An introduction to probability theory

Christel Geiss and Stefan Geiss
Department of Mathematics and Statistics
University of Jyväskylä

June 3, 2009

Contents

1 Probability spaces **9**
- 1.1 Definition of σ-algebras . . . 10
- 1.2 Probability measures . . . 14
- 1.3 Examples of distributions . . . 25
  - 1.3.1 Binomial distribution with parameter $0 < p < 1$ . . . 25
  - 1.3.2 Poisson distribution with parameter $\lambda > 0$ . . . 25
  - 1.3.3 Geometric distribution with parameter $0 < p < 1$ . . . 25
  - 1.3.4 Lebesgue measure and uniform distribution . . . 26
  - 1.3.5 Gaussian distribution on $\mathbb{R}$ with mean $m \in \mathbb{R}$ and variance $\sigma^2 > 0$ . . . 28
  - 1.3.6 Exponential distribution on $\mathbb{R}$ with parameter $\lambda > 0$ . . . 28
  - 1.3.7 Poisson's Theorem . . . 29
- 1.4 A set which is not a Borel set . . . 31

2 Random variables **35**
- 2.1 Random variables . . . 35
- 2.2 Measurable maps . . . 37
- 2.3 Independence . . . 42

3 Integration **45**
- 3.1 Definition of the expected value . . . 45
- 3.2 Basic properties of the expected value . . . 49
- 3.3 Connections to the Riemann-integral . . . 56
- 3.4 Change of variables in the expected value . . . 58
- 3.5 Fubini's Theorem . . . 59
- 3.6 Some inequalities . . . 66
- 3.7 Theorem of Radon-Nikodym . . . 70
- 3.8 Modes of convergence . . . 71

4 Exercises **77**
- 4.1 Probability spaces . . . 77
- 4.2 Random variables . . . 81
- 4.3 Integration . . . 83

Introduction

Probability theory can be understood as a mathematical model for the intuitive notion of *uncertainty*. Without probability theory all the stochastic models in Physics, Biology, and Economics would either not have been developed or would not be rigorous. Also, probability is used in many branches of pure mathematics, even in branches 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 published his modern approach of Probability Theory, including the notion of a measurable space and a probability space. This lecture will start from this notion, to continue with random variables and basic parts of integration theory, and to finish with some first limit theorems. 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 is also used throughout this script.

Let us start with some introducing examples.

Example 1. You stand at a bus-stop knowing that every 30 minutes a bus comes. But you do not know when the last bus came. What is the intuitively *expected* time you have to wait for the next bus? Or you have a light bulb and would like to know how many hours you can expect the light bulb will work? How to treat the above two questions by mathematical tools?

Example 2. Surely many lectures about probability start with rolling a die: you have six possible outcomes $\{1, 2, 3, 4, 5, 6\}$. If you want to roll a certain number, let's say 3, assuming the die is fair (whatever this means) it is intuitively clear that the chance to dice really 3 is 1:6. How to formalize this?

Example 3. Now the situation is a bit more complicated: Somebody has two dice and transmits to you only the sum of the two dice. The set of all outcomes of the experiment is the set

$$\Omega := \{(1,1), ..., (1,6), ..., (6,1), ..., (6,6)\}$$

but what you see is only the result of the map $f : \Omega \to \mathbb{R}$ defined as

$$f((a, b)) := a + b,$$

that means you see $f(\omega)$ but not $\omega = (a, b)$. This is an easy example of a *random variable* f, which we introduce later.

Example 4. In Example 3 we know the set of states $\omega \in \Omega$ explicitly. But this is not always possible nor necessary: Say, that we would like to measure the temperature outside our home. We can do this by an electronic thermometer which consists of a sensor outside and a display, including some electronics, inside. The number we get from the system might not be correct because of several reasons: the electronics is influenced by the inside temperature and the voltage of the power-supply. Changes of these parameters have to be compensated by the electronics, but this can, probably, not be done in a perfect way. Hence, we might not have a systematic error, but some kind of a *random* error which appears as a positive or negative deviation from the exact value. It is impossible to describe all these sources of randomness or uncertainty explicitly.

We denote the exact temperature by T and the displayed temperature by S, so that the difference $T - S$ is influenced by the above sources of uncertainty. If we would measure simultaneously, by using thermometers of the same type, we would get values S_1, S_2, ... with corresponding differences

$$D_1 := T - S_1, \qquad D_2 := T - S_2, \qquad D_3 := T - S_3, ...$$

Intuitively, we get random numbers $D_1, D_2, ...$ having a certain distribution. How to develop an exact mathematical theory out of this?

Firstly, we take an abstract set Ω. Each element $\omega \in \Omega$ will stand for a specific configuration of our sources influencing the measured value. Secondly, we take a function

$$f : \Omega \to \mathbb{R}$$

which gives for all $\omega \in \Omega$ the difference $f(\omega) = T - S(\omega)$. From properties of this function we would like to get useful information of our thermometer and, in particular, about the correctness of the displayed values.

To put Examples 3 and 4 on a solid ground we go ahead with the following questions:

Step 1: How to model the randomness of ω, or how likely an ω is? We do this by introducing the probability spaces in Chapter 1.

Step 2: What mathematical properties of f we need to transport the randomness from ω to $f(\omega)$? This yields to the introduction of the random variables in Chapter 2.

STEP 3: What are properties of f which might be important to know in practice? For example the mean-value and the variance, denoted by

$$\mathbb{E}f \quad \text{and} \quad \mathbb{E}(f - \mathbb{E}f)^2.$$

If the first expression is zero, then in Example 3 the calibration of the thermometer is right, if the second one is small the displayed values are very likely close to the real temperature. To define these quantities one needs the integration theory developed in Chapter 3.

STEP 4: Is it possible to describe the distribution of the values f may take? Or before, what do we mean by a distribution? Some basic distributions are discussed in Section 1.3.

STEP 5: What is a good method to estimate $\mathbb{E}f$? We can take a sequence of *independent* (take this intuitive for the moment) random variables $f_1, f_2, ...$, having the same distribution as f, and expect that

$$\frac{1}{n}\sum_{i=1}^{n} f_i(\omega) \quad \text{and} \quad \mathbb{E}f$$

are close to each other. This lieds us to the *Strong Law of Large Numbers* discussed in Section 3.8.

Notation. Given a set Ω and subsets $A, B \subseteq \Omega$, then the following notation is used:

intersection:	$A \cap B$	=	$\{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$
union:	$A \cup B$	=	$\{\omega \in \Omega : \omega \in A \text{ or (or both) } \omega \in B\}$
set-theoretical minus:	$A \backslash B$	=	$\{\omega \in \Omega : \omega \in A \text{ and } \omega \notin B\}$
complement:	A^c	=	$\{\omega \in \Omega : \omega \notin A\}$
empty set:	$\emptyset$	=	set, without any element
real numbers:	$\mathbb{R}$		
natural numbers:	$\mathbb{N}$	=	$\{1, 2, 3, ...\}$
rational numbers:	$\mathbb{Q}$		
indicator-function:	$\mathbb{1}_A(x)$	=	$\begin{cases} 1 \text{ if } x \in A \\ 0 \text{ if } x \notin A \end{cases}$

Given real numbers α, β, we use $\alpha \wedge \beta := \min\{\alpha, \beta\}$.

Chapter 1

Probability spaces

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

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

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

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

(b) If we flip a coin, then we have either "heads" or "tails" on top, that means

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

If we have two coins, then

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

is the set of all possible outcomes.

(c) For the lifetime of a bulb in hours we can choose

$$\Omega = [0, \infty).$$

(2) A **σ-algebra** $\mathcal{F}$, which is the system of *observable* subsets or *events* $A \subseteq \Omega$. The interpretation is that one can usually not decide whether a system is in the particular state $\omega \in \Omega$, but one can decide whether $\omega \in A$ or $\omega \notin A$.

(3) A **measure** $\mathbb{P}$, which gives a **probability** to all $A \in \mathcal{F}$. This probability is a number $\mathbb{P}(A) \in [0, 1]$ that describes how likely it is that the event A occurs.

For the formal mathematical approach we proceed in two steps: in the first step we define the σ-algebras $\mathcal{F}$, here we do not need any measure. In the second step we introduce the measures.

1.1 Definition of σ-algebras

The σ-algebra is a basic tool in probability theory. It is the set the probability measures are defined on. Without this notion it would be impossible to consider the fundamental Lebesgue measure on the interval $[0,1]$ or to consider Gaussian measures, without which many parts of mathematics can 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**. The elements $A \in \mathcal{F}$ are called **events**. An event A occurs if $\omega \in A$ and it does not occur if $\omega \notin A$.
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**.

Every σ-algebra is an algebra. Sometimes, the terms σ-field and field are used instead of σ-algebra and algebra. We consider some first examples.

Example 1.1.2 (a) The largest σ-algebra on Ω: if $\mathcal{F} = 2^{\Omega}$ is the system of *all* subsets $A \subseteq \Omega$, then $\mathcal{F}$ is a σ-algebra.

(b) The smallest σ-algebra: $\mathcal{F} = \{\Omega, \emptyset\}$.

(c) If $A \subseteq \Omega$, then $\mathcal{F} = \{\Omega, \emptyset, A, A^c\}$ is a σ-algebra.

Some more concrete examples are the following:

Example 1.1.3 (a) Assume a model for a die, i.e. $\Omega := \{1, ..., 6\}$ and $\mathcal{F} := 2^{\Omega}$. The event "the die shows an even number" is described by

$$A = \{2, 4, 6\}.$$

(b) Assume a model with two dice, i.e. $\Omega := \{(a, b) : a, b = 1, ..., 6\}$ and $\mathcal{F} := 2^{\Omega}$. The event "the sum of the two dice equals four" is described by

$$A = \{(1, 3), (2, 2), (3, 1)\}.$$

(*c*) 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 modeled via

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

(*d*) Assume that we want to model the lifetime of a bulb, so that $\Omega := [0,\infty)$. "The bulb works more than 200 hours" we express by

$$A = (200,\infty).$$

But what is the right σ-algebra in this case? It is **not** 2^{Ω} which would be too big.

If $\Omega = \{\omega_1,...,\omega_n\}$, 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 1.1.4 [ALGEBRA, WHICH IS NOT A σ-ALGEBRA] Let $\mathcal{G}$ be the system of subsets $A \subseteq \mathbb{R}$ such that A can be written as

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

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

Unfortunately, most of the important σ–algebras can not be constructed explicitly. Surprisingly, one can work practically with them nevertheless. In the following we describe a simple procedure which generates σ–algebras. We start with the fundamental

Proposition 1.1.5 [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. The proof is very easy, but typical and fundamental. 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 ($\mathcal{F}_j$ are σ–algebras!)

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

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

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

□

Proposition 1.1.6 [SMALLEST σ-ALGEBRA CONTAINING A SET-SYSTEM] *Let Ω be an arbitrary non-empty set and $\mathcal{G}$ be an arbitrary system of subsets $A \subseteq \Omega$. Then there exists a smallest σ-algebra $\sigma(\mathcal{G})$ on Ω such that*

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

Proof. We let

$$J := \{\mathcal{C} \text{ is a } \sigma\text{–algebra on } \Omega \text{ such that } \mathcal{G} \subseteq \mathcal{C}\}.$$

According to Example 1.1.2 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 1.1.5 such that (by construction) $\mathcal{G} \subseteq \sigma(\mathcal{G})$. 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

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

□

The construction is very elegant but has, as already mentioned, the slight disadvantage that one cannot construct all elements of $\sigma(\mathcal{G})$ explicitly. Let us now turn to one of the most important examples, the Borel σ-algebra on $\mathbb{R}$. To do this we need the notion of open and closed sets.

Definition 1.1.7 [OPEN AND CLOSED SETS]

(1) A subset $A \subseteq \mathbb{R}$ is called **open**, if for each $x \in A$ there is an $\varepsilon > 0$ such that $(x - \varepsilon, x + \varepsilon) \subseteq A$.

(2) A subset $B \subseteq \mathbb{R}$ is called **closed**, if $A := \mathbb{R} \backslash B$ is open.

Given $-\infty \leq a \leq b \leq \infty$, the interval (a, b) is open and the interval $[a, b]$ is closed. Moreover, by definition the empty set $\emptyset$ is open and closed.

Proposition 1.1.8 [GENERATION OF THE BOREL σ-ALGEBRA ON $\mathbb{R}$] *We let*

$\mathcal{G}_0$ be the system of all open subsets of $\mathbb{R}$,
$\mathcal{G}_1$ be the system of all closed subsets of $\mathbb{R}$,
$\mathcal{G}_2$ be the system of all intervals $(-\infty, b]$, $b \in \mathbb{R}$,
$\mathcal{G}_3$ be the system of all intervals $(-\infty, b)$, $b \in \mathbb{R}$,
$\mathcal{G}_4$ be the system of all intervals $(a, b]$, $-\infty < a < b < \infty$,
$\mathcal{G}_5$ be the system of all intervals (a, b), $-\infty < a < b < \infty$.

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

Definition 1.1.9 [BOREL σ-ALGEBRA ON $\mathbb{R}$] [1] The σ-algebra constructed in Proposition 1.1.8 is called **Borel** σ-algebra and denoted by $\mathcal{B}(\mathbb{R})$.

In the same way one can introduce BOREL σ-algebras on metric spaces: Given a metric space M with metric d a set $A \subseteq M$ is open provided that for all $x \in A$ there is a $\varepsilon > 0$ such that $\{y \in M : d(x,y) < \varepsilon\} \subseteq A$. A set $B \subseteq M$ is closed if the complement $M \backslash B$ is open. The BOREL σ-algebra $\mathcal{B}(M)$ is the smallest σ-algebra that contains all open (closed) subsets of M.

Proof of Proposition 1.1.8. We only show that

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

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 bounded non-empty open set $A \subseteq \mathbb{R}$. For all $x \in A$ there is a *maximal* $\varepsilon_x > 0$ 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. □

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

1.2 Probability measures

Now we introduce the measures we are going to use:

Definition 1.2.1 [PROBABILITY MEASURE, PROBABILITY SPACE] Let $(\Omega, \mathcal{F})$ be a measurable space.

(1) A map $\mathbb{P} : \mathcal{F} \to [0,1]$ is called **probability measure** if $\mathbb{P}(\Omega) = 1$ and for all $A_1, A_2, ... \in \mathcal{F}$ with $A_i \cap A_j = \emptyset$ for $i \neq j$ one has

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

The triplet $(\Omega, \mathcal{F}, \mathbb{P})$ is called **probability space**.

(2) A map $\mu : \mathcal{F} \to [0, \infty]$ is called **measure** if $\mu(\emptyset) = 0$ and for all $A_1, A_2, ... \in \mathcal{F}$ with $A_i \cap A_j = \emptyset$ for $i \neq j$ one has

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

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

(3) A measure space $(\Omega, \mathcal{F}, \mu)$ or a measure μ is called **σ-finite** provided that there are $\Omega_k \subseteq \Omega$, $k = 1, 2, ...$, such that

(a) $\Omega_k \in \mathcal{F}$ for all $k = 1, 2, ...$,

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

(c) $\Omega = \bigcup_{k=1}^{\infty} \Omega_k$,

(d) $\mu(\Omega_k) < \infty$.

The measure space $(\Omega, \mathcal{F}, \mu)$ or the measure μ are called **finite** if $\mu(\Omega) < \infty$.

Remark 1.2.2 Of course, any probability measure is a finite measure: We need only to check $\mu(\emptyset) = 0$ which follows from $\mu(\emptyset) = \sum_{i=1}^{\infty} \mu(\emptyset)$ (note that $\emptyset \cap \emptyset = \emptyset$) and $\mu(\Omega) < \infty$.

Example 1.2.3 (*a*) 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,

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

(b) 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

$$\mathbb{P}(\{\omega\}) := \frac{1}{4}.$$

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 1.2.4 [Dirac and counting measure] [2]

(a) **Dirac measure**: For $\mathcal{F} = 2^\Omega$ and a fixed $x_0 \in \Omega$ we let

$$\delta_{x_0}(A) := \begin{cases} 1 & : \quad x_0 \in A \\ 0 & : \quad x_0 \notin A \end{cases}.$$

(b) **Counting measure**: Let $\Omega := \{\omega_1, ..., \omega_N\}$ and $\mathcal{F} = 2^\Omega$. Then

$$\mu(A) := \text{cardinality of } A.$$

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

Example 1.2.5 Assume that there are n communication channels between the points A and B. Each of the channels has a communication rate of $\rho > 0$ (say ρ bits per second), which yields to the communication rate ρk, in case k channels are used. Each of the channels fails with probability p, so that we have a random communication rate $R \in \{0, \rho, ..., n\rho\}$. What is the *right* model for this? We use

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

with the interpretation: $\varepsilon_i = 0$ if channel i is failing, $\varepsilon_i = 1$ if channel i is working. $\mathcal{F}$ consists of all unions of

$$A_k := \{\omega \in \Omega : \varepsilon_1 + \cdots + \varepsilon_n = k\}.$$

Hence A_k consists of all ω such that the communication rate is ρk. The system $\mathcal{F}$ is the system of observable sets of events since one can only observe how many channels are failing, but not which channel fails. The measure $\mathbb{P}$ is given by

$$\mathbb{P}(A_k) := \binom{n}{k} p^{n-k}(1-p)^k, \qquad 0 < p < 1.$$

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

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

We continue with some basic properties of a probability measure.

Proposition 1.2.6 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then the following assertions are true:*

(1) *If $A_1, ..., A_n \in \mathcal{F}$ such that $A_i \cap A_j = \emptyset$ if $i \neq j$, then $\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i)$.*

(2) *If $A, B \in \mathcal{F}$, then $\mathbb{P}(A \backslash B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$.*

(3) *If $B \in \mathcal{F}$, then $\mathbb{P}(B^c) = 1 - \mathbb{P}(B)$.*

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

(5) CONTINUITY FROM BELOW*: If $A_1, A_2, ... \in \mathcal{F}$ such that $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$, then*
$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n=1}^\infty A_n\right).$$

(6) CONTINUITY FROM ABOVE*: If $A_1, A_2, ... \in \mathcal{F}$ such that $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$, then*
$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcap_{n=1}^\infty A_n\right).$$

Proof. (1) We let $A_{n+1} = A_{n+2} = \cdots = \emptyset$, so that
$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty \mathbb{P}(A_i) = \sum_{i=1}^n \mathbb{P}(A_i),$$
because of $\mathbb{P}(\emptyset) = 0$.
(2) Since $(A \cap B) \cap (A \backslash B) = \emptyset$, we get that
$$\mathbb{P}(A \cap B) + \mathbb{P}(A \backslash B) = \mathbb{P}((A \cap B) \cup (A \backslash B)) = \mathbb{P}(A).$$
(3) We apply (2) to $A = \Omega$ and observe that $\Omega \backslash B = B^c$ by definition and $\Omega \cap B = B$.
(4) Put $B_1 := A_1$ and $B_i := A_1^c \cap A_2^c \cap \cdots \cap A_{i-1}^c \cap A_i$ for $i = 2, 3, \ldots$ Obviously, $\mathbb{P}(B_i) \leq \mathbb{P}(A_i)$ for all i. Since the B_i's are disjoint and $\bigcup_{i=1}^\infty A_i = \bigcup_{i=1}^\infty B_i$ it follows
$$\mathbb{P}\left(\bigcup_{i=1}^\infty A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^\infty B_i\right) = \sum_{i=1}^\infty \mathbb{P}(B_i) \leq \sum_{i=1}^\infty \mathbb{P}(A_i).$$
(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
$$\bigcup_{n=1}^\infty B_n = \bigcup_{n=1}^\infty A_n \quad \text{and} \quad B_i \cap B_j = \emptyset$$

for $i \neq j$. Consequently,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(B_n) = \lim_{N\to\infty} \sum_{n=1}^{N} \mathbb{P}(B_n) = \lim_{N\to\infty} \mathbb{P}(A_N)$$

since $\bigcup_{n=1}^{N} B_n = A_N$. (6) is an exercise. □

The sequence $a_n = (-1)^n$, $n = 1, 2, \ldots$ does not converge, i.e. the limit of $(a_n)_{n=1}^{\infty}$ does not exist. But the limit superior and the limit inferior for a given sequence of real numbers exists always.

Definition 1.2.7 [$\liminf_n a_n$ AND $\limsup_n a_n$] For $a_1, a_2, ... \in \mathbb{R}$ we let

$$\liminf_n a_n := \lim_n \inf_{k\geq n} a_k \quad \text{and} \quad \limsup_n a_n := \lim_n \sup_{k\geq n} a_k.$$

Remark 1.2.8 (1) The value $\liminf_n a_n$ is the infimum of all c such that there is a subsequence $n_1 < n_2 < n_3 < \cdots$ such that $\lim_k a_{n_k} = c$.

(2) The value $\limsup_n a_n$ is the supremum of all c such that there is a subsequence $n_1 < n_2 < n_3 < \cdots$ such that $\lim_k a_{n_k} = c$.

(3) By definition one has that

$$-\infty \leq \liminf_n a_n \leq \limsup_n a_n \leq \infty.$$

Moreover, if $\liminf_n a_n = \limsup_n a_n = a \in \mathbb{R}$, then $\lim_n a_n = a$.

(4) For example, taking $a_n = (-1)^n$, gives

$$\liminf_n a_n = -1 \quad \text{and} \quad \limsup_n a_n = 1.$$

As we will see below, also for a sequence of sets one can define a limit superior and a limit inferior.

Definition 1.2.9 [$\liminf_n A_n$ AND $\limsup_n A_n$] Let $(\Omega, \mathcal{F})$ be a measurable space and $A_1, A_2, ... \in \mathcal{F}$. Then

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

The definition above says that $\omega \in \liminf_n A_n$ if and only if all events A_n, except a finite number of them, occur, and that $\omega \in \limsup_n A_n$ if and only if infinitely many of the events A_n occur.

Proposition 1.2.10 [LEMMA OF FATOU] [3] *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A_1, A_2, ... \in \mathcal{F}$. Then*

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

The proposition will be deduced from Proposition 3.2.6.

Remark 1.2.11 If $\liminf_n A_n = \limsup_n A_n = A$, then the Lemma of Fatou gives that

$$\liminf_n \mathbb{P}(A_n) = \limsup_n \mathbb{P}(A_n) = \lim_n \mathbb{P}(A_n) = \mathbb{P}(A).$$

Examples for such systems are obtained by assuming $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$ or $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$.

Now we turn to the fundamental notion of independence.

Definition 1.2.12 [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 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. Independence can be also expressed through conditional probabilities. Let us first define them:

Definition 1.2.13 [CONDITIONAL PROBABILITY] Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$. Then

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

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

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

It is now obvious that A and B are independent if and only if $\mathbb{P}(B|A) = \mathbb{P}(B)$. An important place of the conditional probabilities in Statistics is guaranteed by Bayes' formula. Before we formulate this formula in Proposition 1.2.15 we consider $A, B \in \mathcal{F}$, with $0 < \mathbb{P}(B) < 1$ and $\mathbb{P}(A) > 0$. Then

$$A = (A \cap B) \cup (A \cap B^c),$$

where $(A \cap B) \cap (A \cap B^c) = \emptyset$, and therefore,

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) \\ &= \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c). \end{aligned}$$

This implies

$$\begin{aligned} \mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} &= \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)} \\ &= \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)}. \end{aligned}$$

Example 1.2.14 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 we get

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

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

Proposition 1.2.15 [BAYES' FORMULA] [4] *Assume $A, B_j \in \mathcal{F}$, with $\Omega = \bigcup_{j=1}^n B_j$, where $B_i \cap B_j = \emptyset$ for $i \neq j$ and $\mathbb{P}(A) > 0$, $\mathbb{P}(B_j) > 0$ for $j = 1, \ldots, n$. Then*

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

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

The proof is an exercise. 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.

Now we continue with the fundamental Lemma of Borel-Cantelli.

Proposition 1.2.16 [Lemma of Borel-Cantelli] [5] *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A_1, A_2, ... \in \mathcal{F}$. Then one has the following:*

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

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

Proof. (1) It holds by definition $\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$. By

$$\bigcup_{k=n+1}^{\infty} A_k \subseteq \bigcup_{k=n}^{\infty} A_k$$

and the continuity of $\mathbb{P}$ from above (see Proposition 1.2.6) we get that

$$\begin{aligned} \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) &= \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) \\ &= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \\ &\leq \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathbb{P}\left(A_k\right) = 0, \end{aligned}$$

where the last inequality follows again from Proposition 1.2.6.
(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

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

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

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 $A_1^c, A_2^c, ...$, we finally get (setting $p_n := \mathbb{P}(A_n)$) that

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

where we have used that $1 - x \leq e^{-x}$. □

Although the definition of a measure is not difficult, to prove existence and uniqueness of measures may sometimes be difficult. The problem lies in the fact that, in general, the σ-algebras are not constructed explicitly, one only knows their existence. To overcome this difficulty, one usually exploits

Proposition 1.2.17 [Carathéodory's extension theorem] [6]
Let Ω be a non-empty set and $\mathcal{G}$ an algebra *on Ω such that*

$$\mathcal{F} := \sigma(\mathcal{G}).$$

Assume that $\mathbb{P}_0 : \mathcal{G} \to [0, \infty)$ satisfies:

(1) $\mathbb{P}_0(\Omega) < \infty$.

(2) *If $A_1, A_2, ... \in \mathcal{G}$, $A_i \cap A_j = \emptyset$ for $i \neq j$,* **and** $\bigcup_{i=1}^{\infty} A_i \in \mathcal{G}$, *then*

$$\mathbb{P}_0\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mathbb{P}_0(A_i).$$

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

Then there exists a unique finite measure $\mathbb{P}$ *on* $\mathcal{F}$ *such that*

$$\mathbb{P}(A) = \mathbb{P}_0(A) \quad \textit{for all} \quad A \in \mathcal{G}.$$

Proof. See [3] (Theorem 3.1). □

As an application we construct (more or less without rigorous proof) the product space

$$(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mathbb{P}_1 \times \mathbb{P}_2)$$

of two probability spaces $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$. We do this as follows:

(1) $\Omega_1 \times \Omega_2 := \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$.

(2) $\mathcal{F}_1 \otimes \mathcal{F}_2$ is the smallest σ-algebra on $\Omega_1 \times \Omega_2$ which contains all sets of type

$$A_1 \times A_2 := \{(\omega_1, \omega_2) : \omega_1 \in A_1, \omega_2 \in A_2\} \quad \text{with} \quad A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2.$$

(3) As algebra $\mathcal{G}$ we take all sets of type

$$A := \left(A_1^1 \times A_2^1\right) \cup \cdots \cup \left(A_1^n \times A_2^n\right)$$

with $A_1^k \in \mathcal{F}_1$, $A_2^k \in \mathcal{F}_2$, and $\left(A_1^i \times A_2^i\right) \cap \left(A_1^j \times A_2^j\right) = \emptyset$ for $i \neq j$. Finally, we define $\mathbb{P}_0 : \mathcal{G} \to [0,1]$ by

$$\mathbb{P}_0\left(\left(A_1^1 \times A_2^1\right) \cup \cdots \cup \left(A_1^n \times A_2^n\right)\right) := \sum_{k=1}^{n} \mathbb{P}_1(A_1^k)\mathbb{P}_2(A_2^k).$$

Proposition 1.2.18 *The system* $\mathcal{G}$ *is an algebra. The map* $\mathbb{P}_0 : \mathcal{G} \to [0,1]$ *is correctly defined and satisfies the assumptions of Carathéodory's extension theorem Proposition 1.2.17.*

Proof. (i) Assume

$$A = \left(A_1^1 \times A_2^1\right) \cup \cdots \cup \left(A_1^n \times A_2^n\right) = \left(B_1^1 \times B_2^1\right) \cup \cdots \cup \left(B_1^m \times B_2^m\right)$$

where the $(A_1^k \times A_2^k)_{k=1}^n$ and the $(B_1^l \times B_2^l)_{l=1}^m$ are pair-wise disjoint, respectively. We find partitions $C_1^1, ..., C_1^{N_1}$ of Ω_1 and $C_2^1, ..., C_2^{N_2}$ of Ω_2 so that A_i^k and B_i^l can be represented as disjoint unions of the sets C_i^r, $r = 1, ..., N_i$. Hence there is a representation

$$A = \bigcup_{(r,s)\in I} (C_1^r \times C_2^s)$$

for some $I \subseteq \{1, ..., N_1\} \times \{1, ..., N_2\}$. By drawing a picture and using that $\mathbb{P}_1$ and $\mathbb{P}_2$ are measures one observes that

$$\sum_{k=1}^{n} \mathbb{P}_1(A_1^k)\mathbb{P}_2(A_2^k) = \sum_{(r,s)\in I} \mathbb{P}_1(C_1^r)\mathbb{P}_2(C_2^s) = \sum_{l=1}^{m} \mathbb{P}_1(B_1^l)\mathbb{P}_2(B_2^l).$$

(ii) To check that $\mathbb{P}_0$ is σ-additive on $\mathcal{G}$ it is sufficient to prove the following: For $A_i, A_i^k \in \mathcal{F}_i$ with

$$A_1 \times A_2 = \bigcup_{k=1}^{\infty} (A_1^k \times A_2^k)$$

and $(A_1^k \times A_2^k) \cap (A_1^l \times A_2^l) = \emptyset$ for $k \neq l$ one has that

$$\mathbb{P}_1(A_1)\mathbb{P}_2(A_2) = \sum_{k=1}^{\infty} \mathbb{P}_1(A_1^k)\mathbb{P}_2(A_2^k).$$

Since the inequality

$$\mathbb{P}_1(A_1)\mathbb{P}_2(A_2) \geq \sum_{k=1}^{N} \mathbb{P}_1(A_1^k)\mathbb{P}_2(A_2^k)$$

can be easily seen for all $N = 1, 2, ...$ by an argument like in step (i) we concentrate ourselves on the converse inequality

$$\mathbb{P}_1(A_1)\mathbb{P}_2(A_2) \leq \sum_{k=1}^{\infty} \mathbb{P}_1(A_1^k)\mathbb{P}_2(A_2^k).$$

We let

$$\varphi(\omega_1) := \sum_{n=1}^{\infty} \mathbb{1}_{\{\omega_1 \in A_1^n\}} \mathbb{P}_2(A_2^n) \quad \text{and} \quad \varphi_N(\omega_1) := \sum_{n=1}^{N} \mathbb{1}_{\{\omega_1 \in A_1^n\}} \mathbb{P}_2(A_2^n)$$

for $N \geq 1$, so that $0 \leq \varphi_N(\omega_1) \uparrow_N \varphi(\omega_1) = \mathbb{1}_{A_1}(\omega_1)\mathbb{P}_2(A_2)$. Let $\varepsilon \in (0,1)$ and

$$B_\varepsilon^N := \{\omega_1 \in \Omega_1 : (1-\varepsilon)\mathbb{P}_2(A_2) \leq \varphi_N(\omega_1)\} \in \mathcal{F}_1.$$

The sets B_ε^N are non-decreasing in N and $\bigcup_{N=1}^{\infty} B_\varepsilon^N = A_1$ so that

$$(1-\varepsilon)\mathbb{P}_1(A_1)\mathbb{P}_2(A_2) \;=\; \lim_N (1-\varepsilon)\mathbb{P}_1(B_\varepsilon^N)\mathbb{P}_2(A_2).$$

Because $(1-\varepsilon)\mathbb{P}_2(A_2) \leq \varphi_N(\omega_1)$ for all $\omega_1 \in B_\varepsilon^N$ one gets (after some calculation...) $(1-\varepsilon)\mathbb{P}_2(A_2)\mathbb{P}_1(B_\varepsilon^N) \leq \sum_{k=1}^{N} \mathbb{P}_1(A_1^k)\mathbb{P}_2(A_2^k)$ and therefore

$$\lim_N (1-\varepsilon)\mathbb{P}_1(B_\varepsilon^N)\mathbb{P}_2(A_2) \leq \lim_N \sum_{k=1}^{N} \mathbb{P}_1(A_1^k)\mathbb{P}_2(A_2^k) = \sum_{k=1}^{\infty} \mathbb{P}_1(A_1^k)\mathbb{P}_2(A_2^k).$$

Since $\varepsilon \in (0,1)$ was arbitrary, we are done. □

Definition 1.2.19 [PRODUCT OF PROBABILITY SPACES] The extension of $\mathbb{P}_0$ to $\mathcal{F}_1 \otimes \mathcal{F}_2$ according to Proposition 1.2.17 is called **product measure** and denoted by $\mathbb{P}_1 \times \mathbb{P}_2$. The probability space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mathbb{P}_1 \times \mathbb{P}_2)$ is called **product probability space**.

One can prove that

$$(\mathcal{F}_1 \otimes \mathcal{F}_2) \otimes \mathcal{F}_3 = \mathcal{F}_1 \otimes (\mathcal{F}_2 \otimes \mathcal{F}_3) \text{ and } (\mathbb{P}_1 \times \mathbb{P}_2) \times \mathbb{P}_3 = \mathbb{P}_1 \times (\mathbb{P}_2 \times \mathbb{P}_3),$$

which we simply denote by

$$(\Omega_1 \times \Omega_2 \times \Omega_3, \mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \mathcal{F}_3, \mathbb{P}_1 \times \mathbb{P}_2 \times \mathbb{P}_3).$$

Iterating this procedure, we can define *finite* products

$$(\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n, \mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \cdots \otimes \mathcal{F}_n, \mathbb{P}_1 \times \mathbb{P}_2 \times \cdots \times \mathbb{P}_n)$$

by iteration. The case of *infinite* products requires more work. Here the interested reader is referred, for example, to [5] where a special case is considered.

Now we define the the Borel σ-algebra on $\mathbb{R}^n$.

Definition 1.2.20 For $n \in \{1, 2, ...\}$ we let

$$\mathcal{B}(\mathbb{R}^n) := \sigma\left((a_1, b_1) \times \cdots \times (a_n, b_n) : a_1 < b_1, ..., a_n < b_n\right).$$

Letting $|x - y| := (\sum_{k=1}^n |x_k - y_k|^2)^{\frac{1}{2}}$ for $x = (x_1, ..., x_n) \in \mathbb{R}^n$ and $y = (y_1, ..., y_n) \in \mathbb{R}^n$, we say that a set $A \subseteq \mathbb{R}^n$ is open provided that for all $x \in A$ there is an $\varepsilon > 0$ such that

$$\{y \in \mathbb{R}^n : |x - y| < \varepsilon\} \subseteq A.$$

As in the case $n = 1$ one can show that $\mathcal{B}(\mathbb{R}^n)$ is the smallest σ-algebra which contains all open subsets of $\mathbb{R}^n$. Regarding the above product spaces there is

Proposition 1.2.21 $\mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R})$.

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 1.2.22 [π-SYSTEM] A system $\mathcal{G}$ of subsets $A \subseteq \Omega$ is called **π-system**, provided that

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

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

Proposition 1.2.23 *Let $(\Omega, \mathcal{F})$ be a measurable space with $\mathcal{F} = \sigma(\mathcal{G})$, where $\mathcal{G}$ is a π-system. Assume two probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ on $\mathcal{F}$ such that*

$$\mathbb{P}_1(A) = \mathbb{P}_2(A) \quad \textit{for all} \quad A \in \mathcal{G}.$$

Then $\mathbb{P}_1(B) = \mathbb{P}_2(B)$ for all $B \in \mathcal{F}$.

1.3 Examples of distributions

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

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

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

(3) $\mathbb{P}(B) = \mu_{n,p}(B) := \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta_k(B)$, where δ_k is the Dirac measure introduced in Example 1.2.4.

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

1.3.2 Poisson distribution with parameter $\lambda > 0$

(1) $\Omega := \{0, 1, 2, 3, ...\}$.

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

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

The Poisson distribution [7] is used, for example, to model stochastic processes with a continuous time parameter and jumps: the probability that the process jumps k times between the time-points s and t with $0 \leq s < t < \infty$ is equal to $\pi_{\lambda(t-s)}(\{k\})$.

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

(1) $\Omega := \{0, 1, 2, 3, ...\}$.

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

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

INTERPRETATION: The probability that an electric light bulb breaks down is $p \in (0, 1)$. The bulb does not have a "memory", that means the break down is independent of the time the bulb has been already switched on. So, we get the following model: at day 0 the probability of breaking down is p. If the bulb survives day 0, it breaks down again with probability p at the first day so that the total probability of a break down at day 1 is $(1-p)p$. If we continue in this way we get that breaking down at day k has the probability $(1-p)^k p$.

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

1.3.4 Lebesgue measure and uniform distribution

Using CARATHÉODORY'S extension theorem, we first construct the Lebesgue measure on the intervals $(a, b]$ with $-\infty < a < b < \infty$. For this purpose we let

(1) $\Omega := (a, b]$,

(2) $\mathcal{F} = \mathcal{B}((a, b]) := \sigma((c, d] : a \leq c < d \leq b)$,

(3) As generating algebra $\mathcal{G}$ for $\mathcal{B}((a, b])$ we take the system of subsets $A \subseteq (a, b]$ such that A can be written as

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

where $a \leq a_1 \leq b_1 \leq \cdots \leq a_n \leq b_n \leq b$. For such a set A we let

$$\mathbb{P}_0(A) := \frac{1}{b-a} \sum_{i=1}^{n} (b_i - a_i).$$

Proposition 1.3.1 *The system $\mathcal{G}$ is an algebra. The map $\mathbb{P}_0 : \mathcal{G} \to [0, 1]$ is correctly defined and satisfies the assumptions of Carathéodory's extension theorem Proposition 1.2.17.*

Proof. For notational simplicity we let $a = 0$ and $b = 1$. After a standard reduction (check!) we have to show the following: given $0 \leq a \leq b \leq 1$ 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 there is a finite subcover:

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

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

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

Letting $\varepsilon \downarrow 0$ we arrive at

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

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

Definition 1.3.2 [UNIFORM DISTRIBUTION] The unique extension $\mathbb{P}$ of $\mathbb{P}_0$ to $\mathcal{B}((a,b])$ according to Proposition 1.2.17 is called **uniform distribution** on $(a,b]$.

Hence $\mathbb{P}$ is the unique measure on $\mathcal{B}((a,b])$ such that $\mathbb{P}((c,d]) = \frac{d-c}{b-a}$ for $a \leq c < d \leq b$. To get the Lebesgue measure on $\mathbb{R}$ we observe that $B \in \mathcal{B}(\mathbb{R})$ implies that $B \cap (a,b] \in \mathcal{B}((a,b])$ (check!). Then we can proceed as follows:

Definition 1.3.3 [LEBESGUE MEASURE] [8] Given $B \in \mathcal{B}(\mathbb{R})$ we define the Lebesgue measure on $\mathbb{R}$ by

$$\lambda(B) := \sum_{n=-\infty}^{\infty} \mathbb{P}_n(B \cap (n-1, n])$$

where $\mathbb{P}_n$ is the uniform distribution on $(n-1, n]$.

Accordingly, λ is the unique σ-finite measure on $\mathcal{B}(\mathbb{R})$ such that $\lambda((c,d]) = d - c$ for all $-\infty < c < d < \infty$. We can also write $\lambda(B) = \int_B d\lambda(x)$. Now we can go backwards: In order to obtain the Lebesgue measure on a set $I \subseteq \mathbb{R}$ with $I \in \mathcal{B}(\mathbb{R})$ we let $\mathcal{B}_I := \{B \subseteq I : B \in \mathcal{B}(\mathbb{R})\}$ and

$$\lambda_I(B) := \lambda(B) \quad \text{for} \quad B \in \mathcal{B}_I.$$

Given that $\lambda(I) > 0$, then $\lambda_I/\lambda(I)$ is the uniform distribution on I. Important cases for I are the closed intervals $[a,b]$. Furthermore, for $-\infty < a < b < \infty$ we have that $\mathcal{B}_{(a,b]} = \mathcal{B}((a,b])$ (check!).

[8] 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).

1.3.5 Gaussian distribution on $\mathbb{R}$ with mean $m \in \mathbb{R}$ and variance $\sigma^2 > 0$

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

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

(3) We take the algebra $\mathcal{G}$ considered in Example 1.1.4 and define

$$\mathbb{P}_0(A) := \sum_{i=1}^{n} \int_{a_i}^{b_i} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

for $A := (a_1, b_1] \cup (a_2, b_2] \cup \cdots \cup (a_n, b_n]$ where we consider the RIEMANN-integral [9] on the right-hand side. One can show (we do not do this here, but compare with Proposition 3.5.8 below) that $\mathbb{P}_0$ satisfies the assumptions of Proposition 1.2.17, so that we can extend $\mathbb{P}_0$ to a probability measure $\mathcal{N}_{m,\sigma^2}$ on $\mathcal{B}(\mathbb{R})$.

The measure $\mathcal{N}_{m,\sigma^2}$ is called **Gaussian distribution** [10] (**normal distribution**) with mean m and variance σ^2. Given $A \in \mathcal{B}(\mathbb{R})$ we write

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

The function $p_{m,\sigma^2}(x)$ is called **Gaussian density**.

1.3.6 Exponential distribution on $\mathbb{R}$ with parameter $\lambda > 0$

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

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

(3) For A and $\mathcal{G}$ as in Subsection 1.3.5 we define, via the RIEMANN-integral,

$$\mathbb{P}_0(A) := \sum_{i=1}^{n} \int_{a_i}^{b_i} p_\lambda(x)dx \quad \text{with} \quad p_\lambda(x) := \mathbb{1}_{[0,\infty)}(x)\lambda e^{-\lambda x}$$

Again, $\mathbb{P}_0$ satisfies the assumptions of Proposition 1.2.17, so that we can extend $\mathbb{P}_0$ to the **exponential distribution** μ_λ with parameter λ and density $p_\lambda(x)$ on $\mathcal{B}(\mathbb{R})$.

---

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

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

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

$$\mu_\lambda(A) = \int_A p_\lambda(x)dx.$$

The exponential distribution can be considered as a *continuous time* version of the geometric distribution. In particular, we see that the distribution does not have a memory in the sense that for $a, b \geq 0$ we have

$$\mu_\lambda([a+b,\infty)|[a,\infty)) = \mu_\lambda([b,\infty))$$

with the conditional probability on the left-hand side. In words: 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, it holds

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

Example 1.3.4 Suppose that the amount of time one spends in a post office is exponential distributed with $\lambda = \frac{1}{10}$.

(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 $\mu_\lambda([15,\infty)) = e^{-15\frac{1}{10}} \approx 0.220$. For (b) we get $\mu_\lambda([15,\infty)|[10,\infty)) = \mu_\lambda([5,\infty)) = e^{-5\frac{1}{10}} \approx 0.604$.

1.3.7 Poisson's Theorem

For large n and small p the Poisson distribution provides a good approximation for the binomial distribution.

Proposition 1.3.5 [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 $k = 0, 1, \ldots,$*

$$\mu_{n,p_n}(\{k\}) \to \pi_\lambda(\{k\}), \quad n \to \infty.$$

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

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

Of course, $\lim_{n\to\infty}(np_n)^k = \lambda^k$ and $\lim_{n\to\infty} \frac{n(n-1)\dots(n-k+1)}{n^k} = 1$. 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 $\varepsilon_0 > 0$ and $n_0 \geq 1$ such that $|\varepsilon_n| \leq \varepsilon_0$ for all $n \geq n_0$. Then

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

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

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

In the same way we get that

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

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

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

□

1.4 A set which is not a Borel 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

Definition 1.4.1 [λ-SYSTEM] A class $\mathcal{L}$ is a **λ-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) $A_1, A_2, \cdots \in \mathcal{L}$ and $A_n \subseteq A_{n+1}$, $n = 1, 2, \ldots$ imply $\bigcup_{n=1}^{\infty} A_n \in \mathcal{L}$.

Proposition 1.4.2 [π-λ-THEOREM] *If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system, then $\mathcal{P} \subseteq \mathcal{L}$ implies $\sigma(\mathcal{P}) \subseteq \mathcal{L}$.*

Definition 1.4.3 [EQUIVALENCE RELATION] An relation $\sim$ on a set X is called **equivalence relation** if and only if

(1) $x \sim x$ for all $x \in X$ (reflexivity),

(2) $x \sim y$ implies $y \sim x$ for all $x, y \in X$ (symmetry),

(3) $x \sim y$ and $y \sim z$ imply $x \sim z$ for all $x, y, z \in X$ (transitivity).

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

$$\mathcal{L} := \{A \in \mathcal{B}((0,1]) : \\ A \oplus x \in \mathcal{B}((0,1]) \text{ and } \lambda(A \oplus x) = \lambda(A) \text{ for all } x \in (0,1]\} \quad (1.2)$$

where λ is the Lebesgue measure on $(0,1]$.

Lemma 1.4.4 *The system $\mathcal{L}$ from (1.2) 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 $B \setminus A \in \mathcal{L}$. By the definition of $\oplus$ it is easy to see that $A \subseteq B$ implies $A \oplus x \subseteq B \oplus x$ and

$$(B \oplus x) \setminus (A \oplus x) = (B \setminus A) \oplus x,$$

and therefore, $(B \setminus A) \oplus x \ \in \mathcal{B}((0,1])$. Since λ is a probability measure it follows

$$\begin{aligned} \lambda(B \setminus A) &= \lambda(B) - \lambda(A) \\ &= \lambda(B \oplus x) - \lambda(A \oplus x) \\ &= \lambda((B \oplus x) \setminus (A \oplus x)) \\ &= \lambda((B \setminus A) \oplus x) \end{aligned}$$

and $B \setminus A \in \mathcal{L}$. Property (3) is left as an exercise. □

Finally, we need the axiom of choice.

Proposition 1.4.5 [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.

Proposition 1.4.6 *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 (1.2). If $(a,b] \subseteq [0,1]$, then $(a,b] \in \mathcal{L}$. Since

$$\mathcal{P} := \{(a,b] : 0 \le a < b \le 1\}$$

is a π-system which generates $\mathcal{B}((0,1])$ it follows by the π-λ-Theorem 1.4.2 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 $\lambda(H \oplus r) = \lambda(H) = a \geq 0$ 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 $H \notin \mathcal{B}((0,1])$. $\square$

Chapter 2

Random variables

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, in many stochastic models functions $f : \Omega \to \mathbb{R}$ which describe certain random phenomena are considered and one is interested in the computation of 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 we will introduce now.

2.1 Random variables

We start with the most simple random variables.

Definition 2.1.1 [(MEASURABLE) STEP-FUNCTION] Let $(\Omega, \mathcal{F})$ be a measurable space. A function $f : \Omega \to \mathbb{R}$ is called **measurable step-function** or **step-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

$$f(\omega) = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}(\omega),$$

where

$$\mathbb{1}_{A_i}(\omega) := \begin{cases} 1 & : \quad \omega \in A_i \\ 0 & : \quad \omega \notin A_i \end{cases}.$$

Some particular examples for step-functions are

$$\begin{aligned} \mathbb{1}_{\Omega} &= 1, \\ \mathbb{1}_{\emptyset} &= 0, \\ \mathbb{1}_{A} + \mathbb{1}_{A^c} &= 1, \end{aligned}$$

$$\begin{aligned} \mathbb{1}_{A\cap B} &= \mathbb{1}_A \mathbb{1}_B, \\ \mathbb{1}_{A\cup B} &= \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A\cap B}. \end{aligned}$$

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

Definition 2.1.2 [RANDOM VARIABLES] Let $(\Omega, \mathcal{F})$ be a measurable space. A map $f : \Omega \to \mathbb{R}$ is called **random variable** provided that there is a sequence $(f_n)_{n=1}^{\infty}$ of measurable step-functions $f_n : \Omega \to \mathbb{R}$ such that

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

Does our definition give what we would like to have? Yes, as we see from

Proposition 2.1.3 *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 a random variable.*

(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 measurable step-functions. For a measurable step-function one has that

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

so that

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

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

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

so that we can use the step-functions

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

□

Sometimes the following proposition is useful which is closely connected to Proposition 2.1.3.

Proposition 2.1.4 *Assume a measurable space $(\Omega, \mathcal{F})$ and a sequence of random variables $f_n : \Omega \to \mathbb{R}$ such that $f(\omega) := \lim_n f_n(\omega)$ exists for all $\omega \in \Omega$. Then $f : \Omega \to \mathbb{R}$ is a random variable.*

The proof is an exercise.

Proposition 2.1.5 [PROPERTIES OF RANDOM VARIABLES] *Let $(\Omega, \mathcal{F})$ be a measurable space and $f, g : \Omega \to \mathbb{R}$ 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 a random variable.*

(2) $(fg)(\omega) := f(\omega)g(\omega)$ *is a random-variable.*

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

(4) $|f|$ *is a random variable.*

Proof. (2) We find measurable step-functions $f_n, g_n : \Omega \to \mathbb{R}$ such that

$$f(\omega) = \lim_{n\to\infty} f_n(\omega) \quad \text{and} \quad g(\omega) = \lim_{n\to\infty} g_n(\omega).$$

Hence

$$(fg)(\omega) = \lim_{n\to\infty} f_n(\omega)g_n(\omega).$$

Finally, we remark, that $f_n(\omega)g_n(\omega)$ is a measurable step-function. In fact, assuming that

$$f_n(\omega) = \sum_{i=1}^{k} \alpha_i \mathbb{1}_{A_i}(\omega) \quad \text{and} \quad g_n(\omega) = \sum_{j=1}^{l} \beta_j \mathbb{1}_{B_j}(\omega),$$

yields

$$(f_n g_n)(\omega) = \sum_{i=1}^{k}\sum_{j=1}^{l} \alpha_i\beta_j \mathbb{1}_{A_i}(\omega)\mathbb{1}_{B_j}(\omega) = \sum_{i=1}^{k}\sum_{j=1}^{l} \alpha_i\beta_j \mathbb{1}_{A_i\cap B_j}(\omega)$$

and we again obtain a step-function, since $A_i \cap B_j \in \mathcal{F}$. Items (1), (3), and (4) are an exercise. □

2.2 Measurable maps

Now we extend the notion of random variables to the notion of measurable maps, which is necessary in many considerations and even more natural.

Definition 2.2.1 [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 random variables is given by

Proposition 2.2.2 *Let $(\Omega, \mathcal{F})$ be a measurable space and $f : \Omega \to \mathbb{R}$. Then the following assertions are equivalent:*

(1) *The map f is a random variable.*

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

For the proof we need

Lemma 2.2.3 *Let $(\Omega, \mathcal{F})$ and (M, Σ) be measurable spaces and let $f : \Omega \to M$. Assume that $\Sigma_0 \subseteq \Sigma$ is a system of subsets such that $\sigma(\Sigma_0) = \Sigma$. If*

$$f^{-1}(B) \in \mathcal{F} \quad \textit{for all} \quad B \in \Sigma_0,$$

then

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

Proof. Define

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

By assumption, $\Sigma_0 \subseteq \mathcal{A}$. We show that $\mathcal{A}$ is a σ–algebra.
(1) $f^{-1}(M) = \Omega \in \mathcal{F}$ implies that $M \in \mathcal{A}$.
(2) If $B \in \mathcal{A}$, then

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

(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 $\Sigma = \sigma(\Sigma_0)$ this implies that $\Sigma \subseteq \mathcal{A}$, which implies our lemma. □

Proof of Proposition 2.2.2. (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 2.2.3 since $\mathcal{B}(\mathbb{R}) = \sigma((a, b) : -\infty < a < b < \infty)$. □

Example 2.2.4 *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 $f^{-1}((a,b))$ is open for all $-\infty < a < b < \infty$, so that $f^{-1}((a,b)) \in \mathcal{B}(\mathbb{R})$. Since the open intervals generate $\mathcal{B}(\mathbb{R})$ we can apply Lemma 2.2.3. □

Now we state some general properties of measurable maps.

Proposition 2.2.5 *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 the following is satisfied:*

(1) *$g \circ f : \Omega_1 \to \Omega_3$ defined by*

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

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

(2) *Assume that $\mathbb{P}_1$ is a probability measure on $\mathcal{F}_1$ and define*

$$\mathbb{P}_2(B_2) := \mathbb{P}_1\left(\{\omega_1 \in \Omega_1 : f(\omega_1) \in B_2\}\right).$$

Then $\mathbb{P}_2$ is a probability measure on $\mathcal{F}_2$.

The proof is an exercise.

Example 2.2.6 We want to simulate the flipping of an (unfair) coin by the random number generator: the random number generator of the computer gives us a number which has (a discrete) uniform distribution on $[0,1]$. So we take the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$ and define for $p \in (0,1)$ the random variable

$$f(\omega) := \mathbb{1}_{[0,p)}(\omega).$$

Then it holds

$$\begin{aligned} \mathbb{P}_2(\{1\}) &:= \mathbb{P}_1\left(\{\omega \in \Omega : f(\omega) = 1\}\right) = \lambda([0,p)) = p, \\ \mathbb{P}_2(\{0\}) &:= \mathbb{P}_1\left(\{\omega_1 \in \Omega_1 : f(\omega_1) = 0\}\right) = \lambda([p,1]) = 1 - p. \end{aligned}$$

Assume the random number generator gives out the number x. If we would write a program such that "output" = "heads" in case $x \in [0,p)$ and "output" = "tails" in case $x \in [p,1]$, "output" would simulate the flipping of an (unfair) coin, or in other words, "output" has binomial distribution $\mu_{1,p}$.

Definition 2.2.7 [LAW OF A RANDOM VARIABLE] Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f : \Omega \to \mathbb{R}$ be a random variable. Then

$$\mathbb{P}_f(B) := \mathbb{P}\left(\{\omega \in \Omega : f(\omega) \in B\}\right)$$

is called the **law** or **image measure** of the random variable f.

The law of a random variable is completely characterized by its distribution function which we introduce now.

Definition 2.2.8 [DISTRIBUTION-FUNCTION] Given a random variable $f : \Omega \to \mathbb{R}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the function

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

is called **distribution function** of f.

Proposition 2.2.9 [PROPERTIES OF DISTRIBUTION-FUNCTIONS] *The distribution-function $F_f : \mathbb{R} \to [0,1]$ is a right-continuous non-decreasing function such that*

$$\lim_{x \to -\infty} F(x) = 0 \quad \textit{and} \quad \lim_{x \to \infty} F(x) = 1.$$

Proof. (i) F is non-decreasing: given $x_1 < x_2$ one has that

$$\{\omega \in \Omega : f(\omega) \leq x_1\} \subseteq \{\omega \in \Omega : f(\omega) \leq x_2\}$$

and

$$F(x_1) = \mathbb{P}(\{\omega \in \Omega : f(\omega) \leq x_1\}) \leq \mathbb{P}(\{\omega \in \Omega : f(\omega) \leq x_2\}) = F(x_2).$$

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

$$\begin{aligned} F(x) &= \mathbb{P}(\{\omega \in \Omega : f(\omega) \leq x\}) \\ &= \mathbb{P}\left(\bigcap_{n=1}^{\infty} \{\omega \in \Omega : f(\omega) \leq x_n\}\right) \\ &= \lim_n \mathbb{P}\left(\{\omega \in \Omega : f(\omega) \leq x_n\}\right) \\ &= \lim_n F(x_n). \end{aligned}$$

(iii) The properties $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$ are an exercise. □

Proposition 2.2.10 *Assume that $\mathbb{P}_1$ and $\mathbb{P}_2$ are probability measures on $\mathcal{B}(\mathbb{R})$ and F_1 and F_2 are the corresponding distribution functions. Then the following assertions are equivalent:*

(1) $\mathbb{P}_1 = \mathbb{P}_2$.

(2) $F_1(x) = \mathbb{P}_1((-\infty, x]) = \mathbb{P}_2((-\infty, x]) = F_2(x)$ *for all* $x \in \mathbb{R}$.

Proof. (1) $\Rightarrow$ (2) is of course trivial. We consider (2) $\Rightarrow$ (1): For sets of type

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

one can show that

$$F_1(b) - F_1(a) = \mathbb{P}_1(A) = \mathbb{P}_2(A) = F_2(b) - F_2(a).$$

Now one can apply Proposition 1.2.23. □

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 1.1.8 or
any other system such that $\sigma(\mathcal{G}) = \mathcal{B}(\mathbb{R})$.

$\Updownarrow$ Lemma 2.2.3

f is measurable: $f^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{B}(\mathbb{R})$

$\Updownarrow$ Proposition 2.2.2

f is a random variable i.e.
there exist measurable step functions $(f_n)_{n=1}^{\infty}$ i.e.
$f_n = \sum_{k=1}^{N_n} a_k^n \mathbb{1}_{A_k^n}$
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 $n \to \infty$.

$\Updownarrow$ Proposition 2.1.3

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

2.3 Independence

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

Definition 2.3.1 [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, that means for example $I = \{1, ..., n\}$, then the definition above is equivalent to

Definition 2.3.2 [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

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

The connection between the independence of random variables and of events is obvious:

Proposition 2.3.3 *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.*

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 2.3.4 [GROUPING OF INDEPENDENT RANDOM VARIABLES] *Let $f_k : \Omega \to \mathbb{R}$, $k = 1, 2, 3, ...$ be independent random variables. Assume Borel functions $g_i : \mathbb{R}^{n_i} \to \mathbb{R}$ for $i = 1, 2, ...$ and $n_i \in \{1, 2, ...\}$. Then the random variables $g_1(f_1(\omega), ..., f_{n_1}(\omega))$, $g_2(f_{n_1+1}(\omega), ..., f_{n_1+n_2}(\omega))$, $g_3(f_{n_1+n_2+1}(\omega), ..., f_{n_1+n_2+n_3}(\omega))$, ... are independent.*

The proof is an exercise.

Proposition 2.3.5 [INDEPENDENCE AND PRODUCT OF LAWS] *Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and that $f, g : \Omega \to \mathbb{R}$ are random variables with laws $\mathbb{P}_f$ and $\mathbb{P}_g$ and distribution-functions F_f and F_g, respectively. Then the following assertions are equivalent:*

(1) *f and g are independent.*

(2) *$\mathbb{P}((f,g) \in B) = (\mathbb{P}_f \times \mathbb{P}_g)(B)$ for all $B \in \mathcal{B}(\mathbb{R}^2)$.*

(3) *$\mathbb{P}(f \leq x, g \leq y) = F_f(x)F_g(y)$ for all $x, y \in \mathbb{R}$.*

The proof is an exercise.

Remark 2.3.6 Assume that there are Riemann-integrable functions $p_f, p_g : \mathbb{R} \to [0, \infty)$ such that

$$\int_{\mathbb{R}} p_f(x)dx = \int_{\mathbb{R}} p_g(x)dx = 1,$$

$$F_f(x) = \int_{-\infty}^{x} p_f(y)dy, \quad \text{and} \quad F_g(x) = \int_{-\infty}^{x} p_g(y)dy$$

for all $x \in \mathbb{R}$ (one says that the distribution-functions F_f and F_g are absolutely continuous with densities p_f and p_g, respectively). Then the independence of f and g is also equivalent to the representation

$$F_{(f,g)}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p_f(u)p_g(v)d(u)d(v).$$

In other words: the distribution-function of the random *vector* (f, g) has a density which is the product of the densities of f and g.

Often one needs the existence of sequences of independent random variables $f_1, f_2, ... : \Omega \to \mathbb{R}$ having a certain distribution. How to construct such sequences? First we let

$$\Omega := \mathbb{R}^{\mathbb{N}} = \{x = (x_1, x_2, ...) : x_n \in \mathbb{R}\}.$$

Then we define the projections $\pi_n : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ given by

$$\pi_n(x) := x_n,$$

that means π_n filters out the n-th coordinate. Now we take the smallest σ-algebra such that all these projections are random variables, that means we take

$$\mathcal{B}(\mathbb{R}^{\mathbb{N}}) = \sigma\left(\pi_n^{-1}(B) : n = 1, 2, ..., B \in \mathcal{B}(\mathbb{R})\right),$$

see Proposition 1.1.6. Finally, let $\mathbb{P}_1, \mathbb{P}_2, ...$ be a sequence of probability measures on $\mathcal{B}(\mathbb{R})$. Using CARATHÉODORY's extension theorem (Proposition 1.2.17) we find an unique probability measure $\mathbb{P}$ on $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$ such that

$$\mathbb{P}(B_1 \times B_2 \times \cdots \times B_n \times \mathbb{R} \times \mathbb{R} \times \cdots) = \mathbb{P}_1(B_1) \cdots \mathbb{P}_n(B_n)$$

for all $n = 1, 2, ...$ and $B_1, ..., B_n \in \mathcal{B}(\mathbb{R})$, where

$$B_1 \times B_2 \times \cdots \times B_n \times \mathbb{R} \times \mathbb{R} \times \cdots := \left\{ x \in \mathbb{R}^{\mathbb{N}} : x_1 \in B_1, ..., x_n \in B_n \right\}.$$

Proposition 2.3.7 [REALIZATION OF INDEPENDENT RANDOM VARIABLES] *Let $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}), \mathbb{P})$ and $\pi_n : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ be defined as above. Then $(\pi_n)_{n=1}^{\infty}$ is a sequence of independent random variables such that the law of π_n is $\mathbb{P}_n$, that means*

$$\mathbb{P}(\pi_n \in B) = \mathbb{P}_n(B)$$

for all $B \in \mathcal{B}(\mathbb{R})$.

Proof. Take Borel sets $B_1, ..., B_n \in \mathcal{B}(\mathbb{R})$. Then

$$\begin{aligned} &\mathbb{P}(\{x : \pi_1(x) \in B_1, ..., \pi_n(x) \in B_n\}) \\ &\qquad = \mathbb{P}(B_1 \times B_2 \times \cdots \times B_n \times \mathbb{R} \times \mathbb{R} \times \cdots) \\ &\qquad = \mathbb{P}_1(B_1) \cdots \mathbb{P}_n(B_n) \\ &\qquad = \prod_{k=1}^{n} \mathbb{P}(\mathbb{R} \times \cdots \times \mathbb{R} \times B_k \times \mathbb{R} \times \cdots) \\ &\qquad = \prod_{k=1}^{n} \mathbb{P}(\{x : \pi_k(x) \in B_k\}). \end{aligned}$$

□

Chapter 3

Integration

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $f : \Omega \to \mathbb{R}$, we define the expectation or integral

$$\mathbb{E}f = \int_\Omega f d\mathbb{P} = \int_\Omega f(\omega) d\mathbb{P}(\omega)$$

and investigate its basic properties.

3.1 Definition of the expected value

The definition of the integral is done within three steps.

Definition 3.1.1 [STEP ONE, f IS A STEP-FUNCTION] Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an $\mathcal{F}$-measurable $g : \Omega \to \mathbb{R}$ with representation

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

where $\alpha_i \in \mathbb{R}$ and $A_i \in \mathcal{F}$, we let

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

We have to check that the definition is correct, since it might be that different representations give different expected values $\mathbb{E}g$. However, this is not the case as shown by

Lemma 3.1.2 *Assuming measurable step-functions*

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

one has that $\sum_{i=1}^{n} \alpha_i \mathbb{P}(A_i) = \sum_{j=1}^{m} \beta_j \mathbb{P}(B_j)$.

Proof. By subtracting in both equations the right-hand side from the left-hand one we only need to show that

$$\sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i} = 0$$

implies that

$$\sum_{i=1}^{n} \alpha_i \mathbb{P}(A_i) = 0.$$

By taking all possible intersections of the sets A_i and by adding appropriate complements we find a system of sets $C_1, ..., C_N \in \mathcal{F}$ such that

(a) $C_j \cap C_k = \emptyset$ if $j \neq k$,

(b) $\bigcup_{j=1}^{N} C_j = \Omega$,

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

Now we get that

$$0 = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i} = \sum_{i=1}^{n} \sum_{j \in I_i} \alpha_i \mathbb{1}_{C_j} = \sum_{j=1}^{N} \left(\sum_{i: j \in I_i} \alpha_i \right) \mathbb{1}_{C_j} = \sum_{j=1}^{N} \gamma_j \mathbb{1}_{C_j}$$

so that $\gamma_j = 0$ if $C_j \neq \emptyset$. From this we get that

$$\sum_{i=1}^{n} \alpha_i \mathbb{P}(A_i) = \sum_{i=1}^{n} \sum_{j \in I_i} \alpha_i \mathbb{P}(C_j) = \sum_{j=1}^{N} \left(\sum_{i: j \in I_i} \alpha_i \right) \mathbb{P}(C_j) = \sum_{j=1}^{N} \gamma_j \mathbb{P}(C_j) = 0.$$

□

Proposition 3.1.3 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f, g : \Omega \to \mathbb{R}$ be measurable step-functions. Given $\alpha, \beta \in \mathbb{R}$ one has that*

$$\mathbb{E}\left(\alpha f + \beta g\right) = \alpha \mathbb{E} f + \beta \mathbb{E} g.$$

Proof. The proof follows immediately from Lemma 3.1.2 and the definition of the expected value of a step-function since, for

$$f = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i} \quad \text{and} \quad g = \sum_{j=1}^{m} \beta_j \mathbb{1}_{B_j},$$

one has that

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

and

$$\mathbb{E}(\alpha f + \beta g) = \alpha \sum_{i=1}^{n} \alpha_i \mathbb{P}(A_i) + \beta \sum_{j=1}^{m} \beta_j \mathbb{P}(B_j) = \alpha \mathbb{E} f + \beta \mathbb{E} g.$$

□

Definition 3.1.4 [STEP TWO, f IS NON-NEGATIVE] Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $f : \Omega \to \mathbb{R}$ with $f(\omega) \geq 0$ for all $\omega \in \Omega$. Then

$$\begin{aligned}\mathbb{E}f &= \int_\Omega f d\mathbb{P} = \int_\Omega f(\omega) d\mathbb{P}(\omega) \\ &:= \sup\left\{\mathbb{E}g : 0 \leq g(\omega) \leq f(\omega), g \text{ is a measurable step-function}\right\}.\end{aligned}$$

Note that in this definition the case $\mathbb{E}f = \infty$ is allowed. In the last step we will define the expectation for a general random variable. To this end we decompose a random variable $f : \Omega \to \mathbb{R}$ into its positive and negative part

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

with

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

Definition 3.1.5 [STEP THREE, f IS GENERAL] Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f : \Omega \to \mathbb{R}$ be a random variable.

(1) If $\mathbb{E}f^+ < \infty$ or $\mathbb{E}f^- < \infty$, then we say that the **expected value of f exists** and set
$$\mathbb{E}f := \mathbb{E}f^+ - \mathbb{E}f^- \in [-\infty, \infty].$$

(2) The random variable f is called **integrable** provided that
$$\mathbb{E}f^+ < \infty \quad \text{and} \quad \mathbb{E}f^- < \infty.$$

(3) If the expected value of f exists and $A \in \mathcal{F}$, then
$$\int_A f d\mathbb{P} = \int_A f(\omega) d\mathbb{P}(\omega) := \int_\Omega f(\omega) \mathbb{1}_A(\omega) d\mathbb{P}(\omega).$$

The expression $\mathbb{E}f$ is called **expectation** or **expected value** of the random variable f.

Remark 3.1.6 The fundamental LEBESGUE-integral on the real line can be introduced by the means, we have to our disposal so far, as follows: Assume a function $f : \mathbb{R} \to \mathbb{R}$ which is $(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}))$-measurable. Let $f_n : (n-1, n] \to \mathbb{R}$ be the restriction of f which is a random variable with respect to $\mathcal{B}((n-1, n]) = \mathcal{B}_{(n-1,n]}$. In Section 1.3.4 we have introduced the Lebesgue measure $\lambda = \lambda_n$ on $(n-1, n]$. Assume that f_n is integrable on $(n-1, n]$ for all $n = 1, 2, ...$ and that

$$\sum_{n=-\infty}^{\infty} \int_{(n-1,n]} |f(x)| d\lambda(x) < \infty,$$

then $f : \mathbb{R} \to \mathbb{R}$ is called integrable with respect to the LEBESGUE-measure on the real line and the LEBESGUE-integral is defined by

$$\int_{\mathbb{R}} f(x)d\lambda(x) := \sum_{n=-\infty}^{\infty} \int_{(n-1,n]} f(x)d\lambda(x).$$

Now we go the opposite way: Given a Borel set I and a map $f : I \to \mathbb{R}$ which is $(\mathcal{B}_I, \mathcal{B}(\mathbb{R}))$ measurable, we can extend f to a $(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}))$-measurable function $\widetilde{f}$ by $\widetilde{f}(x) := f(x)$ if $x \in I$ and $\widetilde{f}(x) := 0$ if $x \notin I$. If $\widetilde{f}$ is LEBESGUE-integrable, then we define

$$\int_I f(x)d\lambda(x) := \int_{\mathbb{R}} \widetilde{f}(x)d\lambda(x).$$

Example 3.1.7 A basic example for our integration is as follows: Let $\Omega = \{\omega_1, \omega_2, ...\}$, $\mathcal{F} := 2^\Omega$, and $\mathbb{P}(\{\omega_n\}) = q_n \in [0, 1]$ with $\sum_{n=1}^{\infty} q_n = 1$. Given $f : \Omega \to \mathbb{R}$ we get that f is integrable if and only if

$$\sum_{n=1}^{\infty} |f(\omega_n)|q_n < \infty,$$

and the expected value exists if either

$$\sum_{\{n:f(\omega_n)\geq 0\}} f(\omega_n)q_n < \infty \quad \text{or} \quad \sum_{\{n:f(\omega_n)\leq 0\}} (-f(\omega_n))q_n < \infty.$$

If the expected value exists, then it computes to

$$\mathbb{E}f = \sum_{n=1}^{\infty} f(\omega_n)q_n \in [-\infty, \infty].$$

A simple example for the expectation is the expected value while rolling a die:

Example 3.1.8 Assume that $\Omega := \{1, 2, \ldots, 6\}$, $\mathcal{F} := 2^\Omega$, and $\mathbb{P}(\{k\}) := \frac{1}{6}$, which models rolling a die. If we define $f(k) = k$, i.e.

$$f(k) := \sum_{i=1}^{6} i\mathbb{1}_{\{i\}}(k),$$

then f is a measurable step-function and it follows that

$$\mathbb{E}f = \sum_{i=1}^{6} i\mathbb{P}(\{i\}) = \frac{1 + 2 + \cdots + 6}{6} = 3.5.$$

Besides the expected value, the variance is often of interest.

Definition 3.1.9 [VARIANCE] Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f : \Omega \to \mathbb{R}$ be an integrable random variable. Then

$$\mathrm{var(f)} = \sigma_{\mathrm{f}}^2 = \mathbb{E}[\mathrm{f} - \mathbb{E}\mathrm{f}]^2 \in [0, \infty]$$

is called **variance**.

Let us summarize some simple properties:

Proposition 3.1.10 (1) *If f is integrable and $\alpha, c \in \mathbb{R}$, then*

$$\mathrm{var}(\alpha \mathrm{f} - \mathrm{c}) = \alpha^2 \mathrm{var(f)}.$$

(2) *If $\mathbb{E}f^2 < \infty$ then* $\mathrm{var(f)} = \mathbb{E}\mathrm{f}^2 - (\mathbb{E}\mathrm{f})^2 < \infty$.

Proof. (1) follows from

$$\mathrm{var}(\alpha \mathrm{f} - \mathrm{c}) = \mathbb{E}[(\alpha \mathrm{f} - \mathrm{c}) - \mathbb{E}(\alpha \mathrm{f} - \mathrm{c})]^2 = \mathbb{E}[\alpha \mathrm{f} - \alpha \mathbb{E}\mathrm{f}]^2 = \alpha^2 \mathrm{var(f)}.$$

(2) First we remark that $\mathbb{E}|f| \leq (\mathbb{E}f^2)^{\frac{1}{2}}$ as we shall see later by Hölder's inequality (Corollary 3.6.6), that means any square integrable random variable is integrable. Then we simply get that

$$\mathrm{var(f)} = \mathbb{E}[\mathrm{f} - \mathbb{E}\mathrm{f}]^2 = \mathbb{E}\mathrm{f}^2 - 2\mathbb{E}(\mathrm{f}\mathbb{E}\mathrm{f}) + (\mathbb{E}\mathrm{f})^2 = \mathbb{E}\mathrm{f}^2 - 2(\mathbb{E}\mathrm{f})^2 + (\mathbb{E}\mathrm{f})^2.$$

□

3.2 Basic properties of the expected value

We say that a property $\mathcal{P}(\omega)$, depending on ω, holds $\mathbb{P}$-**almost surely** or **almost surely** (a.s.) if

$$\{\omega \in \Omega : \mathcal{P}(\omega) \text{ holds}\}$$

belongs to $\mathcal{F}$ and is of measure one. Let us start with some first properties of the expected value.

Proposition 3.2.1 *Assume a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variables $f, g : \Omega \to \mathbb{R}$.*

(1) *If $0 \leq f(\omega) \leq g(\omega)$, then $0 \leq \mathbb{E}f \leq \mathbb{E}g$.*

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

$$|\mathbb{E}f| \leq \mathbb{E}|f|.$$

(3) *If* $f = 0$ *a.s., then* $\mathbb{E}f = 0$.

(4) *If* $f \geq 0$ *a.s. and* $\mathbb{E}f = 0$, *then* $f = 0$ *a.s.*

(5) *If* $f = g$ *a.s. and* $\mathbb{E}f$ *exists, then* $\mathbb{E}g$ *exists and* $\mathbb{E}f = \mathbb{E}g$.

Proof. (1) follows directly from the definition. Property (2) can be seen as follows: by definition, the random variable f is integrable if and only if $\mathbb{E}f^+ < \infty$ and $\mathbb{E}f^- < \infty$. Since

$$\big\{\omega \in \Omega : f^+(\omega) \neq 0\big\} \cap \big\{\omega \in \Omega : f^-(\omega) \neq 0\big\} = \emptyset$$

and since both sets are measurable, it follows that $|f| = f^+ + f^-$ is integrable if and only if f^+ and f^- are integrable and that

$$|\mathbb{E}f| = |\mathbb{E}f^+ - \mathbb{E}f^-| \leq \mathbb{E}f^+ + \mathbb{E}f^- = \mathbb{E}|f|.$$

(3) If $f = 0$ a.s., then $f^+ = 0$ a.s. and $f^- = 0$ a.s., so that we can restrict ourselves to the case $f(\omega) \geq 0$. If g is a measurable step-function with $g = \sum_{k=1}^n a_k \mathbb{1}_{A_k}$, $g(\omega) \geq 0$, and $g = 0$ a.s., then $a_k \neq 0$ implies $\mathbb{P}(A_k) = 0$. Hence

$$\mathbb{E}f = \sup\{\mathbb{E}g : 0 \leq g \leq f, g \text{ is a measurable step-function}\} = 0$$

since $0 \leq g \leq f$ implies $g = 0$ a.s. Properties (4) and (5) are exercises. □

The next lemma is useful later on. In this lemma we use, as an approximation for f, the so-called *staircase*-function. This idea was already exploited in the proof of Proposition 2.1.3.

Lemma 3.2.2 *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and* $f : \Omega \to \mathbb{R}$ *be a random variable.*

(1) *Then there exists a sequence of measurable step-functions* $f_n : \Omega \to \mathbb{R}$ *such that, for all* $n = 1, 2, \ldots$ *and for all* $\omega \in \Omega$,

$$|f_n(\omega)| \leq |f_{n+1}(\omega)| \leq |f(\omega)| \quad \textit{and} \quad f(\omega) = \lim_{n\to\infty} f_n(\omega).$$

If $f(\omega) \geq 0$ *for all* $\omega \in \Omega$, *then one can arrange* $f_n(\omega) \geq 0$ *for all* $\omega \in \Omega$.

(2) *If* $f \geq 0$ *and if* $(f_n)_{n=1}^\infty$ *is a sequence of measurable step-functions with* $0 \leq f_n(\omega) \uparrow f(\omega)$ *for all* $\omega \in \Omega$ *as* $n \to \infty$, *then*

$$\mathbb{E}f = \lim_{n\to\infty} \mathbb{E}f_n.$$

Proof. (1) It is easy to verify that the *staircase*-functions

$$f_n(\omega) := \sum_{k=0}^{4^n-1} \frac{k}{2^n} \mathbb{I}_{\{\frac{k}{2^n} \le f < \frac{k+1}{2^n}\}}(\omega) + \sum_{k=-4^n}^{-1} \frac{k+1}{2^n} \mathbb{I}_{\{\frac{k}{2^n} \le f < \frac{k+1}{2^n}\}}(\omega)$$

fulfill all the conditions.

(2) Letting

$$f_n^0(\omega) := \sum_{k=0}^{4^n-1} \frac{k}{2^n} \mathbb{I}_{\{\frac{k}{2^n} \le f < \frac{k+1}{2^n}\}}(\omega)$$

we get $0 \le f_n^0(\omega) \uparrow f(\omega)$ for all $\omega \in \Omega$. On the other hand, by the definition of the expectation there exists a sequence $0 \le g_n(\omega) \le f(\omega)$ of measurable step-functions such that $\mathbb{E}g_n \uparrow \mathbb{E}f$. Hence

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

is a measurable step-function with $0 \le g_n(\omega) \le h_n(\omega) \uparrow f(\omega)$, so that

$$\mathbb{E}g_n \le \mathbb{E}h_n \le \mathbb{E}f, \quad \text{and} \quad \lim_{n\to\infty} \mathbb{E}g_n = \lim_{n\to\infty} \mathbb{E}h_n = \mathbb{E}f.$$

Now we will show that for every sequence $(f_k)_{k=1}^{\infty}$ of measurable step-functions with $0 \le f_k \uparrow f$ it holds $\lim_{k\to\infty} \mathbb{E}f_k = \mathbb{E}f$. 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 $k \to \infty$. Let

$$z_{k,n} := \arctan \mathbb{E}d_{k,n}$$

so that $0 \le z_{k,n} \le 1$. Since $(z_{k,n})_{k=1}^{\infty}$ is increasing for fixed n and $(z_{k,n})_{n=1}^{\infty}$ is increasing for fixed k one quickly checks that

$$\lim_k \lim_n z_{k,n} = \lim_n \lim_k z_{k,n}.$$

Hence

$$\mathbb{E}f = \lim_n \mathbb{E}h_n = \lim_n \lim_k \mathbb{E}d_{k,n} = \lim_k \lim_n \mathbb{E}d_{k,n} = \lim_k \mathbb{E}f_k$$

where we have used the following fact: if $0 \le \varphi_n(\omega) \uparrow \varphi(\omega)$ for *step-functions* φ_n and φ, then

$$\lim_n \mathbb{E}\varphi_n = \mathbb{E}\varphi.$$

To check this, it is sufficient to assume that $\varphi(\omega) = \mathbb{I}_A(\omega)$ for some $A \in \mathcal{F}$. Let $\varepsilon \in (0,1)$ and

$$B_\varepsilon^n := \{\omega \in A : 1 - \varepsilon \le \varphi_n(\omega)\}.$$

Then

$$(1-\varepsilon)\mathbb{1}_{B_\varepsilon^n}(\omega) \leq \varphi_n(\omega) \leq \mathbb{1}_A(\omega).$$

Since $B_\varepsilon^n \subseteq B_\varepsilon^{n+1}$ and $\bigcup_{n=1}^\infty B_\varepsilon^n = A$ we get, by the monotonicity of the measure, that $\lim_n \mathbb{P}(B_\varepsilon^n) = \mathbb{P}(A)$ so that

$$(1-\varepsilon)\mathbb{P}(A) \leq \lim_n \mathbb{E}\varphi_n.$$

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

$$\mathbb{E}\varphi = \mathbb{P}(A) \leq \lim_n \mathbb{E}\varphi_n \leq \mathbb{E}\varphi$$

and are done. □

Now we continue with some basic properties of the expectation.

Proposition 3.2.3 [PROPERTIES OF THE EXPECTATION] *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f, g : \Omega \to \mathbb{R}$ be random variables such that $\mathbb{E}f$ and $\mathbb{E}g$ exist.*

(1) *If $f \geq 0$ and $g \geq 0$, then $\mathbb{E}(f+g) = \mathbb{E}f + \mathbb{E}g$.*

(2) *If $c \in \mathbb{R}$, then $\mathbb{E}(cf)$ exists and $\mathbb{E}(cf) = c\mathbb{E}f$.*

(3) *If $\mathbb{E}f^+ + \mathbb{E}g^+ < \infty$ or $\mathbb{E}f^- + \mathbb{E}g^- < \infty$, then $\mathbb{E}(f+g)^+ < \infty$ or $\mathbb{E}(f+g)^- < \infty$ and $\mathbb{E}(f+g) = \mathbb{E}f + \mathbb{E}g$.*

(4) *If $f \leq g$, then $\mathbb{E}f \leq \mathbb{E}g$.*

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

Proof. (1) Here we use Lemma 3.2.2 (2) by finding step-functions $0 \leq f_n(\omega) \uparrow f(\omega)$ and $0 \leq g_n(\omega) \uparrow g(\omega)$ such that $0 \leq f_n(\omega) + g_n(\omega) \uparrow f(\omega) + g(\omega)$ and

$$\mathbb{E}(f+g) = \lim_n \mathbb{E}(f_n + g_n) = \lim_n (\mathbb{E}f_n + \mathbb{E}g_n) = \mathbb{E}f + \mathbb{E}g$$

by Proposition 3.1.3.

(2) is an exercise.

(3) We only consider the case that $\mathbb{E}f^+ + \mathbb{E}g^+ < \infty$. Because of $(f+g)^+ \leq f^+ + g^+$ one gets that $\mathbb{E}(f+g)^+ < \infty$. Moreover, one quickly checks that

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

so that $\mathbb{E}f^- + \mathbb{E}g^- = \infty$ if and only if $\mathbb{E}(f+g)^- = \infty$ if and only if $\mathbb{E}f + \mathbb{E}g = \mathbb{E}(f+g) = -\infty$. Assuming that $\mathbb{E}f^- + \mathbb{E}g^- < \infty$ gives that $\mathbb{E}(f+g)^- < \infty$ and

$$\mathbb{E}[(f+g)^+ + f^- + g^-] = \mathbb{E}[f^+ + g^+ + (f+g)^-]$$

which implies that $\mathbb{E}(f+g) = \mathbb{E}f + \mathbb{E}g$ because of (1).
(4) If $\mathbb{E}f^- = \infty$ or $\mathbb{E}g^+ = \infty$, then $\mathbb{E}f = -\infty$ or $\mathbb{E}g = \infty$ so that nothing is to prove. Hence assume that $\mathbb{E}f^- < \infty$ and $\mathbb{E}g^+ < \infty$. The inequality $f \leq g$ gives $0 \leq f^+ \leq g^+$ and $0 \leq g^- \leq f^-$ so that f and g are integrable and

$$\mathbb{E}f = \mathbb{E}f^+ - \mathbb{E}f^- \leq \mathbb{E}g^+ - \mathbb{E}g^- = \mathbb{E}g.$$

(5) Since $(af+bg)^+ \leq |a||f| + |b||g|$ and $(af+bg)^- \leq |a||f| + |b||g|$ we get that $af + bg$ is integrable. The equality for the expected values follows from (2) and (3). □

Proposition 3.2.4 [MONOTONE CONVERGENCE] *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f, f_1, f_2, ... : \Omega \to \mathbb{R}$ be random variables.*

(1) *If $0 \leq f_n(\omega) \uparrow f(\omega)$ a.s., then $\lim_n \mathbb{E}f_n = \mathbb{E}f$.*

(2) *If $0 \geq f_n(\omega) \downarrow f(\omega)$ a.s., then $\lim_n \mathbb{E}f_n = \mathbb{E}f$.*

Proof. (a) First suppose

$$0 \leq f_n(\omega) \uparrow f(\omega) \quad \text{for all} \quad \omega \in \Omega.$$

For each f_n take a sequence of step functions $(f_{n,k})_{k\geq 1}$ such that $0 \leq f_{n,k} \uparrow f_n$, as $k \to \infty$. Setting

$$h_N := \max_{\substack{1\leq k\leq N \\ 1\leq n\leq N}} f_{n,k}$$

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

$$f_{n,N} \leq h_N \leq f_N$$

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

$$f_n \leq h \leq f,$$

and therefore

$$f = \lim_{n\to\infty} f_n \leq h \leq f.$$

Since h_N is a step function for each N and $h_N \uparrow f$ we have by Lemma 3.2.2 that $\lim_{N\to\infty} \mathbb{E}h_N = \mathbb{E}f$ and therefore, since $h_N \leq f_N$,

$$\mathbb{E}f \leq \lim_{N\to\infty} \mathbb{E}f_N.$$

On the other hand, $f_n \leq f_{n+1} \leq f$ implies $\mathbb{E}f_n \leq \mathbb{E}f$ and hence

$$\lim_{n\to\infty} \mathbb{E}f_n \leq \mathbb{E}f.$$

(b) Now let $0 \leq f_n(\omega) \uparrow f(\omega)$ a.s. By definition, this means that

$$0 \leq f_n(\omega) \uparrow f(\omega) \quad \text{for all} \quad \omega \in \Omega \setminus A,$$

where $\mathbb{P}(A) = 0$. Hence $0 \leq f_n(\omega)\mathbb{1}_{A^c}(\omega) \uparrow f(\omega)\mathbb{1}_{A^c}(\omega)$ for all ω and step (a) implies that

$$\lim_n \mathbb{E} f_n \mathbb{1}_{A^c} = \mathbb{E} f \mathbb{1}_{A^c}.$$

Since $f_n\mathbb{1}_{A^c} = f_n$ a.s. and $f\mathbb{1}_{A^c} = f$ a.s. we get $\mathbb{E}(f_n\mathbb{1}_{A^c}) = \mathbb{E}f_n$ and $\mathbb{E}(f\mathbb{1}_{A^c}) = \mathbb{E}f$ by Proposition 3.2.1 (5).

(c) Assertion (2) follows from (1) since $0 \geq f_n \downarrow f$ implies $0 \leq -f_n \uparrow -f$. □

Corollary 3.2.5 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $g, f, f_1, f_2, ... : \Omega \to \mathbb{R}$ be random variables, where g is integrable. If*

(1) *$g(\omega) \leq f_n(\omega) \uparrow f(\omega)$ a.s. or*

(2) *$g(\omega) \geq f_n(\omega) \downarrow f(\omega)$ a.s.,*

then $\lim_{n\to\infty} \mathbb{E}f_n = \mathbb{E}f$.

Proof. We only consider (1). Let $h_n := f_n - g$ and $h := f - g$. Then

$$0 \leq h_n(\omega) \uparrow h(\omega) \quad a.s.$$

Proposition 3.2.4 implies that $\lim_n \mathbb{E}h_n = \mathbb{E}h$. Since f_n^- and f^- are integrable Proposition 3.2.3 (1) implies that $\mathbb{E}h_n = \mathbb{E}f_n - \mathbb{E}g$ and $\mathbb{E}h = \mathbb{E}f - \mathbb{E}g$ so that we are done. □

In Definition 1.2.7 we defined $\limsup_{n\to\infty} a_n$ for a sequence $(a_n)_{n=1}^\infty \subset \mathbb{R}$. Naturally, one can extend this definition to a sequence $(f_n)_{n=1}^\infty$ of random variables: For any $\omega \in \Omega$ $\limsup_{n\to\infty} f_n(\omega)$ and $\liminf_{n\to\infty} f_n(\omega)$ exist. Hence $\limsup_{n\to\infty} f_n$ and $\liminf_{n\to\infty} f_n$ are random variables.

Proposition 3.2.6 [Lemma of Fatou] *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $g, f_1, f_2, ... : \Omega \to \mathbb{R}$ be random variables with $|f_n(\omega)| \leq g(\omega)$ a.s. Assume that g is integrable. Then* $\limsup f_n$ *and* $\liminf f_n$ *are integrable and one has that*

$$\mathbb{E}\liminf_{n\to\infty} f_n \;\leq\; \liminf_{n\to\infty} \mathbb{E}f_n \;\leq\; \limsup_{n\to\infty} \mathbb{E}f_n \;\leq\; \mathbb{E}\limsup_{n\to\infty} f_n.$$

Proof. We only prove the first inequality. The second one follows from the definition of lim sup and lim inf, the third one can be proved like the first one. So we let

$$Z_k := \inf_{n\geq k} f_n$$

so that $Z_k \uparrow \liminf_n f_n$ and, a.s.,

$$|Z_k| \leq g \quad \text{and} \quad |\liminf_n f_n| \leq g.$$

Applying monotone convergence in the form of Corollary 3.2.5 gives that

$$\mathbb{E}\liminf_n f_n = \lim_k \mathbb{E}Z_k = \lim_k \left(\mathbb{E}\inf_{n\geq k} f_n\right) \leq \lim_k \left(\inf_{n\geq k} \mathbb{E}f_n\right) = \liminf_n \mathbb{E}f_n.$$

□

Proposition 3.2.7 [Lebesgue's Theorem, dominated convergence] *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $g, f, f_1, f_2, ... : \Omega \to \mathbb{R}$ be random variables with $|f_n(\omega)| \leq g(\omega)$ a.s. Assume that g is integrable and that $f(\omega) = \lim_{n\to\infty} f_n(\omega)$ a.s. Then f is integrable and one has that*

$$\mathbb{E}f = \lim_n \mathbb{E}f_n.$$

Proof. Applying Fatou's Lemma gives

$$\mathbb{E}f = \mathbb{E}\liminf_{n\to\infty} f_n \ \leq \ \liminf_{n\to\infty} \mathbb{E}f_n \ \leq \ \limsup_{n\to\infty} \mathbb{E}f_n \ \leq \ \mathbb{E}\limsup_{n\to\infty} f_n = \mathbb{E}f.$$

□

Finally, we state a useful formula for independent random variables.

Proposition 3.2.8 *If f and g are independent and $\mathbb{E}|f| < \infty$ and $\mathbb{E}|g| < \infty$, then $\mathbb{E}|fg| < \infty$ and*

$$\mathbb{E}fg = \mathbb{E}f\mathbb{E}g.$$

The proof is an exercise. Concerning the variance of the sum of independent random variables we get the fundamental

Proposition 3.2.9 *Let $f_1, ..., f_n$ be independent random variables with finite second moment. Then one has that*

$$\mathrm{var}(f_1 + \cdots + f_n) = \mathrm{var}(f_1) + \cdots + \mathrm{var}(f_n).$$

Proof. The formula follows from

$$\begin{aligned}
\mathrm{var}(f_1 + \cdots + f_n) &= \mathbb{E}((f_1 + \cdots + f_n) - \mathbb{E}(f_1 + \cdots + f_n))^2 \\
&= \mathbb{E}\left(\sum_{i=1}^n (f_i - \mathbb{E}f_i)\right)^2 \\
&= \mathbb{E}\sum_{i,j=1}^n (f_i - \mathbb{E}f_i)(f_j - \mathbb{E}f_j) \\
&= \sum_{i,j=1}^n \mathbb{E}\left((f_i - \mathbb{E}f_i)(f_j - \mathbb{E}f_j)\right)
\end{aligned}$$

$$\begin{aligned} &= \sum_{i=1}^{n} \mathbb{E}(f_i - \mathbb{E}f_i)^2 + \sum_{i \neq j} \mathbb{E}\left((f_i - \mathbb{E}f_i)(f_j - \mathbb{E}f_j)\right) \\ &= \sum_{i=1}^{n} \mathrm{var}(f_i) + \sum_{i \neq j} \mathbb{E}(f_i - \mathbb{E}f_i)\mathbb{E}(f_j - \mathbb{E}f_j) \\ &= \sum_{i=1}^{n} \mathrm{var}(f_i) \end{aligned}$$

because $\mathbb{E}(f_i - \mathbb{E}f_i) = \mathbb{E}f_i - \mathbb{E}f_i = 0$. □

3.3 Connections to the Riemann-integral

In two typical situations we formulate (without proof) how our expected value connects to the Riemann-integral. For this purpose we use the Lebesgue measure defined in Section 1.3.4.

Proposition 3.3.1 *Let $f : [0,1] \to \mathbb{R}$ be a continuous function. Then*

$$\int_0^1 f(x)dx = \mathbb{E}f$$

with the Riemann-integral on the left-hand side and the expectation of the random variable f with respect to the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$, where λ is the Lebesgue measure, on the right-hand side.

Now we consider a continuous function $p : \mathbb{R} \to [0, \infty)$ such that

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

and define a measure $\mathbb{P}$ on $\mathcal{B}(\mathbb{R})$ by

$$\mathbb{P}((a_1, b_1] \cap \cdots \cap (a_n, b_n]) := \sum_{i=1}^{n} \int_{a_i}^{b_i} p(x)dx$$

for $-\infty \leq a_1 \leq b_1 \leq \cdots \leq a_n \leq b_n \leq \infty$ (again with the convention that $(a, \infty] = (a, \infty)$) via Carathéodory's Theorem (Proposition 1.2.17). The function p is called **density** of the measure $\mathbb{P}$.

Proposition 3.3.2 *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function such that*

$$\int_{-\infty}^{\infty} |f(x)|p(x)dx < \infty.$$

Then

$$\int_{-\infty}^{\infty} f(x)p(x)dx = \mathbb{E}f$$

with the Riemann-integral on the left-hand side and the expectation of the random variable f with respect to the probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$ on the right-hand side.

Let us consider two examples indicating the difference between the Riemann-integral and our expected value.

Example 3.3.3 We give the standard example of a function which has an expected value, but which is not Riemann-integrable. Let

$$f(x) := \begin{cases} 1, & x \in [0,1] \text{ irrational} \\ 0, & x \in [0,1] \text{ rational} \end{cases}.$$

Then f is not Riemann integrable, but Lebesgue integrable with $\mathbb{E}f = 1$ if we use the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$.

Example 3.3.4 The expression

$$\lim_{t\to\infty} \int_0^t \frac{\sin x}{x} dx = \frac{\pi}{2}$$

is defined as limit in the Riemann sense although

$$\int_0^\infty \left(\frac{\sin x}{x}\right)^+ dx = \infty \quad \text{and} \quad \int_0^\infty \left(\frac{\sin x}{x}\right)^- dx = \infty.$$

Transporting this into a probabilistic setting we take the exponential distribution with parameter $\lambda > 0$ from Section 1.3.6. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = 0$ if $x \leq 0$ and $f(x) := \frac{\sin x}{\lambda x} e^{\lambda x}$ if $x > 0$ and recall that the exponential distribution μ_λ with parameter $\lambda > 0$ is given by the density $p_\lambda(x) = \mathbb{1}_{[0,\infty)}(x)\lambda e^{-\lambda x}$. The above yields that

$$\lim_{t\to\infty} \int_0^t f(x)p_\lambda(x)dx = \frac{\pi}{2}$$

but

$$\int_{\mathbb{R}} f(x)^+ d\mu_\lambda(x) = \int_{\mathbb{R}} f(x)^- d\mu_\lambda(x) = \infty.$$

Hence the expected value of f does not exists, but the Riemann-integral gives a way to define a value, which makes sense. The point of this example is that the Riemann-integral takes more information into the account than the rather abstract expected value.

3.4 Change of variables in the expected value

We want to prove a *change of variable formula* for the integrals $\int_\Omega f d\mathbb{P}$. In many cases, only by this formula it is possible to compute explicitly expected values.

Proposition 3.4.1 [CHANGE OF VARIABLES] *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $(E, \mathcal{E})$ be a measurable space, $\varphi : \Omega \to E$ be a measurable map, and $g : E \to \mathbb{R}$ be a random variable. Assume that $\mathbb{P}_\varphi$ is the image measure of $\mathbb{P}$ with respect to φ [1], that means*

$$\mathbb{P}_\varphi(A) = \mathbb{P}(\{\omega : \varphi(\omega) \in A\}) = \mathbb{P}(\varphi^{-1}(A)) \quad \text{for all} \quad A \in \mathcal{E}.$$

Then

$$\int_A g(\eta) d\mathbb{P}_\varphi(\eta) = \int_{\varphi^{-1}(A)} g(\varphi(\omega)) d\mathbb{P}(\omega)$$

for all $A \in \mathcal{E}$ in the sense that if one integral exists, the other exists as well, and their values are equal.

Proof. (i) Letting $\widetilde{g}(\eta) := \mathbb{1}_A(\eta) g(\eta)$ we have

$$\widetilde{g}(\varphi(\omega)) = \mathbb{1}_{\varphi^{-1}(A)}(\omega) g(\varphi(\omega))$$

so that it is sufficient to consider the case $A = E$. Hence we have to show that

$$\int_E g(\eta) d\mathbb{P}_\varphi(\eta) = \int_\Omega g(\varphi(\omega)) d\mathbb{P}(\omega).$$

(ii) Since, for $f(\omega) := g(\varphi(\omega))$ one has that $f^+ = g^+ \circ \varphi$ and $f^- = g^- \circ \varphi$ it is sufficient to consider the positive part of g and its negative part separately. In other words, we can assume that $g(\eta) \geq 0$ for all $\eta \in E$.

(iii) Assume now a sequence of measurable step-function $0 \leq g_n(\eta) \uparrow g(\eta)$ for all $\eta \in E$ which does exist according to Lemma 3.2.2 so that $g_n(\varphi(\omega)) \uparrow g(\varphi(\omega))$ for all $\omega \in \Omega$ as well. If we can show that

$$\int_E g_n(\eta) d\mathbb{P}_\varphi(\eta) = \int_\Omega g_n(\varphi(\omega)) d\mathbb{P}(\omega)$$

then we are done. By additivity it is enough to check $g_n(\eta) = \mathbb{1}_B(\eta)$ for some $B \in \mathcal{E}$ (if this is true for this case, then one can multiply by real numbers and can take sums and the equality remains true). But now we get

$$\begin{aligned}\int_E g_n(\eta) d\mathbb{P}_\varphi(\eta) = \mathbb{P}_\varphi(B) = \mathbb{P}(\varphi^{-1}(B)) &= \int_\Omega \mathbb{1}_{\varphi^{-1}(B)}(\omega) d\mathbb{P}(\omega) \\ &= \int_\Omega \mathbb{1}_B(\varphi(\omega)) d\mathbb{P}(\omega) = \int_\Omega g_n(\varphi(\omega)) d\mathbb{P}(\omega).\end{aligned}$$

□

Let us give an example for the change of variable formula.

[1] In other words, $\mathbb{P}_\varphi$ is the law of φ.

Definition 3.4.2 [MOMENTS] Assume that $n \in \{1, 2, ...\}$.

(1) For a random variable $f : \Omega \to \mathbb{R}$ the expected value $\mathbb{E}|f|^n$ is called **n-th absolute moment** of f. If $\mathbb{E}f^n$ exists, then $\mathbb{E}f^n$ is called **n-th moment** of f.

(2) For a probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ the expected value

$$\int_{\mathbb{R}} |x|^n d\mu(x)$$

is called **n-th absolute moment** of μ. If $\int_{\mathbb{R}} x^n d\mu(x)$ exists, then $\int_{\mathbb{R}} x^n d\mu(x)$ is called **n-th moment** of μ.

Corollary 3.4.3 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f : \Omega \to \mathbb{R}$ be a random variable with law $\mathbb{P}_f$. Then, for all $n = 1, 2, ...,$*

$$\mathbb{E}|f|^n = \int_{\mathbb{R}} |x|^n d\mathbb{P}_f(x) \quad \textit{and} \quad \mathbb{E}f^n = \int_{\mathbb{R}} x^n d\mathbb{P}_f(x),$$

where the latter equality has to be understood as follows: if one side exists, then the other exists as well and they coincide. If the law $\mathbb{P}_f$ has a density p in the sense of Proposition 3.3.2, then $\int_{\mathbb{R}} |x|^n d\mathbb{P}_f(x)$ can be replaced by $\int_{\mathbb{R}} |x|^n p(x)dx$ and $\int_{\mathbb{R}} x^n d\mathbb{P}_f(x)$ by $\int_{\mathbb{R}} x^n p(x)dx$.

3.5 Fubini's Theorem

In this section we consider iterated integrals, as they appear very often in applications, and show in FUBINI's [2] 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. Before we start with FUBINI's Theorem we need some preparations. First we recall the notion of a vector space.

Definition 3.5.1 [VECTOR SPACE] A set L equipped with operations "$+$" : $L \times L \to L$ and "$\cdot$" : $\mathbb{R} \times L \to L$ is called **vector space** over $\mathbb{R}$ if the following conditions are satisfied:

(1) $x + y = y + x$ for all $x, y \in L$.

(2) $x + (y + z) = (x + y) + z$ for all $x, y, z \in L$.

(3) There exists a $0 \in L$ such that $x + 0 = x$ for all $x \in L$.

(4) For all $x \in L$ there exists a $-x$ such that $x + (-x) = 0$.

[2] Guido Fubini, 19/01/1879 (Venice, Italy) - 06/06/1943 (New York, USA).

(5) $1\,x = x$.

(6) $\alpha(\beta x) = (\alpha\beta)x$ for all $\alpha, \beta \in \mathbb{R}$ and $x \in L$.

(7) $(\alpha + \beta)x = \alpha x + \beta x$ for all $\alpha, \beta \in \mathbb{R}$ and $x \in L$.

(8) $\alpha(x + y) = \alpha x + \alpha y$ for all $\alpha \in \mathbb{R}$ and $x, y \in L$.

Usually one uses the notation $x - y := x + (-y)$ and $-x + y := (-x) + y$ etc. Now we state the *Monotone Class Theorem*. It is a powerful tool by which, for example, measurability assertions can be proved.

Proposition 3.5.2 [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) *$\mathbb{1}_\Omega \in H$.*

(3) *If $f_n \in H$, $f_n \geq 0$, 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. See for example [5] (Theorem 3.14). □

For the following it is convenient to allow that the random variables may take infinite values.

Definition 3.5.3 [extended random variable] Let $(\Omega, \mathcal{F})$ be a measurable space. A function $f : \Omega \to \mathbb{R} \cup \{-\infty, \infty\}$ is called **extended random variable** iff

$$f^{-1}(B) := \{\omega : f(\omega) \in B\} \in \mathcal{F} \quad \text{for all} \quad B \in \mathcal{B}(\mathbb{R}) \text{ or } B = \{-\infty\}.$$

If we have a non-negative extended random variable, we let (for example)

$$\int_\Omega f d\mathbb{P} = \lim_{N \to \infty} \int_\Omega [f \wedge N] d\mathbb{P}.$$

For the following, we recall that the product space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mathbb{P}_1 \times \mathbb{P}_2)$ of the two probability spaces $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ was defined in Definition 1.2.19.

Proposition 3.5.4 [Fubini's Theorem for non-negative functions] *Let $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ be a non-negative $\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable function such that*

$$\int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) d(\mathbb{P}_1 \times \mathbb{P}_2)(\omega_1, \omega_2) < \infty. \tag{3.1}$$

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 $\omega_i^0 \in \Omega_i$.*

(2) *The functions*

$$\omega_1 \to \int_{\Omega_2} f(\omega_1, \omega_2) d\mathbb{P}_2(\omega_2) \quad \text{and} \quad \omega_2 \to \int_{\Omega_1} f(\omega_1, \omega_2) d\mathbb{P}_1(\omega_1)$$

are extended $\mathcal{F}_1$-measurable and $\mathcal{F}_2$-measurable, respectively, random variables.

(3) *One has that*

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) d(\mathbb{P}_1 \times \mathbb{P}_2) &= \int_{\Omega_1} \left[\int_{\Omega_2} f(\omega_1, \omega_2) d\mathbb{P}_2(\omega_2) \right] d\mathbb{P}_1(\omega_1) \\ &= \int_{\Omega_2} \left[\int_{\Omega_1} f(\omega_1, \omega_2) d\mathbb{P}_1(\omega_1) \right] d\mathbb{P}_2(\omega_2). \end{aligned}$$

It should be noted, that item (3) together with Formula (3.1) automatically implies that

$$\mathbb{P}_2 \left\{ \omega_2 : \int_{\Omega_1} f(\omega_1, \omega_2) d\mathbb{P}_1(\omega_1) = \infty \right\} = 0$$

and

$$\mathbb{P}_1 \left\{ \omega_1 : \int_{\Omega_2} f(\omega_1, \omega_2) d\mathbb{P}_2(\omega_2) = \infty \right\} = 0.$$

Proof of Proposition 3.5.4.

(i) 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 via $N \to \infty$ if we use Proposition 2.1.4 to get the necessary measurabilities (which also works for our extended random variables) and the monotone convergence formulated in Proposition 3.2.4 to get to values of the integrals. Hence we can assume for the following that $\sup_{\omega_1, \omega_2} f(\omega_1, \omega_2) < \infty$.

(ii) We want to apply the *Monotone Class Theorem* Proposition 3.5.2. Let $\mathcal{H}$ be the class of bounded $\mathcal{F}_1 \times \mathcal{F}_2$-measurable functions $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ such that

(*a*) 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 $\omega_i^0 \in \Omega_i$,

(*b*) the functions

$$\omega_1 \to \int_{\Omega_2} f(\omega_1, \omega_2) d\mathbb{P}_2(\omega_2) \quad \text{and} \quad \omega_2 \to \int_{\Omega_1} f(\omega_1, \omega_2) d\mathbb{P}_1(\omega_1)$$

are $\mathcal{F}_1$-measurable and $\mathcal{F}_2$-measurable, respectively,

(*c*) one has that

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) d(\mathbb{P}_1 \times \mathbb{P}_2) &= \int_{\Omega_1} \left[\int_{\Omega_2} f(\omega_1, \omega_2) d\mathbb{P}_2(\omega_2) \right] d\mathbb{P}_1(\omega_1) \\ &= \int_{\Omega_2} \left[\int_{\Omega_1} f(\omega_1, \omega_2) d\mathbb{P}_1(\omega_1) \right] d\mathbb{P}_2(\omega_2). \end{aligned}$$

Again, using Propositions 2.1.4 and 3.2.4 we see that $\mathcal{H}$ satisfies the assumptions (1), (2), and (3) of Proposition 3.5.2. As π-system I we take the system of all $F = A \times B$ with $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$. Letting $f(\omega_1, \omega_2) = \mathbb{1}_A(\omega_1)\mathbb{1}_B(\omega_2)$ we easily can check that $f \in \mathcal{H}$. For instance, property (c) follows from

$$\int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) d(\mathbb{P}_1 \times \mathbb{P}_2) = (\mathbb{P}_1 \times \mathbb{P}_2)(A \times B) = \mathbb{P}_1(A)\mathbb{P}_2(B)$$

and, for example,

$$\begin{aligned} \int_{\Omega_1} \left[\int_{\Omega_2} f(\omega_1, \omega_2) d\mathbb{P}_2(\omega_2) \right] d\mathbb{P}_1(\omega_1) &= \int_{\Omega_1} \mathbb{1}_A(\omega_1)\mathbb{P}_2(B) d\mathbb{P}_1(\omega_1) \\ &= \mathbb{P}_1(A)\mathbb{P}_2(B). \end{aligned}$$

Applying the *Monotone Class Theorem* Proposition 3.5.2 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 random variables $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$.

Proposition 3.5.5 [Fubini's Theorem] *Let $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ be an $\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable function such that*

$$\int_{\Omega_1 \times \Omega_2} |f(\omega_1, \omega_2)| d(\mathbb{P}_1 \times \mathbb{P}_2)(\omega_1, \omega_2) < \infty. \tag{3.2}$$

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* $\mathbb{P}_i(M_i) = 1$ *such that the integrals*

$$\int_{\Omega_1} f(\omega_1, \omega_2^0) d\mathbb{P}_1(\omega_1) \quad \text{and} \quad \int_{\Omega_2} f(\omega_1^0, \omega_2) d\mathbb{P}_2(\omega_2)$$

exist and are finite for all $\omega_i^0 \in M_i$.

(3) *The maps*

$$\omega_1 \to \mathbb{1}_{M_1}(\omega_1) \int_{\Omega_2} f(\omega_1, \omega_2) d\mathbb{P}_2(\omega_2)$$

and

$$\omega_2 \to \mathbb{1}_{M_2}(\omega_2) \int_{\Omega_1} f(\omega_1, \omega_2) d\mathbb{P}_1(\omega_1)$$

are $\mathcal{F}_1$*-measurable and* $\mathcal{F}_2$*-measurable, respectively, random variables.*

(4) *One has that*

$$\begin{aligned} &\int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) d(\mathbb{P}_1 \times \mathbb{P}_2) \\ &= \int_{\Omega_1} \left[\mathbb{1}_{M_1}(\omega_1) \int_{\Omega_2} f(\omega_1, \omega_2) d\mathbb{P}_2(\omega_2) \right] d\mathbb{P}_1(\omega_1) \\ &= \int_{\Omega_2} \left[\mathbb{1}_{M_2}(\omega_2) \int_{\Omega_1} f(\omega_1, \omega_2) d\mathbb{P}_1(\omega_1) \right] d\mathbb{P}_2(\omega_2). \end{aligned}$$

Remark 3.5.6 (1) Our understanding is that writing, for example, an expression like

$$\mathbb{1}_{M_2}(\omega_2) \int_{\Omega_1} f(\omega_1, \omega_2) d\mathbb{P}_1(\omega_1)$$

we only consider and compute the integral for $\omega_2 \in M_2$.

(2) The expressions in (3.1) and (3.2) can be replaced by

$$\int_{\Omega_1} \left[\int_{\Omega_2} f(\omega_1, \omega_2) d\mathbb{P}_2(\omega_2) \right] d\mathbb{P}_1(\omega_1) < \infty,$$

and the same expression with $|f(\omega_1, \omega_2)|$ instead of $f(\omega_1, \omega_2)$, respectively.

Proof of Proposition 3.5.5. The proposition follows by decomposing $f = f^+ - f^-$ and applying Proposition 3.5.4. □

In the following example we show how to compute the integral

$$\int_{-\infty}^{\infty} e^{-x^2} dx$$

by FUBINI's Theorem.

Example 3.5.7 Let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a non-negative continuous function. FUBINI's Theorem applied to the uniform distribution on $[-N, N]$, $N \in \{1, 2, ...\}$ gives that

$$\int_{-N}^{N} \left[\int_{-N}^{N} f(x,y) \frac{d\lambda(y)}{2N} \right] \frac{d\lambda(x)}{2N} = \int_{[-N,N]\times[-N,N]} f(x,y) \frac{d(\lambda \times \lambda)(x,y)}{(2N)^2}$$

where λ is the Lebesgue measure. Letting $f(x,y) := e^{-(x^2+y^2)}$, the above yields that

$$\int_{-N}^{N} \left[\int_{-N}^{N} e^{-x^2} e^{-y^2} d\lambda(y) \right] d\lambda(x) = \int_{[-N,N]\times[-N,N]} e^{-(x^2+y^2)} d(\lambda \times \lambda)(x,y).$$

For the left-hand side we get

$$\begin{aligned}
&\lim_{N\to\infty} \int_{-N}^{N} \left[\int_{-N}^{N} e^{-x^2} e^{-y^2} d\lambda(y) \right] d\lambda(x) \\
&\quad = \lim_{N\to\infty} \int_{-N}^{N} e^{-x^2} \left[\int_{-N}^{N} e^{-y^2} d\lambda(y) \right] d\lambda(x) \\
&\quad = \left[\lim_{N\to\infty} \int_{-N}^{N} e^{-x^2} d\lambda(x) \right]^2 \\
&\quad = \left[\int_{-\infty}^{\infty} e^{-x^2} d\lambda(x) \right]^2.
\end{aligned}$$

For the right-hand side we get

$$\begin{aligned}
&\lim_{N\to\infty} \int_{[-N,N]\times[-N,N]} e^{-(x^2+y^2)} d(\lambda \times \lambda)(x,y) \\
&\quad = \lim_{R\to\infty} \int_{x^2+y^2\le R^2} e^{-(x^2+y^2)} d(\lambda \times \lambda)(x,y) \\
&\quad = \lim_{R\to\infty} \int_0^R \int_0^{2\pi} e^{-r^2} r dr d\varphi \\
&\quad = \pi \lim_{R\to\infty} \left(1 - e^{-R^2} \right) \\
&\quad = \pi
\end{aligned}$$

where we have used polar coordinates. Comparing both sides gives

$$\int_{-\infty}^{\infty} e^{-x^2} d\lambda(x) = \sqrt{\pi}.$$

As corollary we show that the definition of the Gaussian measure in Section 1.3.5 was "correct".

Proposition 3.5.8 *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, $\int_{\mathbb{R}} p_{m,\sigma^2}(x)dx = 1$,

$$\int_{\mathbb{R}} x p_{m,\sigma^2}(x)dx = m, \quad \textit{and} \quad \int_{\mathbb{R}} (x-m)^2 p_{m,\sigma^2}(x)dx = \sigma^2. \tag{3.3}$$

In other words: if a random variable $f : \Omega \to \mathbb{R}$ has as law the normal distribution $\mathcal{N}_{m,\sigma^2}$, then

$$\mathbb{E}f = m \quad \textit{and} \quad \mathbb{E}(f - \mathbb{E}f)^2 = \sigma^2. \tag{3.4}$$

Proof. By the change of variable $x \to m + \sigma x$ it is sufficient to show the statements for $m = 0$ and $\sigma = 1$. Firstly, by putting $x = z/\sqrt{2}$ one gets

$$1 = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-x^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz$$

where we have used Example 3.5.7 so that $\int_{\mathbb{R}} p_{0,1}(x)dx = 1$. Secondly,

$$\int_{\mathbb{R}} x p_{0,1}(x)dx = 0$$

follows from the symmetry of the density $p_{0,1}(x) = p_{0,1}(-x)$. Finally, by $(x \exp(-x^2/2))' = \exp(-x^2/2) - x^2 \exp(-x^2/2)$ one can also compute that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = 1.$$

□

We close this section with a "counterexample" to FUBINI's Theorem.

Example 3.5.9 Let $\Omega = [-1,1] \times [-1,1]$ and μ be the uniform distribution on $[-1,1]$ (see Section 1.3.4). The function

$$f(x,y) := \frac{xy}{(x^2+y^2)^2}$$

for $(x,y) \neq (0,0)$ and $f(0,0) := 0$ is not integrable on Ω, even though the iterated integrals exist end are equal. In fact

$$\int_{-1}^{1} f(x,y)d\mu(x) = 0 \quad \text{and} \quad \int_{-1}^{1} f(x,y)d\mu(y) = 0$$

so that

$$\int_{-1}^{1} \left(\int_{-1}^{1} f(x,y)d\mu(x) \right) d\mu(y) = \int_{-1}^{1} \left(\int_{-1}^{1} f(x,y)d\mu(y) \right) d\mu(x) = 0.$$

On the other hand, using polar coordinates we get

$$\begin{aligned} 4\int_{[-1,1]\times[-1,1]} |f(x,y)|d(\mu\times\mu)(x,y) &\geq \int_0^1\int_0^{2\pi} \frac{|\sin\varphi\cos\varphi|}{r}d\varphi dr \\ &= 2\int_0^1 \frac{1}{r}dr = \infty. \end{aligned}$$

The inequality holds because on the right hand side we integrate only over the area $\{(x,y) : x^2+y^2 \leq 1\}$ which is a subset of $[-1,1]\times[-1,1]$ and

$$\int_0^{2\pi} |\sin\varphi\cos\varphi|d\varphi = 4\int_0^{\pi/2} \sin\varphi\cos\varphi d\varphi = 2$$

follows by a symmetry argument.

3.6 Some inequalities

In this section we prove some basic inequalities.

Proposition 3.6.1 [CHEBYSHEV'S INEQUALITY] [3] *Let f be a non-negative integrable random variable defined on a probability space $(\Omega,\mathcal{F},\mathbb{P})$. Then, for all $\lambda > 0$,*

$$\mathbb{P}(\{\omega : f(\omega) \geq \lambda\}) \leq \frac{\mathbb{E}f}{\lambda}.$$

Proof. We simply have

$$\lambda\mathbb{P}(\{\omega : f(\omega) \geq \lambda\}) = \lambda\mathbb{E}\mathbb{1}_{\{f\geq\lambda\}} \leq \mathbb{E}f\mathbb{1}_{\{f\geq\lambda\}} \leq \mathbb{E}f.$$

□

Definition 3.6.2 [CONVEXITY] A function $g : \mathbb{R} \to \mathbb{R}$ is **convex** if and only if

$$g(px+(1-p)y) \leq pg(x)+(1-p)g(y)$$

for all $0 \leq p \leq 1$ and all $x, y \in \mathbb{R}$. A function g is **concave** if $-g$ is convex.

Every convex function $g : \mathbb{R} \to \mathbb{R}$ is continuous (check!) and hence $(\mathcal{B}(\mathbb{R}),\mathcal{B}(\mathbb{R}))$-measurable.

Proposition 3.6.3 [JENSEN'S INEQUALITY] [4] *If $g : \mathbb{R} \to \mathbb{R}$ is convex and $f : \Omega \to \mathbb{R}$ a random variable with $\mathbb{E}|f| < \infty$, then*

$$g(\mathbb{E}f) \leq \mathbb{E}g(f)$$

where the expected value on the right-hand side might be infinity.

[3] Pafnuty Lvovich Chebyshev, 16/05/1821 (Okatovo, Russia) - 08/12/1894 (St Petersburg, Russia)

[4] 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

$$ax_0 + b = g(x_0) \quad \text{and} \quad ax + b \leq g(x)$$

for all $x \in \mathbb{R}$. It follows $af(\omega) + b \leq g(f(\omega))$ for all $\omega \in \Omega$ and

$$g(\mathbb{E}f) = a\mathbb{E}f + b = \mathbb{E}(af + b) \leq \mathbb{E}g(f).$$

□

Example 3.6.4 (1) The function $g(x) := |x|$ is convex so that, for any integrable f,

$$|\mathbb{E}f| \leq \mathbb{E}|f|.$$

(2) For $1 \leq p < \infty$ the function $g(x) := |x|^p$ is convex, so that JENSEN's inequality applied to $|f|$ gives that

$$(\mathbb{E}|f|)^p \leq \mathbb{E}|f|^p.$$

For the second case in the example above there is another way we can go. It uses the famous HÖLDER-inequality.

Proposition 3.6.5 [HÖLDER'S INEQUALITY] [5] *Assume a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variables $f, g : \Omega \to \mathbb{R}$. If $1 < p, q < \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$, then*

$$\mathbb{E}|fg| \leq (\mathbb{E}|f|^p)^{\frac{1}{p}} (\mathbb{E}|g|^q)^{\frac{1}{q}}.$$

Proof. We can assume that $\mathbb{E}|f|^p > 0$ and $\mathbb{E}|g|^q > 0$. For example, assuming $\mathbb{E}|f|^p = 0$ would imply $|f|^p = 0$ a.s. according to Proposition 3.2.1 so that $fg = 0$ a.s. and $\mathbb{E}|fg| = 0$. Hence we may set

$$\tilde{f} := \frac{f}{(\mathbb{E}|f|^p)^{\frac{1}{p}}} \quad \text{and} \quad \tilde{g} := \frac{g}{(\mathbb{E}|g|^q)^{\frac{1}{q}}}.$$

We notice that

$$x^a y^b \leq ax + by$$

for $x, y \geq 0$ and positive a, b with $a + b = 1$, which follows from the concavity of the logarithm (we can assume for a moment that $x, y > 0$)

$$\ln(ax + by) \geq a \ln x + b \ln y = \ln x^a + \ln y^b = \ln x^a y^b.$$

[5] Otto Ludwig Hölder, 22/12/1859 (Stuttgart, Germany) - 29/08/1937 (Leipzig, Germany).

Setting $x := |\tilde{f}|^p$, $y := |\tilde{g}|^q$, $a := \frac{1}{p}$, and $b := \frac{1}{q}$, we get

$$|\tilde{f}\tilde{g}| = x^a y^b \leq ax + by = \frac{1}{p}|\tilde{f}|^p + \frac{1}{q}|\tilde{g}|^q$$

and

$$\mathbb{E}|\tilde{f}\tilde{g}| \leq \frac{1}{p}\mathbb{E}|\tilde{f}|^p + \frac{1}{q}\mathbb{E}|\tilde{g}|^q = \frac{1}{p} + \frac{1}{q} = 1.$$

On the other hand side,

$$\mathbb{E}|\tilde{f}\tilde{g}| = \frac{\mathbb{E}|fg|}{(\mathbb{E}|f|^p)^{\frac{1}{p}}\,(\mathbb{E}|g|^q)^{\frac{1}{q}}}$$

so that we are done. □

Corollary 3.6.6 *For $0 < p < q < \infty$ one has that $(\mathbb{E}|f|^p)^{\frac{1}{p}} \leq (\mathbb{E}|f|^q)^{\frac{1}{q}}$.*

The proof is an exercise.

Corollary 3.6.7 [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*

$$\sum_{n=1}^{\infty} |a_n b_n| \leq \left(\sum_{n=1}^{\infty} |a_n|^p\right)^{\frac{1}{p}} \left(\sum_{n=1}^{\infty} |b_n|^q\right)^{\frac{1}{q}}.$$

Proof. It is sufficient to prove the inequality for finite sequences $(b_n)_{n=1}^{N}$ since by letting $N \to \infty$ we get the desired inequality for infinite sequences. Let $\Omega = \{1, ..., N\}$, $\mathcal{F} := 2^{\Omega}$, and $\mathbb{P}(\{k\}) := 1/N$. Defining $f, g : \Omega \to \mathbb{R}$ by $f(k) := a_k$ and $g(k) := b_k$ we get

$$\frac{1}{N}\sum_{n=1}^{N} |a_n b_n| \leq \left(\frac{1}{N}\sum_{n=1}^{N} |a_n|^p\right)^{\frac{1}{p}} \left(\frac{1}{N}\sum_{n=1}^{N} |b_n|^q\right)^{\frac{1}{q}}$$

from Proposition 3.6.5. Multiplying by N and letting $N \to \infty$ gives our assertion. □

Proposition 3.6.8 [Minkowski inequality] [6] *Assume a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, random variables $f, g : \Omega \to \mathbb{R}$, and $1 \leq p < \infty$. Then*

$$(\mathbb{E}|f + g|^p)^{\frac{1}{p}} \leq (\mathbb{E}|f|^p)^{\frac{1}{p}} + (\mathbb{E}|g|^p)^{\frac{1}{p}}. \tag{3.5}$$

[6] Hermann Minkowski, 22/06/1864 (Alexotas, Russian Empire; now Kaunas, Lithuania) - 12/01/1909 (Göttingen, Germany).

Proof. For $p = 1$ the inequality follows from $|f + g| \leq |f| + |g|$. So assume that $1 < p < \infty$. The convexity of $x \to |x|^p$ gives that

$$\left|\frac{a+b}{2}\right|^p \leq \frac{|a|^p + |b|^p}{2}$$

and $(a+b)^p \leq 2^{p-1}(a^p+b^p)$ for $a, b \geq 0$. Consequently, $|f+g|^p \leq (|f|+|g|)^p \leq 2^{p-1}(|f|^p + |g|^p)$ and

$$\mathbb{E}|f + g|^p \leq 2^{p-1}(\mathbb{E}|f|^p + \mathbb{E}|g|^p).$$

Assuming now that $(\mathbb{E}|f|^p)^{\frac{1}{p}} + (\mathbb{E}|g|^p)^{\frac{1}{p}} < \infty$, otherwise there is nothing to prove, we get that $\mathbb{E}|f+g|^p < \infty$ as well by the above considerations. Taking $1 < q < \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$, we continue with

$$\begin{aligned}
\mathbb{E}|f + g|^p &= \mathbb{E}|f + g||f + g|^{p-1} \\
&\leq \mathbb{E}(|f| + |g|)|f + g|^{p-1} \\
&= \mathbb{E}|f||f + g|^{p-1} + \mathbb{E}|g||f + g|^{p-1} \\
&\leq (\mathbb{E}|f|^p)^{\frac{1}{p}} \left(\mathbb{E}|f + g|^{(p-1)q}\right)^{\frac{1}{q}} + (\mathbb{E}|g|^p)^{\frac{1}{p}} \left(\mathbb{E}|f + g|^{(p-1)q}\right)^{\frac{1}{q}},
\end{aligned}$$

where we have used HÖLDER's inequality. Since $(p-1)q = p$, (3.5) follows by dividing the above inequality by $(\mathbb{E}|f + g|^p)^{\frac{1}{q}}$ and taking into the account $1 - \frac{1}{q} = \frac{1}{p}$. □

We close with a simple deviation inequality for f.

Corollary 3.6.9 *Let f be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{E}f^2 < \infty$. Then one has, for all $\lambda > 0$,*

$$\mathbb{P}(|f - \mathbb{E}f| \geq \lambda) \leq \frac{\mathbb{E}(f - \mathbb{E}f)^2}{\lambda^2} \leq \frac{\mathbb{E}f^2}{\lambda^2}.$$

Proof. From Corollary 3.6.6 we get that $\mathbb{E}|f| < \infty$ so that $\mathbb{E}f$ exists. Applying Proposition 3.6.1 to $|f - \mathbb{E}f|^2$ gives that

$$\mathbb{P}(\{|f - \mathbb{E}f| \geq \lambda\}) = \mathbb{P}(\{|f - \mathbb{E}f|^2 \geq \lambda^2\}) \leq \frac{\mathbb{E}|f - \mathbb{E}f|^2}{\lambda^2}.$$

Finally, we use that $\mathbb{E}(f - \mathbb{E}f)^2 = \mathbb{E}f^2 - (\mathbb{E}f)^2 \leq \mathbb{E}f^2$. □

3.7 Theorem of Radon-Nikodym

Definition 3.7.1 (Signed measures) Let $(\Omega, \mathcal{F})$ be a measurable space.

(*i*) A map $\mu : \mathcal{F} \to \mathbb{R}$ is called (finite) *signed measure* if and only if

$$\mu = \alpha\mu^+ - \beta\mu^-,$$

where $\alpha, \beta \geq 0$ and μ^+ and μ^- are probability measures on $\mathcal{F}$.

(*ii*) Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and that μ is a signed measure on $(\Omega, \mathcal{F})$. Then $\mu \ll \mathbb{P}$ (μ is *absolutely continuous* with respect to $\mathbb{P}$) if and only if

$$\mathbb{P}(A) = 0 \quad \text{implies} \quad \mu(A) = 0.$$

Example 3.7.2 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $L : \Omega \to \mathbb{R}$ be an integrable random variable, and

$$\mu(A) := \int_A L d\mathbb{P}.$$

Then μ is a signed measure and $\mu \ll \mathbb{P}$.

Proof. We let $L^+ := \max\{L, 0\}$ and $L^- := \max\{-L, 0\}$ so that $L = L^+ - L^-$. Assume that $\int_\Omega L^\pm d\mathbb{P} > 0$ and define

$$\mu^\pm(A) = \frac{\int_\Omega \mathbb{1}_A L^\pm d\mathbb{P}}{\int_\Omega L^\pm d\mathbb{P}}.$$

Now we check that $\mu^\pm$ are probability measures. First we have that

$$\mu^\pm(\Omega) = \frac{\int_\Omega \mathbb{1}_\Omega L^\pm d\mathbb{P}}{\int_\Omega L^\pm d\mathbb{P}} = 1.$$

Then assume $A_n \in \mathcal{F}$ to be disjoint sets such that $A = \bigcup_{n=1}^{\infty} A_n$. Set $\alpha := \int_\Omega L^+ d\mathbb{P}$. Then

$$\begin{aligned}
\mu^+\left(\bigcup_{n=1}^{\infty} A_n\right) &= \frac{1}{\alpha}\int_\Omega \mathbb{1}_{\bigcup_{n=1}^{\infty} A_n} L^+ d\mathbb{P} \\
&= \frac{1}{\alpha}\int_\Omega \left(\sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega)\right) L^+ d\mathbb{P} \\
&= \frac{1}{\alpha}\int_\Omega \lim_{N\to\infty}\left(\sum_{n=1}^{N} \mathbb{1}_{A_n}\right) L^+ d\mathbb{P} \\
&= \frac{1}{\alpha}\lim_{N\to\infty}\int_\Omega \left(\sum_{n=1}^{N} \mathbb{1}_{A_n}\right) L^+ d\mathbb{P}
\end{aligned}$$

$$= \sum_{n=1}^{\infty} \mu^+(A_n)$$

where we have used LEBESGUE's dominated convergence theorem. The same can be done for L^-. □

Theorem 3.7.3 (Radon-Nikodym) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and μ a signed measure with $\mu \ll \mathbb{P}$. Then there exists an integrable random variable $L : \Omega \to \mathbb{R}$ such that*

$$\mu(A) = \int_A L(\omega) d\mathbb{P}(\omega), \ A \in \mathcal{F}. \tag{3.6}$$

The random variable L is unique in the following sense. If L and L' are random variables satisfying (3.6), then

$$\mathbb{P}(L \neq L') = 0.$$

The Radon-Nikodym theorem was proved by Radon [7] in 1913 in the case of $\mathbb{R}^n$. The extension to the general case was done by Nikodym [8] in 1930.

Definition 3.7.4 L is called RADON-NIKODYM derivative. We shall write

$$L = \frac{d\mu}{d\mathbb{P}}.$$

We should keep in mind the rule

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

so that '$d\mu = L d\mathbb{P}$'.

3.8 Modes of convergence

First we introduce some basic types of convergence.

Definition 3.8.1 [TYPES OF CONVERGENCE] Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f, f_1, f_2, ... : \Omega \to \mathbb{R}$ random variables.

(1) The sequence $(f_n)_{n=1}^{\infty}$ **converges almost surely** (a.s.) or **with probability 1** to f ($f_n \to f$ a.s. or $f_n \to f$ $\mathbb{P}$-a.s.) if and only if

$$\mathbb{P}(\{\omega : f_n(\omega) \to f(\omega) \ \text{ as } \ n \to \infty\}) = 1.$$

[7] Johann Radon, 16/12/1887 (Tetschen, Bohemia; now Decin, Czech Republic) - 25/05/1956 (Vienna, Austria).

[8] Otton Marcin Nikodym, 13/08/1887 (Zablotow, Galicia, Austria-Hungary; now Ukraine) - 04/05/1974 (Utica, USA).

(2) The sequence $(f_n)_{n=1}^{\infty}$ **converges in probability** to f ($f_n \xrightarrow{\mathbb{P}} f$) if and only if for all $\varepsilon > 0$ one has

$$\mathbb{P}(\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}) \to 0 \text{ as } n \to \infty.$$

(3) If $0 < p < \infty$, then the sequence $(f_n)_{n=1}^{\infty}$ **converges with respect to** L_p or **in the** L_p**-mean** to f ($f_n \xrightarrow{L_p} f$) if and only if

$$\mathbb{E}|f_n - f|^p \to 0 \text{ as } n \to \infty.$$

Note that $\{\omega : f_n(\omega) \to f(\omega) \text{ as } n \to \infty\}$ and $\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}$ are measurable sets.
For the above types of convergence the random variables have to be defined on the same probability space. There is a variant without this assumption.

Definition 3.8.2 [Convergence in distribution] Let $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ and $(\Omega, \mathcal{F}, \mathbb{P})$ be probability spaces and let $f_n : \Omega_n \to \mathbb{R}$ and $f : \Omega \to \mathbb{R}$ be random variables. Then the sequence $(f_n)_{n=1}^{\infty}$ **converges in distribution** to f ($f_n \xrightarrow{d} f$) if and only if

$$\mathbb{E}\psi(f_n) \to \mathbb{E}\psi(f) \text{ as } n \to \infty$$

for all bounded and continuous functions $\psi : \mathbb{R} \to \mathbb{R}$.

We have the following relations between the above types of convergence.

Proposition 3.8.3 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f, f_1, f_2, ... : \Omega \to \mathbb{R}$ be random variables.*

(1) *If $f_n \to f$ a.s., then $f_n \xrightarrow{\mathbb{P}} f$.*

(2) *If $0 < p < \infty$ and $f_n \xrightarrow{L_p} f$, then $f_n \xrightarrow{\mathbb{P}} f$.*

(3) *If $f_n \xrightarrow{\mathbb{P}} f$, then $f_n \xrightarrow{d} f$.*

(4) *One has that $f_n \xrightarrow{d} f$ if and only if $F_{f_n}(x) \to F_f(x)$ at each point x of continuity of $F_f(x)$, where F_{f_n} and F_f are the distribution-functions of f_n and f, respectively.*

(5) *If $f_n \xrightarrow{\mathbb{P}} f$, then there is a subsequence $1 \leq n_1 < n_2 < n_3 < \cdots$ such that $f_{n_k} \to f$ a.s. as $k \to \infty$.*

Proof. See [4]. □

Example 3.8.4 Assume $([0,1], \mathcal{B}([0,1]), \lambda)$ where λ is the Lebesgue measure. We take
$f_1 = \mathbb{1}_{[0,\frac{1}{2})}$, $f_2 = \mathbb{1}_{[\frac{1}{2},1]}$,
$f_3 = \mathbb{1}_{[0,\frac{1}{4})}$, $f_4 = \mathbb{1}_{[\frac{1}{4},\frac{1}{2}]}$, $f_5 = \mathbb{1}_{[\frac{1}{2},\frac{3}{4})}$, $f_6 = \mathbb{1}_{[\frac{3}{4},1]}$,
$f_7 = \mathbb{1}_{[0,\frac{1}{8})}, \ldots$
This implies $\lim_{n\to\infty} f_n(x) \not\to 0$ for all $x \in [0,1]$. But it holds convergence in probability $f_n \xrightarrow{\lambda} 0$: choosing $0 < \varepsilon < 1$ we get

$$\begin{aligned} \lambda(\{x \in [0,1] : |f_n(x)| > \varepsilon\}) &= \lambda(\{x \in [0,1] : f_n(x) \neq 0\}) \\ &= \begin{cases} \frac{1}{2} & \text{if } n = 1,2 \\ \frac{1}{4} & \text{if } n = 3,4,\ldots,6 \\ \frac{1}{8} & \text{if } n = 7,\ldots \\ \vdots \end{cases} \end{aligned}$$

As a preview on the next probability course we give some examples for the above concepts of convergence. We start with the *weak law of large numbers* as an example of the convergence in probability:

Proposition 3.8.5 [Weak law of large numbers] *Let $(f_n)_{n=1}^{\infty}$ be a sequence of independent random variables with*

$$\mathbb{E}f_k = m \quad \text{and} \quad \mathbb{E}(f_k - m)^2 = \sigma^2 \quad \text{for all } k = 1,2,\ldots.$$

Then

$$\frac{f_1 + \cdots + f_n}{n} \xrightarrow{\mathbb{P}} m \quad \text{as} \quad n \to \infty,$$

that means, for each $\varepsilon > 0$,

$$\lim_n \mathbb{P}\left(\left\{\omega : |\frac{f_1 + \cdots + f_n}{n} - m| > \varepsilon\right\}\right) \to 0.$$

Proof. By Chebyshev's inequality (Corollary 3.6.9) we have that

$$\begin{aligned} \mathbb{P}\left(\left\{\omega : \left|\frac{f_1 + \cdots + f_n - nm}{n}\right| > \varepsilon\right\}\right) &\leq \frac{\mathbb{E}|f_1 + \cdots + f_n - nm|^2}{n^2\varepsilon^2} \\ &= \frac{\mathbb{E}\left(\sum_{k=1}^n (f_k - m)\right)^2}{n^2\varepsilon^2} \\ &= \frac{n\sigma^2}{n^2\varepsilon^2} \to 0 \end{aligned}$$

as $n \to \infty$. □

Using a stronger condition, we get easily more: the almost sure convergence instead of the convergence in probability. This gives a form of the *strong law of large numbers*.

Proposition 3.8.6 [Strong law of large numbers]
Let $(f_n)_{n=1}^{\infty}$ be a sequence of independent random variables with $\mathbb{E}f_k = 0$, $k = 1, 2, \ldots,$ and $c := \sup_n \mathbb{E}f_n^4 < \infty$. Then

$$\frac{f_1 + \cdots + f_n}{n} \overset{a.s.}{\to} 0.$$

Proof. Let $S_n := \sum_{k=1}^{n} f_k$. It holds

$$\begin{aligned}\mathbb{E}S_n^4 = \mathbb{E}\left(\sum_{k=1}^{n} f_k\right)^4 &= \mathbb{E}\sum_{i,j,k,l,=1}^{n} f_i f_j f_k f_l \\ &= \sum_{k=1}^{n}\mathbb{E}f_k^4 + 3\sum_{\substack{k,l=1 \\ k\neq l}}^{n}\mathbb{E}f_k^2\mathbb{E}f_l^2,\end{aligned}$$

because for distinct $\{i, j, k, l\}$ it holds

$$\mathbb{E}f_i f_j^3 = \mathbb{E}f_i f_j^2 f_k = \mathbb{E}f_i f_j f_k f_l = 0$$

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 $\mathbb{E}|f_j|^3 \leq (\mathbb{E}|f_j|^4)^{\frac{3}{4}} \leq c^{\frac{3}{4}}$. Moreover, by Jensen's inequality,

$$\left(\mathbb{E}f_k^2\right)^2 \leq \mathbb{E}f_k^4 \leq c.$$

Hence $\mathbb{E}f_k^2 f_l^2 = \mathbb{E}f_k^2\mathbb{E}f_l^2 \leq c$ for $k \neq l$. Consequently,

$$\mathbb{E}S_n^4 \leq nc + 3n(n-1)c \leq 3cn^2,$$

and

$$\mathbb{E}\sum_{n=1}^{\infty}\frac{S_n^4}{n^4} = \sum_{n=1}^{\infty}\mathbb{E}\frac{S_n^4}{n^4} \leq \sum_{n=1}^{\infty}\frac{3c}{n^2} < \infty.$$

This implies that $\frac{S_n^4}{n^4} \overset{\text{a.s.}}{\to} 0$ and therefore $\frac{S_n}{n} \overset{\text{a.s.}}{\to} 0$. □

There are several *strong laws of large numbers* with other, in particular weaker, conditions. We close with a fundamental example concerning the convergence in distribution: the Central Limit Theorem (CLT). For this we need

Definition 3.8.7 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability spaces. A sequence of **I**ndependent random variables $f_n : \Omega \to \mathbb{R}$ is called **I**dentically **D**istributed (i.i.d.) provided that the random variables f_n have the same law, that means

$$\mathbb{P}(f_n \leq \lambda) = \mathbb{P}(f_k \leq \lambda)$$

for all $n, k = 1, 2, ...$ and all $\lambda \in \mathbb{R}$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(f_n)_{n=1}^{\infty}$ be a sequence of i.i.d. random variables with $\mathbb{E}f_1 = 0$ and $\mathbb{E}f_1^2 = \sigma^2$. By the law of large numbers we know

$$\frac{f_1 + \cdots + f_n}{n} \xrightarrow{\mathbb{P}} 0.$$

Hence the law of the limit is the DIRAC-measure δ_0. Is there a right scaling factor $c(n)$ such that

$$\frac{f_1 + \cdots + f_n}{c(n)} \to g,$$

where g is a non-degenerate random variable in the sense that $\mathbb{P}_g \neq \delta_0$? And in which sense does the convergence take place? The answer is the following

Proposition 3.8.8 [CENTRAL LIMIT THEOREM] *Let $(f_n)_{n=1}^{\infty}$ be a sequence of i.i.d. random variables with $\mathbb{E}f_1 = 0$ and $\mathbb{E}f_1^2 = \sigma^2 > 0$. Then*

$$\mathbb{P}\left(\frac{f_1 + \cdots + f_n}{\sigma\sqrt{n}} \leq x\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du$$

for all $x \in \mathbb{R}$ as $n \to \infty$, that means that

$$\frac{f_1 + \cdots + f_n}{\sigma\sqrt{n}} \xrightarrow{d} g$$

for any g with $\mathbb{P}(g \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du$.

Chapter 4

Exercises

4.1 Probability spaces

1. Prove that $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

2. Prove that $\left(\bigcup_{i \in I} A_i\right)^c = \bigcap_{i \in I} A_i^c$ where $A_i \subseteq \Omega$ and I is an arbitrary index set.

3. Given a set Ω and two non-empty sets $A, B \subseteq \Omega$ such that $A \cap B = \emptyset$. Give all elements of the smallest σ-algebra $\mathcal{F}$ on Ω which contains A and B.

4. Let $x \in \mathbb{R}$. Is it true that $\{x\} \in \mathcal{B}(\mathbb{R})$, where $\{x\}$ is the set, consisting of the element x only?

5. Assume that $\mathbb{Q}$ is the set of rational numbers. Is it true that $\mathbb{Q} \in \mathcal{B}(\mathbb{R})$?

6. Given two dice with numbers $\{1, 2, ..., 6\}$. Assume that the probability that one die shows a certain number is $\frac{1}{6}$. What is the probability that the sum of the two dice is $m \in \{1, 2, ..., 12\}$?

7. There are three students. Assuming that a year has 365 days, what is the probability that at least two of them have their birthday at the same day ?

8.* **Definition:** The system $\mathcal{F} \subseteq 2^\Omega$ is a **monotonic class**, if

   (a) $A_1, A_2, ... \in \mathcal{F}$, $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \Longrightarrow \bigcup_n A_n \in \mathcal{F}$, and

   (b) $A_1, A_2, ... \in \mathcal{F}$, $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots \Longrightarrow \bigcap_n A_n \in \mathcal{F}$.

   Show that if $\mathcal{F} \subseteq 2^\Omega$ is an *algebra* and a *monotonic class*, then $\mathcal{F}$ is a σ-algebra.

9. Let E, F, G, be three events. Find expressions for the events that of E, F, G,

(a) only F occurs,

(b) both E and F but not G occur,

(c) at least one event occurs,

(d) at least two events occur,

(e) all three events occur,

(f) none occurs,

(g) at most one occurs,

(h) at most two occur.

10. Show that in the definition of an algebra (Definition 1.1.1 one can replace

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

    by

    (3") $A, B \in \mathcal{F}$ implies that $A \cap B \in \mathcal{F}$.

11. Prove that $A \setminus B \in \mathcal{F}$ if $\mathcal{F}$ is an algebra and $A, B \in \mathcal{F}$.

12. Prove that $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$ if $\mathcal{F}$ is a σ-algebra and $A_1, A_2, ... \in \mathcal{F}$.

13. Give an example where the union of two σ-algebras is not a σ-algebra.

14. Let $\mathcal{F}$ be a σ-algebra and $A \in \mathcal{F}$. Prove that

$$\mathcal{G} := \{B \cap A : B \in \mathcal{F}\}$$

    is a σ-algebra.

15. Prove that

$$\{B \cap [\alpha, \beta] : B \in \mathcal{B}(\mathbb{R})\} = \sigma\{[a, b] : \alpha \leq a < b \leq \beta\}$$

    and that this σ-algebra is the smallest σ-algebra generated by the subsets $A \subseteq [\alpha, \beta]$ which are open within $[\alpha, \beta]$. The generated σ-algebra is denoted by $\mathcal{B}([\alpha, \beta])$.

16. Show the equality $\sigma(\mathcal{G}_2) = \sigma(\mathcal{G}_4) = \sigma(\mathcal{G}_0)$ in Proposition 1.1.8 holds.

17. Show that $\mathcal{A} \subseteq \mathcal{B}$ implies that $\sigma(\mathcal{A}) \subseteq \sigma(\mathcal{B})$.

18. Prove $\sigma(\sigma(\mathcal{G})) = \sigma(\mathcal{G})$.

19. Let $\Omega \neq \emptyset$, $A \subseteq \Omega$, $A \neq \emptyset$ and $\mathcal{F} := 2^{\Omega}$. Define

$$\mathbb{P}(B) := \begin{cases} 1 & : \quad B \cap A \neq \emptyset \\ 0 & : \quad B \cap A = \emptyset \end{cases}.$$

    Is $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space?

20. Prove Proposition 1.2.6 (7).

21. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A \in \mathcal{F}$ such that $\mathbb{P}(A) > 0$. Show that $(\Omega, \mathcal{F}, \mu)$ is a probability space, where

$$\mu(B) := \mathbb{P}(B|A).$$

22. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Show that if $A, B \in \mathcal{F}$ are independent this implies

    (a) A and B^c are independent,

    (b) A^c and B^c are independent.

23. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the model of rolling two dice, i.e. $\Omega = \{(k,l) : 1 \leq k, l \leq 6\}$, $\mathcal{F} = 2^\Omega$, and $\mathbb{P}((k,l)) = \frac{1}{36}$ for all $(k,l) \in \Omega$. Assume

$$\begin{aligned} A &:= \{(k,l) : l = 1, 2 \text{ or } 5\}, \\ B &:= \{(k,l) : l = 4, 5 \text{ or } 6\}, \\ C &:= \{(k,l) : k + l = 9\}. \end{aligned}$$

    Show that

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

    Are A, B, C independent ?

24. (a) Let $\Omega := \{1, ..., 6\}$, $\mathcal{F} := 2^\Omega$ and $\mathbb{P}(B) := \frac{1}{6}\text{card}(B)$. We define $A := \{1, 4\}$ and $B := \{2, 5\}$. Are A and B independent?

    (b) Let $\Omega := \{(k,l) : k, l = 1, ..., 6\}$, $\mathcal{F} := 2^\Omega$ and $\mathbb{P}(B) := \frac{1}{36}\text{card}(B)$. We define

$$A := \{(k,l) : k = 1 \text{ or } k = 4\} \text{ and } B := \{(k,l) : l = 2 \text{ or } l = 5\}.$$

    Are A and B independent?

25. In a certain community 60 % of the families own their own car, 30 % own their own home, and 20% own both (their own car and their own home). If a family is randomly chosen, what is the probability that this family owns a car or a house but not both?

26. Prove Bayes' formula: Proposition 1.2.15.

27. Suppose we have 10 coins which are such that if the ith one is flipped then heads will appear with probability $\frac{i}{10}$, $i = 1, 2, \ldots 10$. One chooses randomly one of the coins. What is the conditionally probability that one has chosen the fifth coin given it had shown heads after flipping? **Hint:** Use Bayes' formula.

28.* A class that is both, a π-system and a λ-system, is a σ-algebra.

29. Prove Property (3) in Lemma 1.4.4.

30.* Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and assume $A_1, A_2, ..., A_n \in \mathcal{F}$ are independent. Show that then $A_1^c, A_2^c, ..., A_n^c$ are independent.

31. Let us play the following game: If we roll a die it shows each of the numbers $\{1, 2, 3, 4, 5, 6\}$ with probability $\frac{1}{6}$. Let $k_1, k_2, ... \in \{1, 2, 3, ...\}$ be a given sequence.

    (a) First go: One has to roll the die k_1 times. If it did show all k_1 times 6 we won.

    (b) Second go: We roll k_2 times. Again, we win if it would all k_2 times show 6.

    And so on... Show that

    (a) The probability of winning infinitely many often is 1 if and only if
    $$\sum_{n=1}^{\infty} \left(\frac{1}{6}\right)^{k_n} = \infty,$$

    (b) The probability of loosing infinitely many often is always 1.

    Hint: Use the lemma of Borel-Cantelli.

32. Let $n \geq 1$, $k \in \{0, 1, ..., n\}$ and
    $$\binom{n}{k} := \frac{n!}{k!(n-k)!}$$
    where $0! := 1$ and $k! := 1 \cdot 2 \cdots k$, for $k \geq 1$.

    Prove that one has $\binom{n}{k}$ possibilities to choose k elements out of n elements.

33. Binomial distributon: Assume $0 < p < 1$, $\Omega := \{0, 1, 2, ..., n\}$, $n \geq 1$, $\mathcal{F} := 2^{\Omega}$ and
    $$\mu_{n,p}(B) := \sum_{k \in B} \binom{n}{k} p^{n-k}(1-p)^k.$$

    (a) Is $(\Omega, \mathcal{F}, \mu_{n,p})$ a probability space?

    (b) Compute $\max_{k=0,...,n} \mu_{n,p}(\{k\})$ for $p = \frac{1}{2}$.

34. Geometric distribution: Let $0 < p < 1$, $\Omega := \{0, 1, 2, ...\}$, $\mathcal{F} := 2^{\Omega}$ and
    $$\mu_p(B) := \sum_{k \in B} p(1-p)^k.$$

(a) Is $(\Omega, \mathcal{F}, \mu_p)$ a probability space?

(b) Compute $\mu_p(\{0, 2, 4, 6, ...\})$.

35. POISSON DISTRIBUTION: Let $\lambda > 0$, $\Omega := \{0, 1, 2, ...\}$, $\mathcal{F} := 2^\Omega$ and

$$\pi_\lambda(B) := \sum_{k \in B} e^{-\lambda} \frac{\lambda^k}{k!}.$$

(a) Is $(\Omega, \mathcal{F}, \pi_\lambda)$ a probability space?

(b) Compute $\sum_{k \in \Omega} k \pi_\lambda(\{k\})$.

4.2 Random variables

1. Let $A_1, A_2, ... \subseteq \Omega$. Show that

   (a) $\liminf_{n\to\infty} \mathbb{1}_{A_n}(\omega) = \mathbb{1}_{\liminf_{n\to\infty} A_n}(\omega)$,

   (b) $\limsup_{n\to\infty} \mathbb{1}_{A_n}(\omega) = \mathbb{1}_{\limsup_{n\to\infty} A_n}(\omega)$,

   for all $\omega \in \Omega$.

2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A \subseteq \Omega$ a set. Show that $A \in \mathcal{F}$ if and only if $\mathbb{1}_A : \Omega \to \mathbb{R}$ is a random variable.

3. Show the assertions (1), (3) and (4) of Proposition 2.1.5.

4. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f : \Omega \to \mathbb{R}$.

   (a) Show that, if $\mathcal{F} = \{\emptyset, \Omega\}$, then f is measurable if and only if f is constant.

   (b) Show that, if $\mathbb{P}(A)$ is 0 or 1 for every $A \in \mathcal{F}$ and f is measurable, then
   $$\mathbb{P}(\{\omega : f(\omega) = c\}) = 1 \text{ for a constant } c.$$

5. Let $(\Omega, \mathcal{F})$ be a measurable space and $f_n, n = 1, 2, \ldots$ a sequence of random variables. Show that

$$\liminf_{n\to\infty} f_n \quad \text{and} \quad \limsup_{n\to\infty} f_n$$

   are random variables.

6. Complete the proof of Proposition 2.2.9 by showing that for a distribution function $F_g(x) = \mathbb{P}(\{\omega : g \leq x\})$ of a random variable g it holds

$$\lim_{x\to-\infty} F_g(x) = 0 \text{ and } \lim_{x\to\infty} F_g(x) = 1.$$

7. Let $f : \Omega \to \mathbb{R}$ be a map. Show that for $A_1, A_2, \cdots \subseteq \mathbb{R}$ it holds

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

8. 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. Show that then $g \circ f : \Omega_1 \to \Omega_3$ defined by

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

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

9. Prove Proposition 2.1.4.

10. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, (M, Σ) a measurable space and assume that $f : \Omega \to M$ is $(\mathcal{F}, \Sigma)$-measurable. Show that μ with

$$\mu(B) := \mathbb{P}\left(\{\omega : f(\omega) \in B\}\right) \quad \text{for } B \in \Sigma$$

is a probability measure on Σ.

11. Assume the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $f, g : \Omega \to \mathbb{R}$ be measurable step-functions. Show that f and g are independent if and only if for all $x, y \in \mathbb{R}$ it holds

$$\mathbb{P}\left(\{f = x, g = y\}\right) = \mathbb{P}\left(\{f = x\}\right)\mathbb{P}\left(\{g = y\}\right)$$

12. Assume the product space $([0,1] \times [0,1], \mathcal{B}([0,1]) \otimes \mathcal{B}([0,1]), \lambda \times \lambda)$ and the random variables $f(x,y) := \mathbb{1}_{[0,p)}(x)$ and $g(x,y) := \mathbb{1}_{[0,p)}(y)$, where $0 < p < 1$. Show that

    (a) f and g are independent,

    (b) the law of $f + g$, $\mathbb{P}_{f+g}(\{k\})$, for $k = 0, 1, 2$ is the binomial distribution $\mu_{2,p}$.

13. Let $([0,1], \mathcal{B}([0,1]))$ be a measurable space and define the functions

$$f(x) := \mathbb{1}_{[0,1/2)}(x) + 2\mathbb{1}_{[1/2,1])}(x),$$

and

$$g(x) := \mathbb{1}_{[0,1/2)}(x) - \mathbb{1}_{\{1/4\}}(x).$$

Compute (a) $\sigma(f)$ and (b) $\sigma(g)$.

14. Assume a measurable space $(\Omega, \mathcal{F})$ and random variables $f, g : \Omega \to \mathbb{R}$. Is the set $\{\omega : f(\omega) = g(\omega)\}$ measurable?

15. Let $\mathcal{G} := \sigma\{(a,b),\ 0 < a < b < 1\}$ be a σ -algebra on $[0,1]$. Which continuous functions $f : [0,1] \to \mathbb{R}$ are measurable with respect to $\mathcal{G}$?

16. Let $f_k : \Omega \to \mathbb{R}$, $k = 1, \ldots, n$ be independent random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and assume the functions $g_k : \mathbb{R} \to \mathbb{R}$, $k = 1, \ldots, n$ are $\mathcal{B}(\mathbb{R})$–measurable. Show that $g_1(f_1), g_2(f_2), \ldots, g_n(f_n)$ are independent.

17. Assume $n \in \mathbb{N}$ and define by $\mathcal{F} := \sigma\{\left(\frac{k}{n}, \frac{k+1}{n}\right]$ for $k = 0, 1, \ldots, n-1\}$ a σ–algebra on $[0,1]$.

    (a) Why is the function $f(x) = x, \quad x \in [0,1]$ not $\mathcal{F}$–measurable?

    (b) Give an example of an $\mathcal{F}$–measurable function.

18. Prove Proposition 2.3.4.

19.* Show that families of random variables are independent iff (if and only if) their generated σ-algebras are independent.

20. Show Proposition 2.3.5.

4.3 Integration

1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f, g : \Omega \to \mathbb{R}$ random variables with

$$f = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i} \quad \text{and} \quad g = \sum_{j=1}^{m} b_j \mathbb{1}_{B_j},$$

where $a_i, b_j \in \mathbb{R}$ and $A_i, B_j \in \mathcal{F}$. Assume that $\{A_1, \ldots, A_n\}$ and $\{B_1, \ldots B_m\}$ are independent. Show that

$$\mathbb{E}fg = \mathbb{E}f\mathbb{E}g.$$

2. Prove assertion (4) and (5) of Proposition 3.2.1
Hints: To prove (4), set $A_n := \left\{\omega : f(\omega) > \frac{1}{n}\right\}$ and show that

$$0 = \mathbb{E}f\mathbb{1}_{A_n} \geq \mathbb{E}\frac{1}{n}\mathbb{1}_{A_n} = \frac{1}{n}\mathbb{E}\mathbb{1}_{A_n} = \frac{1}{n}\mathbb{P}(A_n).$$

This implies $\mathbb{P}(A_n) = 0$ for all n. Using this one gets

$$\mathbb{P}\{\omega : f(\omega) > 0\} = 0.$$

To prove (5) use

$$\mathbb{E}f = \mathbb{E}(f\mathbb{1}_{\{f=g\}} + f\mathbb{1}_{\{f\neq g\}}) = \mathbb{E}f\mathbb{1}_{\{f=g\}} + \mathbb{E}f\mathbb{1}_{\{f\neq g\}} = \mathbb{E}f\mathbb{1}_{\{f=g\}}.$$

3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A_1, A_2, ... \in \mathcal{F}$. Show that

$$\begin{aligned} \mathbb{P}\left(\liminf_n A_n\right) &\leq \liminf_n \mathbb{P}(A_n) \\ &\leq \limsup_n \mathbb{P}(A_n) \leq \mathbb{P}\left(\limsup_n A_n\right). \end{aligned}$$

Hint: Use Fatou's lemma and Exercise 4.2.1.

4. Assume the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$ and the random variable

$$f = \sum_{k=0}^{\infty} k \mathbb{1}_{[a_{k-1}, a_k)}$$

with $a_{-1} := 0$ and

$$a_k := e^{-\lambda} \sum_{m=0}^{k} \frac{\lambda^m}{m!}, \quad \text{for } k = 0, 1, 2, \ldots$$

where $\lambda > 0$. Compute the law $\mathbb{P}_f(\{k\})$, $k = 0, 1, 2, \ldots$ of f. Which distribution has f?

5. Assume the probability space $((0,1], \mathcal{B}((0,1]), \lambda)$ and

$$f(x) := \begin{cases} 1 \text{ for irrational } x \in (0,1] \\ \frac{1}{x} \text{ for rational } x \in (0,1] \end{cases}$$

Compute $\mathbb{E}f$.

6. UNIFORM DISTRIBUTION:

Assume the probability space $([a,b], \mathcal{B}([a,b]), \frac{\lambda}{b-a})$, where $-\infty < a < b < \infty$.

(a) Show that

$$\mathbb{E}f = \int_{[a,b]} f(\omega) \frac{1}{b-a} d\lambda(\omega) = \frac{a+b}{2} \quad \text{where} \quad f(\omega) := \omega.$$

(b) Compute

$$\mathbb{E}g = \int_{[a,b]} g(\omega) \frac{1}{b-a} d\lambda(\omega) = \quad \text{where} \quad g(\omega) := \omega^2.$$

(c) Compute the variance $\mathbb{E}(f - \mathbb{E}f)^2$ of the random variable $f(\omega) := \omega$.

Hint: Use a) and b).

7. Poisson distribution: Let $\lambda > 0$, $\Omega := \{0, 1, 2, 3, ...\}$, $\mathcal{F} := 2^\Omega$ and

$$\pi_\lambda = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \delta_k.$$

   (a) Show that

$$\mathbb{E}f = \int_\Omega f(\omega) d\pi_\lambda(\omega) = \lambda \quad \text{where} \quad f(k) := k.$$

   (b) Show that

$$\mathbb{E}g = \int_\Omega g(\omega) d\mu_\lambda(\omega) = \lambda^2 \quad \text{where} \quad g(k) := k(k-1).$$

   (c) Compute the variance $\mathbb{E}(f - \mathbb{E}f)^2$ of the random variable $f(k) := k$.

   Hint: Use a) and b).

8. Binomial distribution: Let $0 < p < 1$, $\Omega := \{0, 1, ..., n\}$, $n \geq 2$, $\mathcal{F} := 2^\Omega$ and

$$\mu_{n,p} = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{(n-k)} \delta_k.$$

   (a) Show that

$$\mathbb{E}f = \int_\Omega f(\omega) d\mu_{n,p}(\omega) = np \quad \text{where} \quad f(k) := k.$$

   (b) Show that

$$\mathbb{E}g = \int_\Omega g(\omega) d\mu_{n,p}(\omega) = n(n-1)p^2 \quad \text{where} \quad g(k) := k(k-1).$$

   (c) Compute the variance $\mathbb{E}(f - \mathbb{E}f)^2$ of the random variable $f(k) := k$.

   Hint: Use a) and b).

9. Assume integrable, independent random variables $f_i : \Omega \to \mathbb{R}$ with $\mathbb{E}f_i^2 < \infty$, mean $\mathbb{E}f_i = m_i$ and variance $\sigma_i^2 = \mathbb{E}(f_i - m_i)^2$, for $i = 1, 2, \ldots, n$. Compute the mean and the variance of

   (a) $g = af_1 + b$, where $a, b \in \mathbb{R}$,

   (b) $g = \sum_{i=1}^n f_i$.

10. Let the random variables f and g be independent and Poisson distrubuted with parameters λ_1 and λ_2, respectively. Show that $f+g$ is Poisson distributed

$$\mathbb{P}_{f+g}(\{k\}) = \mathbb{P}(f+g=k) = \sum_{l=0}^{k} \mathbb{P}(f=l, g=k-l) = \dots.$$

Which parameter does this Poisson distribution have?

11. Assume that f ja g are independent random variables where $\mathbb{E}|f| < \infty$ and $\mathbb{E}|g| < \infty$. Show that $\mathbb{E}|fg| < \infty$ and

$$\mathbb{E}fg = \mathbb{E}f\mathbb{E}g.$$

Hint:

(a) Assume first $f \geq 0$ and $g \geq 0$ and show using the "stair-case functions" to approximate f and g that

$$\mathbb{E}fg = \mathbb{E}f\mathbb{E}g.$$

(b) Use (a) to show $\mathbb{E}|fg| < \infty$, and then Lebesgue's Theorem for

$$\mathbb{E}fg = \mathbb{E}f\mathbb{E}g.$$

in the general case.

12. Use Hölder's inequality to show Corollary 3.6.6.

13. Let $f_1, f_2, \dots$ be non-negative random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Show that

$$\mathbb{E}\sum_{k=1}^{\infty} f_k = \sum_{k=1}^{\infty} \mathbb{E}f_k \quad (\leq \infty).$$

14. Use Minkowski's inequality to show that for sequences of real numbers $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$ and $1 \leq p < \infty$ it holds

$$\left(\sum_{n=1}^{\infty} |a_n + b_n|^p\right)^{\frac{1}{p}} \leq \left(\sum_{n=1}^{\infty} |a_n|^p\right)^{\frac{1}{p}} + \left(\sum_{n=1}^{\infty} |b_n|^p\right)^{\frac{1}{p}}.$$

15. Prove assertion (2) of Proposition 3.2.3.

16. Assume a sequence of i.i.d. random variables $(f_k)_{k=1}^{\infty}$ with $\mathbb{E}f_1 = m$ and variance $\mathbb{E}(f_1 - m)^2 = \sigma^2$. Use the Central Limit Theorem to show that

$$\mathbb{P}\left(\left\{\omega : \frac{f_1 + f_2 + \cdots + f_n - nm}{\sigma\sqrt{n}} \leq x\right\}\right) \to \frac{1}{2\pi}\int_{-\infty}^{x} e^{-\frac{u^2}{2}} du$$

as $n \to \infty$.

17. Let $([0,1], \mathcal{B}([0,1]), \lambda)$ be a probability space. Define the functions $f_n : \Omega \to \mathbb{R}$ by

$$f_{2n}(x) := n^3 \mathbb{1}_{[0,1/2n]}(x) \quad \text{and} \quad f_{2n-1}(x) := n^3 \mathbb{1}_{[1-1/2n,1]}(x),$$

where $n = 1, 2, \ldots$.

(a) Does there exist a random variable $f : \Omega \to \mathbb{R}$ such that $f_n \to f$ almost surely?

(b) Does there exist a random variable $f : \Omega \to \mathbb{R}$ such that $f_n \xrightarrow{\mathbb{P}} f$?

(c) Does there exist a random variable $f : \Omega \to \mathbb{R}$ such that $f_n \xrightarrow{L_P} f$?

Index

λ-system, 31
$\liminf_n A_n$, 17
$\liminf_n a_n$, 17
$\limsup_n A_n$, 17
$\limsup_n a_n$, 17
π-system, 24
π-systems and uniqueness of measures, 24
$\pi - \lambda$-Theorem, 31
σ-algebra, 10
σ-finite, 14

algebra, 10
axiom of choice, 32

Bayes' formula, 19
binomial distribution, 25
Borel σ-algebra, 13
Borel σ-algebra on $\mathbb{R}^n$, 24

Carathéodory's extension theorem, 21
central limit theorem, 75
change of variables, 58
Chebyshev's inequality, 66
closed set, 12
conditional probability, 18
convergence almost surely, 71
convergence in L_p, 71
convergence in distribution, 72
convergence in probability, 71
convexity, 66
counting measure, 15

Dirac measure, 15
distribution-function, 40
dominated convergence, 55

equivalence relation, 31
event, 10
existence of sets, which are not Borel, 32
expectation of a random variable, 47
expected value, 47
exponential distribution on $\mathbb{R}$, 28
extended random variable, 60

Fubini's Theorem, 61, 62

Gaussian distribution on $\mathbb{R}$, 28
geometric distribution, 25

Hölder's inequality, 67

i.i.d. sequence, 74
independence of a family of random variables, 42
independence of a finite family of random variables, 42
independence of a sequence of events, 18

Jensen's inequality, 66

law of a random variable, 39
Lebesgue integrable, 47
Lebesgue measure, 26, 27
Lebesgue's Theorem, 55
lemma of Borel-Cantelli, 20
lemma of Fatou, 18
lemma of Fatou for random variables, 54

measurable map, 38
measurable space, 10
measurable step-function, 35
measure, 14
measure space, 14
Minkowski's inequality, 68

moments, absolute moments, 59
monotone class theorem, 60
monotone convergence, 53

open set, 12

Poisson distribution, 25
Poisson's Theorem, 29
probability measure, 14
probability space, 14
product of probability spaces, 23

random variable, 36
realization of independent
    random variables, 44

step-function, 35
strong law of large numbers, 74

uniform distribution, 26, 27

variance, 49
vector space, 59

weak law of large numbers, 73

Bibliography

[1] H. Bauer. *Probability theory*. Walter de Gruyter, 1996.

[2] H. Bauer. *Measure and integration theory*. Walter de Gruyter, 2001.

[3] P. Billingsley. *Probability and Measure*. Wiley, 1995.

[4] A.N. Shiryaev. *Probability*. Springer, 1996.

[5] D. Williams. Probability with martingales. Cambridge University Press, 1991.