{2}$.

(2) If we assume to have two coins, i.e.

$$\Omega = \{(T,T),(H,T),(T,H),(H,H)\}$$

and $\mathcal{F} = 2^\Omega$, then the intuitive assumption 'fair' leads to

$$\mathbb{P}(\{\omega\}) := \frac{1}{4} \quad \text{for all} \quad \omega \in \Omega.$$

That means, for example, the probability that exactly one of two coins shows heads is $\mathbb{P}(\{(H,T),(T,H)\}) = \frac{1}{2}$.

Example 1.3.5 (Dirac and counting measure [1]). Let Ω be an arbitrary non-empty set and $\mathcal{F} = 2^\Omega$.

(1) **Dirac measure**: For a fixed $\omega_0 \in \Omega$ we let

$$\delta_{\omega_0}(A) := \begin{cases} 1 & : \quad \omega_0 \in A \\ 0 & : \quad \omega_0 \notin A \end{cases}.$$

(2) **Counting measure**: Define

$$\mu(A) := \#A,$$

the cardinality of the set A. This measure is σ-finite if and only if Ω is countable.

(3) Let $B_0 \subseteq \Omega$ with $\#B_0 > 1$, and let

$$\mu(B) := \begin{cases} 1 & : \quad B_0 \cap B \neq \emptyset \\ 0 & : \quad \text{otherwise} \end{cases}.$$

One notes that μ is *not* a measure.

Let us now discuss a typical example in which the σ–algebra $\mathcal{F}$ is not the set of *all* subsets of Ω.

[1] Paul Adrien Maurice Dirac, 08/08/1902 (Bristol, England) - 20/10/1984 (Tallahassee, Florida, USA), Nobelprice in Physics 1933.

Example 1.3.6. Assume n communication channels between locations A and B. Each channel has a communication rate $\rho > 0$ (say ρ bits per second), which is assumed to yield to the rate ρk, in case k channels are working. Each channel fails independently with a probability $p \in (0, 1)$, so that we have a random communication rate $R \in \{0, \rho, ..., n\rho\}$. What is the *right* model for this from the viewpoint of a customer at A or B, who is not familiar with the technical details of the communication? We use

$$\Omega := \{\omega = (\varepsilon_1, ..., \varepsilon_n) : \varepsilon_i \in \{0, 1\}\}$$

with the interpretation that $\varepsilon_i = 0$ if channel i fails, $\varepsilon_i = 1$ if channel i works. The σ-algebra $\mathcal{F}$ consists of all unions of

$$A_k := \{\omega = (\varepsilon_1, \ldots, \varepsilon_n) \in \Omega : \varepsilon_1 + \cdots + \varepsilon_n = k\}$$

including the empty set. Hence A_k consists of all ω such that the communication rate is ρk. The system $\mathcal{F}$ is the system of observable sets (events) since one can only observe how many channels fail, but - as a customer in A or B - not which channels fail. The measure $\mathbb{P}$ is given by

$$\mathbb{P}(A_k) := \binom{n}{k} p^{n-k}(1-p)^k.$$

Note that $\mathbb{P}$ describes the **binomial distribution** with parameter p on $\{0, ..., n\}$ if we identify A_k with the natural number k. The binomial distribution is introduced in Section 2.1.1 below.

We close this section with some basic properties of measures.

Proposition 1.3.7. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Then the following assertions are true:*

(1) *If $A_1, \ldots, A_n \in \mathcal{F}$ are pair-wise disjoint, then $\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i)$.*

(2) *If $A, B \in \mathcal{F}$, then $\mu(B \cap A) + \mu(B \backslash A) = \mu(B)$. In particular, $\mu(A) \leqslant \mu(B)$ if $A \subseteq B$.*

(3) *If $A \in \mathcal{F}$ and μ is a probability measure, then $\mu(A^c) = 1 - \mu(A)$.*

(4) *If $A_1, A_2, ... \in \mathcal{F}$ then $\mu\left(\bigcup_{i=1}^\infty A_i\right) \leqslant \sum_{i=1}^\infty \mu(A_i)$.*

(5) Continuity from below: *If $A_1, A_2, \ldots \in \mathcal{F}$ such that $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$, then*

$$\lim_{n\to\infty} \mu(A_n) = \mu\left(\bigcup_{i=1}^{\infty} A_i\right).$$

(6) Continuity from above: *If $A_1, A_2, \ldots \in \mathcal{F}$ such that $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$ and $\mu(A_1) < \infty$, then*

$$\lim_{n\to\infty} \mu(A_n) = \mu\left(\bigcap_{i=1}^{\infty} A_i\right).$$

Proof. (1) We let $A_{n+1} = A_{n+2} = \cdots = \varnothing$ and use $\mu(\varnothing) = 0$, to get

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

(2) Since $(B \cap A) \cap (B \backslash A) = \varnothing$, we get that

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

(3) We apply (2) to $B = \Omega$ and exploit $\mu(\Omega) = 1$.

(4) Put $B_1 := A_1$ and $B_i := A_i \backslash (A_1 \cup \cdots \cup A_{i-1})$ for $i = 2, 3, \ldots$ Obviously, $\mu(B_i) \leqslant \mu(A_i)$ for all $i \geqslant 1$ because $B_i \subseteq A_i$. Since the sets B_i are pair-wise disjoint and $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$ it follows that

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \mu\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mu(B_i) \leqslant \sum_{i=1}^{\infty} \mu(A_i).$$

(5) We define $B_1 := A_1$, $B_2 := A_2 \backslash A_1$, $B_3 := A_3 \backslash A_2$, $B_4 := A_4 \backslash A_3$, ... and get that

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i \quad \text{and} \quad B_i \cap B_j = \varnothing$$

for $i \neq j$. Consequently,

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

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

(6) is an exercise (Exercise 3). □

1.4 Infinitely often occurring events: The lemmas of Fatou and Borel-Cantelli

We start with the notion of the limit superior and a limit inferior for sets:

Definition 1.4.1. [$\liminf_n A_n$ and $\limsup_n A_n$] Let Ω be non-empty and $A_1, A_2, \ldots$ subsets of Ω. Then

$$\liminf_n A_n := \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i \quad \text{and} \quad \limsup_n A_n := \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i.$$

The following lemma gives the interpretation of $\liminf_n A_n$ and $\limsup_n A_n$:

Lemma 1.4.2. (1) *We have that $\omega \in \liminf_n A_n$ if and only if there is an $n \geqslant 1$ such that for all $i \geqslant n$ we have $\omega \in A_i$.*

(2) *We have that $\omega \in \limsup_n A_n$ if and only if ω belongs to infinitely many of the A_n.*

Proof. (1) follows directly from the definition. Regarding (2) we observe that, again by the definition, we have that $\omega \in \limsup_n A_n$ if and only if for all $n \geqslant 1$ there is a $i \geqslant n$ such that $\omega \in A_i$. Hence, if ω belongs to infinitely many of the A_n, then this condition is satisfied. Conversely, assume that for all $n \geqslant 1$ there is a $i \geqslant n$ such that $\omega \in A_i$: We start by $n_1 = 1$ and find an $i_1 \geqslant 1$ with $\omega \in A_{i_1}$. Then, letting $n_2 := i_1 + 1$, we find an $i_2 \geqslant n_2$ with $\omega \in A_{i_2}$. Continuing this construction, we find $1 \leqslant i_1 < i_2 < \cdots$ with $\omega \in A_{i_l}$ for $l \geqslant 1$. □

Proposition 1.4.3 (Lemma of Fatou[2])**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A_1, A_2, ... \in \mathcal{F}$. Then*

$$\mathbb{P}\left(\liminf_n A_n\right) \leqslant \liminf_n \mathbb{P}(A_n) \leqslant \limsup_n \mathbb{P}(A_n) \leqslant \mathbb{P}\left(\limsup_n A_n\right).$$

Proof. From Proposition 1.3.7(5) we conclude

$$\mathbb{P}\left(\liminf_n A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{i=n}^{\infty} A_i\right), \tag{1.2}$$

[2] Pierre Joseph Louis Fatou, 28/02/1878-10/08/1929, French mathematician (dynamical systems, Mandelbrot-set).

where we use that $\bigcup_{i=n+1}^{\infty} A_i \subseteq \bigcup_{i=n}^{\infty} A_i$. Then the first inequality of the assertion follows from

$$\lim_{n\to\infty} \mathbb{P}\left(\bigcap_{i=n}^{\infty} A_i\right) = \liminf_{n\to\infty} \mathbb{P}\left(\bigcap_{i=n}^{\infty} A_i\right) \leqslant \liminf_{n\to\infty} \mathbb{P}(A_n).$$

The second inequality is true since the limit inferior of a real valued sequence is always less than or equal to its limit superior. The last relation follows from Proposition 1.3.7(6) as

$$\mathbb{P}\left(\limsup_n A_n\right) = \mathbb{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} A_i\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=n}^{\infty} A_i\right)$$
$$= \limsup_{n\to\infty} \mathbb{P}\left(\bigcup_{i=n}^{\infty} A_i\right) \geqslant \limsup_{n\to\infty} \mathbb{P}(A_n).$$

□

Corollary 1.4.4. *For a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and* $A_1, A_2, \ldots \in \mathcal{F}$ *one has that*

(1) $\liminf_n \mathbb{P}(A_n) = 0$ *implies* $\mathbb{P}(\liminf_n A_n) = 0$,

(2) $\limsup_n \mathbb{P}(A_n) = 1$ *implies* $\mathbb{P}(\limsup_n A_n) = 1$.

Now we turn to the fundamental notion of independence.

Definition 1.4.5 (INDEPENDENCE OF EVENTS)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The events $(A_i)_{i\in I} \subseteq \mathcal{F}$, where I is an arbitrary non-empty index set, are called **independent**, provided that for all $n \geqslant 2$ and distinct $i_1, ..., i_n \in I$ one has that

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_n}) = \mathbb{P}(A_{i_1})\,\mathbb{P}(A_{i_2})\cdots\mathbb{P}(A_{i_n}).$$

Given $A_1, ..., A_n \in \mathcal{F}$, one can easily see that only demanding

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\,\mathbb{P}(A_2)\cdots\mathbb{P}(A_n)$$

would not yield to an appropriate notion for the independence of $A_1, ..., A_n$: for example, taking A and B with

$$\mathbb{P}(A \cap B) \neq \mathbb{P}(A)\mathbb{P}(B)$$

and $C = \emptyset$ gives

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C),$$

which is surely not, what we had in mind. Now we continue with the fundamental Lemma of Borel-Cantelli:

Proposition 1.4.6 (Lemma of Borel-Cantelli[3]). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A_1, A_2, \ldots \in \mathcal{F}$. Then one has the following:*

(1) *If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$.*

(2) *If $A_1, A_2, ...$ are assumed to be independent and $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$.*

Consequently, if $A_1, A_2, \ldots$ are independent, then

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \Longleftrightarrow \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0$$

and

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty \Longleftrightarrow \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 1.$$

As a by-product we obtain our first **zero-one law**, namely that

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) \in \{0, 1\}$$

in case $A_1, A_2, \ldots$ are independent.

[3] Francesco Paolo Cantelli, 20/12/1875-21/07/1966, Italian mathematician.

Proof of Proposition 1.4.6. (1) From (1.2) we have that

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \leqslant \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathbb{P}\left(A_k\right) = 0,$$

where the last inequality follows again from Proposition 1.3.7.
(2) It holds that

$$\left(\limsup_n A_n\right)^c = \liminf_n A_n^c = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_n^c.$$

So, we would need to show

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_n^c\right) = 0.$$

Letting $B_n := \bigcap_{k=n}^{\infty} A_k^c$ we get that $B_1 \subseteq B_2 \subseteq B_3 \subseteq \cdots$, so that

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_n^c\right) = \lim_n \mathbb{P}(B_n)$$

so that it suffices to show

$$\mathbb{P}(B_n) = \mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right) = 0.$$

Since the independence of $A_1, A_2, ...$ implies the independence of $A_1^c, A_2^c, ...$ (Exercise 4), we finally get (setting $p_n := \mathbb{P}(A_n)$) that

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right) &= \lim_{N\to\infty, N\geqslant n} \mathbb{P}\left(\bigcap_{k=n}^{N} A_k^c\right) = \lim_{N\to\infty, N\geqslant n} \prod_{k=n}^{N} \mathbb{P}\left(A_k^c\right) \\
&= \lim_{N\to\infty, N\geqslant n} \prod_{k=n}^{N} (1-p_k) \leqslant \lim_{N\to\infty, N\geqslant n} \prod_{k=n}^{N} e^{-p_k} \\
&= \lim_{N\to\infty, N\geqslant n} e^{-\sum_{k=n}^{N} p_k} = e^{-\sum_{k=n}^{\infty} p_k} \\
&= 0
\end{aligned}$$

because of $\sum_{k=n}^{\infty} p_k = \infty$, where we have used that $1 - x \leqslant e^{-x}$. □

1.5 A first look at conditional probabilities and Bayes' rule

Independence can be also expressed through conditional probabilities.

Definition 1.5.1 (CONDITIONAL PROBABILITY)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$. Then

$$\mathbb{P}(B|A) := \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} \quad \text{for} \quad B \in \mathcal{F},$$

is called **conditional probability of** B **given** A.

Proposition 1.5.2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$. Then the following holds:*

(1) *For $\mathbb{P}_A(B) := \mathbb{P}(B|A)$, $B \in \mathcal{F}$, the triplet $(\Omega, \mathcal{F}, \mathbb{P}_A)$ is a probability space.*

(2) *For $B \in \mathcal{F}$ we have that A is independent from B if and only if $\mathbb{P}(B|A) = \mathbb{P}(B)$.*

The proof is subject to Exercise 5. Now we formulate Bayes' formula:

Proposition 1.5.3 (BAYES' FORMULA[4])**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\Omega = \bigcup_{i=1}^n B_i$ with $n \geqslant 2$, $B_i \in \mathcal{F}$, $B_i \cap B_j = \emptyset$ for $i \neq j$, and $\mathbb{P}(B_i) > 0$ for $i = 1, \ldots, n$. Then, for $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$ one has*

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{k=1}^n \mathbb{P}(A|B_k)\mathbb{P}(B_k)}.$$

Proof. We get that

$$\frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_{k=1}^n \mathbb{P}(A|B_k)\mathbb{P}(B_k)} = \frac{\mathbb{P}(A \cap B_j)}{\sum_{k=1}^n \mathbb{P}(A \cap B_k)} = \frac{\mathbb{P}(A \cap B_j)}{\mathbb{P}(A)}.$$

□

[4] Thomas Bayes, 1702-17/04/1761, English mathematician.

Remark 1.5.4. An event B_j is also called **hypothesis**, the probabilities $\mathbb{P}(B_j)$ the **prior probabilities** (or a priori probabilities), and the probabilities $\mathbb{P}(B_j|A)$ the **posterior probabilities** (or a posteriori probabilities) of B_j.

Example 1.5.5. We apply Proposition 1.5.3 for $n = 2$, $B_1 = B$, and $B_2 = B^c$, and get that

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)}.$$

A laboratory blood test is 95% effective in detecting a certain disease when it is, in fact, present. However, the test also yields a "false positive" result for 1% of the healthy persons tested. If 0.5% of the population actually has the disease, what is the probability a person has the disease given his test result is positive? We set

$$\begin{aligned} B &:= \text{"the person has the disease"}, \\ A &:= \text{"the test result is positive"}. \end{aligned}$$

Hence we have

$$\begin{aligned} \mathbb{P}(A|B) &= \mathbb{P}(\text{"a positive test result"}|\text{"person has the disease"}) = 0.95, \\ \mathbb{P}(A|B^c) &= 0.01, \\ \mathbb{P}(B) &= 0.005. \end{aligned}$$

Applying the above formula gives that

$$\mathbb{P}(B|A) = \frac{0.95 \times 0.005}{0.95 \times 0.005 + 0.01 \times 0.995} \approx 0.323.$$

That means only 32% of the persons whose test results are positive actually have the disease.

1.6 * The completion of measure spaces

In applications one needs often the following property of a measure space:

Definition 1.6.1. A probability space $(\Omega, \mathcal{F}, \mu)$ is called complete provided that $A \in \mathcal{F}$, $\mu(A) = 0$ and $B \subseteq A$ implies that $B \in \mathcal{F}$.

Given an arbitrary measure space $(\Omega, \mathcal{F}, \mu)$, we obtain a completion as follows:

Proposition 1.6.2. *For a measure space $(\Omega, \mathcal{F}, \mu)$ define $\overline{\mathcal{F}}^{\mu}$ to be the system of all sets $B \subseteq \Omega$ such that there are $A, B \in \mathcal{F}$ with $A \subseteq B \subseteq C$ and $\mu(C \backslash A) = 0$. Moreover, define $\overline{\mu}(B) := \mu(A) = \mu(B)$. Then the following assertions hold true:*

(1) *$\overline{\mu}$ is well-defined.*

(2) *$(\Omega, \overline{\mathcal{F}}, \overline{\mu})$ is a complete measure space.*

The proof is subject to Exercise 6.

1.7 Exercises

Ex 1: Given an algebra $\mathcal{A}$ and $A, B \in \mathcal{A}$, prove that $A \backslash B \in \mathcal{A}$.

Ex 2: Prove that the system $\mathcal{A}$ from Example 1.1.7 is an algebra but not a σ-algebra.

Ex 3: Proof statement (6) from Proposition 1.3.7.

Ex 4: For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and indpendent events $A_1, \ldots, A_n \in \mathcal{F}$ prove that $A_1^c, \ldots, A_n^n$ are independent as well.

Ex 5: Verify Proposition 1.5.2.

Ex 6: Verify Proposition 1.6.2.

Chapter 2

Construction of measure spaces

Distributions, which can be modelled on $\Omega \subseteq \mathbb{N}_0$, are often called discrete distributions. We start this chapter with a list of common examples of discrete distributions. Then we construct measures and distributions where $\Omega = \mathbb{R}$, for example the LEBESGUE measure and the GAUSSIAN distribution. We will learn that those measures need a smaller domain of definition than $2^{\mathbb{R}}$.

At the end of this chapter, in Section 2.3, we will have our first limit theorem, which states the convergence of the binomial distribution to the POISSON distribution.

In Section 8.4 one can see how some of the distributions, introduced below, are used in non-life insurance: The POISSON process, that models the number of incoming claims as time goes on, can be defined by the POISSON distribution. It is then somehow surprising that this process also relates to the exponential distribution.

2.1 Distributions on $\mathbb{N}_0$

2.1.1 Binomial distribution with parameter $0 < p < 1$

(1) $\Omega := \{0, 1, ..., n\}$.

(2) $\mathcal{F} := 2^{\Omega}$ (system of all subsets of Ω).

(3) $\mathbb{P}(B) = \mathrm{Bin}_{n,p}(B) := \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta_k(B)$, where

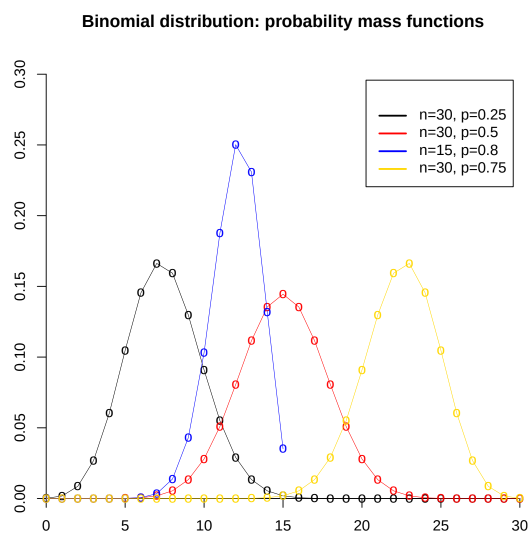
$$\delta_k(B) = \begin{cases} 1 & : \ k \in B \\ 0 & : \ k \notin B \end{cases}$$

is the DIRAC measure already introduced in Example 1.3.5.

INTERPRETATION: Coin-tossing with one coin, such that one has *heads* with probability p and *tails* with probability $1-p$. Then $\mathrm{Bin}_{n,p}(\{k\})$ equals the probability, that within n trials one has k-times *heads*. The function

$$\{0,1,...,n\} \ni k \mapsto \mathrm{Bin}_{n,p}(\{k\})$$

is called the probability mass function.



2.1.2 POISSON distribution with parameter $\lambda > 0$

(1) $\Omega := \mathbb{N}_0$.

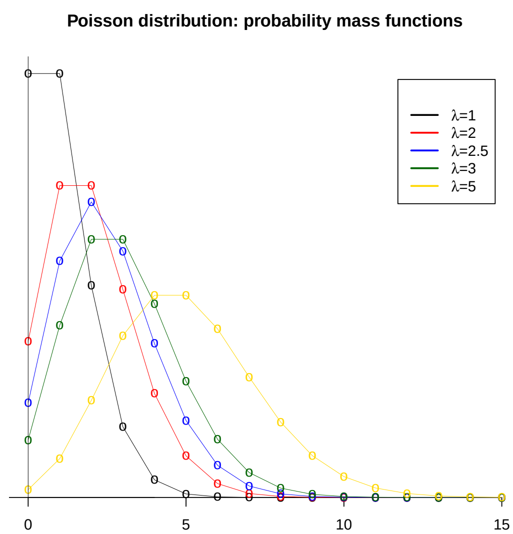
(2) $\mathcal{F} := 2^{\Omega}$ (system of all subsets of Ω).

(3) $\mathbb{P}(B) = \mathrm{Pois}_{\lambda}(B) := \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \delta_k(B)$.

The POISSON distribution[1] is used, for example, to model the POISSON process, a stochastic processes with a continuous time parameter and jumps:

[1] Siméon Denis Poisson, 21/06/1781 (Pithiviers, France) - 25/04/1840 (Sceaux, France).

The probability that a homogeneous POISSON process with intensity $\lambda > 0$ jumps k times between the time-points s and t with $0 \leqslant s < t < \infty$ is equal to $\text{Pois}_{\lambda(t-s)}(\{k\})$.



2.1.3 Geometric distribution with parameter $0 < p < 1$

(1) $\Omega := \mathbb{N}_0$.

(2) $\mathcal{F} := 2^\Omega$ (system of all subsets of Ω).

(3) $\mathbb{P}(B) = \text{Geom}_p(B) := \sum_{k=0}^{\infty}(1-p)^k p \delta_k(B)$.

INTERPRETATION: The probability that an electric light bulb has gone is $p \in (0,1)$. We assume that the bulb does not have a 'memory', that means its failure is independent of the time the bulb has been already switched on. So, we get the following model: At day 0 the probability of failure is p. If the bulb has not gone at day 0, it will fail again with probability p at the first day so that the total probability of failure at day 1 is $(1-p)p$. If we continue in this way, then we get that failure at day k has the probability $(1-p)^k p$.

2.2 Measures on $\mathbb{R}$

2.2.1 The Borel σ-algebra

Later we introduce the fundamental Lebesgue measure λ on the real line which simply measures the length of an interval, i.e. we want that

$$\lambda((a,b)) = b - a$$

for $-\infty < a < b < \infty$. So we have a set of elementary events, $\Omega = \mathbb{R}$ and a candidate for a natural measure, the Lebesgue measure. The question arises which sets can be measured by this type of measure. In particular, can we take the system of all subsets $\mathcal{F} = 2^{\Omega}$? Surprisingly it turns out that we cannot use this σ-algebra as it is too big. It can be shown that the Lebesgue measure cannot be constructed so that it can measure all subsets in an appropriate way (cf. the counter example in Section 10.3.3). This leads us to a fundamental task in measure theory: the construction of a reasonable σ-algebra for the Lebesgue measure. If one is modest, then the minimal requirement should be that this σ-algebra contains all open intervals. To get this idea working we use Proposition 2.2.2 below, which is based on the following fundamental construction:

Proposition 2.2.1 (INTERSECTION OF σ-ALGEBRAS IS A σ-ALGEBRA)**.** *Let Ω be an arbitrary non-empty set and let $\mathcal{F}_j$, $j \in J$, $J \neq \emptyset$, be a family of σ-algebras on Ω, where J is an arbitrary index set. Then*

$$\mathcal{F} := \bigcap_{j \in J} \mathcal{F}_j$$

is a σ-algebra as well.

Proof. First we notice that $\emptyset, \Omega \in \mathcal{F}_j$ for all $j \in J$, so that $\emptyset, \Omega \in \bigcap_{j\in J} \mathcal{F}_j$. Now let $A, A_1, A_2, ... \in \bigcap_{j\in J} \mathcal{F}_j$. Hence $A, A_1, A_2, ... \in \mathcal{F}_j$ for all $j \in J$, so that

$$A^c = \Omega \backslash A \in \mathcal{F}_j \quad \text{and} \quad \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}_j$$

for *all* $j \in J$. Consequently,

$$A^c \in \bigcap_{j\in J} \mathcal{F}_j \quad \text{and} \quad \bigcup_{i=1}^{\infty} A_i \in \bigcap_{j\in J} \mathcal{F}_j,$$

and $\bigcap_{j\in J} \mathcal{F}_j$ is a σ-algebra. □

Proposition 2.2.2 (Smallest σ-algebra containing a set-system). *Let $\Omega \neq \emptyset$, $\mathcal{G}$ be a non-empty system of subsets of Ω, and*

$$\sigma(\mathcal{G}) := \bigcap_{\mathcal{F}\ \sigma\text{-algebra and } \mathcal{F}\supseteq\mathcal{G}} \mathcal{F}.$$

Then $\sigma(\mathcal{G})$ is the smallest σ-algebra containing $\mathcal{G}$:

(1) *The system $\sigma(\mathcal{G})$ is a σ-algebra containing all sets from $\mathcal{G}$.*

(2) *If $\mathcal{F}$ is any σ-algebra on Ω containing all $G \in \mathcal{G}$, then $\sigma(\mathcal{G}) \subseteq \mathcal{F}$.*

Proof. We let

$$J := \left\{\mathcal{C} \subseteq 2^{\Omega} : \mathcal{C} \text{ is a } \sigma\text{–algebra and } \mathcal{G} \subseteq \mathcal{C}\right\}.$$

One has $J \neq \emptyset$, because $\mathcal{G} \subseteq 2^{\Omega}$ and 2^{Ω} is a σ-algebra. Hence

$$\sigma(\mathcal{G}) := \bigcap_{\mathcal{C}\in J} \mathcal{C}$$

yields to a σ-algebra according to Proposition 2.2.1 such that $\mathcal{G} \subseteq \sigma(\mathcal{G})$ by construction. It remains to show that $\sigma(\mathcal{G})$ is the smallest σ-algebra containing $\mathcal{G}$. Assume another σ-algebra $\mathcal{F}$ with $\mathcal{G} \subseteq \mathcal{F}$. By definition of J we have that $\mathcal{F} \in J$ so that

$$\sigma(\mathcal{G}) = \bigcap_{\mathcal{C}\in J} \mathcal{C} \subseteq \mathcal{F}. \qquad \square$$

The above construction of the smallest σ-algebra containing a set system is elegant and the basis for various fundamental measure spaces. One of the most important σ-algebras is the Borel σ-algebra on $\mathbb{R}$. There are several ways to define this Borel-σ-algebra. Below we illustrate some of them, where for the notion of open and closed sets we refer to Section 10.1.

Proposition 2.2.3 (Generating the Borel σ-algebra on $\mathbb{R}$). *We let*

(0) *$\mathcal{G}_0$ be the system of all open subsets of $\mathbb{R}$,*
(1) *$\mathcal{G}_1$ be the system of all closed subsets of $\mathbb{R}$,*
(2) *$\mathcal{G}_2$ be the system of all intervals $(-\infty, b]$, $b \in \mathbb{R}$,*
(3) *$\mathcal{G}_3$ be the system of all intervals $(-\infty, b)$, $b \in \mathbb{R}$,*

(4) $\mathcal{G}_4$ *be the system of all intervals* $(a, b]$, $-\infty < a < b < \infty$,
(5) $\mathcal{G}_5$ *be the system of all intervals* (a, b), $-\infty < a < b < \infty$.

Then $\sigma(\mathcal{G}_0) = \sigma(\mathcal{G}_1) = \sigma(\mathcal{G}_2) = \sigma(\mathcal{G}_3) = \sigma(\mathcal{G}_4) = \sigma(\mathcal{G}_5)$.

Proof. We only show that

$$\sigma(\mathcal{G}_0) = \sigma(\mathcal{G}_1) = \sigma(\mathcal{G}_3) = \sigma(\mathcal{G}_5),$$

the remaining parts are intended to be an exercise. Because of $\mathcal{G}_3 \subseteq \mathcal{G}_0$ one has

$$\sigma(\mathcal{G}_3) \subseteq \sigma(\mathcal{G}_0).$$

Moreover, for $-\infty < a < b < \infty$ one has that

$$(a, b) = \bigcup_{n=1}^{\infty} \left((-\infty, b) \backslash (-\infty, a + \frac{1}{n}) \right) \in \sigma(\mathcal{G}_3)$$

so that $\mathcal{G}_5 \subseteq \sigma(\mathcal{G}_3)$ and

$$\sigma(\mathcal{G}_5) \subseteq \sigma(\mathcal{G}_3).$$

Now let us assume a non-empty open set $A \subseteq \mathbb{R}$. For all $x \in A$ there is a *maximal* $\varepsilon_x > 0$ (cf. Exercise 1) such that

$$(x - \varepsilon_x, x + \varepsilon_x) \subseteq A.$$

Hence

$$A = \bigcup_{x \in A \cap \mathbb{Q}} (x - \varepsilon_x, x + \varepsilon_x),$$

which proves $\mathcal{G}_0 \subseteq \sigma(\mathcal{G}_5)$ and

$$\sigma(\mathcal{G}_0) \subseteq \sigma(\mathcal{G}_5).$$

Finally, $A \in \mathcal{G}_0$ implies $A^c \in \mathcal{G}_1 \subseteq \sigma(\mathcal{G}_1)$ and $A \in \sigma(\mathcal{G}_1)$. Hence $\mathcal{G}_0 \subseteq \sigma(\mathcal{G}_1)$ and

$$\sigma(\mathcal{G}_0) \subseteq \sigma(\mathcal{G}_1).$$

The remaining inclusion $\sigma(\mathcal{G}_1) \subseteq \sigma(\mathcal{G}_0)$ can be shown in the same way. □

Definition 2.2.4 (Borel σ-algebra on ℝ [2])**.** The σ-algebra on $\mathbb{R}$ generated in Proposition 2.2.3 is called **Borel** σ-algebra and denoted by $\mathcal{B}(\mathbb{R})$.

[2]Félix Edouard Justin Émile Borel, 07/01/1871-03/02/1956, French mathematician.

2.2.2 * The Lebesgue σ-algebra

2.2.3 Prerequisite 1: Carathéodory's extension

Although the definition of a measure is not difficult, to prove existence and uniqueness of measures may sometimes be demanding. The reason for this lies in the fact that, in certain cases, one can not construct σ-algebras explicitly, one only knows their existence. We will state Carathéodory's extension theorem and use it in the sequel as the prerequisite to construct, for instance, the Lebesgue measure and the Gaussian distribution.

Proposition 2.2.5 (Carathéodory's extension theorem[3]). *Let $\Omega \neq \emptyset$, $\mathcal{A}$ be an algebra such that $\mathcal{F} = \sigma(\mathcal{A})$. Assume a map $\mu_0 : \mathcal{A} \to [0,\infty]$ such that*

(1) $\mu_0(\Omega_n) < \infty$, $n = 1, 2, ...,$ *for some partition* $(\Omega_n)_{n=1}^\infty \subseteq \mathcal{A}$ *of* Ω,

(2) $\mu_0\left(\bigcup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \mu_0(A_n)$ *for pair-wise disjoint* $A_n \in \mathcal{A}$ *such that*

$$\bigcup_{n=1}^\infty A_n \in \mathcal{A}.$$

Then there exists a unique measure $\mu : \mathcal{F} \to [0,\infty]$ such that

$$\mu(A) = \mu_0(A) \quad \textit{for all} \quad A \in \mathcal{A}.$$

The measure μ in the theorem above is σ-finite because $\mu(\Omega_n) < \infty$ for all $n = 1, 2, \ldots$. A proof of Proposition 2.2.5 can be found in Section 10.3.2 below. From this proof we also get the following *outer regularity* of the measure μ:

Proposition 2.2.6. *Assume the setting and the notation from Proposition 2.2.5. Then, for all $A \in \mathcal{A}$ one has that*

$$\mu(A) = \inf\left\{\sum_{n=1}^\infty \mu(A_n) : A \subseteq \bigcup_{n=1}^\infty A_n, A_n \in \mathcal{A}.\right\}.$$

[3] Constantin Carathéodory, 13/09/1873 (Berlin, Germany) - 02/02/1950 (Munich, Germany).

If one is only interested in the uniqueness of measures one can also use the following approach as a replacement of CARATHÉODORY's extension theorem.

Definition 2.2.7. [π-SYSTEM] A system $\mathcal{P}$ of subsets $A \subseteq \Omega$ is called **π-system**, provided that

$$A \cap B \in \mathcal{P} \quad \text{for all} \quad A, B \in \mathcal{P}.$$

Any algebra is a π-system but a π-system is not an algebra in general, take for example the π-system $\mathcal{P} := \{(a,b) : -\infty < a < b < \infty\} \cup \{\emptyset\}$.

Proposition 2.2.8 (UNIQUENESS OF MEASURES)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space such that the following is satisfied:*

(1) *$\mathcal{F} = \sigma(\mathcal{P})$, where $\mathcal{P}$ is a π-system containing Ω.*

(2) *μ and ν are finite measures on $\mathcal{F}$ such that*

$$\mu(A) = \nu(A) \quad \textit{for all} \quad A \in \mathcal{P}.$$

Then $\mu(B) = \nu(B)$ for all $B \in \mathcal{F}$.

The proof can be found in Section 10.3.1.

2.2.4 LEBESGUE measure on $\mathbb{R}$

We will construct the Lebesgue measure on $\mathbb{R}$ using CARATHÉODORY'S extension theorem. For this purpose we let:

(1) $\Omega := \mathbb{R}$.

(2) $\mathcal{F} := \mathcal{B}(\mathbb{R})$.

(3) As generating algebra $\mathcal{A}$ for $\mathcal{B}(\mathbb{R})$ we take the algebra from Example 1.1.7, i.e. the system of subsets $A \subseteq \mathbb{R}$ such that A can be written as

$$A = (a_1, b_1] \cup (a_2, b_2] \cup \cdots \cup (a_n, b_n]$$

or

$$A = (a_1, b_1] \cup (a_2, b_2] \cup \cdots \cup (a_n, \infty)$$

where $n \in \mathbb{N}$ and $-\infty \leqslant a_1 \leqslant b_1 \leqslant \cdots \leqslant a_n \leqslant b_n < \infty$. For such a set A we let

$$\lambda_0(A) := \sum_{i=1}^{n}(b_i - a_i).$$

Proposition 2.2.9. *The system $\mathcal{A}$ is an algebra. The map $\lambda_0 : \mathcal{A} \to [0, \infty]$ is well-defined and satisfies the assumptions of* Carathéodory*'s extension theorem Proposition 2.2.5.*

Proof. The map λ_0 is well-defined as $\lambda_0(A)$ does not depend on the representation $A = (a_1, b_1] \cup (a_2, b_2] \cup \cdots \cup (a_n, b_n]$. We need to prove that assumption (2) of Proposition 2.2.5 is satisfied. It is easy to see that it is sufficient to show the following: Given $-\infty < a < b < \infty$ and pair-wise disjoint intervals $(a_n, b_n]$ with

$$(a, b] = \bigcup_{n=1}^{\infty}(a_n, b_n],$$

we have that $b - a = \sum_{n=1}^{\infty}(b_n - a_n)$. Let $\varepsilon \in (0, b - a)$ and observe that

$$[a + \varepsilon, b] \subseteq \bigcup_{n=1}^{\infty}\left(a_n, b_n + \frac{\varepsilon}{2^n}\right).$$

Hence we have an open covering of a compact set, and by the Heine-Borel theorem there exists a finite sub-cover:

$$[a + \varepsilon, b] \subseteq \bigcup_{n \in I(\varepsilon)}\left(a_n, b_n + \frac{\varepsilon}{2^n}\right) = \bigcup_{n \in I(\varepsilon)}(a_n, b_n] \cup \left(b_n, b_n + \frac{\varepsilon}{2^n}\right)$$

for some *finite* set $I(\varepsilon)$. The total length of the intervals $\left(b_n, b_n + \frac{\varepsilon}{2^n}\right)$ is at most $\varepsilon > 0$, so that

$$b - a - \varepsilon \leqslant \sum_{n \in I(\varepsilon)}(b_n - a_n) + \varepsilon \leqslant \sum_{n=1}^{\infty}(b_n - a_n) + \varepsilon.$$

Letting $\varepsilon \downarrow 0$ we arrive at

$$b - a \leqslant \sum_{n=1}^{\infty}(b_n - a_n).$$

Since $b - a \geqslant \sum_{n=1}^{N}(b_n - a_n)$ for all $N \geqslant 1$, the opposite inequality is obvious. □

Definition 2.2.10 (LEBESGUE[4] MEASURE)**.** The unique extension λ of λ_0 to $\mathcal{B}(\mathbb{R})$ according to Proposition 2.2.5 is called Lebesgue measure.

Accordingly, λ is the unique σ-finite measure on $\mathcal{B}(\mathbb{R})$ such that

$$\lambda((a,b]) = b - a \quad \text{for all } -\infty < a < b < \infty.$$

We close by an easy to handle characterization of null-sets from $\mathcal{B}(\mathbb{R})$:

Proposition 2.2.11. *A set $N \in \mathcal{B}(\mathbb{R})$ satisfies $\lambda(N) = 0$ if and only if for all $\varepsilon > 0$ there are open intervals (a_n, b_n) with $-\infty < a_n < b_n < \infty$ such that*

$$N \subseteq \bigcup_{n=1}^{\infty} (a_n, b_n) \quad \textit{and} \quad \sum_{n=1}^{\infty} |b_n - a_n| < \varepsilon.$$

The proof is subject to Exercise 2.

2.2.5 GAUSSian distribution on $\mathbb{R}$

(1) $\Omega := \mathbb{R}$.

(2) $\mathcal{F} := \mathcal{B}(\mathbb{R})$ Borel σ-algebra.

(3) We take the algebra $\mathcal{A}$ considered in Example 1.1.7 and define for $m \in \mathbb{R}$ and $\sigma^2 > 0$
$$\mathcal{N}^0_{m,\sigma^2}(A) := \sum_{i=1}^{n} \int_{a_i}^{b_i} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$
for $A := (a_1, b_1] \cup (a_2, b_2] \cup \cdots \cup (a_n, b_n]$ with $-\infty \leqslant a_1 \leqslant b_1 \leqslant \cdots \leqslant a_n \leqslant b_n < \infty$, where we consider the RIEMANN-integral [5] on the right-hand side. One can show (we do not do this here, but compare with Corollary 5.8.1 below) that $\mathcal{N}^0_{m,\sigma^2}$ satisfies the assumptions of Proposition 2.2.5, so that we can extend $\mathcal{N}^0_{m,\sigma^2}$ to a probability measure $\mathcal{N}_{m,\sigma^2}$ on $\mathcal{B}(\mathbb{R})$.

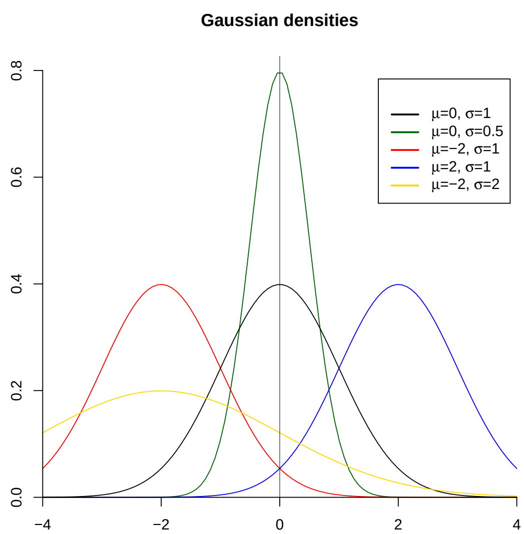
[4] Henri Léon Lebesgue, 28/06/1875-26/07/1941, French mathematician (generalized the Riemann integral by the Lebesgue integral; continuation of work of Emile Borel and Camille Jordan).

[5] Georg Friedrich Bernhard Riemann, 17/09/1826 (Germany) - 20/07/1866 (Italy), Ph.D. thesis under Gauss.

The measure $\mathcal{N}_{m,\sigma^2}$ is called **Gaussian distribution** [6] (**normal distribution**) with mean m and variance σ^2. Given $A \in \mathcal{B}(\mathbb{R})$ we write formally

$$\mathcal{N}_{m,\sigma^2}(A) = \int_A p_{m,\sigma^2}(x)d\lambda(x) \quad \text{with} \quad p_{m,\sigma^2}(x) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}},$$

to indicate that for general sets $A \in \mathcal{B}(\mathbb{R})$ the integral w.r.t. the Lebesgue measure λ is required (we will introduce this integral in Section 5). The function $p_{m,\sigma^2}(x)$ is called **Gaussian density**.



2.2.6 Exponential distribution on $\mathbb{R}$

(1) $\Omega := \mathbb{R}$.

(2) $\mathcal{F} := \mathcal{B}(\mathbb{R})$ Borel σ-algebra.

(3) For $A \in \mathcal{A}$ as in Subsection 2.2.5 we define, via the Riemann-integral,

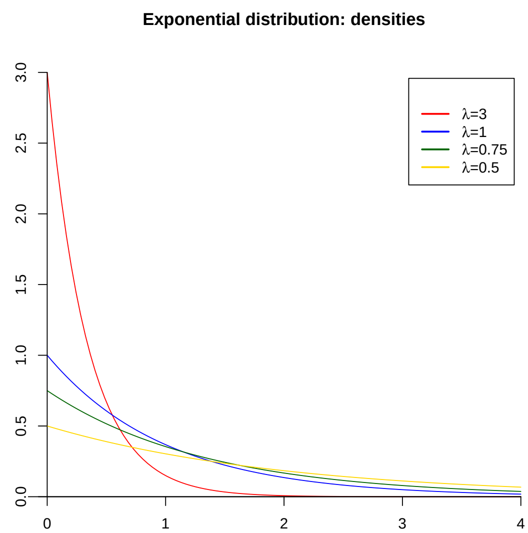
$$\operatorname{Exp}_\lambda^0(A) := \sum_{i=1}^n \int_{a_i}^{b_i} p_\lambda(x)dx \quad \text{with} \quad p_\lambda(x) := \mathbb{1}_{[0,\infty)}(x)\lambda e^{-\lambda x}$$

[6] Johann Carl Friedrich Gauss, 30/04/1777 (Brunswick, Germany) - 23/02/1855 (Göttingen, Hannover, Germany).

Again, ε_0 satisfies the assumptions of Proposition 2.2.5, so that we can extend Exp^0_λ to the **exponential distribution** Exp_λ with parameter $\lambda > 0$ and density $p_\lambda(x)$ on $\mathcal{B}(\mathbb{R})$.

Given $A \in \mathcal{B}(\mathbb{R})$ we write formally

$$\mathrm{Exp}_\lambda(A) = \int_A p_\lambda(x) d\lambda(x).$$



The exponential distribution does not have a memory in the sense that for $a, b \geqslant 0$ we have

$$\mathrm{Exp}_\lambda([a+b,\infty)|[a,\infty)) = \mathrm{Exp}_\lambda([b,\infty))$$

with the conditional probability on the left-hand side. This means, the probability of a realization larger or equal to $a + b$ under the condition that one has already a value larger or equal a is the same as having a realization larger or equal b. Indeed, this follows from

$$\begin{aligned}\mathrm{Exp}_\lambda([a+b,\infty)|[a,\infty)) &= \frac{\mathrm{Exp}_\lambda([a+b,\infty)\cap[a,\infty))}{\mathrm{Exp}_\lambda([a,\infty))} = \frac{\lambda\int_{a+b}^\infty e^{-\lambda x}dx}{\lambda\int_a^\infty e^{-\lambda x}dx} \\ &= \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} = \mathrm{Exp}_\lambda([b,\infty)).\end{aligned}$$

Example 2.2.12. Suppose that the amount of time one spends in a post office is exponential distributed with $\lambda = \frac{1}{10}$. We ask: (a) What is the probability, that a customer will spend more than 15 minutes? (b) What is the probability, that a customer will spend more than 15 minutes from the beginning in the post office, given that the customer already spent at least 10 minutes? The answer for (a) is $\mathrm{Exp}_\lambda([15,\infty)) = e^{-15\frac{1}{10}} \approx 0.220$. For (b) we get $\mathrm{Exp}_\lambda([15,\infty)|[10,\infty)) = \mathrm{Exp}_\lambda([5,\infty)) = e^{-5\frac{1}{10}} \approx 0.604$.

2.2.7 Prerequisite 2: The trace σ-algebra

Another tool to construct measures consists in the concept of a trace σ-algebra: Let $(\Omega, \mathcal{F})$ be a measurable space. Sometimes one needs a σ-algebra on a subset $A \subseteq \Omega$. This can be constructed as follows.

Lemma 2.2.13. *If $(\Omega, \mathcal{F})$ is a measurable space and $\emptyset \neq A \subseteq \Omega$, then*

$$\mathcal{F}_A := \{F \cap A : F \in \mathcal{F}\}$$

is a σ-algebra on A.

Proof. Choosing $F = \emptyset$ and $F = \Omega$ implies $\emptyset \in \mathcal{F}_A$ and $A \in \mathcal{F}_A$, respectively. Any element $G \in \mathcal{F}_A$ can be written as $G = F \cap A$ for some $F \in \mathcal{F}$. Therefore $A \backslash G \in \mathcal{F}_A$ follows from $F^c \in \mathcal{F}$ and

$$A \backslash G = A \backslash (F \cap A) = F^c \cap A.$$

Assume that for $k = 1, 2, \ldots$ the G_k are given by $G_k = F_k \cap A$ with $F_k \in \mathcal{F}$. Then

$$\bigcup_{k=1}^{\infty} G_k = \bigcup_{k=1}^{\infty} (F_k \cap A) = \left(\bigcup_{k=1}^{\infty} F_k \right) \cap A \in \mathcal{F}_A$$

since $\bigcup_{k=1}^{\infty} F_k \in \mathcal{F}$. □

Definition 2.2.14. The σ-algebra $\mathcal{F}_A$ on A constructed in Lemma 2.2.13 is called the **trace of $\mathcal{F}$ on A.**

Lemma 2.2.15. *If $\Omega \neq \emptyset$, if $\mathcal{G}$ is a non-empty system of subsets of Ω, and if $A \subseteq \Omega$ in non-empty, then*

$$\sigma(\mathcal{G})_A = \sigma(\mathcal{G}_A)$$

where $\mathcal{G}_A := \{G \cap A : G \in \mathcal{G}\}$ and $\sigma(\mathcal{G}_A)$ denotes the smallest σ-algebra on A which contains $\mathcal{G}_A$.

Proof. (a) We show the inclusion $\sigma(\mathcal{G}_A) \subseteq \sigma(\mathcal{G})_A$: We have that $\mathcal{G} \subseteq \sigma(\mathcal{G})$ so that $\mathcal{G}_A \subseteq \sigma(\mathcal{G})_A$. Lemma 2.2.13 implies that

$$\sigma(\mathcal{G}_A) \subseteq \sigma(\mathcal{G})_A.$$

(b) For the inclusion $\sigma(\mathcal{G}_A) \supseteq \sigma(\mathcal{G})_A$ we use the *principle of good sets* and define

$$\mathcal{G}_0 := \{G \in \sigma(\mathcal{G}) : G \cap A \in \sigma(\mathcal{G}_A)\}$$

so that

$$\mathcal{G} \subseteq \mathcal{G}_0 \subseteq \sigma(\mathcal{G}).$$

If $\mathcal{G}_0$ would be a σ-algebra, then we would have $\mathcal{G}_0 = \sigma(\mathcal{G})$ and $G \cap A \in \sigma(\mathcal{G}_A)$ would hold for any $G \in \sigma(\mathcal{G})$, i.e. $\sigma(\mathcal{G}_A) \supseteq \sigma(\mathcal{G})_A$. So it remains to show that $\mathcal{G}_0$ is a σ-algebra:

(1) Since $A, \emptyset \in \sigma(\mathcal{G}_A)$ we conclude from the definition of $\mathcal{G}_0$ that $\Omega, \emptyset \in \mathcal{G}_0$.
(2) Assume $G \in \mathcal{G}_0$. Then $G \cap A \in \sigma(\mathcal{G}_A)$, and since $\sigma(\mathcal{G}_A)$ is a σ-algebra on A, also $G^c \cap A = A \backslash (G \cap A) \in \sigma(\mathcal{G}_A)$. Hence $G^c \in \mathcal{G}_0$.
(3) Similarly as in the proof of Lemma 2.2.13 one derives from $G_1, G_2, \ldots \in \mathcal{G}_0$ that $\bigcup_{n=1}^{\infty} G_n \in \mathcal{G}_0$. □

2.2.8 Uniform distribution on $[0, 1]$

Now we obtain the uniform distribution on $[0, 1]$ from the Lebesgue-measure on $\mathbb{R}$ constructed before:

(1) $\Omega := [0, 1]$.

(2) $\mathcal{F}$ is the trace σ-algebra of $\mathcal{B}(\mathbb{R})$ on $[0, 1]$ which is denoted by $\mathcal{B}([0, 1])$.

(3) $\mathcal{U}_{[0,1]}(A) := \lambda(A)$ for $A \in \mathcal{B}([0, 1])$, where λ is the Lebesgue-measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

2.3 A first limit theorem: Poisson's Theorem

For large n and small p the Poisson distribution provides a good approximation for the binomial distribution in the sense of Proposition 2.3.1 below. This proposition describes the *weak convergence* investigated later in Section 9.7:

Proposition 2.3.1 (Poisson's Theorem)**.** *Let* $\lambda > 0$*,* $p_n \in (0,1)$*,* $n = 1, 2, ...,$ *and assume that* $np_n \to \lambda$ *as* $n \to \infty$*. Then, for all* $k = 0, 1, \ldots,$

$$\mathrm{Bin}_{n,p_n}(\{k\}) \to \mathrm{Pois}_\lambda(\{k\}), \quad n \to \infty.$$

Proof. Fix an integer $k \geqslant 0$. Then

$$\begin{aligned} \mathrm{Bin}_{n,p_n}(\{k\}) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\ &= \frac{n(n-1)\ldots(n-k+1)}{k!} p_n^k (1-p_n)^{n-k} \\ &= \frac{1}{k!} \frac{n(n-1)\ldots(n-k+1)}{n^k} (np_n)^k \, (1-p_n)^{n-k}. \end{aligned}$$

Of course, $\lim_{n\to\infty}(np_n)^k = \lambda^k$ and $\lim_{n\to\infty} \frac{n(n-1)\ldots(n-k+1)}{n^k} = 1$. So we have to show that $\lim_{n\to\infty}(1-p_n)^{n-k} = e^{-\lambda}$. By $np_n \to \lambda$ we get that there exists a sequence ε_n such that

$$np_n = \lambda + \varepsilon_n \text{ with } \lim_{n\to\infty} \varepsilon_n = 0.$$

Choose $\varepsilon_0 \in (0, \lambda)$ and $n_0 \geqslant 1$ such that $|\varepsilon_n| \leqslant \varepsilon_0$ for all $n \geqslant n_0$. Then

$$\left(1 - \frac{\lambda+\varepsilon_0}{n}\right)^{n-k} \leqslant \left(1 - \frac{\lambda+\varepsilon_n}{n}\right)^{n-k} \leqslant \left(1 - \frac{\lambda-\varepsilon_0}{n}\right)^{n-k}.$$

Using l'Hospital's rule we get

$$\begin{aligned} \lim_{n\to\infty} \ln\left(1 - \frac{\lambda+\varepsilon_0}{n}\right)^{n-k} &= \lim_{n\to\infty}(n-k)\ln\left(1 - \frac{\lambda+\varepsilon_0}{n}\right) \\ &= \lim_{n\to\infty} \frac{\ln\left(1 - \frac{\lambda+\varepsilon_0}{n}\right)}{1/(n-k)} \\ &= \lim_{n\to\infty} \frac{\left(1 - \frac{\lambda+\varepsilon_0}{n}\right)^{-1} \frac{\lambda+\varepsilon_0}{n^2}}{-1/(n-k)^2} \\ &= -(\lambda+\varepsilon_0). \end{aligned}$$

Hence

$$e^{-(\lambda+\varepsilon_0)} = \lim_{n\to\infty}\left(1 - \frac{\lambda+\varepsilon_0}{n}\right)^{n-k} \leqslant \liminf_{n\to\infty}\left(1 - \frac{\lambda+\varepsilon_n}{n}\right)^{n-k}.$$

In the same way we get that

$$\limsup_{n\to\infty}\left(1-\frac{\lambda+\varepsilon_n}{n}\right)^{n-k}\leqslant e^{-(\lambda-\varepsilon_0)}.$$

Finally, since we can choose $\varepsilon_0>0$ arbitrarily small

$$\lim_{n\to\infty}(1-p_n)^{n-k}=\lim_{n\to\infty}\left(1-\frac{\lambda+\varepsilon_n}{n}\right)^{n-k}=e^{-\lambda}.$$ □

2.4 Exercises

Ex 1: One proves that a set $G\subseteq\mathbb{R}$ is open if and only if there are countable many disjoint open intervals (a_i,b_i), $i\in I$, with $-\infty\leqslant a_i<b_i\leqslant\infty$ such that $G=\bigcup_{i\in I}(a_i,b_i)$.

Ex 2: Verfiy Proposition 2.2.11.

Chapter 3

Random variables and measurable maps

So far we made ourselves familiar with probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$. Our next step is to introduce the concept of random variables, which reveals the full strength of probability and its applications in other branches of mathematics, in statistics, or in stochastic modeling. For example, given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we wish to consider functions $f : \Omega \to \mathbb{R}$ that describe certain random phenomena. In order to describe probabilistic properties of these phenomena one is naturally interested in expressions like

$$\mathbb{P}\left(\{\omega \in \Omega : f(\omega) \in (a,b)\}\right) \quad \text{where} \quad a < b.$$

This leads us to the condition

$$\{\omega \in \Omega : f(\omega) \in (a,b)\} \in \mathcal{F}$$

and hence to random variables as we will see now.

3.1 Random variables

We start with the most simple random variables.

Definition 3.1.1 (SIMPLE FUNCTION)**.** Let $(\Omega, \mathcal{F})$ be a measurable space. A function $f : \Omega \to \mathbb{R}$ is called **simple function**, provided that there are $\alpha_1, ..., \alpha_n \in \mathbb{R}$ and $A_1, ..., A_n \in \mathcal{F}$ such that f can be written as

$$f(\omega) = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}(\omega) \quad \text{where} \quad \mathbb{1}_{A_i}(\omega) = \begin{cases} 1 & : \ \omega \in A_i \\ 0 & : \ \omega \notin A_i \end{cases}.$$

The definition above concerns only functions that take finitely many values, which will be too restrictive in future. So we wish to extend this definition.

Definition 3.1.2 (MEASURABLE FUNCTIONS, RANDOM VARIABLES)**.** Let $(\Omega, \mathcal{F})$ be a measurable space. A function $f : \Omega \to \mathbb{R}$ is called **measurable** provided that there is a sequence $(f_n)_{n=1}^{\infty}$ of simple functions $f_n : \Omega \to \mathbb{R}$ such that

$$f(\omega) = \lim_{n\to\infty} f_n(\omega) \quad \text{for all} \quad \omega \in \Omega.$$

It is common to call a measurable function a **random variable** in the presence of a probability measure $\mathbb{P}$ on the measurable space $(\Omega, \mathcal{F})$, i.e. when we are given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. One can also couple the notion of a random variable on the underlying measure $\mathbb{P}$ by only requiring that there is set $\Omega_0 \in \mathcal{F}$ with $\mathbb{P}(\Omega_0) = 1$ such that $f(\omega) = \lim_{n\to\infty} f_n(\omega)$ only for $\omega \in \Omega_0$. This notion would not be independent of the measure and we will *not* follow this approach.

Does Definition 3.1.2 give what we would like to have? Yes, as we see from the following result.

Proposition 3.1.3. *Let $(\Omega, \mathcal{F})$ be a measurable space and let $f : \Omega \to \mathbb{R}$ be a function. Then the following conditions are equivalent:*

(1) *f is measurable.*

(2) *For all $-\infty < a < b < \infty$ one has that*

$$f^{-1}((a,b)) := \{\omega \in \Omega : a < f(\omega) < b\} \in \mathcal{F}.$$

Proof. (1) $\Longrightarrow$ (2) Assume that

$$f(\omega) = \lim_{n\to\infty} f_n(\omega)$$

where $f_n : \Omega \to \mathbb{R}$ are simple functions. For a simple function one has that

$$f_n^{-1}((a,b)) \in \mathcal{F}$$

so that

$$f^{-1}((a,b)) \quad = \quad \left\{\omega \in \Omega : a < \lim_n f_n(\omega) < b\right\}$$

$$= \bigcup_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \left\{ \omega \in \Omega : a + \frac{1}{m} < f_n(\omega) < b - \frac{1}{m} \right\} \in \mathcal{F}.$$

(2) $\Longrightarrow$ (1) First we observe that we also have that

$$\begin{aligned} f^{-1}([a,b)) &= \{\omega \in \Omega : a \leqslant f(\omega) < b\} \\ &= \bigcap_{m=1}^{\infty} \left\{ \omega \in \Omega : a - \frac{1}{m} < f(\omega) < b \right\} \in \mathcal{F} \end{aligned}$$

so that we can use the simple functions

$$f_n(\omega) := \sum_{k=-4^n}^{4^n-1} \frac{k}{2^n} \mathbb{I}_{\{\frac{k}{2^n} \leqslant f < \frac{k+1}{2^n}\}}(\omega).$$ □

The property of being measurable is stable with respect to point-wise convergence:

Proposition 3.1.4. *Assume a measurable space* $(\Omega, \mathcal{F})$ *and a sequence of measurable* $f_n : \Omega \to \mathbb{R}$ *such that* $f(\omega) := \lim_n f_n(\omega)$ *exists for all* $\omega \in \Omega$. *Then* $f : \Omega \to \mathbb{R}$ *is measurable.*

Proof. In fact, we did already prove this statement in the proof of Proposition 3.1.3 as for $-\infty < a < b < \infty$ we again have that

$$f^{-1}((a,b)) = \bigcup_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \left\{ \omega \in \Omega : a + \frac{1}{m} < f_n(\omega) < b - \frac{1}{m} \right\} \in \mathcal{F}.$$ □

As a direct consequence of Proposition 3.1.3 we obtain

Corollary 3.1.5. *Let* $(\Omega, \mathcal{F})$ *be a measurable space,* $f_1, \ldots, f_N : \Omega \to \mathbb{R}$ *be* Borel*-measurable, and* $F : \mathbb{R}^N \to \mathbb{R}$ *be continuous. Then* $F(f_1, \ldots, f_N) : \Omega \to \mathbb{R}$ *is measurable.*

Proof. For all $l \in \{1, \ldots, N\}$ assume measurable simple functions $f_n^l : \Omega \to \mathbb{R}$ such that

$$\lim_{n\to\infty} f_n^l(\omega) = f_l(\omega) \quad \text{for all} \quad \omega \in \Omega.$$

Then one can show that the function $F(f_1^l, \ldots, f_N^l) : \Omega \to \mathbb{R}$ is a simple function and, by the continuity of F, one has that

$$\lim_{n\to\infty} F(f_n^1(\omega), \ldots, f_n^N(\omega)) = F(f_1(\omega), \ldots, f_N(\omega)).$$

□

Later we will see that we can weaken the assumption on F and require BOREL measurability instead of continuity.

Proposition 3.1.6 (PROPERTIES OF MEASURABLE MAPS)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space, $f, g : \Omega \to \mathbb{R}$ random variables, and $\alpha, \beta \in \mathbb{R}$. Then the following is true:*

(1) $(\alpha f + \beta g)(\omega) := \alpha f(\omega) + \beta g(\omega)$ *is measurable.*

(2) $(fg)(\omega) := f(\omega)g(\omega)$ *is measurable.*

(3) *If $g(\omega) \neq 0$ for all $\omega \in \Omega$, then $\left(\frac{f}{g}\right)(\omega) := \frac{f(\omega)}{g(\omega)}$ is measurable.*

(4) $|f|$ *is measurable.*

Proof. The proof of (1), (2), and (4) follows directly from Corollary 3.1.5: In (1) we take $\Phi(x, y) := \alpha x + \beta y$, in (2) $\Phi(x, y) := xy$, and in (4) $\Phi(x) := |x|$. The proof of (3) is similarly (Exercise 1). □

3.2 Measurable maps

Now we extend the notion of random variables to the notion of measurable maps by using the condition in Proposition 3.1.3(2). This extension enables us to consider push forward measures and, at the same time, simplifies various proofs in the sequel.

Definition 3.2.1 (MEASURABLE MAP)**.** Let $(\Omega, \mathcal{F})$ and (M, Σ) be measurable spaces. A map $f : \Omega \to M$ is called **$(\mathcal{F}, \Sigma)$-measurable**, provided that

$$f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{F} \quad \text{for all} \quad B \in \Sigma.$$

The connection to the previously introduced random variables is given by the following equivalence:

Proposition 3.2.2. *Let $(\Omega, \mathcal{F})$ be a measurable space and $f : \Omega \to \mathbb{R}$. Then the following assertions are equivalent:*

(1) *The function f is measurable.*

(2) *The map f is $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-measurable.*

For the proof we need the lemma below.

Lemma 3.2.3. *Let* $(\Omega, \mathcal{F})$ *and* (M, Σ) *be measurable spaces and let* $f : \Omega \to M$*. Assume that* $\mathcal{G} \subseteq \Sigma$ *is a system of subsets such that* $\sigma(\mathcal{G}) = \Sigma$*. If*

$$f^{-1}(B) \in \mathcal{F} \quad \text{for all} \quad B \in \mathcal{G},$$

then

$$f^{-1}(B) \in \mathcal{F} \quad \text{for all} \quad B \in \Sigma.$$

Proof. Define

$$\mathcal{A} := \left\{ B \subseteq M : f^{-1}(B) \in \mathcal{F} \right\}.$$

By assumption, $\mathcal{G} \subseteq \mathcal{A}$. We show that $\mathcal{A}$ is a σ-algebra. Because $f^{-1}(M) = \Omega \in \mathcal{F}$, we have that $M \in \mathcal{A}$. If $B \in \mathcal{A}$, then

$$\begin{aligned} f^{-1}(B^c) &= \{\omega : f(\omega) \in B^c\} \\ &= \{\omega : f(\omega) \notin B\} \\ &= \Omega \backslash \{\omega : f(\omega) \in B\} \\ &= f^{-1}(B)^c \in \mathcal{F}. \end{aligned}$$

(3) If $B_1, B_2, \cdots \in \mathcal{A}$, then

$$f^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right) = \bigcup_{i=1}^{\infty} f^{-1}(B_i) \quad \in \mathcal{F}.$$

By definition of $\Sigma = \sigma(\mathcal{G})$ this implies that $\Sigma \subseteq \mathcal{A}$, which implies our lemma. □

Proof of Proposition 3.2.2. (2) $\Longrightarrow$ (1) follows from $(a, b) \in \mathcal{B}(\mathbb{R})$ for $a < b$ which implies that $f^{-1}((a, b)) \in \mathcal{F}$.

(1) $\Longrightarrow$ (2) is a consequence of Lemma 3.2.3 since $\mathcal{B}(\mathbb{R}) = \sigma((a, b) : -\infty < a < b < \infty)$. □

Directly from the definition of measurable maps one obtains:

Proposition 3.2.4. *Let* $(\Omega_1, \mathcal{F}_1)$, $(\Omega_2, \mathcal{F}_2)$, $(\Omega_3, \mathcal{F}_3)$ *be measurable spaces. Assume that* $f : \Omega_1 \to \Omega_2$ *is* $(\mathcal{F}_1, \mathcal{F}_2)$*-measurable and that* $g : \Omega_2 \to \Omega_3$ *is* $(\mathcal{F}_2, \mathcal{F}_3)$*-measurable. Then* $g \circ f : \Omega_1 \to \Omega_3$ *with*

$$(g \circ f)(\omega_1) := g(f(\omega_1))$$

is $(\mathcal{F}_1, \mathcal{F}_3)$*-measurable.*

Proof. For $A_3 \in \mathcal{F}_3$ we get $A_2 := g^{-1}(A_3) \in \mathcal{F}_2$ and $(g \circ f)^{-1}(A_3) = f^{-1}(A_2) \in \mathcal{F}_1$. □

The next proposition shows how useful the general concept of measurable maps is. Assume two metric spaces, (M_1, d_1) and (M_2, d_2), and a map $f : M_1 \to M_2$. Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be the open sets of M_1 and M_2, respectively. One knows that $f : M_1 \to M_2$ is continuous if and only if $f^{-1}(G_2)$ is open for all open $G_2 \subseteq M_2$ (one can take this as definition or equivalent condition, see []). This fits very well with our approach of measurability:

Proposition 3.2.5. *Let* (M_1, d_1) *and* (M_2, d_2) *be metric spaces with the systems of open sets* $\mathcal{G}_1$ *and* $\mathcal{G}_2$*, respectively. Define the* σ*-algebras* $\mathcal{F}_i := \sigma(\mathcal{G}_i)$ *on* M_i*. Then any continuous map* $f : M_1 \to M_2$ *is measurable as map between* $(M_1, \mathcal{F}_1)$ *and* $(M_2, \mathcal{F}_2)$*. In particular, if* $f : \mathbb{R} \to \mathbb{R}$ *is continuous, then* f *is* $(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}))$*-measurable.*

Proof. Since f is continuous we know that $f^{-1}(G_2) \in \mathcal{G}_1 \subseteq \mathcal{F}_1$ for all $G_2 \in \mathcal{G}_2$ and we can apply Lemma 3.2.3. □

3.3 Summary

Let $(\Omega, \mathcal{F})$ be a measurable space and $f : \Omega \to \mathbb{R}$ be a function. Then the following relations hold true:

$f^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{G}$
where $\mathcal{G}$ is one of the systems given in Proposition 2.2.3 or any other system such that $\sigma(\mathcal{G}) = \mathcal{B}(\mathbb{R})$.

$\Updownarrow$ Lemma 3.2.3

$$\boxed{f \text{ is } (\mathcal{F}, \mathcal{B}(\mathbb{R}))\text{-measurable, i.e. } f^{-1}(A) \in \mathcal{F} \text{ for all } A \in \mathcal{B}(\mathbb{R})}$$

$$\Updownarrow \text{ Proposition 3.2.2}$$

f is measurable i.e.
there exist simple functions $(f_n)_{n=1}^\infty$ i.e.
$f_n = \sum_{k=1}^{N_n} a_k^n \mathbb{I}_{A_k^n}$
with $a_k^n \in \mathbb{R}$ and $A_k^n \in \mathcal{F}$ such that
$f_n(\omega) \to f(\omega)$ for all $\omega \in \Omega$ as $n \to \infty$.

$$\Updownarrow \text{ Proposition 3.1.3}$$

$$\boxed{f^{-1}((a,b)) \in \mathcal{F} \text{ for all } -\infty < a < b < \infty}$$

3.4 Image measures and distribution functions

The notion of an image measure or push forward measure is central in probability. Our first proposition is the basis for its definition:

Proposition 3.4.1 (Image measures)**.** *Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces, $f : \Omega_1 \to \Omega_2$ be a measurable map, and μ_1 be a measure on $\mathcal{F}_1$. Define*

$$\mu_2(B_2) := \mu_1(f^{-1}(B_2)) \text{ for } B_2 \in \mathcal{F}_2.$$

Then $(\Omega_2, \mathcal{F}_2, \mu_2)$ is a measure space with $\mu_1(\Omega_1) = \mu_2(\Omega_2)$.

Proof. First we observe that $\mu_2(\emptyset) = \mu_1(f^{-1}(\emptyset)) = \mu_1(\emptyset) = 0$. Moreover, let $A_1, A_2, \ldots \in \mathcal{F}_2$ with $A_i \cap A_j = \emptyset$ for $i \neq j$. Then $f^{-1}(A_i) \in \mathcal{F}_1$ with $f^{-1}(A_i) \cap f^{-1}(A_j) = \emptyset$ for $i \neq j$ and

$$\mu_2\left(\bigcup_{i=1}^\infty A_i\right) = \mu_1\left(f^{-1}\left(\bigcup_{i=1}^\infty A_i\right)\right) = \mu_1\left(\bigcup_{i=1}^\infty f^{-1}(A_i)\right)$$

$$= \sum_{i=1}^{\infty} \mu_1\left(f^{-1}(A_i)\right) = \sum_{i=1}^{\infty} \mu_2(A_i).$$

□

Remark 3.4.2. *In Proposition 3.4.1 one has that μ_2 is a finite measure (probability measure) whenever μ_1 is a finite measure (probability measure). However, there are examples that μ_1 is σ-finite, but μ_2 is not.*

Definition 3.4.3. The measure μ_2 is called **image measure** or **push forward measure** of μ_1. In case μ_1 is a probability measure, the measure μ_2 is also called **law** or **distribution** of the random variable f.

Now we show that the law of a random variable is completely characterized by its distribution function.

Definition 3.4.4 (DISTRIBUTION-FUNCTION)**.** (1) Let μ be a finite measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then

$$F_\mu(x) := \mu((-\infty, x])$$

is called **distribution function of** μ.

(2) Given a random variable $f : \Omega \to \mathbb{R}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the function

$$F_f(x) := \mathbb{P}(\{\omega \in \Omega : f(\omega) \leqslant x\})$$

is called **distribution function of** f.

If μ is the law of a random variable $f : \Omega \to \mathbb{R}$, then it follows from the definition that

$$F_f(x) = F_\mu(x) \quad \text{for all} \quad x \in \mathbb{R}.$$

Proposition 3.4.5 (PROPERTIES OF DISTRIBUTION-FUNCTIONS)**.** *The distribution-function $F_\mu : \mathbb{R} \to [0, \infty)$ for a finite measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ has the following properties:*

(1) *F_μ is right-continuous.*

(2) *F_μ non-decreasing.*

(3) *One has* $\lim_{x\to-\infty} F_\mu(x) = 0$ *and* $\lim_{x\to\infty} F_\mu(x) = \mu(\mathbb{R})$.

Proof. (1) F_μ is non-decreasing as for $x_1 < x_2$ one has that $(-\infty, x_1] \subseteq (-\infty, x_2]$ and therefore

$$F_\mu(x_1) = \mu((-\infty, x_1]) \leqslant \mu((-\infty, x_2]) = F_\mu(x_2).$$

(2) F_f is right-continuous: let $x \in \mathbb{R}$ and $x_n \downarrow x$. Then

$$F_\mu(x) = \mu((-\infty, x]) = \mu\left(\bigcap_{n=1}^{\infty}(-\infty, x_n]\right) = \lim_n \mu\left((-\infty, x_n]\right) = \lim_n F_\mu(x_n).$$

(iii) This item is subject to Exercise 2. □

Proposition 3.4.6. *Assume finite measures μ_1 and μ_2 on $\mathcal{B}(\mathbb{R})$ and F_1 and F_2 are the corresponding distribution functions. Then the following assertions are equivalent:*

(1) $\mu_1 = \mu_2$.

(2) $F_1(x) = F_2(x)$ *for all* $x \in \mathbb{R}$.

Proof. (1) $\Longrightarrow$ (2) is obvious, so we consider (2) $\Longrightarrow$ (1): For sets of type

$$A := (a, b]$$

one can show that

$$\mu_1(A) = F_1(b) - F_1(a) = F_2(b) - F_2(a) = \mu_2(A).$$

Now one can apply Proposition 2.2.8. □

We close this section with the important converse of Proposition 3.4.5:

Proposition 3.4.7. *Assume a function* $F : \mathbb{R} \to [0,1]$ *that is a right-continuous, non-decreasing, and such that*

$$\lim_{x \to -\infty} F(x) = 0 \quad \textit{and} \quad \lim_{x \to \infty} F(x) = m < \infty.$$

Then there is exists a unique measure μ *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *such that*

$$\mu((-\infty, x]) = F(x).$$

For this measure we have that $\mu(\mathbb{R}) = m$.

Proof. The uniqueness follows from Proposition 3.4.6. Regarding the existence, let $\mathcal{A}$ be the algebra defined in Section 2.2.4, i.e. the system of subsets $A \subseteq \mathbb{R}$ such that A can be written as

$$(a_1, b_1] \cup (a_2, b_2] \cup \cdots \cup (a_n, b_n] \quad \text{or} \quad (a_1, b_1] \cup (a_2, b_2] \cup \cdots \cup (a_n, \infty)$$

where $-\infty \leqslant a_1 \leqslant b_1 \leqslant \cdots \leqslant a_n \leqslant b_n < \infty$. For such a set A we let

$$\mu_0(A) := \sum_{i=1}^{n} (F(b_i) - F(a_i))$$

and verify as in Proposition 2.2.9 that μ_0 satisfies the assumptions of Carathéodory's extension theorem Proposition 2.2.5. □

3.5 Exercises

Ex 1: Verify Proposition 3.1.6 (3).

Ex 2: Verify Proposition 3.4.5 (3).

Chapter 4

Independence

In real live *independence* is an intuitive notion, which might have different meanings depending on the context. Even in mathematics *independence* is used in different meanings. In vector spaces vectors might be *linearly* independent, in mathematical logic *sentences* might be independent. Here in probability we speak about *stochastic independence*. In Definition 1.4.5 we did already introduce the (stochastic) *independence of events*, which has to be understood as *one possibility* in order to transport the intuitive notion of independence from real live to probability. Now we extend this concept to the (stochastic) *independence of random variables*.

4.1 The basic concept

Let us first start with the notion of a family of independent random variables.

Definition 4.1.1 (INDEPENDENCE OF A FINITE FAMILY OF RANDOM VARIABLES)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f_i : \Omega \to \mathbb{R}$, $i = 1, \ldots, n$, random variables. The random variables $f_1, \ldots, f_n$ are called **independent** provided that for all $B_1, ..., B_n \in \mathcal{B}(\mathbb{R})$ one has that

$$\mathbb{P}(f_1 \in B_1, ..., f_n \in B_n) = \mathbb{P}(f_1 \in B_1) \cdots \mathbb{P}(f_n \in B_n).$$

In case, we have an index set I, which is not necessarily finite, then the definition is changed into:

Definition 4.1.2 (INDEPENDENCE OF A FAMILY OF RANDOM VARIABLES)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f_i : \Omega \to \mathbb{R}$, $i \in I$, be random variables where I is a non-empty index-set. The family $(f_i)_{i\in I}$ is called **independent** provided that for all distinct $i_1, ..., i_n \in I$, $n = 1, 2, ...,$ and all $B_1, ..., B_n \in \mathcal{B}(\mathbb{R})$ one has that

$$\mathbb{P}\left(f_{i_1} \in B_1, ..., f_{i_n} \in B_n\right) = \mathbb{P}\left(f_{i_1} \in B_1\right) \cdots \mathbb{P}\left(f_{i_n} \in B_n\right).$$

In case, we have a finite index set I, then both definitions coincide as in Definition 4.1.1 one can chose $B_i = \mathbb{R}$ which corresponds to leaving out the particular index i. Moroever, as an exercise we will see that Definition 4.1.2 directly extends Definition 1.4.5 because, for a family $(A_i)_{i\in I} \subseteq \mathcal{F}$ we have the equivalence of the following assertions:

(1) The events $(A_i)_{i\in I}$ are independent.

(2) The random variables $(\mathbb{1}_{A_i})_{i\in I}$ are independent.

The connection between the independence of random variables and of events is obvious and follows directly from the corresponding definitions:

Proposition 4.1.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f_i : \Omega \to \mathbb{R}$, $i \in I$, be random variables where I is a non-empty index-set. Then the following assertions are equivalent.*

(1) *The family $(f_i)_{i\in I}$ is independent.*

(2) *For all families $(B_i)_{i\in I}$ of* Borel *sets $B_i \in \mathcal{B}(\mathbb{R})$ one has that the events $(\{\omega \in \Omega : f_i(\omega) \in B_i\})_{i\in I}$ are independent.*

4.2 Kolmogorov's 0-1 law

Kolmogorov's 0-1 law is a first surprising, but important, example what independence may imply. Let us motivate this law by a simple to formulate problem: It is known that $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$ but that $\sum_{n=1}^{\infty} (-1)^n \frac{1}{n}$ does converge. What is going to happen if we take a random sequence of signs?

Definition 4.2.1. We denote by $\varepsilon_1, \varepsilon_2, \ldots : \Omega \to \mathbb{R}$ a sequence of independent random variables such that

$$\mathbb{P}\left(\varepsilon_n = -1\right) = \mathbb{P}\left(\varepsilon_n = 1\right) = \frac{1}{2}$$

and call this sequence a sequence of BERNOULLI [1] random variables.

Now we are interested in

$$\mathbb{P}\left(\sum_{n=1}^{\infty}\frac{\varepsilon_n}{n} \text{ converges}\right). \tag{4.1}$$

Remark 4.2.2. Let $\xi_1, \xi_2, ... : \Omega \to \mathbb{R}$ be random variables over $(\Omega, \mathcal{F}, \mathbb{P})$. Then

$$A := \left\{\omega \in \Omega : \sum_{n=1}^{\infty} \xi_n(\omega) \text{ converges}\right\} \in \mathcal{F}.$$

This is not difficult to verify because $\omega \in A$ if and only if

$$\bigcap_{N=1,2,...}\ \bigcup_{n_0=1,2,...}\ \bigcap_{m>n\geqslant n_0} \left\{\omega \in \Omega : \left|\sum_{k=n+1}^{m} \xi_k(\omega)\right| < \frac{1}{N}\right\}.$$

What is a typical property of the set $A = \left\{\omega \in \Omega : \sum_{n=1}^{\infty} \xi_n(\omega) \text{ converges}\right\}$? The condition does not depend on the first realizations $\xi_1(\omega), ..., \xi_N(\omega)$ since the convergence of $\sum_{n=1}^{\infty} \xi_n(\omega)$ and $\sum_{n=N+1}^{\infty} \xi_n(\omega)$ are equivalent so that

$$A = \left\{\omega \in \Omega : \sum_{n=N+1}^{\infty} \xi_n(\omega) \text{ converges}\right\}.$$

We shall formulate this in a more abstract way. For this we need

Definition 4.2.3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\xi_\alpha : \Omega \to \mathbb{R}$ be a family of random variables. Then $\sigma(\xi_\alpha : \alpha \in I)$ is the smallest σ-algebra which contains all sets of form

$$\{\omega \in \Omega : \xi_\alpha(\omega) \in B\} \quad \text{where } \alpha \in I \text{ and } B \in \mathcal{B}(\mathbb{R}).$$

Remark 4.2.4. The σ-algebra $\sigma(\xi_\alpha : \alpha \in I)$ is the smallest σ-algebra on Ω such that all random variables $\xi_\alpha : \Omega \to \mathbb{R}$ are measurable.

[1] Jacob Bernoulli, 27/12/1654 (Basel, Switzerland)- 16/08/1705 (Basel, Switzerland), Swiss mathematician.

Example 4.2.5. For random variables $\xi_1, \xi_2, ... : \Omega \to \mathbb{R}$ and

$$A = \left\{ \omega \in \Omega : \sum_{n=1}^{\infty} \xi_n(\omega) \text{ converges} \right\}$$

one has that $A \in \bigcap_{N=1}^{\infty} \sigma(\xi_N, \xi_{N+1}, \xi_{N+2}, ...)$.

The above example leads straight to the Zero-One law of Kolmogorov [2]:

Proposition 4.2.6 (Zero-One law of Kolmogorov)**.** *Assume independent random variables* $\xi_1, \xi_2, ... : \Omega \to \mathbb{R}$ *on some probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and let*

$$\mathcal{F}_N^{\infty} := \sigma(\xi_N, \xi_{N+1}, ...) \quad \textit{and} \quad \mathcal{F}^{\infty} := \bigcap_{N=1}^{\infty} \mathcal{F}_N^{\infty}.$$

Then $\mathbb{P}(A) \in \{0, 1\}$ *for all* $A \in \mathcal{F}^{\infty}$.

For the proof the following lemma is needed:

Lemma 4.2.7. *Let* $\mathcal{A} \subseteq \mathcal{F}$ *be an algebra and let* $\sigma(\mathcal{A}) = \mathcal{F}$*. Then, for all* $\varepsilon > 0$ *and* $B \in \mathcal{F}$*, there is an* $A \in \mathcal{A}$ *such that*

$$\mathbb{P}(A \Delta B) < \varepsilon.$$

Proof. We define Σ to be the system of all $B \in \mathcal{F}$ such that for all $\varepsilon > 0$ there exists an $A_\varepsilon \in \mathcal{A}$ with $\mathbb{P}(A_\varepsilon \Delta B) < \varepsilon$. By definition, $\mathcal{A} \subseteq \Sigma$. We show that Σ is a σ-algebra which immediately implies that $\Sigma = \mathcal{F}$. In fact, $\emptyset \in \mathcal{A} \subseteq \Sigma$. Moreover, for $B \in \Sigma$ and $\mathbb{P}(A_\varepsilon \Delta B) < \varepsilon$, we have

$$(A_\varepsilon)^c \Delta B^c = A_\varepsilon \Delta B \quad \text{and} \quad (A_\varepsilon)^c \in \mathcal{A}$$

so that $B^c \in \Sigma$. Finally, let $B_1, B_2, \ldots \in \Sigma$. Given $\varepsilon > 0$, we first find an $N \geqslant 1$ such that

$$\mathbb{P}\left(\left(\bigcup_{i=1}^{\infty} B_i \right) \setminus \left(\bigcup_{i=1}^{N} B_i \right) \right) < \frac{\varepsilon}{2}.$$

Moreover, we find $A_i \in \mathcal{A}$ with

$$\mathbb{P}(B_i \Delta A_i) < \frac{\varepsilon}{2^{i+1}}$$

[2]Andrey Nikolaevich Kolmogorov 25/04/1903 (Tambov, Russia) - 20/10/1987 (Moscow, Russia), one of the founders of modern probability theory, Wolf prize 1980.

for $i = 1, \ldots, N$. Then $A := \bigcup_{i=1}^N A_i \in \mathcal{A}$ and

$$\mathbb{P}\left(A\Delta\left(\bigcup_{i=1}^{\infty} B_i\right)\right) < \varepsilon.$$

□

Proof of Proposition 4.2.6. The idea of the proof is to show that $\mathbb{P}(A) = \mathbb{P}(A)^2$. Define the algebra

$$\mathcal{A} := \bigcup_{n=1}^{\infty} \sigma(\xi_1, ..., \xi_n).$$

We have that $\mathcal{F}^\infty \subseteq \sigma(\mathcal{A})$. Hence Lemma 4.2.7 implies that for $A \in \mathcal{F}^\infty$ there are $A_n \in \sigma(\xi_1, ..., \xi_{N_n})$ such that

$$\mathbb{P}(A\Delta A_n) \to_{n\to\infty} 0.$$

We get also that

$$\mathbb{P}(A_n \cap A) \to_{n\to\infty} \mathbb{P}(A) \quad \text{and} \quad \mathbb{P}(A_n) \to_{n\to\infty} \mathbb{P}(A).$$

The first relation can be seen as follows: since

$$\mathbb{P}(A_n \cap A) + \mathbb{P}(A_n \Delta A) = \mathbb{P}(A_n \cup A) \geqslant \mathbb{P}(A)$$

we get that

$$\liminf_n \mathbb{P}(A_n \cap A) \geqslant \mathbb{P}(A) \geqslant \limsup \mathbb{P}(A_n \cap A).$$

The second relation can be also checked easily. But now we get, by independence, that

$$\mathbb{P}(A) = \lim_n \mathbb{P}(A \cap A_n) = \lim_n \mathbb{P}(A)\mathbb{P}(A_n) = \mathbb{P}(A)^2$$

so that $\mathbb{P}(A) \in \{0, 1\}$. □

Corollary 4.2.8. *Let $\xi_1, \xi_2, ... : \Omega \to \mathbb{R}$ be independent random variables over $(\Omega, \mathcal{F}, \mathbb{P})$. Then*

$$\mathbb{P}\left(\sum_{n=1}^{\infty} \xi_n \textit{ converges}\right) \in \{0, 1\}.$$

Returning to our initial problem in (4.1) we get that

$$\mathbb{P}\left(\sum_{n=1}^{\infty}\frac{\varepsilon_n}{n} \text{ converges}\right) \in \{0,1\}$$

without any explicit computation. Later in Example 8.1.5 we will solve the problem whether we get zero or one as probability.

4.3 Product spaces

Products of measure spaces, in particular probability spaces, might look slightly technical at first glance. This is mainly due to the notation, otherwise the concept is natural. Products of measure spaces are used in various places, we will use them to show the existence of independent random variables, and they are the basis of FUBINI's theorem, one of the main theorems in integration theory.

In this section we construct in Proposition 4.3.1 below products of probability spaces by an application of Carathéodory's extension theorem. In the following we let $I := \{1,\ldots,d\}$ or $I := \{1,2,\ldots\}$, i.e. we consider finite or countable product spaces at the same time. Our main statement can be formulated in a rather short way:

Proposition 4.3.1 (PRODUCTS OF PROBABILITY SPACES). *Assume probability spaces $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$, $i \in I$. Let*

$$\Omega := \bigtimes_{i\in I} \Omega_i = \big\{(\omega_i)_{i\in I} : \omega_i \in \Omega_i \ \text{for}\ i \in I\big\},$$

and define the σ-algebra $\mathcal{F} := \otimes_{i\in I}\mathcal{F}_i$ on Ω to be generated by all sets

$$\bigtimes_{i\in I} A_i := \{(\omega_i)_{i\in I} \in \Omega : \omega_i \in A_i\} \quad \text{where} \quad A_i \in \mathcal{F}_i.$$

Then there exists a unique probability measure $\mathbb{P} = \otimes_{i\in I}\mathbb{P}_i$ on $\mathcal{F}$ such that

$$\mathbb{P}\left(\bigtimes_{i\in I} A_i\right) = \prod_{i\in I}\mathbb{P}_i(A_i).$$

Definition 4.3.2 (PRODUCT OF PROBABILITY SPACES). The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called product space and denoted by $\otimes_{i\in I}(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$.

Proof of Proposition 4.3.1. (a) Existence of $\mathbb{P}$: We define the algebra $\mathcal{A}$ generated by all finite unions of 'cuboids'

$$A = \bigtimes_{i \in I} A_i$$

with $A_i \in \mathcal{F}_i$ and where for only finitely many of the A_i it holds that $A_i \neq \Omega_i$ (in the case $I = \{1, \ldots, d\}$ this is always satisfied). The algebra $\mathcal{A}$ can be described by all finite unions of such cuboids. Let

$$A^{(l)} = \bigtimes_{i \in I} A_i^{(l)} \in \mathcal{A}$$

for $l = 1, \ldots, n$ we define $\mathbb{P}^0 : \mathcal{A} \to [0,1]$ by

$$\mathbb{P}^0 \left(\bigcup_{l=1}^{L} (\bigtimes_{i \in I} A_i^{(l)}) \right) := \sum_{l=1}^{L} \left(\prod_{i \in I} \mathbb{P}_i(A_i^{(l)}) \right)$$

By construction, the map $\mathbb{P}^0$ is (i) well-defined, (ii) monotone, and (iii) finitely additive, which we leave as an exercise. Here we prove the main part, the σ-additivity of $\mathbb{P}^0$ on $\mathcal{A}$, i.e. that $A^{(1)}, A^{(2)}, \ldots \in \mathcal{A}$ with $A^{(k)} \cap A^{(l)} = \varnothing$ for $k \neq l$ implies

$$\mathbb{P}^0 \left(\bigcup_{l=1}^{\infty} A^{(l)} \right) = \sum_{l=1}^{\infty} \mathbb{P}^0 \left(A^{(l)} \right) \quad \textbf{whenever} \quad \bigcup_{l=1}^{\infty} A^{(l)} \in \mathcal{A}.$$

One inequality is immediate by monotonicity and finite-additivity, because

$$\mathbb{P}^0 \left(\bigcup_{l=1}^{\infty} A^{(l)} \right) \geqslant \mathbb{P}^0 \left(\bigcup_{l=1}^{L} A^{(l)} \right) = \sum_{l=1}^{L} \mathbb{P}^0 \left(A^{(l)} \right),$$

so that, by $L \to \infty$,

$$\mathbb{P}^0 \left(\bigcup_{l=1}^{\infty} A^{(l)} \right) \geqslant \sum_{l=1}^{\infty} \mathbb{P}^0 \left(A^{(l)} \right).$$

To prove the converse inequality we assume an $\varepsilon > 0$ with

$$\varepsilon + \sum_{l=1}^{\infty} \mathbb{P}^0 \left(A^{(l)} \right) = \varepsilon + \lim_{L} \mathbb{P}^0 \left(\bigcup_{l=1}^{L} A^{(l)} \right) \leqslant \mathbb{P}^0 \left(\bigcup_{l=1}^{\infty} A^{(l)} \right)$$

which implies

$$\varepsilon \leqslant \lim_L \mathbb{P}^0 \left(\bigcup_{l=L+1}^{\infty} A^{(l)} \right).$$

If $B^{(L)} := \bigcup_{l=L+1}^{\infty} A^{(l)}$, then $B^{(1)} \supseteq B^{(2)} \supseteq \cdots$, and since the sets $A^{(l)}, A^{(l)}, \ldots$ are disjoint, it holds $\bigcap_{L=1}^{\infty} B^{(L)} = \varnothing$. We will use the assumption

$$\lim_L \mathbb{P}^0(B^{(L)}) \geqslant \varepsilon > 0$$

to construct a sequence $(\omega_i)_{i \in I} \in B^{(L)}$ for all $L = 1, 2, \ldots$ which contradicts $\bigcap_{L=1}^{\infty} B^{(L)} = \varnothing$ - and proves therefore our remaining inequality.

We define $\Omega^{(1)} := \times_{i \in I \setminus \{1\}} \Omega_i$, an algebra $\mathcal{A}^{(1)}$ and a map $\mathbb{P}^1$ just as we have done with Ω, $\mathcal{A}$ and $\mathbb{P}^0$, but without the first component $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$. Moreover, we define for each $\omega_1^0 \in \Omega_1$ the sections

$$B^{(L)}(\omega_1^0) := \{(\omega_i)_{i \in I \setminus \{1\}} : (\omega_1^0, \omega_2, \omega_3, \ldots) \in B^{(L)}\} \in \mathcal{A}^{(1)}$$

and the sets

$$C_1^{(L)} := \left\{ \omega_1 \in \Omega_1 : \mathbb{P}^1(B^{(L)}(\omega_1)) > \frac{\varepsilon}{2} \right\}.$$

Because

$$\varepsilon \leqslant \mathbb{P}^0(B^{(L)}) \leqslant \mathbb{P}_1(C_1^{(L)}) + \mathbb{P}_1((C_1^{(L)})^c) \frac{\varepsilon}{2} \leqslant \mathbb{P}_1(C_1^{(L)}) + \frac{\varepsilon}{2}$$

we have that

$$\mathbb{P}_1(C_1^{(L)}) \geqslant \frac{\varepsilon}{2} \quad \text{and} \quad C_1^{(1)} \supseteq C_1^{(2)} \supseteq \cdots .$$

By continuity from above, we conclude that

$$\lim_{L \to \infty} \mathbb{P}_1(C_1^{(L)}) = \mathbb{P}_1 \left(\bigcap_{L=1}^{\infty} C_1^{(L)} \right) \geqslant \frac{\varepsilon}{2}.$$

This implies that $\bigcap_{L=1}^{\infty} C_1^{(L)}$ contains at least one element, which we denote by ω_1. We proceed with this construction using now $\Omega^{(1)}, \mathcal{A}^{(1)}, \mathbb{P}^1$ and the system $(B^{(L)}(\omega_1))_{L=1}^{\infty}$. Continuing in this way we obtain a sequence $(\omega_i)_{i \in I} \in B^{(L)}$ for all $L = 1, 2, \ldots$ which contradicts the assumption that an $\varepsilon > 0$ exists with $\varepsilon + \sum_{l=1}^{\infty} \mathbb{P}^0 \left(A^{(l)} \right) \leqslant \mathbb{P}^0 \left(\bigcup_{l=1}^{\infty} A^{(l)} \right)$.

Now we can apply Proposition 2.2.5 to obtain an extension of $\mathbb{P}^0 : \mathcal{A} \to [0, 1]$ to $\mathbb{P} : \mathcal{F} = \sigma(\mathcal{A}) \to [0, 1]$.

(b) Uniqueness of $\mathbb{P}$: The uniqueness follows from Proposition 2.2.8 if we take the π-system of all $A = A_1 \times A_2 \times A_3 \times \cdots$ with $A_i \in \mathcal{F}_i$ where only finitely many of the A_i differ from Ω_i. □

Now we deduce a similar statement for σ-finite measure spaces, which we exploit to derive the existence of the d-dimensional LEBESGUE-measure.

Corollary 4.3.3 (PRODUCTS OF σ-FINITE MEASURE SPACES)**.** *Assume σ-finite measure spaces $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i = 1, \ldots, d$. Define $\Omega := \times_{i=1}^d \Omega_i$ and $\mathcal{F} := \otimes_{i=1}^d \mathcal{F}_i$ as in Proposition 4.3.1. Then there exists a unique measure $\mu = \otimes_{i=1}^d \mu_i$ on $\mathcal{F}$ such that*

$$\mu\left(\bigtimes_{i=1}^d A_i\right) = \prod_{i=1}^d \mu_i(A_i) \tag{4.2}$$

for $A_i \in \mathcal{F}_i$ with $\mu_i(A_i) < \infty$. The measure μ is σ-finite as well.

Proof. (a) Existence of μ: By the σ-finiteness we find disjoint partitions $\Omega_i = \bigcup_{j \in J_i} \Omega_{i,j}$ with $0 < p_{i,j} := \mu_i(\Omega_{i,j}) < \infty$. By Proposition 4.3.1 we define, for $j_1 \in J_1, \ldots, j_n \in J_n$, the products

$$\left(\times_{i=1}^n \Omega_{i,j_i}, \otimes_{i=1}^n \mathcal{F}_i|_{\Omega_{i,j_i}}, \mathbb{P}_{j_1,\ldots,j_n}\right) = \otimes_{i=1}^n \left(\Omega_{i,j_i}, \mathcal{F}_i|_{\Omega_{i,j_i}}, \frac{\mu_i|_{\Omega_{i,j_i}}}{p_{i,j_i}}\right).$$

Then we define

$$\mu(B) := \sum_{j_1 \in J_1, \ldots, j_n \in J_n} \left[\prod_{i=1}^n p_{i,j_i}\right] \mathbb{P}_{j_1,\ldots,j_n}\left(B \cap (\times_{i=1}^n \Omega_{i,j_i})\right).$$

It is obvious that μ is a σ-finite measure. Moreover,

$$\begin{aligned}
& \mu(A_1 \times \cdots \times A_d) \\
= & \sum_{j_1 \in J_1, \ldots, j_n \in J_n} \left[\prod_{i=1}^n p_{i,j_i}\right] \mathbb{P}_{j_1,\ldots,j_n}\left((A_1 \times \cdots \times A_d) \cap (\times_{i=1}^n \Omega_{i,j_i})\right) \\
= & \sum_{j_1 \in J_1, \ldots, j_n \in J_n} \mu_1(A_1 \cap \Omega_{1,j_1}) \cdots \mu_n(A_n \cap \Omega_{n,j_n}) \\
= & \mu_1(A_1) \cdots \mu_n(A_n).
\end{aligned}$$

(b) Uniqueness of μ: Assuming two different measures μ and ν that satisfy (4.2), there has to be one cuboid $\Omega_{1,j_1} \times \cdots \times \Omega_{n,j_n}$ where the restrictions of μ and ν do not coincide. But then we can use Proposition 4.3.1 to see that this is not possible under our assumptions. □

Directly from Proposition 4.3.3 we obtain

Corollary 4.3.4. *There is a unique measure* λ^d *on* $\otimes_{i=1}^d \mathcal{B}(\mathbb{R})$ *such that*

$$\lambda^d(B_1 \times \ldots \times B_d) = \prod_{i=1}^d \lambda(B_i)$$

for all $B_1, \ldots, B_d \in \mathcal{B}(\mathbb{R})$ *of finite measure. The measure is* σ*-finite and it holds that* $\lambda^d = \otimes_{i=1}^d \lambda$*, where* λ *is the one-dimensional Lebesgue measure.*

Definition 4.3.5. We let $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d) := \otimes_{i=1}^d (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, the measure λ^d is called d-dimensional Lebesgue measure.

Finally it is convenient to know that the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$ can be obtained in various ways:

Proposition 4.3.6. *For a* σ*-algebra* $\mathcal{F}$ *on* $\mathbb{R}^d$ *the following assertions are equivalent:*

(1) $\mathcal{F} = \mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R})$.

(2) $\mathcal{F}$ *is the smallest* σ*-algebra on* $\mathbb{R}^d$ *that contains all cuboids* $(a_1, b_1] \times \ldots \times (a_d, b_d]$ *with* $-\infty < a_i < b_i < \infty$.

(3) $\mathcal{F}$ *is the smallest* σ*-algebra on* $\mathbb{R}^d$ *that contains all open sets from* $\mathbb{R}^d$.

Proof. Let us denote by $\mathcal{F}_k$ the σ-algebra considered in item (k) above. Then $\mathcal{F}_3 \subseteq \mathcal{F}_2$ as any open set in $\mathbb{R}^d$ can be written as a countable union of the cuboids considered in (2). We also have $\mathcal{F}_2 \subseteq \mathcal{F}_1$ as $(a_i, b_i] \in \mathcal{B}(\mathbb{R})$. So it remains to check $\mathcal{F}_1 \subseteq \mathcal{F}_3$. The σ-algebra $\mathcal{F}_3$ contains all sets of type $\mathbb{R} \times \cdots \mathbb{R} \times (a_i, b_i) \times \cdots \times \mathbb{R}$ for $-\infty < a_i < b_i < \infty$, so that $\mathcal{F}_3$ contains all sets of type $\mathbb{R} \times \cdots \mathbb{R} \times B_i \times \cdots \times \mathbb{R}$ for a Borel set $B_i \in \mathcal{B}(\mathbb{R})$ (this has to be formally verified and is subject to an exercise). But then $\mathcal{F}_1 \subseteq \mathcal{F}_3$. □

Remark 4.3.7. In Corollary 4.3.3 the condition that the measure spaces $(\Omega_i, \mathcal{F}_i, \mu_i)$ are σ-finite is essential for the uniqueness of the product measure. The problem is to give products of type $0 \cdot \infty$ a meaning. Let us give an example for $d = 2$ where we do not have uniqueness if one $(\Omega_i, \mathcal{F}_i, \mu_i)$ is not σ-finite:

Let $(\Omega_1, \mathcal{F}_1, \mu_1) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, $\Omega_2 := \mathbb{R}$, $\mathcal{F}_2 := \sigma(\mathcal{B}((-\infty, 0)), [0, \infty))$, i.e. $[0, \infty) \in \mathcal{F}_2$ cannot be divided, and let

$$\mu_2(B_2) := \begin{cases} \lambda(B_2) & : B \subseteq (-\infty, 0) \\ +\infty & : \text{else} \end{cases}.$$

As product measure we can choose $\lambda \otimes \lambda$ as well as

$$\nu(B) := \begin{cases} (\lambda \otimes \lambda)(B) & : B \subseteq \mathbb{R} \times (-\infty, 0) \\ +\infty & : \text{else} \end{cases}.$$

The measures $\lambda \otimes \lambda$ and ν do not coincide because (for example)

$$(\lambda \otimes \lambda)(\{0\} \times \mathbb{R}) = 0 \quad \text{but} \quad \nu(\{0\} \times \mathbb{R}) = \infty.$$

4.4 Do independent random variables exist?

When dealing with families of independent random variables for the first time one might easily overlook the question, whether such families do exist. In this section we deduce the existence straight from Proposition 4.3.1:

Corollary 4.4.1 (Existence of independent random variables)**.** *Let $I = \{1, 2, ..., n\}$ or $I = \mathbb{N}$. Given probability measures μ_i on $\mathcal{B}(\mathbb{R})$, $i \in I$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and independent random variables $(f_i)_{i \in I}$, $f_i : \Omega \to \mathbb{R}$ such that the law of f_i is μ_i, that means*

$$\mathbb{P}(f_i \in B) = \mu_i(B) \quad \textit{for all} \quad B \in \mathcal{B}(\mathbb{R}).$$

Proof. Using Proposition 4.3.1 we set

$$(\Omega, \mathcal{F}, \mathbb{P}) := \otimes_{i \in I}(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_i)$$

and define the maps

$$f_i : \Omega \to \mathbb{R} \quad \text{with} \quad f_i(x) := x_i \quad \text{if} \quad x = (x_j)_{j \in I}.$$

The maps f_i are random variables as for $B \in \mathcal{B}(\mathbb{R})$ we have that

$$f_i^{-1}(B) = \bigtimes_{j \in I} A_j \in \mathcal{A} \subseteq \mathcal{F} \quad \text{with} \quad A_j := \begin{cases} B & j = i \\ \mathbb{R} & j \neq i \end{cases}.$$

Finally, for any $n \in I$ and $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}(f_1 \in B_1, \ldots, f_n \in B_n) = \mathbb{P}(B_1 \times \cdots \times B_n \times \mathbb{R} \times \mathbb{R} \times \cdots) = \prod_{i=1}^{n} \mu_i(B_i)$$

and, by fixing i and letting $B_j = \mathbb{R}$ for $j \neq i$, we have

$$\begin{aligned}
\mathbb{P}(f_i \in B_i) &= \mathbb{P}(f_1 \in \mathbb{R}, \ldots, f_{i-1} \in \mathbb{R}, f_i \in B_i, f_{i+1} \in \mathbb{R}, \ldots, f_n \in \mathbb{R}) \\
&= \cdots \\
&= \mu_1(\mathbb{R}) \cdots \mu_{i-1}(\mathbb{R}) \mu_i(B_i) \mu_{i+1}(\mathbb{R}) \cdots \mu_n(\mathbb{R}) \\
&= \mathbb{P}(f_i \in B_i)
\end{aligned}$$

which proves that the law of f_i is μ_i and, by the computation before, that

$$\mathbb{P}(f_1 \in B_1, \ldots, f_n \in B_n) = \prod_{i=1}^{n} \mathbb{P}(f_i \in B_i). \qquad \square$$

The next proposition shows that the Independence of random variables $f_1, \ldots, f_n$ is only a property of the distribution, or in other words the law, of the random vector

$$(f_1, \ldots, f_n) : \Omega \to \mathbb{R}^d.$$

In particular, this implies the following: If $f_1, \ldots, f_n : \Omega \to \mathbb{R}$ are independent and if $f_1', \ldots, f_n' : \Omega' \to \mathbb{R}$ are random variables such that the laws of $(f_1, \ldots, f_n)$ and $(f_1', \ldots, f_n')$ coincide, then $f_1', \ldots, f_n' : \Omega' \to \mathbb{R}$ are independent as well.

Proposition 4.4.2 (INDEPENDENCE AND PRODUCT OF LAWS)**.** *Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and that $f_1, \ldots, f_d : \Omega \to \mathbb{R}$ are random variables with laws $\mu_1, \ldots, \mu_d$, and distribution-functions $F_1, \ldots, F_d$. Then the following assertions are equivalent:*

(1) *$f_1, \ldots, f_d$ are independent.*

(2) *$\mathbb{P}((f_1, \ldots, f_d) \in B) = \otimes_{i=1}^{d} \mu_i(B)$ for all $B \in \mathcal{B}(\mathbb{R}^d)$.*

(3) *$\mathbb{P}(f_1 \leqslant x_1, \ldots, f_d \leqslant x_d) = F_1(x_1) \cdots F_d(x_d)$ for all $x_1, \ldots, x_d \in \mathbb{R}$.*

The proof is subject to an exercise. Sometimes we need to group independent random variables. In this respect the following proposition turns out to be useful. For the following we say that $g : \mathbb{R}^n \to \mathbb{R}$ is Borel-measurable (or a Borel function) provided that g is $(\mathcal{B}(\mathbb{R}^n), \mathcal{B}(\mathbb{R}))$-measurable.

Proposition 4.4.3. [Grouping of independent random variables] *Let $f_i : \Omega \to \mathbb{R}$, $i = 1, 2, 3, ...$ be independent random variables. Assume Borel functions $g_k : \mathbb{R}^{n_k} \to \mathbb{R}$ for $k = 1, 2, ...$ and $n_k \in \{1, 2, ...\}$. Then the random variables $g_1(f_1, ..., f_{n_1})$, $g_2(f_{n_1+1}, ..., f_{n_1+n_2})$, $g_3(f_{n_1+n_2+1}, ..., f_{n_1+n_2+n_3})$, ... are independent as well.*

This proposition is verified within an exercise.

4.5 * Hewitt-Savage's 0-1 law

In Definition 4.2.1 we introduced the Bernoulli random variables, now we extend this to the non-symmetric case.

Definition 4.5.1. *For $p \in (0, 1)$ we denote by $\varepsilon_1^{(p)}, \varepsilon_2^{(p)}, ... : \Omega \to \mathbb{R}$ a sequence of independent random variables such that*

$$\mathbb{P}\left(\varepsilon_n^{(p)} = -1\right) = p \quad \textit{and} \quad \mathbb{P}\left(\varepsilon_n^{(p)} = 1\right) = 1 - p.$$

Let $p \in (0, 1)$ and $f_n := \sum_{i=1}^n \varepsilon_i^{(p)}$ where $n = 1, 2, ...$ Consider the event

$$A := \{\omega : \# \{n : f_n(\omega) = 0\} = \infty\}.$$

In words, A is the event, that the path of the process $(f_n)_{n=1}^\infty$ reaches 0 infinitely many often. We would like to have a Zero-One law for this event. However, we are not able to apply Kolmogorov's Zero-One law (Proposition 4.2.6). But what is the typical property of A? We can rearrange *finitely* many elements of the sum $\sum_{i=1}^n \varepsilon_i^{(p)}$ and we will get the same event since from some N on, this rearrangement does not influence the sum anymore. To give a formal definition of this property we need

Definition 4.5.2. A map $\pi : \mathbb{N} \to \mathbb{N}$ is called *finite permutation* if

(1) the map π is a bijection,

(2) there is some $N \in \mathbb{N}$ such that $\pi(n) = n$ for all $n \geqslant N$.

Moreover, let $\mathcal{B}(\mathbb{R}^\mathbb{N})$ be the smallest σ-algebra on $\mathbb{R}^\mathbb{N}$ which contains all cylinder sets Z of form

$$Z := \{(\xi_1, \xi_2, ...) : a_n < \xi_n < b_n, n = 1, 2, ...\}$$

for some $-\infty < a_n < b_n < \infty$. Now the symmetry property, needed for the next Zero-One law, looks as follows:

Definition 4.5.3. Let $(\xi_n)_{n=1}^{\infty}$, $\xi_n : \Omega \to \mathbb{R}$, be a sequence of independent random variables over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ and

$$A := \{\omega \in \Omega : (\xi_1(\omega), \xi_2(\omega), ...) \in B\}.$$

The set A is called *symmetric* if for all finite permutations $\pi : \mathbb{N} \to \mathbb{N}$ one has that

$$A = \left\{\omega \in \Omega : (\xi_{\pi(1)}(\omega), \xi_{\pi(2)}(\omega), ...) \in B\right\}.$$

A typical set B which serves as *basis* for the set A is given by

Example 4.5.4. We let $B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ be the set of all sequences $(\xi_n)_{n=1}^{\infty}$, $\xi \in \{-1, 1\}$, such that

$$\#\{n : \xi_1 + \cdots + \xi_n = 0\} = \infty.$$

The next Zero-One law works for identically distributed random variables.

Proposition 4.5.5 (Zero-One law of HEWITT and SAVAGE). [3] *Assume a sequence of independent random variables* $\xi_1, \xi_2, ... : \Omega \to \mathbb{R}$ *over some probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *that have the* same *law (in other words that are identically distributed). If the event* $A \in \mathcal{F}$ *is symmetric, then* $\mathbb{P}(A) \in \{0, 1\}$.

Proof. (a) We are going to use the permutations

$$\pi_n(k) := \begin{cases} n+k & : \quad 1 \leqslant k \leqslant n \\ k-n & : \quad n+1 \leqslant k \leqslant 2n \\ k & : \quad k > 2n \end{cases}.$$

Now we approximate the set A. By Lemma 4.2.7 we find $B_n \in \mathcal{B}(\mathbb{R}^n)$ such that

$$\mathbb{P}(A_n \Delta A) \to_n 0 \quad \text{for} \quad A_n := \{\omega : (\xi_1(\omega), \ldots, \xi_n(\omega)) \in B_n\}.$$

[3] Leonard Jimmie Savage 20/11/1917 (Detroit, USA) - 1/11/1971 (New Haven, USA), American mathematician and statistician.

(b) Our goal is to show that

$$\begin{aligned} \mathbb{P}\left(A_n\right) &\to \mathbb{P}(A), \\ \mathbb{P}\left(\pi_n(A_n)\right) &\to \mathbb{P}(A), \\ \mathbb{P}\left(A_n \cap \pi_n(A_n)\right) &\to \mathbb{P}(A), \end{aligned}$$

as $n \to \infty$, where

$$\begin{aligned} \pi_n\left(A_n\right) &:= \left\{\omega : \left(\xi_{\pi_n(k)}(\omega)\right)_{k=1}^n \in B_n\right\} \\ &= \left\{\omega : \left(\xi_{n+1}(\omega), \ldots, \xi_{2n}(\omega)\right) \in B_n\right\}. \end{aligned}$$

(c) Assuming for a moment that (b) is proved we derive

$$\mathbb{P}\left(A_n \cap \pi_n(A_n)\right) = \mathbb{P}(A_n)\mathbb{P}(\pi_n(A_n))$$

since A_n is a condition on $\xi_1, ..., \xi_n$, $\pi_n(A_n)$ is a condition on $\xi_{n+1}, ..., \xi_{2n}$, and $\xi_1, ..., \xi_{2n}$ are independent. By $n \to \infty$ this equality turns into $\mathbb{P}(A) = \mathbb{P}(A)\mathbb{P}(A)$ so that $\mathbb{P}(A) \in \{0, 1\}$.

(d) Now we prove (b). The convergence

$$\mathbb{P}(A_n) \to \mathbb{P}(A)$$

is obvious since $\mathbb{P}(A_n \Delta A) \to 0$. This implies

$$\mathbb{P}(\pi_n(A_n)) \to \mathbb{P}(A)$$

since $\mathbb{P}(\pi_n(A_n)) = \mathbb{P}(A_n)$ which follows from the fact that

$$(\xi_{n+1}, ..., \xi_{2n}, \xi_1, ..., \xi_n) \quad \text{and} \quad (\xi_1, ..., \xi_n, \xi_{n+1}, ..., \xi_{2n})$$

have the same law in $\mathbb{R}^{2n}$ as a consequence that we have an identically distributed sequence of random variables. Finally

$$\begin{aligned} \mathbb{P}(A\Delta A_n) &= \mathbb{P}\left(\pi_n(A\Delta A_n)\right) \\ &= \mathbb{P}\left(\pi_n(A)\Delta\pi_n(A_n)\right) \\ &= \mathbb{P}\left(A\Delta\pi_n(A_n)\right) \end{aligned}$$

where the first equality follows because the random variables $(\xi_1, \xi_2, ...)$ have the same distribution and the last equality is a consequence of the symmetry of A. Hence

$$\mathbb{P}\left(A\Delta A_n\right) \to_n 0 \quad \text{and} \quad \mathbb{P}\left(A\Delta\pi_n(A_n)\right) \to_n 0$$

which implies

$$\mathbb{P}\left(A\Delta\left(A_n \cap \pi_n(A_n)\right)\right) \to_n 0 \quad \text{and} \quad \mathbb{P}\left(A_n \cap \pi_n(A_n)\right) \to \mathbb{P}(A).$$

□

As an application we obtain for the symmetric random walk:

Proposition 4.5.6. *Let $\varepsilon_1, \varepsilon_2, ... : \Omega \to \mathbb{R}$ be independent* Bernoulli *random variables and $f_n = \sum_{i=1}^n \varepsilon_i$, $n = 1, 2, ...$ Then*

$$\mathbb{P}(\#\{n : f_n = 0\} = \infty) = 1$$

Proof. Consider the sets

$$\begin{aligned} A^+ &:= \{\omega \in \Omega : \#\{n : f_n(\omega) = 0\} < \infty\} \cap \{\omega \in \Omega : \liminf f_n(\omega) > 0\}, \\ A &:= \{\omega \in \Omega : \#\{n : f_n(\omega) = 0\} = \infty\}, \\ A_- &:= \{\omega \in \Omega : \#\{n : f_n(\omega) = 0\} < \infty\} \cap \{\omega \in \Omega : \limsup f_n(\omega) < 0\}. \end{aligned}$$

Since the random walk is symmetric we have

$$\mathbb{P}(A_+) = \mathbb{P}(A_-).$$

Moreover, A_+, A, and A_- are symmetric, so that

$$\mathbb{P}(A_+), \mathbb{P}(A), \mathbb{P}(A_-) \in \{0, 1\}$$

by the Zero-One law of Hewitt-Savage. As the only solution to that we obtain

$$\mathbb{P}(A) = 1 \quad \text{and} \quad \mathbb{P}(A_+) = \mathbb{P}(A_-) = 0.$$

□

The non-symmetric random walk behaves differently as the symmetric one:

Proposition 4.5.7. *Let $p \neq 1/2$ and $f_n = \sum_{i=1}^n \varepsilon_i^{(p)}$, $n = 1, 2, ...$, where the random variables $\varepsilon_1^{(p)}, \varepsilon_2^{(p)}, ... : \Omega \to \mathbb{R}$ are given by Definition 4.5.1. Then*

$$\mathbb{P}\left(\#\{n : f_n = 0\} = \infty\right) = 0.$$

Proof. First we recall STIRLING [4]'s formula

$$n! = \sqrt{2\pi n} \quad \left(\frac{n}{e}\right)^n \quad e^{\frac{\theta}{12n}}$$

for $n = 1, 2, ...$ and some $\theta \in (0,1)$ depending on n. This gives

$$\binom{2n}{n} = \frac{(2n)!}{(n!)^2} = \frac{\sqrt{2\pi(2n)}\left(\frac{2n}{e}\right)^{2n} e^{\frac{\theta_1}{12\cdot 2n}}}{(\sqrt{2\pi n})^2 \left(\frac{n}{e}\right)^{2n} \left(e^{\frac{\theta_2}{12n}}\right)^2} \sim \frac{4^n}{\sqrt{\pi n}}.$$

Letting $B_n := \{f_{2n} = 0\}$ we obtain

$$\mathbb{P}(B_n) = \binom{2n}{n}(pq)^n \sim \frac{(4pq)^n}{\sqrt{\pi n}}$$

by STIRLING's formula. Since $p \neq q$ gives $4pq < 1$, we have

$$\sum_{n=1}^{\infty} \mathbb{P}(B_n) < \infty$$

so that the Lemma of BOREL-CANTELLI implies that

$$\mathbb{P}(\omega \in B_n \text{ infinitely often }) = 0.$$

□

Remark 4.5.8. Let us illustrate by some additional very simple examples the difference of the assumptions for the Zero-One law of KOLMOGOROV and the Zero-One law of HEWITT-SAVAGE. Assume a sequence of independent and identically distributed random variables $\xi_1, \xi_2, ... : \Omega \to \mathbb{R}$ over some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and that $\mathcal{F}^\infty = \bigcap_{n=1}^{\infty} \sigma(\xi_n, \xi_{n+1}, \ldots)$.

(1) Given $B \in \mathcal{B}(\mathbb{R})$ we have that $\{\limsup_n \xi_n \in B\} \in \mathcal{F}^\infty$. Here we do not need the same distribution of the random variables $\xi_1, \xi_2, ...$

(2) The set $A := \{\omega \in \Omega : \ \xi_n(\omega) = 0 \text{ for all } n = 1, 2, \ldots\}$ does not belong to $\mathcal{F}^\infty$ but is symmetric, because

$$\begin{aligned} A &= \left\{\omega \in \Omega : (\xi_n(\omega))_{n=1}^{\infty} \in B\right\} \\ &= \left\{\omega \in \Omega : \left(\xi_{\pi(n)}(\omega)\right)_{n=1}^{\infty} \in B\right\} \end{aligned}$$

for $B = \{(0, 0, 0, \ldots)\} \subseteq \mathcal{B}(\mathbb{R}^{\mathbb{N}})$.

[4] James Stirling May 1692 (Garden, Scotland) - 5/12/1770 (Edinburgh, Scotland), Scottish mathematician whose most important work Methodus Differentialis in 1730 is a treatise on infinite series, summation, interpolation and quadrature.

(3) The set $A := \{\omega \in \Omega : \Sigma_{n=1}^{\infty}|\xi_n(\omega)|$ exists and $\Sigma_{n=1}^{\infty}\xi_n(\omega) < 1\}$ does not belong to $\mathcal{F}^\infty$ but is symmetric, because

$$\begin{aligned} A &= \left\{\omega \in \Omega : (\xi_n(\omega))_{n=1}^{\infty} \in B\right\} \\ &= \left\{\omega \in \Omega : \left(\xi_{\pi(n)}(\omega)\right)_{n=1}^{\infty} \in B\right\} \end{aligned}$$

if B is the set of all summable sequences with sum strictly less than one.

4.6 Exercises

Ex 1:

Chapter 5

Integration

The introduction of the RIEMANN-integral in calculus is motivated by the simple concept of upper- and lower sums to define this integral. We recall this concept in Section 5.5.2 below. The RIEMANN-integration is coming up against its limits in simple cases: Take the function

$$f : [0,1] \to \mathbb{R} \quad \text{given by} \quad f(x) := \mathbb{1}_{\mathbb{Q}\cap[0,1]}(x).$$

This function is a pro-type of a function that is not RIEMANN-integrable, but will be LEBESGUE-integrable. Changing this function into the DIRICHLET function

$$g : [0,1] \to \mathbb{R} \quad \text{given by} \quad g(x) := \begin{cases} \frac{1}{q} & : x = \frac{p}{q} \text{ cannot be reduced} \\ 0 & : x \notin \mathbb{Q} \end{cases},$$

gives - in contrast to f - a RIEMANN-integrable function. This chapter is devoted to introduce the LEBESGUE-integral

$$\int_\Omega f d\mu = \int_\Omega f(\omega) d\mu(\omega)$$

for a measure space $(\Omega, \mathcal{F}, \mu)$ and a BOREL-measurable map $f : \Omega \to \mathbb{R}$ and to discuss LEBESGUE's characterization of RIEMANN-integrability by LEBESGUE-integrability in Section 5.5.2. From now on we use the

Convention 5.0.1. Given a measure space $(\Omega, \mathcal{F}, \mu)$, we say that a property $\mathcal{P}(\omega)$, depending on ω, holds **μ-almost everywhere** (a.e.) (or **almost surely** (a.s.) in case μ is a probability measure) if $\{\omega \in \Omega : \mathcal{P}(\omega) \text{ holds}\} \in \mathcal{F}$ and that

$$\mu\left(\{\omega \in \Omega : \mathcal{P}(\omega) \text{ does not hold}\}\right) = 0.$$

5.1 Definition of the LEBESGUE-integral

In the following we assume a measure space $(\Omega, \mathcal{F}, \mu)$. The definition of the integral $\int_\Omega f d\mu$ (also called expected value if μ is a probability measure) is done within three steps.

Definition 5.1.1. [STEP ONE: NON-NEGATIVE SIMPLE FUNCTIONS] Given a measurable $g : \Omega \to \mathbb{R}$ with representation

$$g = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i} \tag{5.1}$$

where $\alpha_i \in [0, \infty)$, $A_i \in \mathcal{F}$, and $\Omega = \bigcup_{i=1}^n A_i$ with $A_i \cap A_j = \emptyset$ for $i \neq j$, we let

$$\int_\Omega g d\mu = \int_\Omega g(\omega) d\mu(\omega) := \sum_{i=1}^{n} \alpha_i \mu(A_i).$$

We have to check that $\int_\Omega g d\mu$ is well-defined, since it might be that different representations of g give different integrals $\int_\Omega g d\mu$. However, this is not the case:

Lemma 5.1.2. *If one has two representations*

$$g = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i} = \sum_{j=1}^{m} \beta_j \mathbb{1}_{B_j}$$

where $\alpha_i, \beta_j \geqslant 0$ and $A_i, B_j \in \mathcal{F}$ [1], *then $\sum_{i=1}^n \alpha_i \mu(A_i) = \sum_{j=1}^m \beta_j \mu(B_j)$.*

Proof. We define a system of sets $C_1, ..., C_N \in \mathcal{F}$ where each C_k is given by

$$C_k = D_1 \cap ... \cap D_{n+m}$$

where $D_i \in \{A_i, A_i^c\}$ for $i = 1, ..., n$ and $D_{n+j} \in \{B_j, B_j^c\}$ for $j = 1, ..., m$. It is easy to see that $N = 2^{n+m}$. Moreover, it holds

(a) $C_k \cap C_l = \emptyset$ if $k \neq l$,

(b) $\bigcup_{k=1}^N C_k = \Omega$,

[1] We do not assume that the $(A_i)_{i=1}^n$ and $(B_j)_{j=1}^m$ are pair-wise disjoint or cover Ω.

(c) for all A_i there is a set $I_i \subseteq \{1, ..., N\}$ such that $A_i = \bigcup_{k \in I_i} C_k$,

(d) for all B_j there is a set $J_j \subseteq \{1, ..., N\}$ such that $B_j = \bigcup_{k \in J_j} C_k$.

Now we get that

$$\sum_{i: k \in I_i} \alpha_i = \sum_{j: k \in J_j} \beta_j \quad \text{whenever} \quad C_k \neq \emptyset$$

and

$$\begin{aligned}
\sum_{i=1}^{n} \alpha_i \mu(A_i) &= \sum_{i=1}^{n} \sum_{k \in I_i} \alpha_i \mu(C_k) = \sum_{k=1}^{N} \left(\sum_{i: k \in I_i} \alpha_i \right) \mu(C_k) \\
&= \sum_{k=1}^{N} \left(\sum_{j: k \in J_j} \beta_j \right) \mu(C_k) = \sum_{j=1}^{n} \beta_j \mu(B_j).
\end{aligned}$$

□

One might wonder why we assumed that the simple functions are non-negative. If we do not assume this, then the definition $\int_\Omega g d\mu$ might fail as we could get in the sum $\sum_{i=1}^{n} \alpha_i \mu(A_i)$ a situation like $\infty - \infty$.

Definition 5.1.3. [STEP TWO: NON-NEGATIVE FUNCTIONS] Given a BOREL-measurable map $f : \Omega \to \mathbb{R}$ with $f(\omega) \geqslant 0$ for all $\omega \in \Omega$. Then

$$\begin{aligned}
\int_\Omega f d\mu = \int_\Omega f(\omega) d\mu(\omega) & \\
&:= \sup \left\{ \int_\Omega g d\mu : 0 \leqslant g(\omega) \leqslant f(\omega), g \text{ is a simple function} \right\}.
\end{aligned}$$

In the last step we define the LEBESGUE-integral for a general BOREL-measurable map. To this end we decompose $f : \Omega \to \mathbb{R}$ into its positive and negative part

$$f(\omega) = f^+(\omega) - f^-(\omega)$$

with

$$f^+(\omega) := \max\{f(\omega), 0\} \geqslant 0 \quad \text{and} \quad f^-(\omega) := \max\{-f(\omega), 0\} \geqslant 0.$$

Definition 5.1.4. [STEP THREE: THE GENERAL CASE] Let $f : \Omega \to \mathbb{R}$ be a BOREL-measurable map.

(1) If $\int_\Omega f^+ d\mu < \infty$ or $\int_\Omega f^- d\mu < \infty$, then we say that the LEBESGUE-integral $\int_\Omega f d\mu$ exists and set

$$\int_\Omega f d\mu := \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu \in [-\infty, \infty].$$

(2) The map f is called **integrable** provided that

$$\int_\Omega f^+ d\mu < \infty \quad \text{and} \quad \int_\Omega f^- d\mu < \infty.$$

(3) If the LEBESGUE-integral $\int_\Omega f d\mu$ exists and $A \in \mathcal{F}$, then

$$\int_A f d\mu = \int_A f(\omega) d\mu(\omega) := \int_\Omega f(\omega) 1\!\!1_A(\omega) d\mu(\omega).$$

The LEBESGUE-integral $\int_\Omega f d\mu$ is called **expectation** or **expected value** of the random variable f in case μ is a probability measure, and then also denoted by $\mathbb{E}f$ or $\mathbb{E}_\mu f$.

Remark 5.1.5 (Exercise 1)**.** In Definition 5.1.4(3) we used the fact that if $\int_\Omega f d\mu$ exists, then $\int_\Omega f 1\!\!1_A d\mu$ exists as well.

Lemma 5.1.6 (Exercise 2)**.** *If $f, g : \Omega \to \mathbb{R}$ are measurable with $\mu(\{\omega \in \Omega : f(\omega) \neq g(\omega)\}) = 0$ and if $\int_\Omega f d\mu \in \mathbb{R} \cup \{-\infty, \infty\}$ exists, then $\int_\Omega g d\mu$ exists and one has that $\int_\Omega g d\mu = \int_\Omega f d\mu$.*

5.2 Theorem about monotone convergence

The first basic theorem about integration is Theorem 5.2.2 below about monotone convergence. It is used in two different directions: firstly we need the theorem later to derive basic properties of the Lebesgue integral, secondly it opens the way to compute various concrete examples. We start with an elementary version:

Lemma 5.2.1. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $f_n, f : \Omega \to \mathbb{R}$ be non-negative measurable functions such that the f_n are* simple functions *and $0 \leqslant f_n(\omega) \uparrow f(\omega)$ for all $\omega \in \Omega$ as $n \to \infty$. Then*

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

Proof. (a) First we assume that $f = 1\!\!1_A$ for some $A \in \mathcal{F}$. Let $\varepsilon \in (0,1)$ and

$$B_n^\varepsilon := \{\omega \in A : 1-\varepsilon \leqslant f_n(\omega)\}\,.$$

Then

$$(1-\varepsilon)1\!\!1_{B_n^\varepsilon}(\omega) \leqslant f_n(\omega) \leqslant 1\!\!1_A(\omega).$$

Since $B_n^\varepsilon \subseteq B_{n+1}^\varepsilon$ and $\bigcup_{n=1}^\infty B_n^\varepsilon = A$ we get, by the continuity of μ from below that $\lim_n \mu(B_n^\varepsilon) = \mu(A)$ so that

$$(1-\varepsilon)\mu(A) \leqslant \lim_n \int_\Omega f_n d\mu.$$

Since this is true for all $\varepsilon > 0$ we finally derive

$$\int_\Omega f d\mu = \mu(A) \leqslant \lim_n \int_\Omega f_n d\mu \leqslant \int_\Omega f d\mu.$$

(b) From (a) we deduce the statement when f is a simple function.

(c) Now let us assume that f is general. We construct simple functions $0 \leqslant h_n \uparrow f$ and show $\int_\Omega f d\mu = \lim_{n\to\infty} \int_\Omega h_n d\mu$. For this let

$$f_n^0(\omega) := \sum_{k=0}^{4^n-1} \frac{k}{2^n} 1\!\!1_{\left\{\frac{k}{2^n} \leqslant f < \frac{k+1}{2^n}\right\}}(\omega)$$

so that $0 \leqslant f_n^0(\omega) \uparrow f(\omega)$ for all $\omega \in \Omega$. By the definition of the Lebesgue-integral there exists a sequence $0 \leqslant g_n(\omega) \leqslant f(\omega)$ of simple functions such that $\int_\Omega g_n d\mu \uparrow \int_\Omega f d\mu$. Hence

$$h_n := \max\left\{f_n^0, g_1, \ldots, g_n\right\}$$

is a simple function with $0 \leqslant g_n(\omega) \leqslant h_n(\omega) \uparrow f(\omega)$, so that

$$\int_\Omega g_n d\mu \leqslant \int_\Omega h_n d\mu \leqslant \int_\Omega f d\mu \quad \text{and} \quad \lim_{n\to\infty} \int_\Omega g_n d\mu = \lim_{n\to\infty} \int_\Omega h_n d\mu = \int_\Omega f d\mu.$$

(d) Finally we show the assertion. Consider

$$d_{k,n} := f_k \wedge h_n.$$

Clearly, $d_{k,n} \uparrow f_k$ as $n \to \infty$ and $d_{k,n} \uparrow h_n$ as $k \to \infty$. Therefore we get

$$\int_\Omega f d\mu = \lim_n \int_\Omega h_n d\mu = \lim_n \lim_k \int_\Omega d_{k,n} d\mu = \lim_k \lim_n \int_\Omega d_{k,n} d\mu = \lim_k \int f_k d\mu$$

where we use $\lim_k \lim_n z_{k,n} = \lim_n \lim_k z_{k,n}$ for non-negative $z_{k,n} \in \mathbb{R}$ that increase with respect to k and n (where the remaining k or l is fixed). $\square$

Theorem 5.2.2. [MONOTONE CONVERGENCE] *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f, f_1, f_2, \ldots : \Omega \to \mathbb{R}$ be measurable.*

(1) *If $0 \leqslant f_n(\omega) \uparrow f(\omega)$ a.e., then $\lim_n \int_\Omega f_n d\mu = \int_\Omega f d\mu$.*

(2) *If $0 \geqslant f_n(\omega) \downarrow f(\omega)$ a.e., then $\lim_n \int_\Omega f_n d\mu = \int_\Omega f d\mu$.*

Proof. (1) By definition we have $0 \leqslant f_n(\omega) \uparrow f(\omega)$ for all $\omega \in \Omega \backslash A$, where $\mu(A) = 0$. Hence

$$0 \leqslant g_n(\omega) := f_n(\omega) 1\!\!1_{A^c}(\omega) \uparrow f(\omega) 1\!\!1_{A^c}(\omega) =: g(\omega)$$

for all $\omega \in \Omega$. By Lemma 5.1.6 we also have $\int f_n 1\!\!1_{A^c} d\mu = \int f_n d\mu$ and $\int f 1\!\!1_{A^c} d\mu = \int f d\mu$, so that it suffices to verify the statement for g_n and g. For each g_n take a sequence of step functions $(g_{n,k})_{k \geqslant 1}$ such that $0 \leqslant g_{n,k} \uparrow g_n$, as $k \to \infty$. Setting

$$h_N := \max_{\substack{1 \leqslant k \leqslant N \\ 1 \leqslant n \leqslant N}} g_{n,k}$$

we get $h_{N-1} \leqslant h_N \leqslant \max_{1 \leqslant n \leqslant N} g_n = g_N$. Define $h := \lim_{N \to \infty} h_N$. For $1 \leqslant n \leqslant N$ it holds that

$$g_{n,N} \leqslant h_N \leqslant g_N$$

so that, by $N \to \infty$,

$$g_n \leqslant h \leqslant g,$$

and therefore

$$g = \lim_{n \to \infty} g_n \leqslant h \leqslant g.$$

Since h_N is a step function for each N and $h_N \uparrow g$ we have by Lemma 5.2.1 that $\int_\Omega g d\mu = \lim_{N \to \infty} \int_\Omega h_N d\mu$ and therefore, since $h_N \leqslant g_N$,

$$\int_\Omega g d\mu \leqslant \lim_{N \to \infty} \int_\Omega g_N d\mu.$$

On the other hand, $g_N \leqslant g_{N+1} \leqslant g$ implies

$$\lim_{N \to \infty} \int_\Omega g_N d\mu \leqslant \int_\Omega g d\mu.$$

(2) Assertion (2) follows from (1) since $0 \geqslant f_n \downarrow f$ implies $0 \leqslant -f_n \uparrow -f$. $\square$

5.3 Basic properties of the LEBESGUE-integral

Now we state a first set of basic properties of the expected value:

Proposition 5.3.1. [PROPERTIES OF THE LEBESGUE-INTEGRAL] *For a measure space* $(\Omega, \mathcal{F}, \mu)$ *and measurable* $f, g : \Omega \to \mathbb{R}$, *such that* $\int_\Omega f d\mu$ *and* $\int_\Omega g d\mu$ *exist, the following assertions hold:*

(1) *If* $f \leqslant g$ *a.e., then* $\int_\Omega f d\mu \leqslant \int_\Omega g d\mu$.

(2) *The map* f *is integrable if and only if* $|f|$ *is integrable. In this case one has*

$$\left| \int_\Omega f d\mu \right| \leqslant \int_\Omega |f| d\mu = \int_\Omega f^+ d\mu + \int_\Omega f^- d\mu < \infty.$$

(3) *If* $\int_\Omega f^+ d\mu + \int_\Omega g^+ d\mu < \infty$ *or* $\int_\Omega f^- d\mu + \int_\Omega g^- d\mu < \infty$, *then* $\int_\Omega (f+g)^+ d\mu < \infty$ *or* $\int_\Omega (f+g)^- d\mu < \infty$ *and* $\int_\Omega (f+g) d\mu = \int_\Omega f d\mu + \int_\Omega g d\mu$.

(4) *If* $a \in \mathbb{R}$, *then* $\int_\Omega (af) d\mu$ *exists and* $\int_\Omega (af) d\mu = a \int_\Omega f d\mu$.

(5) *If* f *and* g *are integrable and* $a, b \in \mathbb{R}$, *then* $af + bg$ *is integrable.*

(6) *If* $f \geqslant 0$ *a.e. and* $\int_\Omega f d\mu = 0$, *then* $f = 0$ *a.e.*

Proof. (a) We assume $f(\omega) \geqslant 0$ and $g(\omega) \geqslant 0$ for all $\omega \in \Omega$. Using the approximations f_n^0 and g_n^0 from the proof of Lemma 5.2.1 we see that

$$0 \leqslant f_n^0 \leqslant f, \quad 0 \leqslant g_n^0 \leqslant g, \quad \lim_n f_n^0 = f, \quad \lim_n g_n^0 = g.$$

Hence $\lim_n (f_n^0 + g_n^0) = f + g$ as well and $\int_\Omega (f_n^0 + g_n^0) d\mu = \int_\Omega f_n^0 d\mu + \int_\Omega g_n^0 d\mu$ (because we have simple functions) and Lemma 5.2.1 gives therefore that

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

(b) <u>Proof of (3)</u>: We have that

$$(f+g)^+ + f^- + g^- = f^+ + g^+ + (f+g)^-$$

and, by (a),

$$\int_\Omega (f+g)^+ d\mu + \int_\Omega f^- d\mu + \int_\Omega g^- d\mu = \int_\Omega f^+ d\mu + \int_\Omega g^+ d\mu + \int_\Omega (f+g)^- d\mu.$$

Under the assumptions of (3) we can rearrange this equation to

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

and are done (note that $\int_\Omega (f+g)^\pm d\mu \leqslant \int_\Omega f^\pm d\mu + \int_\Omega g^\pm d\mu$ because of $(f+g)^\pm \leqslant f^\pm + g^\pm$).

(c) Proof of (2): Here we note that $|f| = f^+ + f^-$ which implies by (3) that

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

The required inequality reads as $|\int_\Omega f^+ d\mu - \int_\Omega f^- d\mu| \leqslant \int_\Omega f^+ d\mu + \int_\Omega f^- d\mu$.

(d) Proof of (1): According to Lemma 5.1.6 we only have to check the case $f(\omega) \leqslant g(\omega)$ for all $\omega \in \Omega$ by setting f and g to zero when $f(\omega) > g(\omega)$. Moreover $\int_\Omega g^- d\mu = \infty$ gives $\int_\Omega g d\mu = -\infty$ and there is nothing to prove. So let us assume that $\int_\Omega g^- d\mu < \infty$ and define $h(\omega) := g(\omega) - f(\omega) \geqslant 0$ so that $f + h = g$ and $\int_\Omega f^- d\mu < \infty$ as well. By (3) we deduce that $\int_\Omega g d\mu = \int_\Omega f d\mu + \int_\Omega h d\mu \geqslant \int_\Omega f d\mu$.

(e) Proof of (6): Again by Lemma 5.1.6 we can assume that $f(\omega) \geqslant 0$ for all $\omega \in \Omega$. Assume that $\mu(f > 0) > 0$. By considering $\{f > 0\} = \bigcup_{n=1}^\infty \{f \geqslant 1/n\}$ and using the monotonicity of μ from below we get $\mu(f > 0) = \lim_n \mu(f \geqslant 1/n)$ and some $n_0 \geqslant 1$ with $\mu(f \geqslant 1/n_0) > 0$. Now

$$0 \leqslant \frac{1}{n_0} \mathbb{I}_{\{f \geqslant 1/n_0\}} \leqslant f$$

and therefore $\int_\Omega f d\mu \geqslant \frac{1}{n_0} \mu(f \geqslant 1/n_0) > 0$ which is a contradiction.

(d) Proof of (5): Since $(af + bg)^+ \leqslant |a||f| + |b||g|$ and $(af + bg)^- \leqslant |a||f| + |b||g|$ we get that $af + bg$ is integrable.

(e) The proof of (4) is subject to Exercise 3. □

As a corollary we extend Theorem 5.2.2 about monotone convergence:

Corollary 5.3.2. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and assume measurable maps $g, f, f_1, f_2, ... : \Omega \to \mathbb{R}$. If*

(1) $g \leqslant f_n \uparrow f$ *a.e. and* $\int_\Omega g^- d\mu < \infty$ *or*

(2) $g \geqslant f_n \downarrow f$ *a.e. and* $\int_\Omega g^+ d\mu < \infty$,

then $\lim_{n\to\infty} \int_\Omega f_n d\mu = \int_\Omega f d\mu$.

Proof. It is enough to consider (1), (2) is equivalent to (1) by a multiplication by -1. Let $h_n := f_n - g$ and $h := f - g$. Then

$$0 \leqslant h_n \uparrow h \quad a.e.$$

Theorem 5.2.2 implies that $\lim_n \int_\Omega h_n d\mu = \int_\Omega h d\mu$. Since f_n^- and f^- are integrable Proposition 5.3.1 (1) implies that $\int_\Omega h_n d\mu = \int_\Omega f_n d\mu - \int_\Omega g d\mu$ and $\int_\Omega h d\mu = \int_\Omega f d\mu - \int_\Omega g d\mu$ so that we are done. □

5.4 Lemma of Fatou and Lebesgue's theorem

In the following it is convenient to work with extended random variables.

Definition 5.4.1 (EXTENDED RANDOM VARIABLE)**.** Let $(\Omega, \mathcal{F})$ be a measurable space. A function $f : \Omega \to \mathbb{R} \cup \{-\infty, \infty\}$ is called an **extended measurable map** if and only if

(1) $f^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}(\mathbb{R})$,

(2) $f^{-1}(\{+\infty\}) \in \mathcal{F}$ and $f^{-1}(\{-\infty\}) \in \mathcal{F}$.

Given an extended map f, the positive and negative part are defined as before with the convention that $+\infty$ is a positive number and $-\infty$ is a negative number. In the following it is convenient to have the Lebesgue integral for extended Borel-measurable maps. For this reason we do not need to redo our steps toward the Lebesgue-integral, we use the following definition instead:

Definition 5.4.2. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f : \Omega \to \mathbb{R} \cup \{+\infty, -\infty\}$ be an extended measurable map.

(1) If $f(\omega) \in [0, \infty]$ for all $\omega \in \Omega$, then

$$\int_\Omega f d\mu := \begin{cases} \int_\Omega \mathbb{I}_{\{f<\infty\}} f d\mu & : \mu(f = \infty) = 0 \\ +\infty & : \mu(f = \infty) > 0 \end{cases}.$$

(2) If $\int_\Omega f^+ d\mu < \infty$ or $\int_\Omega f^- d\mu < \infty$, then

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

Moreover, we need

Lemma 5.4.3 (Exercise 4)**.** *For a measurable space* $(\Omega, \mathcal{F})$ *and measurable maps* $f_n : \Omega \to \mathbb{R}$ *the* ω*-wise defined expressions* $\liminf_{n\to\infty} f_n$ *and* $\limsup_{n\to\infty} f_n$ *are extended measurable maps.*

Proposition 5.4.4 (Lemma of Fatou)**.** *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space and* $g, f_1, f_2, ... : \Omega \to \mathbb{R}$ *be measurable.*

(1) *If* $g \leqslant f_n$ *a.e. and* $\int_\Omega g^- d\mu < \infty$*, then* $\int_\Omega (\liminf_{n\to\infty} f_n)^- d\mu < \infty$ *and*

$$\int_\Omega \liminf_{n\to\infty} f_n d\mu \leqslant \liminf_{n\to\infty} \int_\Omega f_n d\mu.$$

(2) *If* $f_n \leqslant g$ *a.e. and* $\int_\Omega g^+ d\mu < \infty$*, then* $\int_\Omega (\liminf_{n\to\infty} f_n)^+ d\mu < \infty$ *and*

$$\limsup_{n\to\infty} \int_\Omega f_n d\mu \leqslant \int_\Omega \limsup_{n\to\infty} f_n d\mu.$$

Proof. We only prove (1) as (2) can be deduced from (1) by passing from f_n to $-f_n$. We let $Z_k := \inf_{n\geqslant k} f_n$ so that $Z_k \uparrow \liminf_n f_n$, $g \leqslant Z_k$ a.e. and $g \leqslant \liminf_n f_n$ a.e. This explains $\int_\Omega (\liminf_{n\to\infty} f_n)^- d\mu \leqslant \int_\Omega g^- d\mu < \infty$. Furthermore, applying Corollary 5.3.2 gives that

$$\begin{aligned}\int_\Omega \liminf_n f_n d\mu = \lim_n \int_\Omega Z_n d\mu &= \lim_n \left(\int_\Omega \inf_{k\geqslant n} f_k d\mu \right) \\ &\leqslant \lim_n \left(\inf_{k\geqslant n} \int_\Omega f_k d\mu \right) = \liminf_n \int_\Omega f_n d\mu.\end{aligned}$$

□

Proposition 5.4.5 (Lebesgue's Theorem, dominated convergence). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $g, f, f_1, f_2, ... : \Omega \to \mathbb{R}$ be measurable $|f_n| \leqslant |g|$ a.e. Assume that g is integrable and that $f = \lim_{n\to\infty} f_n$ a.e. Then f is integrable and one has that*

$$\int_\Omega f d\mu = \lim_n \int_\Omega f_n d\mu.$$

Proof. Applying Fatou's Lemma gives

$$\begin{aligned}\int_\Omega f d\mu = \int_\Omega \liminf_{n\to\infty} f_n d\mu \leqslant \liminf_{n\to\infty} \int_\Omega f_n d\mu \\ \leqslant \limsup_{n\to\infty} \int_\Omega f_n d\mu \leqslant \int_\Omega \limsup_{n\to\infty} f_n d\mu = \int_\Omega f d\mu.\end{aligned}$$

□

5.5 Examples: discrete and continuous

One feature of the Lebesgue-integral is that it provides an unified framework to different mathematical objects, in our case to sums and to the Riemann-integral.

5.5.1 Countable probability spaces

Let Ω be finite or countably infinite, $\mathcal{F} = 2^\Omega$, assume $p_\omega \in [0, 1]$ for $\omega \in \Omega$, and define the probability measure $\mathbb{P}$ by

$$\mathbb{P}(A) := \sum_{\omega \in A} p_\omega.$$

Any function $f : \Omega \to \mathbb{R}$ is measurable and for a non-negative f we have

$$\mathbb{E}f = \int_\Omega f d\mathbb{P} = \sum_{\omega\in\Omega} f(\omega) p_\omega.$$

The Lebesgue-integral of the map $f : \Omega \to \mathbb{R}$ exists if and only if

$$\sum_{\{\omega: f(\omega)\geqslant 0\}} f(\omega)p_\omega < \infty \quad \text{or} \quad \sum_{\{\omega: f(\omega)\leqslant 0\}} f(\omega)p_\omega > -\infty.$$

If the LEBESGUE-integral exists, then it computes to

$$\mathbb{E}f = \int_\Omega f d\mathbb{P} = \sum_{\{\omega : f(\omega) \geqslant 0\}} f(\omega)p_\omega + \sum_{\{\omega : f(\omega) \leqslant 0\}} f(\omega)p_\omega.$$

The map f is integrable if and only if $\sum_{\omega\in\Omega} |f(\omega)|p_\omega < \infty$.

5.5.2 The RIEMANN-integral

For the following let $-\infty < a < b < \infty$. We call $\mathcal{P} = (t_n)_{n=0}^N$ be a partition of $[a, b]$ provided that

$$a = t_0 < \cdots < t_N = b.$$

For a bounded function $f : [a, b] \to \mathbb{R}$ we define the upper and lower sums by

$$\begin{aligned} U_\mathcal{P}(f) &:= \sum_{i=1}^n \Big[\sup_{t\in[t_{i-1},t_i]} f(t) \Big] (t_i - t_{i-1}), \\ L_\mathcal{P}(f) &:= \sum_{i=1}^n \Big[\inf_{t\in[t_{i-1},t_i]} f(t) \Big] (t_i - t_{i-1}), \end{aligned}$$

and the corresponding upper and lower integrals

$$U(f) := \inf_\mathcal{P} U_\mathcal{P}(f) \quad \text{and} \quad L(f) := \sup_\mathcal{P} L_\mathcal{P}(f).$$

The function f is called RIEMANN-integrable provided that $U(f) = L(f)$. In this case we let

$$R - \int_a^b f(t)dt := U(f) = L(f)$$

be the RIEMANN-integral.

Proposition 5.5.1 (LEBESGUE-criterion for RIEMANN-integrability)**.** *Let $-\infty < a < b < \infty$ and $f : [a, b] \to \mathbb{R}$ be a bounded and measurable function. Then the following assertions are equivalent:*

(1) *The function f is* RIEMANN-*integrable.*

(2) *There exists a* BOREL-*set $\mathcal{N} \in \mathcal{B}(\mathbb{R})$ with $\mathcal{N} \subseteq [a, b]$ and $\lambda(\mathcal{N}) = 0$ such that f is continuous in all $x \in [a, b]\backslash\mathcal{N}$.*

In this case one has that $R - \int_a^b f(t)dt = \int_{[a,b]} f d\lambda$.

Proof. (1) $\Longrightarrow$ (2) By assumption there are partitions $\mathcal{P}_N$ and $\mathcal{Q}_N$ such that

$$R - \int_a^b f(t)dt = \lim_N U_{\mathcal{P}_N}(f) = \lim_N L_{\mathcal{Q}_N}(f).$$

Replacing $\mathcal{P}_N$ and $\mathcal{Q}_N$ by the union $\mathcal{P}_1 \cup \cdots \cup \mathcal{P}_N \cup \mathcal{Q}_1 \cup \cdots \cup \mathcal{Q}_N$, we may assume that $\mathcal{P}_N = \mathcal{Q}_N$ and that $\mathcal{P}_N$ is a sub-partition of $\mathcal{P}_{N+1}$. So we can choose $\mathcal{P}_N$ with

$$a = t_0^N < t_1^N < \cdots < t_{n_N}^N = b$$

such that

$$L_{\mathcal{P}_1}(f) \leqslant L_{\mathcal{P}_2}(f) \leqslant \cdots \leqslant R - \int_a^b f(t)dt \leqslant \cdots \leqslant U_{\mathcal{P}_2}(f) \leqslant U_{\mathcal{P}_1}(f).$$

For a subset $I \subseteq [a,b]$ define the oscillation

$$\operatorname{osc}(f, I) := \sup_{s,t \in I} |f(s) - f(t)|$$

so that

$$\lim_N \left[\sum_{n=1}^{n_N} \operatorname{osc}(f, [t_{i-1}^N, t_i^N])(t_i^N - t_{i-1}^N) \right] = 0.$$

We define the simple functions $g_N : [a,b] \to \mathbb{R}$ by

$$g_N(t) := \sum_{n=1}^{n_N} \operatorname{osc}(f, [t_{i-1}^N, t_i^N]) 1\!\!1_{(t_i^N - t_{i-1}^N]}(t)$$

so that $g_1(t) \geqslant g_2(t) \geqslant \cdots \geqslant 0$ for all $t \in [a,b]$ and

$$0 = \lim_N \int_{[a,b]} g_N(t) d\lambda(t) = \int_{[a,b]} \lim_N g_N(t) d\lambda(t)$$

by dominated convergence. Hence

$$\lambda\left(\left\{ t \in [a,b] : \lim_N g_N(t) \neq 0 \right\}\right) = 0.$$

Now let $t \in [a,b]$ be such that $\lim_N g_N(t) = 0$ and that t is not a grid-point of one of the partitions $\mathcal{P}_N$. Assume that f is not continuous in t. Then

$$\lim_N g_N(t) \geqslant \operatorname{osc}(f, t) > 0$$

with

$$\operatorname{osc}(f,t) := \lim_{\delta\to 0}\left[\sup\{|f(s_0)-f(s_1)| : s_0,s_1\in[a,b]\cap[t-\delta,t+\delta]\}\right],$$

which is a contradiction.

(1) $\Longrightarrow$ (2) We fix arbitrary $\varepsilon,\eta>0$. By definition we have that $\operatorname{osc}(f,t)=0$ if and only if f is continuous at t. Moreover $G_\varepsilon := \{t\in[a,b] : \operatorname{osc}(f,t)<\varepsilon\}$ is open in $[a,b]$ for $\varepsilon>0$, so that $F_\varepsilon := \{t\in[a,b] : \operatorname{osc}(f,t)\geqslant\varepsilon\}$ is closed in $[a,b]$ and in $\mathbb{R}$. By assumption, $\lambda(F_\varepsilon)=0$ so that the compactness of F_ε and Proposition 2.2.11 imply that we find finitely many open (in $[a,b]$) intervals $O_1,\ldots,O_K$ such that

$$F_\varepsilon\subseteq\bigcup_{k=1}^K O_k \quad\text{and}\quad \sum_{k=1}^K |O_k|<\eta.$$

Without loss of generality we can arrange that the closures of O_k are pair-wise disjoint. The complement $[a,b]\backslash\bigcup_{k=1}^K O_k\subseteq\{t\in[a,b] : \operatorname{osc}(f,t)<\varepsilon\}$ is a finite union of closed intervals. For each $t\in[a,b]\backslash\bigcup_{k=1}^K O_k$ we find an (in $[a,b]$) open interval $I_t\ni t$ such that $\operatorname{osc}(f,I_t)<\varepsilon$. Therefore we can cover $[a,b]\backslash\bigcup_{k=1}^K O_k$ as a finite union of of $I_{t_1},\ldots,I_{t_K}$. As $[a,b]\backslash\bigcup_{k=1}^K O_k$ is a finite union of closed intervals we can contruct from the $I_{t_1},\ldots,I_{t_K}$ a finite cover of closed intervals $J_1,\ldots,J_L$ with pair-wise disjoint interiors and such that $\operatorname{osc}(f,J_l)<\varepsilon$ for $l=1,\ldots,L$. Consequently,

$$U_{\mathcal{P}}(f)-L_{\mathcal{P}}(f)\leqslant\varepsilon[b-a]+2\sup_{t\in[a,b]}|f(t)|\eta. \qquad \square$$

Let us return to an example considered before which is the pro-type of example to illustrate the difference between the LEBESGUE-integral and the RIEMANN-integral:

Example 5.5.2 (Introducing examples of Chapter 5 continued)**.** The function

$$f(t) := \begin{cases} 1, & t\in[0,1]\text{ irrational}\\ 0, & t\in[0,1]\text{ rational}\end{cases}$$

fails to be RIEMANN-integrable, which can be seen directly from

$$L(f)=0 \quad\text{and}\quad U(f)=1$$

or it can be seen from Proposition 5.5.1 because f is discontinuous in all $t \in [0,1]$. On the other hand, the DIRICHLET function

$$g : [0,1] \to \mathbb{R} \quad \text{with} \quad g(x) := \begin{cases} \frac{1}{q} & : x = \frac{p}{q} \text{ cannot be reduced} \\ 0 & : x \notin \mathbb{Q} \end{cases},$$

is continues at all $[0,1] \backslash \mathbb{Q}$, so that this function RIEMANN-integrable.

5.5.3 The LEBESGUE-STIELTJES-integral

In our setup the LEBESGUE-STIELTJES[2]-integral is only a notation for the integral we already defined. Here we only treat one particular case, similar cases can be treated in the same way.

Assume an increasing and right-continuous function $F : \mathbb{R} \to \mathbb{R}$ such that $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$. Consider the unique probability measure $\mathbb{P}$ on $\mathcal{B}(\mathbb{R})$ with

$$\mathbb{P}((a,b]) := F(b) - F(a).$$

The measure μ is called LEBESGUE-STIELTJES-measure and one writes

$$\int f(t) dF(t) := \int_{\mathbb{R}} f(t) d\mathbb{P}(t).$$

The LEBESGUE-integral in the above notation is called LEBESGUE-STIELTJES-integral.

5.6 Change of variable formula

We want to prove a *change of variable formula* for the integrals $\int_\Omega g d\mu$. In many cases, only by this formula it is possible to compute explicitly expected values.

Proposition 5.6.1 (CHANGE OF VARIABLE). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $(E, \mathcal{E})$ be a measurable space, $f : \Omega \to E$ be a measurable map, and*

[2] Thomas Jan Stieltjes 29/12/1856 (Zwolle, Overijssel, The Netherlands) - 31/12/1894 (Toulouse, France); analysis, number theory.

$\varphi : E \to \mathbb{R}$ be $(\mathcal{E}, \mathcal{B}(\mathbb{R}))$-measurable. Assume that ν is the image measure of μ with respect to f, that means

$$\nu(B) = \mu(\{\omega \in \Omega : f(\omega) \in B\}) = \mu(f^{-1}(B)) \quad \text{for all} \quad B \in \mathcal{E}.$$

Then, for $\mu \circ f^{-1} := \nu$, one has

$$\int_B \varphi d(\mu \circ f^{-1}) = \int_{f^{-1}(B)} \varphi \circ f d\mu$$

for all $B \in \mathcal{E}$ in the sense that if one integral exists, the other exists as well, and their values are equal.

In the situation of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random variable $f : \Omega \to \mathbb{R}$, and a Borel-measurable function $\varphi : \mathbb{R} \to \mathbb{R}$ the above statement reads as

$$\mathbb{E}_{\mathbb{P}_f}\varphi = \mathbb{E}_{\mathbb{P}}(\varphi \circ f) \quad \text{where} \quad \mathbb{P}_f := \mathbb{P} \circ f^{-1} \text{ is the law of } f.$$

Proof of Proposition 5.6.1. By replacing φ by φ^+ and φ^- we can restrict ourselves to non-negative $\varphi : E \to \mathbb{R}$. Moreover, letting $\widetilde{\varphi}(x) := \mathbb{1}_B(x)\varphi(x)$ we have

$$\widetilde{\varphi}(f(\omega)) = \mathbb{1}_{f^{-1}(B)}(\omega)\varphi(f(\omega))$$

so that it is sufficient to consider the case $B = E$. Hence we have to show that

$$\int_E \varphi(x) d\nu(x) = \int_\Omega \varphi(f(\omega)) d\mu(\omega).$$

Assume now a sequence of simple functions $0 \leqslant \varphi_n(x) \uparrow \varphi(x)$ for all $x \in E$ so that $\varphi_n(f(\omega)) \uparrow \varphi(f(\omega))$ for all $\omega \in \Omega$ as well. If we can show that

$$\int_E \varphi_n d\nu = \int_\Omega \varphi_n(f) d\mu,$$

then the proof is complete. By linearity it is enough to show the equality for $\varphi_n = \mathbb{1}_B$ for some $B \in \mathcal{E}$. But now we get

$$\int_E \varphi_n d\nu = \nu(B) = \mu(f^{-1}(B)) = \int_\Omega \mathbb{1}_{f^{-1}(B)} d\mu$$
$$= \int_\Omega \mathbb{1}_B(f) d\mu = \int_\Omega \varphi_n(f) d\mu.$$

□

Let us give an example for the change of variable formula.

Definition 5.6.2 (MOMENTS). Assume that $n \in \mathbb{N}$ and that $(\Omega, \mathcal{F}, \mu)$ is a measure space.

(1) For a measurable map $f : \Omega \to \mathbb{R}$ the integral $\int_\Omega |f|^n d\mu$ is called n**-th absolute moment** of f. If $\int_\Omega f^n d\mu$ exists, then $\int_\Omega f^n d\mu$ is called n**-th moment** of f.

(2) For a measure ν on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ the integral $\int_\mathbb{R} |x|^n d\nu(x)$ is called n**-th absolute moment** of ν. If $\int_\mathbb{R} x^n d\nu(x)$ exists, then $\int_\mathbb{R} x^n d\nu(x)$ is called n**-th moment** of ν.

Corollary 5.6.3. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f : \Omega \to \mathbb{R}$ be* BOREL-*measurable with image measure ν. Then, for all $n = 1, 2, ...,$*

$$\int_\Omega |f|^n d\mu = \int_\mathbb{R} |x|^n d\nu(x) \quad \text{and} \quad \int_\Omega f^n d\mu = \int_\mathbb{R} x^n d\nu(x),$$

where the latter equality has to be understood as follows: if one side exists, then the other exists as well and they coincide.

5.7 FUBINI's Theorem

In this section we consider iterated integrals, as they appear often in applications, and show in FUBINI's [3] Theorem that integrals with respect to product measures can be written as iterated integrals and that one can change the order of integration in these iterated integrals. In many cases this provides an appropriate tool for the computation of integrals.

Theorem 5.7.1 (FUBINI'S THEOREM FOR NON-NEGATIVE FUNCTIONS). *Let $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i = 1, 2$ be σ-finite measure spaces and $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ be a non-negative $\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable function. Then one has the following:*

(1) *The functions $\omega_1 \to f(\omega_1, \omega_2^0)$ and $\omega_2 \to f(\omega_1^0, \omega_2)$ are $\mathcal{F}_1$-measurable and $\mathcal{F}_2$-measurable, respectively, for all $\omega_i^0 \in \Omega_i$.*

[3] Guido Fubini, 19/01/1879 (Venice, Italy) - 06/06/1943 (New York, USA).

(2) *The functions*

$$\omega_1 \to \int_{\Omega_2} f(\omega_1,\omega_2) d\mu_2(\omega_2) \quad \text{and} \quad \omega_2 \to \int_{\Omega_1} f(\omega_1,\omega_2) d\mu_1(\omega_1)$$

are extended $\mathcal{F}_1$-measurable and $\mathcal{F}_2$-measurable maps, respectively.

(3) *One has that*

$$\begin{aligned}
&\int_{\Omega_1\times\Omega_2} f(\omega_1,\omega_2) d(\mu_1\otimes\mu_2)((\omega_1,\omega_2)) \\
&\qquad = \int_{\Omega_1}\left[\int_{\Omega_2} f(\omega_1,\omega_2) d\mu_2(\omega_2)\right] d\mu_1(\omega_1) \\
&\qquad = \int_{\Omega_2}\left[\int_{\Omega_1} f(\omega_1,\omega_2) d\mu_1(\omega_1)\right] d\mu_2(\omega_2).
\end{aligned}$$

Remark 5.7.2. Under the condition

$$\int_{\Omega_1\times\Omega_2} f(\omega_1,\omega_2) d(\mu_1\otimes\mu_2)(\omega_1,\omega_2) < \infty$$

part (3) implies that

$$\mu_2\left(\left\{\omega_2 : \int_{\Omega_1} f(\omega_1,\omega_2) d\mu_1(\omega_1) = \infty\right\}\right) = 0$$

and

$$\mu_1\left(\left\{\omega_1 : \int_{\Omega_2} f(\omega_1,\omega_2) d\mu_2(\omega_2) = \infty\right\}\right) = 0.$$

Proof of Theorem 5.7.1. (a) First we remark it is sufficient to prove the assertions for

$$f_N(\omega_1,\omega_2) := \min\{f(\omega_1,\omega_2), N\}$$

which is bounded. The statements (1), (2), and (3) can be obtained by $N \to \infty$ if we use Proposition 3.1.4 to get the necessary measurabilities (which also works for our extended random variables) and the monotone convergence formulated in Theorem 5.2.2 to get to values of the integrals. Hence we can assume for the following that $\sup_{\omega_1,\omega_2} f(\omega_1,\omega_2) < \infty$.

(b) We want to apply the *Monotone Class Theorem* Proposition 10.3.3. Let $\mathcal{H}$ be the class of bounded $\mathcal{F}_1\otimes\mathcal{F}_2$-measurable functions $f : \Omega_1\times\Omega_2 \to \mathbb{R}$ such that the following holds:

(i) The functions $\omega_1 \to f(\omega_1, \omega_2^0)$ and $\omega_2 \to f(\omega_1^0, \omega_2)$ are $\mathcal{F}_1$-measurable and $\mathcal{F}_2$-measurable, respectively, for all $\omega_i^0 \in \Omega_i$.

(ii) The functions

$$\omega_1 \to \int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \quad \text{and} \quad \omega_2 \to \int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1)$$

are $\mathcal{F}_1$-measurable and $\mathcal{F}_2$-measurable, respectively.

(iii) One has that

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) d(\mu_1 \otimes \mu_2) &= \int_{\Omega_1} \left[\int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \right] d\mu_1(\omega_1) \\ &= \int_{\Omega_2} \left[\int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) \right] d\mu_2(\omega_2). \end{aligned}$$

Again, using Propositions 3.1.4 and Theorem 5.2.2 we see that $\mathcal{H}$ satisfies the assumptions (1), (2), and (3) of Proposition 10.3.3. As π-system I we take the system of all $F = A_1 \times A_2$ with $A_i \in \mathcal{F}_i$ and $\mu_i(A_i) < \infty$. Letting $f(\omega_1, \omega_2) = 1\!\!I_{A_1}(\omega_1) 1\!\!I_{A_2}(\omega_2)$ we easily can check that $f \in \mathcal{H}$. For instance, property (iii) follows from

$$\int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) d(\mu_1 \otimes \mu_2) = (\mu_1 \otimes \mu_2)(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$$

and, for example,

$$\begin{aligned} \int_{\Omega_1} \left[\int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \right] d\mu_1(\omega_1) &= \int_{\Omega_1} 1\!\!I_{A_1}(\omega_1) \mu_2(A_2) d\mu_1(\omega_1) \\ &= \mu_1(A_1)\mu_2(A_2). \end{aligned}$$

Applying the *Monotone Class Theorem* Proposition 10.3.3 gives that $\mathcal{H}$ contains all bounded functions $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ measurable with respect $\mathcal{F}_1 \otimes \mathcal{F}_2$. Hence we are done. □

Now we state FUBINI's Theorem for general measurable maps $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$.

Theorem 5.7.3 (Fubini's Theorem). *Let $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i = 1, 2$ be σ-finite measure spaces and $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ be an $\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable function such that*

$$\int_{\Omega_1 \times \Omega_2} |f(\omega_1, \omega_2)| d(\mu_1 \otimes \mu_2)(\omega_1, \omega_2) < \infty. \tag{5.2}$$

Then the following holds:

(1) *The functions $\omega_1 \to f(\omega_1, \omega_2^0)$ and $\omega_2 \to f(\omega_1^0, \omega_2)$ are $\mathcal{F}_1$-measurable and $\mathcal{F}_2$-measurable, respectively, for all $\omega_i^0 \in \Omega_i$.*

(2) *There are $M_i \in \mathcal{F}_i$ with $\mu_i(M_i^c) = 0$ such that the integrals*

$$\int_{\Omega_1} f(\omega_1, \omega_2^0) d\mu_1(\omega_1) \quad \text{and} \quad \int_{\Omega_2} f(\omega_1^0, \omega_2) d\mu_2(\omega_2)$$

exist and are finite for all $\omega_i^0 \in M_i$ and the functions

$$\begin{aligned} \omega_1 &\to \mathbb{1}_{M_1}(\omega_1) \int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2), \\ \omega_2 &\to \mathbb{1}_{M_2}(\omega_2) \int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) \end{aligned}$$

are $\mathcal{F}_1$-measurable and $\mathcal{F}_2$-measurable maps, respectively.

(3) *One has that*

$$\begin{aligned} &\int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) d(\mu_1 \otimes \mu_2) \\ &\qquad = \int_{\Omega_1} \left[\mathbb{1}_{M_1}(\omega_1) \int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \right] d\mu_1(\omega_1) \\ &\qquad = \int_{\Omega_2} \left[\mathbb{1}_{M_2}(\omega_2) \int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) \right] d\mu_2(\omega_2). \end{aligned}$$

Remark 5.7.4. Our understanding is that writing, for example, an expression like

$$\mathbb{1}_{M_2}(\omega_2) \int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1)$$

we only consider and compute the integral for $\omega_2 \in M_2$.

Proof of Theorem 5.7.3. The proposition follows by decomposing $f = f^+ - f^-$ and applying Theorem 5.7.1. □

We close this section with an example that should remind us to check the assumptions of FUBINI's Theorem in applications carefully.

Example 5.7.5. Let $\Omega := [-1,1]\backslash\{0\}$ and μ be the uniform distribution on Ω (see Section 2.2.8). The function $f : \Omega \to \Omega \to \mathbb{R}$ with

$$f(x,y) := \frac{xy}{(x^2+y^2)^2}$$

is not integrable on $\Omega \times \Omega$, even though the iterated integrals exist end are equal. In fact

$$\int_\Omega f(x,y)d\mu(x) = 0 \quad \text{and} \quad \int_\Omega f(x,y)d\mu(y) = 0$$

so that

$$\int_\Omega \left(\int_\Omega f(x,y)d\mu(x) \right) d\mu(y) = \int_\Omega \left(\int_\Omega f(x,y)d\mu(y) \right) d\mu(x) = 0.$$

On the other hand

$$\begin{aligned} 4\int_{\Omega\times\Omega} |f(x,y)|d(\mu\times\mu)(x,y) &\geqslant \int_0^1 \int_0^{2\pi} \frac{|\sin\varphi\cos\varphi|}{r} d\lambda(\varphi)d\lambda(r) \\ &= 2\int_0^1 \frac{1}{r} d\lambda(r) = \infty. \end{aligned}$$

The inequality holds because on the right hand side we integrate only over the area $\{(x,y) : x^2+y^2 \leqslant 1\}$ which is a subset of $[-1,1]\times[-1,1]$ and

$$\int_0^{2\pi} |\sin\varphi\cos\varphi|d\lambda(\varphi) = 4\int_0^{\pi/2} \sin\varphi\cos\varphi d\lambda(\varphi) = 2$$

follows by a symmetry argument.

5.8 Some applications of FUBINI's Theorem

Our first consequence concerns Gaussian measures introduced before in Section 2.2.5.

Corollary 5.8.1. *The following assertions hold true:*

(1) *One has* $\int_{-\infty}^{\infty} e^{-x^2} d\lambda(x) = \sqrt{\pi}$.

(2) *For* $\sigma > 0$ *and* $m \in \mathbb{R}$ *let*

$$p_{m,\sigma^2}(x) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Then, $\int_{\mathbb{R}} p_{m,\sigma^2}(x) d\lambda(x) = 1$,

$$\int_{\mathbb{R}} x p_{m,\sigma^2}(x) d\lambda(x) = m, \quad \textit{and} \quad \int_{\mathbb{R}} (x-m)^2 p_{m,\sigma^2}(x) d\lambda(x) = \sigma^2. \quad (5.3)$$

Consequently, if a random variable $f : \Omega \to \mathbb{R}$ *has as law the normal distribution* $\mathcal{N}_{m,\sigma^2}$*, then*

$$\mathbb{E}f = m \quad \textit{and} \quad \mathbb{E}(f - \mathbb{E}f)^2 = \sigma^2. \quad (5.4)$$

Proof. (1) Let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a non-negative continuous function. Fubini's Theorem gives that

$$\int_{\mathbb{R}\times\mathbb{R}} f(x,y) d(\lambda \otimes \lambda)(x,y) = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} f(x,y) d\lambda(y) \right] d\lambda(x)$$

where λ is the Lebesgue measure. Letting $f(x,y) := e^{-(x^2+y^2)}$, the above yields that

$$\begin{aligned} \int_{\mathbb{R}\times\mathbb{R}} e^{-(x^2+y^2)} d(\lambda \otimes \lambda)(x,y) &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} e^{-(x^2+y^2)} d\lambda(y) \right] d\lambda(x) \\ &= \int_{\mathbb{R}} e^{-x^2} \left[\int_{\mathbb{R}} e^{-y^2} d\lambda(y) \right] d\lambda(x) \\ &= \left[\int_{-\infty}^{\infty} e^{-x^2} d\lambda(x) \right]^2. \end{aligned}$$

On the other side, by monotone convergence we get that

$$\begin{aligned} \int_{\mathbb{R}^2} e^{-(x^2+y^2)} d(\lambda \otimes \lambda)(x,y) &= \lim_{R\to\infty} \int_{x^2+y^2 \leqslant R^2} e^{-(x^2+y^2)} (\lambda \otimes \lambda)(x,y) \\ &= \lim_{R\to\infty} \int_0^R \int_0^{2\pi} e^{-r^2} r d\lambda(r) d\lambda(\varphi) \end{aligned}$$

$$= \pi \lim_{R\to\infty}\left(1-e^{-R^2}\right)$$
$$= \pi$$

where we use Exercise 5 regarding the polar coordinates.

(2) By the change of variables $y = (x-m)/(\sqrt{2}\sigma)$ we get that

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \int_{\mathbb{R}} \frac{1}{\sqrt{\pi}} e^{-y^2} dy = 1.$$

Secondly,

$$\begin{aligned}\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} x e^{-\frac{(x-m)^2}{2\sigma^2}} dx &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} (x-m) e^{-\frac{(x-m)^2}{2\sigma^2}} dx + m \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} y e^{-\frac{y^2}{2\sigma^2}} dy + m \\ &= m\end{aligned}$$

where the integral term vanishes because the function that is integrated is odd. The remaining part is subject to Exercise 6. □

Another corollary of FUBINI's theorem is the following formula that allows to compute the mean of a measurable function by its distribution function:

Corollary 5.8.2. Assume that $(\Omega, \mathcal{F}, \mu)$ is a σ-finite measure space and let $f : \Omega \to \mathbb{R}$ be non-negative and measurable. Then,

$$\int_{\Omega} f d\mu = \int_{[0,\infty)} \mu(f > t) d\lambda(t).$$

Proof. By FUBINI's theorem,

$$\begin{aligned}\int_{\Omega} f d\mu &= \int_{\Omega}\int_{[0,\infty)} \mathbb{I}_{\{t\in[0,\infty):t<f(\omega)\}} d\lambda(t) d\mu(\omega) \\ &= \int_{[0,\infty)}\int_{\Omega} \mathbb{I}_{\{t\in[0,\infty):t<f(\omega)\}} d\mu(\omega) d\lambda(t) \\ &= \int_{[0,\infty)} \mu(f > t) d\lambda(t).\end{aligned}$$

□

Another corollary is a formula to compute the expected value of products of independent random variables:

Corollary 5.8.3. *If* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a probability space and* $f, g : \Omega \to \mathbb{R}$ *are independent random variables such that* $\mathbb{E}|f| < \infty$ *and* $\mathbb{E}|g| < \infty$. *Then*

$$\mathbb{E}|fg| < \infty \quad \textit{and} \quad \mathbb{E}fg = \mathbb{E}f\mathbb{E}g.$$

Proof. We consider the product space $(\Omega, \mathcal{F}, \mathbb{P}) \otimes (\Omega, \mathcal{F}, \mathbb{P})$ and the random variable $h : \Omega \times \Omega \to \mathbb{R}$ with $h(\omega, \eta) := f(\omega)g(\eta)$. Fubini's theorem gives

$$\begin{aligned}\int_{\Omega\times\Omega} |h(\omega,\eta)|(\mathbb{P}\otimes\mathbb{P})(\omega,\eta) &= \int_\Omega \int_\Omega |h(\omega,\eta)|d\mathbb{P}(\omega)d\mathbb{P})(\eta) \\ &= \int_\Omega |f(\omega)|d\mathbb{P}(\omega) \int_\Omega |g(\eta)|d\mathbb{P}(\eta) = \mathbb{E}|f|\mathbb{E}|g| < \infty\end{aligned}$$

and, in the same way, $\mathbb{E}h = \mathbb{E}f\mathbb{E}g$. On the other hand, by independence we know that $h : \Omega \times \Omega \to \mathbb{R}$ and $fg : \Omega \to \mathbb{R}$ have the same distribution, so that $\mathbb{E}|fg| = \int_{\Omega\times\Omega} |h(\omega,\eta)|(\mathbb{P}\otimes\mathbb{P})(\omega,\eta)$ and $\mathbb{E}fg = \int_{\Omega\times\Omega} h(\omega,\eta)(\mathbb{P}\otimes\mathbb{P})(\omega,\eta)$. □

5.9 The variance of a random variable

Besides the expected value for a random variable, its variance is another important quantity frequently used.

Definition 5.9.1 (VARIANCE). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f : \Omega \to \mathbb{R}$ be an integrable random variable. Then

$$\mathrm{var}(\mathrm{f}) = \mathbb{E}[\mathrm{f} - \mathbb{E}\mathrm{f}]^2 \in [0, \infty]$$

is called **variance**.

Now we summarize some basic properties of the variance:

Theorem 5.9.2. (1) *If* f *is integrable and* $\alpha, c \in \mathbb{R}$, *then*

$$\mathrm{var}(\alpha \mathrm{f} - \mathrm{c}) = \alpha^2 \mathrm{var}(\mathrm{f}).$$

(2) *If* $\mathbb{E}f^2 < \infty$, *then* $\mathrm{var}(\mathrm{f}) = \mathbb{E}\mathrm{f}^2 - (\mathbb{E}\mathrm{f})^2 \leqslant \mathbb{E}\mathrm{f}^2 < \infty$.

(3) *If* $f_1, ..., f_n$ *be independent random variables with finite second moment, then*

$$\mathrm{var}(f_1 + \cdots + f_n) = \mathrm{var}(f_1) + \cdots + \mathrm{var}(f_n).$$

Proof. (1) follows from

$$\mathrm{var}(\alpha \mathrm{f} - \mathrm{c}) = \mathbb{E}[(\alpha \mathrm{f} - \mathrm{c}) - \mathbb{E}(\alpha \mathrm{f} - \mathrm{c})]^2 = \mathbb{E}[\alpha \mathrm{f} - \alpha \mathbb{E}\mathrm{f}]^2 = \alpha^2 \mathrm{var}(\mathrm{f}).$$

(2) First we remark that $\mathbb{E}|f| \leqslant (\mathbb{E}f^2)^{\frac{1}{2}}$ as we shall see later by Hölder's inequality (Corollary 5.10.6), that means that any square integrable random variable is integrable. Then we simply get that

$$\mathrm{var}(\mathrm{f}) = \mathbb{E}[\mathrm{f} - \mathbb{E}\mathrm{f}]^2 = \mathbb{E}\mathrm{f}^2 - 2\mathbb{E}(\mathrm{f}\mathbb{E}\mathrm{f}) + (\mathbb{E}\mathrm{f})^2 = \mathbb{E}\mathrm{f}^2 - 2(\mathbb{E}\mathrm{f})^2 + (\mathbb{E}\mathrm{f})^2.$$

(3) The formula follows from

$$\begin{aligned}
\mathrm{var}(f_1 + \cdots + f_n) &= \mathbb{E}((f_1 + \cdots + f_n) - \mathbb{E}(f_1 + \cdots + f_n))^2 \\
&= \mathbb{E}\left(\sum_{i=1}^n (f_i - \mathbb{E}f_i)\right)^2 \\
&= \mathbb{E}\sum_{i,j=1}^n (f_i - \mathbb{E}f_i)(f_j - \mathbb{E}f_j) \\
&= \sum_{i,j=1}^n \mathbb{E}\left((f_i - \mathbb{E}f_i)(f_j - \mathbb{E}f_j)\right) \\
&= \sum_{i=1}^n \mathbb{E}(f_i - \mathbb{E}f_i)^2 + \sum_{i \neq j} \mathbb{E}\left((f_i - \mathbb{E}f_i)(f_j - \mathbb{E}f_j)\right) \\
&= \sum_{i=1}^n \mathrm{var}(f_i) + \sum_{i \neq j} \mathbb{E}(f_i - \mathbb{E}f_i)\mathbb{E}(f_j - \mathbb{E}f_j) \\
&= \sum_{i=1}^n \mathrm{var}(f_i)
\end{aligned}$$

where we use Corollary 5.8.3 to verify $\mathbb{E}\left((f_i - \mathbb{E}f_i)(f_j - \mathbb{E}f_j)\right) = \mathbb{E}(f_i - \mathbb{E}f_i)\mathbb{E}(f_j - \mathbb{E}f_j)$ and because $\mathbb{E}(f_i - \mathbb{E}f_i) = \mathbb{E}f_i - \mathbb{E}f_i = 0$. □

5.10 Important inequalities

In this section we prove some of the basic inequalities.

5.10.1 Chebyshev's inequality

Proposition 5.10.1 (Chebyshev's inequality). [4] *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f : \Omega \to \mathbb{R}$ be non-negative, measurable, and integrable. Then, for all $\lambda > 0$,*

$$\mu(\{\omega : f(\omega) \geqslant \lambda\}) \leqslant \frac{\int_\Omega f d\mu}{\lambda}.$$

Proof. We simply have

$$\lambda\mu(\{\omega : f(\omega) \geqslant \lambda\}) = \lambda \int_\Omega \mathbb{1}_{\{f \geqslant \lambda\}} d\mu \leqslant \int_\Omega f \mathbb{1}_{\{f \geqslant \lambda\}} d\mu \leqslant \int_\Omega f d\mu.$$

□

5.10.2 Jensen's inequality

Definition 5.10.2 (convexity). A function $g : \mathbb{R} \to \mathbb{R}$ is **convex** if and only if

$$g(\theta x + (1-\theta)y) \leqslant \theta g(x) + (1-\theta)g(y)$$

for all $0 \leqslant \theta \leqslant 1$ and all $x, y \in \mathbb{R}$. A function g is **concave** if $-g$ is convex.

Any convex function $g : \mathbb{R} \to \mathbb{R}$ is continuous (see Exercise 7) and therefore $(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}))$-measurable.

Proposition 5.10.3 (Jensen's inequality). [5] *If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, $g : \mathbb{R} \to \mathbb{R}$ convex, and $f : \Omega \to \mathbb{R}$ be a random variable with $\mathbb{E}|f| < \infty$, then*

$$g(\mathbb{E}f) \leqslant \mathbb{E}g(f)$$

where the expected value on the right-hand side might be infinite.

[4] Pafnuty Lvovich Chebyshev, 16/05/1821 (Okatovo, Russia) - 08/12/1894 (St Petersburg, Russia)

[5] Johan Ludwig William Valdemar Jensen, 08/05/1859 (Nakskov, Denmark)- 05/03/1925 (Copenhagen, Denmark).

Proof. Let $x_0 = \mathbb{E}f$. Since g is convex we find a *supporting line* in x_0, that means $a, b \in \mathbb{R}$ such that

$$ax_0 + b = g(x_0) \quad \text{and} \quad ax + b \leqslant g(x)$$

for all $x \in \mathbb{R}$. One approach to find this support line is a follows: One takes the limit of the monotone sequence $(n(g(x_0 + \frac{1}{n}) - g(x_0)))_{n=1}^{\infty}$, say $a \in \mathbb{R}$, and computes $b \in \mathbb{R}$ such that $ax_0 + b = g(x_0)$. It follows $af(\omega) + b \leqslant g(f(\omega))$ for all $\omega \in \Omega$ and

$$g(\mathbb{E}f) = a\mathbb{E}f + b = \mathbb{E}(af + b) \leqslant \mathbb{E}g(f). \qquad \square$$

Example 5.10.4. Frequently used examples of convex functions are the following:

(1) The function $g(x) := |x|$ is convex so that, for any integrable f,

$$|\mathbb{E}f| \leqslant \mathbb{E}|f|.$$

(2) For $1 \leqslant p < \infty$ the function $g(x) := |x|^p$ is convex, so that Jensen's inequality applied to $|f|$ gives that

$$(\mathbb{E}|f|)^p \leqslant \mathbb{E}|f|^p.$$

5.10.3 Hölder's inequality and Minkowski's inequality

For Example 5.10.4(2) there is another way we can go. It uses the Hölder inequality.

Proposition 5.10.5 (Hölder's inequality). [6] *Assume a measure space $(\Omega, \mathcal{F}, \mu)$ and measurable maps $f, g : \Omega \to \mathbb{R}$. If $1 < p, q < \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$, then*

$$\int_\Omega |fg| d\mu \leqslant \left(\int_\Omega |f|^p d\mu \right)^{\frac{1}{p}} \left(\int_\Omega |g|^q d\mu \right)^{\frac{1}{q}}.$$

[6]Otto Ludwig Hölder, 22/12/1859 (Stuttgart, Germany) - 29/08/1937 (Leipzig, Germany).

Proof. We can assume that $\int_\Omega |f|^p d\mu > 0$ and $\int_\Omega |g|^q d\mu > 0$. For example, assuming $\int_\Omega |f|^p d\mu = 0$ would imply $|f|^p = 0$ a.e. according to Proposition 5.3.1 so that $fg = 0$ a.e. and $\int_\Omega |fg| d\mu = 0$. Similarly, we may assume that $\int_\Omega |f|^p d\mu + \int_\Omega |g|^q d\mu < \infty$, otherwise there is nothing to prove. Hence we may set

$$\tilde{f} := \frac{f}{\left(\int_\Omega |f|^p d\mu\right)^{\frac{1}{p}}} \quad \text{and} \quad \tilde{g} := \frac{g}{\left(\int_\Omega |g|^q d\mu\right)^{\frac{1}{q}}}.$$

We notice that

$$x^a y^b \leqslant ax + by$$

for $x, y \geqslant 0$ and positive a, b with $a + b = 1$, which follows from the concavity of the logarithm (we can assume for a moment that $x, y > 0$)

$$\ln(ax + by) \geqslant a \ln x + b \ln y = \ln x^a + \ln y^b = \ln x^a y^b.$$

Setting $x := |\tilde{f}|^p$, $y := |\tilde{g}|^q$, $a := \frac{1}{p}$, and $b := \frac{1}{q}$, we get

$$|\tilde{f}\tilde{g}| = x^a y^b \leqslant ax + by = \frac{1}{p}|\tilde{f}|^p + \frac{1}{q}|\tilde{g}|^q$$

and

$$\int_\Omega |\tilde{f}\tilde{g}| d\mu \leqslant \frac{1}{p}\int_\Omega |\tilde{f}|^p d\mu + \frac{1}{q}\int_\Omega |\tilde{g}|^q d\mu = \frac{1}{p} + \frac{1}{q} = 1.$$

On the other hand side,

$$\int_\Omega |\tilde{f}\tilde{g}| d\mu = \frac{\int_\Omega |fg| d\mu}{\left(\int_\Omega |f|^p d\mu\right)^{\frac{1}{p}} \left(\int_\Omega |g|^q d\mu\right)^{\frac{1}{q}}}$$

so that we are done. □

Corollary 5.10.6. *For $0 < p < q < \infty$, a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variables $f, g : \Omega \to \mathbb{R}$ one has that $(\mathbb{E}|f|^p)^{\frac{1}{p}} \leqslant (\mathbb{E}|f|^q)^{\frac{1}{q}}$.*

The proof is subject to Exercise 8.

Corollary 5.10.7 (Hölder's inequality for sequences)**.** *Let $(a_n)_{n=1}^\infty$ and $(b_n)_{n=1}^\infty$ be sequences of real numbers. If $1 < p, q < \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$, then*

$$\sum_{n=1}^\infty |a_n b_n| \leqslant \left(\sum_{n=1}^\infty |a_n|^p\right)^{\frac{1}{p}} \left(\sum_{n=1}^\infty |b_n|^q\right)^{\frac{1}{q}}.$$

Proof. The statement is a special case of Proposition 5.10.5 when using the measure space $\Omega := \mathbb{N}$, $\mathcal{F} := 2^{\mathbb{N}}$, and $\mu(A)$ is the cardinality of A. □

Proposition 5.10.8 (Minkowski inequality). [7] *Assume a measure space* $(\Omega, \mathcal{F}, \mu)$*, measurable* $f, g : \Omega \to \mathbb{R}$*, and* $0 < p < \infty$*. Then*

$$\left(\int_\Omega |f+g|^p d\mu\right)^{\frac{1}{p}} \leqslant c_p \left[\left(\int_\Omega |f|^p d\mu\right)^{\frac{1}{p}} + \left(\int_\Omega |g|^p d\mu\right)^{\frac{1}{p}}\right]. \tag{5.5}$$

where $c_p = 1$ *for* $1 \leqslant p < \infty$ *and* $c_p = 2^{\frac{1}{p}-1}$ *for* $0 < p < 1$*.*

Proof. The case $p = 1$ follows from $|f+g| \leqslant |f| + |g|$.

For the case $1 < p < \infty$ we note that the convexity of $x \to |x|^p$ gives that

$$\left|\frac{a+b}{2}\right|^p \leqslant \frac{|a|^p + |b|^p}{2}$$

and $(a+b)^p \leqslant 2^{p-1}(a^p+b^p)$ for $a, b \geqslant 0$. Consequently, $|f+g|^p \leqslant (|f|+|g|)^p \leqslant 2^{p-1}(|f|^p + |g|^p)$ and

$$\int_\Omega |f+g|^p d\mu \leqslant 2^{p-1}\left(\int_\Omega |f|^p d\mu + \int_\Omega |g|^p d\mu\right).$$

Assuming now that $\left(\int_\Omega |f|^p d\mu\right)^{\frac{1}{p}} + \left(\int_\Omega |g|^p d\mu\right)^{\frac{1}{p}} < \infty$, otherwise there is nothing to prove, we get that $\int_\Omega |f+g|^p d\mu < \infty$ as well by the above considerations. Taking $1 < q < \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$, we continue with

$$\begin{aligned}
\int_\Omega |f+g|^p d\mu &= \int_\Omega |f+g||f+g|^{p-1} d\mu \\
&\leqslant \int_\Omega (|f|+|g|)|f+g|^{p-1} d\mu \\
&= \int_\Omega |f||f+g|^{p-1} d\mu + \int_\Omega |g||f+g|^{p-1} d\mu
\end{aligned}$$

[7] Hermann Minkowski, 22/06/1864 (Alexotas, Russian Empire; now Kaunas, Lithuania) - 12/01/1909 (Göttingen, Germany).

$$\leqslant \left(\int_\Omega |f|^p d\mu\right)^{\frac{1}{p}} \left(\int_\Omega |f+g|^{(p-1)q} d\mu\right)^{\frac{1}{q}}$$
$$+ \left(\int_\Omega |g|^p d\mu\right)^{\frac{1}{p}} \left(\int_\Omega |f+g|^{(p-1)q} d\mu\right)^{\frac{1}{q}},$$

where we have used HÖLDER's inequality. Since $(p-1)q = p$, (5.5) follows by dividing the above inequality by $\left(\int_\Omega |f+g|^p d\mu\right)^{\frac{1}{q}}$ (in case this expression is positive, otherwise there is nothing to prove) and taking into the account $1 - \frac{1}{q} = \frac{1}{p}$.

In the remaining case $0 < p \leqslant 1$ we get $|a+b|^p \leqslant |a|^p + |b|^p$ for all $a, b \in \mathbb{R}$, so that

$$\left(\int_\Omega |f+g|^p d\mu\right)^{\frac{1}{p}} \leqslant \left(\int_\Omega |f|^p d\mu + \int_\Omega |g|^p d\mu\right)^{\frac{1}{p}}$$
$$\leqslant 2^{\frac{1}{p}-1}\left[\left(\int_\Omega |f|^p d\mu\right)^{\frac{1}{p}} + \left(\int_\Omega |g|^p d\mu\right)^{\frac{1}{p}}\right],$$

where, for $1 \leqslant q = \frac{1}{p} < \infty$, we have used $|a+b|^q \leqslant 2^{q-1}(|a|^q + |b|^q)$. □

We close with a simple deviation inequality for f used frequently in probability:

Corollary 5.10.9. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f : \Omega \to \mathbb{R}$ be a random variable such that $\mathbb{E}f^2 < \infty$. Then one has, for all $\lambda > 0$,*

$$\mathbb{P}(|f - \mathbb{E}f| \geqslant \lambda) \leqslant \frac{\mathbb{E}(f - \mathbb{E}f)^2}{\lambda^2} \leqslant \frac{\mathbb{E}f^2}{\lambda^2}.$$

Proof. From Corollary 5.10.6 we get that $\mathbb{E}|f| < \infty$ so that $\mathbb{E}f$ exists. Applying Proposition 5.10.1 to $|f - \mathbb{E}f|^2$ gives that

$$\mathbb{P}(\{|f - \mathbb{E}f| \geqslant \lambda\}) = \mathbb{P}(\{|f - \mathbb{E}f|^2 \geqslant \lambda^2\}) \leqslant \frac{\mathbb{E}|f - \mathbb{E}f|^2}{\lambda^2}.$$

Finally, we use that $\mathbb{E}(f - \mathbb{E}f)^2 = \mathbb{E}f^2 - (\mathbb{E}f)^2 \leqslant \mathbb{E}f^2$. □

5.11 Exercises

Ex 1: Prove Remark 5.1.5.

Ex 2: Proof Lemma 5.1.6.

Ex 3: Proof part (4) of Proposition 5.3.1.

Ex 4: Prove Lemma 5.4.3.

Ex 5: Polar coordinates: Given a BOREL-measurable function $f : \mathbb{R}^2 \to \mathbb{R}$ that is non-negative. Prove

$$\int_{R^2} f(x,y)d(\lambda \otimes \lambda)(x,y) = \int_{[0,\infty)\times[0,2\pi)} f(x(r,\varphi), y(r,\varphi))rd(\lambda \otimes \lambda)(r,\varphi)$$

where $x(r,\varphi) := r\cos(\varphi)$ and $y(r,\varphi) := r\sin(\varphi)$.

Ex 6: Prove the second relation in (5.3).

Ex 7: Prove that all convex functions $f : \mathbb{R} \to \mathbb{R}$ are continuous.

Ex 8: Prove Corollary 5.10.6

Chapter 6

What does it mean that random variables converge?

Assume that we perform some experiment several times and get, each time, a measurement denoted by $f_1, f_2, f_3, \ldots$ To get the *true* quantity (whatever this means) we naturally consider

$$S_n = \frac{1}{n}(f_1 + \cdots + f_n)$$

for large n and hope that S_n converges to this true value as $n \to \infty$. To make this precise we have at least to clarify in which sense the convergence takes place. This will take us to the *almost sure convergence* and to the *convergence in probability*. But these are not the only relevant types of convergence. The aim of this chapter is to discuss the most important types of convergence and their relations to each other. For this it is convenient to define the space of random variables.

Definition 6.0.1. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ let

$$\mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P}) := \{f : \Omega \to \mathbb{R} \text{ random variable}\}.$$

6.1 Almost sure convergence

Definition 6.1.1. Let $f, f_1, f_2, \ldots \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$. We say that f_n *converges almost surely to* f if

$$\mathbb{P}(\{\omega \in \Omega : |f_n(\omega) - f(\omega)| \xrightarrow[n]{} 0\}) = 1.$$

We write $f_n \xrightarrow[a.s.]{} f$.

Remark 6.1.2.

(1) To formulate the above definition we need that $\{\omega \in \Omega : |f_n(\omega) - f(\omega)| \xrightarrow[n]{} 0\}) \in \mathcal{F}$. This follows from

$$\begin{aligned} & \{\omega \in \Omega : |f_n(\omega) - f(\omega)| \xrightarrow[n]{} 0\} \\ = & \left\{\omega \in \Omega : \forall\ m \geqslant 1\ \exists\ k \geqslant 1\ s.t.\ \forall\ n \geqslant k\ \ |f_n(\omega) - f(\omega)| < \frac{1}{m}\right\} \\ = & \bigcap_{m=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \left\{\omega : |f_n(\omega) - f(\omega)| < \frac{1}{m}\right\} \in \mathcal{F}. \end{aligned}$$

(2) The above definition depends on the measure $\mathbb{P}$. In general one does not have that

$$\mathbb{P}(\{\omega \in \Omega : |f_n(\omega) - f(\omega)| \xrightarrow[n]{} 0\}) = 1$$

if and only if

$$\mathbb{Q}(\{\omega \in \Omega : |f_n(\omega) - f(\omega)| \xrightarrow[n]{} 0\}) = 1$$

if $\mathbb{Q}$ is another probability measure on $\mathcal{F}$.

(3) Only few properties of f_n are transferred to f by almost sure convergence. Take, for example, $\Omega = [0,1]$, $\mathcal{F} = \mathbb{B}([0,1])$, and λ to be the Lebesgue measure ($\lambda([a,b]) = b - a$). Let f_n be

$$f_n(\omega) := \begin{cases} n^2 2^{n+2}\omega, & \omega \in \left[0, \frac{1}{2n}\right] \\ n 2^{n+2} - n^2 2^{n+2}\omega, & \omega \in \left(\frac{1}{2n}, \frac{1}{n}\right] \\ 0, & \omega \in \left(\frac{1}{n}, 1\right]. \end{cases}$$

The function $f_n : [0,1] \to \mathbb{R}$ is continuous so that f_n is a random variable. Moreover $\lim_n f_n(\omega) = 0$ for all $\omega \in [0,1]$. On the other side

$$\int_0^1 f_n(\omega)\, d\lambda(\omega) = \int_0^1 f_n(t)\, dt = 2^n \xrightarrow[n]{} \infty.$$

A useful characterization of the almost sure convergence is given by

Proposition 6.1.3. *Let $f, f_1, f_2, ... \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$. Then the following assertions are equivalent:*

(1) $f_n \xrightarrow[a.s.]{} f$.

(2) $\lim_{n\to\infty} \mathbb{P}(\{\omega \in \Omega : \sup_{k\geqslant n} |f_k(\omega) - f(\omega)| > \varepsilon\}) = 0$ *for all* $\varepsilon > 0$.

Proof. For $\varepsilon > 0$ and $k \geqslant 1$ define $A_k^\varepsilon := \{\omega \in \Omega : |f_k(\omega) - f(\omega)| > \varepsilon\}$ and notice that

$$\left\{\omega \in \Omega : \sup_{k\geqslant n} |f_k(\omega) - f(\omega)| > \varepsilon\right\} = \bigcup_{k=n}^\infty A_k^\varepsilon \in \mathcal{F}$$

and $\Omega_0 := \{\omega \in \Omega : \lim_{n\to\infty} f_n(\omega) = f(\omega)\} \in \mathcal{F}$.

(1)$\Longrightarrow$(2) For all $\varepsilon > 0$ it holds

$$\Omega_0 \subseteq \liminf_{n\to\infty} (A_n^\varepsilon)^c = \bigcup_{n=1}^\infty \bigcap_{k=n}^\infty \{\omega \in \Omega : |f_k(\omega) - f(\omega)| \leqslant \varepsilon\}.$$

If $f_n \xrightarrow[a.s.]{} f$, then

$$\begin{aligned} 1 = \mathbb{P}(\Omega_0) &\leqslant \mathbb{P}(\liminf_{n\to\infty} (A_n^\varepsilon)^c) \\ &= \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k=n}^\infty (A_k^\varepsilon)^c\right) \\ &= \lim_{n\to\infty} \mathbb{P}\left(\{\omega \in \Omega : \sup_{k\geqslant n} |f_k(\omega) - f(\omega)| \leqslant \varepsilon\}\right). \end{aligned}$$

Hence $\lim_{n\to\infty} \mathbb{P}(\{\omega \in \Omega : \sup_{k\geqslant n} |f_k(\omega) - f(\omega)| > \varepsilon\}) = 0$.

(2)$\Longrightarrow$(1) For all $\varepsilon > 0$ we have that

$$0 = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=n}^\infty A_k^\varepsilon\right) = \mathbb{P}(\limsup_{n\to\infty} A_n^\varepsilon) = 1 - \mathbb{P}(\liminf_{n\to\infty} (A_n^\varepsilon)^c).$$

This implies

$$\mathbb{P}(\liminf_{n\to\infty} (A_n^\varepsilon)^c) = 1 \quad \text{and therefore} \quad 1 = \mathbb{P}\left(\bigcap_{N=1}^\infty \liminf_{n\to\infty} (A_n^{1/N})^c\right) = \mathbb{P}(\Omega_0).$$

Here the last equality follows from $\omega \in \bigcap_{N=1}^{\infty} \liminf_{n\to\infty} (A_n^{1/N})^c$ if and only if for all $N = 1, 2, \ldots$ one has $\omega \in \liminf_{n\to\infty} (A_n^{1/N})^c$, which means that for all $N = 1, 2, \ldots$ there is an $n_N(\omega)$ such that $\omega \in \bigcap_{i=n_N}^{\infty} (A_n^{1/N})^c$ which is exactly $\omega \in \Omega_0$. □

An analogous statement is true that describes CAUCHY [1] sequences:

Proposition 6.1.4. *Let $f_1, f_2, \ldots \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$. Then the following conditions are equivalent:*

(1) $\mathbb{P}\left(\{\omega \in \Omega : (f_n(\omega))_{n=1}^{\infty} \text{ is a CAUCHY sequence }\}\right) = 1.$

(2) *For all $\varepsilon > 0$ one has that*

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{k,l\geqslant n} |f_k - f_l| \geqslant \varepsilon\right) = 0.$$

(3) *For all $\varepsilon > 0$ one has that*

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{k\geqslant n} |f_k - f_n| \geqslant \varepsilon\right) = 0.$$

Proof. (2)$\Longleftrightarrow$ (3) follows from

$$\begin{aligned} \sup_{k\geqslant n} |f_k - f_n| &\leqslant \sup_{k,l\geqslant n} |f_k - f_l| \\ &\leqslant \sup_{k\geqslant n} |f_k - f_n| + \sup_{l\geqslant n} |f_n - f_l| = 2 \sup_{k\geqslant n} |f_k - f_n|. \end{aligned}$$

(1)$\Longleftrightarrow$ (2) Let

$$\begin{aligned} A &:= \{\omega \in \Omega : (f_n(\omega))_{n=1}^{\infty} \text{ is a CAUCHY sequence }\} \\ &= \bigcap_{N=1,2,\ldots} \bigcup_{n=1,2,\ldots} \bigcap_{k>l\geqslant n} \left\{\omega \in \Omega : |f_k(\omega) - f_l(\omega)| \leqslant \frac{1}{N}\right\}. \end{aligned}$$

Consequently, we have that $\mathbb{P}(A) = 1$ if and only if

$$\mathbb{P}\left(\bigcup_{n=1,2,\ldots} \bigcap_{k>l\geqslant n} \left\{\omega \in \Omega : |f_k(\omega) - f_l(\omega)| \leqslant \frac{1}{N}\right\}\right) = 1$$

[1] Augustin Louis Cauchy, 21/08/1789 (Paris, France)- 23/05/1857 (Sceaux, France), study of real and complex analysis, theory of permutation groups.

for all $N = 1, 2, ...,$ if and only if

$$\lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k>l\geqslant n} \left\{\omega \in \Omega : |f_k(\omega) - f_l(\omega)| \leqslant \frac{1}{N}\right\}\right) = 1$$

for all $N = 1, 2, ...,$ if and only if

$$\lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k>l\geqslant n} \left\{\omega \in \Omega : |f_k(\omega) - f_l(\omega)| > \frac{1}{N}\right\}\right) = 0$$

for all $N = 1, 2, ...$ We can finish the proof by remarking that

$$\bigcup_{k>l\geqslant n} \left\{\omega \in \Omega : |f_k(\omega) - f_l(\omega)| > \frac{1}{N}\right\} = \left\{\sup_{k>l\geqslant n} |f_k(\omega) - f_l(\omega)| > \frac{1}{N}\right\}.$$

□

A first example of the almost sure convergence is the following form of the *strong law of large numbers.*

Proposition 6.1.5 (Strong law of large numbers)**.** *Let $(f_n)_{n=1}^\infty$ be a sequence of independent random variables with $\mathbb{E}f_n = 0$, $n = 1, 2, \ldots,$ and $c := \sup_n \mathbb{E}f_n^4 < \infty$. Then*

$$\frac{f_1 + \cdots + f_n}{n} \xrightarrow[a.s.]{} 0.$$

Proof. For $S_n := \sum_{k=1}^n f_k$ it holds that

$$\begin{aligned} \mathbb{E}S_n^4 = \mathbb{E}\left(\sum_{k=1}^n f_k\right)^4 &= \mathbb{E}\sum_{i,j,k,l,=1}^n f_i f_j f_k f_l \\ &= \sum_{k=1}^n \mathbb{E}f_k^4 + 3\sum_{\substack{k,l=1,\ldots,n \\ k\neq l}} \mathbb{E}f_k^2 \mathbb{E}f_l^2, \end{aligned}$$

because for distinct $\{i, j, k, l\}$ we have

$$\mathbb{E}f_i f_j^3 = \mathbb{E}f_i f_j^2 f_k = \mathbb{E}f_i f_j f_k f_l = 0$$

by independence. For example, $\mathbb{E}f_i f_j^3 = \mathbb{E}f_i \mathbb{E}f_j^3 = 0 \cdot \mathbb{E}f_j^3 = 0$, where one gets that f_j^3 is integrable by $\mathbb{E}|f_j|^3 \leqslant (\mathbb{E}|f_j|^4)^{\frac{3}{4}} \leqslant c^{\frac{3}{4}}$. Moreover, by JENSEN's inequality,

$$\left(\mathbb{E}f_k^2\right)^2 \leqslant \mathbb{E}f_k^4 \leqslant c.$$

Hence $\mathbb{E}f_k^2 f_l^2 = \mathbb{E}f_k^2 \mathbb{E}f_l^2 \leqslant c$ for $k \neq l$. Consequently,

$$\mathbb{E}S_n^4 \leqslant nc + 3n(n-1)c \leqslant 3cn^2$$

and

$$\mathbb{E}\sum_{n=1}^{\infty} \frac{S_n^4}{n^4} = \sum_{n=1}^{\infty} \mathbb{E}\frac{S_n^4}{n^4} \leqslant \sum_{n=1}^{\infty} \frac{3c}{n^2} < \infty.$$

This implies that $\frac{S_n^4}{n^4} \xrightarrow[a.s.]{} 0$ and therefore $\frac{S_n}{n} \xrightarrow[a.s.]{} 0$. □

6.2 Convergence in probability

6.2.1 Probabilistic formulation

Although we saw in Remark 6.1.2(3) that a.s. convergence may be a weak notion, there are examples where this notion turns out to be too strong.

Example 6.2.1. Let $\Omega := [0,1]$, $\mathcal{F} := \mathbb{B}([0,1])$, and λ be the LEBESGUE measure on $\mathcal{F}$. Define

$$\begin{aligned} f_1(\omega) &:= \mathbb{1}_{[0,\frac{1}{2})}(\omega)\,,\ f_2(\omega) := \mathbb{1}_{[\frac{1}{2},1)}(\omega)\,, \\ f_3(\omega) &:= \mathbb{1}_{[0,\frac{1}{4})}(\omega)\,,\ f_4(\omega) := \mathbb{1}_{[\frac{1}{4},\frac{1}{2})}(\omega)\,, \dots,\ f_6(\omega) := \mathbb{1}_{[\frac{3}{4},1)}(\omega)\,, \\ f_7(\omega) &:= \mathbb{1}_{[0,\frac{1}{8})}(\omega)\,,\ \dots \end{aligned}$$

We do not have $f_n \xrightarrow[a.s.]{} 0$ as $\#\{n : f_n(\omega) = 1\} = \infty$ for all $\omega \in [0,1)$.

Nevertheless, in Example 6.2.1 one has the feeling that $\lim_n f_n = 0$, but in what sense? One solution consists in the concept of the *convergence in probability*:

Definition 6.2.2. Let $f, f_1, f_2, ... \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$.

(1) *The sequence* $(f_n)_{n=1}^{\infty}$ *converges to* f *in probability* (we write $f_n \xrightarrow[\mathbb{P}]{} f$) if

$$\lim_{n\to\infty} \mathbb{P}(|f_n - f| > \varepsilon) = 0 \quad \text{for all} \quad \varepsilon > 0.$$

(2) The sequence $(f_n)_{n=1}^{\infty}$ is a *Cauchy sequence in probability* [2] provided that for all $\varepsilon > 0$ there exists $n(\varepsilon) \geqslant 1$ such that for all $m, n \geqslant n(\varepsilon)$ one has that

$$\mathbb{P}(|f_m - f_n| > \varepsilon) \leqslant \varepsilon.$$

Basic properties of the convergence in probability are given by

Proposition 6.2.3. *Let* $f, g, f_1, f_2, \ldots \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$.

(1) *If* $f_n \underset{\mathbb{P}}{\longrightarrow} f$, *then* $(f_n)_{n=1}^{\infty}$ *is a Cauchy sequence in probability.*

(2) Uniqueness of the limit*: If* $f_n \underset{\mathbb{P}}{\longrightarrow} f$ *and* $f_n \underset{\mathbb{P}}{\longrightarrow} g$, *then* $\mathbb{P}(f = g) = 1$.

(3) Completeness: *If* $(f_n)_{n=1}^{\infty}$ *is a Cauchy sequence in probability, then there exist a subsequence* $1 \leqslant n_1 < n_2 < \cdots$ *and* $h \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$ *with*

$$f_n \underset{\mathbb{P}}{\longrightarrow} h \quad \textit{and} \quad f_{n_k} \underset{a.s.}{\longrightarrow} h.$$

Proof. We only proof part (3), the remaining parts are subject to Exercise 1. We fix a sequence $\varepsilon_1 > \varepsilon_2 > \varepsilon_3 > \cdots > 0$ such that

$$\sum_{j=1}^{\infty} \varepsilon_j < \infty$$

and find $1 \leqslant n_1 < n_2 < \ldots$ such that

$$\mathbb{P}\left(\{\omega \in \Omega : |f_k(\omega) - f_l(\omega)| > \varepsilon_j\}\right) < \varepsilon_j$$

for $k, l \geqslant n_j$. Taking the sequence $(f_{n_j})_{j=1}^{\infty}$ we get that

$$\mathbb{P}\left(\left\{\omega \in \Omega : |f_{n_{j+1}}(\omega) - f_{n_j}(\omega)| > \varepsilon_j\right\}\right) < \varepsilon_j$$

and

$$\sum_{j=1}^{\infty} \mathbb{P}\left(\left\{\omega \in \Omega : |f_{n_{j+1}}(\omega) - f_{n_j}(\omega)| > \varepsilon_j\right\}\right) < \infty.$$

Applying the Borel-Cantelli Lemma implies

$$\mathbb{P}\left(\left\{\omega \in \Omega : |f_{n_{j+1}}(\omega) - f_{n_j}(\omega)| > \varepsilon_j \text{ infinitely often}\right\}\right) = 0.$$

[2] or *fundamental in probability*

Hence

$$\mathbb{P}\left(\left\{\omega \in \Omega : \sum_{j=1}^{\infty} |f_{n_{j+1}}(\omega) - f_{n_j}(\omega)| < \infty\right\}\right) = 1.$$

We set

$$h(\omega) := \begin{cases} f_{n_1}(\omega) + \sum_{j=1}^{\infty}(f_{n_{j+1}}(\omega) - f_{n_j}(\omega)) & : \sum_{j=1}^{\infty} |f_{n_{j+1}} - f_{n_j}| < \infty \\ 0 & : \text{else} \end{cases}$$

and get that $f_{n_j} \xrightarrow[a.s.]{} h$. Finally, we have to check that $f_n \xrightarrow[\mathbb{P}]{} h$. For $\varepsilon > 0$ one gets

$$\begin{aligned} \mathbb{P}(|f_n - h| > \varepsilon) &\leqslant \mathbb{P}(|f_n - f_{n_j}| + |f_{n_j} - h| > \varepsilon) \\ &\leqslant \mathbb{P}\left(|f_n - f_{n_j}| > \frac{\varepsilon}{2}\right) + \mathbb{P}\left(|f_{n_j} - h| > \frac{\varepsilon}{2}\right), \end{aligned}$$

where the last inequality follows from

$$\left\{|f_n - f_{n_j}| \leqslant \frac{\varepsilon}{2}\right\} \cap \left\{|f_{n_j} - h| \leqslant \frac{\varepsilon}{2}\right\} \subseteq \{|f_n - f_{n_j}| + |f_{n_j} - h| \leqslant \varepsilon\}.$$

Then $\lim_{j\to\infty} \mathbb{P}\left(|f_{n_j} - h| > \frac{\varepsilon}{2}\right) = 0$ follows from Proposition 6.2.4 (1). We conclude by

$$\mathbb{P}\left(|f_n - f_{n_j}| > \frac{\varepsilon}{2}\right) \leqslant \eta$$

whenever $n, n_j \geqslant n(\eta, \varepsilon) \geqslant 1$ and $\eta > 0$ is arbitrary. □

Example (Example 6.2.1 continued)**.** We have that $f_n \xrightarrow[\lambda]{} 0$. In fact,

$$\lim_n \lambda(\{\omega \in [0,1] : |f_n(\omega)| > \varepsilon\}) \leqslant \lim_n \lambda(\{\omega \in [0,1] : f_n(\omega) \neq 0\}) = 0$$

since

$$\lambda(\{\omega \in [0,1] : f_n(\omega) \neq 0\}) = \begin{cases} \frac{1}{2} & \text{for } n = 1, 2 \\ \frac{1}{4} & \text{for } n = 3, 4, 5, 6 \\ \frac{1}{8} & \text{for } n = 7, \ldots \\ \vdots & \end{cases}.$$

Now we can clarify the relation between the almost sure convergence and the convergence in probability.

Proposition 6.2.4. *For $f, f_1, f_2, \ldots \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$ one has:*

(1) *If* $f_n \xrightarrow[a.s.]{} f$*, then* $f_n \xrightarrow[\mathbb{P}]{} f$*.*

(2) *If* $f_n \xrightarrow[\mathbb{P}]{} f$*, then there exists* $1 \leqslant n_1 < n_2 < n_3 < \cdots$ *such that* $f_{n_k} \xrightarrow[a.s.]{} f$ *as* $k \to \infty$*.*

(3) *One has* $f_n \xrightarrow[\mathbb{P}]{} f$ *if and only if for all sub-sequences* $1 \leqslant n_1 < n_2 < n_3 < \ldots$ *there is a sub-sequence* $1 \leqslant n_{k_1} < n_{k_2} < n_{k_3} < \cdots$ *such that*

$$f_{n_{k_l}} \xrightarrow[a.s.]{} f \quad as \quad l \to \infty.$$

(4) Algebraic operations*: If* $f_n \xrightarrow[\mathbb{P}]{} f$ *and* $g_n \xrightarrow[\mathbb{P}]{} g$ *and* $\lambda, \mu \in \mathbb{R}$*, then*

$$\lambda f_n + \mu g_n \xrightarrow[\mathbb{P}]{} \lambda f + \mu g \quad and \quad f_n g_n \xrightarrow[\mathbb{P}]{} fg.$$

Proof. (1) The first assertion follows from Proposition 6.1.3 and

$$\mathbb{P}(|f_n - f| > \varepsilon) \leqslant \mathbb{P}\left(\sup_{k \geqslant n} |f_k - f| > \varepsilon\right).$$

(2) The sequence $(f_n)_{n=1}^{\infty}$ is a CAUCHY sequence in probability so that by Proposition 6.2.3 there is a subsequence $(n_k)_{k=1}^{\infty}$ and a random variable h such that $f_{n_k} \xrightarrow[a.s.]{} h$ as $k \to \infty$. But this implies $f_{n_k} \xrightarrow[\mathbb{P}]{} h$ and $f_{n_k} \xrightarrow[\mathbb{P}]{} f$ as $k \to \infty$ so that $f = h$ a.s.

(3) Let us first assume that $f_n \xrightarrow[\mathbb{P}]{} f$. Then $f_{n_k} \xrightarrow[\mathbb{P}]{} f$ as $k \to \infty$ and part (2) implies the existence of one more sub-sequence $(n_{k_l})_{l=1}^{\infty}$ such that $f_{n_{k_l}} \xrightarrow[a.s.]{} f$ as $l \to \infty$. In order to prove the opposite implication, assume that f_n does not converge to f in probability. We find an $\varepsilon > 0$ and a subsequence $(n_k)_{k=1}^{\infty}$ such that $\mathbb{P}(|f_{n_k} - f| > \varepsilon) > \varepsilon$. However, by assumption there is a subsequence $(n_{k_l})_{l=1}^{\infty}$ such that $f_{n_{k_l}} \xrightarrow[a.s.]{} f$ as $l \to \infty$ and the same in probability. But this implies that $\mathbb{P}(|f_{n_{k_l}} - f| > \varepsilon) \to 0$ as $l \to \infty$ which is a contradiction.

(4) See Exercise 2. □

Example (Example 6.2.1 continued)**.** What is a possible sub-sequence for a.s. convergence? One can take

$$f_1 = \mathbb{1}_{[0\frac{1}{2})},\ f_3 = \mathbb{1}_{[0\frac{1}{4})},\ f_7 = \mathbb{1}_{[0\frac{1}{8})}, \ldots.$$

Proposition 6.2.5 (Continuous mapping theorem). *Assume that for random variables* $f, f_1, f_2, \ldots \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$ *we have* $f_n \xrightarrow[\mathbb{P}]{} f$. *If* $\varphi : \mathbb{R} \to \mathbb{R}$ *is continuous, then*

$$\varphi(f_n) \xrightarrow[\mathbb{P}]{} \varphi(f).$$

Proof. The assertion follows directly from Proposition 6.2.4 (iii). □

We conclude this subsection with a version of the weak law of large numbers.

Proposition 6.2.6 (Weak law of large numbers). *Let* $(f_n)_{n=1}^{\infty}$ *be a sequence of random variables such that* $\mathbb{E}f_n^2 < \infty$, $\operatorname{cov}(f_k, f_l) = 0$ *for* $k \neq l$, $\mathbb{E}f_k = m$ *and*

$$\lim_n \frac{1}{n^2}\left[\operatorname{var}(f_1) + \cdots + \operatorname{var}(f_n)\right] = 0.$$

Then

$$\frac{f_1 + \cdots + f_n}{n} \xrightarrow[\mathbb{P}]{} m \quad \text{as} \quad n \to \infty.$$

Proof. By Chebyshev's inequality (Corollary 5.10.9) we have that

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{f_1 + \cdots + f_n - nm}{n}\right| > \varepsilon\right) &\leqslant \frac{\mathbb{E}|f_1 + \cdots + f_n - nm|^2}{n^2\varepsilon^2} \\
&= \frac{\mathbb{E}\left(\sum_{k=1}^n (f_k - m)\right)^2}{n^2\varepsilon^2} \\
&\leqslant \frac{\sum_{k=1}^n \operatorname{var}(f_k)}{n^2\varepsilon^2} \to 0
\end{aligned}$$

as $n \to \infty$. □

6.2.2* Analytical formulation

By the following Ky-Fan (pseudo) metric we can express this convergence in probability in terms of a metric:

Definition 6.2.7 (Ky Fan [3]-metric). Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. For $f, g \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$ we define

$$d(f, g) := \inf\{\varepsilon > 0 : \mathbb{P}(|f - g| > \varepsilon) \leqslant \varepsilon\}.$$

[3] Ky Fan, 19/12/1914 (Hangzhou, China) - 22/03/2010 (Santa Barbara, USA).

The idea behind this definition is that $(\mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P}), d)$ becomes a pseudo metric space:

Lemma 6.2.8. *For $f, g, h \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$ one has*

(1) *$d(f, g) = 0$ if and only $\mathbb{P}(f = g) = 1$,*

(2) *$d(f, h) \leqslant d(f, g) + d(g, h)$,*

(3) *$d(f, g) = d(g, f)$,*

(4) *$d(f, g) = d(f + h, g + h)$.*

The proof is subject to Exercise 3. The link to Section 6.2.1 is as follows:

Lemma 6.2.9. *For $f, f_1, f_2, \ldots \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$ the following holds:*

(1) $\lim_n d(f_n, f) = 0$ *if and only if $f_n \xrightarrow[\mathbb{P}]{} f$.*

(2) *For all $\varepsilon > 0$ there is an $n(\varepsilon) \geqslant 1$ such that for all $m, n \geqslant n(\varepsilon)$ one has $d(f_n, f_m) \leqslant \varepsilon$ if and only if $(f_n)_{n=1}^{\infty}$ is a* CAUCHY *sequence in probability.*

The proof is subject to Exercise 4. We conclude by formulating the convergence in probability in terms of metric spaces instead of pseudo metric spaces.

Definition 6.2.10. (1) For $f, g \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$ we define the equivalence relation $f \sim g$ by $\mathbb{P}(f = g) = 1$. We denote by $[f]$ the class of all random variables equivalent to f.

(2) $L_0(\Omega, \mathcal{F}, \mathbb{P})$ is the space of all equivalence classes from $\mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$.

(3) For $f, g \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$ and $\lambda \in \mathbb{R}$ we introduce the linear operations

$$\lambda[f] := [\lambda f] \quad \text{and} \quad [f] + [g] := [f + g].$$

(4) We define $d([f], [g]) := d(f, g)$.

The final result is

Proposition 6.2.11. *The space $[L_0(\Omega, \mathcal{F}, \mathbb{P}), d]$ is a complete linear metric space.*

6.3 Convergence in mean

6.3.1 Probabilistic formulation

Definition 6.3.1. Let $f, f_1, f_2, \ldots \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$.

(1) For $p \in (0, \infty)$ we say that f_n converges to f *with respect to the p-th mean* ($f_n \xrightarrow[L_p]{} f$) provided that

$$\lim_{n\to\infty} \int_\Omega |f_n - f|^p d\mathbb{P} = 0.$$

(2) For $p = \infty$ we say that f_n converges to f *uniformly* ($f_n \xrightarrow[L_\infty]{} f$) provided that

$$\lim_{n\to\infty} \operatorname*{ess\,sup}_{\omega\in\Omega} |f_n(\omega) - f(\omega)| = 0$$

where

$$\operatorname*{ess\,sup}_{\Omega} |g| := \inf\left\{ \sup_{\omega\in\Omega\setminus N} |g(\omega)| : N \in \mathcal{F}, \mathbb{P}(N) = 0 \right\}.$$

Since f_n and f are random variables and $x \mapsto |x|^p$ is a continuous function, $\omega \mapsto |f_n(\omega) - f(\omega)|^p$ is a non-negative random variable and we may integrate. Let us summarize some basic properties of this type of convergence.

Proposition 6.3.2. *Let $0 < p < q < \infty$ and $f, g, f_1, g_1, f_2, g_2, \ldots \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$.*

(1) *If $f_n \xrightarrow[L_p]{} f$, then $f_n \xrightarrow[\mathbb{P}]{} f$.*

(2) *If $f_n \xrightarrow[L_p]{} f$ and $g_n \xrightarrow[L_p]{} g$, then $f_n + g_n \xrightarrow[L_p]{} f + g$.*

(3) *If $f_n \xrightarrow[L_q]{} f$, then $f_n \xrightarrow[L_p]{} f$.*

(4) *If $f_n \xrightarrow[\mathbb{P}]{} f$ and $\mathbb{E}\sup_n |f_n|^p < \infty$, then $f \in \mathcal{L}_p(\Omega, \mathcal{F}, \mathbb{P})$ and*

$$\lim_{n\to\infty} \mathbb{E}|f_n - f|^p = 0.$$

Proof. (1) Let $\varepsilon > 0$. Then

$$\begin{aligned}\mathbb{P}(\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}) &= \mathbb{P}(\{\omega : |f_n(\omega) - f(\omega)|^p > \varepsilon^p\}) \\ &\leqslant \frac{1}{\varepsilon^p}\int_\Omega |f_n - f|^p d\mathbb{P} \xrightarrow[n]{} 0.\end{aligned}$$

(2) follows directly from the MINKOWSKI inequality Proposition 5.10.8.

(3) is an application of the HÖLDER inequality Proposition 5.10.5: Let $r := \frac{q}{p} \in (1,\infty)$ and let $1 = \frac{1}{r} + \frac{1}{s}$. Then

$$\begin{aligned}\int_\Omega |f_n - f|^p d\mathbb{P} &= \int_\Omega |f_n - f|^p \cdot 1 d\mathbb{P} \\ &\leqslant \left(\int_\Omega (|f_n - f|^p)^r \, d\mathbb{P}\right)^{\frac{1}{r}} \left(\int_\Omega 1^s d\mathbb{P}\right)^{\frac{1}{s}} \\ &= \left(\int_\Omega |f_n - f|^q d\mathbb{P}\right)^{\frac{p}{q}}.\end{aligned}$$

(4) Assume first, that the claim fails to be true. Then we find a sub-sequence $n_1 < n_2 < \cdots$ and a $\delta > 0$ such that

$$\int_\Omega |f_{n_k} - f|^p d\mathbb{P} \geqslant \delta > 0.$$

But there is also one more sub-sequence $(n_{k_l})_{l=1}^\infty$ such that $f_{n_{k_l}} \xrightarrow[a.s.]{} f$ as $l \to \infty$. Hence to get the contradiction, it suffices to prove the original claim under the stronger assumption that $f_n \xrightarrow[a.s.]{} f$. By LEBESGUE's theorem of dominated convergence we get that

$$\mathbb{E}|f|^p = \lim_{n\to\infty} \mathbb{E}|f_n|^p < \infty.$$

Hence

$$\mathbb{E}\sup_n |f_n - f|^p \leqslant c_p \mathbb{E}\left(\sup_n |f_n|^p + |f|^p\right) < \infty$$

and, again by dominated convergence,

$$0 = \mathbb{E}\lim_{n\to\infty} |f_n - f|^p = \lim_{n\to\infty} \mathbb{E}|f_n - f|^p. \qquad \square$$

6.3.2* Analytical formulation

Similarly to the case of convergence in probability in Section 6.2.2 we translate the convergence in mean more into analysis. For this reason we use the following spaces:

Definition 6.3.3. Letting $p \in (0,\infty]$ and

$$\|f\|_{L_p} := \begin{cases} \left(\int_\Omega |f|^p d\mathbb{P}\right)^{\frac{1}{p}} & : \quad p \in (0,\infty), \\ \operatorname{ess\,sup}_\Omega |f| & : \quad p = \infty, \end{cases}$$

we define

$$\begin{aligned} \mathcal{L}_p(\Omega,\mathcal{F},\mathbb{P}) &:= \left\{ f : \Omega \to \mathbb{R} \text{ random variable}, \|f\|_{L_p} < \infty \right\}, \\ L_p(\Omega,\mathcal{F},\mathbb{P}) &:= \{[f] : f \in \mathcal{L}_p(\Omega,\mathcal{F},\mathbb{P})\} \text{ with } \|[f]\|_{L_p} := \|f\|_{L_p}. \end{aligned}$$

Proposition 6.3.4. *The space $\left[L_p(\Omega,\mathcal{F},\mathbb{P}), \|\cdot\|_{L_p}\right]$ is a Banach space if $p \in [1,\infty]$, and a quasi-Banach space if $p \in (0,1)$.*

Proof. We check the properties according to Defintion 10.1.4. (1) $\|[f]\|_{L_p} = 0$ if and only if $f = 0$ a.s. if and only if $[f] = 0$. (2) is obvious and (3) is a consequence of the MINKOWSKI inequality. (4) Assume a CAUCHY sequence $([f_n])_{n=1}^\infty \subset L_p$. Then $(f_n)_{n=1}^\infty$ is a Cauchy sequence in probability because

$$\mathbb{P}\left(\{\omega : |f_n(\omega) - f_m(\omega)| > \lambda\}\right) \leqslant \frac{1}{\lambda^p}\|[f_n] - [f_m]\|_{L_p}^p \leqslant \varepsilon$$

for $n,m \geqslant n(\lambda,\varepsilon) \geqslant 1$. Hence there is a limit $f : \Omega \to \mathbb{R}$ such that $f_n \xrightarrow[\mathbb{P}]{} f$. It remains to show that $f_n \xrightarrow[L_p]{} f$ as well. We pick a sub-sequence $(f_{n_k})_{k=1}^\infty$ such that $f_{n_k} \xrightarrow[a.s.]{} f$ as $k \to \infty$. Applying the lemma of FATOU gives

$$\int_\Omega |f_m - f|^p d\mathbb{P} \leqslant \liminf_k \int_\Omega |f_m - f_{n_k}|^p d\mathbb{P} \leqslant \varepsilon$$

for $m \geqslant m(\varepsilon) \geqslant 1$. □

6.4 Uniform integrability

Now we want to understand the condition $\mathbb{E}\sup_n |f_n|^p < \infty$ for $p = 1$ from Proposition 6.3.2 (4) better. We recall that $\int_\Omega |f| d\mathbb{P} < \infty$ implies that

$$\lim_{c\to\infty} \int_{\{|f| \geqslant c\}} |f| d\mathbb{P} = 0.$$

In fact, the latter condition is (of course) equivalent to $\int_\Omega |f| d\mathbb{P} < \infty$. This leads us to the following definition of uniform integrability.

Definition 6.4.1. Let $(f_i)_{i\in I} \subseteq \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$, where I is an arbitrary index set. Then the family $(f_i)_{i\in I}$ is called *uniformly integrable* provided that for all $\varepsilon > 0$ there is a constant $c > 0$ such that

$$\sup_{i\in I} \int_{\{|f_i| \geqslant c\}} |f_i| d\mathbb{P} \leqslant \varepsilon,$$

or equivalently,

$$\lim_{c\uparrow\infty} \sup_{i\in I} \int_{\{|f_i| \geqslant c\}} |f_i| d\mathbb{P} = 0.$$

Example 6.4.2. Let $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \lambda)$ and $f_n(t) := n\mathbb{1}_{\left[0,\frac{1}{n}\right]}(t)$, $n \in I = \{1, 2, 3, ...\}$. This family is not uniformly integrable because for any $c > 0$ we have that

$$\sup_n \int_{\{|f_n| \geqslant c\}} |f_n(t)| dt = 1.$$

The name *uniformly integrable* suggests that the expected values of a uniform integrable family of random variables is uniformly bounded. In fact, we have

Lemma 6.4.3. *Let $f_i : \Omega \to \mathbb{R}$, $i \in I$ be a uniformly integrable family, then*

$$\sup_{i\in I} \mathbb{E}|f_i| < \infty.$$

Proof. We choose an $\varepsilon > 0$ and find an $c > 0$ such that

$$\sup_{i\in I} \int_{\{|f_i| \geqslant c\}} |f_i| d\mathbb{P} \leqslant \varepsilon.$$

Then, for all $i \in I$,

$$\mathbb{E}|f_i| = \int_{\{|f_i| \geqslant c\}} |f_i| d\mathbb{P} + \int_{\{|f_i| < c\}} |f_i| d\mathbb{P} \leqslant \varepsilon + \int_{\{|f_i| \leqslant c\}} c d\mathbb{P} \leqslant \varepsilon + c. \qquad \square$$

An important sufficient criteria for uniform integrability is

Lemma 6.4.4. *Let* $G : [0,\infty) \to [0,\infty)$ *non-negative and increasing such that*

$$\lim_{t\to\infty} \frac{G(t)}{t} = \infty$$

and $(f_i)_{i\in I}$ *be a family of random variables* $f_i : \Omega \to \mathbb{R}$ *such that*

$$\sup_{i\in I} \mathbb{E}G(|f_i|) < \infty.$$

Then $(f_i)_{i\in I}$ *is uniformly integrable.*

Proof. Since an increasing function is Borel measurable the expression $G(|f_i|)$ is a random variable. We let $\varepsilon > 0$ and $M := \sup_{i\in I} \mathbb{E}G(|f_i|)$ and find a $c > 0$ such that

$$\frac{M}{\varepsilon} \leqslant \frac{G(t)}{t}$$

for $t \geqslant c$. Then

$$\int_{\{|f_i|\geqslant c\}} |f_i| d\mathbb{P} \leqslant \frac{\varepsilon}{M} \int_{\{|f_i|\geqslant c\}} G(|f_i|) d\mathbb{P} \leqslant \varepsilon.$$ □

Example 6.4.5. (1) Examples for functions in Lemma 6.4.4 are $G(t) := t^p$ with $1 < p < \infty$ and $G(t) := t\log(1+t)$. Hence $\sup_{i\in I} \mathbb{E}|f_i|^p < \infty$ or $\sup_{i\in I} \mathbb{E}[|f_i|\log(1+|f_i|)] < \infty$ imply that $(f_i)_{i\in I}$ is uniformly integrable.

(2) In Example 6.4.2 we have seen that $\mathbb{E}|f_n| = 1$ does not guarantee that $(f_n)_{n=1}^\infty$ is u.i. Hence one can not take the function $G(t) = t$.

The main statement of this section is

Proposition 6.4.6. *Let* $p \geqslant 1$ *and asuume that* $f, f_1, f_2, ... \in \mathcal{L}_p(\Omega, \mathcal{F}, \mathbb{P})$ *are such that* $f_n \xrightarrow[\mathbb{P}]{} f$. *Then the following assertions are equivalent:*

(1) $f_n \xrightarrow[L_p]{} f$.

(2) $(|f_n|^p)_{n=1}^\infty$ *is u.i.*

(3) $\lim_{n\to\infty} \|f_n\|_{L_p} = \|f\|_{L_p}$.

The main application of the equivalence above is

Corollary 6.4.7. *Assume that* $f, f_1, f_2, ...\mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$ *are such that* $(f_n)_{n=1}^{\infty}$ *is u.i.and* $f_n \xrightarrow[\mathbb{P}]{} f$. *Then* $\mathbb{E}|f| < \infty$ *and*

$$\lim_{n\to\infty} |\mathbb{E}f_n - Ef| \leqslant \lim_{n\to\infty} \mathbb{E}|f_n - f| = 0.$$

Proof. According to Proposition 6.2.3 there is a sub-sequence $1 \leqslant n_1 < n_2 < \cdots$ such that $f_{n_k} \xrightarrow[a.s.]{} f$ as $k \to \infty$. Applying the Lemma of Fatou, Proposition 5.4.4 and Lemma 6.4.3 we get that

$$\mathbb{E}|f| \leqslant \liminf_k \mathbb{E}|f_{n_k}| \leqslant \sup_n \mathbb{E}|f_n| < \infty.$$

Hence we can apply Proposition 6.4.6 and use the implication (ii) $\Rightarrow$ (i). □

For the proof of Proposition 6.4.6 we need some lemmata:

Lemma 6.4.8. *The conditions* $f_n \xrightarrow[\mathbb{P}]{} f$, $|f_n| \leqslant g$ *and* $|f| \leqslant g$ *for some* g *with* $\mathbb{E}g < \infty$ *imply that*

$$\lim_n \int_\Omega f_n d\mathbb{P} = \int_\Omega f d\mathbb{P}.$$

Proof. Assume that the conclusion is not true. Then there is a $\varepsilon > 0$ and a sub-sequence $n_1 < n_2 < n_3 < \cdots$ such that

$$\left| \int_\Omega f_{n_k} d\mathbb{P} - \int_\Omega f d\mathbb{P} \right| \geqslant \varepsilon.$$

But we can find one more sub-sequence n_{k_l} such that

$$f_{n_{k_l}} \xrightarrow[a.s.]{} f \quad \text{as} \quad l \to \infty.$$

Applying dominated convergence yields in this case that

$$\lim_l \int_\Omega f_{n_{k_l}} d\mathbb{P} = \int_\Omega f d\mathbb{P}$$

which is a contradiction. □

Lemma 6.4.9. *For* $f \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ *and* $\mathbb{P}(A_n) \to_n 0$ *one has*

$$\lim_n \int_{A_n} f d\mathbb{P} = 0.$$

Proof. We apply Lemma 6.4.8 to $\widetilde{f}_n := f\mathbb{1}_{A_n}$ and get

$$|\widetilde{f}_n| \leqslant |f| =: g \quad \text{and} \quad \widetilde{f}_n \xrightarrow[\mathbb{P}]{} 0$$

because, for $\varepsilon > 0$,

$$\mathbb{P}(|\widetilde{f}_n| > \varepsilon) \leqslant \mathbb{P}(A_n) \to 0 \quad \text{as} \quad n \to \infty.$$

□

Proof of Proposition 6.4.6. (1) $\Longrightarrow$ (2). We have for $p \geqslant 1$ that

$$\begin{aligned}\int_{\{|f_n|^p \geqslant c\}} |f_n|^p d\mathbb{P} &\leqslant 2^{p-1}\left(\int_{\{|f_n|^p \geqslant c\}} |f_n - f|^p d\mathbb{P} + \int_{\{|f_n|^p \geqslant c\}} |f|^p d\mathbb{P}\right) \\ &\leqslant 2^{p-1}\Bigg(\|f_n - f\|_{L_p}^p + \int_{\{|f_n - f|^p \geqslant c/2^p\}} |f|^p d\mathbb{P} \\ &\qquad\qquad + \int_{\{|f|^p \geqslant c/2^p\}} |f|^p d\mathbb{P}\Bigg)\end{aligned}$$

since

$$\{|f_n|^p \geqslant c\} \subseteq \{2^{p-1}|f_n - f|^p \geqslant c/2\} \cup \{2^{p-1}|f|^p \geqslant c/2\}.$$

By assumption it holds

$$\lim_n \|f_n - f\|_{L_p}^p = \lim_{c \uparrow \infty} \int_{\{|f|^p \geqslant c/2^p\}} |f|^p d\mathbb{P} = 0,$$

and for $\int_{\{|f_n - f|^p \geqslant c/2^p\}} |f|^p d\mathbb{P}$ we use Lemma 6.4.9. We deduce that

$$\int_{\{|f_n|^p \geqslant c\}} |f_n|^p d\mathbb{P} \leqslant \varepsilon$$

for $c \geqslant c(\varepsilon)$ and $n \geqslant n(c(\varepsilon), \varepsilon)$.

(2) $\Longrightarrow$ (3) By Lemma 6.4.8 we get that

$$\int_\Omega (|f_n|^p \wedge c) d\mathbb{P} \to \int_\Omega (|f|^p \wedge c) d\mathbb{P}$$

because $(|f_n|^p \wedge c) \xrightarrow[\mathbb{P}]{} (|f|^p \wedge c)$ as $x \to |x|^p \wedge c$ is a continuous function so that the continuous mapping theorem Proposition 6.2.5 applies. But then

$$|\mathbb{E}|f_n|^p - \mathbb{E}|f|^p|$$

$$
\begin{aligned}
&\leqslant |\mathbb{E}|f_n|^p - \mathbb{E}(|f_n|^p \wedge c)| + |\mathbb{E}(|f_n|^p \wedge c) - \mathbb{E}(|f|^p \wedge c)| \\
&\qquad\qquad + |\mathbb{E}(|f|^p \wedge c) - \mathbb{E}|f|^p| \\
&\leqslant \int_{\{|f_n|^p \geqslant c\}} |f_n|^p d\mathbb{P} + |\mathbb{E}(|f_n|^p \wedge c) - \mathbb{E}(|f|^p \wedge c)| + \int_{\{|f|^p \geqslant c\}} |f|^p d\mathbb{P}.
\end{aligned}
$$

(3) $\Longrightarrow$ (1) We have that

$$\big||f_n - f|^p - 2^{p-1}|f_n|^p + 2^{p-1}|f|^p\big| \leqslant 2^p|f|^p$$

and

$$|f_n - f|^p - 2^{p-1}|f_n|^p + 2^{p-1}|f|^p \xrightarrow[\mathbb{P}]{} 0.$$

Hence, by Lemma 6.4.8,

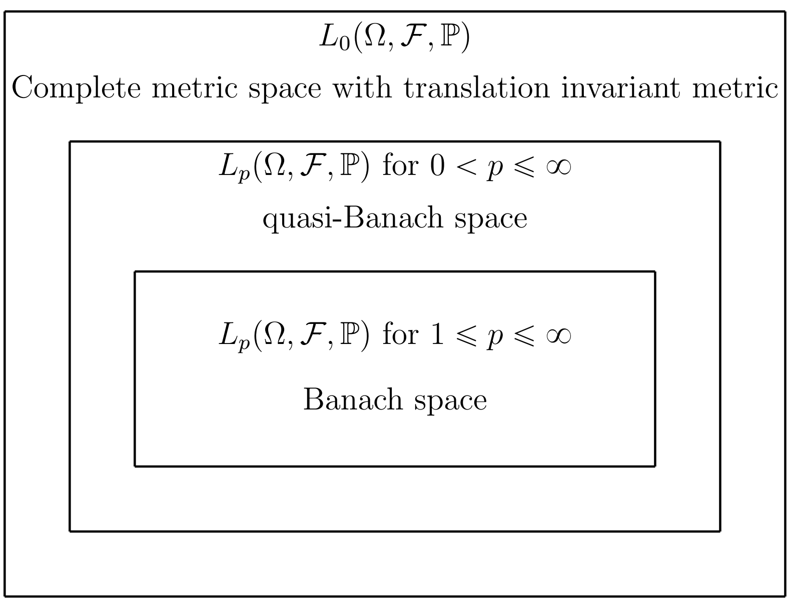
$$\lim_{n\to\infty} \mathbb{E}[|f_n - f|^p - |f_n|^p + |f|^p] = 0,$$

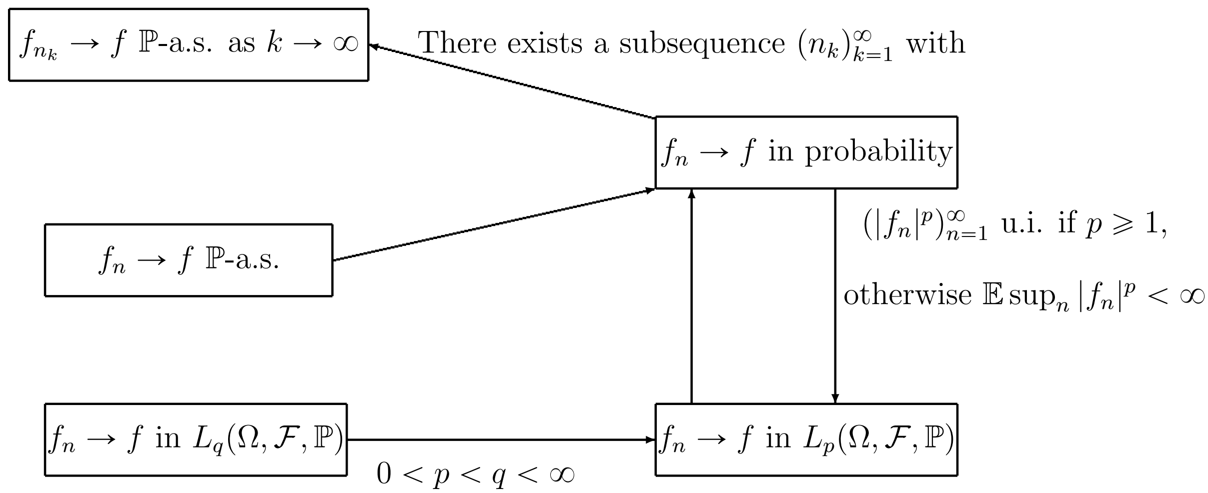
and because $\lim_{n\to\infty} \mathbb{E}|f_n|^p = \mathbb{E}|f|^p$, we conclude that

$$\lim_{n\to\infty} \mathbb{E}|f_n - f|^p = 0.$$

□

6.5 Summary





6.6 Exercises

Ex 1: Prove parts (1), and (2) of Proposition 6.2.3. prop

Ex 2: Prove (4) of Proposition 6.2.4 directly and by applying Proposition 6.2.4 (3).

Ex 3: Verify Lemma 6.2.8.

Ex 4: Verify Lemma 6.2.9.

Ex 5: For $f, g \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$ define

$$D(f,g) := \int_\Omega \frac{|f-g|}{1+|f-g|} d\mathbb{P}.$$

Prove that D generates a translation invariant metric on $L_0(\Omega, \mathcal{F}, \mathbb{P})$ and that $f_n \xrightarrow[\mathbb{P}]{} f$ if and only if $\lim_{n\to\infty} D(f_n, f) = 0$.

Chapter 7

The theorem of Radon-Nikodym and conditional expectation

So far, given a measurable space $(\Omega, \mathcal{F})$, we mainly did consider one measure on this space. However, in many situations the consideration of different measures and their relation to each other is of interest. In fact, we have been in this situation already as we defined the GAUSSIAN distribution p_{m,σ^2} for $\sigma^2 > 0$ by its density with respect to the LEBESGUE measure on $\mathbb{R}$. Could one do the same for the Dirac measure as well, i.e. does there exist a Borel function f on the real line such that

$$\delta_1((-\infty, b]) = \int_0^b f(x) d\lambda(x) \quad \text{for all} \quad b \in \mathbb{R}?$$

As one expects, the answer is *no*. What is the intrinsic difference between the GAUSSIAN distribution with positive variance and the delta distribution? This section will provide us with the theory to answer this question.

The central concept will be *absolute continuity*: Assuming two measures μ and ν on $(\Omega, \mathcal{F})$, we say that ν is absolutely continuous with respect to μ if and only if $\mu(A) = 0$ implies that $\nu(A) = 0$ and we write $\nu \ll \mu$. At first glance this concept seems to be rather weak as it only concerns null-sets, however the RADON-NIKODYM Theorem (see Theorem 7.2.1 below) demonstrates that this notion is powerful and the right one. The RADON-NIKODYM Theorem makes it possible to relate different measures to each other by densities. An application is the construction of conditional expectations, one of the fundamental concepts in stochastic process theory and filtering.

7.1 Signed measures

Definition 7.1.1 (Signed measures and absolute continuity)**.** Let $(\Omega, \mathcal{F})$ be a measurable space.

(1) A map $\nu : \mathcal{F} \to \mathbb{R}$ is called a (finite) *signed measure* if and only if

$$\nu = \nu^+ - \nu^-,$$

where ν^+ and ν^- are finite measures on $\mathcal{F}$.

(2) Assume that μ is a measure on the measurable space $(\Omega, \mathcal{F})$ and ν is a signed measure on $(\Omega, \mathcal{F})$. Then ν is *absolutely continuous* with respect to μ if and only if

$$\mu(A) = 0 \quad \text{implies} \quad \nu(A) = 0.$$

We shall write $\nu \ll \mu$.

Example 7.1.2. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $L : \Omega \to \mathbb{R}$ be integrable, and

$$\nu(A) := \int_A L d\mu.$$

Then ν is a signed measure and $\nu \ll \mu$.

Proof. We let $L^+ := \max\{L, 0\}$ and $L^- := \max\{-L, 0\}$ so that $L = L^+ - L^-$. Define

$$\nu^\pm(A) := \int_\Omega \mathbb{1}_A L^\pm d\mu.$$

Now we check that $\nu^\pm$ are finite measures. First we have that

$$\nu^\pm(\Omega) = \int_\Omega \mathbb{1}_\Omega L^\pm d\mu \leqslant \int_\Omega |L| d\mu < \infty.$$

Assume $(A_n)_{n=1}^\infty \subseteq \mathcal{F}$ to be disjoint sets.Then

$$\nu^+\left(\overset{\infty}{\underset{n=1}{\cup}} A_n\right) = \int_\Omega 1\!\!1_{\overset{\infty}{\underset{n=1}{\cup}} A_n} L^+ d\mu = \int_\Omega \left(\sum_{n=1}^{\infty} 1\!\!1_{A_n}(\omega)\right) L^+ d\mu$$
$$= \int_\Omega \lim_{N\to\infty} \left(\sum_{n=1}^{N} 1\!\!1_{A_n}\right) L^+ d\mu = \lim_{N\to\infty} \int_\Omega \left(\sum_{n=1}^{N} 1\!\!1_{A_n}\right) L^+ d\mu = \sum_{n=1}^{\infty} \nu^+(A_n)$$

where we have used the theorem about dominated convergence (Theorem 5.2.2). The same can be done for L^-. □

Next we prove the HAHN- and the JORDAN-decompositions of a signed measure which give an *optimal* decomposition $\nu = \nu^+ - \nu^-$ of a signed measure ν into finite measures ν^+, ν^-. Note the decomposition from Definition 7.1.1 (1) is not unique.

Proposition 7.1.3 (HAHN-DECOMPOSITION)**.** *Let ν be a signed measure on $(\Omega, \mathcal{F})$. Then there exists a measurable partition $\Omega = \Omega^- \cup \Omega^+$ such that*

(1) $\nu(A) \geqslant 0$ *for all* $A \in \mathcal{F}$ *with* $A \subseteq \Omega^+$,

(2) $\nu(A) \leqslant 0$ *for all* $A \in \mathcal{F}$ *with* $A \subseteq \Omega^-$.

Given another partition $\Omega = \tilde{\Omega}^- \cup \tilde{\Omega}^+$ with these properties, one has that

$$\nu(\Omega^+ \Delta\, \tilde{\Omega}^+) = \nu(\Omega^- \Delta\, \tilde{\Omega}^-) = 0.$$

Proof. We follow the idea of [3, Theorem 32.1]. Let $\kappa := \sup_{B\in\mathcal{F}} \nu(B)$. Then there exists a sequence $(\Omega_n)_{n=1}^\infty \subseteq \mathcal{F}$ such that $\lim_n \nu(\Omega_n) = \kappa$. We will construct by the help of this sequence a set Ω^+ for which ν attains the supremum κ. Let $C: = \bigcup_{k=1}^\infty \Omega_k$ and consider for each $n \in \mathbb{N}$ the partition $\mathcal{C}_n$ of C given by

$$\mathcal{C}_n = \{C_1 \cap ... \cap C_n : C_k \in \{\Omega_k, C\backslash\Omega_k\}\}.$$

Define

$$\mathcal{C}_n^- := \{B \in \mathcal{C}_n : \nu(B) < 0\} \quad \text{and} \quad \mathcal{C}_n^+ := \mathcal{C}_n \backslash \mathcal{C}_n^-.$$

Notice that $\{B \in \mathcal{C}_n : B \subseteq \Omega_n\}$ is a partition of Ω_n. Put

$$\Omega_n^+ := \bigcup_{B\in\mathcal{C}_n^+} B.$$

Since for all sets $B \in \mathcal{C}_n^+$ the measure ν is non-negative, we have that

$$\nu(\Omega_n^+) = \sum_{B \in \mathcal{C}_n^+} \nu(B) \geqslant \sum_{B \in \mathcal{C}_n^+, B \subseteq \Omega_n} \nu(B) \geqslant \sum_{B \in \mathcal{C}_n, B \subseteq \Omega_n} \nu(B) = \nu(\Omega_n).$$

This implies

$$\nu\left(\bigcup_{k=n}^{\infty} \Omega_k^+\right) = \nu(\Omega_n^+) + \sum_{k=n+1}^{\infty} \nu\left(\Omega_k^+ \backslash \left(\bigcup_{l=n}^{k-1} \Omega_l^+\right)\right) \geqslant \nu(\Omega_n^+).$$

Setting

$$\Omega^+ := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \Omega_k^+$$

we conclude that

$$\nu\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \Omega_k^+\right) = \lim_{n\to\infty} \nu\left(\bigcup_{k=n}^{\infty} \Omega_k^+\right) \geqslant \lim_{n\to\infty} \nu\left(\Omega_n^+\right) \geqslant \lim_{n\to\infty} \nu\left(\Omega_n\right) = \kappa.$$

We show that Ω^+ fulfils property (1): Assuming a measurable subset $A \subseteq \Omega^+$ with $\nu(A) < 0$ would yield to a contradiction because in this case,

$$\kappa = \nu(\Omega^+) < \nu(\Omega^+ \backslash A) \leqslant \kappa.$$

Set $\Omega^- := \Omega \backslash \Omega^+$ and asume a measurable $A \subseteq \Omega^-$ with $\nu(A) > 0$. This would again yield a contradiction because then

$$\kappa = \nu(\Omega^+) < \nu(\Omega^+ \cup A) \leqslant \kappa.$$

The last arguments also show that any exchange between Ω^+ and Ω^-, which concerns a subset A of positive measure, yields to a contradiction. This proves the uniqueness. □

From the HAHN [1]-decomposition we derive the JORDAN [2] -decomposition:

Definition 7.1.4 (JORDAN-DECOMPOSITION)**.** *Given a signed measure μ on $(\Omega, \mathcal{F})$ and $\Omega^{\pm}$ from Proposition 7.1.3, we call $\nu := \nu_J^+ - \nu_J^-$ with*

$$\nu_J^+(A) := \nu(A \cap \Omega^+) \quad \textit{and} \quad \nu_J^-(A) := -\nu(A \cap \Omega^-)$$

the JORDAN-decomposition of ν.

[1] Hans Hahn, 27/09/1879 (Vienna, Austria)- 24/07/1934 (Vienna, Austria).
[2] Camille Jordan, 05/01/1838 (Lyon, France)- 22/01/1922 (Paris, France).

The JORDAN-decomposition is unique in the following sense:

Theorem 7.1.5. *Assume a signed measure ν on $(\Omega, \mathcal{F})$ with decomposition $\nu = \nu^+ - \nu^-$. Then one has*

$$\nu_J^+(A) + \nu_J^-(A) \leqslant \nu^+(A) + \nu^-(A) \quad \textit{for all} \quad A \in \mathcal{F},$$

where the equality holds if and only if $\nu_J^+ = \nu^+$ and $\nu_J^- = \nu^-$.

Proof. We assume a HAHN-decomposition $\Omega = \Omega^- \cup \Omega^+$ for ν. With this decomposition we get that

$$\begin{aligned}\nu^+(A) + \nu^-(A) &= \nu^+(A \cap \Omega^+) + \nu^+(A \cap \Omega^-) + \nu^-(A \cap \Omega^+) + \nu^-(A \cap \Omega^-) \\ &\geqslant \nu^+(A \cap \Omega^+) - \nu^-(A \cap \Omega^+) + \nu^+(A \cap \Omega^-) - \nu^-(A \cap \Omega^-) \\ &= \nu(A \cap \Omega^+) + \nu(A \cap \Omega^-) \\ &= \nu_J^+(A \cap \Omega^+) + \nu_J^-(A \cap \Omega^-).\end{aligned}$$

Therefore the inequality is verified and we have an equality if and only if $\nu^-(A \cap \Omega^+) = \nu^+(A \cap \Omega^-) = 0$. However this condition implies that

$$\nu^+(A) = \nu^+(A \cap \Omega^+) \geqslant \nu^+(A \cap \Omega^+) - \nu^-(A \cap \Omega^+) = \nu(A \cap \Omega^+) = \nu_J^+(A)$$

and

$$\nu^-(A) = \nu^-(A \cap \Omega^-) \geqslant \nu^-(A \cap \Omega^-) - \nu^+(A \cap \Omega^-) = -\nu(A \cap \Omega^-) = \nu_J^-(A).$$

Because we have $\nu_J^-(\Omega^+) = \nu_J^+(\Omega^-) = 0$ as well, we can use the same argument, to deduce $\nu_J^+(A) \geqslant \nu^+(A)$ and $\nu_J^-(A) \geqslant \nu^-(A)$ which proves that $\nu_J^+ = \nu^+$ and $\nu_J^- = \nu^-$. □

Definition 7.1.6. *Given a signed measure ν on $(\Omega, \mathcal{F})$ with* JORDAN-*decomposition $\nu = \nu_J^+ - \nu_J^-$ we let*

$$|\nu| := \nu_J^+ + \nu_J^- \quad \textit{and} \quad |\nu|_{\mathrm{TV}} := \nu_J^+(\Omega) + \nu_J^-(\Omega)$$

be the total variation of ν.

For more information the reader is referred to [16].

7.2 Theorem of Radon-Nikodym

Theorem 7.2.1 (Radon-Nikodym)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a σ-finite measure space and ν be a signed measure with $\nu \ll \mu$. Then there exists a measurable and integrable $L : \Omega \to \mathbb{R}$ such that*

$$\nu(A) = \int_A L d\mu \quad \text{for all} \quad A \in \mathcal{F}. \tag{7.1}$$

The map L is unique in the following sense: if L and L' are maps as above satisfying (7.1), then $\mu(L \neq L') = 0$.

The Radon-Nikodym theorem was proved by Radon[3] in 1913 in the case of $\mathbb{R}^n$. The extension to the general case was done by Nikodym[4] in 1930.

Definition 7.2.2. The measurable map L is called Radon-Nikodym derivative and we write

$$L = \frac{d\nu}{d\mu}.$$

We should keep in mind the rule

$$\nu(A) = \int_\Omega \mathbb{I}_A d\nu = \int_\Omega \mathbb{I}_A L d\mu,$$

so that '$d\nu = Ld\mu$'.

Proof of Theorem 7.2.1. It is sufficient to prove the statement for finite measures μ so that we may even assume that $\mu = \mathbb{P}$ is a probability measure. Indeed, in the general case we can use a partition $\Omega = \bigcup_{n=1}^\infty \Omega_n$ with $\mu(\Omega_n) < \infty$, get $L_n : \Omega_n \to \mathbb{R}$, define $L = \sum_n \mathbb{I}_{\Omega_n} L_n$, and observe that

$$\int_\Omega |L| d\mu = \sum_{n=1}^\infty \int_{\Omega_n} |L_n| d\mu = \sum_{n=1}^\infty (\nu^+(\Omega_n) + \nu^-(\Omega_n)) = \nu^+(\Omega) + \nu^-(\Omega) < \infty,$$

where $\nu = \nu^+ - \nu^-$ is the Jordan decomposition.

[3] Johann Radon, 16/12/1887 (Tetschen, Bohemia; now Decin, Czech Republic) - 25/05/1956 (Vienna, Austria).

[4] Otton Marcin Nikodym, 13/08/1887 (Zablotow, Galicia, Austria-Hungary; now Ukraine) - 04/05/1974 (Utica, USA).

So let us assume that $\mu = \mathbb{P}$. By the JORDAN-decomposition we have $\nu = \nu^+ - \nu^-$ with $\nu^+ \ll \mathbb{P}$ and $\nu^- \ll \mathbb{P}$ because, for example, $\mathbb{P}(A) = 0$ implies $\mathbb{P}(A \cap \Omega^+) = 0$ and therefore $0 = \nu(A \cap \Omega^+) = \nu^+(A)$. Hence we can consider ν^+ and ν^- separately, i.e. we may assume that ν is a non-negative finite measure. We start our proof by defining

$$\mathcal{M} := \left\{ L \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}) : L \geqslant 0, \int_A L d\mathbb{P} \leqslant \nu(A) \text{ for all } A \in \mathcal{F} \right\}.$$

Because $0 \in \mathcal{M}$ the set $\mathcal{M}$ is non-empty. Moreover, $L_1, L_2 \in \mathcal{M}$ implies

$$L_1 \vee L_2 \leqslant L_1 + L_2 \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$$

and, with $\Omega_1 := \{L_1 \geqslant L_2\}$ and $\Omega_2 := \{L_2 > L_1\}$, the estimates

$$\int_A L_1 \vee L_2 d\mathbb{P} = \int_{A \cap \Omega_1} L_1 d\mathbb{P} + \int_{A \cap \Omega_2} L_2 d\mathbb{P} \leqslant \nu(A \cap \Omega_1) + \nu(A \cap \Omega_2) = \nu(A).$$

Consequently, $L_1 \vee L_2 \in \mathcal{M}$. Observe that

$$\kappa := \sup_{L \in \mathcal{M}} \int_\Omega L d\mathbb{P} \leqslant \nu(\Omega) < \infty$$

and find $L_n \in \mathcal{M}$ such that

$$\sup_n \int_\Omega L_n d\mathbb{P} = \kappa.$$

Letting $K_n := \max\{L_1, ..., L_n\}$, we obtain $K_n \in \mathcal{M}$, $0 \leqslant K_1 \leqslant K_2 \leqslant \cdots$, and

$$\lim_n \int_\Omega K_n d\mathbb{P} = \kappa.$$

Define the extended random variable

$$L(\omega) := \lim_n K_n(\omega) \in [0, \infty].$$

From monotone convergence we get that

$$\int_\Omega L d\mathbb{P} = \kappa < \infty$$

and may redefine L on a null set such that $L : \Omega \to \mathbb{R}$, $L \in \mathcal{M}$, and

$$\int_A L d\mathbb{P} \leqslant \nu(A) \quad \text{for} \quad A \in \mathcal{F}.$$

We want to show that $\int_A L d\mathbb{P} = \nu(A)$ for $A \in \mathcal{F}$. For this we assume that there is an $A_0 \in \mathcal{F}$ with

$$\int_{A_0} L d\mathbb{P} < \nu(A_0). \tag{7.2}$$

Since $\nu(A_0) > 0$ we conclude from $\nu \ll \mathbb{P}$ that $\mathbb{P}(A_0) > 0$. Choose an $\varepsilon > 0$ such that for

$$\nu'(A) := \nu(A) - \int_A (L + \varepsilon) d\mathbb{P}, \quad A \in \mathcal{F},$$

one has $\nu'(A_0) > 0$. The HAHN-decomposition with respect to ν' yields to a partition $\Omega = \Omega^+ \cup \Omega^-$ with $\nu'(\Omega^+) > 0$. Because $\nu' \ll \mathbb{P}$ as well, we obtain that $\mathbb{P}(\Omega^+) > 0$. Then for $L' := L(1 + \varepsilon 1\!\!1_{\Omega^+})$ we have $L' \geqslant 0$ and $L' \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$. To show that $L' \in \mathcal{M}$, we let $A \in \mathcal{F}$ and get that

$$\begin{aligned}
\int_A L' d\mathbb{P} &= \int_{A \cap \Omega^+} L' d\mathbb{P} + \int_{A \cap (\Omega^+)^c} L' d\mathbb{P} \\
&= \int_{A \cap \Omega^+} (L + \varepsilon) d\mathbb{P} + \int_{A \cap (\Omega^+)^c} L d\mathbb{P} \\
&= \int_{A \cap \Omega^+} (L + \varepsilon) d\mathbb{P} - \nu(A \cap \Omega^+) + \int_{A \cap (\Omega^+)^c} L d\mathbb{P} + \nu(A \cap \Omega^+) \\
&= -\nu'(A \cap \Omega^+) + \int_{A \cap (\Omega^+)^c} L d\mathbb{P} + \nu(A \cap \Omega^+) \\
&\leqslant \int_{A \cap (\Omega^+)^c} L d\mathbb{P} + \nu(A \cap \Omega^+) \\
&\leqslant \nu(A \cap (\Omega^+)^c) + \nu(A \cap \Omega^+) \\
&= \nu(A).
\end{aligned}$$

But this leads to a contradiction since

$$\kappa = \int_\Omega L d\mathbb{P} < \int_\Omega L' d\mathbb{P} = \int_\Omega L d\mathbb{P} + \varepsilon \mathbb{P}(\Omega^+) \leqslant \sup_{L'' \in \mathcal{M}} \int_\Omega L'' d\mathbb{P} = \kappa$$

so that a set $A_0 \in \mathcal{F}$ satisfying (7.2) does not exist.

The a.s. uniqueness of L is subject to Exercise 1. □

7.3 Conditional expectation

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, a random variable $f : \Omega \to \mathbb{R}$ that is $\mathcal{G}$-measurable is automatically $\mathcal{F}$-measurable. The converse is not true in general. Roughly speaking, one can say that an $\mathcal{F}$-measurable random variable is more complex or does contain more information than a $\mathcal{G}$-measurable random variable. There is a canonical way to project an $\mathcal{F}$-measurable random variable onto a subspace of $\mathcal{G}$-measurable random variables, that is called *conditional expectation*. The notion of conditional expectation plays a central role in the theory of stochastic processes and their applications. In particular, the notion of a martingale is derived from this. Martingales are used, for example, to describe fair games, for the definition of a stochastic integral, and to show why one can solve PDEs by Monte-Carlo methods.

Let us explain the conditional expectation in an easy - but typical - situation as an average procedure. Assume that the σ-algebra $\mathcal{G}$ is obtained by all possible combinations of the elements of a partition

$$\Omega = \bigcup_{i=1}^{n} \Omega_i$$

where the Ω_i are pair-wise disjoint and of positive measure. Given an random variable $f : \Omega \to \mathbb{R}$ such that $\int_\Omega |f| d\mathbb{P} < \infty$ we average f over the partition and obtain a new random variable

$$g(\omega) := \frac{1}{\mathbb{P}(\Omega_i)} \int_{\Omega_i} f d\mathbb{P} \quad \text{if} \quad \omega \in \Omega_i.$$

Because g is constant on the Ω_i's it is measurable with respect to $\mathcal{G}$. Moreover, one can easily see that one cannot distinguish between f and g if one integrates over sets from $\mathcal{G}$, i.e.

$$\int_B f d\mathbb{P} = \int_B g d\mathbb{P} \quad \text{for all} \quad B \in \mathcal{G}.$$

The random variable g will be called *conditional expectation of f given $\mathcal{G}$*. The following proposition guaranties the existence of a conditional expectation in the general case, i.e. when $\mathcal{G}$ is not generated by a finite or countable partition of Ω.

Proposition 7.3.1 (Conditional expectation)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra and $f \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$.*

(1) *There exists a $g \in \mathcal{L}_1(\Omega, \mathcal{G}, \mathbb{P})$ such that*

$$\int_B f d\mathbb{P} = \int_B g d\mathbb{P} \text{ for all } B \in \mathcal{G}.$$

(2) *If g and g' are as in* (i)*, then $\mathbb{P}(g \neq g') = 0$.*

Proof. Define

$$\nu(A) := \int_A f d\mathbb{P} \quad \text{for} \quad A \in \mathcal{G}$$

so that ν is a signed measure on $\mathcal{G}$. Applying the Theorem of Radon-Nikodym gives a $g \in \mathcal{L}_1(\Omega, \mathcal{G}, \mathbb{P})$ such that

$$\nu(A) = \int_A g d\mathbb{P} \quad \text{and} \quad \int_A g d\mathbb{P} = \nu(A) = \int_A f d\mathbb{P}.$$

Assume now another $g' \in \mathcal{L}_1(\Omega, \mathcal{G}, \mathbb{P})$ with

$$\int_A g d\mathbb{P} = \int_A g' d\mathbb{P}$$

for all $A \in \mathcal{G}$ and assume that $\mathbb{P}(g \neq g') > 0$. Hence we find a set $A \in \mathcal{G}$ with $\mathbb{P}(A) > 0$ and real numbers $\alpha < \beta$ such that

$$g(\omega) \leqslant \alpha < \beta \leqslant g'(\omega) \quad \text{for} \quad \omega \in A$$

or

$$g'(\omega) \leqslant \alpha < \beta \leqslant g(\omega) \quad \text{for} \quad \omega \in A.$$

Consequently (for example in the first case)

$$\int_A g d\mathbb{P} \leqslant \alpha \mathbb{P}(A) < \beta \mathbb{P}(A) \leqslant \int_A g' d\mathbb{P}$$

which is a contradiction. □

The main point of the theorem above is that g is $\mathcal{G}$-measurable. This leads to the following definition:

Definition 7.3.2 (Conditional expectation)**.** The $\mathcal{G}$-measurable and integrable random variable g from Proposition 7.3.1 is called *conditional expectation of f given $\mathcal{G}$* and is denoted by

$$g = \mathbb{E}[f \mid \mathcal{G}].$$

One has to keep in mind that the conditional expectation is only unique up to null sets from $\mathcal{G}$. We continue with some basic properties of conditional expectations.

Proposition 7.3.3. *Let $f, g \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ be sub-σ-algebras of $\mathcal{F}$. Then the following holds true:*

(1) Linearity*: If $\mu, \lambda \in \mathbb{R}$, then*

$$\mathbb{E}[\lambda f + \mu g \mid \mathcal{G}] = \lambda \mathbb{E}[f \mid \mathcal{G}] + \mu \mathbb{E}[g \mid \mathcal{G}] \text{ a.s.}$$

(2) Monotonicity*: If $f \leqq g$ a.s., then $\mathbb{E}[f \mid \mathcal{G}] \leqq \mathbb{E}[g \mid \mathcal{G}]$ a.s.*

(3) Positivity*: If $g \geqq 0$ a.s., then $\mathbb{E}[g \mid \mathcal{G}] \geqq 0$ a.s.*

(4) Convexity*: One has that $|\mathbb{E}[f \mid \mathcal{G}]| \leqq \mathbb{E}[|f| \mid \mathcal{G}]$ a.s.*

(5) Projection property I*: If f is $\mathcal{G}$-measurable, then $\mathbb{E}[f \mid \mathcal{G}] = f$ a.s.*

(6) Projection property II*:*

$$\mathbb{E}[\mathbb{E}[f \mid \mathcal{G}] \mid \mathcal{H}] = \mathbb{E}[\mathbb{E}[f \mid \mathcal{H}] \mid \mathcal{G}] = \mathbb{E}[f \mid \mathcal{H}] \text{ a.s.}$$

(7) *If $h : \Omega \to \mathbb{R}$ is $\mathcal{G}$-measurable and $fh \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$, then*

$$\mathbb{E}[hf \mid \mathcal{G}] = h\mathbb{E}[f \mid \mathcal{G}] \text{ a.s.}$$

(8) *If $\mathcal{G} = \{\emptyset, \Omega\}$, then $\mathbb{E}[f|\mathcal{G}] = \mathbb{E}f$ a.s.*

(9) *If for all $B \in \mathcal{B}(\mathbb{R})$ and all $A \in \mathcal{G}$, then one has that*

$$\mathbb{P}(\{f \in B\} \cap A) = \mathbb{P}(f \in B)\mathbb{P}(A),$$

i.e. if f is independent from $\mathcal{G}$, then $\mathbb{E}[f|\mathcal{G}] = \mathbb{E}f$ a.s.

(10) MONOTONE CONVERGENCE: *Assume $f \geqslant 0$ a.s. and random variables $0 \leqslant h_n \uparrow f$ a.s. Then*

$$\lim_n \mathbb{E}[h_n|\mathcal{G}] = \mathbb{E}[f|\mathcal{G}] \text{ a.s.}.$$

Proof. (1), (9) and (10) are exercises (see Exercise 2).

(2) Set $A := \{\omega \in \Omega : \mathbb{E}[f|\mathcal{G}](\omega) > \mathbb{E}[g|\mathcal{G}](\omega)\}$. The relation $f \leqslant g$ a.s. implies $\mathbb{E}[\mathbb{1}_A(f-g)] \leqslant 0$. Noticing that $A \in \mathcal{G}$, by the definition of conditional expectation we conclude

$$0 \geqslant \mathbb{E}\mathbb{1}_A f - \mathbb{E}\mathbb{1}_A g = \mathbb{E}\mathbb{1}_A\mathbb{E}[f|\mathcal{G}] - \mathbb{E}\mathbb{1}_A\mathbb{E}[g|\mathcal{G}] = \mathbb{E}\mathbb{1}_A\big(\mathbb{E}[f|\mathcal{G}] - \mathbb{E}[g|\mathcal{G}]\big).$$

Therefore $\mathbb{P}(A) = 0$.

(3) Apply (2) to $0 = f \leqslant g$. (4) The inequality $f \leqslant |f|$ gives $\mathbb{E}[f|\mathcal{G}] \leqq \mathbb{E}[|f||\mathcal{G}]$ a.s. and $-f \leqslant |f|$ gives $-\mathbb{E}[f|\mathcal{G}] \leqq \mathbb{E}[|f||\mathcal{G}]$ a.s., so that we are done. (5) follows directly from the definition.

(6) Since $\mathbb{E}[f|\mathcal{H}]$ is $\mathcal{H}$-measurable and hence $\mathcal{G}$-measurable, item (iv) implies that $\mathbb{E}[\mathbb{E}[f|\mathcal{H}]|\mathcal{G}] = \mathbb{E}[f|\mathcal{H}]$ a.s. so that one equality is shown. For the other equality we have to show that

$$\int_A \mathbb{E}[\mathbb{E}[f|\mathcal{G}]|\mathcal{H}]d\mathbb{P} = \int_A f d\mathbb{P}$$

for $A \in \mathcal{H}$. Letting $h := \mathbb{E}[f|\mathcal{G}]$, this follows from

$$\int_A \mathbb{E}[\mathbb{E}[f|\mathcal{G}]|\mathcal{H}]d\mathbb{P} = \int_A \mathbb{E}[h|\mathcal{H}]d\mathbb{P} = \int_A h d\mathbb{P} = \int_A \mathbb{E}[f|\mathcal{G}]d\mathbb{P} = \int_A f d\mathbb{P}$$

since $A \in \mathcal{H} \subseteq \mathcal{G}$.

(7) Assume first that $h = \sum_{n=1}^N \alpha_n \mathbb{1}_{A_n}$, where $\bigcup_{n=1}^N A_n = \Omega$ is a partition with $A_n \in \mathcal{G}$. For $A \in \mathcal{G}$ we get, a.s., that

$$\begin{aligned}\int_A h f d\mathbb{P} &= \sum_{n=1}^N \alpha_n \int_A \mathbb{1}_{A_n} f d\mathbb{P} = \sum_{n=1}^N \alpha_n \int_{A\cap A_n} f d\mathbb{P} \\ &= \sum_{n=1}^N \alpha_n \int_{A\cap A_n} \mathbb{E}[f|\mathcal{G}]d\mathbb{P} = \int_A \left(\sum_{n=1}^N \alpha_n \mathbb{1}_{A_n}\right) \mathbb{E}[f|\mathcal{G}]d\mathbb{P}\end{aligned}$$

$$= \int_A h\mathbb{E}(f|\mathcal{G})d\mathbb{P}.$$

Hence $\mathbb{E}[hf|\mathcal{G}] = h\mathbb{E}[f|\mathcal{G}]$ a.s. For the general case we can assume that $f, h \geqslant 0$ since we can decompose $f = f^+ - f^-$ and $h = h^+ - h^-$ with $f^+ := \max\{f, 0\}$ and $f^- := \min\{-f, 0\}$ (and in the same way we proceed with h). We find simple functions $0 \leqslant h_n \leqslant h$ such that $h_n(\omega) \uparrow h(\omega)$. Then, by our fist step, we get that

$$h_n\mathbb{E}[f|\mathcal{G}] = \mathbb{E}[h_n f|\mathcal{G}] \text{ a.s.}$$

By $n \to \infty$ the left-hand side follows. The right-hand side is a consequence of the monotone convergence given in (10).

(8) Clearly, we have that

$$\int_\varnothing f d\mathbb{P} = 0 = \int_\varnothing (\mathbb{E}f) d\mathbb{P} \quad \text{and} \quad \int_\Omega f d\mathbb{P} = \mathbb{E}f = \int_\Omega (\mathbb{E}f) d\mathbb{P}$$

so that (8) follows. □

7.4 A representation of conditional expectations

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f : \Omega \to \mathbb{R}$ be an integrable random variable. Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra obtained by a measurable map $g : \Omega \to E$, where $(E, \mathcal{E})$ is a measurable space, i.e. $\mathcal{G} = \sigma(g) = \{g^{-1}(B) : B \in \mathcal{E}\}$. Let $\mathbb{P}_g$ be the image measure of $\mathbb{P}$ with respect to g. In the following we want to give a proper sense to the intuitive equality

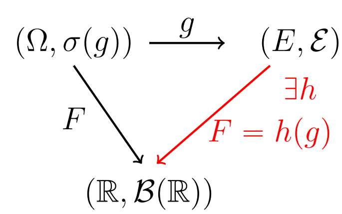
$$\int_{\{g \in B\}} f d\mathbb{P} = \int_B \mathbb{E}[f|g = y] d\mathbb{P}_g(y)$$

that represents our conditional expectation on the left-hand side by computing averages over the sets $\{g = y\} \in \mathcal{E}$ on the right-hand side. The problem consists in the fact that the sets $\{g = y\}$ might be of $\mathbb{P}_g$-measure zero. We start with a factorisation lemma, that is of independent interest.

Lemma 7.4.1 (Factorization Lemma, [1, Section II.11.7] or [12]). *Assume $\Omega \neq \emptyset$, a measurable space $(E, \mathcal{E})$ and maps $g : \Omega \to E$ and $F : \Omega \to \mathbb{R}$. Let $\sigma(g)$ denote the sigma-algebra generated by g. Then the following assertions are equivalent:*

(1) *The map F is $(\Omega, \sigma(g)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable.*

(2) *There exists a measurable $h : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $F = h \circ g$.*



Proof. (2) $\Longrightarrow$ (1) We have

$$g : (\Omega, \sigma(g)) \to (E, \mathcal{E}) \quad \text{and} \quad h : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})),$$

so that for the composition $F = h \circ g$ it holds that $F : (\Omega, \sigma(g)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
(1) $\Longrightarrow$ (2) We first assume that

$$F = \sum_{i=1}^{n} \alpha_i \mathbb{I}_{A_i} \text{ with } \alpha_i \in \mathbb{R}, \alpha_i \geqslant 0 \text{ and } A_i \in \sigma(g) = \{g^{-1}(B) : B \in \mathcal{E}\}.$$

We put $h := \sum_{i=1}^{n} \alpha_i \mathbb{I}_{B_i}$ where $B_i \in \mathcal{E}$ is such that $A_i = g^{-1}(B_i)$. Then it holds $h : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and

$$F = \sum_{i=1}^{n} \alpha_i \mathbb{I}_{A_i} = \sum_{i=1}^{n} \alpha_i \mathbb{I}_{g^{-1}(B_i)} = \sum_{i=1}^{n} \alpha_i \mathbb{I}_{B_i}(g(\cdot)) = h \circ g.$$

If $F \geqslant 0$, then we can approximate it by measurable step functions F_n such that $F_n(\omega) \uparrow F(\omega)$ for all $\omega \in \Omega$. For example, one can use $F_n := \sum_{k=0}^{4^n} \frac{k}{2^n} \mathbb{I}_{F^{-1}((k/2^n, (k+1)/2^n])}$. For each n we construct a function $h_n : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with $F_n = h_n \circ g$ as above. Then also $h := \sup_n h_n : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and it holds that

$$h \circ g = (\sup_n h_n) \circ g = \sup_n (h_n \circ g) = \sup_n F_n = F.$$

For a general F we write $F = F^+ - F^-$ with $F^+ := \max\{F, 0\}$ and $F^- := \max\{-F, 0\}$. We construct $h^+, h^- : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with $F^+ = h^+ \circ g$ and $F^{=} h^- \circ g$ as before. $\square$

Lemma 7.4.2. *Let $f : \Omega \to \mathbb{R}$ be an integrable random variable and $g : \Omega \to E$ be measurable. Consider a version F of $\mathbb{E}[f|\sigma(g)]$ that is $\sigma(g)$ measurable and a factorization $F = h \circ g$ according to Lemma 7.4.1. If F' is another version with a corresponding factorization $F' = h' \circ g$, then*

$$\mathbb{P}_g(h = h') = 1,$$

where $\mathbb{P}_g$ is the image measure of $\mathbb{P}$ with respect to g.

Proof. For $B \in \mathcal{E}$ we obtain by a change of variables that

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_g} \mathbb{1}_B (h - h') &= \int_B [h(y) - h'(y)] d\mathbb{P}_g(y) \\ &= \int_\Omega \mathbb{1}_B(g(\omega))[h(g(\omega)) - h'(g(\omega))] d\mathbb{P}(\omega) = 0 \end{aligned}$$

because $\mathbb{1}_B(g)$ is $\sigma(g)$-measurable. Testing with $B = \{y \in E : h(y) > h'(y)\}$ and $B = \{y \in E : h(y) < h'(y)\}$ leads to $h = h'$ $\mathbb{P}_g$-a.s. □

Now we can define the conditional probability we are interested in:

Definition 7.4.3. We let

$$\mathbb{E}[f|g = y] := h(y) \quad \text{and} \quad \mathbb{P}[A|g = y] := \mathbb{E}[\mathbb{1}_A | g = y].$$

Proposition 7.4.4. *For a set $B \in \mathcal{E}$ it holds that*

$$\int_{\{g \in B\}} f d\mathbb{P} = \int_B \mathbb{E}[f|g = y] d\mathbb{P}_g(y).$$

In particular, for $f = \mathbb{1}_A$ we get that

$$\mathbb{P}(A \cap \{g \in B\}) = \int_B \mathbb{E}[\mathbb{1}_A | g = y] d\mathbb{P}_g(y).$$

Proof. Because $\{g \in B\} \in \sigma(g)$ and the definition of $\mathbb{E}[f|g = y]$ we have that

$$\begin{aligned} \int_{\{g \in B\}} f d\mathbb{P} &= \int_{\{g \in B\}} \mathbb{E}[f|\sigma(g)] d\mathbb{P} \\ &= \int_{\{g \in B\}} h \circ g d\mathbb{P} \end{aligned}$$

$$
\begin{aligned}
&= \int_B h(y) d\mathbb{P}_g(y) \\
&= \int_B \mathbb{E}[f|g=y] d\mathbb{P}_g(y).
\end{aligned}
$$

□

With respect to densities we obtain the following representation:

Proposition 7.4.5. *Let $f : \Omega \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}^d$ be random variables whose joint distribution has a density $h_{(f,g)} \in \mathcal{L}_1(\mathbb{R}^{1+d}, \mathcal{B}(\mathbb{R}^{1+d}), \lambda_{1+d})$, i.e. for $A \in \mathcal{B}(\mathbb{R})$ and $B \in \mathcal{B}(\mathbb{R}^d)$ one has that*

$$
\mathbb{P}(\{\omega \in \Omega : (f(\omega), g(\omega)) \in A \times B\}) = \int_{A\times B} h_{(f,g)}(x,y) d\lambda_{d+1}(x,y).
$$

Then the following assertion hold:

(1) *f and g have densities h_f and h_g, respectively.*

(2) *If*

$$
h_{f|g}(x|y) := \begin{cases} \frac{h_{(f,g)}(x,y)}{h_g(y)} & \text{if } h_g(y) \neq 0, \\ 0 & \text{otherwise.} \end{cases},
$$

then $\mathbb{P}[f \in A|g=y] = \int_A h_{f|g}(x|y)\, d\lambda(x)$ for all $A \in \mathcal{B}(\mathbb{R})$ and $\mathbb{P}_g$-a.a. $y \in \mathbb{R}^d$.

Proof. (1) The densities one can compute by integrating over the remaining coordinate, i.e.

$$
\begin{aligned}
h_f(x) &:= \int_{\mathbb{R}^d} h_{(f,g)}(x,y) d\lambda_d(y), \\
h_g(y) &:= \int_{\mathbb{R}} h_{(f,g)}(x,y) d\lambda(x).
\end{aligned}
$$

(ii) First we observe that

$$
\begin{aligned}
\int_B \int_A h_{f|g}(x|y)\, d\lambda(x) d\mathbb{P}_g(y) &= \int_B \int_A h_{f|g}(x|y)\, d\lambda(x) h_g(y) d\lambda_d(y) \\
&= \int_B \int_A h_{f|g}(x|y)\, d\lambda(x) \mathbb{1}_{\{h_g(y)>0\}} h_g(y) d\lambda_d(y)
\end{aligned}
$$

$$
\begin{aligned}
&= \int_B \int_A h_{(f,g)}(x,y)\, d\lambda(x) \mathbb{1}_{\{h_g(y)>0\}} d\lambda_d(y) \\
&= \int_B \int_A h_{(f,g)}(x,y)\, d\lambda(x) d\lambda_d(y) \\
&= \mathbb{P}(\{f \in A\} \cap \{g \in B\}).
\end{aligned}
$$

On the other hand, we have by Proposition 7.4.4 that

$$\mathbb{P}(\{f \in A\} \cap \{g \in B\}) = \int_B \mathbb{P}[f \in A | g = y] d\mathbb{P}_g(y),$$

so that

$$\int_B \left[\int_A h_{f|g}(x|y)\, d\lambda(x) \right] d\mathbb{P}_g(y) = \int_B \mathbb{P}[f \in A | g = y] d\mathbb{P}_g(y)$$

for all $B \in \mathcal{B}(\mathbb{R}^d)$ which implies the assertion. □

Another form of a representation of a conditional expectation is the following:

Proposition 7.4.6 (cf. [15, Appendix]). *Let $f, g : \Omega \to \mathbb{R}$ be random variables and $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ algebra, where we assume that f is independent from $\mathcal{G}$ and g is $\mathcal{G}$-measurable. If φ is a Borel function on $\mathbb{R}^2$ such that $\mathbb{E}|\varphi(f,g)| < \infty$, then*

$$\mathbb{E}[\varphi(f,g)|\mathcal{G}] = (\mathbb{E}\varphi(f,y))\big|_{y=g}.$$

Proof. We have to check that

$$\int_G \varphi(f,g) d\mathbb{P} = \int_G (\mathbb{E}\varphi(f,y))\big|_{y=g} d\mathbb{P}$$

for all $G \in \mathcal{G}$. To verify this equation, we start with a change of variables to get for $G \in \mathcal{G}$ that

$$\mathbb{E}\big(\varphi(f,g)\mathbb{1}_G\big) = \int_{\mathbb{R}^3} \varphi(x,y) z d\mathbb{P}_{(f,g,\mathbf{1}_G)}(x,y,z).$$

Since g is $\mathcal{G}$-measurable and f is independent from $\mathcal{G}$ we get that f and $(g, \mathbb{1}_G)$ are independent, which means $\mathbb{P}_{(f,g,\mathbf{1}_G)} = \mathbb{P}_f \otimes \mathbb{P}_{(g,\mathbf{1}_G)}$. By Fubini's Theorem and another change of variables,

$$\int_{\mathbb{R}^3} \varphi(x,y) z d\mathbb{P}_{(f,g,\mathbf{1}_G)}(x,y,z) \quad = \quad \int_{\mathbb{R}^2} \left[\int_{\mathbb{R}} \varphi(x,y) d\mathbb{P}_f(x) \right] z d\mathbb{P}_{(g,\mathbf{1}_G)}(y,z)$$

$$
\begin{aligned}
&= \int_{\mathbb{R}^2} [\mathbb{E}\varphi(f,y)] z \, d\mathbb{P}_{(g,\mathbb{1}_G)}(y,z) \\
&= \mathbb{E}\big[(\mathbb{E}\varphi(f,y)]) \big|_{y=g} \mathbb{1}_G \big]
\end{aligned}
$$

which implies the assertion. □

7.5 Exercises

Ex 1: Prove the the map L from the RADON-NIKODYM theoerem 7.2.1 is unique.

Ex 2: Prove items (1), (9) and (10) of Proposition 7.3.3.

Chapter 8

Sums of independent random variables and strong laws of large numbers

In this chapter we systematically investigate sums of independent random variables and will deduce two versions of *Strong Laws of Large Numbers*. In order to prove these theorems we shall *not* use the method of *Characteristic Functions* (introduced in Chapter 9 below). This is in contrast to other fundamental limit theorems, the *Law of Iterated Logarithm* and the *Central Limit Theorem*, which we shall verify by the help of characteristic functions.

8.1 Sums of independent random variables

If $\varepsilon_1, \varepsilon_2, \ldots$ are independent BERNOULLI random variables, that means $\mathbb{P}(\varepsilon_n = 1) = \mathbb{P}(\varepsilon_i = -1) = \frac{1}{2}$, then we know from Corollary 4.2.8 that

$$\mathbb{P}\left(\sum_{n=1}^{\infty} \frac{\varepsilon_n}{n} \text{ converges}\right) \in \{0, 1\}.$$

But, do we get probability zero or one? For this there is a beautiful complete answer in the Two- and Three-Series-Theorems of KOLMOGOROV which are the main subject of this section. Let us start with the more special Two-Series-Theorem:

Proposition 8.1.1 (Two-Series-Theorem of Kolmogorov). *Assume a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and random variables* $\xi_1, \xi_2, ... : \Omega \to \mathbb{R}$ *which are assumed to be* independent *and such that* $\mathbb{E}\xi_i^2 < \infty$.

(1) *If* $\sum_{n=1}^{\infty} \mathbb{E}\xi_n$ *converges and* $\sum_{n=1}^{\infty} \mathbb{E}\left(\xi_n - \mathbb{E}\xi_n\right)^2 < \infty$, *then*

$$\mathbb{P}\left(\sum_{n=1}^{\infty} \xi_n \text{ converges}\right) = 1.$$

(2) *Let* $|\xi_n(\omega)| \leqslant d$ *for all* $n = 1, 2, ...$ *and all* $\omega \in \Omega$, *where* $d > 0$ *is some constant. Then the following assertions are equivalent:*

   (a) $\mathbb{P}\left(\sum_{n=1}^{\infty} \xi_n \text{ converges}\right) = 1$.

   (b) $\sum_{n=1}^{\infty} \mathbb{E}\xi_n$ *converges and* $\sum_{n=1}^{\infty} \mathbb{E}\left(\xi_n - \mathbb{E}\xi_n\right)^2 < \infty$.

To prove the Two-Series-Theorem, we need a deviation inequality due to Kolmogorov. Before we state and prove this inequality, let us motivate it. Assume a random variable $f : \Omega \to \mathbb{R}$ with $\mathbb{E}f^2 < \infty$. Then, by Chebyshev's inequality

$$\varepsilon^2\, \mathbb{P}\left(|f| \geqslant \varepsilon\right) = \varepsilon^2\, \mathbb{P}\left(|f|^2 \geqslant \varepsilon^2\right) \leqslant \int_{\Omega} |f|^2 \, d\,\mathbb{P}$$

for $\varepsilon > 0$ so that

$$\mathbb{P}\left(|f| \geqslant \varepsilon\right) \leqslant \frac{\mathbb{E}\,|f|^2}{\varepsilon^2}.$$

Assuming independent random variables $\xi_1, \xi_2, ... : \Omega \to \mathbb{R}$ and $S_n := \xi_1 + \cdots + \xi_n$ this gives that

$$\mathbb{P}\left(|S_n| \geqslant \varepsilon\right) \leqslant \frac{\mathbb{E}|S_n|^2}{\varepsilon^2}.$$

We can enlarge the left-hand side by replacing $|S_n|$ by $\sup_{k=1,...,n} |S_k|$ so that we get a *maximal inequality*:

Lemma 8.1.2 (Inequalities of Kolmogorov). *Let* $\xi_1, \ldots, \xi_n : \Omega \to \mathbb{R}$ *be independent with* $\mathbb{E}\xi_n^2 < \infty$ *and* $\mathbb{E}\xi_n = 0$. *Then, for* $S_k := \xi_1 + \cdots + \xi_k$, $S_n^* := \max_{1 \leqslant k \leqslant n} |S_k|$, *and* $\varepsilon > 0$,

$$\frac{\mathbb{E}S_n^2 - \varepsilon^2}{\mathbb{E}S_n^2 + 2c\varepsilon + c^2} \leqslant \mathbb{P}\left(S_n^* \geqslant \varepsilon\right) \leqslant \frac{1}{\varepsilon^2} \int_{\{S_n^* \geqslant \varepsilon\}} S_n^2 d\mathbb{P}$$

where $c \in [0, \infty]$ *is such that* $|\xi(\omega)| \leqslant c$ *for all* $k = 1, \ldots, n$ *and* $\omega \in \Omega$.

Proof. The proof is taken from [17]. Let $A := \{S_n^* \geqslant \varepsilon\}$ and $A_k := \{|S_1| < \varepsilon, \ldots, |S_{k-1}| < \varepsilon, |S_k| \geqslant \varepsilon\}$. Then

$$\begin{aligned}
\int_A S_n^2 d\mathbb{P} &= \sum_{k=1}^n \int_{A_k} S_n^2 d\mathbb{P} \\
&= \sum_{k=1}^n \left[\int_{A_k} S_k^2 d\mathbb{P} + 2\int_{A_k} S_k \left(\sum_{i=k+1}^n \xi_i \right) d\mathbb{P} + \int_{A_k} \left(\sum_{i=k+1}^n \xi_i \right)^2 d\mathbb{P} \right] \\
&= \sum_{k=1}^n \left[\int_{A_k} S_k^2 d\mathbb{P} + \int_{A_k} \left(\sum_{i=k+1}^n \xi_i \right)^2 d\mathbb{P} \right].
\end{aligned}$$

Now the upper bound is obtained by

$$\varepsilon^2 \mathbb{P}\left(S_n^* \geqslant \varepsilon\right) = \varepsilon^2 \sum_{k=1}^n \mathbb{P}\left(A_k\right) \leqslant \sum_{k=1}^n \int_{A_k} S_k^2 d\mathbb{P} \leqslant \int_A S_n^2 d\mathbb{P}.$$

For the lower bound only the case $c < \infty$ has to be considered. Here we get

$$\begin{aligned}
\mathbb{E}S_n^2 - \varepsilon^2 \mathbb{P}(A^c) &\leqslant \mathbb{E}S_n^2 - \int_{A^c} S_n^2 d\mathbb{P} = \int_A S_n^2 d\mathbb{P} \\
&= \sum_{k=1}^n \left[\int_{A_k} S_k^2 d\mathbb{P} + \int_{A_k} \left(\sum_{i=k+1}^n \xi_i \right)^2 d\mathbb{P} \right] \\
&\leqslant (\varepsilon + c)^2 \sum_{k=1}^n \mathbb{P}(A_k) + \sum_{k=1}^n \mathbb{P}(A_k) \sum_{j=k+1}^n \mathbb{E}\xi_j^2 \\
&\leqslant \mathbb{P}(A) \left[(\varepsilon + c)^2 + \sum_{j=1}^n \mathbb{E}\xi_j^2 \right] \\
&= \mathbb{P}(A)[(\varepsilon + c)^2 + \mathbb{E}S_n^2].
\end{aligned}$$

This implies

$$\mathbb{P}(A) \geqslant \frac{\mathbb{E}S_n^2 - \varepsilon^2}{\mathbb{E}S_n^2 - \varepsilon^2 + (\varepsilon + c)^2}. \qquad \square$$

Example 8.1.3. For BERNOULLI variables $(\varepsilon_n)_{n=1}^\infty$ it follows from Lemma 8.1.2 that

$$\mathbb{P}\left(\max_{1\leqslant k\leqslant n}|\varepsilon_1+\cdots+\varepsilon_k|\geqslant\varepsilon\right)\leqslant\frac{\mathbb{E}\,f_n^2}{\varepsilon^2}=\frac{n}{\varepsilon^2}.$$

Letting $\varepsilon = \theta\sqrt{n}$, $\theta>0$, this gives

$$\mathbb{P}\left(\max_{1\leqslant k\leqslant n}|\varepsilon_1+\cdots+\varepsilon_k|\geqslant\theta\sqrt{n}\right)\leqslant\frac{1}{\theta^2}.$$

The left-hand side describes the probability that the random walk exceeds $-\theta\sqrt{n}$ or $\theta\sqrt{n}$ *up* to step n (not *only* at step n).

Proof of Proposition 8.1.1. (1) We let $\eta_n := \xi_n - \mathbb{E}\xi_n$ so that $\mathbb{E}\eta_n = 0$ and

$$\sum_{n=1}^\infty \mathbb{E}\eta_n^2 = \sum_{n=1}^\infty \mathbb{E}(\xi_n-\mathbb{E}\xi_n)^2<\infty.$$

Since $\sum_{n=1}^\infty\mathbb{E}\xi_n$ converges, the convergence of $\sum_{n=1}^\infty\eta_n(\omega)$ implies the convergence of $\sum_{n=1}^\infty\xi_n(\omega)=\sum_{n=1}^\infty(\eta_n(\omega)+\mathbb{E}\xi_n)$. Hence it is sufficient to show that

$$\mathbb{P}\left(\sum_{n=1}^\infty\eta_n \text{ converges}\right)=1.$$

Let $g_n := \eta_1+\cdots+\eta_n$. Applying Proposition 6.1.4 we see that it is enough to prove

$$\lim_{n\to\infty}\mathbb{P}\left(\sup_{k\geqslant n}|g_k-g_n|\geqslant\varepsilon\right)=0$$

for all $\varepsilon>0$. But this follows from

$$\begin{aligned}\mathbb{P}\left(\sup_{k\geqslant n}|g_k-g_n|\geqslant\varepsilon\right) &= \lim_{N\to\infty}\mathbb{P}\left(\sup_{n\leqslant k\leqslant n+N}|g_k-g_n|\geqslant\varepsilon\right)\\ &\leqslant \lim_{N\to\infty}\frac{\mathbb{E}(g_{n+N}-g_n)^2}{\varepsilon^2}\\ &= \lim_{N\to\infty}\frac{\sum_{k=1}^N\mathbb{E}\eta_{n+k}^2}{\varepsilon^2}\\ &\leqslant \frac{\sum_{l=n+1}^\infty\mathbb{E}\eta_l^2}{\varepsilon^2}\end{aligned}$$

where we have used the KOLMOGOROV inequality Lemma 8.1.2. Obviously, the last term converges to zero as $n \to \infty$.

(2) Because of step (1) we only have to prove that $(a) \Longrightarrow (b)$. We use again a symmetrization argument and consider a new sequence $\xi_1', \xi_2', ... : \Omega' \to \mathbb{R}$ of independent random variables on $(\Omega', \mathcal{F}', \mathbb{P}')$ having the same distribution as the original sequence $\xi_1, \xi_2, ...$, that means

$$\mathbb{P}(\xi_n \leqslant \lambda) = \mathbb{P}'(\xi_n' \leqslant \lambda)$$

for all $n = 1, 2, ...$ and all $\lambda \in \mathbb{R}$. We also may assume that $|\xi_n'(\omega')| \leqslant d$ for all $\omega' \in \Omega'$ and $n = 1, 2, ...$ Taking the product space $(M, \Sigma, \mu) = (\Omega, \mathcal{F}, \mathbb{P}) \times (\Omega', \mathcal{F}', \mathbb{P}')$ we may consider $\xi_n, \xi_n' : M \to \mathbb{R}$ with the convention that $\xi_n(\omega, \omega') = \xi_n(\omega)$ and $\xi_n'(\omega, \omega') = \xi_n'(\omega')$. Now we let

$$\eta_n(\omega, \omega') = \xi_n(\omega, \omega') - \xi_n'(\omega, \omega')$$

and get

$$\begin{aligned} \mathbb{E}_\mu \eta_n &= \mathbb{E}\xi_n - \mathbb{E}\xi_n' = 0, \\ |\eta_n(\omega, \omega')| &\leqslant |\xi_n(\omega, \omega')| + |\xi_n'(\omega, \omega')| \leqslant 2d, \end{aligned}$$

and

$$\begin{aligned} & \mu\left(\left\{(\omega, \omega') \in M : \sum_{n=1}^{\infty} \eta_n(\omega, \omega') \text{ converges }\right\}\right) \\ = \;& \mathbb{P} \times \mathbb{P}'\left(\left\{(\omega, \omega') \in \Omega \times \Omega' : \sum_{n=1}^{\infty} (\xi_n(\omega) - \xi_n'(\omega')) \text{ converges }\right\}\right) \\ = \;& 1. \end{aligned}$$

Letting $g_n := \eta_1 + \cdots + \eta_n$ and $\varepsilon > 0$, Proposition 6.1.4 implies that there is an $n \in \{1, 2, ...\}$ such that

$$\mu\left(\sup_{k \geqslant n} |g_k - g_n| \geqslant \varepsilon\right) < \frac{1}{2}.$$

Exploiting Lemma 8.1.2 gives, for $N \geqslant 1$, that

$$1 - \frac{(2d+\varepsilon)^2}{\sum_{k=n+1}^{n+N} \mathbb{E}\eta_k^2} = 1 - \frac{(2d+\varepsilon)^2}{\mathbb{E}(g_{n+N} - g_n)^2} \leqslant \mu\left(\sup_{k=n,...,n+N} |g_k - g_n| \geqslant \varepsilon\right) < \frac{1}{2}.$$

But from this it follows that

$$\sum_{k=n+1}^{n+N} \mathbb{E}\eta_k^2 < 2(2d+\varepsilon)^2 < \infty$$

for all N, so that

$$\sum_{k=1}^{\infty} \mathbb{E}\eta_k^2 < \infty.$$

This gives that

$$\sum_{n=1}^{\infty} \mathbb{E}(\xi_n - \mathbb{E}\xi_n)^2 = \frac{1}{2}\sum_{n=1}^{\infty} \mathbb{E}([\xi_n - \mathbb{E}\xi_n] - [\xi_n' - \mathbb{E}\xi_n'])^2 = \sum_{n=1}^{\infty} \mathbb{E}\eta_n^2 < \infty.$$

It remains to show that $\sum_{n=1}^{\infty} \mathbb{E}\xi_n$ exists. Since $\sum_{n=1}^{\infty} \xi_n$ converges almost surely and $\sum_{n=1}^{\infty}(\xi_n - \mathbb{E}\xi_n)$ converges almost surely because of step (i) and $\sum_{n=1}^{\infty} \mathbb{E}(\xi_n - \mathbb{E}\xi_n)^2 < \infty$ proved right now, we have to have that $\sum_{n=1}^{\infty} \mathbb{E}\xi_n$ converges as well. □

As we saw, the Two-Series-Theorem only gives an equivalence in the case that $\sup_{n,\omega} |\xi_n(\omega)| < \infty$. For the general case we have an equivalence as well, the Three-Series-Theorem, which one can quickly deduce from the Two-Series-Theorem. For its formulation we introduce for a random variable $f : \Omega \to \mathbb{R}$ and some constant $c > 0$ the *truncated* random variable

$$f^c(\omega) := \begin{cases} f(\omega) & : \ |f(\omega)| \leqslant c \\ c & : \ f(\omega) > c \\ -c & : \ f(\omega) < -c \end{cases}.$$

Corollary 8.1.4 (Three-Series-Theorem of Kolmogorov)**.** *Assume a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and* independent *random variables $\xi_1, \xi_2, ... : \Omega \to \mathbb{R}$. Then the following conditions are equivalent:*

(1) $\mathbb{P}\left(\sum_{n=1}^{\infty} \xi_n \text{ converges }\right) = 1.$

(2) *For* all *constants $c > 0$ the following three conditions are satisfied:*

   (a) $\sum_{n=1}^{\infty} \mathbb{E}\xi_n^c$ *converges,*

(b) $\sum_{n=1}^{\infty} \mathbb{E}\left(\xi_n^c - \mathbb{E}\xi_n^c\right)^2 < \infty$,

(c) $\sum_{n=1}^{\infty} \mathbb{P}\left(|\xi_n| \geqslant c\right) < \infty$.

(3) *There* exists *one constant* $c > 0$ *such that conditions* (a), (b), *and* (c) *of item* (ii) *are satisfied.*

Proof. (3) $\Longrightarrow$ (1). The Two-Series-Theorem implies that

$$\mathbb{P}\left(\sum_{n=1}^{\infty} \xi_n^c \text{ converges }\right) = 1.$$

Consider $\eta_n := \xi_n - \xi_n^c$ and $B_n := \{\omega \in \Omega : \eta_n(\omega) \neq 0\}$. Then

$$\sum_{n=1}^{\infty} \mathbb{P}(B_n) = \sum_{n=1}^{\infty} \mathbb{P}\left(|\xi_n| > c\right) < \infty.$$

The lemma of Borel-Cantelli implies that

$$\mathbb{P}\left(\{\omega \in \Omega : \#\{n : \omega \in B_n\} = \infty\}\right) = 0.$$

This implies that $\mathbb{P}\left(\sum_{n=1}^{\infty} \eta_n \text{ converges }\right) = 1$ so that

$$\mathbb{P}\left(\sum_{n=1}^{\infty} \xi_n \text{ converges }\right) = \mathbb{P}\left(\sum_{n=1}^{\infty} [\eta_n + \xi_n^c] \text{ converges }\right) = 1.$$

(2) $\Longrightarrow$ (3) is trivial.
(1) $\Longrightarrow$ (2) The almost sure convergence of $\sum_{n=1}^{\infty} \xi_n(\omega)$ implies that $\lim_n |\xi_n(\omega)| = 0$ a.s. so that

$$\mathbb{P}\left(\limsup_n \{\omega \in \Omega : |\xi_n(\omega)| \geqslant c\}\right) = 0.$$

The Lemma of Borel-Cantelli (note that the random variables ξ_n are independent) gives that

$$\sum_{n=1}^{\infty} \mathbb{P}\left(|\xi_n| \geqslant c\right) < \infty$$

so that we obtain condition (c). Next, the almost sure convergence of $\sum_{n=1}^{\infty} \xi_n(\omega)$ implies the almost sure convergence of $\sum_{n=1}^{\infty} \xi_n^c(\omega)$ since $\xi_n(\omega) = \xi_n^c(\omega)$ for $n \geqslant n(\omega)$ for almost all $\omega \in \Omega$. Hence we can apply the Two-Series-Theorem Proposition 8.1.1 to obtain items (a) and (b). □

Now we consider an example.

Example 8.1.5. Let $\varepsilon_1, \varepsilon_2, ... : \Omega \to \mathbb{R}$ be independent Bernoulli random variables, $\alpha_1, \alpha_2, ... \in \mathbb{R}$ and $\beta_1, \beta_2, ... \in \mathbb{R}$. Then

$$\mathbb{P}\left(\sum_{n=1}^{\infty}[\alpha_n + \beta_n\varepsilon_n] \text{ converges}\right) = 1 \tag{8.1}$$

if and only if $\sum_{n=1}^{\infty}\alpha_n$ converges and $\sum_{n=1}^{\infty}\beta_n^2 < \infty$.

Proof. (a) Assuming the conditions on $(\alpha_n)_{n=1}^{\infty}$ and $(\beta_n)_{n=1}^{\infty}$ we get, for $\xi_n := \alpha_n + \beta_n\varepsilon_n$, that

$$\sum_{n=1}^{\infty}\mathbb{E}\xi_n = \sum_{n=1}^{\infty}\alpha_n \quad \text{and} \quad \sum_{n=1}^{\infty}\mathbb{E}[\xi_n - \mathbb{E}\xi_n]^2 = \sum_{n=1}^{\infty}\beta_n^2 < \infty.$$

The Two-Series-Theorem gives the almost sure convergence of $\sum_{n=1}^{\infty}[\alpha_n + \beta_n\varepsilon_n]$.

(b) Assuming (8.1), we define $\eta_n(\omega) := -\varepsilon_n(\omega)$ and obtain independent Bernoulli random variables. Since (8.1) is an assumption on the *distribution* of $(\varepsilon_1, \varepsilon_2, ...)$, which is the same as that of $(\eta_1, \eta_2, ...)$, we get (8.1) for $(\eta_n)_{n=1}^{\infty}$ as well. Hence there are $\Omega_\varepsilon, \Omega_\eta \in \mathcal{F}$ of measure one such that

$$\sum_{n=1}^{\infty}[\alpha_n + \beta_n\varepsilon_n(\omega)] \quad \text{and} \quad \sum_{n=1}^{\infty}[\alpha_n + \beta_n\eta_n(\omega')]$$

converge for $\omega \in \Omega_\varepsilon$ and $\omega' \in \Omega_\eta$. Since $\Omega_\varepsilon \cap \Omega_\eta$ is of measure one, there exists at least one $\omega_0 \in \Omega_\varepsilon \cap \Omega_\eta$ such that

$$\sum_{n=1}^{\infty}[\alpha_n + \beta_n\varepsilon_n(\omega_0)] \quad \text{and} \quad \sum_{n=1}^{\infty}[\alpha_n + \beta_n\eta_n(\omega_0)]$$

converge. Taking the sum, we get that

$$\sum_{n=1}^{\infty}[\alpha_n + \beta_n\varepsilon_n(\omega_0)] + \sum_{n=1}^{\infty}[\alpha_n + \beta_n\eta_n(\omega_0)] = \sum_{n=1}^{\infty}[\alpha_n + \beta_n\varepsilon_n(\omega_0)] + \sum_{n=1}^{\infty}[\alpha_n + \beta_n(-\varepsilon_n(\omega_0))] = 2\sum_{n=1}^{\infty}\alpha_n$$

converges. From that we can deduce in turn that

$$\mathbb{P}\left(\sum_{n=1}^{\infty} \beta_n \varepsilon_n \text{ converges}\right) = 1.$$

Picking again an $\omega_0' \in \Omega$ such that $\sum_{n=1}^{\infty} \beta_n \varepsilon_n(\omega_0')$ converges, we get that $\sup_n |\beta_n| = \sup_n |\beta_n||\varepsilon_n(\omega_0)| < \infty$. Hence, by Proposition 8.1.1,

$$\sum_{n=1}^{\infty} \beta_n^2 = \sum_{n=1}^{\infty} \mathbb{E}|\beta_n \varepsilon_n - \mathbb{E}\beta_n \varepsilon_n|^2 < \infty. \qquad \square$$

The following examples demonstrate the basic usage of the Three-Series-Theorem: it shows that not the size of the large values of the random variables is sometimes important, but only their probability.

Example 8.1.6. Assume independent random variables $\xi_1, \xi_2, \ldots : \Omega \to \mathbb{R}$ such that

$$\mathbb{P}(\xi_n = \alpha_n) = \mathbb{P}(\xi_n = -\alpha_n) = p_n \in \left(0, \frac{1}{2}\right)$$

with $\sum_{n=1}^{\infty} p_n < \infty$, $\mathbb{P}(\xi_n = 0) = 1 - 2p_n$, and $\alpha_n \geqslant 0$. Then one has that

$$\mathbb{P}\left(\sum_{n=1}^{\infty} \xi_n \text{ converges}\right) = 1.$$

Proof. For $c = 1$ we apply the Three-Series-Theorem. By symmetry we have that

$$\sum_{n=1}^{\infty} \mathbb{E}\xi_n^1 = \sum_{n=1}^{\infty} 0 = 0$$

so that item (a) follows. Moreover,

$$\sum_{n=1}^{\infty} \mathbb{E}(\xi_n^1 - \mathbb{E}\xi_n^1)^2 = \sum_{n=1}^{\infty} \mathbb{E}(\xi_n^1)^2 \leqslant 2 \sum_{n=1}^{\infty} p_n < \infty$$

and

$$\sum_{n=1}^{\infty} \mathbb{P}(|\xi_n| \geqslant 1) \leqslant 2 \sum_{n=1}^{\infty} p_n < \infty$$

so that items (b) and (c) follow as well. $\square$

8.2 Two strong laws of large numbers

Given a sequence of independent random variables $\xi_1, \xi_2, \ldots : \Omega \to \mathbb{R}$ such that

$$\frac{1}{n}(\xi_1 + \cdots + \xi_n) \to c \in \mathbb{R} \quad \text{as} \quad n \to \infty$$

we speak about a *Weak Law of Large Numbers* (WLLN) if the convergence is in probability ands about a *Strong Law of Large Numbers* (SLLN) if the convergence is almost surely. We did already meet a SLNN under a 4-th moment condition in Proposition 6.1.5 and a WLLN in Proposition 6.2.6. In the following we prove in Proposition 8.2.1 a SLLN which only needs a first moment condition but the condition that the random variables are identically distributed, and in Proposition 8.2.5 a variant where the random variables do not need to be identically distributed.

Proposition 8.2.1 (Kolmogorov). *Let $\xi_1, \xi_2, ... \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$ such that*

(1) $\xi_1, \xi_2, \ldots$ *are independent,*

(2) $\xi_1, \xi_2, \ldots$ *have the same distribution, and*

(3) $\mathbb{E}|\xi_1| < \infty$.

Then one has $\lim_{n\to\infty} \frac{1}{n} (\xi_1(\omega) + \cdots + \xi_n(\omega)) = \mathbb{E}\xi_1$ *a.s.*

For the proof we need some preparations.

Lemma 8.2.2 (Töplitz). [1] *Let $a_n \geqslant 0$, $b_n := a_1 + a_2 + \cdots + a_n > 0$, $\lim_n b_n = \infty$. Assume $(x_n)_{n=1}^\infty \subseteq \mathbb{R}$ such that $\lim_n x_n = x$. Then*

$$\lim_{n\to\infty} \frac{1}{b_n} \sum_{i=1}^n a_i x_i = x.$$

The proof is subject to Exercise 1.

Lemma 8.2.3 (Kronecker). [2] *Let $0 < b_1 \leqslant b_2 \leqslant b_3 \leqslant \ldots$ with $b_n \to \infty$ and $(x_n)_{n=1}^\infty \subseteq \mathbb{R}$ so that $\sum_{n=1}^\infty x_n$ is convergent. Then*

$$\lim_{n\to\infty} \frac{1}{b_n} \sum_{j=1}^n b_j x_j = 0.$$

[1] Otto Töplitz, 01/08/1881 (Wroclaw, Poland) - 15/02/1940 (Jerusalem, Israel).

[2] Leopold Kronecker, 07/12/1823 (Legnica, Poland) - 29/12/1891 (Berlin, Germany).

Proof. We let $b_0 = 0$, $S_0 = 0$ and $S_n := x_1 + \cdots + x_n$. Then

$$\begin{aligned}\frac{1}{b_n}\sum_{j=1}^n b_j x_j &= \frac{1}{b_n}\sum_{j=1}^n b_j(S_j - S_{j-1}) \\ &= \frac{1}{b_n}\left(b_n S_n - b_0 S_0 - \sum_{j=1}^n S_{j-1}(b_j - b_{j-1})\right) \\ &= S_n - \frac{1}{b_n}\sum_{j=1}^n S_{j-1}(b_j - b_{j-1}) \to x - x = 0\end{aligned}$$

as $n \to \infty$. □

Lemma 8.2.4. *Let $f : \Omega \to \mathbb{R}$ be a random variable. Then*

$$\sum_{n=1}^\infty \mathbb{P}(|f| \geqslant n) \leqslant \int_\Omega |f| d\mathbb{P} \leqslant 1 + \sum_{n=1}^\infty \mathbb{P}(|f| \geqslant n).$$

Proof. We simply have that

$$\sum_{n=1}^\infty \mathbb{P}(|f| \geqslant n) = \sum_{n=1}^\infty \sum_{k \geqslant n} \mathbb{P}(|f| \in [k, k+1)) = \sum_{k=1}^\infty k\mathbb{P}(|f| \in [k, k+1)) \leqslant \mathbb{E}|f|.$$

The other inequality is proved in the same way. □

Next we deduce a variant of the SLLN:

Proposition 8.2.5 (Kolmogorov)**.** *Let $\xi_1, \xi_2, ... \in \mathcal{L}_2(\Omega, \mathcal{F}, \mathbb{P})$ and $\beta_n > 0$ with $\beta_n \uparrow \infty$ such that*

(1) *$\xi_1, \xi_2, ...$ are independent,*

(2) *$\mathbb{E}\xi_n = 0$,*

(3) *$\sum_{n=1}^\infty \frac{\mathbb{E}\xi_n^2}{\beta_n^2} < \infty$.*

Then one has that $\mathbb{P}\left(\lim_{n\to\infty}\frac{1}{\beta_n}(\xi_1+\cdots+\xi_n)=0\right)=1$. *Consequently, for independent* $\xi_1,\xi_2,...:\Omega\to\mathbb{R}$ *such that*

$$\sum_{n=1}^{\infty}\frac{1}{n^2}\mathrm{var}(\xi_n)<\infty \quad \textit{and} \quad m=\mathbb{E}\xi_1=\mathbb{E}\xi_2=\cdots$$

one has that $\mathbb{P}\left(\lim_{n\to\infty}\frac{1}{n}(\xi_1+\cdots+\xi_n)=m\right)=1$.

Proof. From the Two-Series-Theorem we know that

$$\mathbb{P}\left(\sum_{n=1}^{\infty}\frac{\xi_n}{\beta_n}\text{ converges}\right)=1.$$

Hence Lemma 8.2.3 gives that

$$1=\mathbb{P}\left(\lim_n\frac{1}{\beta_n}\sum_{i=1}^{n}\beta_i\frac{\xi_i}{\beta_i}=0\right)=\mathbb{P}\left(\lim_n\frac{1}{\beta_n}\sum_{i=1}^{n}\xi_i=0\right).$$ □

Proof of Proposition 8.2.1. We can assume that $\mathbb{E}\xi_1=0$. The idea is to truncate the random variables and to apply Proposition 8.2.5. We let

$$\tilde{\xi}_n(\omega):=\begin{cases}\xi_n(\omega) & : \ |\xi_n(\omega)|<n\\ 0 & : \ |\xi_n(\omega)|\geqslant n\end{cases}$$

and $\eta_n(\omega):=\tilde{\xi}_n(\omega)-\mathbb{E}\tilde{\xi}_n$.
(i) Now we compute

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{1}{n^2}\mathbb{E}\eta_n^2 &\leqslant \sum_{n=1}^{\infty}\frac{1}{n^2}\mathbb{E}\tilde{\xi}_n^2\\ &= \sum_{n=1}^{\infty}\frac{1}{n^2}\mathbb{E}\mathbb{1}_{\{|\xi_n|<n\}}\xi_n^2\\ &= \sum_{n=1}^{\infty}\sum_{k=1}^{n}\frac{1}{n^2}\mathbb{E}\mathbb{1}_{\{k-1\leqslant|\xi_1|<k\}}\xi_1^2\\ &= \sum_{k=1}^{\infty}\left(\sum_{n=k}^{\infty}\frac{1}{n^2}\right)\mathbb{E}\mathbb{1}_{\{k-1\leqslant|\xi_1|<k\}}\xi_1^2\end{aligned}$$

$$
\begin{aligned}
&\leqslant 2\sum_{k=1}^{\infty}\frac{1}{k}\mathbb{E}\mathbb{1}_{\{k-1\leqslant|\xi_1|<k\}}\xi_1^2 \\
&\leqslant 2\sum_{k=1}^{\infty}\mathbb{E}\mathbb{1}_{\{k-1\leqslant|\xi_1|<k\}}|\xi_1| \\
&= 2\mathbb{E}|\xi_1| < \infty,
\end{aligned}
$$

where we have used that

$$
\sum_{n=k}^{\infty}\frac{1}{n^2} \leqslant \frac{1}{k} + \int_k^{\infty}\frac{1}{x^2}dx = \frac{1}{k} + \left(-x^{-1}|_k^{\infty}\right) = \frac{2}{k}.
$$

(ii) Applying Proposition 8.2.5 gives that $\frac{1}{n}(\eta_1 + \cdots + \eta_n) \xrightarrow[a.s.]{} 0$.
(iii) To replace η by $\tilde{\xi}$ we need to show that $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\tilde{\xi}_k = 0$. According to the TÖPLITZ-Lemma it is sufficient to prove that $\lim_{n\to\infty}\mathbb{E}\tilde{\xi}_n = 0$. But this follows from

$$
\mathbb{E}\tilde{\xi}_n = \int_{\{|\xi_n|<n\}}\xi_n d\mathbb{P} = \int_{\{|\xi_1|<n\}}\xi_1 d\mathbb{P} \to \int_{\Omega}\xi_1 d\mathbb{P} = 0
$$

as $n\to\infty$ because of dominated convergence. So we get $\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\tilde{\xi}_k \xrightarrow[a.s.]{} 0$.
(iv) To replace $\tilde{\xi}$ by ξ we use the Lemma of BOREL-CANTELLI. We observe

$$
\begin{aligned}
\mathbb{E}|\xi_1| < \infty \quad &\Longleftrightarrow \quad \sum_{n=1}^{\infty}\mathbb{P}(|\xi_1| \geqslant n) < \infty \\
&\Longleftrightarrow \quad \sum_{n=1}^{\infty}\mathbb{P}(|\xi_n| \geqslant n) < \infty \\
&\Longleftrightarrow \quad \mathbb{P}(\{\omega : \#\{n : |\xi_n(\omega)| \geqslant n\} < \infty\}) = 1
\end{aligned}
$$

and let $\Omega_0 := \{\omega : \#\{n : |\xi_n(\omega)| \geqslant n\} < \infty\}$. Hence, for all $\omega \in \Omega_0$ we get some $n(\omega)$ such that for $n \geqslant n(\omega)$ one has $\xi_n(\omega) = \tilde{\xi}_n(\omega)$ and

$$
\lim_{n\to\infty}\frac{1}{n}(\xi_1(\omega) + \cdots + \xi_n(\omega)) = \lim_{n\to\infty}\frac{1}{n}\left(\tilde{\xi}_1(\omega) + \cdots + \tilde{\xi}_n(\omega)\right).
$$

This completes the proof. □

Now we consider a converse to the SLLN (Proposition 8.2.1):

Proposition 8.2.6 (Kolomogorov). *Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and that $\xi_1, \xi_2, \ldots : \Omega \to \mathbb{R}$ are independent random variables having the same distribution. If there is a constant $c \in \mathbb{R}$ such that*

$$\frac{1}{n}(\xi_1 + \cdots + \xi_n) \xrightarrow[a.s.]{} c$$

as $n \to \infty$, then $\mathbb{E}|\xi_1| < \infty$ and $\mathbb{E}\xi_1 = c$.

The proof is subject to Exercise 2.

As an application we deduce a result of Borel. Let $t \in [0,1)$ and write

$$t = \sum_{n=1}^{\infty} \frac{1}{2^n} t_n, \ t_n \in \{0,1\},$$

where $\#\{n : t_n = 0\} = \infty$. Hence

$$\{t : t_1 = x_1, \ldots, t_n = x_n\} = \left\{ t : \frac{x_1}{2} + \cdots + \frac{x_n}{2^n} \leqslant t < \frac{x_1}{2} + \cdots + \frac{x_n}{2^n} + \frac{1}{2^n} \right\}.$$

Let $\Omega = [0,1)$, $\mathcal{F} = \mathbb{B}([0,1))$, λ be the Lebesgue measure.

Proposition 8.2.7 (Borel). *Given $t \in [0,1)$ we let*

$$Z_n(t) := \# \{1 \leqslant k \leqslant n : t_k = 1\}.$$

Then

$$\lambda \left(\left\{ t \in [0,1) : \frac{1}{n} Z_n(t) \to \frac{1}{2} \right\} \right) = 1.$$

Proof. Letting $f_n(t) := t_n$ we simply have $\frac{1}{n} Z_n(t) = \frac{1}{n}(f_1(t) + \cdots + f_n(t))$. The random variables $f_1, f_2, \ldots : \Omega \to \mathbb{R}$ satisfy

$$\lambda(f_1 = \theta_1, \ldots, f_n = \theta_N) = \frac{1}{2^N} = \lambda(f_1 = \theta_1) \cdots \lambda(f_n = \theta_N)$$

for all $\theta_1, \ldots, \theta_N \in \{0,1\}$ and are therefore independent as one can replace the condition $\{f_k = \theta_k\}$ by $\{f \in B_k\}$ for $B_k \in \mathcal{B}(\mathbb{R})$. Hence we can apply the SLLN and are done. □

8.3 The law of iterated logarithm

The Law of Iterated Logarithm gives the precise asymptotics of a random walk:

Proposition 8.3.1 (Law of iterated logarithm)**.** *Let $\xi_1, \xi_2, ... : \Omega \to \mathbb{R}$ be a sequence of independent and identically distributed random variables such that $\mathbb{E}\xi_n = 0$ and $\mathbb{E}\xi_n^2 = \sigma^2 \in (0, \infty)$. Then, for $f_n := \xi_1 + \cdots + \xi_n$, one has that*

$$\mathbb{P}\left(\limsup_{\binom{n\to\infty}{n\geqslant 3}} \frac{f_n}{\psi(n)} = 1, \quad \liminf_{\binom{n\to\infty}{n\geqslant 3}} \frac{f_n}{\psi(n)} = -1 \right) = 1$$

with $\psi(n) := \sqrt{2\sigma^2 n \log\log n}$.

Remark 8.3.2. (1) The conditions

$$\limsup_{\substack{n\to\infty\\ n\geqslant 3}} \frac{f_n}{\psi(n)} = 1 \text{ a.s.} \quad \text{and} \quad \liminf_{\substack{n\to\infty\\ n\geqslant 3}} \frac{f_n}{\psi(n)} = -1 \text{ a.s.}$$

are equivalent since one may consider the random variables $(-\xi_n)_{n=1}^\infty$ which satisfy the assumptions of the (LIL) as well.

(2) The statement can be reformulated in terms of the two conditions

$$\begin{aligned} \mathbb{P}\left(\#\{n \geqslant 3 : f_n \geqslant (1-\varepsilon)\psi(n)\} = \infty\right) &= 1, \\ \mathbb{P}\left(\#\{n \geqslant 3 : f_n \geqslant (1+\varepsilon)\psi(n)\} = \infty\right) &= 0 \end{aligned}$$

for all $\varepsilon \in (0,1)$.

(3) Khinchin proved the (LIL) in 1924 in the case that $|\xi_n(\omega)| \leqslant c$. Later on, Kolmogorov extended the law to other random variables in 1929. Finally, the above version was proved by Wiener [3] and Hartman in 1941.

[3] Norbert Wiener, 26/11/1894 (Columbia, USA) - 18/03/1964 (Stockholm, Sweden), worked on Brownian motion, from where he progressed to harmonic analysis, and won the Bôcher prize from his studies on Tauberian theorems.

We will give the idea of the proof of the (LIL) in the case that $\xi_n = g_n \sim N(0,1)$. Let us start with an estimate for the distribution of g_n.

Lemma 8.3.3. *For $g \sim N(0,\sigma^2)$ with $\sigma > 0$ one has that*

$$\lim_{\substack{\lambda\to\infty\\ \lambda>0}} \frac{\mathbb{P}(g>\lambda)}{\frac{\sigma}{\sqrt{2\pi}\lambda}e^{-\frac{\lambda^2}{2\sigma^2}}} = 1.$$

Proof. By the change of variables $y = \sigma x$ we get that

$$\lim_{\substack{\lambda\to\infty\\ \lambda>0}} \frac{\mathbb{P}(g>\lambda)}{\frac{\sigma}{\sqrt{2\pi}\lambda}e^{-\frac{\lambda^2}{2\sigma^2}}} = \lim_{\substack{\lambda\to\infty\\ \lambda>0}} \frac{\int_\lambda^\infty e^{-\frac{y^2}{2\sigma^2}}\frac{dy}{\sqrt{2\pi}\sigma}}{\frac{\sigma}{\sqrt{2\pi}\lambda}e^{-\frac{\lambda^2}{2\sigma^2}}}$$

$$= \lim_{\substack{\lambda\to\infty\\ \lambda>0}} \frac{\int_{\frac{\lambda}{\sigma}}^\infty e^{-\frac{x^2}{2}}\frac{dx}{\sqrt{2\pi}}}{\frac{\sigma}{\sqrt{2\pi}\lambda}e^{-\frac{\lambda^2}{2\sigma^2}}} = \lim_{\substack{\lambda\to\infty\\ \lambda>0}} \frac{\int_{\lambda}^\infty e^{-\frac{x^2}{2}}\frac{dx}{\sqrt{2\pi}}}{\frac{1}{\sqrt{2\pi}\lambda}e^{-\frac{\lambda^2}{2}}} = 1$$

where we apply the rule of L'Hospital [4]. □

We need the following maximal inequality:

Lemma 8.3.4. *Let $\xi_1,...,\xi_n : \Omega \to \mathbb{R}$ be independent random variables which are symmetric, that means*

$$\mathbb{P}(\xi_k \leqslant \lambda) = \mathbb{P}(-\xi_k \leqslant \lambda)$$

for all $k = 1,...,n$ and $\lambda \in \mathbb{R}$. Then one has that

$$\mathbb{P}\left(\max_{k=1,...,n} f_k > \varepsilon\right) \leqslant 2\mathbb{P}(f_n > \varepsilon)$$

where $f_k := \xi_1 + \cdots + \xi_k$ and $\varepsilon > 0$.

[4] Guillaume Francois Antoine Marquis de L'Hôpital, 1661 (Paris, France)- 2/2/1704 (Paris, France), French mathematician who wrote the first textbook on calculus, which consisted of the lectures of his teacher Johann Bernoulli.

Proof. Let $1 \leqslant k \leqslant n$. Because of $\mathbb{P}(f_n - f_k > 0) = \mathbb{P}(f_n - f_k < 0)$ and

$$1 = \mathbb{P}(f_n - f_k > 0) + \mathbb{P}(f_n - f_k < 0) + \mathbb{P}(f_n - f_k = 0)$$

we get that $\mathbb{P}(f_n - f_k \geqslant 0) \geqslant 1/2$. Now, again as in the proof of KOLMOGOROV's maximal inequality, let $B_1 := \{f_1 > \varepsilon\}$ and

$$B_k := \{f_1 \leqslant \varepsilon, ..., f_{k-1} \leqslant \varepsilon, f_k > \varepsilon\}$$

for $k = 2, ..., n$. Then we get that, where $0/0 := 1$,

$$\begin{aligned}\mathbb{P}\left(\max_{k=1,\ldots,n} f_k > \varepsilon\right) &= \sum_{k=1}^{n} \mathbb{P}(B_k) \\ &= \sum_{k=1}^{n} \mathbb{P}(B_k)\frac{\mathbb{P}(f_n \geqslant f_k)}{\mathbb{P}(f_n \geqslant f_k)} \\ &= \sum_{k=1}^{n} \frac{\mathbb{P}(B_k \cap \{f_n \geqslant f_k\})}{\mathbb{P}(f_n \geqslant f_k)}\end{aligned}$$

where we used the independence of B_k and $f_n - f_k = \xi_{k+1} + \cdots + \xi_n$ if $k < n$. Using that

$$\frac{1}{\mathbb{P}(f_n \geqslant f_k)} \leqslant 2$$

we end up with

$$\mathbb{P}\left(\max_{k=1,\ldots,n} f_k > \varepsilon\right) \leqslant 2\sum_{k=1}^{n} \mathbb{P}(B_k \cap \{f_n \geqslant f_k\}) \leqslant 2\mathbb{P}(f_n > \varepsilon). \qquad \square$$

Proof. of Proposition 8.3.1 for $\xi_n = g_n$. By symmetry we only need to show that

$$\mathbb{P}\left(\limsup_{\substack{n\to\infty \\ n\geqslant 3}} \frac{f_n}{\psi(n)} \leqslant 1\right) = 1 \quad \text{and} \quad \mathbb{P}\left(\limsup_{\substack{n\to\infty \\ n\geqslant 3}} \frac{f_n}{\psi(n)} \geqslant 1\right) = 1.$$

This is equivalent that for all $\varepsilon \in (0,1)$ one has that

$$\mathbb{P}\Big(\Big\{\omega \in \Omega : \exists n_0 \geqslant 3 \forall\ n \geqslant n_0\ \ f_n(\omega) \leqslant (1+\varepsilon)\psi(n)\Big\}\Big) = 1 \tag{8.2}$$

and

$$\mathbb{P}\Big(\Big\{\omega \in \Omega : \#\{n \geqslant 3\ \ f_n(\omega) \geqslant (1-\varepsilon)\psi(n)\}\Big\} = \infty\Big) = 1. \tag{8.3}$$

Let $n_k := (1+\varepsilon)^k$ for $k \geqslant k_0$ such that $n_{k_0} \geqslant 3$.

Equation (8.2): Let $\varepsilon > 0$ and define

$$A_k := \{\omega \in \Omega : \exists n \in (n_k, n_{k+1}]\ \ f_n(\omega) > (1+\varepsilon)\psi(n)\}$$

for $k \geqslant k_0$. Then $\mathbb{P}(\limsup_k A_k) = 0$ would imply (8.2). According to the Lemma of Borel-Cantelli it is sufficient to show that $\sum_{k=k_0}^{\infty} \mathbb{P}(A_k) < \infty$. This follows from

$$\begin{aligned} \mathbb{P}(A_k) &\leqslant \mathbb{P}\left(\exists n \in [1, n_{k+1}] : f_n > (1+\varepsilon)\psi(n_k)\right) \\ &\leqslant 2\mathbb{P}\left(f_{[n_{k+1}]} > (1+\varepsilon)\psi(n_k)\right) \\ &\leqslant 2c\frac{\sigma(f_{[n_{k+1}]})}{\sqrt{2\pi}((1+\varepsilon)\psi(n_k))} e^{-\frac{((1+\varepsilon)\psi(n_k))^2}{2\sigma(f_{[n_{k+1}]})^2}} \end{aligned}$$

where we have used the maximal inequality from Lemma 8.3.4 and the estimate from Lemma 8.3.3 and where $\sigma(f_n)^2$ is the variance of f_n. Finally, by $\sigma(f_n)^2 = n$ we get (after some computation) that

$$\sum_{k=k_0}^{\infty} \mathbb{P}(A_k) \leqslant \sum_{k=k_0}^{\infty} c' k^{-(1+\varepsilon)} < \infty.$$

Equation (8.3). Applying (8.2) to $(-f_n)_{n=1}^{\infty}$ and $\varepsilon = 1$ gives that

$$\mathbb{P}\Big(\big\{\omega \in \Omega : \exists n_0 \geqslant 3 \forall\ \ n \geqslant n_0\ \ -f_n(\omega) \leqslant 2\psi(n)\big\}\Big) = 1. \tag{8.4}$$

We set $Y_k := f_{[n_k]} - f_{[n_{k-1}]}$ for $k > k_0$ and $\varepsilon \in (0,1)$. Assume that we can show that

$$Y_k > (1-\varepsilon)\psi(n_k) + 2\psi(n_{k-1}) \tag{8.5}$$

happens infinitely often with probability one, which means by definition that

$$f_{[n_k]} - f_{[n_{k-1}]} > (1-\varepsilon)\psi(n_k) + 2\psi(n_{k-1})$$

happens infinitely often with probability one. Together with (8.4) this would imply that

$$f_{[n_k]} > (1-\varepsilon)\psi(n_k)$$

happens infinitely often with probability one. Hence we have to show (8.5). For this one can prove that

$$\mathbb{P}\left(Y_k > (1-\varepsilon)\psi(n_k) + 2\psi(n_{k-1})\right) \geqslant \frac{c_2}{k \log k}$$

so that
$$\sum_{k=k_0+1}^{\infty} \mathbb{P}\left(Y_k > (1-\varepsilon)\psi(n_k) + 2\psi(n_{k-1})\right) = \infty.$$
An application of the Lemma of BOREL-CANTELLI implies (8.5). □

8.4 An application to insurance

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\lambda > 0$ and $\Delta_1, \Delta_2, ... : \Omega \to [0, \infty)$ be independent random variables having an exponential distribution with parameter $\lambda > 0$, i.e.

$$\mathbb{P}(\Delta_k \in B) = \lambda \int_0^\infty \mathbb{I}_B(t) e^{-\lambda t} dt$$

for a Borel set $B \in \mathcal{B}(\mathbb{R})$. We can assume that the random variables Δ_k do not attend the value zero. Then we define the jump times

$$T_0 := 0 \quad \text{and} \quad T_n := \Delta_1 + \cdots + \Delta_n$$

for $n \geqslant 1$. We know that $\mathbb{E}\Delta_k = 1/\lambda$.

Definition 8.4.1 (POISSON process)**.** *The stochastic process $N = (N_t)_{t \geqslant 0}$, $N_t : \Omega \to \mathbb{R}$, with $N_0 := 0$ and*

$$N_t := n \quad \text{if} \quad T_n \leqslant t < T_{n+1}$$

is called POISSON *Process with intensity λ.*

Assume independent non-negative random variables $L_1, L_2, ... : \Omega \to \mathbb{R}$ having the same distribution such that the random variables of both families $(\Delta_k)_{k \geqslant 1}$ and $(L_k)_{k \geqslant 1}$ are independent, and some $\kappa \in \mathbb{R}$, and define

$$X_t := \kappa t - \sum_{k=1}^{N_t} L_k,$$

where an empty sum is treated as zero. The process $(X_t)_{t \geqslant 0}$ belongs to the family of Lévy processes, that means the following properties hold:

Theorem 8.4.2. (1) *For all $0 \leqslant t_1 < \cdots < t_n \leqslant s < t$ the increment $X_t - X_s$ is independent from the family $(X_{t_1}, ..., X_{t_n})$.*

(2) *For all* $0 \leqslant s < t < \infty$ *the random variables* $X_t - X_s$ *and* X_{t-s} *have the same distribution.*

(3) *All trajectories* $t \to X_t(\omega)$ *are right-continuous and have left-hand side paths.*

Letting $C, \kappa > 0$ and assume that the L_k take only non-negative values. Then the process

$$R_t := C + \kappa t - \sum_{k=0}^{N_t} L_k = C + X_t$$

can be interpreted as risk process in insurance, where

(1) $C > 0$ is the starting capital,

(2) $\kappa > 0$ is the rate for charging the premium ,

(3) L_1 is the claim size distribution.

The event

$$\mathcal{R} := \{\omega \in \Omega : \exists t \geqslant 0 \quad R_t(\omega) < 0\}$$

is called *ruin.*

Theorem 8.4.3. *If* $\mathbb{E}|L_k| < \infty$ *and* $\kappa < \lambda \mathbb{E} L_1$ *or if* $\mathbb{E}|L_k|^2 < \infty$ *and* $\kappa = \lambda \mathbb{E} L_1$*, then* $\mathbb{P}(\mathcal{R}) = 1$.

Proof. We get that

$$\begin{aligned}
& \{\omega \in \Omega : \exists t \geqslant 0 \quad R_t(\omega) < 0\} \\
= & \{\omega \in \Omega : \exists n \geqslant 1 \quad R_{N_n}(\omega) < 0\} \\
= & \left\{\omega \in \Omega : \exists n \geqslant 1 \quad C + \kappa N_n(\omega) - \sum_{k=1}^{n} L_k(\omega) < 0\right\} \\
= & \left\{\omega \in \Omega : \exists n \geqslant 1 \quad C + \sum_{k=1}^{n} \left[\kappa \Delta_k(\omega) - L_k(\omega)\right] < 0\right\} \\
= & \left\{\omega \in \Omega : \exists n \geqslant 1 \quad \sum_{k=1}^{n} \xi_k(\omega) > C\right\}
\end{aligned}$$

with $\xi_k(\omega) := L_k(\omega) - \kappa\Delta_k(\omega)$. If $\mathbb{E}|L_k| < \infty$ and $\kappa < \lambda\mathbb{E}L_1$, then

$$m := \mathbb{E}\xi_1 = \mathbb{E}L_1 - \frac{\kappa}{\lambda} > 0.$$

By the strong law of large numbers we get that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \xi_k = m > 0 \quad a.s.$$

so that $\lim_{n\to\infty}\sum_{k=1}^{n}\xi_k = \infty$ a.s. and $\mathbb{P}(\mathcal{R}) = 1$. If $\kappa = \lambda\mathbb{E}L_1$ and $\mathbb{E}|L_k|^2 < \infty$ we obtain $m = 0$. Under this condition we can apply the Law of Iterated Logarithm to obtain that

$$\mathbb{P}(\limsup_n \frac{\sum_{k=1}^{n}\xi_k}{\psi(n)} = 1 \quad \text{and} \quad \liminf_n \frac{\sum_{k=1}^{n}\xi_k}{\psi(n)} = -1) = 1$$

which also implies that $\lim_{n\to\infty}\sum_{k=1}^{n}\xi_k = \infty$ a.s. and therefore $\mathbb{P}(\mathcal{R}) = 1$. □

The only condition which remains possible for an insurance company not go bankruptcy is

$$\kappa > \lambda\mathbb{E}L_1$$

which is called *Net Profit Condition*.

8.5 * An ergodic Theorem of Birkhoff and Khinchin

Extensions of the SLLN can be obtained by ergodic theory.

Definition 8.5.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

(1) A measurable map $T : \Omega \to \Omega$ is called *measure preserving* provided that

$$\mathbb{P}(T^{-1}(A)) = \mathbb{P}(A) \quad \text{for all} \quad A \in \mathcal{F}.$$

(2) A measure preserving map $T : \Omega \to \Omega$ is called *ergodic* provided that, for $A \in \mathcal{F}$, the condition

$$T^{-1}(A) = A \quad \text{implies} \quad \mathbb{P}(A) \in \{0, 1\}.$$

Now we get the following ergodic theorem:

Proposition 8.5.2 (Birkhoff [5] and Khinchin [6]). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $T : \Omega \to \Omega$ be ergodic, and $f : \Omega \to \mathbb{R}$ be a random variable such that $\mathbb{E}|f| < \infty$. Then one has that*

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} f(T^k) = \mathbb{E}f \quad a.s. \quad where \quad f(T^0) := f.$$

Why does the Ergodic Theorem of Birkhoff and Khinchin extend the Strong Law of Large Numbers? Let

$$(\Omega, \mathcal{F}, \mathbb{P}) := \otimes_{n=1}^{\infty} (M, \Sigma, \mu)$$

for some probability space (M, Σ, μ). Define the shift $T : \Omega \to \Omega$ by

$$T(\omega_1, \omega_2, \omega_3, ...) := (\omega_2, \omega_3, ...).$$

The shift T is ergodic:

(a) T is measure-preserving: This can be checked on the π-system of cylinder-sets with a finite-dimensional basis.

(b) Assume now that $T^{-1}(A) = A$ for some $A \in \mathcal{F}$. By iteration we get that $T^{-k}A = A$ for all $k = 1, 2, ...$ In other words,

$$A \in \bigcap_{k=1}^{\infty} \sigma(P_k, P_{k+1}, ...)$$

where the P_k are the coordinate functionals $P_k : \Omega \to M$, i.e. $P_k(\omega_1, \omega_2, \omega_3, ...) := \omega_k$. By the zero-one law of Kolmogorov we get that $\mathbb{P}(A) \in \{0, 1\}$.

Finally, we remark that the family $(f(T^k))_{k=0}^{\infty}$ forms an i.i.d. sequence of random variables having the distribution of f we were starting from.
To prove Theorem 8.5.2 we need the following theorem:

[5] George David Birkhoff, 21/03/1884 (Overisel, Michigan, USA)- 12/11/1944 (Cambridge, Massachusetts, USA).

[6] Aleksandr Yakovlevich Khinchin, 19/07/1894 (Kondrovo, Kaluzhskaya guberniya, Russia)- 18/11/1959 (Moscow, USSR).

Proposition 8.5.3 (Maximal ergodic theorem)**.** *Let $T : \Omega \to \Omega$ be a measure preserving map and $f : \Omega \to \mathbb{R}$ be an integrable random variable. Define*

$$S_n := f + \cdots + f(T^{n-1}) \quad \textit{and} \quad M_n := \max\{0, S_1, ..., S_n\}$$

for $n \in \mathbb{N}$. Then $\mathbb{E}(f 1\!\!1_{\{M_n>0\}}) \geqslant 0$.

Proof. We follow [17], which goes back to Garsia [8]. First we observe that

$$f + M_n(T) \geqslant f + S_k(T) = S_{k+1} \quad \text{for} \quad 1 \leqslant k \leqslant n < \infty$$

and $f + M_n(T) \geqslant f = S_1$ so that

$$(f + M_n(T))^+ \geqslant M_{n+1} \geqslant M_n \quad \text{for} \quad n \in \mathbb{N}.$$

On $\{M_n > 0\}$ this implies $(f + M_n(T))^+ = f + M_n(T)$. Therefore,

$$\begin{aligned} \mathbb{E} f 1\!\!1_{\{M_n>0\}} &\geqslant \mathbb{E} M_n 1\!\!1_{\{M_n>0\}} - \mathbb{E} M_n(T) 1\!\!1_{\{M_n>0\}} \\ &\geqslant \mathbb{E} M_n 1\!\!1_{\{M_n>0\}} - \mathbb{E} M_n(T) 1\!\!1_{\{M_n(T)>0\}} \\ &= 0. \end{aligned}$$

□

Proof of Proposition 8.5.2. We follow the argument from [17]. By normalization we can assume that $\mathbb{E} f = 0$. Let

$$\eta' := \liminf_n \frac{1}{n} \sum_{k=0}^{n-1} f(T^k) \leqslant \limsup_n \frac{1}{n} \sum_{k=0}^{n-1} f(T^k) =: \eta.$$

We show that $0 \leqslant \eta' \leqslant \eta \leqslant 0$ a.s. By the symmetry of the problem (replace f by $-f$) it is sufficient to show that $\eta \leqslant 0$ a.s. For $\varepsilon > 0$ and $n \in \mathbb{N}$ let

$$\begin{aligned} A_\varepsilon &:= \{\eta > \varepsilon\}, \\ f_\varepsilon &:= (f - \varepsilon) 1\!\!1_{A_\varepsilon}, \\ S_{n,\varepsilon} &:= f_\varepsilon + \cdots + f_\varepsilon(T^{n-1}), \\ M_{n,\varepsilon} &:= \max\{0, S_{1,\varepsilon}, ..., S_{n,\varepsilon}\}. \end{aligned}$$

From Proposition 8.5.3 it follows that $\mathbb{E}(f_\varepsilon 1\!\!1_{M_{n,\varepsilon}>0}) \geqslant 0$. Moreover,

$$\bigcup_{n=1}^{\infty}\{M_{n,\varepsilon}>0\} = \left\{\sup_{n\in\mathbb{N}} S_{n,\varepsilon}>0\right\} = \left\{\sup_{n\in\mathbb{N}} \frac{S_{n,\varepsilon}}{n}>0\right\}$$
$$= \left\{\sup_{n\in\mathbb{N}} \frac{S_n}{n}>\varepsilon\right\}\cap A_\varepsilon = A_\varepsilon$$

because of $\sup_{n\in\mathbb{N}} \frac{S_n}{n} \geqslant \eta$. By LEBESGUE's dominated convergence,

$$0 \leqslant \mathbb{E}(f_\varepsilon 1\!\!1_{\{M_{n,\varepsilon}>0\}}) \to \mathbb{E}(f_\varepsilon 1\!\!1_{A_\varepsilon}) = \mathbb{E}(f 1\!\!1_{A_\varepsilon}) - \varepsilon\mathbb{P}(A_\varepsilon)$$

as $n\to\infty$. If J is the σ-algebra of T-invariant sets, we get that $A_\varepsilon \in J$ and

$$\mathbb{E}(f 1\!\!1_{A_\varepsilon}) = \mathbb{E}(1\!\!1_{A_\varepsilon}\mathbb{E}(f|J)) = \mathbb{E}(1\!\!1_{A_\varepsilon}0) = 0$$

as all sets from J have measure 0 or 1 and therefore

$$0 \leqslant \lim_{n\to\infty} \mathbb{E}(f_\varepsilon 1\!\!1_{\{M_{n,\varepsilon}>0\}}) = -\varepsilon\mathbb{P}(A_\varepsilon)$$

so that $\mathbb{P}(A_\varepsilon)=0$ for all $\varepsilon>0$ and we are done. □

8.6 Exercises

Ex 1: Prove Lemma 8.2.2.

Ex 2: Prove Proposition 8.2.6.

Hint: Because of Proposition 8.2.1 we only need to show that $\mathbb{E}|f_1|<\infty$. According to Lemma 8.2.4 this is equivalent to $\sum_{n=1}^{\infty}\mathbb{P}(|f_n|\geqslant n)<\infty$. Using the Lemma of BOREL-CANTELLI, this is equivalent to

$$\mathbb{P}\left(\{\omega : \#\{n : |f_n(\omega)| \geqslant n\} = \infty\}\right) = 0. \tag{8.6}$$

Now use that

$$\frac{f_n}{n} = \frac{S_n}{n} - \frac{n-1}{n}\frac{S_{n-1}}{n-1} \xrightarrow[a.s.]{} 0$$

as $n\to\infty$ with $S_n := f_1+\cdots+f_n$, which implies 8.6.

Ex 3: Assume quadratic integrable random variables $\xi_1,\xi_2,\ldots:\Omega\to\mathbb{R}$ such that $\mathbb{E}\xi_n = m\in\mathbb{R}$ and $\mathrm{var}(\xi_n) = n/\log(n+1)$. Check whether the conditions of the WLLN in Proposition 6.2.6, i.e.

$$\lim_{n\to\infty}\frac{\mathrm{var}(\xi_1)+\cdots+\mathrm{var}(\xi_n)}{n^2} = 0,$$

and the condition of the SLLN in Proposition 8.2.5, i.e.

$$\sum_{n=1}^{\infty} \frac{\mathrm{var}(\xi_n)}{n^2} < \infty,$$

is satisfied.

Chapter 9

Characteristic functions

The concept of characteristic functions is one of the most important tools in probability theory. From the perspective of analysis it is the FOURIER-transform of measures and the idea is to describe properties of random variables $f : \Omega \to \mathbb{R}^d$ by the Fourier transform of their laws. This method is, at the same time, very elegant and extremely useful. And the method is not obvious at all and will provide us with unexpected insights into probability and alternative options of proving our statements.

To recall some facts about complex numbers and to set up complex valued random variables, the reader is referred to Section 10.7 in the appendix.

9.1 Definition and basic properties of characteristic functions

Definition 9.1.1. Given $d \in \mathbb{N}$, we denote the collection of finite signed measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ (see Definition 7.1.1) by $\mathcal{M}(\mathbb{R}^d)$. The subset of all probability measures is denoted by $\mathcal{M}_1^+(\mathbb{R}^d)$.

The notation '1' stands for $\mu(\mathbb{R}^d) = 1$ and '+' for $\mu(A) \geqslant 0$.

Definition 9.1.2. (1) For $\mu \in \mathcal{M}(\mathbb{R}^d)$ we let

$$\widehat{\mu}(x) := \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\mu(y),\ x \in \mathbb{R}^d,$$

where $\langle x, y\rangle = \sum_{k=1}^d x_k y_k$, is called *Fourier transform* of μ.

(2) Let $f : \Omega \to \mathbb{R}^d$ be a random variable. Then

$$\varphi_f(x) := \mathbb{E}e^{i\langle x,f\rangle} = \int_\Omega e^{i\langle x,f(\omega)\rangle} d\mathbb{P}(\omega), \; x \in \mathbb{R}^d,$$

is called *characteristic function* of f.

Remark 9.1.3. (1) The transforms $\widehat{\mu}$ and φ_f do exist, because $\left|e^{i\langle x,y\rangle}\right| = \left|e^{i\langle x,f\rangle}\right| = 1$ and $y \to e^{i\langle x,y\rangle}$ is continuous, so that $y \to e^{i\langle x,y\rangle}$ and $\omega \to e^{i\langle x,f(\omega)\rangle}$ are measurable and bounded. Moreover, assuming *any* decompostion of $\mu = \mu^+ - \mu-$ into finite and non-negative measures one gets that

$$\int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\mu(y) = \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\mu^+(y) - \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\mu^-(y)$$

because the integrals on the right hand side are always finite due to the boundedness of the integrand $e^{i\langle x,y\rangle}$.

(2) If $\mathbb{P}_f$ is the law of $f : \Omega \to \mathbb{R}^d$, then $\varphi_f(x) = \widehat{\mathbb{P}}_f(x)$ for $x \in \mathbb{R}^d$. In fact, this follows from the change of variable formula where we get, for $\psi(y) = e^{i\langle x,y\rangle}$,

$$\int_\Omega e^{i\langle x,f(\omega)\rangle} d\mathbb{P}(\omega) = \int_{\mathbb{R}^d} \psi(y) d\mathbb{P}_f(y) = \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\mathbb{P}_f(y).$$

Let us consider some first examples that already illustrate the connection to harmonic analysis:

Example 9.1.4. (a) Let $a \in \mathbb{R}^d$ and δ_a be the DIRAC-measure with

$$\delta_a(B) = \begin{cases} 1, & a \in B \\ 0, & a \notin B. \end{cases}$$

Then

$$\widehat{\delta}_a(x) = \int_{\mathbb{R}} e^{i\langle x,y\rangle} d\delta_a(y) = e^{i\langle x,a\rangle}.$$

(b) Let $a_1, \ldots, a_n \in \mathbb{R}^d$, $0 \leqslant \theta_i \leqslant 1$, $\sum_{i=1}^n \theta_i = 1$ and $\mu = \sum_{i=1}^n \theta_i \delta_{a_i}$. Then

$$\widehat{\mu}(x) = \sum_{i=1}^n \theta_i e^{i\langle x,a_i\rangle} = \sum_{i=1}^n \theta_i \left(\cos\langle x,a_i\rangle + i \sin\langle x,a_i\rangle\right),$$

which is a trigonometric polynomial.

(c) Binomial distribution: Let $0 < p < 1$, $d = 1$, $n \in \{1, 2, \ldots\}$ and

$$\mu(\{k\}) := \binom{n}{k} p^{n-k}(1-p)^k, \quad \text{for} \quad k = 0, \ldots, n.$$

Then

$$\begin{aligned}
\widehat{\mu}(x) &= \int_{\mathbb{R}} e^{ixy} d\mu(y) \\
&= \sum_{k=0}^{n} \binom{n}{k} p^{n-k}(1-p)^k e^{ixk} \\
&= \sum_{k=0}^{n} \binom{n}{k} p^{n-k} \left((1-p)e^{ix}\right)^k \\
&= \left(p + (1-p)e^{ix}\right)^n.
\end{aligned}$$

Now we give some basic properties of $\widehat{\mu}$ for $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$. The following items (1) and (2) can be modified to obtain statements about the case $\mu \in \mathcal{M}(\mathbb{R}^d)$ as well.

Proposition 9.1.5. *For $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$ the following is true:*

(1) *The function $\widehat{\mu} : \mathbb{R}^d \to \mathbb{C}$ is uniformly continuous, that means that for all $\varepsilon > 0$ there exists $\delta > 0$ such that $|\widehat{\mu}(x) - \widehat{\mu}(y)| \leqslant \varepsilon$ whenever $|x - y| = \left(\sum_{i=1}^d |x_i - y_i|^2\right)^{\frac{1}{2}} \leqslant \delta$.*

(2) *For all $x \in \mathbb{R}^d$ one has $|\widehat{\mu}(x)| \leqslant \widehat{\mu}(0) = 1$.*

(3) *The function $\widehat{\mu}$ is positive semi-definite, that means that for all vectors $x_1, \ldots, x_n \in \mathbb{R}^d$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ it follows that*

$$\sum_{k,l=1}^{n} \lambda_k \bar{\lambda}_l \widehat{\mu}(x_k - x_l) \geqslant 0.$$

Proof. (1) Assume that $g : \mathbb{R}^d \to \mathcal{C}$ is Borel measurable and $\int_{\mathbb{R}^d} |g(y)| d\mu(y) < \infty$. Then, by HÖLDER's inequality,

$$\left| \int_{\mathbb{R}^d} g(y) d\mu(y) \right|^2 = \left(\int_{\mathbb{R}^d} Re\, g(y) d\mu(y) \right)^2 + \left(\int_{\mathbb{R}^d} Im\, g(y) d\mu(y) \right)^2$$

$$\leqslant \left(\int_{\mathbb{R}^d} \frac{|Re\, g(y)|^2}{|g(y)|} d\mu(y) + \int_{\mathbb{R}^d} \frac{|Im\, g(y)|^2}{|g(y)|} d\mu(y) \right) \times \int_{\mathbb{R}^d} |g(y)| d\mu(y),$$

so that

$$\left| \int_{\mathbb{R}^d} g(y) d\mu(y) \right| \leqslant \int_{\mathbb{R}^d} |g(y)| d\mu(y).$$

Let $\varepsilon > 0$. Choose a ball of radius $R > 0$ such that $\mu(\mathbb{R}^d \backslash B_R(0)) < \frac{\varepsilon}{3}$, where

$$B_R(0) := \left\{ x \in \mathbb{R}^d : \left(\sum_{i=1}^{d} |x_i|^2 \right)^{\frac{1}{2}} \leqslant R \right\}.$$

Take $\delta := \frac{\varepsilon}{3R}$. Then, since $\left| e^{i\alpha} - e^{i\beta} \right| \leqslant |\alpha - \beta|$, and if $|x_1 - x_2| \leqslant \delta$,

$$\begin{aligned}
&|\widehat{\mu}(x_1) - \widehat{\mu}(x_2)| \\
\leqslant\; & \int_{B_R(0)} \left| e^{i\langle x_1, y\rangle} - e^{i\langle x_2, y\rangle} \right| d\mu(y) + \int_{\mathbb{R}^d \backslash B_R(0)} \left| e^{i\langle x_1, y\rangle} - e^{i\langle x_2, y\rangle} \right| d\mu(y) \\
\leqslant\; & \int_{B_R(0)} |\langle x_1 - x_2, y\rangle| \, d\mu(y) + \int_{\mathbb{R}^d \backslash B_R(0)} 2 d\mu(y) \\
\leqslant\; & |x_1 - x_2| \int_{B_R(0)} |y| d\mu(y) + 2\frac{\varepsilon}{3} \\
\leqslant\; & \frac{\varepsilon}{3R} R + 2\frac{\varepsilon}{3} = \varepsilon.
\end{aligned}$$

(2) This part follows from

$$\left| \int_{\mathbb{R}^d} e^{i\langle x, y\rangle} d\mu(y) \right| \leqslant \int_{\mathbb{R}^d} |e^{i\langle x, y\rangle}| d\mu(y) = 1.$$

(3) Here we have that

$$\begin{aligned}
\sum_{k,l=1}^{n} \lambda_k \bar{\lambda}_l \widehat{\mu}(x_k - x_l) &= \sum_{k,l=1}^{n} \lambda_k \bar{\lambda}_l \int_{\mathbb{R}^d} e^{i\langle x_k - x_l, y\rangle} d\mu(y) \\
&= \sum_{k,l=1}^{n} \lambda_k \bar{\lambda}_l \int_{\mathbb{R}^d} e^{i\langle x_k, y\rangle} e^{-i\langle x_l, y\rangle} d\mu(y).
\end{aligned}$$

Since $\overline{e^{i\alpha}} = \overline{\cos\alpha + i\sin\alpha} = \cos\alpha - i\sin\alpha = \cos(-\alpha) + i\sin(-\alpha) = e^{-i\alpha}$, we can continue to

$$\begin{aligned}
\sum_{k,l=1}^{n} \lambda_k \bar{\lambda}_l \widehat{\mu}(x_k - x_l) &= \int_{\mathbb{R}^d} \left[\sum_{k,l=1}^{n} \lambda_k \bar{\lambda}_l e^{i\langle x_k, y\rangle} \overline{e^{i\langle x_l, y\rangle}} \right] d\mu(y) \\
&= \int_{\mathbb{R}^d} \left[\sum_{k=1}^{n} \lambda_k e^{i\langle x_k, y\rangle} \right] \left[\sum_{l=1}^{n} \bar{\lambda}_l \overline{e^{i\langle x_l, y\rangle}} \right] d\mu(y) \\
&= \int_{\mathbb{R}^d} \left| \sum_{k=1}^{n} \lambda_k e^{i\langle x_k, y\rangle} \right|^2 d\mu(y) \geqslant 0.
\end{aligned}$$

□

Now we establish the connection of our FOURIER-transform to the FOURIER-transform for functions $\psi : \mathbb{R}^d \to \mathbb{C}$.

Definition 9.1.6. For a BOREL-measurable $\psi : \mathbb{R}^d \to \mathbb{C}$ such that $\int_{\mathbb{R}^d} |\psi(y)| d\lambda(y) < \infty$ we let the FOURIER-transform $\widehat{\psi} : \mathbb{R}^d \to \mathbb{C}$ be given by

$$\widehat{\psi}(x) := \int_{\mathbb{R}^d} e^{i\langle x, y\rangle} \psi(y) d\lambda^d(y)$$

where λ^d is the LEBESGUE measure on $\mathbb{R}^d$.

Proposition 9.1.7. *Let $\psi : \mathbb{R}^d \to \mathbb{R}$ be Borel-measurable such that $\int_{\mathbb{R}^d} |\psi(x)| d\lambda^d(x) < \infty$. Define*

$$\mu(B) := \int_{\mathbb{R}^d} 1\!\!I_B(x) \psi(x) d\lambda^d(x).$$

Then one has the following:

(1) *$\mu \in \mathcal{M}(\mathbb{R}^d)$ with $\mu = \mu^+ - \mu^-$, where*

$$\mu^{\pm}(B) := \int_{\mathbb{R}^d} 1\!\!I_B(x) \psi^{\pm}(x) d\lambda^d(x),$$

$\psi^+ := \max\{\psi, 0\}$ and $\psi^- := \max\{-\psi, 0\}$.

(2) *$\widehat{\psi}(x) = \widehat{\mu}(x)$ for all $x \in \mathbb{R}^d$.*

Proof. Assertion (1) is obvious, cf. Example 7.1.2. Assertion (2) follows from

$$\begin{aligned}\widehat{\mu}(x) &= \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\mu^+(y) - \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\mu^-(y) \\ &= \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} \psi^+(y) d\lambda(y) - \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} \psi^-(y) d\lambda(y) \\ &= \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} \psi(y) d\lambda(y).\end{aligned}$$

□

9.2 Convolutions

The problem to obtain information about the distribution of $f_1 + \cdots + f_n$ where $f_1, ..., f_n : \Omega \to \mathbb{R}^d$ are independent random variables occurs frequently, we already meet this problem in the *Strong Law of Large Numbers* (SLLN). Now we introduce a technique especially designed for this purpose: the convolution.

Definition 9.2.1 (Convolution of measures)**.** *For $\mu_1, \ldots, \mu_n \in \mathcal{M}(\mathbb{R}^d)$ we define $\mu_1 * \cdots * \mu_n : \mathcal{B}(\mathbb{R}^d) \to \mathbb{R}$ by*

$$(\mu_1 * \cdots * \mu_n)(B) = (\mu_1 \otimes \cdots \otimes \mu_n)\left(\{(x_1, \ldots, x_n) : x_1 + \cdots + x_n \in B\}\right).$$

The map $\mu_1 * \cdots * \mu_n$ is called convolution of $\mu_1, \ldots, \mu_n$.

The connection of the definition above to independent random variables is as follows:

Proposition 9.2.2. *Assume that $f_1, \ldots, f_n : \Omega \to \mathbb{R}^d$ are independent random variables. Then*

$$\mathbb{P}_{f_1} * \cdots * \mathbb{P}_{f_n} = \mathbb{P}_{f_1 + \cdots + f_n}.$$

Proof. Analogously to Proposition 4.4.2, where real-valued random variables were considered, we have that by independence, the distribution of $(f_1, \ldots, f_n)$ is given by

$$\mathbb{P}_{(f_1,\ldots,f_n)} = \mathbb{P}_{f_1} \otimes \cdots \otimes \mathbb{P}_{f_n}.$$

Consequently, by change of variable, for any $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned}
\mathbb{P}(f_1 + \cdots + f_n \in B) &= \mathbb{P}_{(f_1,\ldots,f_n)}(\{(x_1,\ldots,x_n) : x_1 + \cdots + x_n \in B\}) \\
&= \mathbb{P}_{f_1} \otimes \cdots \otimes \mathbb{P}_{f_n}(\{(x_1,\ldots,x_n) : x_1 + \cdots + x_n \in B\}) \\
&= \mathbb{P}_{f_1} * \cdots * \mathbb{P}_{f_n}(B).
\end{aligned}$$

□

Proposition 9.2.3. *For $\mu_1, \ldots, \mu_n \in \mathcal{M}(\mathbb{R}^d)$ one has $\mu_1 * \cdots * \mu_n \in \mathcal{M}(\mathbb{R}^d)$.*

Proof. We can write each μ_k as $\mu_k = \alpha_k^+ \mu_k^+ - \alpha_k^- \mu_k^-$ where $\alpha_k^\pm \in [0,\infty)$ and $\mu_k^\pm \in \mathcal{M}_1^+(\mathbb{R}^d)$. Therefore, $\mu_1 * \cdots * \mu_n$ can be written as a linear combination of convolutions of probability measures. For this reason we can assume w.l.o.g. that $\mu_1, \ldots, \mu_n \in \mathcal{M}_1^+(\mathbb{R}^d)$. But for this case the assertion follows directly from Definition 9.2.1. □

Let us list some basic properties of the convolution. For $a \in \mathbb{R}^d$ we recall that

$$\delta_a(B) := \begin{cases} 1, & a \in B \\ 0, & a \notin B \end{cases}$$

is the Dirac-measure $\delta_a \in \mathcal{M}_1^+(\mathbb{R}^d)$.

Proposition 9.2.4. *Let $\mu, \mu_1, \mu_2, \mu_3 \in \mathcal{M}(\mathbb{R}^d)$. Then the following assertions hold true:*

(1) $\mu_1 * \mu_2 = \mu_2 * \mu_1$.

(2) $\mu_1 * (\mu_2 * \mu_3) = (\mu_1 * \mu_2) * \mu_3 = \mu_1 * \mu_2 * \mu_3$.

(3) $\delta_a * \mu(B) = \mu(B - a)$.

(4) $\delta_0 * \mu = \mu$, *that means that δ_0 is a unit with respect to the convolution.*

(5) $\delta_{a_1} * \cdots * \delta_{a_n} = \delta_{a_1 + \cdots + a_n}$ *for $a_1, \ldots, a_n \in \mathbb{R}^d$.*

Proof. (1) and (2) By multi-linearity we can reduce this statement to $\mu_1, \mu_2, \mu_3 \in \mathcal{M}_1^+(\mathbb{R}^d)$. But then we can use Proposition 9.2.2 and simply use that $f_1 + f_2 = f_2 + f_1$ and that $f_1 + (f_2 + f_3) = (f_1 + f_2) + f_3 = f_1 + f_2 + f_3$.

(3) By Fubini's theorem

$$\begin{aligned}(\delta_a * \mu)(B) &= (\delta_a \otimes \mu)\left(\{(x_1,x_2) : x_1 + x_2 \in B\}\right) \\ &= \int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d} \mathbb{I}_B(x_1+x_2)d\delta_a(x_1)\right]d\mu(x_2) \\ &= \int_{\mathbb{R}^d} \mathbb{I}_B(a+x_2)d\mu(x_2) \\ &= \mu(B-a).\end{aligned}$$

(4) and (5) follow directly from (3). □

Next we want to connect the notation of convolutions for measures to the convolutions of functions. Behind this there is the embedding

$$J : \mathcal{L}_1(\mathbb{R}^d, \lambda^d) \hookrightarrow \mathcal{M}(\mathbb{R}^d)$$

given by

$$\mu(B) := \int_B \psi(x)d\lambda^d(x) \quad \text{with} \quad |\mu|_{\mathrm{TV}} = \int_{\mathbb{R}^d} |\psi(x)|d\lambda^d(x).$$

Given $\psi_1, \psi_2 \in \mathcal{L}_1(\mathbb{R}^d, \lambda^d)$ it turns out that there is a $\psi \in \mathcal{L}_1(\mathbb{R}^d, \lambda^d)$ such that

$$(J\psi_1) * (J\psi_2) = J\psi.$$

To get the candidate for this ψ assume continuous $\psi_1, \psi_2 : \mathbb{R}^d \to [0,\infty) \in \mathcal{L}_1(\mathbb{R}^d, \lambda^d)$ and let us *formally* compute

$$\begin{aligned}(\mu_1 * \mu_2)(B) &= (\mu_1 \otimes \mu_2)(\{(x_1,x_2) : x_1 + x_2 \in B\}) \\ &= \int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d} \mathbb{I}_B(x_1+x_2)\psi_1(x_1)d\lambda^d(x_1)\right]\psi_2(x_2)d\lambda^d(x_2) \\ &= \int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d} \mathbb{I}_B(x)\psi_1(x-x_2)d\lambda^d(x)\right]\psi_2(x_2)d\lambda^d(x_2) \\ &= \int_{\mathbb{R}^d} \mathbb{I}_B(x)\left[\int_{\mathbb{R}^d} \psi_1(x-x_2)\psi_2(x_2)d\lambda^d(x_2)\right]d\lambda^d(x)\end{aligned}$$

by FUBINI's theorem. Hence $(\psi_1 * \psi_2)(x) := \int_{\mathbb{R}} \psi_1(x-y)\psi_2(y)d\lambda^d(y)$ seems to be a good candidate:

Definition 9.2.5 (Convolutions of functions)**.**

(1) Let $\psi_1, \psi_2 : \mathbb{R}^d \to \mathbb{R}$ be Borel-functions such that $\psi_1(x) \geqslant 0$ and $\psi_2(x) \geqslant 0$ for all $x \in \mathbb{R}^d$. Then $\psi_1 * \psi_2 : \mathbb{R}^d \to [0, \infty]$ with

$$(\psi_1 * \psi_2)(x) := \int_{\mathbb{R}^d} \psi_1(x-y)\psi_2(y) d\lambda^d(y)$$

is called *convolution* of ψ_1 and ψ_2.

(2) For $\psi_1, \psi_2 \in \mathcal{L}_1(\mathbb{R}^d, \lambda^d)$ we let

$$(\psi_1 * \psi_2)(x) = (\psi_1^+ * \psi_2^+)(x) - (\psi_1^+ * \psi_2^-)(x) - (\psi_1^- * \psi_2^+)(x) + (\psi_1^- * \psi_2^-)(x)$$

for those $x \in \mathbb{R}^d$ where all terms on the right-hand side are finite.

Now we analyze the definition above:

Proposition 9.2.6. *Let* $\psi_1, \psi_2 : \mathbb{R}^d \to \mathbb{R}$ *be as in* Definition 9.2.5 (1)*. Then one has the following:*

(1) $\psi_1 * \psi_2 : \mathbb{R}^d \to [0, \infty]$ *is an extended measurable function.*

(2) $\lambda_d(\{x \in \mathbb{R}^d : (\psi_1 * \psi_2)(x) = \infty\}) = 0$ *if* $\psi_1, \psi_2 \in \mathcal{L}_1(\mathbb{R}^d, \lambda^d)$.

(3) $\int_{\mathbb{R}^d} (\psi_1 * \psi_2)(x) d\lambda^d(x) = \int_{\mathbb{R}^d} \psi_1(x) d\lambda^d(x) \int_{\mathbb{R}^d} \psi_2(x) d\lambda^d(x)$.

Proof. Consider the map $g : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined by

$$g(x,y) := \psi_1(x-y)\psi_2(y).$$

We get a non-negative measurable function $g : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and can apply Fubini's theorem. We observe that

$$\begin{aligned} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} g(x,y) d\lambda^d(x) d\lambda^d(y) &= \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} \psi_1(x-y)\psi_2(y) d\lambda^d(x) \right] d\lambda^d(y) \\ &= \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} \psi_1(x-y) d\lambda^d(x) \right] \psi_2(y) d\lambda^d(y) \\ &= \int_{\mathbb{R}^d} \psi_2(y) d\lambda^d(y) \int_{\mathbb{R}^d} \psi_1(x) d\lambda^d(x). \end{aligned}$$

implies (3). Moreover, (1) and (2) follow as a byproduct from Fubini's theorem. □

Corollary 9.2.7. *For $\psi_1, \psi_2 \in \mathcal{L}_1(\mathbb{R}^d, \lambda^d)$ consider*

$$(\psi_1 * \psi_2)(x) = (\psi_1^+ * \psi_2^+)(x) - (\psi_1^+ * \psi_2^-)(x) - (\psi_1^- * \psi_2^+)(x) + (\psi_1^- * \psi_2^-)(x).$$

The complement of the set of those $x \in \mathbb{R}$ such that all terms on the right-hand side are finite is of λ^d-measure zero.

Convention 9.2.8. For $\psi_1, \psi_2 \in \mathcal{L}_1(\mathbb{R}^d, \lambda^d)$ we define

$$(\psi_1 * \psi_2)(x) := 0$$

if one of the terms $(\psi_1^+ * \psi_2^+)(x)$, $(\psi_1^+ * \psi_2^-)(x)$, $(\psi_1^- * \psi_2^+)(x)$, $(\psi_1^- * \psi_2^-)(x)$ is infinite.

Now we close this section with a re-formulation of the previous observations in terms of probability densities:

Proposition 9.2.9. *Let $\mu_1, \mu_2 \in \mathcal{M}_1^+(\mathbb{R}^d)$ with*

$$\mu_i(B) := \int_B p_i(x) d\lambda_d(x), \ \ i = 1, 2,$$

for Borel-measurable and non-negative densities $p_1, p_2 : \mathbb{R}^d \to \mathbb{R}$, i.e.

$$\int_{\mathbb{R}^d} p_1(x) d\lambda^d(x) = \int_{\mathbb{R}^d} p_2(x) d\lambda^d(x) = 1.$$

Then

$$(\mu_1 * \mu_2)(B) = \int_B (p_1 * p_2)(x) d\lambda^d(x),$$

i.e. $\mu_1 * \mu_2$ has the density $p_1 * p_2$.

9.3 Some important properties

Proposition 9.3.1. *For $\mu, \nu \in \mathcal{M}_1^+(\mathbb{R}^d)$ one has the following:*

(1) $\widehat{\mu} + \widehat{\nu} = \widehat{\mu + \nu}$ *where* $(\mu + \nu)(B) := \mu(B) + \nu(B)$.

(2) $\widehat{\mu * \nu} = \widehat{\mu}\widehat{\nu}$.

(3) *If* $A = (a_{ij})_{i,j=1}^d : \mathbb{R}^d \to \mathbb{R}^d$ *is a linear transformation, then*

$$\widehat{\mu_A}(x) = \widehat{\mu}(A^{\mathrm{T}}x),$$

where $\mu_A(B) := \mu(x \in \mathbb{R}^d : Ax \in B)$.

(4) *If* $S_a : \mathbb{R}^d \to \mathbb{R}^d$ *is the shift operator* $S_a x := x + a$, $a \in \mathbb{R}^d$, *then*

$$\widehat{\mu_{S_a}} = \widehat{\delta}_a \widehat{\mu}.$$

Proof. (1) is clear, (2) follows from

$$\begin{aligned} \widehat{\mu * \nu}(x) &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{i\langle x, y+z\rangle} d\mu(y) d\nu(z) \\ &= \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\mu(y) \int_{\mathbb{R}^d} e^{i\langle x,z\rangle} d\nu(z) \\ &= \widehat{\mu}(x)\widehat{\nu}(x). \end{aligned}$$

(3) is an exercise and

$$\widehat{\mu_{S_a}}(x) = \int_{\mathbb{R}^d} e^{i\langle x, y+a\rangle} d\mu(y) = e^{i\langle x,a\rangle}\widehat{\mu}(x)$$

implies (4). □

A motivating example.

Example 9.3.2. Let $f_1, f_2, \ldots : \Omega \to \mathbb{R}$ be independent random variables having the same distribution. Let

$$S_n := \frac{1}{\sqrt{n}}(f_1 + \cdots + f_n).$$

We are interested in the convergence of S_n and compute their characteristic functions. Here we get that

$$\begin{aligned} \varphi_{S_n}(t) &= \varphi_{\left(\frac{f_1}{\sqrt{n}} + \cdots + \frac{f_n}{\sqrt{n}}\right)}(t) = \left(\widehat{\mu_{\frac{f_1}{\sqrt{n}}} * \cdots * \mu_{\frac{f_n}{\sqrt{n}}}}\right)(t) \\ &= \left(\widehat{\mu_{\frac{f_1}{\sqrt{n}}}}(t)\right)^n = \left(\mathbb{E}e^{i\frac{f_1}{\sqrt{n}}t}\right)^n = \left(\varphi_{f_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n, \end{aligned}$$

where $\mu_{\frac{f_k}{\sqrt{n}}}$ is the law of $\frac{f_k}{\sqrt{n}}$.

Now we ask:

(Q1) Under which conditions does $(\varphi_{f_1}(\frac{t}{\sqrt{n}}))^n$ converge to a function φ?

(Q2) If $(\varphi_{f_1}(\frac{t}{\sqrt{n}}))^n$ converges, is the limit φ a characteristic function, i.e. does there exist a probability measure μ such that $\widehat{\mu} = \varphi$?

(Q3) And finally, if there is a measure μ, what is its connection to the distributions of S_n?

As a preparation for the Uniqueness Theorem for Fourier transforms we prove next the Theorem of Riemann & Lebesgue.

Definition 9.3.3. Let

$$C_0(\mathbb{R}^d;\mathbb{C}) := \left\{ g : \mathbb{R}^d \to \mathbb{C} \text{ continuous and } \lim_{|x|\to\infty} |g(x)| = 0 \right\}$$

and

$$\|g\|_{C_0} := \sup_{x\in\mathbb{R}^d} |g(x)|.$$

Proposition 9.3.4 (Riemann & Lebesgue)**.** *For all $f \in L_1(\mathbb{R}^d;\mathbb{C})$ one has $\widehat{f} \in C_0(\mathbb{R}^d;\mathbb{C})$.*

Proof. First of all we remark that by decomposing f into the positive and negative parts of the real and imaginary part (so that we have four parts) and applying Proposition 9.1.5 we get that $\widehat{f}$ is continuous. Next we recall that

$$\begin{aligned} \left|\widehat{f}(x) - \widehat{g}(x)\right| &= \left|\int_{\mathbb{R}^d} [f(y) - g(y)]\, e^{ixy} d\lambda_d(y)\right| \\ &\leqslant \int_{\mathbb{R}^d} |f(y) - g(y)|\, d\lambda_d(y) \\ &= \|f - g\|_{L_1}. \end{aligned}$$

Assume that $E \subseteq L_1(\mathbb{R}^d;\mathbb{C})$ is a dense subset such that $\widehat{f_0} \in C_0(\mathbb{R}^d;\mathbb{C})$ for $f_0 \in E$. Letting $f \in L_1(\mathbb{R}^d;\mathbb{C})$ we find $f_n \in E$ such that $\lim_n \|f_n - f\|_{L_1} = 0$

so that $\lim_n \sup_{x\in\mathbb{R}^d} |\widehat{f}_n(x) - \widehat{f}(x)| = 0$. Given $\varepsilon > 0$ we find an $n \geqslant 1$ such that $\|\widehat{f}_n - \widehat{f}\|_{C_0} < \varepsilon$ so that

$$\overline{\lim_{|x|\to\infty}} \left|\widehat{f}(x)\right| = \overline{\lim_{|x|\to\infty}} \left|\widehat{f}(x) - \widehat{f}_n(x)\right| \leqslant \varepsilon.$$

Since this holds for all $\varepsilon > 0$, we are done. What is the set E? We can take all linear combinations of indicator functions $g(x_1, \ldots, x_d) = \mathbb{1}_{(a_1,b_1)}(x_1) \cdots \mathbb{1}_{(a_d,b_d)}(x_d)$ for $-\infty < a_k < b_k < \infty$. In this case we obtain that $\widehat{g}(x_1, \ldots, x_d) = \widehat{\mathbb{1}}_{(a_1,b_1)}(x_1) \cdots \widehat{\mathbb{1}}_{(a_d,b_d)}(x_d)$ and

$$\widehat{\mathbb{1}}_{(a_k,b_k)}(x_k) = \int_{a_k}^{b_k} e^{ix_k y_k} dy_k = \frac{1}{ix_k}\left(e^{ix_k b_k} - e^{ix_k a_k}\right) \longrightarrow 0, \quad \text{as } |x_k| \longrightarrow \infty.$$

(To show that this set E ist a dense subset of $L_1(\mathbb{R}^d;\mathbb{C})$ is an exercise.) □

The second result we need is

Corollary 9.3.5. *One has that*

$$A := \left\{\widehat{f} : \mathbb{R}^d \to \mathbb{C} : f \in L_1(\mathbb{R}^d;\mathbb{C})\right\} \subseteq C_0(\mathbb{R}^d;\mathbb{C})$$

is dense.

Proof. Proposition 9.3.4 implies $A \subseteq C_0(\mathbb{R}^d;\mathbb{C})$. Now we have to check that the conditions of Proposition 10.5.1 are satisfied.

(i) is clear.

(ii) $\widehat{f}_1\widehat{f}_2 = \widehat{f_1 * f_2}$ and $\|f_1 * f_2\|_{L_1} \leqslant \|f_1\|_{L_1}\|f_2\|_{L_1}$

(iii) Here we have that

$$\overline{\widehat{f}}(x) = \overline{\int_{\mathbb{R}^d} e^{i\langle x,y\rangle} f(y) d\lambda_d(y)} = \int_{\mathbb{R}^d} e^{-i\langle x,y\rangle}\overline{f}(y) d\lambda_d(y) = \widehat{\overline{f}}(-x).$$

(iv) Let $x_0 \in \mathbb{R}^d$, $V := \{x_0 + y : |y| \leqslant 1\}$, $f(x) := \mathbb{1}_V(x)e^{-i\langle x_0,x\rangle} \in L_1(\mathbb{R}^d;\mathbb{C})$. Then

$$\widehat{f}(x_0) = \int_{\mathbb{R}^d} e^{i\langle x_0,x\rangle}\mathbb{1}_V(x)e^{-i\langle x_0,x\rangle} d\lambda_d(x) = \lambda_d(V) > 0.$$

(v) Let $x_0 \neq x_1$ and $z_0 := (x_0 - x_1)/|x_0 - x_1|^2$. Then $e^{i\langle x_0, z_0\rangle} \neq e^{i\langle x_1, z_0\rangle}$ because of $e^{i\langle x_0 - x_1, z_0\rangle} = e^i \neq 1$ and $e^{i\langle x_0, z\rangle} \neq e^{i\langle x_1, z\rangle}$ for all $z \in W$, where W is a small neighborhood of z_0. Let $f(x) := \mathbb{I}_W(x)(e^{-i\langle x_0, x\rangle} - e^{-i\langle x_1, x\rangle})$. Then

$$\begin{aligned}
\widehat{f}(x_0) - \widehat{f}(x_1) &= \int_W \left[e^{-i\langle x_0, y\rangle} - e^{-i\langle x_1, y\rangle}\right]\left[e^{i\langle x_0, y\rangle} - e^{i\langle x_1, y\rangle}\right] d\lambda_d(y) \\
&= \int_W \left|e^{i\langle x_0, y\rangle} - e^{i\langle x_1, y\rangle}\right|^2 d\lambda_d(y) > 0.
\end{aligned}$$

□

Proposition 9.3.6 (Uniqueness)**.** *Let $\mu, \nu \in \mathcal{M}_1^+(\mathbb{R}^d)$. Then $\mu = \nu$, if and only if $\widehat{\mu} = \widehat{\nu}$.*

Proof. It is clear that if $\mu = \nu$, then $\widehat{\mu} = \widehat{\nu}$. Let us show the other direction.
Step 1: We show that

$$\widehat{\mu} = \widehat{\nu} \implies \int_{\mathbb{R}^d} \widehat{f}(x) d\mu(x) = \int_{\mathbb{R}^d} \widehat{f}(x) d\nu(x), \quad \text{for all } f \in L_1(\mathbb{R}^d; \mathbb{C}) :$$

Assume that $f \in L_1(\mathbb{R}^d; \mathbb{C})$. By Fubini's theorem we get

$$\begin{aligned}
\int_{\mathbb{R}^d} \widehat{\mu}(y) f(y) d\lambda_d(y) &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{i\langle x, y\rangle} f(y) d\mu(x) d\lambda_d(y) \\
&= \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} e^{i\langle x, y\rangle} f(y) d\lambda_d(y)\right] d\mu(x) \\
&= \int_{\mathbb{R}^d} \widehat{f}(x) d\mu(x).
\end{aligned}$$

This implies the claim.

Step 2: From Step 1 we have that

$$\int_{\mathbb{R}^d} \widehat{f}(x) d\mu(x) = \int_{\mathbb{R}^d} \widehat{f}(x) d\nu(x), \quad \text{for all } f \in L_1(\mathbb{R}^d; \mathbb{C})$$

and want to show now that $\mu = \nu$. Assuming $g \in C_0(\mathbb{R}^d; \mathbb{C})$ and $f_n \in L_1(\mathbb{R}^d; \mathbb{C})$ such that $\|g - \widehat{f_n}\|_{C_0} \underset{n}{\longrightarrow} 0$, we get that

$$\left|\int_{\mathbb{R}^d} g(x) d\mu(x) - \int_{\mathbb{R}^d} g(x) d\nu(x)\right|$$

$$
\begin{aligned}
&= \left| \int_{\mathbb{R}^d} \left[g(x) - \widehat{f}_n(x) \right] d\mu(x) - \int_{\mathbb{R}^d} \left[g(x) - \widehat{f}_n(x) \right] d\nu(x) \right| \\
&\leqslant 2 \left\| g - \widehat{f}_n \right\|_{C_0} \xrightarrow[n]{} 0.
\end{aligned}
$$

Let us define now the system

$$
\Pi := \{(a_1, b_1] \times \cdots \times (a_d, b_d] : -\infty < a_k \leqslant b_k < \infty,\ k = 1, \ldots, d\}
$$

For large n we find continuous functions

$$
g_k^{(n)}(x) := \begin{cases} 1 : & x \in [a_k + \frac{1}{n}, b_k] \\ 0 : & x \leqslant a_k \text{ or } x \geqslant b_k + \frac{1}{n} \\ \text{linear} : & \text{otherwise} \end{cases}
$$

so that

$$
\lim_{n \to \infty} \prod_{k=1}^{d} g_k^{(n)}(x_k) = \mathbb{1}_{\prod_{k=1}^{d}(a_k, b_k]}(x).
$$

Since $g^n(x) := \prod_{k=1}^{d} g_k^{(n)}(x_k) \in C_0(\mathbb{R}^d; \mathbb{C})$ we get by dominated convergence that

$$
\begin{aligned}
\mu\left(\prod_{k=1}^{d} (a_k, b_k] \right) &= \int_{\mathbb{R}^d} \lim_{n \to \infty} g^n(x) d\mu(x) \\
&= \lim_{n \to \infty} \int_{\mathbb{R}^d} g^n(x) d\mu(x) \\
&= \lim_{n \to \infty} \int_{\mathbb{R}^d} g^n(x) d\nu(x) \\
&= \nu\left(\prod_{k=1}^{d} (a_k, b_k] \right).
\end{aligned}
$$

Since Π is a π-system which generates $\mathcal{B}(\mathbb{R}^d)$, we are done. □

Next, we state the important

Proposition 9.3.7 (Bochner & Khintchin). *Assume that $\varphi : \mathbb{R}^d \to \mathbb{C}$ is continuous with $\varphi(0) = 1$. Then the following assertions are equivalent* *(i)* φ *is the Fourier transform of some* $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$.

(ii) φ *is positive semi-definite, i.e. for all* $n = 1, 2, \ldots$ *for all* $x_1, \ldots, x_n \in \mathbb{R}^d$ *and* $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$

$$\sum_{k,l=1}^{n} \lambda_k \bar{\lambda}_l \varphi(x_k - x_l) \geqslant 0.$$

Next we have an explicit inversion formula.

Proposition 9.3.8. *(i) Let* $\mu \in \mathcal{M}_1^+(\mathbb{R})$ *and let* $F(b) = \mu((-\infty, b])$ *be its distribution function. Then*

$$F(b) - F(a) = \lim_{c\to\infty} \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-iya} - e^{-iyb}}{iy} \widehat{\mu}(y) d\lambda(y),$$

if $a < b$ *and* a *and* b *are points of continuity of* F.

(ii) If $\mu \in \mathcal{M}_1^+(\mathbb{R})$ *and* $\int_{\mathbb{R}} |\widehat{\mu}(x)| d\lambda(x) < \infty$, *then* μ *has a continuous density* $f : \mathbb{R} \to [0, \infty)$, *i.e.*

$$\mu(B) = \int_B f(x) d\lambda(x).$$

Moreover,

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ixy} \widehat{\mu}(y) d\lambda(y).$$

Proposition 9.3.9. *Let* $\mu \in \mathcal{M}_1^+(\mathbb{R})$. *Then* $\mu(B) = \mu(-B)$ *for all* $B \in \mathcal{B}(\mathbb{R})$, *if and only if* $\widehat{\mu}(x) \in \mathbb{R}$ *for all* $x \in \mathbb{R}$.

The proof is an exercise.

The Bochner & Khinchin assumption *positive semi-definite* is sometimes difficult to check. For a real-valued function, there is an easier sufficient condition:

Proposition 9.3.10 (Polya)**.** *Let* $\varphi : \mathbb{R} \to [0, \infty)$ *be*

1. *continuous,*
2. *even (i.e. $\varphi(x) = \varphi(-x)$),*
3. *convex on $[0, \infty)$,*
4. *and assume that $\varphi(0) = 1$ and $\lim_{x\to\infty} \varphi(x) = 0$.*

Then there exists some $\mu \in \mathcal{M}_1^+(\mathbb{R})$ such that $\widehat{\mu} = \varphi$.

9.4 Examples

Normal distribution on $\mathbb{R}$. Recall that

$$\gamma(B) := \int_B e^{-\frac{x^2}{2}} \frac{d\lambda_1(x)}{\sqrt{2\pi}}, \quad B \in \mathcal{B}(\mathbb{R}),$$

is the standard normal distribution on $\mathbb{R}$.

Lemma 9.4.1. *One has that*

$$\widehat{\gamma}(x) = e^{-\frac{x^2}{2}}.$$

Proof. We will not give all details. By definition

$$\widehat{\gamma}(x) = \int_{\mathbb{R}} e^{ixy} e^{-\frac{y^2}{2}} \frac{d\lambda_1(y)}{\sqrt{2\pi}} = \int_{\mathbb{R}} \cos(xy) e^{-\frac{y^2}{2}} \frac{d\lambda_1(y)}{\sqrt{2\pi}}.$$

But then

$$\widehat{\gamma}'(x) = -\int_{\mathbb{R}} \sin(xy) y e^{-\frac{y^2}{2}} \frac{d\lambda_1(y)}{\sqrt{2\pi}} = -x \int_{\mathbb{R}} \cos(xy) e^{-\frac{y^2}{2}} \frac{d\lambda_1(y)}{\sqrt{2\pi}} = -x\widehat{\gamma}(x)$$

by integration by parts. Letting $y \geqslant 0$ and $E(y) := \widehat{\gamma}(\sqrt{2y})$ gives

$$E'(y) = \sqrt{2}\frac{1}{2}\frac{1}{\sqrt{y}}\widehat{\gamma}'(\sqrt{2y}) = \frac{1}{\sqrt{2y}}\widehat{\gamma}'(\sqrt{2y}) = -\frac{\sqrt{2y}}{\sqrt{2y}}\widehat{\gamma}(\sqrt{2y}) = -E(y)$$

for $y > 0$. Since $E(0) = 1$, we get $E(y) = e^{-y}$ and $\widehat{\gamma}(x) = E(\frac{x^2}{2}) = e^{-\frac{x^2}{2}}$. $\square$

Now we shift and stretch the normal distribution. Let $\sigma > 0$ and $m \in \mathbb{R}$. Then

$$\gamma_{m,\sigma^2}(B) := \int_B e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}} \frac{dx}{\sqrt{2\pi\sigma^2}}.$$

Proposition 9.4.2.

(1) $\gamma_{m,\sigma^2} \in \mathcal{M}_1^+(\mathbb{R})$.

(2) $\int_{\mathbb{R}} x d\gamma_{m,\sigma^2}(x) = m$ *is the mean.*

(3) $\int_{\mathbb{R}} (x-m)^2 d\gamma_{m,\sigma^2}(x) = \sigma^2$ *is the variance.*

(4) $\widehat{\gamma}_{m,\sigma^2}(x) = e^{imx} e^{-\frac{1}{2}\sigma^2 x^2}$.

Proof. (1) follows from

$$\int_{\mathbb{R}} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}} \frac{dx}{\sqrt{2\pi\sigma^2}} = \int_{\mathbb{R}} e^{-\frac{y^2}{2}} \frac{dy}{\sqrt{2\pi}} = 1,$$

where we have used $y := \frac{x-m}{\sigma}$ and $\sigma dy = dx$ and the last equality was shown in the basic course.

(2) is a consequence of

$$\int_{\mathbb{R}} x d\gamma_{m,\sigma^2}(x) = \int_{\mathbb{R}} x e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}} \frac{dx}{\sqrt{2\pi\sigma^2}} = \int_{\mathbb{R}} (\sigma y + m) e^{-\frac{1}{2}y^2} \frac{dy}{\sqrt{2\pi}} = m$$

for $y := \frac{x-m}{\sigma}$.

(3) is an exercise.

(4) Here we get

$$\begin{aligned} \widehat{\gamma}_{m,\sigma^2}(x) &= \int_{\mathbb{R}} e^{ixy} e^{-\frac{1}{2}\frac{(y-m)^2}{\sigma^2}} \frac{dy}{\sqrt{2\pi\sigma^2}} \\ &= \int_{\mathbb{R}} e^{ix(\sigma z + m)} e^{-\frac{z^2}{2}} \frac{dz}{\sqrt{2\pi}} \\ &= e^{ixm} \widehat{\gamma}(x\sigma) \\ &= e^{imx} e^{-\frac{1}{2}\sigma^2 x^2}. \end{aligned}$$

□

Definition 9.4.3. A measure $\mu \in \mathcal{M}_1^+(\mathbb{R})$ is a *Gaussian measure,* provided that $\mu = \delta_m$ for some $m \in \mathbb{R}$ or $\mu = \gamma_{m,\sigma^2}$ for $\sigma^2 > 0$ and $m \in \mathbb{R}$.

Normal distribution on $\mathbb{R}^d$. There are various ways to introduce this distribution.

Lemma 9.4.4. *One has that* $\int_{\mathbb{R}^d} e^{-\frac{\langle x,x\rangle}{2}} \frac{dx}{\sqrt{2\pi}^d} = 1$.

Proof. By Fubini's theorem we get that

$$\begin{aligned} \int_{\mathbb{R}^d} e^{-\frac{\langle x,x\rangle}{2}} \frac{dx}{\sqrt{2\pi}^d} &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} e^{-\frac{1}{2}x_1^2} \cdots e^{-\frac{1}{2}x_d^2} \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_d}{\sqrt{2\pi}} \\ &= \left(\int_{\mathbb{R}} e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \right)^d = 1. \end{aligned}$$

□

Definition 9.4.5. The measure $\gamma = \gamma^{(d)} \in \mathcal{M}_1^+(\mathbb{R}^d)$ given by

$$\gamma(B) := \int_B e^{-\frac{\langle x,x\rangle}{2}} \frac{dx}{\sqrt{2\pi}^d}, \quad B \in \mathcal{B}(\mathbb{R}^d),$$

is called *standard Gaussian measure* on $\mathbb{R}^d$.

Definition 9.4.6. A matrix $R = (r_{ij})_{i,j=1}^d$ is called *positive semi-definite*, provided that

$$\langle Rx, x\rangle = \sum_{k,l=1}^d r_{kl} x_k x_l \geqslant 0$$

for all $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$. A matrix $R = (r_{ij})_{i,j=1}^d$ is called *symmetric*, provided that $r_{kl} = r_{lk}$ for all $k, l = 1, \ldots, d$.

Proposition 9.4.7. *Let $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$. Then the following assertions are equivalent*

(1) There exist a matrix $A = (\alpha_{kl})_{k,l=1}^d$ and a vector $m \in \mathbb{R}^d$ such that μ is the image measure of $\gamma^{(d)}$ with respect to the map $x \mapsto Ax + m$, i.e.

$$\mu(B) = \gamma^{(d)} \left(\left\{ x \in \mathbb{R}^d : Ax + m \in B \right\} \right).$$

(2) There exist a positive semi-definite and symmetric matrix $R = (r_{kl})_{k,l=1}^d$ *and a vector* $m' \in \mathbb{R}^d$ *such that*

$$\widehat{\mu}(x) = e^{i\langle x,m'\rangle - \frac{1}{2}\langle Rx,x\rangle}.$$

(3) For all $b = (b_1, \ldots, b_d) \in \mathbb{R}^d$ *the law of* $\psi_b : \mathbb{R}^d \to \mathbb{R}$, $\psi_b(x) := \langle x, b\rangle$, *with respect to* $\left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu\right)$ *is a Gaussian measure on the real line* $\mathbb{R}$.

In particular, we have that m, R *and* m' *are unique and that*

(a) $m = m'$ *and* $R = AA^{\mathrm{T}}$,

(b) $\int_{\mathbb{R}^d} x_k d\mu(x) = m_k$ *and*

(c) $\int_{\mathbb{R}^d} (x_k - m_k)(x_l - m_l)\, d\mu(x) = r_{kl}$.

Definition 9.4.8. A measure $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$ is called Gaussian measure on $\mathbb{R}^d$ with mean $m = (m_k)_{k=1}^d$ and covariance $R = (r_{kl})_{k,l=1}^d$ if there exists a constant $m \in \mathbb{R}^d$ and a symmetric, positiv semi-definite $d \times d$ matrix R such that

$$\widehat{\mu}(x) = e^{i\langle x,m\rangle - \frac{1}{2}\langle Rx,x\rangle}.$$

Proof of Proposition 9.4.7. (i) ⇒ (ii) follows from

$$\begin{aligned}
\widehat{\mu}(x) &= \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\mu(y) = \int_{\mathbb{R}^d} e^{i\langle x,m+Ay\rangle} d\gamma^{(d)}(y) \\
&= e^{i\langle x,m\rangle} \int_{\mathbb{R}} e^{i\left\langle A^{\mathrm{T}}x,y\right\rangle} d\gamma^{(d)}(y) = e^{i\langle x,m\rangle} \widehat{\gamma}^{(d)}\left(A^{\mathrm{T}}x\right) \\
&= e^{i\langle x,m\rangle} e^{-\frac{1}{2}\left\langle A^{\mathrm{T}}x,A^{\mathrm{T}}x\right\rangle} = e^{i\langle x,m\rangle} e^{-\frac{1}{2}\left\langle AA^{\mathrm{T}}x,x\right\rangle}
\end{aligned}$$

so that $AA^{\mathrm{T}} = R$ and $m = m'$ where we have used that

$$\begin{aligned}
\widehat{\gamma}^{(d)}(x) &= \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\gamma^{(d)}(y) \\
&= \int_{\mathbb{R}} e^{ix_1 y_1} d\gamma^{(1)}(y_1) \cdots \int_{\mathbb{R}} e^{ix_d y_d} d\gamma^{(1)}(y_d) \\
&= e^{-\frac{x_1^2}{2}} \cdots e^{-\frac{x_d^2}{2}} = e^{-\frac{1}{2}\langle x,x\rangle}.
\end{aligned}$$

(ii) ⇒ (iii): We compute the Fourier transform of law(ψ_b):

$$\begin{aligned}\widehat{\mathrm{law}(\psi_b)}(t) &= \int_{\mathbb{R}} e^{its} d\mathrm{law}(\psi_b)(s) = \int_{\mathbb{R}^d} e^{it\psi_b(y)} d\mu(y)\\ &= \int_{\mathbb{R}^d} e^{it\langle b,y\rangle} d\mu(y) = \widehat{\mu}(tb)\\ &= e^{i\langle tb,m'\rangle - \frac{1}{2}\langle Rtb,tb\rangle} = e^{it\langle b,m'\rangle - \frac{1}{2}t^2\langle Rb,b\rangle}.\end{aligned}$$

But this is the Fourier transform of a Gaussian measure on $\mathbb{R}$.

(iii) ⇒ (i): For all $b \in \mathbb{R}^d$ there are $m_b \in \mathbb{R}^d$ and $\sigma_b \geqslant 0$ such that

$$\int_{\mathbb{R}^d} e^{it\langle b,y\rangle} d\mu(y) = e^{im_b t - \frac{1}{2}\sigma_b^2 t^2}.$$

Now we compute m_b and σ_b. From Proposition 9.4.2. We know that $\int_{\mathbb{R}^d} \langle b,x\rangle \, d\mu(x) = m_b$ and $\int_{\mathbb{R}^d} (\langle b,x\rangle - m_b)^2 \, d\mu(x) = \sigma_b^2$. The first equation implies that

$$\left\langle b, \left(\int_{\mathbb{R}^d} \langle e_1, x\rangle \, d\mu(x), \ldots, \int_{\mathbb{R}^d} \langle e_d, x\rangle \, d\mu(x) \right) \right\rangle = m_b$$

and $\langle b, m''\rangle = m_b$ with $m'' = \left(\int_{\mathbb{R}^d} \langle e_k, x\rangle \, d\mu(x)\right)_{k=1}^d$. Knowing this, we can rewrite the second equation as

$$\int_{\mathbb{R}^d} \langle b, x - m''\rangle^2 \, d\mu(x) = \sigma_b^2.$$

Defining

$$r'_{kl} := \int_{\mathbb{R}^d} \langle e_k, x - m''\rangle \langle e_l, x - m''\rangle \, d\mu(x)$$

we get that $R' = (r'_{kl})_{k,l=1}^d$ is symmetric and that

$$\begin{aligned}\langle R'b, b\rangle &= \sum_{k,l=1}^d r'_{kl} b_k b_l\\ &= \sum_{k,l=1}^d \int_{\mathbb{R}^d} \langle e_k, x - m''\rangle \langle e_l, x - m''\rangle \, d\mu(x) b_k b_l\end{aligned}$$

$$\begin{aligned} &= \int_{\mathbb{R}^d} \left(\sum_{k=1}^d b_k \langle e_k, x - m'' \rangle \right)^2 d\mu(x) \\ &= \int_{\mathbb{R}^d} \langle b, x - m'' \rangle^2 d\mu(x) = \sigma_b^2. \end{aligned}$$

Consequently,

$$\int_{\mathbb{R}^d} e^{i\langle b,y\rangle} d\mu(y) = e^{i\langle b,m''\rangle - \frac{1}{2}\langle R'b,b\rangle},$$

where R' is positive semi-definite because of $\langle R'b, b\rangle = \sigma_b^2 \geqslant 0$. To get (i) we use algebra: There exists a matrix A such that $R' = AA^{\mathrm{T}}$. Hence

$$\int_{\mathbb{R}^d} e^{i\langle b,y\rangle} d\mu(y) = e^{i\langle b,m''\rangle - \frac{1}{2}\left\langle A^{\mathrm{T}}b, A^{\mathrm{T}}b\right\rangle} = \widehat{\mathrm{law}(\psi)}(b),$$

where $\psi(x) = m'' + Ax$ and $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \gamma^{(d)})$ is taken as probability space.

Finally, we check (a), (b) and (c). Using (ii) $\Rightarrow$ (iii) gives that

$$\int_{\mathbb{R}^d} \langle x, b\rangle \, d\mu(x) = \langle b, m'\rangle \quad \text{and} \quad \int_{\mathbb{R}^d} \langle x - m', b\rangle^2 \, d\mu(x) = \langle Rb, b\rangle,$$

so that $\langle Rb, b\rangle = \langle R'b, b\rangle$ for all $b \in \mathbb{R}^d$. Since R and R' are symmetric, we may deduce

$$\begin{aligned} \langle Rb_1, b_2\rangle &= \frac{1}{4}\left[\langle R(b_1+b_2), b_1+b_2\rangle - \langle R(b_1-b_2), b_1-b_2\rangle\right] \\ &= \frac{1}{4}\left[\langle R'(b_1+b_2), b_1+b_2\rangle - \langle R'(b_1-b_2), b_1-b_2\rangle\right] = \langle R'b_1, b_2\rangle \end{aligned}$$

and $R = R'$ which proves (c). Using (i), we get

$$\int_{\mathbb{R}^d} x d\mu(x) = \int_{\mathbb{R}^d} (m + Ax)\, d\gamma^{(d)}(x) = m,$$

so that (b) is proved and that $m = m'$. Finally, $R = AA^{\mathrm{T}}$ follows now from (i) $\Rightarrow$ (ii). □

Definition 9.4.9. A Gaussian measure $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$ with mean m and covariance R is *degenerated* if $\mathrm{rank}(R) < d$. The Gaussian measure is called *non-degenerated* if $\mathrm{rank}(R) = d$.

Examples

(a) $\gamma^{(d)}$ is non-degenerated since

$$R = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

has rank d.

(b) Let $d = 2$ and $\mu = \gamma^{(1)} \otimes \delta_{x_{2,0}}$, i.e.

$$\mu(B) := \int_{\mathbb{R}} e^{-\frac{1}{2}x_1^2} \mathbb{I}_{(x_1, x_{2,0}) \in B} \frac{dx_1}{\sqrt{2\pi}}.$$

Let us compute the mean:

$$\int_{\mathbb{R}^2} x_1 d\mu(x) = 0 \quad \text{and} \quad \int_{\mathbb{R}^2} x_2 d\mu(x) = x_{2,0}.$$

Moreover

$$\int_{\mathbb{R}^2} (x_1 - 0)^2 d\mu(x) = 1,$$

$$\int_{\mathbb{R}^2} (x_2 - x_{2,0})^2 d\mu(x) = \int_{\mathbb{R}} (x_2 - x_{2,0})^2 d\delta_{x_{2,0}} = 0,$$

and

$$\int_{\mathbb{R}^2} (x_1 - 0)(x_2 - x_{2,0}) d\mu(x) = 0.$$

Consequently, $R = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\text{rank}(R) = 1$.

In the case of non-degenerate measures we have the

Proposition 9.4.10. *Assume that $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$ is a non-degenerate Gaussian measure with covariance R and mean m. Then one has that*

$$\mu(B) = \int_B e^{-\frac{1}{2}\langle R^{-1}(x-m), x-m \rangle} \frac{dx}{(2\pi)^{\frac{d}{2}} |det\ R|^{\frac{1}{2}}}.$$

We will not prove this. The proof is a computation.

Cauchy distribution on $\mathbb{R}$. We start with an experiment. Take a point source which sends small particles to a wall. The distance between the point source and the wall is $\alpha > 0$. The angle $\psi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ is distributed uniformly, that means that it has the uniform distribution on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

Problem: What is the probability that a particle hits a set B? Here we get, for $0 \leqslant \psi < \frac{\pi}{2}$, $\mu_\alpha\left([0, \alpha \tan \psi]\right) = \frac{\psi}{\pi}$ and, for $0 \leqslant x < \infty$,

$$\begin{aligned}\mu_\alpha\left([0, x]\right) &= \frac{\arctan \frac{x}{\alpha}}{\pi} = \frac{1}{\pi}\int_0^{\frac{x}{\alpha}} \frac{d\xi}{1+\xi^2} \\ &= \frac{1}{\pi\alpha}\int_0^x \frac{d\eta}{1+\left(\frac{\eta}{\alpha}\right)^2} = \frac{\alpha}{\pi}\int_0^x \frac{d\eta}{\alpha^2+\eta^2}.\end{aligned}$$

Definition 9.4.11. For $\alpha > 0$ the distribution

$$d\mu_\alpha(x) := \frac{\alpha}{\pi}\frac{1}{\alpha^2 + x^2}dx \in \mathcal{M}_1^+(\mathbb{R}^d)$$

is called *Cauchy distribution with parameter $\alpha > 0$.*

Proposition 9.4.12. *One has that*

$$\widehat{\mu}_\alpha(x) = e^{-\alpha|x|}.$$

Proof. We prove the statement for $\alpha = 1$. (The rest can be done by a change of variables.) Since $\widehat{\mu}_1(-x) = \overline{\widehat{\mu}_1(x)}$ we can restrict our proof to $x \geqslant 0$. We consider the meromorphic function $f : \mathbb{C} \to \mathbb{C}$, $f(z) := \frac{e^{ixz}}{1+z^2}$, which has its residuals in the $z \in \mathbb{C}$ such that $1 + z^2 = 0$, that means $z_1 = i$ and $z_2 = -i$. From complex analysis it is known that

$$\lim_{z\to i}(z-i)f(z) = \frac{1}{2\pi i}\left[\int_{-R}^R f(z)dz + \int_{S_R} f(z)dz\right]$$

where S_R denotes the upper half circle on the complex plane from R to $-R$. Since

$$\int_{S_R} \frac{e^{ixz}}{1+z^2}dz \longrightarrow 0, \text{ as } R \longrightarrow \infty$$

and

$$\lim_{z\to i}(z-i)f(z) = \lim_{z\to i}\frac{e^{ixz}}{z+i} = \frac{1}{2i}e^{-x},$$

we obtain that

$$\frac{1}{2i}e^{-x} = \frac{1}{2\pi i}\lim_{R\to\infty}\int_{-R}^{R}\frac{e^{ixz}}{1+z^2}dz = \frac{1}{2i}\widehat{\mu}_1(x).$$

□

9.5 A characterization of independent random variables

In Corollary 4.4.1 we did prove the existence of independent random variables, now we prove a characterization of independence.

Proposition 9.5.1. *Let $f_1, ..., f_d : \Omega \to \mathbb{R}$ be random variables. Then the following assertions are equivalent.*

(i) $f_1, ..., f_d$ *are independent.*

(ii) $\varphi_{(f_1,...,f_d)}(x) = \varphi_{f_1}(x_1)\dots\varphi_{f_d}(x_d)$ *for all* $x = (x_1..., x_d) \in \mathbb{R}^d$.

Proof. (i)⇒(ii): By Corollary 5.8.3 Define $h : \Omega \to \mathbb{R}^d$ by $h(\omega) := (f_1(\omega), ..., f_d(\omega))$. Then

$$\begin{aligned}\varphi_h(x) &= \int_\Omega e^{i(x_1 f_1(\omega)+\cdots+x_d f_d(\omega))}d\mathbb{P}(\omega)\\ &= \int_\Omega e^{ix_1 f_1(\omega)}d\mathbb{P}(\omega)\cdots\int_\Omega e^{ix_d f_d(\omega)}d\mathbb{P}(\omega) = \varphi_{f_1}(x_1)\cdots\varphi_{f_d}(x_d).\end{aligned}$$

(ii)⇒(i): Define $H : \Omega\times\cdots\times\Omega :\to \mathbb{R}^d$ by

$$H(\omega_1,\dots\omega_d) := (f_1(\omega_1),\dots,f_d(\omega_d)).$$

Then the coordinates are independent and $\varphi_H(x_1,\dots,x_d) = \varphi_h(x_1,\dots,x_d) = \varphi_{f_1}(x_1)\cdots\varphi_{f_d}(x_d)$ so that the law of H and h is the same. But this implies that

$$\mathbb{P}(f_1 \in B_1,\dots,f_d\in B_d) = \mathbb{P}(h\in B_1\times\cdots\times B_d)$$

$$\begin{aligned} &= (\mathbb{P} \otimes \cdots \otimes \mathbb{P}) (H \in B_1 \times \cdots \times B_d) \\ &= \mathbb{P}(f_1 \in B_1) \cdots \mathbb{P}(f_d \in B_d). \end{aligned}$$

□

We consider an application of this:

Definition 9.5.2. The random variables $f_1, \ldots f_d : \Omega \to \mathbb{R}$ with $\mathbb{E}f_1^2 + \cdots + \mathbb{E}f_d^2 < \infty$ are called *uncorrelated*, provided that $\mathbb{E}(f_k - \mathbb{E}f_k)(f_l - \mathbb{E}f_l) = 0$ for all $k, l = 1, ..., d$ with $k \neq l$.

Remark 9.5.3. If $f_1, \ldots f_d : \Omega \to \mathbb{R}$ are independent and if $\mathbb{E}f_1^2 + \cdots + \mathbb{E}f_d^2 < \infty$, then they are uncorrelated. In fact, we have (by Corollary 5.8.3) that

$$\mathbb{E}(f_k - \mathbb{E}f_k)(f_l - \mathbb{E}f_l) = [\mathbb{E}(f_k - \mathbb{E}f_k)] [\mathbb{E}(f_l - \mathbb{E}f_l)] = 0$$

for all $k, l = 1, ..., d$ with $k \neq l$.

Proposition 9.5.4. *Let $f_1, ..., f_d : \Omega \to \mathbb{R}$ be random variables such that $(f_1, ..., f_d) : \Omega \to \mathbb{R}^d$ is a Gaussian random variable. Then the following assertions are equivalent.*

(i) $f_1, ..., f_d$ are uncorrelated.

(ii) $f_1, ..., f_d$ are independent.

Proof. We only have to check that (i) implies (ii). We know from Proposition 9.4.7 that for $x = (x_1, ..., x_d)$ one has

$$\varphi_{(f_1,...,f_d)}(x_1, ..., x_d) = e^{i(x_1 \mathbb{E}f_1 + ... + x_d \mathbb{E}f_d) - \frac{1}{2}\langle Rx, x \rangle},$$

with $R = (r_{k,l})_{k,l=1}^d$ and

$$r_{k,l} = \mathbb{E}(f_k - \mathbb{E}f_k)(f_l - \mathbb{E}f_l) = 0 \quad \text{for} \quad k \neq l.$$

Consequently,

$$\begin{aligned} \varphi_{(f_1,...,f_d)}(x_1, ..., x_d) &= e^{ix_1 \mathbb{E}f_1 - \frac{1}{2}x_1^2 \mathrm{var}(f_1)} \ldots e^{ix_d \mathbb{E}f_d - \frac{1}{2}x_d^2 \mathrm{var}(f_d)} \\ &= \varphi_{f_1}(x_1) \ldots \varphi_{f_d}(x_d). \end{aligned}$$

Applying Proposition 9.5.1 we get the independence of $f_1, ..., f_d$. □

Warning: We need that the joint distribution of $f_1, ..., f_d$ (in other words $\mathrm{law}(f_1, ..., f_d) \in \mathcal{M}_1^+(\mathbb{R}^d)$) is Gaussian.

9.6 Moments of measures

There are different types of moments of a measure $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$. Given integers $l_1, \ldots, l_d \geqslant 0$ and $0 < p < \infty$ we have for example that

$$\begin{aligned}
\int_{\mathbb{R}^d} x_1^{l_1} \cdots x_d^{l_d} d\mu(x_1, \ldots, x_d) &= \text{moment of order } (l_1, \ldots, l_d), \\
\int_{\mathbb{R}^d} \left| x_1^{l_1} \cdots x_d^{l_d} \right| d\mu(x_1, \ldots, x_d) &= \text{absolute moment of order } (l_1, \ldots, l_d), \\
\int_{\mathbb{R}} \left| x - \int_{\mathbb{R}} x d\mu(x) \right|^p d\mu(x) &= \text{centred absolute } p\text{-th moment.}
\end{aligned}$$

We are interested in the first type and show that one can use the Fourier transform to compute these moments:

Proposition 9.6.1. *Let $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$ and assume integers $k_1, \ldots, k_d \geqslant 0$ such that for all integers $0 \leqslant l_j \leqslant k_j$ one has that*

$$\int_{\mathbb{R}^d} \left| x_1^{l_1} \cdots x_d^{l_d} \right| d\mu(x_1, \ldots, x_d) < \infty.$$

Then one has the following:

(i)
$$\frac{\partial^{l_1 + \cdots + l_d}}{\partial x_1^{l_1} \cdots \partial x_d^{l_d}} \widehat{\mu} \in C_b(\mathbb{R}^d).$$

(ii)
$$\frac{\partial^{l_1 + \cdots + l_d}}{\partial x_1^{l_1} \cdots \partial x_d^{l_d}} \widehat{\mu}(x) = i^{l_1 + \cdots + l_d} \int_{\mathbb{R}^d} e^{i\langle x, y \rangle} y_1^{l_1} \cdots y_d^{l_d} d\mu(y).$$

(iii)
$$\frac{\partial^{l_1 + \cdots + l_d}}{\partial x_1^{l_1} \cdots \partial x_d^{l_d}} \widehat{\mu}(0) = i^{l_1 + \cdots + l_d} \int_{\mathbb{R}^d} y_1^{l_1} \cdots y_d^{l_d} d\mu(y).$$

(iv) The partial derivatives of $\widehat{\mu}$ are uniformly continuous.

Example 9.6.2. (a) Binomial distribution: $0 < p < 1$, $d = 1$, $n \in \{1, 2, \ldots\}$.

$$\mu(\{k\}) := \binom{n}{k} p^{n-k}(1-p)^k, \ k = 1, \ldots, n.$$

$$\widehat{\mu}(x) = \left[p + (1-p)e^{ix}\right]^n, \quad \widehat{\mu}'(x) = n\left[p + (1-p)e^{ix}\right]^{n-1}(1-p)ie^{ix},$$

$$\widehat{\mu}'(0) = n(1-p)i, \quad \frac{\widehat{\mu}'(0)}{i} = n(1-p).$$

(b) Gaussian measure $\gamma_{0,\sigma^2} \in \mathcal{M}_1^+(\mathbb{R})$: We have that $\widehat{\gamma}_{0,\sigma^2}(x) = e^{-\frac{1}{2}\sigma^2 x^2}$ and get

$$\widehat{\gamma}'_{0,\sigma^2}(x) = -x\sigma^2 e^{-\frac{1}{2}\sigma^2 x^2}, \quad \widehat{\gamma}'_{0,\sigma^2}(0) = 0,$$

$$\widehat{\gamma}''_{0,\sigma^2}(x) = (x^2\sigma^4 - \sigma^2)e^{-\frac{1}{2}\sigma^2 x^2}, \quad \widehat{\gamma}''_{0,\sigma^2}(0) = -\sigma^2.$$

(c) Cauchy distribution $\mu_\alpha \in \mathcal{M}_1^+(\mathbb{R})$ with $\alpha > 0$: Recall that

$$d\mu_\alpha(x) = \frac{\alpha}{\pi}\frac{1}{\alpha^2 + x^2} \quad \text{and} \quad \widehat{\mu}(x) = e^{-\alpha|x|}.$$

Proposition 9.6.3. *For all $\alpha > 0$ one has*

$$\int_{\mathbb{R}} |x|^k d\mu_\alpha(x) = \infty \quad \textit{for} \quad k = 1, 2, \ldots$$

Proof. A first variant of the proof is

$$\lim_{x \downarrow 0} \frac{e^{-\alpha|x|} - 1}{x} = -\alpha \neq \alpha = \lim_{x \uparrow 0} \frac{e^{-\alpha|x|} - 1}{x}.$$

Or we can use

$$\int_{\mathbb{R}} \frac{|x|^k}{(\alpha^2 + x^2)} dx \geqslant \frac{1}{\alpha^2 + 1} \int_{|x| \geqslant 1} \frac{|x|^k}{x^2} dx = 2\int_1^\infty x^{k-2} dx = \infty.$$

□

Proof of Proposition 9.6.1. We only prove the case $l_1 = 1, l_2 = \cdots = l_d = 0$ (the rest follows by induction). Fix $x_2, \ldots, x_d \in \mathbb{R}$ and define $f(x_1, y) := e^{i\langle (x_1,\ldots,x_d),y\rangle}$. Then

$$\frac{\partial}{\partial x_1}\int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\mu(y) = \int_{\mathbb{R}^d} \frac{\partial}{\partial x_1} e^{i\langle x,y\rangle} d\mu(y) = i\int_{\mathbb{R}^d} e^{i\langle x,y\rangle} y_1 d\mu(y),$$

where the first inequality has to be justified. Now we define $d\upsilon_+(y) = \mathbb{1}_{\{y_1 \geqslant 0\}} y_1 d\mu(y)$ and $d\upsilon_-(y) = -\mathbb{1}_{\{y_1<0\}} y_1 d\mu(y)$ and obtain bounded measures, so that

$$x \mapsto \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\upsilon_\pm(y)$$

are uniformly continuous and bounded and (iv) follows. □

For the equality we have to justify, we need

Lemma 9.6.4. *Let $f : \mathbb{R} \times \Omega \to \mathbb{C}$ be such that*

(i) $\frac{\partial f}{\partial x}(\cdot, \omega)$ *is continuous for all* $\omega \in \Omega$,

(ii) $\frac{\partial f}{\partial x}(x, \cdot)$ *and* $f(x, \cdot)$ *are random variables,*

(iii) There exists a $g : \Omega \to \mathbb{R}$, $g(\omega) \geqslant 0$, *such that* $\left|\frac{\partial f}{\partial x}(x, \omega)\right| \leqslant g(\omega)$ *for all* $\omega \in \Omega$, $x \in \mathbb{R}$ *and* $\mathbb{E}g < \infty$,

(iv) $\int_\Omega |f(x, \omega)| d\mathbb{P}(\omega) < \infty$ *for all* $x \in \mathbb{R}$.

Then

$$\frac{\partial}{\partial x}\int_\Omega f(x, \omega) d\mathbb{P}(\omega) = \int_\Omega \frac{\partial f}{\partial x}(x, \omega) d\mathbb{P}(\omega).$$

9.7 Weak convergence

In the beginning of the lecture we considered the following types of convergence: almost sure convergence, convergence in probability and L_p-convergence. Up to now we needed that the underlying probability spaces are the same. This will be relaxed by the weak convergence, where we only consider the convergence of the laws.

Proposition 9.7.1 (Portmanteau Theorem)**.** *Let $\mu_n, \mu \in \mathcal{M}_1^+(\mathbb{R}^d)$. Then the following assertions are equivalent.*

(i) For all continuous and bounded functions $\varphi : \mathbb{R}^d \to \mathbb{R}$ one has that $\int \varphi(x) d\mu_n(x) \underset{n}{\longrightarrow} \int \varphi(x) d\mu(x)$.

(ii) For all closed sets $A \in \mathcal{B}(\mathbb{R}^d)$ one has $\overline{\lim}_n \mu_n(A) \leqslant \mu(A)$.

(iii) For all open sets $B \in \mathcal{B}(\mathbb{R}^d)$ one has $\underline{lim}_n \mu_n(B) \geqslant \mu(B)$.

(iv) If $d = 1$ and if $F_n(x) := \mu_n((-\infty, x])$ and $F(x) := \mu((-\infty, x])$, then $F_n(x) \underset{n}{\longrightarrow} F(x)$ for all points $x \in \mathbb{R}$ of continuity of F.

(v) $\widehat{\mu}_n(x) \underset{n}{\longrightarrow} \widehat{\mu}(x)$ for $x \in \mathbb{R}^d$.

Definition 9.7.2. (i) For $\mu_n, \mu \in \mathcal{M}_1^+(\mathbb{R}^d)$ we say that μ_n converges weakly to μ ($\mu_n \Rightarrow \mu$ or $\mu_n \xrightarrow{w} \mu$) provided that the conditions of Proposition 9.7.1 are satisfied.

(ii) Let $f_n : \Omega_n \to \mathbb{R}^d$ and $f : \Omega \to \mathbb{R}^d$ be random variables over probability spaces $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ and $(\Omega, \mathcal{F}, \mathbb{P})$. Then f_n converges to f *weakly* or *in distribution* ($f_n \xrightarrow{d} f$) provided that the corresponding laws $\mu_n(B) = \mathbb{P}_n(f_n \in B)$ and $\mu(B) = \mathbb{P}(f \in B)$ are converging weakly.

What is the connection to our earlier types of convergence?

Proposition 9.7.3. *For $f, f_1, f_2, ... : \Omega \to \mathbb{R}^d$ one has that if f_n converges to f in probability, then f_n converges to f in distribution.*

Proof. Letting $\varphi : \mathbb{R}^d \to \mathbb{R}$ be continuous and bounded, we need to show that

$$\mathbb{E}\varphi(f_n) = \int_{\mathbb{R}^d} \varphi(x) d\mu_n(x) \underset{n}{\longrightarrow} \int_{\mathbb{R}^d} \varphi(x) d\mu(x) = \mathbb{E}\varphi(f).$$

But this follows from (defining $\|\varphi\|_\infty := \sup_x |\varphi(x)|$)

$$|\mathbb{E}\varphi(f_n) - \mathbb{E}\varphi(f)| \leqslant \int_{\{|f_n - f| > \varepsilon\}} d\mathbb{P} 2\|\varphi\|_\infty + \int_{\{|f_n - f| \leqslant \varepsilon, |f| > N\}} d\mathbb{P} 2\|\varphi\|_\infty$$

$$
\begin{aligned}
&+\int_{\{|f_n-f|\leqslant\varepsilon,|f|\leqslant N\}} |\varphi(f_n)-\varphi(f)|\,d\mathbb{P}\\
\leqslant\ & 2\|\varphi\|_\infty\left(\mathbb{P}(|f_n-f|>\varepsilon)+\mathbb{P}(|f|>N)\right)\\
&+\sup_{|x-y|\leqslant\varepsilon,\ |y|\leqslant N}|\varphi(x)-\varphi(y)|\\
\leqslant\ & 3\delta,
\end{aligned}
$$

for $N = N(\delta)$, $\varepsilon = \varepsilon(N(\delta), \delta))$ and $n \geqslant n(\varepsilon)$. □

9.8 Central limit theorem

Proposition 9.8.1 (Central limit theorem)**.** *Let $f_1, f_2, \ldots : \Omega \to \mathbb{R}$ be a sequence of independent random variables which have the same distribution such that $0 < \mathbb{E}(f_k - \mathbb{E}f_k)^2 = \sigma^2 < \infty$ and $\mathbb{E}f_k = m$. Then*

$$\mathbb{P}\left(\frac{1}{\sqrt{k\sigma^2}}\left((f_1-m)+\cdots+(f_k-m)\right)\leqslant x\right)\xrightarrow[k]{}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{\xi^2}{2}}d\xi.$$

Proof. Let $f_n^0 := \frac{f_n-m}{\sigma}$. Then we get that $\mathbb{E}f_n^0 = \frac{1}{\sigma}(\mathbb{E}f_n - m) = 0$ and $\mathbb{E}(f_n^0)^2 = \frac{1}{\sigma^2}\mathbb{E}(f_n-m)^2 = 1$. We have to show that

$$\mathbb{P}\left(\frac{1}{\sqrt{n}}\left(f_1^0+\cdots+f_n^0\right)\leqslant x\right)\xrightarrow[n]{}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{\xi^2}{2}}d\xi$$

or, for $S_n := f_1^0 + \cdots + f_n^0$, $\frac{1}{\sqrt{n}}S_n \Rightarrow g \sim N(0,1)$. By Proposition 9.7.1 this is equivalent to

$$\varphi_{\frac{1}{\sqrt{n}}S_n}(t)\xrightarrow[n]{}\varphi_g(t)=e^{-\frac{t^2}{2}}$$

for all $t \in \mathbb{R}$. Now

$$\varphi_{\frac{1}{\sqrt{n}}S_n}(t)=\varphi_{\frac{1}{\sqrt{n}}\left(f_1^0+\cdots+f_n^0\right)}(t)=\varphi_{\frac{f_1^0}{\sqrt{n}}}(t)\cdots\varphi_{\frac{f_n^0}{\sqrt{n}}}(t)=\varphi\left(\frac{t}{\sqrt{n}}\right)^n,$$

if $\varphi(t) = \varphi_{f_1^0}(t)$. Since $\mathbb{E}(f_1^0)^2 < \infty$, Proposition 9.6.1 implies that $\varphi'' \in C_b(\mathbb{R})$ and

$$\varphi(t) = \varphi(0) + t\varphi'(0) + \frac{t^2}{2}\varphi''(0) + o(t^2)$$

$$
\begin{aligned}
&= \varphi(0) + ti\mathbb{E}f_1^0 + \frac{t^2}{2} i^2 \mathbb{E}(f_1^0)^2 + o(t^2) \\
&= 1 - \frac{t^2}{2} + o(t^2)
\end{aligned}
$$

for $t \in \mathbb{R}$ with

$$\lim_{t\to 0} \frac{o(t^2)}{t^2} = 0.$$

So it remains to show that

$$\left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \xrightarrow[n]{} e^{-\frac{t^2}{2}}.$$

Letting $\theta = \frac{t^2}{2}$, this is done by

$$\left(1 - \frac{\theta}{n} + o\left(\frac{\theta}{n}\right)\right)^n \xrightarrow[n]{} e^{-\theta}.$$

Fix $\theta \geqslant 0$, let $\varepsilon \in (0,1)$, and choose $n(\varepsilon,\theta) \in \mathbb{N}$ such that

$$\left| o\left(\frac{\theta}{n}\right)\right| \leqslant \varepsilon \left|\frac{\theta}{n}\right|$$

for $n \geqslant n(\varepsilon,\theta)$. Then

$$\left(1 - \frac{\theta(1+\varepsilon)}{n}\right)^n \leqslant \left(1 - \frac{\theta}{n} + o\left(\frac{\theta}{n}\right)\right)^n \leqslant \left(1 - \frac{\theta(1-\varepsilon)}{n}\right)^n$$

for $n \geqslant n(\varepsilon,\theta)$ such that $\frac{\theta(1+\varepsilon)}{n} < 1$. Because

$$\lim_n \left(1 - \frac{\theta(1+\varepsilon)}{n}\right)^n = e^{-\theta(1+\varepsilon)} \quad \text{and} \quad \lim_n \left(1 - \frac{\theta(1-\varepsilon)}{n}\right)^n = e^{-\theta(1-\varepsilon)}$$

it follows that

$$
\begin{aligned}
e^{-\theta(1+\varepsilon)} \leqslant \liminf_n \left(1 - \frac{\theta}{n} + o\left(\frac{\theta}{n}\right)\right)^n &\leqslant \limsup_n \left(1 - \frac{\theta}{n} + o\left(\frac{\theta}{n}\right)\right)^n \\
&\leqslant e^{-\theta(1-\varepsilon)}.
\end{aligned}
$$

Since this is true for all $\varepsilon > 0$ we end up with

$$e^{-\theta} \leqslant \lim_n \left(1 - \frac{\theta}{n} + o\left(\frac{\theta}{n}\right)\right)^n \leqslant e^{-\theta}$$

which finishes the proof. □

Proposition 9.8.2 (POISSON). *Let $f_{n,1},\ldots,f_{n,n}:\Omega\to\mathbb{R}$ be independent random variables such that $\mathbb{P}(f_{n,k}=1)=p_{nk}$ and $\mathbb{P}(f_{n,k}=0)=q_{nk}$ with $p_{nk}+q_{nk}=1$. Assume that*

$$\max_{1\leqslant k\leqslant n} p_{nk}\to_n 0 \text{ and } \sum_{k=1}^n p_{nk}\to_n \lambda>0.$$

Then, for $S_n:=f_{n,1}+\cdots+f_{n,n}$, the laws $\mu_n:=\mathrm{law(S_n)}$ converge weakly to the POISSON *distribution*

$$\pi_\lambda(B)=\sum_{k=0}^\infty e^{-\lambda}\frac{\lambda^k}{k!}\delta_{\{k\}}(B).$$

Proof. For $\theta=e^{it}-1$ we get that

$$\begin{aligned}
\varphi_{S_n}(t) &= \prod_{k=1}^n \varphi_{f_{n,k}}(t)=\prod_{k=1}^n\left(p_{nk}e^{it}+q_{nk}\right)\\
&= \prod_{k=1}^n\left(1+p_{nk}\left(e^{it}-1\right)\right)\\
&= \prod_{k=1}^n(1+p_{nk}\theta)\\
&= 1+\theta\left(\sum_{k=1}^n p_{nk}\right)+\theta^2\left(\sum_{1\leqslant k_1<k_2\leqslant n}p_{nk_1}p_{nk_2}\right)\\
&\quad +\cdots+\theta^l\left(\sum_{1\leqslant k_1<\cdots<k_l\leqslant n}p_{nk_1}\cdots p_{nk_l}\right)+\cdots\\
&\quad +\theta^n p_{n1}\cdots p_{nn}\\
&= 1+\sum_{l=1}^n\frac{(\theta\lambda)^l}{l!}b_{ln}
\end{aligned}$$

with

$$b_{ln}:=\frac{l!}{\lambda^l}\sum_{1\leqslant k_1<\cdots<k_l\leqslant n}p_{nk_1}\cdots p_{nk_l}.$$

We do not prove POISSON's Theorem, we just show that

$$\lim_n b_{ln}=1$$

for *fixed* l. This can be easily seen by

$$\begin{aligned} & \frac{l!}{\lambda^l} \sum_{1 \leqslant k_1 < \cdots < k_l \leqslant n} p_{nk_1} \cdots p_{nk_l} \\ = & \left(\frac{p_{n1} + \cdots + p_{nn}}{\lambda} \right)^l \frac{\sum_{\substack{1 \leqslant k_1, \ldots, k_l \leqslant n \\ \text{all indices are distinct}}} p_{nk_1} \cdots p_{nk_l}}{\sum_{1 \leqslant k_1, \ldots, k_l \leqslant n} p_{nk_1} \cdots p_{nk_l}}. \end{aligned}$$

The first factor converges to 1 as $n \to \infty$. The second one we write as

$$\frac{\sum_{\substack{1 \leqslant k_1, \ldots, k_l \leqslant n \\ \text{all indices are distinct}}} p_{nk_1} \cdots p_{nk_l}}{\sum_{1 \leqslant k_1, \ldots, k_l \leqslant n} p_{nk_1} \cdots p_{nk_l}} = 1 - \frac{\sum_{\substack{1 \leqslant k_1, \ldots, k_l \leqslant n \\ \text{not all indices are distinct}}} p_{nk_1} \cdots p_{nk_l}}{\sum_{1 \leqslant k_1, \ldots, k_l \leqslant n} p_{nk_1} \cdots p_{nk_l}}$$

and can bound the second term by

$$\begin{aligned} \sum_{\substack{1 \leqslant k_1, \ldots, k_l \leqslant n \\ \text{not all indices are distinct}}} p_{nk_1} \cdots p_{nk_l} \leqslant & \ (p_{n1} + \cdots + p_{nn})^{l-1} \max_k p_{nk} \\ & + (p_{n1} + \cdots + p_{nn})^{l-2} \max_k p_{nk}^2 \\ & + \cdots \\ & + (p_{n1} + \cdots + p_{nn}) \max_k p_{nk}^{l-1}. \end{aligned}$$

This should motivate the convergence

$$\varphi_{S_n}(t) \underset{n}{\longrightarrow} e^{\lambda\theta} = e^{\lambda(e^{it}-1)} = \widehat{\pi}_\lambda(t).$$

□

Example 9.8.3. $p_{n,1} = \cdots = p_{n,n} := \frac{\lambda}{n}$ for $n > \lambda$.

Now we consider a limit theorem which is an extension of the central limit theorem.

Proposition 9.8.4. *Assume that $f_{n1}, f_{n2}, ..., f_{nn}$ are independent random variables such that*

$$\mathbb{E} f_{nk} = 0 \quad \textit{and} \quad \sum_{k=1}^n \mathbb{E} f_{nk}^2 = 1$$

for all $n = 1, 2, ...$ *Let* $\mu_{nk} := \mathrm{law}(f_{nk})$ *and assume that the* Lindeberg condition

$$\sum_{k=1}^{n} \int_{|x|>\varepsilon} x^2 d\mu_{nk}(x) \to_n 0$$

is satisfied for all $\varepsilon > 0$. *Then one has that*

$$S_n \to_d N(0,1)$$

where $S_n := f_{n1} + \cdots + f_{nn}$.

Example 9.8.5. Let us check that the Lindeberg condition is satisfied in the case that we consider the weak limit of

$$\frac{1}{\sqrt{n}}(f_1 + \cdots + f_n)$$

where

(1) $\mathbb{E} f_k = 0$,

(2) $\mathbb{E} f_k^2 = 1$,

(3) $f_1^2, f_2^2, f_3^2, ...$ is a uniformly integrable family of random variables.

In the above notation we can write (in distribution) that

$$f_{nk} =_d \frac{f_k}{\sqrt{n}},$$

so that

$$\begin{aligned}
\sum_{k=1}^{n} \int_{|x|>\varepsilon} x^2 d\mu_{nk}(x) &= \sum_{k=1}^{n} \int_{\left|\frac{f_k}{\sqrt{n}}\right|>\varepsilon} \left(\frac{f_k}{\sqrt{n}}\right)^2 d\mathbb{P} \\
&= \frac{1}{n} \sum_{k=1}^{n} \int_{|f_k|>\sqrt{n}\varepsilon} f_k^2 d\mathbb{P} \\
&\leqslant \sup_{1 \leqslant k \leqslant n} \int_{|f_k|>\sqrt{n}\varepsilon} f_k^2 d\mathbb{P} \\
&\leqslant \sup_{k \geqslant 1} \int_{|f_k|>\sqrt{n}\varepsilon} f_k^2 d\mathbb{P} \\
&\to 0
\end{aligned}$$

as $n \to \infty$. Examples that the family $(f_k^2)_{k=1}^{\infty}$ is uniformly integrable are

(a) $f_1, f_2, ...$ are identical distributed,

(b) $\sup_k \mathbb{E}|f_k|^p < \infty$ for some $2 < p < \infty$.

9.9 Exercises

Ex 1: Compute $\mu_1 * \mu_2$, where μ_i is the uniform distribution on $[a_i, a_i+1] \subseteq \mathbb{R}$ for $i = 1, 2$.

Chapter 10

Appendix

10.1 Some analysis

Definition 10.1.1. [$\liminf_n a_n$ AND $\limsup_n a_n$] For $a_1, a_2, ... \in \mathbb{R}$ we let

$$\begin{aligned} \liminf_n a_n &:= \lim_n \inf_{k \geqslant n} a_k \in [-\infty, \infty], \\ \limsup_n a_n &:= \lim_n \sup_{k \geqslant n} a_k \in [-\infty, \infty] \end{aligned}$$

be the *limit inferior* and the *limit superior* of the sequence $(a_n)_{n=1}^{\infty}$, respectively.

Remark 10.1.2. (1) The $\liminf_n a_n$ does exists because $\inf_{k \geqslant n} a_k$ is a non-decreasing sequence and the limit of this monotone sequence does exist. Similarly, $\limsup_n a_n$ does exist.

(2) The value $\liminf_n a_n$ is the smallest of all $c \in [-\infty, \infty]$ such that there is a subsequence $n_1 < n_2 < n_3 < \cdots$ with $\lim_k a_{n_k} = c$. In other words, it is the smallest of all accumulation points of the sequence $(a_n)_{n=1}^{\infty}$.

(3) Similarly, the value $\limsup_n a_n$ is the largest of all $c[-\infty, \infty]$ such that there is a subsequence $n_1 < n_2 < n_3 < \cdots$ with $\lim_k a_{n_k} = c$. In other words, it is the largest of all accumulation points of the sequence $(a_n)_{n=1}^{\infty}$.

(4) By definition one has that

$$-\infty \leqslant \liminf_n a_n \leqslant \limsup_n a_n \leqslant \infty.$$

Moreover, if $\liminf_n a_n = \limsup_n a_n = a \in \mathbb{R}$, then $\lim_n a_n = a$.

Definition 10.1.3. Let $M \neq \varnothing$. The pair (M, d) is called *metric space,* if $d : M \times M \to [0, \infty)$ satisfies

(i) $d(x, y) = 0$ if and only if $x = y$ (reflexivity),

(ii) $d(x, y) = d(y, x)$ (symmetry),

(iii) $d(x, z) \leqslant d(x, y) + d(y, z)$ (triangle inequality).

Definition 10.1.4. Let X be a linear space and $\|\cdot\| : X \to [0, \infty)$. Then $[X, \|\cdot\|]$ is called a quasi BANACH space, provided that

(1) $\|x\| = 0$ if and only if $x = 0$,

(2) $\|\lambda x\| = |\lambda|\|x\|$ for all $\lambda \in \mathbb{R}$ ($\lambda \in \mathbb{C}$) and $x \in X$,

(3) $\|x + y\| \leqslant c[\|x\| + \|y\|]$ for some $c \geqslant 1$, and

(4) if $(x_n)_{n=1}^{\infty} \subseteq X$ is a CAUCHY sequence, i.e. for all $\varepsilon > 0$ there exists $n(\varepsilon) \geqslant 1$ with $\|x_m - x_n\| \leqslant \varepsilon$ whenever $m, n \geqslant n(\varepsilon)$, then there is an $x \in X$ such that
$$\lim_n \|x_n - x\| = 0.$$

In case of $c = 1$, the space $[X, \|\cdot\|]$ is called BANACH space.

Remark 10.1.5. Properties (1), (2), and (3) say that $\|\cdot\|$ is a norm.

Example 10.1.6. For $p \in [1, \infty)$, $X = \mathbb{R}^n$, and $\|x\|_p = \|(x_1, ..., x_n)\| := (|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}}$ we obtain a norm on $\mathbb{R}^n$.

Definition 10.1.7. [OPEN AND CLOSED SETS]

(1) Given $x, y \in \mathbb{R}^d$, the Euclidean distance is given by
$$d(x, y) = |x - y| := \left(\sum_{i=1}^{d} |x_i - y_i|^2 \right)^{\frac{1}{2}}.$$

For $x \in \mathbb{R}^d$ and $\varepsilon > 0$ we define the **open ball** with radius $\varepsilon > 0$ centered at $x \in \mathbb{R}^d$ by
$$U_\varepsilon(x) := \{y \in \mathbb{R}^d : |x - y| < \varepsilon\}.$$

(2) A subset $A \subseteq \mathbb{R}^d$ is called **open** if for each $x \in A$ there is an $\varepsilon > 0$ such that $U_\varepsilon(x) \subseteq A$.

(3) A subset $B \subseteq \mathbb{R}^d$ is called **closed** if $A := \mathbb{R} \backslash B$ is open.

10.2 Some set theory

Definition 10.2.1. Let $M \neq \varnothing$ be an arbitrary set. We say that a relation $x \sim y$ is *an equivalence class relation* on M, provided that

(i) $x \sim x$ for all $x \in M$ (reflexivity),

(ii) if $x \sim y$, then $y \sim x$ (symmetry),

(iii) if $x \sim y$ and $y \sim z$, then $x \sim z$ (transitivity).

Lemma 10.2.2. *Let $M \neq \varnothing$ equipped with an equivalence class relation $x \sim y$. Then*

$$M = \bigcup_{i \in I} M_i,$$

with

(i) $M_i \neq \varnothing$,

(ii) $M_i \cap M_j = \varnothing$, *if* $i \neq j$,

(iii) $x, y \in M$ *belong to the same set* M_i *if and only if* $x \sim y$.

Definition 10.2.3. The sets M_i are called *equivalence classes*, an element $x_i \in M_i$ is called *representative.* We also use the notation $M_i = [x_i]$ if $x_i \in M_i$.

The Axiom of choice is part of the set theory according to Zermelo-Fraenkel. It has been formulated by Ernst Zermelo in 1904.

Axiom 10.2.4. [Axiom of choice] *Let I be a non-empty set and $(M_\alpha)_{\alpha \in I}$ be a system of non-empty sets M_α. Then there is a function φ on I such that*

$$\varphi : \alpha \to m_\alpha \in M_\alpha.$$

In other words, one can form a set by choosing of each set M_α a representative m_α.

10.3 Some measure and integration theory

10.3.1 Monotone class and π-λ-Theorem

Definition 10.3.1. [λ-SYSTEM] Given $\Omega \neq \emptyset$, a system $\mathcal{L}$ of subsets of Ω is called λ-*system* or DYNKIN [1] *system if*

(1) $\Omega \in \mathcal{L}$,

(2) $A, B \in \mathcal{L}$ *and* $A \subseteq B$ *imply* $B \backslash A \in \mathcal{L}$,

(3) $A_1, A_2, \dots \in \mathcal{L}$ *and* $A_i \subseteq A_{i+1}$, $i = 1, 2, \dots$ *imply* $\bigcup_{i=1}^{\infty} A_n \in \mathcal{L}$.

Definition 10.3.2. [π-SYSTEM] Given $\Omega \neq \emptyset$, a non-empty system $\mathcal{P}$ of subsets of Ω is called π-*system* provided that

$$A \cap B \in \mathcal{P} \quad \text{for all} \quad A, B \in \mathcal{P}.$$

Any algebra is a π-system but a π-system is not an algebra in general, take for example the π-system $\{(a,b) : -\infty < a < b < \infty\} \cup \{\emptyset\}$.

Definition 10.3.3. Given $\Omega \neq \emptyset$, a system $\mathcal{M}$ of subsets of Ω is called μ-*system* or *monotone class* provided that

(1) $\bigcup_{i=1}^{\infty} A_i \in \mathcal{M}$ for all $A_1, A_2, \dots \in \mathcal{M}$ with $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$,

(2) $\bigcap_{i=1}^{\infty} A_i \in \mathcal{M}$ for all $A_1, A_2, \dots \in \mathcal{M}$ with $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$.

Now we collect some set-operations and agree about the following convention: If $\mathcal{G}$ be a non-empty system of subsets of Ω, then we say that $\mathcal{G}$ satisfies the set-rule $\mathcal{R}$ if $\mathcal{R}$ is one of the following conditions ($\mathcal{C}1$) - ($\mathcal{C}10$):

($\mathcal{C}1$) $\Omega \in \mathcal{G}$.

($\mathcal{C}2$) $\emptyset \in \mathcal{G}$.

($\mathcal{C}3$) $A \in \mathcal{G}$ implies $A^c \in \mathcal{G}$.

[1] Eugene Borisovich Dynkin, born 11/05/1924 (Leningrad, SSSR, today St Petersburg, Russia, student of KOLMOGOROV, major contributions in the theory of Lie algebras and to probability theory (for example Markov processes).

($\mathcal{C}4$) $A, B \in \mathcal{G}$ with $A \subseteq B$ implies $B \backslash A \in \mathcal{G}$.

($\mathcal{C}5$) $A, B \in \mathcal{G}$ implies $A \cup B \in \mathcal{G}$.

($\mathcal{C}6$) $A, B \in \mathcal{G}$ implies $A \cap B \in \mathcal{G}$.

($\mathcal{C}7$) $A_1, A_2, \dots \in \mathcal{G}$ and $A_i \subseteq A_{i+1}$, $i = 1, 2, \dots$ imply $\bigcup_{i=1}^{\infty} A_n \in \mathcal{G}$.

($\mathcal{C}8$) $A_1, A_2, \dots \in \mathcal{G}$ and $A_i \supseteq A_{i+1}$, $i = 1, 2, \dots$ imply $\bigcap_{i=1}^{\infty} A_n \in \mathcal{G}$.

($\mathcal{C}9$) $A_1, A_2, \dots \in \mathcal{G}$ implies $\bigcup_{i=1}^{\infty} A_n \in \mathcal{G}$.

($\mathcal{C}10$) $A_1, A_2, \dots \in \mathcal{G}$ implies $\bigcap_{i=1}^{\infty} A_n \in \mathcal{G}$.

definition	associated with
algebra	$\mathcal{C}1$, $\mathcal{C}2$, $\mathcal{C}3$, $\mathcal{C}5$
σ-algebra	$\mathcal{C}1$, $\mathcal{C}2$, $\mathcal{C}3$, $\mathcal{C}9$
λ-system	$\mathcal{C}1$, $\mathcal{C}4$, $\mathcal{C}7$
μ-system	$\mathcal{C}7$, $\mathcal{C}8$
π-system	$\mathcal{C}6$

Lemma 10.3.4 (a meta lemma)**.** *Let $\Omega \neq \emptyset$, let $\mathcal{G}$ be a non-empty system of subsets of Ω, and let $(\mathcal{R}_j)_{j \in J}$ be a family of set-rules on $\mathcal{G}$. Then there is a smallest system $\rho(\mathcal{G})$ of subsets of Ω such that*

(1) $\mathcal{G} \subseteq \rho(\mathcal{G})$,

(2) $\rho(\mathcal{G})$ *satisfies the same family of set-rules* $(\mathcal{R}_j)_{j \in J}$.

The aim of this section is to prove the following fundamental theorems:

Proposition 10.3.5. [Dynkin's π-λ-Theorem] *For a π-system $\mathcal{P}$ one has*

$$\sigma(\mathcal{P}) = \lambda(\mathcal{P}).$$

Proof. As any σ-algebra is a λ-system, one has $\sigma(\mathcal{P}) \supseteq \lambda(\mathcal{P})$. So it remains to show that $\sigma(\mathcal{P}) \subseteq \lambda(\mathcal{P})$. For this it is sufficient to check that $\lambda(\mathcal{A})$ is a σ-algebra. Because an algebra that is a λ-system is a σ-algebra, we only need to verfy that $\lambda(\mathcal{P})$ is an algebra. By properties (1) and (2) we have $\Omega, \emptyset \in \lambda(\mathcal{P})$

and that $A^c \in \lambda(\mathcal{P})$ whenever $A \in \lambda(\mathcal{P})$. Assume now $A, B \in \lambda(\mathcal{P})$. We wish to show that $A \cap B \in \lambda(\mathcal{P})$. We verify this within two steps:

Step 1: We fix $B \in \mathcal{P}$ and define

$$\mathcal{S}_B := \{A \in \lambda(\mathcal{P}) : A \cap B \in \lambda(\mathcal{P})\}.$$

From the definition of a π-system we see that $\mathcal{P} \subseteq \mathcal{S}_B$. Moreover, it is easy to check that $\mathcal{S}_B$ is a λ-system. Therefore $\mathcal{S}_B = \lambda(\mathcal{P})$, i.e. we checked that $A \cap B \in \lambda(\mathcal{P})$ for all $A \in \lambda(\mathcal{P})$ and $B \in \mathcal{P}$. Changing the notation, we have $A \cap B \in \lambda(\mathcal{P})$ for all $A \in \mathcal{P}$ and $B \in \lambda(\mathcal{P})$.

Step 2: We fix $B \in \lambda(\mathcal{P})$ and recall the argument from Step 1. This gives finally that $A \cap B \in \lambda(\mathcal{P})$ for all $A, B \in \lambda(\mathcal{P})$ and the proof is complete. □

Proposition 10.3.6. [Monotone Class Theorem] *For an algebra $\mathcal{A}$ one has*

$$\sigma(\mathcal{A}) = \mu(\mathcal{A}).$$

Proof. As any σ-algebra is a μ-system, one has $\sigma(\mathcal{A}) \supseteq \mu(\mathcal{A})$. So it remains to show that $\sigma(\mathcal{A}) \subseteq \mu(\mathcal{A})$. For this it is sufficient to check that $\mu(\mathcal{A})$ is a σ-algebra and by the monotonicity property (1) only that $\mu(\mathcal{A})$ is an algebra.

(a) Since $\mathcal{A} \subseteq \mu(\mathcal{A})$, we have that $\Omega, \emptyset \in \mu(\mathcal{A})$.

(b) We let

$$\mathcal{C} := \{A \in \mu(\mathcal{A}) : A^c \in \mu(\mathcal{A})\}.$$

We have that $\mathcal{A} \subseteq \mathcal{C}$ as $\mathcal{A}$ is an algebra and $\mathcal{A} \subseteq \mu(\mathcal{A})$ and that $\mathcal{C}$ is a monotone class. Therefore, $A \in \mu(\mathcal{A})$ implies that $A^c \in \mu(\mathcal{A})$.

(c) We prove that $A \cup B \in \mu(\mathcal{A})$ whenever $A, B \in \mu(\mathcal{A})$ and do this again within two steps:

Step 1: We fix $B \in \mathcal{A}$ and consider

$$\mathcal{S}_B := \{A \in \mu(\mathcal{A}) : A \cup B \in \mu(\mathcal{A})\}.$$

As $\mathcal{A}$ is an algebra, $\mathcal{A} \subseteq \mathcal{S}_B$. Moreover, $\mathcal{S}_B$ is a monotone class. Therefore, $\mathcal{S}_B = \mu(\mathcal{A})$, or $A \cup B \in \mu(\mathcal{A})$ whenever $B \in \mathcal{A}$ and $A \in \mu(\mathcal{A})$. Changing the notation, gives $A \cup B \in \mu(\mathcal{A})$ whenever $A \in \mathcal{A}$ and $B \in \mu(\mathcal{A})$, so that in

Step 2 we recall Step 1 for $B \in \mu(\mathcal{A})$. This gives the result. □

Proof of Proposition 2.2.8. Define the system

$$\mathcal{L} := \{A \in \mathcal{F} : \mu(A) = \nu(A)\} \supseteq \mathcal{P}.$$

If we can proof that $\mathcal{L}$ is a λ-system, then Proposition 10.3.5 implies that $\mathcal{F} = \sigma(\mathcal{P}) \subseteq \mathcal{L}$ and the statement follows. Property (1) of Definition 10.3.1 follows by the definition of $\mathcal{L}$, property (2) by the finite additivity, and property (3) from the monotonicity of the measures μ and ν from below. □

10.3.2 Outer measures and CARATHÉODORY's Theorem

We start with the notion of an outer measure:

Definition 10.3.7. Let $\Omega \neq \emptyset$. A map $\mu : 2^\Omega \to [0, \infty]$ is called *outer measure* provided that

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

(2) $\mu(A) \leqslant \mu(B)$ for all $A \subseteq B$,

(3) $\mu(\bigcup_{i=1}^\infty A_i) \leqslant \sum_{i=1}^\infty \mu(A_i)$ for all $A_i \subseteq \Omega$.

Moreover, we let

$$\mathcal{F}^\mu := \{B \subseteq \Omega : \mu(A) = \mu(A \cap B) + \mu(A \cap B^c) \quad \text{for all} \quad A \subseteq \Omega\}$$

be the system of *μ-measurable sets*.

The system $\mathcal{F}^\mu$ can be understood as system of good separators B with respect to μ. The point of the construction is

Proposition 10.3.8. *The system $\mathcal{F}^\mu$ forms a σ-algebra and μ is a measure on $\mathcal{F}^\mu$.*

Proof. Before we start we remark that we also have that

$$\mu\left(\bigcup_{i=1}^n A_i\right) \leqslant \sum_{i=1}^n \mu(A_i)$$

by setting $\emptyset = A_{n+1} = A_{n+2} = \cdots$.
(a) Because $\mu(A \cap \emptyset) + \mu(A \cap \Omega) = \mu(A)$ we have that $\emptyset, \Omega \in \mathcal{F}^\mu$.

(b) By the symmetry of the definition of $\mathcal{F}^\mu$ with respect to B we also have that $B \in \mathcal{F}^\mu$ if and only if $B^c \in \mathcal{F}^\mu$.
(c) Now we assume that $B, B' \in \mathcal{F}^\mu$ and prove that $B \cap B' \in \mathcal{F}^\mu$: This follows from

$$\begin{aligned} \mu(A) &= \mu(A \cap B) + \mu(A \cap B^c) \\ &= \mu(A \cap B \cap B') + \mu(A \cap B \cap (B')^c) + \mu(A \cap B^c) \\ &\geqslant \mu(A \cap B \cap B') + \mu([A \cap B \cap (B')^c] \cup [A \cap B^c]) \\ &= \mu(A \cap B \cap B') + \mu(A \cap (B \cap B')^c). \end{aligned}$$

Therefore, $\mathcal{F}^\mu$ is an algebra.
(d) Now let us assume that $B_1, B_2, ... \in \mathcal{F}^\mu$. By using that $\mathcal{F}^\mu$ is an algebra we may assume that the B_n are pair-wise disjoint while proving that

$$\mu(A) \geqslant \mu\left(A \cap \left(\bigcup_{n=1}^\infty B_n\right)\right) + \mu\left(A \cap \left(\bigcup_{n=1}^\infty B_n\right)^c\right).$$

By the fact that $\mathcal{F}^\mu$ is an algebra we know that

$$\begin{aligned} \mu(A) &= \mu\left(A \cap \left(\bigcup_{i=1}^n B_i\right)\right) + \mu\left(A \cap \left(\bigcup_{i=1}^n B_i\right)^c\right) \\ &= \sum_{i=1}^n \mu(A \cap B_i) + \mu\left(A \cap \left(\bigcup_{i=1}^n B_i\right)^c\right) \end{aligned}$$

where the second equality follows from a successive application of the definition of $\mathcal{F}^\mu$. This implies that

$$\begin{aligned} \mu(A) &\geqslant \sum_{i=1}^\infty \mu(A \cap B_i) + \mu\left(A \cap \left(\bigcup_{i=1}^\infty B_i\right)^c\right) \\ &\geqslant \mu\left(A \cap \left(\bigcup_{i=1}^\infty B_i\right)\right) + \mu\left(A \cap \left(\bigcup_{i=1}^\infty B_i\right)^c\right) \\ &\geqslant \mu(A) \end{aligned}$$

so that we are done.
(e) Finally, let us prove that μ is a measure on $\mathcal{F}^\mu$. Assuming pair-wise disjoint $B_1, B_2, ... \in \mathcal{F}^\mu$ we have that

$$\mu\left(\bigcup_{i=1}^\infty B_i\right) \leqslant \sum_{i=1}^\infty \mu(B_i)$$

by the definition of the outer measure. To check the converse inequality, we observe that the definition of $\mathcal{F}^\mu$ inductively implies that

$$\begin{aligned}\mu\left(\bigcup_{i=1}^\infty B_i\right) &= \mu(B_1) + \mu\left(\bigcup_{i=2}^\infty B_i\right)\\ &= \mu(B_1) + \mu(B_2) + \mu\left(\bigcup_{i=3}^\infty B_i\right)\\ &= \mu(B_1) + \cdots + \mu(B_n) + \mu\left(\bigcup_{i=n+1}^\infty B_i\right)\end{aligned}$$

so that

$$\mu\left(\bigcup_{i=1}^\infty B_i\right) \geqslant \sum_{i=1}^\infty \mu(B_i).$$

□

Now we prove Carathéodory's Theorem:

Proof of Proposition 2.2.5. We define

$$\mu_0^*(A) := \inf\left\{\sum_{i=1}^\infty \mu_0(A_i) : A \subseteq \bigcup_{i=1}^\infty A_i, A_i \in \mathcal{G}\right\}.$$

(a) We check the properties (1-3) for μ_0^* being an outer measure: (1) Because $\mu_0(\varnothing) = 0$, we have $\mu_0^*(\varnothing) = 0$. (2) Any covering of B is a covering of A, so that $\mu_0^*(A) \leqslant \mu_0^*(B)$. (3) Let $\varepsilon > 0$ and find coverings

$$A_i \subseteq \bigcup_{j=1}^\infty A_j^i \quad \text{with} \quad A_j^i \in \mathcal{G} \quad \text{and} \quad \sum_{j=1}^\infty \mu_0(A_j^i) \leqslant \frac{\varepsilon}{2^i} + \mu_0^*(A_i).$$

Then $\bigcup_{i=1}^\infty A_i \subseteq \bigcup_{i,j=1}^\infty A_j^i$ and

$$\mu_0^*\left(\bigcup_{i=1}^\infty A_i\right) \leqslant \sum_{i,j=1}^\infty \mu_0(A_j^i) \leqslant \sum_{i=1}^\infty \left[\frac{\varepsilon}{2^i} + \mu_0^*(A_i)\right] = \varepsilon + \sum_{i=1}^\infty \mu_0^*(A_i).$$

Letting $\varepsilon \downarrow 0$, the subadditivity follows.

(b) One has $\mathcal{G} \subseteq \mathcal{F}^{\mu_0^*}$ so that $\sigma(\mathcal{G}) \subseteq \mathcal{F}^{\mu_0^*}$: Let $B \in \mathcal{G}$ and $A \subseteq \Omega$. We have to show that

$$\mu_0^*(A) = \mu_0^*(A \cap B) + \mu_0^*(A \cap B^c).$$

Because $\mu_0^*(A) \leqslant \mu_0^*(A \cap B) + \mu_0^*(A \cap B^c)$ follows from the property that μ_0^* is an outer measure, it remains to check that $\mu_0^*(A) \geqslant \mu_0^*(A \cap B) + \mu_0^*(A \cap B^c)$. For $\varepsilon > 0$ choose a $\mathcal{G}$-cover

$$A \subseteq \bigcup_{i=1}^{\infty} A_i \quad \text{with} \quad \sum_{i=1}^{\infty} \mu_0(A_i) \leqslant \mu_0^*(A) + \varepsilon.$$

For $A \cap B$ and $A \cap B^c$ we obtain the $\mathcal{G}$-covers

$$A \cap B \subseteq \bigcup_{i=1}^{\infty} (A_i \cap B) \quad \text{and} \quad A \cap B^c \subseteq \bigcup_{i=1}^{\infty} (A_i \cap B^c),$$

and that

$$\mu_0^*(A \cap B) \leqslant \sum_{i=1}^{\infty} \mu_0(A_i \cap B) \quad \text{and} \quad \mu_0^*(A \cap B^c) \leqslant \sum_{i=1}^{\infty} \mu_0(A_i \cap B^c).$$

This gives that

$$\begin{aligned} \mu_0^*(A \cap B) + \mu_0^*(A \cap B^c) &\leqslant \sum_{i=1}^{\infty} [\mu_0(A_i \cap B) + \mu_0(A_i \cap B^c)] \\ &= \sum_{i=1}^{\infty} \mu_0(A_i) \\ &\leqslant \mu_0^*(A) + \varepsilon. \end{aligned}$$

Again, letting $\varepsilon \downarrow 0$ yields the assertion.

(c) One has $\mu_0 = \mu_0^*$ on $\mathcal{G}$: Let $A \in \mathcal{G}$. Then A is a $\mathcal{G}$-cover for A so that $\mu_0^*(A) \leqslant \mu_0(A)$. Now we check $\mu_0(A) \leqslant \mu_0^*(A)$, i.e. that for any $\mathcal{G}$-cover $A \subseteq \bigcup_{i=1}^{\infty} A_i$ one has

$$\mu_0(A) \leqslant \sum_{i=1}^{\infty} \mu_0(A_i).$$

This follows from

$$\mu_0(A) = \mu_0 \left(\bigcup_{i=1}^{\infty} (A \cap A_i) \right) = \sum_{i=1}^{\infty} \mu_0 (A \cap A_i) \leqslant \sum_{i=1}^{\infty} \mu_0 (A_i).$$

(d) It remains to show the uniqueness. For the case $\mu_0(\Omega) < \infty$ this follows immediately from Proposition 2.2.8. The general case can be checked as follows: We consider the trace σ-algebras of $\sigma(\mathcal{G})$ on Ω_n, and get that μ and ν coincide on these trace σ-algebras. This implies that μ and ν coincide globally. □

10.3.3 A set which is not a Lebesgue set

In this section we shall construct a set which is a subset of $(0,1]$ but not an element of

$$\mathcal{B}((0,1]) := \{B = A \cap (0,1] : A \in \mathcal{B}(\mathbb{R})\}.$$

Before we start we need

Given $x, y \in (0,1]$ and $A \subseteq (0,1]$, we also need the addition modulo one

$$x \oplus y := \begin{cases} x + y & \text{if } x + y \in (0,1] \\ x + y - 1 & \text{otherwise} \end{cases}$$

and

$$A \oplus x := \{a \oplus x : \ a \in A\}.$$

Now define

$$\begin{aligned} \mathcal{L} := \{ & A \in \mathcal{B}((0,1]) : \\ & A \oplus x \in \mathcal{B}((0,1]) \text{ and } \lambda(A \oplus x) = \lambda(A) \text{ for all } x \in (0,1]\} \end{aligned} \tag{10.1}$$

where λ is the Lebesgue measure on $(0,1]$.

Lemma 10.3.1. *The system $\mathcal{L}$ from (10.1) is a λ-system.*

Proof. The property (1) is clear since $\Omega \oplus x = \Omega$. To check (2) let $A, B \in \mathcal{L}$ and $A \subseteq B$, so that

$$\lambda(A \oplus x) = \lambda(A) \quad \text{and} \quad \lambda(B \oplus x) = \lambda(B).$$

We have to show that $B \backslash A \in \mathcal{L}$. By the definition of $\oplus$ it is easy to see that $A \subseteq B$ implies $A \oplus x \subseteq B \oplus x$ and

$$(B \oplus x) \backslash (A \oplus x) = (B \backslash A) \oplus x,$$

and therefore, $(B\backslash A) \oplus x \in \mathcal{B}((0,1])$. Since λ is a probability measure it follows

$$\begin{aligned}\lambda(B\backslash A) &= \lambda(B) - \lambda(A) \\ &= \lambda(B \oplus x) - \lambda(A \oplus x) \\ &= \lambda((B \oplus x)\backslash(A \oplus x)) \\ &= \lambda((B\backslash A) \oplus x)\end{aligned}$$

and $B\backslash A \in \mathcal{L}$. Property (3) is left as an exercise. □

Finally, we need the axiom of choice.

Proposition 10.3.2. *There exists a subset $H \subseteq (0,1]$ which does not belong to $\mathcal{B}((0,1])$.*

Proof. We take the system $\mathcal{L}$ from (10.1). If $(a,b] \subseteq [0,1]$, then $(a,b] \in \mathcal{L}$. Since

$$\mathcal{P} := \{(a,b] : 0 \leqslant a < b \leqslant 1\}$$

is a π-system which generates $\mathcal{B}((0,1])$ it follows by the π-λ-Theorem (Proposition 10.3.5) that

$$\mathcal{B}((0,1]) \subseteq \mathcal{L}.$$

Let us define the equivalence relation

$$x \sim y \quad \text{if and only if} \quad x \oplus r = y \quad \text{for some rational} \quad r \in (0,1].$$

Let $H \subseteq (0,1]$ be consisting of exactly one representative point from each equivalence class (such set exists under the assumption of the axiom of choice). Then $H \oplus r_1$ and $H \oplus r_2$ are disjoint for $r_1 \neq r_2$: if they were not disjoint, then there would exist $h_1 \oplus r_1 \in (H \oplus r_1)$ and $h_2 \oplus r_2 \in (H \oplus r_2)$ with $h_1 \oplus r_1 = h_2 \oplus r_2$. But this implies $h_1 \sim h_2$ and hence $h_1 = h_2$ and $r_1 = r_2$. So it follows that $(0,1]$ is the countable union of disjoint sets

$$(0,1] = \bigcup_{r \in (0,1] \text{ rational}} (H \oplus r).$$

If we assume that $H \in \mathcal{B}((0,1])$ then $\mathcal{B}((0,1]) \subseteq \mathcal{L}$ implies $H \oplus r \in \mathcal{B}((0,1])$ and

$$\lambda((0,1]) = \lambda\left(\bigcup_{r\in(0,1] \text{ rational}} (H \oplus r)\right) = \sum_{r\in(0,1] \text{ rational}} \lambda(H \oplus r).$$

By $\mathcal{B}((0,1]) \subseteq \mathcal{L}$ we have $\lambda(H \oplus r) = \lambda(H) = a \geqslant 0$ for all rational numbers $r \in (0,1]$. Consequently,

$$1 = \lambda((0,1]) = \sum_{r\in(0,1] \text{ rational}} \lambda(H \oplus r) = a + a + \ldots$$

So, the right hand side can either be 0 (if $a = 0$) or ∞ (if $a > 0$). This leads to a contradiction, so $H \notin \mathcal{B}((0,1])$. □

10.3.4 Monotone class theorem for functions

Now we state the *Monotone Class Theorem*. It is a powerful tool by which, for example, measurability assertions can be proved.

Proposition 10.3.3. [Monotone Class Theorem] *Let H be a class of bounded functions from Ω into $\mathbb{R}$ satisfying the following conditions:*

(1) H is a vector space over $\mathbb{R}$ where the natural point-wise operations " $+$ " and " $\cdot$ " are used.

(2) $1\!\!1_\Omega \in H$.

(3) If $f_n \in H$, $f_n \geqslant 0$, and $f_n \uparrow f$, where f is bounded on Ω, then $f \in H$.

Then one has the following: if H contains the indicator function of every set from some π-system I of subsets of Ω, then H contains every bounded $\sigma(I)$-measurable function on Ω.

Proof. □

10.4 More on the convergence in probability

There are alternatives to the metric of KY FAN. We describe one of them in this section.

Definition 10.4.1. *For $f, g \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$ we let*

$$d_I(f,g) := \int_\Omega \frac{|f-g|}{1+|f-g|} d\mathbb{P}(\omega).$$

Proposition 10.4.2. *For $f, g, h \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$ one has that*

(i) $d_I(f,g) = 0$ *if and only if* $\mathbb{P}(f = g) = 1$,

(ii) $d_I(f,g) = d_I(g,f)$,

(iii) $d_I(f,h) \leqslant d_I(f,g) + d_I(g,h)$.

Proof. (i) We have that

$$d_I(f,g) = 0 \Longleftrightarrow \int_\Omega \frac{|f-g|}{1+|f-g|} d\mathbb{P} = 0 \Longleftrightarrow \mathbb{P}\left(\frac{|f-g|}{1+|f-g|} = 0\right) = 1$$

so that $d_I(f,g) = 0$ if and only if $\mathbb{P}(f = g) = 1$. Assertion (ii) follows by definition and (iii) we can verify by

$$\begin{aligned} d_I(f,h) &= \int_\Omega \frac{|f-h|}{1+|f-h|} d\mathbb{P} \\ &\leqslant \int_\Omega \frac{|f-g|+|g-h|}{1+|f-g|+|g-h|} d\mathbb{P} \\ &\leqslant \int_\Omega \frac{|f-g|}{1+|f-g|} d\mathbb{P} + \int_\Omega \frac{|g-h|}{1+|g-h)|} d\mathbb{P}) \\ &= d_I(f,g) + d_I(g,h). \end{aligned}$$

□

Proposition 10.4.3. *For $f_n, f \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$ the following holds true:*

(i) *We have* $d_I(f_n, f) \underset{n}{\longrightarrow} 0$ *if and only if* $f_n \underset{\mathbb{P}}{\longrightarrow} f$.

(ii) *The sequence* $(f_n)_{n=1}^{\infty}$ *is a Cauchy sequence with respect to* d*, i.e. for all* $\varepsilon > 0$ *there is an* $n \geqslant 1$ *such that for* $k, l \geqslant n$ *one has that* $d_I(f_k, f_l) < \varepsilon$*, if and only if* $(f_n)_{n=1}^{\infty}$ *is a Cauchy sequence in probability.*

Proof. (i) The condition $d_I(f_n, f) \underset{n}{\longrightarrow} 0$ implies

$$\int_\Omega \frac{|f-g|}{1+|f-g|} d\mathbb{P} \underset{n}{\longrightarrow} 0,$$

so that by Ĉebyšev's inequality, for $\lambda > 0$,

$$\lambda \mathbb{P}\left(\frac{|f_n - f|}{1+|f_n - f|} > \lambda\right) \leqslant \int_\Omega \frac{|f_n - f|}{1+|f_n - f|} d\mathbb{P} \underset{n}{\longrightarrow} 0.$$

Given $\varepsilon > 0$ we find $\lambda(\varepsilon) > 0$ such that if $|x| > \varepsilon$, then $\frac{|x|}{1+|x|} > \lambda(\varepsilon)$. Hence

$$\mathbb{P}\left(|f_n - f| > \varepsilon\right) \leqslant \mathbb{P}\left(\frac{|f_n - f|}{1+|f_n - f|} > \lambda(\varepsilon)\right) \leqslant \frac{1}{\lambda(\varepsilon)} \int_\Omega \frac{|f_n - f|}{1+|f_n - f|} d\mathbb{P}$$

and $\mathbb{P}\left(|f_n - f| > \varepsilon\right) \underset{n}{\longrightarrow} 0$.

Conversely, for all $\varepsilon > 0$ we have that

$$\begin{aligned} & \int_\Omega \frac{|f_n - f|}{1+|f_n - f|} d\mathbb{P} \\ = & \int_{\{|f_n - f| > \varepsilon\}} \frac{|f_n - f|}{1+|f_n - f|} d\mathbb{P} + \int_{\{|f_n - f| \leqslant \varepsilon\}} \frac{|f_n - f|}{1+|f_n - f|} d\mathbb{P} \\ \leqslant & \ \mathbb{P}\left(|f_n - f| > \varepsilon\}\right) + \frac{\varepsilon}{1+\varepsilon}, \end{aligned}$$

since the function $\frac{x}{1+x} = 1 - \frac{1}{1+x}$ is monotone for $x \geqslant 0$. Given $\theta > 0$, we take $\varepsilon > 0$ such that $\frac{\varepsilon}{1+\varepsilon} \leqslant \frac{\theta}{2}$ and then $n_0 \geqslant 1$ such that $\mathbb{P}\left(|f_n - f| > \varepsilon\}\right) \leqslant \frac{\theta}{2}$ for $n \geqslant n_0$. Hence $d_I(f_n, f) \leqslant \theta$ for $n \geqslant n_0$.

(ii) Assume that $(f_n)_{n=1}^{\infty}$ is a Cauchy sequence with respect to d_I. For $\lambda > 0$ we have that

$$\lambda \mathbb{P}\left(\frac{|f_k - f_l|}{1+|f_k - f_l|} > \lambda\right) \leqslant d_I(f_k, f_l) \leqslant \eta$$

for $k,l \geqslant n(\eta) \geqslant 1$. For $\lambda := \frac{\varepsilon}{1+\varepsilon}$ with $\varepsilon > 0$ this gives that

$$\frac{\varepsilon}{1+\varepsilon}\mathbb{P}\left(|f_k - f_l| > \varepsilon\right) \leqslant \eta$$

for $k,l \geqslant n(\eta) \geqslant 1$. Choosing $\eta = \varepsilon^2$ we end up with

$$\mathbb{P}\left(|f_k - f_l| > \varepsilon\right) \leqslant \varepsilon(1+\varepsilon)$$

for $k,l \geqslant n(\varepsilon^2) \geqslant 1$. Hence $(f_n)_{n=1}^\infty$ is a Cauchy sequence in probability.

Now assume that $(f_n)_{n=1}^\infty$ is a Cauchy sequence in probability. Then for all $\varepsilon > 0$ there exists $n(\varepsilon) \geqslant 1$ such that for all $k,l \geqslant n(\varepsilon)$ $\mathbb{P}\left(|f_k - f_l| > \varepsilon\right) \leqslant \varepsilon$. Consequently,

$$\int_\Omega \frac{|f_k - f_l|}{1+|f_k - f_l|} d\mathbb{P} \leqslant \mathbb{P}(|f_k - f_l| > \varepsilon) + \frac{\varepsilon}{1+\varepsilon} \leqslant \varepsilon + \frac{\varepsilon}{1+\varepsilon}$$

for $k,l \geqslant n(\varepsilon) \geqslant 1$. □

Proposition 10.4.4. *For Lévy-Prokhorov metric we have that*

$$d_{LP}(\mathrm{law}(f), \mathrm{law}(g)) \leqslant d_{KF}(f,g).$$

Proof. □

10.5 The Theorem of Stone and Weierstrass

Proposition 10.5.1 (Stone & Weierstrass)**.** *Assume that $A \subseteq C_0(\mathbb{R}^d;\mathbb{C})$ satisfies the following properties:*

(i) A is a linear space.

(ii) $g_1, g_2 \in A$ implies $g_1 g_2 \in A$.

(iii) $g \in A$ implies $\bar{g} \in A$.

(iv) For all $x_0 \in \mathbb{R}^d$ there is a $g \in A$ such that $g(x_0) \neq 0$.

(v) For all $x_0 \neq x_1$ there is a $g \in A$ such that $g(x_0) \neq g(x_1)$.

Then A is dense, that means that for all $g \in C_0(\mathbb{R}^d;\mathbb{C})$ there exists $g_n \in A$ such that $\lim_n \sup_x |g_n(x) - g(x)| = 0$.

Proof: John B. Conway: A Course in Functional Analysis, Corollary 8.3.

10.6 Some more inequalities

Proposition 10.6.1 (Lévy[2]-Octaviani inequalities)**.** *Assume independent random variables* $\xi_1, \xi_2, \ldots : \Omega \to \mathbb{R}$ *and* $\varepsilon > 0$*. Then one has*

(1) $\mathbb{P}\left(\max_{1 \leqslant k \leqslant n} |f_k| > \varepsilon\right) \leqslant 3 \max_{1 \leqslant k \leqslant n} \mathbb{P}\left(|f_k| > \frac{\varepsilon}{3}\right)$,

(2) $\mathbb{P}\left(\max_{1 \leqslant k \leqslant n} |f_k| > \varepsilon\right) \leqslant 2\mathbb{P}\left(|f_n| > \varepsilon\right)$ *if the sequence* $(\xi_n)_{n=1}^{\infty}$ *is additionally symmetric, that means that for all signs* $\theta_1, \ldots, \theta_n \in \{-1, 1\}$ *the distributions of the vectors* $(\xi_1, \ldots, \xi_n)$ *and* $(\theta_1\xi_1, \ldots, \theta_n\xi_n)$ *coincide.*

10.7 Complex numbers

We identify
$$\mathbb{C} \cong \mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$$
and write
$$\mathbb{C} \ni z = x + iy \cong (x, y) \in \mathbb{R}^2,$$
where $x = Re(z)$ is the real part of z and $y = Im(z)$ is the imaginary part of z. The complex numbers are an extension of the real numbers by using $\mathbb{R}$ as $\mathbb{C}$ with $x \mapsto (x, 0) = x + i \cdot 0$. We recall some definitions:

Addition. For $z_1 = (x_1, y_1)$ and $z_2 = (x_2, y_2)$ we let
$$z_1 + z_2 := (x_1 + x_2, y_1 + y_2) = (x_1 + x_2) + i(y_1 + y_2).$$

Multiplication. For $z_1 = (x_1, y_1)$ and $z_2 = (x_2, y_2)$ we let
$$z_1 z_2 := (x_1x_2 - y_1y_2, x_1y_2 + x_2y_1) = (x_1x_2 - y_1y_2) + i(x_1y_2 + x_2y_1).$$

Remark 10.7.1. (i) If we interpret $i^2 = -1$, we get this formally by
$$(x_1 + iy_1)(x_2 + iy_2) = x_1x_2 + i^2y_1y_2 + i(x_1y_2 + x_2y_1).$$

(ii) If $z_1 = (x_1, 0)$, then $z_1z_2 = (x_1x_2, x_1y_2) = x_1(x_2, y_2)$ and, in the same way, if $z_2 = (x_2, 0)$, then $z_1z_2 = x_2(x_1, y_1)$.

[2]Paul Pierre Lévy, 15/09/1886 (Paris, France) - 15/12/1971 (Paris, France), influenced greatly probability theory, also worked in functional analysis and partial differential equations.

Length of a complex number: If $z = (x, y)$, then $|z| = \sqrt{x^2 + y^2}$.

Conjugate complex number: If $z = (x, y)$, then $\bar{z} := (x, -y) = x - iy$. We have that $z\bar{z} = x^2 + y^2 = |x|^2$.

Polar coordinates: These coordinates are given as (r, φ) where $r \geqslant 0$ and $\varphi \in [0, 2\pi)$ are determined by

$$\begin{aligned} x &= r\cos\varphi, \\ y &= r\sin\varphi. \end{aligned}$$

Note that $r = |z|$ and that φ is not unique whenever $r = 0$.

Now we recall the notion of the complex **exponential function**.

Definition 10.7.2. *For $z \in \mathbb{C}$ we let*

$$e^z := \sum_{n=0}^{\infty} \frac{z^n}{n!},$$

where the convergence is considered with respect to the euclidean metric in $\mathbb{R}^2$.

Proposition 10.7.3. (i) *For all $z_1, z_2 \in \mathbb{C}$ one has $e^{z_1+z_2} = e^{z_1}e^{z_2}$.*

(ii) Euler*'s* [3] *formula: One has $e^{ix} = \cos x + i\sin x$ for $x \in \mathbb{R}$.*

Proof. (i) is an exercise. (ii) follows from

$$e^{ix} = \frac{x^0}{0!} - \frac{x^2}{2!} + \frac{x^4}{4!} - \ldots + i\left(\frac{x^1}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots\right) = \cos x + i\sin x.$$

□

[3] Leonhard Euler 15/04/1707 (Basel, Switzerland) - 18/09/1783 (St Petersburg, Russia), Swiss mathematician.

Complex valued random variables.

Definition 10.7.4. Let $(\Omega, \mathcal{F})$ be a measurable space.

(i) A map $f : \Omega \to \mathbb{C}$ is called *measurable* provided that $f : \Omega \to \mathbb{R}^2$ is measurable, i.e. for $f = (f_1, f_2)$ the maps $f_1, f_2 : \Omega \to \mathbb{R}$ are measurable.

(ii) A random variable $f : \Omega \to \mathbb{C}$ is called *integrable* provided that

$$\int_\Omega |f(\omega)| d\mathbb{P}(\omega) < \infty.$$

In this case we let

$$\int_\Omega f(\omega) d\mathbb{P}(\omega) = \int_\Omega Re\,(f(\omega))\, d\mathbb{P}(\omega) + i \int_\Omega Im\,(f(\omega))\, d\mathbb{P}(\omega).$$

Index

Bibliography

[1] H. Bauer. Volume 26 of *Studies in Mathematics*. De Gruyter, 2001.

[2] E. Behrends, P. Gritzmann, and G. M. Ziegler. *Pi und Co.* Springer, Berlin, Heidelberg, 2008.

[3] Patrick Billingsley. *Probability and Measure.* Wiley, 1995.

[4] L. Breiman. *Probability*, Addison-Wesley , 1968.

[5] Leclerc de Buffon. Essai d'Arithmétique morale. 1777.

[6] R.M. Dudley. *Real Analysis and Probability*, Cambridge, 2002.

[7] W. Feller. *An introduction to probability theory and its applications*, Volumes I and II, Wiley, 1968 and 1966.

[8] A.M. Garsia. A simple proof of E. Hopf's maximal erogodic theorem. *J. Math. Mech.*, pages 381–382, 1965.

[9] C. Geiss and S. Geiss. An introduction to probability theory. Lectures Notes 60, University of Jyväskylä, Department of Mathematics and Statistics. 2009.

[10] S. Geiss. An introduction to probability theory II. Lectures Notes 61, University of Jyväskylä, Department of Mathematics and Statistics. 2009.

[11] J. Jacod. Probability Essentials.

[12] Olav Kallenberg. *Foundations of Modern Probability.* Springer, New York, second edition, 2002.

[13] A. Klenke. *Probability Theory*, Springer.

[14] A. N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933.

[15] Damien Lamberton and Bernard Lapeyre. *Stochastic Calculus Applied to Finance*. . Chapman & Hall, London, 1996.

[16] Walter Rudin. *Real and Complex Analysis*. McGraw-Hill, 1986.

[17] A.N. Širjaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer, 1996.

[18] D. Williams. *Probability with Martingales*, Cambridge, 1991.