Stochastic Methods for Engineering

Paolo Guiotto

Contents

1

Probability Space

Probability theory arises from the problem of *making predictions under uncertainty*. Historically, probability was developed to analyze *games of chance*, where one has basically to *count* the favorable outcomes over the possible outcomes. Probability largely remained in that realm until the nineteenth century. In 1827, R. Brown first described what is now called *Brownian motion* (hereafter, BM): the irregular movement of small particles suspended in a fluid, caused by incessant collisions with the fluid's molecules. Typical trajectories are highly irregular, with apparently random changes of direction, which makes them hard to model. In the early twentieth century, L. Bachelier proposed a model for stock prices based on Brownian-like paths, drawing the attention of mathematicians. A rigorous mathematical description of BM was later provided by N. Wiener, influenced by the then-recent measure theory developed by H. Lebesgue. Wiener's visionary idea was to build a probabilistic structure on path space, that is on the space of continuous functions $\omega = \omega(t)$. This shed new light on probability and paved the way for N. Wiener's construction Kolmogorov's 1933 axiomatization of modern probability theory.

Since then, Probability has undergone tremendous development in many directions, becoming a central branch of mathematics. It is no coincidence that this growth has occurred alongside advances in science and technology. Indeed, probability has proved to be an effective tool for describing complex phenomena, where deterministic predictions give way to probabilistic ones. This is perhaps why probability is viewed as a practical tool, and probabilistic modeling as a genuine skill rather than a merely illustrative device.

1.1. Basic definitions

From the formal point of view, a *probability space* is a *measure space* with total measure $= 1$.

Definition 1.1.1

A **probability space** is a measure space $(\Omega, \mathscr{F}, \mathbb{P})$ such that $\mathbb{P}(\Omega) = 1$.

- Set Ω is called **sample space**,
- measurable sets are called **events**, their measure $\mathbb{P}(E)$ is called **probability of** E.
- An event is said **certain** if $\mathbb{P}(E) = 1$, **impossible** if $\mathbb{P}(E) = 0$.

If $\mathscr{F}$ contains all the subsets of an impossible event, we say that $(\Omega, \mathscr{F}, \mathbb{P})$ is a **complete probability space**.

Example 1.1.2

Let $\Omega = \mathbb{R}^d$, $\mathscr{F} := \mathscr{M}_d$ (Lebesgue class) and $f \in L^1(\mathbb{R}^d)$ be such that $f \geqslant 0$ a.e. and $\int_{\mathbb{R}^d} f = 1$. We define

$$\mathbb{P}(E) = \int_E f(x)\, dx,\ E \in \mathscr{M}_d =: \mu_f(E)$$

Then $(\mathbb{R}^d, \mathscr{M}_d, \mu_f)$ is a complete probability space.

Example 1.1.3

Let Ω be a generic set, $\mathscr{F} := \mathscr{P}(\Omega)$. Let $\omega_0 \in \Omega$. We define

$$\mathbb{P}(E) = \begin{cases} 0, & \omega_0 \notin E, \\ 1, & \omega_0 \in E \end{cases} =: \delta_{\omega_0}(E)$$

Then $(\Omega, \mathscr{P}(\Omega), \delta_{\omega_0})$ is a probability space.

1.1.1. Discrete Probability Space. The classical Probability is described by the following setup:

Proposition 1.1.4

Let Ω be a *finite* or *countable* set, say $\Omega = \{\omega_n \ :\ n \in \mathbb{N}\}$. Let $(p_n) \subset [0,1]$ be such that

$$\sum_n p_n = 1. \tag{1.1.1}$$

We call such (p_n) a **probability mass distribution**. On $(\Omega, \mathscr{P}(\Omega))$ we define

$$\mathbb{P}(E) := \sum_{\omega_n \in E} p_n \equiv \sum_n p_n 1_E(\omega_n).$$

Space $(\Omega, \mathscr{P}(\Omega), \mathbb{P})$ is a probability space called **discrete probability space**.

Proof. Clearly, by (1.1.1), $\mathbb{P}(E)$ is well defined for every $E \in \mathscr{P}(\Omega)$. We have also that $\mathbb{P}(E) \geqslant 0$ for every event E. To check that $\mathbb{P}$ is a probability measure we need to verify that

i) $\mathbb{P}(\varnothing) = 0$ (trivial) and $\mathbb{P}(\Omega) = \sum_n p_n 1_\Omega(\omega_n) = \sum_n p_n = 1$.

ii) $\mathbb{P}$ is countably additive. Assume $E = \bigsqcup_k E_k$. Notice that, since every ω belongs at most a only one of the E_k,

$$1_E(\omega) = \sum_k 1_{E_k}(\omega)$$

thus

$$\mathbb{P}(E) = \sum_n p_n \sum_k 1_{E_k}(\omega_n) = \sum_k \sum_n p_n 1_{E_k}(\omega_n) = \sum_k \mathbb{P}(E_k). \qquad \square$$

Discrete Probability Space model solves basically any everyday probabilistic framework. It is the classical ancient probabilistic setup. The sample space Ω represents of set of all possible outcomes.

Example 1.1.5: Coin Tossing

The tossing experiment can be described by two possible outcomes, H for head, T for tail. The sample space is $\Omega := \{H, T\}$. For a fair coin, $p_H = p_T = \frac{1}{2}$.

If the coin is unfair, we may have $p_H \neq \frac{1}{2}$, in that case $p_T = 1 - p_H$. This is called Bernoulli model.

Example 1.1.6: Rolling a die

In this case $\Omega := \{1, \ldots, 6\}$ with $p_n = \frac{1}{6}$ for $n = 1, \ldots, 6$.

Example 1.1.7: Rolling two dice

Suppose we want to describe the set of possible outcomes when rolling two dice. We can represent a single trial by a pair (i, j) where $i, j \in \{1, \ldots, 6\}$ are, resp., the outcomes of the first and second die. In this case $\Omega = \{1, \ldots, 6\}^2$ and, if the outcome of each die is independent of the outcome of the other's, $p_{i,j} = \frac{1}{6^2} = \frac{1}{36}$ for $(i, j) \in \Omega$.

Example 1.1.8: Binomial model

In one single day, a stock price can go Up with probability p and Down with probability $1 - p$. Precisely, if w is the value at beginning of the day, at the end of the day it can be either $(1 + r)w$ (with rate $r > 0$) if it goes Up, or $\frac{w}{1+r}$ if it goes Down. Assuming that the $n-$th day behavior is independent of the past, describe the space of outcomes after N days together with their probabilities.

Proof. We notice that, no matter which is the order of Ups and Downs, if the price goes Up n times and Down $N - n$ times, the day N value is

$$(1 + r)^n \left(\frac{1}{1 + r}\right)^{N-n} w = (1 + r)^{2n-N} w.$$

Since $r > 0$, when $n = 0, \ldots, N$, $(1 + r)^{2n-N}$ takes $N + 1$ different values. Therefore, there are $N + 1$ different final states. We can identify these states with the number of Ups, so $\Omega = \{0, \ldots, N\}$. The state n is obtained exactly when tossing N times an unfair coin with Prob. Up=p. For a single trial this probability is $p^n(1 - p)^{N-n}$. The number of different trials is the number of $N-$ples of Up and Down with n Ups, that is the binomial coefficient $\binom{N}{n}$. Thus,

$$p_n := \text{Prob.}(n \text{ Ups}) = \binom{N}{n} p^n (1 - p)^{N-n}, \; n \in \{0, \ldots, N\} = \Omega.$$

Let's check that this is indeed a probability distribution: we have $p_n \geqslant 0$ for every n and

$$\sum_{n=0}^{N} p_n = \sum_{n=0}^{N} \binom{N}{n} p^n (1 - p)^{N-n} = (p + 1 - p)^N = 1.$$

1.1.2. Basic properties of Probability measures. Since a probability measure is a particular type of measure, it fulfills all the basic properties of any measure. We notice that, by additivity, for any event $E \in \mathscr{F}$, we have

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(E \sqcup E^c) = \mathbb{P}(E) + \mathbb{P}(E^c),$$

from which

$$\boxed{\mathbb{P}(E^c) = 1 - \mathbb{P}(E).}$$

As every measure, a probability measure is *continuous from below*. In addition, since the total measure $\mathbb{P}(\Omega) = 1 < +\infty$, a probability measures is always also continuous from above. An interesting fact is the following: *continuity from above at $\varnothing$ is in fact equivalent to countable additivity*.

Proposition 1.1.9

Let $\mathbb{P}$ be a finitely additive probability on $(\Omega, \mathscr{F})$. Then, the following properties are equivalent:

i) $\mathbb{P}$ is countably additive on $\mathscr{F}$.
ii) $\mathbb{P}$ is continuous from above at $\varnothing$, that is

$$(E_n) \subset \mathscr{F}, \,:\, E_n \downarrow \varnothing, \implies \lim_n \mathbb{P}(E_n) = 0.$$

PROOF. i) $\Longrightarrow$ ii). It follows from the continuity from above of any countably additive finite measure. ii) $\Longrightarrow$ i) Let $(E_n) \subset \mathscr{F}$ be such that $E_n \cap E_m = \varnothing$ for $n \neq m$, and let $E := \bigsqcup_n E_n \in \mathscr{F}$. Then, setting $F_n := E \setminus \bigsqcup_{k=0}^n E_k \downarrow \varnothing$, so by the assumption

$$\lim_n \mathbb{P}\left(E \setminus \bigsqcup_{k=0}^n E_k\right) = 0.$$

Now, since

$$E = \left(E \setminus \bigsqcup_{k=0}^n E_k\right) \sqcup \left(\bigsqcup_{k=0}^n E_k\right)$$

by finite additivity of $\mathbb{P}$ we have

$$\mathbb{P}(E) = \mathbb{P}\left(E \setminus \bigsqcup_{k=0}^n E_k\right) + \mathbb{P}\left(\bigsqcup_{k=0}^n E_k\right) = \mathbb{P}\left(E \setminus \bigsqcup_{k=0}^n E_k\right) + \sum_{k=0}^n \mathbb{P}(E_k) \longrightarrow 0 + \sum_{k=0}^\infty \mathbb{P}(E_k). \quad \square$$

1.2. Space of Sequences

A rudimentary model for the BM is the *random walk model*. We assume that a particle starts at the origin, then, at each second it moves left or right with equal probability. A random walk is described by an infinite sequence of L and R, like $LLRLRRLRLRLRRRL\ldots$. The set of all possible sequences of this type,

$$\Omega : \{\omega = (\omega_n) \,:\, \omega_n \in \{R, L\}, \ \forall n \in \mathbb{N}\} \equiv \{R, L\}^{\mathbb{N}}.$$

is the *path space*. Imagine now we aim to introduce a probabilistic structure on this Ω. The problem is that Ω is not countable: $\{0,1\}^{\mathbb{N}}$ are all possible binary sequences, it is in correspondence 1-1 with $[0,1]$.

1.2.1. Sample space. Without no particular effort, we can extend this framework. Let $S = \{s_1, \ldots, s_N\}$ be any **finite** set. This set will be called *state space*. We define, as sample space, the set Ω made of all possible sequences of states, that is

$$\Omega := \{\omega = (\omega_n) \, : \, \omega_n \in S, \ \forall n \in \mathbb{N}\} \equiv S^{\mathbb{N}}.$$

In some models $\mathbb{N}$ can be replaced by $\mathbb{Z}$ without any relevant change. Our goal is to define on Ω a structure of probability space in such a way that natural sets are events to which we can assign a probability. The natural sets we consider are called **cylinders**. These are sets for which only a finite number of components of ω are constrained, while all the remaining are free to take any value of S. Formally, for $k \in \mathbb{N}$, and $E_0, \ldots, E_k \subset S$ we set

$$C(k; E_0 \times \cdots \times E_k) = \{\omega \in \Omega \, : \, \omega_0 \in E_0, \ \ldots, \ \omega_k \in E_k\},$$

or, more in general, for $E \subset S^{k+1}$,

$$C(k; E) := \{\omega \in \Omega \, : \, (\omega_0, \ldots, \omega_k) \in E\}.$$

Sets $C(k; E)$ are called **cylinders** (this because they remind of geometrical cylinders, where some of the coordinates are constrained to some domain, e.g. a disk, and others are free). We also set $\mathscr{C}$ the family of all the cylinders, for all possible choices of k, and $E \subset S^k$. Notice that

i) $\Omega, \varnothing \in \mathscr{C}$: indeed, for example,

$$\varnothing = C(0; \varnothing), \quad \Omega = C(0, S).$$

ii) if $C \in \mathscr{C}$ then also $C^c \in \mathscr{C}$. Indeed,

$$\begin{aligned} C(k; E)^c &= \{\omega \in \Omega \, : \, (\omega_0, \ldots, \omega_k) \notin E\} = \{\omega \in \Omega \, : \, (\omega_0, \ldots, \omega_k) \in E^c\} \\ &= C(k; E^c). \end{aligned}$$

iii) $\mathscr{C}$ is closed wrt finite unions. Let's check this for $C_1 \cup C_2$, $C_j = C(k_j; E_j)$, $j = 1, 2$, cylinders. Let $k := \max\{k_1, k_2\}$, then

$$C(k_j; E_j) = C(k; E_j \times S^{k-k_j})$$

so we can always assume that $C_j = C(k; E_j)$ with the same k. Then

$$C_1 \cup C_2 = \{\omega \in \Omega \, : \, (\omega_0, \ldots, \omega_k) \in E_1 \cup E_2\} = C(k; E_1 \cup E_2).$$

Unfortunately, $\mathscr{C}$ **is not a** $\sigma-$**algebra**. For example, if $\varnothing \subsetneq E \subsetneq S$, then

$$\bigcup_n C(n; E^n) = \left\{\omega \in \Omega \, : \, \omega_j \in E, \ \forall j\right\} \notin \mathscr{C}.$$

Definition 1.2.1

We say that $\mathscr{A}$ is an **algebra of sets** if

- $\varnothing, \Omega \in \mathscr{A}$;
- if $A \in \mathscr{A}$ then also $A^c \in \mathscr{A}$;
- if $A_1, \ldots, A_n \in \mathscr{A}$ then $\bigcup_{k=1}^n A_k \in \mathscr{A}$.

Thus, $\mathscr{C}$ is an algebra of sets. We know that a natural $\sigma-$algebra is always available: it is $\sigma(\mathscr{C})$, the $\sigma-$algebra generated by $\mathscr{C}$, which is also the smallest $\sigma-$algebra containing $\mathscr{C}$.

1.2.2. Probability structure. Let's now introduce a probabilistic structure on $\Omega = S^{\mathbb{N}}$. Let $p : S \longrightarrow [0,1]$ be a probability mass distribution, so

$$\sum_{s \in S} p_s = 1.$$

We now define

$$\mathbb{P} : \mathscr{C} \longrightarrow [0,1].$$

The idea is simple: if for example we consider the cylinder $C(0;\{s\}) = \{\omega \in \Omega : \omega_0 = s\}$. It would be natural to say

$$\mathbb{P}(C(0;\{s\})) = \text{Prob}(\omega_0 = s) = p_s.$$

This yields the definition

$$\mathbb{P}\left(C(k;E)\right) := \sum_{(s_0,\dots,s_k)\in E} p_{s_0} p_{s_2} \dots p_{s_k}. \tag{1.2.1}$$

This quantity is well posed. We need to check this because we can represent any cylinder in infinitely many ways. For instance

$$C(k;E) = C(k+1; E \times S).$$

According to the (1.2.1) we have

$$\begin{aligned}
\mathbb{P}(C(k+1;E\times S) &= \textstyle\sum_{(s_0,\dots,s_k,s_{k+1})\in E\times S} p_{s_1} p_{s_2} \dots p_{s_k} p_{s_{k+1}} \\
&= \textstyle\sum_{(s_0,\dots,s_k)\in E} p_{s_1} p_{s_2} \dots p_{s_k} \underbrace{\displaystyle\sum_{s_{k+1}\in S} p_{s_{k+1}}}_{=1} \\
&= \textstyle\sum_{(s_0,\dots,s_k)\in E} p_{s_1} p_{s_2} \dots p_{s_k} = \mathbb{P}(C(k;E)).
\end{aligned}$$

Similarly, $C(k;E) = C(k+m; E\times S^m)$ and $\mathbb{P}(C(k;E)) = \mathbb{P}(C(k+m;E\times S^m))$. Furthermore, $\mathbb{P}(\emptyset) = 0$ and

$$\mathbb{P}(\Omega) = \mathbb{P}(C(0;S)) = \sum_{s\in S} p_s = 1.$$

$\mathbb{P}$ is also additive. Indeed, if C_1, C_2 are disjoint cylinders, since we can always represent them as $C_1 = C(k;E_1)$ and $C_2 = C(k;E_2)$ then, necessarily, E_1 and E_2 must be disjoint. In this case,

$$\begin{aligned}
\mathbb{P}(C_1 \cup C_2) &= \mathbb{P}(C(k;E_1\cup E_2)) = \textstyle\sum_{(s_1,\dots,s_k)\in E_1\cup E_2} p_{s_1}\cdots p_{s_k} \\
&= \textstyle\sum_{(s_1,\dots,s_k)\in E_1} p_{s_1}\cdots p_{s_k} + \textstyle\sum_{(s_1,\dots,s_k)\in E_2} p_{s_1}\cdots p_{s_k} \\
&= \mathbb{P}(C(k;E_1)) + \mathbb{P}(C(k;E_2)) = \mathbb{P}(C_1) + \mathbb{P}(C_2).
\end{aligned}$$

In general, as we said, $\mathscr{C}$ is not a $\sigma-$algebra. So, a countable union of cylinders might not be a cylinder. If however this happens and the union is disjoint, it turns out that $\mathbb{P}$ is countably additive. To show this, we first state an equivalent condition for countable additivity that holds for probability measures:
So, we need to prove the

Lemma 1.2.2

Let $(C_n) \subset \mathscr{C}$ be a sequence of cylinders such that $C_n \downarrow \varnothing$. Then

$$\lim_n \mathbb{P}(C_n) = 0.$$

Proof. Suppose, by contradiction, that $\mathbb{P}(C_n) \not\longrightarrow 0$. Since $C_n \downarrow$ and $\mathbb{P}$ is finitely additive, easily $\mathbb{P}(C_n) \downarrow$. So, the contradiction means that

$$\exists \varepsilon > 0, : \mathbb{P}(C_n) \geqslant \varepsilon, \ \forall n \in \mathbb{N}.$$

The goal is to prove that

$$\exists \omega \in \bigcap_n C_n.$$

Since $\mathbb{P}(C_n) \geqslant \varepsilon > 0$, $C_n \neq \varnothing$, so

$$\begin{array}{rllll} \exists\omega_1 = (\omega_{1,1} & \omega_{1,2} & \omega_{1,3} & \ldots) \in C_1, \\ \exists\omega_2 = (\omega_{2,1} & \omega_{2,2} & \omega_{2,3} & \ldots) \in C_2 \\ \exists\omega_3 = (\omega_{3,1} & \omega_{3,2} & \omega_{3,3} & \ldots) \in C_3 \\ \vdots & \vdots & \vdots & \vdots \\ \exists\omega_n = (\omega_{n,1} & \omega_{n,2} & \omega_{n,3} & \ldots) \in C_n \\ \vdots & \vdots & \vdots & \vdots \end{array}$$

Let's focus on the sequence of first components $(\omega_{n,1}) \subset S$. Since S is finite, there is at least one of the element of the sequence that repeats infintely many times. In other words,

$$\exists (n_j^1) \subset \mathbb{N} \ : \ \omega_{n_j^1,1} \equiv \widetilde{\omega}_1.$$

Let's now consider the subsequence $\omega_{n_j^1}$ and, in particular, the second coordinates $(\omega_{n_j^1,2}) \subset S$. By the same argument, at least one of the elements of the sequence repeats infinitely many times, that is,

$$\exists (n_j^2) \subset (n_j^1) \ : \ \omega_{n_j^2,2} \equiv \widetilde{\omega}_2.$$

Notice that $(\omega_{n_j^2,1}) \subset (\omega_{n_j^1,1})$ so $\omega_{n_j^2,1} \equiv \widetilde{\omega}_1$. Iterating this procedure we have that

$$\exists (n_j^k) \subset (n_j^{k-1}) \ : \ \omega_{n_j^k,i} \equiv \widetilde{\omega}_k, \ i = 1, \ldots, k.$$

Let finally

$$\widetilde{\omega} := (\widetilde{\omega}_1, \widetilde{\omega}_2, \ldots).$$

We claim that $\widetilde{\omega} \in \bigcap_n C_n$. Indeed, $C_n = C(k_n; E_n)$ for some $k_n \in \mathbb{N}$, $E_n \subset S^{k_n}$. So

$$\widetilde{\omega} \in C_n \iff (\widetilde{\omega}_1, \ldots, \widetilde{\omega}_{k_n}) \in E_n.$$

Notice that $(\widetilde{\omega}_1, \ldots, \widetilde{\omega}_{k_n}) = \left(\omega_{n_j^{k_n},1}, \ldots, \omega_{n_j^{k_n},k_n}\right)$ which are the first k_n components of $\omega_{n_j^{k_n}} \in C_{n_j^{k_n}}$. For j large enough, $n_j^{k_n} \geqslant n$, so $C_{n_j^{k_n}} \subset C_n$, so

$$(\widetilde{\omega}_1, \ldots, \widetilde{\omega}_{k_n}) = (\omega_{n_j^{k_n},1}, \omega_{n_j^{k_n},2}, \ldots, \omega_{n_j^{k_n},k_n}) \in E_n,$$

from which the conclusion follows. □

Definition 1.2.3

Let $\mathscr{A}$ be an algebra of sets of Ω. A set-function $\mathbb{P} : \mathscr{A} \longrightarrow [0,1]$ is a **pre-probability** if

i) $\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1$.

ii) if $(A_n) \subset \mathscr{A}$ is such that $\bigsqcup_n A_n \in \mathscr{A}$ then

$$\mathbb{P}\left(\bigsqcup_n A_n\right) = \sum_n \mathbb{P}(A_n).$$

It can be proved that any pre-probability can be extended in a unique way to the $\sigma-$algebra generated by $\mathscr{A}$:

Theorem 1.2.4: Carathéodory's extension theorem

Let $\mathscr{A} \subset \mathscr{P}(\Omega)$ be an algebra of subsets and $\mathbb{P} : \mathscr{A} \longrightarrow [0,1]$ be a pre-probability. Then $\mathbb{P}$ admits a unique extension to $\sigma(\mathscr{A})$.

1.3. Exercises

Exercise 1.3.1 ($*$ geometric distribution)**.** *Let $\Omega = \mathbb{N}$, $p_n := (1-p)p^n$ with $0 < p < 1$. Check that (p_n) is a probability mass distribution.*

Exercise 1.3.2 ($*$ Poisson distribution)**.** *Let $\Omega = \mathbb{N}$, $p_n := e^{-\lambda}\frac{\lambda^n}{n!}$, for $\lambda > 0$ fixed. Check that (p_n) is a probability mass distribution.*

Exercise 1.3.3 ($*$)**.** *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, E, F two events such that $\mathbb{P}(E) = \frac{3}{4}$ and $\mathbb{P}(F) = \frac{1}{3}$. Show that $\mathbb{P}(E \cap F) \geqslant \frac{1}{12}$.*

Exercise 1.3.4 ($*$)**.** *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, $(E_n) \subset \mathscr{F}$ be sure events, that is $\mathbb{P}(E_n) = 1$ for every n. Then, also $\bigcap_n E_n$ is a sure event.*

Exercise 1.3.5 ($**$)**.** *Let $\Omega = [0,1]$, $\mathscr{F} := \{E \subset [0,1] \ : \ E \text{ countable or } E^c \text{countable}\}$, and*

$$\mathbb{P} : \mathscr{F} \longrightarrow [0,1], \ \mathbb{P}(E) := \begin{cases} 0, & E \text{ countable}, \\ \\ 1, & E^c, \text{ countable}. \end{cases}$$

Determine whether $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space or not.

Exercise 1.3.6 ($**$)**.** *Let Ω be a sample space, $\mathscr{F}$ a $\sigma-$algebra of events, $\mathbb{P}, \mathbb{Q}$ two probability measures on $\mathscr{F}$.*

i) *Check that, if $\mathbb{P}(E) = \mathbb{Q}(E)$ for every $E \in \mathscr{F}$ with $\mathbb{P}(E) \leqslant \frac{1}{2}$, then $\mathbb{P} = \mathbb{Q}$ (that is, $\mathbb{P}(E) = \mathbb{Q}(E)$ for every $E \in \mathscr{F}$).*

ii) *Is i) still true if $\mathbb{P}(E) = \mathbb{Q}(E)$ for every $E \in \mathscr{F}$ with $\mathbb{P}(E) < \frac{1}{2}$?*

Exercise 1.3.7 ($**$)**.** *Let $\Omega = \mathbb{N}$, $\mathscr{F} = \mathscr{P}(\Omega)$ and define*

$$\mathbb{P}_n(E) := \frac{\sharp E \cap \{0, \ldots, n-1\}}{n}.$$

i) *Prove that $\mathbb{P}_n$ is finitely additive. Is also countably additive?*

ii) *Define the class $\widetilde{\mathscr{F}} \subset \mathscr{F}$ as follows:*

$$\widetilde{\mathscr{F}} := \{E \in \mathscr{F} \ : \ \exists \lim_{n\to+\infty} \mathbb{P}_n(E) =: \mathbb{P}(E)\}.$$

Check that $\widetilde{\mathscr{F}}$ is closed for finite unions and $\mathbb{P}$ is finitely additive.

iii) *Is $\mathscr{F}$ a $\sigma-$algebra? Is $\mathbb{P}$ countably additive?*

Exercise 1.3.8 ($**$). *Letting $N \longrightarrow +\infty$ into the binomial model we may obtain the Poisson model. Precisely, consider a binomial model with parameter $p_N = \frac{\lambda}{N}$ and set*

$$p_n^N := \binom{N}{n} p_N^n (1-p_N)^{N-n}, \ \ n = 0, \dots, N.$$

Show that

$$\lim_{N\to+\infty} p_n^N = \frac{\lambda^n}{n!} e^{-\lambda}.$$

(hint: use Stirling's formula $k! \sim_{+\infty} \sqrt{2\pi k}\frac{k^k}{e^k}$.)

Exercise 1.3.9 ($**$). *Let $\Omega = S^{\mathbb{N}}$. For each of the following sets K determine if it is a cylinder and if it belongs to $\sigma(\mathscr{C})$.*

- *A singleton $K := \{\widetilde{\omega}\}$, where $\widetilde{\omega} \in \Omega$.*
- *Let $U \subset S$, and $K := \{\omega \in \Omega \ : \ \omega_n \in U, \ \forall n \geqslant N\}$;*
- *Let $s \in S$, $K := \{\omega \in \Omega \ : \ \omega_{2k} \equiv s, \ \forall k\}$*
- *Let $s \in S$, and $K := \{\omega \in \Omega \ : \ \omega_k = s,$ for infinitely many $k\}$.*

Let $\mathbb{P}$ be the probability that originates from $(p_s)_{s\in S}$ with $0 < p_s < 1$ for every s and $\sum_{s\in S} p_s = 1$ What is the $\mathbb{P}$ of previous examples?

2

Random Variables

2.1. Basic definitions

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. A **random variable** (random variable) is just an $\mathscr{F}-$measurable function $X : \Omega \longrightarrow \mathbb{R}$. Usually, r.vs. are denoted by uppercase letters as $X, Y, \ldots$. Therefore, all properties of measurable functions apply. In particular, sum, difference, product of r.vs. is a random variable, as well as the point-wise limit of a sequence of random variable is a random variable. We say that a property $p = p(\omega)$ holds $\mathbb{P}-$**almost surely** (shortening: $\mathbb{P}-$a.s.) if it holds *almost everywhere* in the ordinary language of measures. So, for example, if X is a random variable,

$$X \geqslant 0,\ \mathbb{P} - a.s., \quad \Longleftrightarrow \quad \mathbb{P}(X < 0) = 0.$$

If $X \geqslant 0$ $\mathbb{P}-$a.s. it is always defined

$$\int_\Omega X \, d\mathbb{P},$$

possibly equal to $+\infty$. If $X \in L^1(\Omega)$, then it is well defined the **expected value of** X (or, also, *expectation of* X)

$$\mathbb{E}[X] := \int_\Omega X \, d\mathbb{P} \in \mathbb{R}.$$

An important value is the **variance** of X. This is defined for $X \in L^2(\Omega)$ as

$$\mathbb{V}[X] := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \int_\Omega X^2 \, d\mathbb{P} - \left(\int_\Omega X \, d\mathbb{P}\right)^2.$$

Notice that, by the CS inequality,

$$|\mathbb{E}[X]| = |\mathbb{E}[1 \cdot X]| \leqslant \mathbb{E}[1]^{1/2}\mathbb{E}[X^2]^{1/2} = \mathbb{E}[X^2]^{1/2},$$

so if $X \in L^2$ then $X \in L^1$. In general, the quantity

$$\mathbb{E}[X^k] = \int_\Omega X^k \, d\mathbb{P},\ (k \in \mathbb{N}),$$

is called $k-$**th moment of** X (to be defined we need $X \in L^k(\Omega)$).

Given X, Y random variables we define the **covariance** of X and Y as the quantity

$$\mathrm{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

By the CS inequality, $\mathrm{Cov}(X, Y)$ is well defined for $X, Y \in \mathbb{L}^2$ and

$$|\mathrm{Cov}(X, Y)| \leqslant \mathbb{E}[(X - \mathbb{E}[X])^2]^{1/2}\mathbb{E}[(Y - \mathbb{E}[Y])^2]^{1/2} = \mathbb{V}[X]^{1/2}\mathbb{V}[Y]^{1/2}.$$

The quantity $\mathbb{V}[X]^{1/2}$ is also called **standard deviation**. The **linear correlation** (or **Pearson's correlation**) is

$$\rho(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\mathbb{V}[X]^{1/2}\mathbb{V}[Y]^{1/2}}.$$

The linear correlation is well defined provided $X, Y \in L^2$ are not a.e. constants and, because of CS inequality

$$|\rho(X,Y)| \leqslant 1.$$

2.2. Law of a random variable

A random variable induces a natural probability on $\mathbb{R}$ equipped with Borel $\sigma-$algebra $\mathscr{B}_{\mathbb{R}}$. We recall that this is the $\sigma-$algebra generated by open sets of $\mathbb{R}$. The idea is to define

$$\mu_X(E) := \mathbb{P}(X \in E), \; E \in \mathscr{B}_{\mathbb{R}}.$$

To be sure that this definition makes sense, we need first to verify that $\{X \in E\} \in \mathscr{F}$ for every E is a Borel set. This is the content of the following Proposition.

Proposition 2.2.1

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, $X : \Omega \longrightarrow \mathbb{R}$ a function. Then, the following properties are equivalent:

i) X is a random variable.
ii) $\{X \in E\} \in \mathscr{F}, \; \forall E \in \mathscr{B}_{\mathbb{R}}$.

Proof. ii) $\Longrightarrow$ i) is trivial: since $\mathscr{B}_{\mathbb{R}}$ is generated by open sets, it contains, in particular, all intervals $I \subset \mathbb{R}$, so by ii) we have $\{X \in I\} \in \mathscr{F}$ for every I interval, and this means that $X \in L(\Omega)$, that is X is a random variable.
Let's prove that i) $\Longrightarrow$ ii). To this aim, define the family of sets

$$\mathscr{G} := \{E \in \mathscr{B}_{\mathbb{R}} \; : \; \{X \in E\} \in \mathscr{F}\} \subset \mathscr{B}_{\mathbb{R}}.$$

The goal is to prove that $\mathscr{G} = \mathscr{B}_{\mathbb{R}}$. To this aim we will verify that

i) $\mathscr{G}$ contains the open sets (of $\mathbb{R}$);
ii) $\mathscr{G}$ is a $\sigma-$algebra.

From this, it follows that $\mathscr{G} \supset \sigma(\{\text{open sets}\}) = \mathscr{B}_{\mathbb{R}}$, and since by definition $\mathscr{G} \subset \mathscr{B}_{\mathbb{R}}$ the conclusion follows.
i) Let E be an open interval. By definition of measurable function $\{X \in E\} \in \mathscr{F}$. Now, if E is a generic open set, we know that for every $x \in E$ there exists I_x open neighbourhood of x (something like $]x-\varepsilon, x+\varepsilon[$) such that $I_x \subset E$. In this way

$$E = \bigcup_x I_x.$$

We need to refine this union to a countable union. By density of $\mathbb{Q}$ in $\mathbb{R}$, for every $x \in E$ we can find $q_x \in \mathbb{Q}$ and an open neighbourhood J_{q_x} such that $x \in J_{q_x} \subset I_x$. Thus

$$E = \bigcup_x J_{q_x},$$

and since the $(q_x)_{x\in E}$ are at most a countable number, say $(q_x)_{x\in E} = (r_n)_{n\in\mathbb{N}}$, we have

$$E = \bigcup_n J_{r_n}.$$

Therefore, since $\{X \in J_{r_n}\} \in \mathscr{F}$ by the first part of the argument and $\mathscr{F}$ is a $\sigma-$algebra, we have

$$\{X \in E\} = \bigcup_n \{X \in J_{r_n}\} \in \mathscr{F}$$

ii) We check that $\mathscr{G}$ is a $\sigma-$algebra of subsets of $\mathbb{R}$. First $\{X \in \mathbb{R}\} = \Omega \in \mathscr{F}$, thus $\mathbb{R} \in \mathscr{G}$. Similarly, $\{X \in \emptyset\} = \emptyset \in \mathscr{F}$, so $\emptyset \in \mathscr{G}$. If $E \in \mathscr{G}$, then, since

$$\{X \in E^c\} = \{X \in E\}^c \in \mathscr{F},$$

we have $E^c \in \mathscr{G}$. Finally, if $(E_n) \subset \mathscr{G}$ we have

$$\left\{X \in \bigcup_n E_n\right\} = \bigcup_n \{X \in E_n\} \in \mathscr{F},$$

thus $\bigcup_n E_n \in \mathscr{G}$.

Definition 2.2.2

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, X a random variable. We call **law of** X the probability measure μ_X on $(\mathbb{R}, \mathscr{B}_\mathbb{R})$ defined by

$$\mu_X(E) := \mathbb{P}(X \in E),\ E \in \mathscr{B}_\mathbb{R}.$$

Example 2.2.3: constant random variable

The simplest possible example of random variable is a *constant* one, namely $X = x_0$ a.s.. In this case the law of X is

$$\mu_X(E) = \mathbb{P}(X \in E) = \begin{cases} 1, & x_0 \in E, \\ 0, & x_0 \notin E, \end{cases} = \delta_{x_0}(E).$$

the delta Dirac δ_{x_0}.

Example 2.2.4: Bernoulli random variable

The simplest non constant random variable is that one who takes two values, say $X = 1$ with probability p and $X = 0$ with prob. $1 - p$. We write $X \sim \mathrm{Ber}(p)$. The law of X is

$$\mu_X(E) = \begin{cases} 0, & E \not\ni 0, 1, \\ p, & E \ni 1,\ E \not\ni 0, \\ 1-p, & E \ni 0,\ E \not\ni 1, \\ 1, & E \ni 0, 1. \end{cases} = p\delta_1(E) + (1-p)\delta_0(E),$$

Example 2.2.5: Uniform random variable

A uniform random variable is a random variable taking values in $[a, b]$ with law

$$\mu_X(E) = \frac{1}{b-a}\lambda_1(E \cap [a, b]),$$

where λ_1 is the one-dimensional Lebesgue measure. We write $X \sim U([a, b])$.

2.3. Change of measure

By definition of law we can write

$$\mathbb{E}[1_E(X)] = \mathbb{P}(X \in E) = \mu_X(E) = \int_{\mathbb{R}} 1_E \, d\mu_X.$$

By linearity, if $s = \sum_{k=0}^n c_k 1_{E_k}$ is a simple Borel function (that is $E_n \in \mathscr{B}_{\mathbb{R}}$ for every n), we have

$$\mathbb{E}[s(X)] = \mathbb{E}\left[\sum_{k=0}^n c_k 1_{E_k}\right] = \sum_{k=0}^n c_k \mathbb{E}[1_{E_k}(X)] = \sum_{k=0}^n c_k \int_{\mathbb{R}} 1_{E_k} \, d\mu_X = \int_{\mathbb{R}} s \, d\mu_X.$$

This formula extends to any function $\phi(X)$ provided ϕ be a Borel measurable function.

Proposition 2.3.1

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, X a random variable. Let $\phi = \phi(x) : \mathbb{R} \longrightarrow \mathbb{R}$ be a Borel measurable function. We have $\phi(X) \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ iff $\phi \in L^1(\mathbb{R}, \mathscr{B}_{\mathbb{R}}, \mu_X)$ and the following identity holds:

$$\mathbb{E}[\phi(X)] = \int_{\mathbb{R}} \phi \, d\mu_X. \tag{2.3.1}$$

In particular,

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \, d\mu_X(x), \quad \mathbb{V}[X] = \int_{\mathbb{R}} x^2 \, d\mu_X(x) - \left(\int_{\mathbb{R}} x \, d\mu_X(x)\right)^2.$$

Proof. Let $\phi \in L_+(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ be a positive Borel-function. As we know for general positive measurable functions, there exists a sequence $(s_n) \subset L_+(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ of simple Borel functions such that $s_n \uparrow \phi$ on $\mathbb{R}$. Then, by monotone convergence,

$$\lim_n \int_{\mathbb{R}} s_n \, d\mu_X = \int_{\mathbb{R}} \phi \, d\mu_X.$$

On the other hand, $s_n(X) \uparrow \phi(X)$ $\mathbb{P}-$a.s. and. by monotone convergence,

$$\lim_n \mathbb{E}[s_n(X)] = \mathbb{E}[\phi(X)].$$

So, formula (2.3.1) holds for every $\phi \in L_+(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$.
Let now $\phi \in L(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ be a generic Borel-measurable function. The, as well knonw,

$$\phi \in L^1(\mathbb{R}, \mu_X) \iff \int_{\mathbb{R}} \phi_\pm \, d\mu_X = \mathbb{E}[\phi_\pm(X)] < +\infty, \iff \phi(X) \in L^1(\Omega, \mathbb{P}).$$

In this case,

$$\int_{\mathbb{R}} \phi \, d\mu_X = \int_{\mathbb{R}} \phi_+ \, d\mu_X - \int_{\mathbb{R}} \phi_- \, d\mu_X = \mathbb{E}[\phi_+(X)] - \mathbb{E}[\phi_-(X)] = \mathbb{E}[\phi(X)]. \quad \square$$

Example 2.3.2

Let $X \sim U([a,b])$. Then

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \, d\mu_X, \text{ where } \mu(E) = \frac{1}{b-a}\lambda_1(E \cap [a,b]).$$

Notice that

$$\mu_X(E) = \frac{1}{b-a}\int_{E\cap[a,b]} dx = \frac{1}{b-a}\int_a^b 1_E(x)\, dx,$$

so

$$\int_{\mathbb{R}} \phi(x)\, d\mu_X(x) = \frac{1}{b-a}\int_a^b \phi(x)\, dx,$$

whence

$$\mathbb{E}[X] = \frac{1}{b-a}\int_a^b x \, dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_{x=a}^{x=b} = \frac{a+b}{2}.$$

Notice also that

$$\mathbb{E}[X^2] = \frac{1}{b-a}\int_a^b x^2 \, dx \frac{1}{b-a}\left[\frac{x^3}{3}\right]_{x=a}^{x=b} = \frac{1}{3}\frac{b^3-a^3}{b-a} = \frac{a^2+ab+b^2}{3}.$$

Therefore

$$\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

2.4. Markowitz's Optimal Portfolio selection

Markowitz's Optimal Portfolio selection model was introduced in 1952 to describe the efficient selection of a portfolio. An investor seeks for the most efficient allocation of a *wealth* w in a portfolio made of N *risky* assets and 1 *risk free* asset. The assets have known values x_k ($k = 1, \ldots, N+1$, $k = N+1$ represents the risk free asset) at moment when the decision on the allocation is made, and *uncertain* values X_k at moment when uncertainty is resolved.

Let $(a_1, \ldots, a_{N+1})$ the array of the allocations. Because of wealth constraint,

$$w = a_1 x_1 + \cdots + a_N x_N + a_{N+1} x_{N+1}. \tag{2.4.1}$$

Final wealth is then

$$W = a_1 X_1 + \cdots + a_N X_N + a_{N+1} X_{N+1}$$

The *rate of return* of each asset is the quantity R_j such that

$$X_j = (1 + R_j) x_j.$$

Notice that, for the risk free asset, $R_{N+1} \equiv r$ is constant. The portfolio return rate is then

$$\begin{aligned} W &= \sum_{j=1}^{N} a_j(1+R_j)x_j + a_{N+1}(1+r)x_{N+1} = w + \sum_{j=1}^{N} a_j x_j R_j + a_{N+1} r x_{N+1} \\ &= \left(1 + \sum_{j=1}^{N} \frac{a_j x_j}{w} R_j + \frac{a_{N+1}x_{N+1}}{w} r\right) w \\ &=: \left(1 + \vec{\theta}\cdot\vec{R} + \theta_{N+1} r\right) w \end{aligned}$$

where $\theta_j := \frac{a_j x_j}{w}$. Notice that, because of the budget constraint (2.4.1), we have

$$\sum_{j=1}^{N} \theta_j + \theta_{N+1} = \frac{1}{w}\left(a_1 x_1 + \cdots + a_{N+1}x_{N+1}\right) = 1.$$

This allows a one variable reduction setting $\theta_{N+1} = 1 - \sum_{j=1}^{N} \theta_j$,

$$W = \left(1 + \vec{\theta}\cdot\vec{R} + (1 - \vec{\theta}\cdot\vec{1})r\right) w,$$

Hereafter we use the notation $W_{\vec{\theta}}$ for the previous value, and we denote by

$$R(\vec{\theta}) := \vec{\theta}\cdot\vec{R} + (1 - \vec{\theta}\cdot\vec{1})r$$

the portfolio return rate. Notice that the array $\vec{\theta}$ is now unconstrained in $\mathbb{R}^n$.

The *efficient investor problem* consists in determining the optimal allocation that combines the highest expected rate of return $R(\vec{\theta})$ with the minimum possible risk. We assume an underlying probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Here $\mathbb{P}$ is also called *physical probability* and, in the model, it reflects the *beliefs* of the investor. The return rates $R_1, \dots, R_N$ of risky assets are r.vs. We assume moreover that the investor is *risk averse*. This means that the investor is disappointed when an investment yields a high risk. We assume as *measure of risk* variance

$$\mathbb{V}[R] = \mathbb{E}\left[(R - \mathbb{E}[R])^2\right]$$

Of course, this is a very limited way of measuring risk. For example, $\mathbb{V}[R]$ does not distinguish between good states – when the rate of return $R > \mathbb{E}[R]$ is above its expectation – from bad states, when the opposite happens. Nonetheless, since pioneering work of Markowitz, it is a very popular measure of risk. The basic idea is that $\mathbb{V}[(R - \mathbb{E}[R])^2]$ emphasizes large displacements from expected return $\mathbb{E}[R]$

To cope expected return with risk, Markowitz proposed a mixed *target functional*

$$\mathbb{MV}[R(\vec{\theta})] = \mathbb{E}[R(\vec{\theta})] - \frac{\varrho}{2}\mathbb{V}[R(\vec{\theta})].$$

Here, $\varrho > 0$ it is called *risk aversion* parameter, it yields a way to weight risk respect to expected return. We can now formalize *Markowitz's Optimal Portfolio Selection Problem*:

$$\max_{\vec{\theta}\in\mathbb{R}^n} \mathbb{MV}[R(\vec{\theta})].$$

We can easily solve the problem now. First notice that

$$R(\vec{\theta}) = \vec{\theta} \cdot (\vec{R} - \vec{r}) + r,$$

and since, as easily checked, $\mathbb{MV}[R(\vec{\theta})] = \mathbb{MV}[\vec{\theta} \cdot (\vec{R} - \vec{r})] + r$. Now,

$$\mathbb{E}[\vec{\theta} \cdot (\vec{R} - \vec{r})] = \vec{\theta} \cdot \mathbb{E}[\vec{R} - \vec{r}] = \vec{\theta} \cdot (\vec{\mu} - \vec{r}),$$

where $\vec{\mu} = \mathbb{E}[\vec{R}]$ is the array of expected return rates. Notice also that

$$\mathbb{V}[\vec{\theta} \cdot (\vec{R} - \vec{r})] = \mathbb{E}\left[\left(\vec{\theta} \cdot (\vec{R} - \vec{\mu})\right)^2\right] = \mathbb{E}\left[\sum_{i,j} \theta_i \theta_j (R_i - \mu_i)(R_j - \mu_j)\right] = C\vec{\theta} \cdot \vec{\theta},$$

where C is the $N \times N$ covariance matrix

$$C = [C_{ij}], \;\; C_{ij} = \mathrm{Cov}(R_i - \mu_i, R_j - \mu_j) = \mathrm{Cov}(R_i, R_j).$$

In particular, C is a positive definite matrix. Therefore

$$\mathbb{MV}[R(\vec{\theta})] = \vec{\theta} \cdot (\vec{\mu} - \vec{r}) - \frac{\varrho}{2} C\vec{\theta} \cdot \vec{\theta}.$$

Assuming the covariance matrix C strictly positive definite, the target function is negative quadratic, hence it has a global maximum $\vec{\theta}^*$ that verifies the first order condition

$$\nabla \mathbb{MV}[R(\vec{\theta}^*)] = \vec{\mu} - \vec{r} - \varrho C\vec{\theta}^* = \vec{0},$$

from which we obtain the well known Markowitz formula:

$$\boxed{\vec{\theta}^* = \frac{1}{\varrho} C^{-1}(\vec{\mu} - \vec{r}).}$$

2.5. Exercises

Exercise 2.5.1 ($**$). *Let $X, Y \in L^2$ be non constant r.vs.*

i) *Check that $\rho(aX + b, cY + d) = \pm\rho(X, Y)$, for every $a, b, c, d \in \mathbb{R}$.*
ii) ($+$) *True or false: is $\rho(X, Y) = \pm 1$ iff $Y = aX + b$ for some $a, b \in \mathbb{R}$? (hint: think to Cauchy-Schwarz inequality)*

Exercise 2.5.2 ($**$). *Show with an example that we may have $\rho(X, Y^2) = \pm 1$ and $\rho(X, Y) = 0$.*

Exercise 2.5.3 ($**$). *Let X, Y be two random variable on $(\Omega, \mathscr{F}, \mathbb{P})$ such that $\mu_X(I) = \mu_Y(I)$ for every $I \subset \mathbb{R}$ interval.*

i) *Let $\mathscr{S} := \{E \in \mathscr{B}_{\mathbb{R}} \;:\; \mu_X(E) = \mu_Y(E)\}$. Check that $\mathscr{S}$ is a $\sigma-$algebra.*
ii) *Deduce $\mu_X = \mu_Y$.*

Exercise 2.5.4 ($**$). *We consider an extension of the concept of measurable function. Let $\Omega_{1,2}$ two sets, $\mathscr{F}_{1,2}$ two $\sigma-$algebras of sets of, resp., $\Omega_{1,2}$. We say that a map $T : \Omega_1 \longrightarrow \Omega_2$ is measurable wrt $\mathscr{F}_{1,2}$ if*

$$\{T \in E\} \in \mathscr{F}_1, \; \forall E \in \mathscr{F}_2.$$

We write $T \in L((\Omega_1, \mathscr{F}_1); (\Omega_2, \mathscr{F}_2))$.

i) *Check that $T^{-1}(\mathscr{F}_2) = \{T^{-1}(E) \;:\; E \in \mathscr{F}_2\}$ is a sub $\sigma-$algebra of $\mathscr{F}_1$.*
ii) *Check that by composing two measurable maps you get a measurable map.*

3

Cumulative Distribution Function (cdf)

3.1. Definition and main properties

From the probabilistic point of view, a random variable X is fully described by its law. This is a measure on $\mathbb{R}$ equipped with the Borel $\sigma-$algebra. However, dealing with measures is not easy. It is therefore natural to find some other mathematical tool to make the handling of a random variable a bit easier. The cdf is a function associated to every random variable.

Definition 3.1.1

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, X a random variable. We call **cumulative distribution function of** X the function $F_X : \mathbb{R} \longrightarrow [0,1]$ defined as

$$F_X(x) := \mathbb{P}(X \leqslant x),\ (= \mu_X(]-\infty, x])),\ x \in \mathbb{R}.$$

The principal properties of cdf are listed in the following Proposition.

Proposition 3.1.2

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, F_X the cdf of the random variable X. The following properties hold:

i) F_X is increasing, that is $F_X(x) \leqslant F_X(y)$ for every $x \leqslant y$.

ii) $F_X(-\infty) = \lim_{x\to-\infty} F_X(x) = 0$ and $F_X(+\infty) = \lim_{x\to+\infty} F_X(x) = 1$.

iii) F_X is **right continuous**, that is

$$\exists \lim_{y\to x+} F_X(y) = F_X(x),\ \forall x \in \mathbb{R}.$$

iv) F_X has **left limit**, that is

$$\exists \lim_{y\to x-} F_X(y) \leqslant F_X(x),\ \forall x \in \mathbb{R}.$$

Proof. i) This is an easy consequence of monotonicity of probability measure $\mathbb{P}$: if $x \leqslant y$ then $\{X \leqslant x\} \subset \{X \leqslant y\}$m thus

$$F_X(x) = \mathbb{P}(X \leqslant x) \leqslant \mathbb{P}(X \leqslant y) = F_X(y).$$

ii) This follows from continuity properties of probability measure $\mathbb{P}$. We first notice that, since by i) F_X is monotonic, unilateral limits always exist. Thus, in particular, $\alpha := \lim_{x\to-\infty} F_X(x)$ and $\beta := \lim_{x\to+\infty} F_X(x)$ exist. Now, for the first take $n \in \mathbb{Z}$ and define $E_n := \{X \leqslant x_n\}$. Notice that $E_n \downarrow \{X \leqslant$

$-\infty\} = \emptyset$ when $n \downarrow -\infty$. Since $\mathbb{P}$ is a finite measure, it is continuous from above, so

$$0 = \mathbb{P}(\emptyset) = \lim_{n\to-\infty} \mathbb{P}(X \leqslant n) = \lim_{n\to-\infty} F_X(n) = \alpha.$$

Similarly, $E_n \uparrow \{X \leqslant +\infty\} = \Omega$ when $n \uparrow +\infty$ thus, by continuity from below,

$$1 = \mathbb{P}(\Omega) = \lim_{n\to+\infty} \mathbb{P}(X \leqslant n) = \lim_{n\to+\infty} F_X(n) = \beta.$$

iii) By i), left limit $\lim_{y\to x+} F_X(y)$ exists, and since $F_X \nearrow$, we have also $\lim_{y\to x+} F_X(y) \geqslant F_X(x)$. To prove the equality, take $y_n \downarrow x$ and define $E_n := \{X \leqslant x_n\}$, in such a way that

$$E_n \downarrow \bigcap_k E_k = \bigcap_k \{X \leqslant y_k\}$$

We claim this intersection is $\{X \leqslant x\}$. Indeed, since $y_k > x$, $\{X \leqslant x\} \subset \{X \leqslant y_k\}$ for every k¡ thus $\{X \leqslant x\} \subset \bigcap_k \{X \leqslant y_k\}$. Conversely, if $\omega \in \bigcap_k \{X \leqslant y_k\}$, then $X(\omega) \leqslant y_k$ for every k. Letting $k \longrightarrow +\infty$ we have $X(\omega) \leqslant \lim_k y_k = x$, thus $\omega \in \{X \leqslant x\}$, and this proves $\bigcap_k \{X \leqslant y_k\} \subset \{X \leqslant x\}$, from which equality follows. Thus

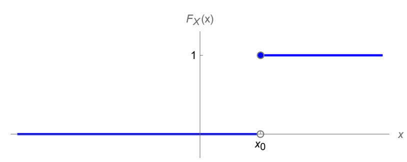
$$E_n \downarrow \{X \leqslant x\},$$

and by continuity from above the conclusion now follows.
iv) Existence of the left limit, once more, follows by the monotonic nature of F_X and since $F_X(y) \leqslant F_X(x)$ for every $y < x$, we deduce $\lim_{y\to x-} F_X(y) \leqslant F_X(x)$. □

Example 3.1.3

Let $X = x_0$ a.s.. In this case the law of X is the delta Dirac δ_{x_0}. The cdf is

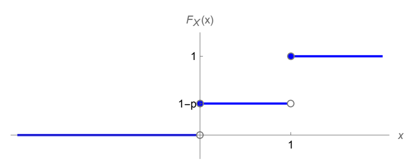
$$F_X(x) = \mathbb{P}(X \leqslant x) = \mu_X(]-\infty, x]) = \delta_{x_0}(]-\infty, x]) = 1_{[x_0,+\infty[}(x).$$



Example 3.1.4

Let $X \sim \mathrm{Ber}(p)$ be a Bernoulli random variable of parameter p. The cdf is

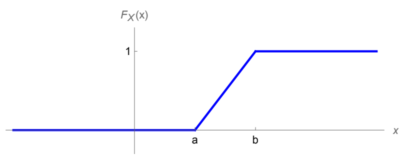
$$F_X(x) = \mu_X(]-\infty, x]) = \begin{cases} 0, & x < 0, \\ 1-p, & 0 \leqslant x < 1, \\ 1, & x \geqslant 1. \end{cases}$$



Example 3.1.5

Let $X \sim U(a, b)$. We have

$$F_X(x) = \mu_X(]-\infty, x]) = \frac{1}{b-a}\lambda(]-\infty, x] \cap [a, b]) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \leqslant x \leqslant b, \\ 1, & x > b. \end{cases}$$



To each random variable X it is associated a cdf F_X. The viceversa also holds:

Proposition 3.1.6

Assume that $F = F(x) : \mathbb{R} \longrightarrow [0, 1]$ verifies properties i)–iv) of Proposition 3.1. There exists then a random variable X on a suitable probability space such that $F_X = F$.

Proof. For simplicity, we assume that F be *strictly increasing* and *continuous*. Let $(\Omega, \mathscr{F}, \mathbb{P}) := ([0, 1], \mathscr{B}([0, 1]), \lambda_1)$ where $\mathscr{B}([0, 1])$ are the Borel sets of $[0, 1]$. Now, since we want

$$\mathbb{P}(X \leqslant x) \equiv \lambda_1(X \leqslant x) = F_X(x),$$

the idea is to define, for $\omega \in [0, 1]$,

$$X(\omega) := F^{-1}(\omega).$$

In this way we would have

$$\mathbb{P}(X \leqslant x) = \lambda_1(\{\omega \in [0, 1] \ : \ F^{-1}(\omega) \leqslant x\}) = \lambda_1(\{\omega \in [0, 1] \ : \ \omega \leqslant F(x)\}) = \lambda_1([0, F(x)]) = F(x),$$

as desired.

3.2. Absolutely continuous random variable

Definition 3.2.1

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, X a random variable. We say that X is **absolutely continuous (a.c.) with density** $f_X \in L^1(\mathbb{R}, \mathscr{B}_{\mathbb{R}}, \lambda_1)$ **with respect to** $\mathbb{P}$ if

$$d\mu_X = f_X \, dx,$$

that is,

$$\mu_X(E) = \int_E f_X(x) \, dx, \ \forall E \in \mathscr{B}_{\mathbb{R}}. \tag{3.2.1}$$

Clearly, a density f_X is non negative and

$$\int_{\mathbb{R}} f_X(x) \, dx = \mu_X(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = 1.$$

Moreover, if X is a.c. random variable with density f, we have

$$\mathbb{E}[\phi(X)] = \int_{\mathbb{R}} \phi(x) f_X(x)\, dx.$$

This means that $\phi(X) \in L^1(\Omega)$ iff $\phi \in L^1(\mathbb{R}, \mu_X)$, iff

$$\int_{\mathbb{R}} |\phi(x)| f_X(x)\, dx < +\infty.$$

So, in particular,

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x)\, dx, \quad \mathbb{E}[X^2] = \int_{\mathbb{R}} x^2 f_X(x)\, dx.$$

provided, respectively,

$$\int_{\mathbb{R}} |x| f_X(x)\, dx < +\infty, \quad \int_{\mathbb{R}} x^2 f_X(x)\, dx < +\infty.$$

If X is an a.c. random variable, then

$$F_X(x) = \int_{]-\infty,x]} f_X(y)\, dy \equiv \int_{-\infty}^{x} f_X(y)\, dy.$$

From this we have the

Proposition 3.2.2

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, F_X the cdf of a random variable X. Then, X is absolutely continuous with density f_X iff

i) $F_X \in \mathscr{C}(\mathbb{R})$;

ii) F_X is a.e. differentiable and

$$F'_X(x) = f_X(x), \; a.e.\, x \in \mathbb{R}.$$

iii) $f_X \in L^1(\mathbb{R})$.

Proof. Necessity: let X be a.c. and let's prove that i),ii),iii) hold. We already know that F_X is right continuous with left limit. If $x_n \uparrow x$ then, by monotone convergence,

$$F_X(x_n) = \int_{\mathbb{R}} 1_{]-\infty,x_n]}(x) f_X(x)\, dx \longrightarrow \int_{\mathbb{R}} 1_{]-\infty,x[}(x) f_X(x)\, dx = \int_{\mathbb{R}} 1_{]-\infty,x]}(x) f_X(x)\, dx = F_X(x),$$

since singletons are null sets for the Lebesgue measure. This proves continuity. Differentiability is more complex. We start noticing that if $\varepsilon > 0$,

$$\frac{F_X(x+\varepsilon) - F_X(x)}{\varepsilon} = \frac{1}{\varepsilon} \int_x^{x+\varepsilon} f_X(y)\, dy = \int_{\mathbb{R}} f_X(y) \frac{1}{\varepsilon} 1_{[0,\varepsilon]}(y-x)\, dy = f * \delta_\varepsilon(x),$$

where $\delta_\varepsilon(u) = \frac{1}{\varepsilon} 1_{[0,1]}(-\frac{u}{\varepsilon})$ is an approximate unit and $f * \delta_\varepsilon \stackrel{L^1}{\longrightarrow} f$. We want pointwise convergence. We notice that the previous property implies that

$$\forall (\varepsilon_n),\; \varepsilon_n \longrightarrow 0,\; \exists (\varepsilon_{n_k}) \;:\; \frac{F_X(x+\varepsilon_{n_k}) - F_X(x)}{\varepsilon_{n_k}} \longrightarrow f_X(x),\; a.e. \tag{3.2.2}$$

Now,, pick an x for which the previous property is true and assume We claim that

$$\exists \lim_{\varepsilon\to 0} \frac{F_X(x+\varepsilon)-F_X(x)}{\varepsilon} = f_X(x),\ a.e.\, x \in \mathbb{R}.$$

Indeed, if false, there would be $\varepsilon_n \longrightarrow 0$ for which

$$\lambda_1\left(\lim_n \frac{F_X(x+\varepsilon_n)-F_X(x)}{\varepsilon_n} \neq f_X(x)\right) > 0.$$

But then,

$$\lambda_1\left(\lim_k \frac{F_X(x+\varepsilon_{n_k})-F_X(x)}{\varepsilon_{n_k}} \neq f_X(x)\right) > 0,$$

that is

$$\lambda_1\left(f_X(x) \neq f_X(x)\right) > 0,$$

which is impossible!

Sufficiency: we assume i),ii) and iii) hold and we prove X is a.c. By ii) and the fundamental theorem of integral calculus (weak form) we have

$$\begin{aligned}\mu_X(]a,b]) &= \mathbb{P}(a < X \leqslant b) = \mathbb{P}(\{X\leqslant b\}\setminus\{X\leqslant a\}) = \mathbb{P}(X\leqslant b) - \mathbb{P}(X \leqslant a)\\ &= F_X(b)-F_X(a) = \int_a^b f_X(y)\, dy,\end{aligned}$$

and since F_X is continuous, from the continuity from above (valud for μ_X probability measure), we have

$$\mu_X([a,b]) = \lim_{\alpha\to a-} \mu_X(]\alpha,b]) = \lim_{\alpha\to a-}(F_X(b)-F_X(\alpha)) = F_X(b)-F_X(a) = \int_a^b f_X(x)\, dx.$$

In other words, $\mu_X(I) = \int_I f_X$ for every I interval, and since this extends to the $\sigma-$algebra generate by intervals, that is to $\mathscr{B}_{\mathbb{R}}$, we have that (3.2.1) holds, that is X is a.c.

Example 3.2.3

Constant and Bernoulli r.vs are not a.c. being their cdf not continuous. For $X \sim U([a,b])$ we have

$$\exists F_X'(x) = \begin{cases} 0, & x, a,\ x > b,\\ \frac{1}{b-a}, & a < x < b.\end{cases} \overset{a.e.}{=} \frac{1}{b-a}1_{[a,b]}(x) = f_X(x).$$

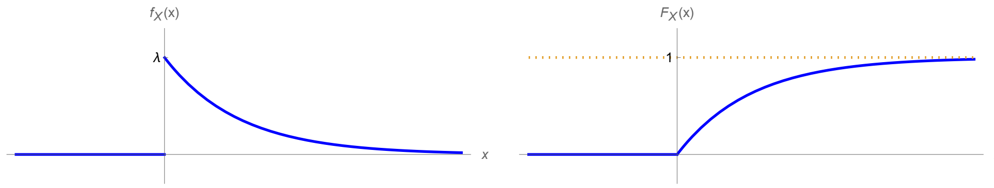
Density is also a practical way to introduce random variables. Here some remarkable examples.

3.2.1. Exponential. An *exponential random variable*, notation $X \sim \exp\lambda$, is a random variable with density

$$\boxed{f_X(x) = \lambda e^{-\lambda x}1_{[0,+\infty[}(x),}$$

with $\lambda > 0$ (notice that $\int_{\mathbb{R}} f_X = \int_0^{+\infty} \lambda e^{-\lambda x}\, dx = [-e^{-\lambda x}]_{x=0}^{x=+\infty} = 0-(-1) = 1$). The cdf is

$$F_X(x) = \int_{-\infty}^x f_X(y)\, dy = \begin{cases} 0, & x \leqslant 0,\\ \int_0^x \lambda e^{-\lambda y}\, dy = [-e^{-\lambda y}]_{x=0}^{x=y} = 1-e^{-\lambda x}, & x > 0.\end{cases}$$



We have also

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X = \int_0^{+\infty} \lambda x e^{-\lambda x}\, dx = \frac{1}{\lambda}\int_0^{+\infty} u e^{-u}\, du = \frac{1}{\lambda},$$

$$\mathbb{V}[X] = \int_0^{+\infty} x^2 \lambda e^{-\lambda x}\, dx - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}\left(\int_0^{+\infty} u^2 e^{-u}\, du - 1\right) = \frac{1}{\lambda^2}.$$

Exponential random variable are used to model the distribution of random occurrence times.

3.2.2. Gaussian. A Gaussian (or normal) random variable, notation $X \sim \mathcal{N}(m, \sigma^2)$ with $m \in \mathbb{R}$ and $\sigma^2 > 0$, is a random variable with density

$$\boxed{f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}.}$$

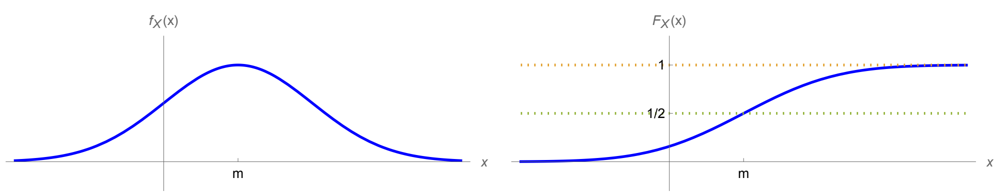
An $X \sim \mathcal{N}(0,1)$ is also called **standard gaussian** random variable. There is not an explicit formula for the cdf

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}\, dx = \Phi\left(\frac{x-m}{\sigma}\right),$$

where Φ is the cdf of the standard Gaussian

$$\Phi(x) = \int_{-\infty}^{x} e^{-\frac{y^2}{2}} \frac{dy}{\sqrt{2\pi}},$$

which is considered as an elementary function.



The function

$$\mathrm{Erf}(x) := \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-y^2}\, dy$$

is called **error function** and

$$\Phi(x) = \frac{1}{2}\left(1 + \mathrm{Erf}\left(\frac{x}{\sqrt{2}}\right)\right)$$

We also have

$$\mathbb{E}[X] = m, \quad \mathbb{V}[X] = \sigma^2.$$

Gaussian random variable are central in Probability Theory. Because of the Central Limit Theorem, averages of *independent random variable* with the same distribution converge to gaussian distributions, these last are used to model phenomena for which no particular information on the randomness is known.

3.2.3. Gamma. This is a class of distributions that extends the exponentials. We recall that the Euler's Γ function is defined by the integral

$$\Gamma(\alpha) := \int_0^{+\infty} u^{\alpha-1} e^{-u}\, du.$$

It is not difficult to show that $\exists \Gamma(\alpha) \in \mathbb{R}$ iff $\alpha - 1 > -1$, that is $\alpha > 0$. We notice also that, if $n \in \mathbb{N}$,

$$\begin{aligned}\Gamma(n+1) &= \int_0^{+\infty} u^n e^{-u}\, du = \int_0^{+\infty} u^n \left(-e^{-u}\right)'\, du = \left[-u^n e^{-u}\right]_{u=0}^{u=+\infty} + \int_0^{+\infty} n u^{n-1} e^{-u}\, d\xi \\ &= n\Gamma(n)\end{aligned}$$

Then,

$$\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)(n-2)\Gamma(n-2) = \ldots = (n-1)(n-2)\cdots 1\Gamma(1),$$

and since

$$\Gamma(1) = \int_0^{+\infty} e^{-\xi}\, d\xi = 1,$$

we have $\Gamma(n) = (n-1)!$. Let now

$$f_{\alpha,\lambda}(x) = C_{\alpha,\lambda} x^{\alpha-1} e^{-\lambda x} 1_{[0,+\infty[}(x), \quad (\alpha > 1,\ \lambda > 0).$$

We determine $C_{\alpha,\lambda}$ in such a way that $f_{\alpha,\lambda}$ be a probability density. Since

$$\int_{\mathbb{R}} f_{\alpha,\lambda}(x)\, dx = \int_0^{+\infty} C_{\alpha,\lambda} x^{\alpha-1} e^{-\lambda x}\, dx \overset{\lambda x = u}{=} C_{\alpha,\lambda} \int_0^{+\infty} \frac{1}{\lambda^{\alpha-1}} u^{\alpha-1} e^{-u}\, du = C_{\alpha,\lambda} \frac{\Gamma(\alpha)}{\lambda^{\alpha-1}},$$

from which

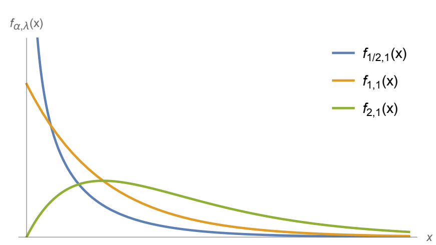
$$\int_{\mathbb{R}} f_{\alpha,\lambda}(\xi)\, d\xi = 1, \quad \Longleftrightarrow \quad C_{\alpha,\lambda} = \frac{\lambda^{\alpha-1}}{\Gamma(\alpha)}.$$

We say that a random variable X has **gamma distribution** $X \sim \Gamma(\alpha, \lambda)$ if it is a.c. with density

$$\boxed{f_{\alpha,\lambda}(x) := \frac{\lambda^{\alpha-1}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} 1_{[0,+\infty[}(x).}$$

Notice that $\Gamma(1, \lambda) = \exp(\lambda)$ and, for $n \in \mathbb{N}$,

$$f_{n,1}(x) = \frac{1}{(n-1)!} x^{n-1} e^{-x} \chi_{[0,+\infty[}(x).$$



It holds

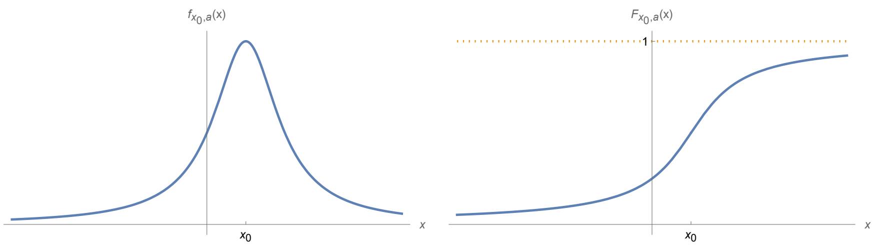
$$\mathbb{E}[X] = \frac{\alpha}{\lambda}, \quad \mathbb{V}[X] = \frac{\alpha}{\lambda^2}.$$

3.2.4. Cauchy. A Cauchy random variable $X \sim C(x_0, a)$ is a random variable with density

$$\boxed{f_{x_0,a}(x) = \frac{1}{\pi}\frac{a}{a^2 + (x - x_0)^2}. \quad (x_0 \in \mathbb{R},\ a > 0)}$$

It is easy to verify that f_X is a probability density. The cdf is

$$\begin{aligned} F_{x_0,a}(x) &= \int_{-\infty}^{x} f_X(y)\,dy = \frac{1}{\pi}\int_{-\infty}^{x} \frac{1}{1+\left(\frac{y-x_0}{a}\right)^2}\,\frac{dy}{a} = \frac{1}{\pi}\left[\arctan \frac{y-x_0}{a}\right]_{y=-\infty}^{y=x} \\ &= \frac{1}{\pi}\left(\arctan \frac{x-x_0}{a} + \frac{\pi}{2}\right) = \frac{1}{\pi}\arctan \frac{x-x_0}{a} + \frac{1}{2}. \end{aligned}$$



If $X \sim C(x_0, a)$ then $\mathbb{E}[X]$ is not defined. Indeed,

$$\mathbb{E}[|X|] = \int_{\mathbb{R}} |x| f_X(x)\,dx = \frac{a}{\pi}\int_{\mathbb{R}} \frac{|x|}{a^2 + (x - x_0)^2}\,dx = +\infty,$$

being the intrgrand not integrable at $\pm\infty$. Similarly, $\mathbb{V}[X] = +\infty$.

3.2.5. Log-Normal. A log-normal random variable, notation $\log X \sim \mathcal{N}(m, \sigma^2)$ is a random variable $X = e^Y$ where $Y \sim \mathcal{N}(m, \sigma^2)$. We need a general result:

Proposition 3.2.4

Let $X = \phi(Y)$ where ϕ is a regular $1-1$ bijection and Y has density f_Y. Then

$$\boxed{f_X(x) = f_Y(\phi^{-1}(x))|(\phi^{-1})'(x)|.}$$

Proof. Just notice that

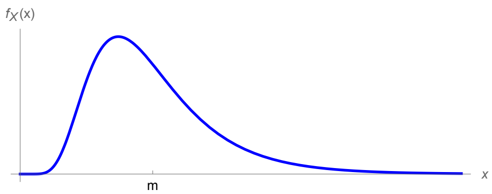
$$\begin{aligned} \mu_X(E) &= \mathbb{P}(X \in E) = \mathbb{P}(\phi(Y) \in E) = \mathbb{P}\left(Y \in \phi^{-1}(E)\right) = \mu_Y(\phi^{-1}(E)) \\ &= \int_{\phi^{-1}(E)} f_Y(y)\,dy \overset{x=\phi(y),\ y=\phi^{-1}(x)}{=} \int_E \underbrace{f_Y(\phi^{-1}(x))|(\phi^{-1})'(x)|}_{=:f_X(x)}\,dx. \quad \square \end{aligned}$$

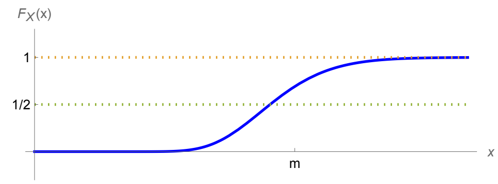
So, in particular, if $X = e^Y$ with $Y \sim \mathcal{N}(m, \sigma^2)$ we have $\phi(y) = e^y$ and $\phi^{-1}(x) = \log x$, so

$$\boxed{f_X(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\log x - m)^2}{2\sigma^2}} 1_{]0,+\infty[}(x).}$$

For the cdf we have

$$F_X(x) = \Phi\left(\frac{\log x - m}{\sigma}\right).$$





We also have

$$\mathbb{E}[X] = e^{m+\frac{\sigma^2}{2}}, \quad \mathbb{V}[X] = (e^{\sigma^2} - 1)e^{2m+\sigma^2}.$$

Log-normal distributions are used in Finance to model prices.

3.2.6. Piecewise regular cdfs. We now derive a representation formula for a regular cdf F_X with a finite number of discontinuities. The conclusion holds under less demanding conditions, but this representation is sufficient for many applied cases.

Theorem 3.2.5

Let F_X be the cdf of a random variable X. Assume that $F_X \in \mathscr{C}^1(\mathbb{R}\backslash\{x_1, \ldots, x_N\})$, where $x_1 < x_2 < \ldots < x_N$ are the discontinuity points of F_X. Set

$$\lambda_k := F_X(x_k) - F_X(x_k-) > 0. \ (k - 1, \ldots, N)$$

There exists then $f \in L^1(\mathbb{R})$ such that

$$F_X(x) = \underbrace{\sum_{k=1}^{N} \lambda_k H(x - x_k)}_{=:F_X^s(x)} + \underbrace{\int_{-\infty}^{x} f(y)\, dy}_{=:F_X^{ac}(x)}, \quad \forall x \in \mathbb{R}, \tag{3.2.3}$$

where $H(u) = 1_{[0,+\infty[}(u)$ is the Heaviside function.

F_X^s is called **singular component** and F_X^{ac} is called **absolutely continuous component**.

Proof. For simplicity, we assume that $F_X \in \mathscr{C}^1(\mathbb{R}\backslash\{x^*\})$. Let

$$f_X(x) := F_X'(x), \quad x \in \mathbb{R}\backslash\{x^*\}$$

By the fundamental thm of Integral Calculus, if $x < x^*$ we have

$$F_X(x) - F_X(-\infty) = \int_{-\infty}^{x} f_X(y)\, dy, \quad \Longleftrightarrow \quad F_X(x) = \int_{-\infty}^{x} f_X(u)\, du, \quad \forall x < x^*.$$

In particular,

$$F_X(x^*-) = \lim_{x \to x^*-} F_X(x) = \int_{-\infty}^{x^*} f_X(u)\, du.$$

Setting $\lambda^* := F(x^*) - F(x^*-) > 0$, we have

$$F_X(x) = \lambda^* H(x - x^*) + \int_{-\infty}^{x} f(u)\, du, \quad \forall x \leqslant x^*.$$

This formula holds also for $x > x^*$. Indeed, if $x^* < y < x$ we have

$$F_X(x) - F_X(y) = \int_{y}^{x} f_X(u)\, du,$$

and letting $y \longrightarrow x^*+$ we have

$$F_X(x) = F_X(x^*) + \int_{x^*}^{x} f_X(u)\, du = \lambda^* + \int_{-\infty}^{x^*} f_X(u)\, du + \int_{x^*}^{x} f_X(u)\, du = \lambda^* H(x - x^*) + \int_{-\infty}^{x} f_X(u)\, du,$$

as claimed.

3.3. The classical Newsvendor model

The *newsvendor* model is a mathematical model in Operations Management used to determine optimal inventory levels. A firm produces a certain quantity q of a good at unit cost $c > 0$ selling at unit price $p > c$ and facing an uncertain demand $D \geqslant 0$. The origin of the name comes by analogy with the situation faced by a newspaper vendor who must decide how many copies of the day's paper to stock in the face of uncertain demand and knowing that unsold copies will be worthless at the end of the day. It is one of the most ancient inventory models dating back to XIX century.

Let us give a mathematical form to this problem. The key ingredient is the profit and loss statement (P&L). Let q be the number of units produced/ordered at unit cost c. This quantity q is assumed to be positive and determined by the firm. In particular, it is not uncertain. The total cost of production is $C(q) := -cq$ and it has a deterministic nature. Revenues come from the sale of goods at a future time, when the demand D is uncertain and described by a random variable. The firm can sell only what produced until the demand is satisfied at unit price p. Therefore, the revenues are

$$R(q) = \begin{cases} pq, & D > q, \\ pD, & D \leqslant q. \end{cases} = p \min(q, D).$$

The P&L is then the *business position*

$$P(q) := R(q) - C(q) = p \min(q, D) - cq.$$

The P&L is then uncertain. We assume the firm aims to maximize the expected P&L, that is solving

$$\max_{q \geqslant 0} \mathbb{E}\left[P(q)\right].$$

Notice that, if D is a.c. with density f_D and cdf F_D, we have

$$\begin{aligned} \mathbb{E}[P(q)] &= p\mathbb{E}[\min(q, D)] - cq = p\mathbb{E}[q \underbrace{1_{D>q}}_{1-1_{D\leqslant q}} + D1_{D\leqslant q}] - cq \\ &= p\left(q(1 - F_D(q)) + \int_0^q x f_D(x)\, dx\right) - cq \end{aligned}$$

Now, being $f_D = F'_D$, we have

$$\int_0^q x f_D(x)\, dx = \int_0^q x F'_D(x)\, dx = [xF_D(x)]_{x=0}^{x=q} - \int_0^q F_D(x)\, dx = qF_D(q) - \int_0^q F_D(x)\, dx,$$

so

$$u(q) := \mathbb{E}[P(q)] = p\left(q - \int_0^q F_D(x)\, dx\right) - cq.$$

Now,

$$u'(q) = p(1 - F_D(q)) - c \geqslant 0, \quad \Longleftrightarrow \quad 1 - F_D(q) \geqslant \frac{c}{p}, \quad \Longleftrightarrow \quad F_D(q) \leqslant 1 - \frac{c}{p},$$

and if F_D is strictly increasing, we obtain

$$q \leqslant F_D^{-1}\left(1 - \frac{c}{p}\right).$$

Therefore, the optimal q is

$$q^* := F_D^{-1}\left(1 - \frac{c}{p}\right).$$

For example, assume D has a uniform distribution in the range $[0, \widehat{d}]$. Then

$$F_D(x) = \begin{cases} 0, & x < 0, \\ \frac{x}{\widehat{d}}, & 0 \leqslant x \leqslant \widehat{d}, \\ 1, & x > \widehat{d}. \end{cases}$$

We have

$$F_D(q^*) = 1 - \frac{c}{p}, \quad \Longleftrightarrow \quad \frac{q^*}{\widehat{d}} = 1 - \frac{c}{p}, \quad \Longleftrightarrow \quad q^* = \left(1 - \frac{c}{p}\right)\widehat{d}.$$

3.4. Exercises

Exercise 3.4.1 ($*$). *Let F_X be the cdf of X. Use F_X to express the following probabilities:*

$$\mathbb{P}(a \leqslant X \leqslant b), \ \mathbb{P}(a < X < b), \ \mathbb{P}(X \geqslant b), \ \mathbb{P}(X = a).$$

Exercise 3.4.2 ($*$). *For each of the following F say if they are cdf of some random variable:*

i) $F(x) := (1 - \frac{e^{-x}}{3})1_{]0,+\infty[}(x)$
ii) $F(x) = \frac{1}{\pi}\arctan x + \frac{1}{2}$.
iii) $F(x) := \frac{1}{2} + \frac{1}{\pi}\arctan(x^3 - x)$

Exercise 3.4.3 ($*$). *Let*

$$F(x) := \begin{cases} 0, & x < 0, \\ & \\ 1 - 3^{-[x]}, & x \geqslant 0. \end{cases}$$

Is F the cdf of a random variable X? If yes, what is the probability $\mathbb{P}(X > 3)$ and of $\mathbb{P}(X = 2)$?

Exercise 3.4.4 ($**$). *Let X be a random variable with cdf F_X. What is the cdf of $|X|$?*

Exercise 3.4.5 ($**$). *Let X be an absolutely continuous random variable with density f_X. Show that also X^2 is a.c., and determine its density in terms of f_X.*

Exercise 3.4.6 $(**)$**.** *Let* $X \sim \mathcal{N}(m, \sigma^2)$.

i) *Determine if* $Y := \sqrt{|X|}$ *is a.c. and, in this case, compute its density* f_Y.
ii) *Determine if* $Y := \frac{1}{1+e^{-X}}$ *is a.c. and, in this case, compute its density* f_Y.

Exercise 3.4.7 $(**)$**.** *Let* $X \sim \exp(\lambda)$. *Determine the cdf of* $Y = [X]$ *(the integer part of* X*).*

Exercise 3.4.8 $(**)$**.** *Let* $P_0 := (x_0, y_0)$ *be a fixed point in the Cartesian plane with* $y_0 > 0$. *Consider the straight line* r_θ *passing through point* P_0*, where* θ *is the angle formed with the vertical line through* P_0. *Assume that* θ *is a uniformly distributed random variable taking values in the open interval* $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. *Determine the distribution of* X*, where* X *is the abscissa of the point where* r_θ *intersects the x-axis.*

Exercise 3.4.9 $(***)$**.** *Let* F_X *be the cdf of a random variable* X. *For each* x^* *discontinuity of* F_X*, we define*

$$I_{x^*} :=]F_X(x^*-), F_X(x^*)[.$$

i) *Check that, if* $x^* \neq y^*$ *then* $I_{x^*} \cap I_{y^*} = \emptyset$.
ii) *Deduce, from i), that the set* D *of discontinuity points of* F_X *is, at most, countable.*

4

Multivariate random variables

4.1. Definitions

Together with scalar random variables, we consider also *vector valued* random variables. Let $X = (X_1, \dots, X_N) : \Omega \longrightarrow \mathbb{R}^N$. We call X a **random vector** (or, **multivariate random variable**) if each X_j is a random variable (notation: $X \in L(\Omega)$ iff $X_j \in L(\Omega)$ for all $j = 1, \dots, N$). If $N = 2$, $X = (X_1, X_2)$ is also called **bi-variate** random variable. We define the **expected value** of X as

$$\mathbb{E}[X] := (\mathbb{E}[X_1], \dots, \mathbb{E}[X_N]) \in \mathbb{R}^N,$$

provided $X_j \in L^1$ for $j = 1, \dots, N$ (in this case we write $X \in L^1(\Omega)$). If $X_j \in L^2(\Omega)$, $j = 1, \dots, N$ (we write $X \in L^2(\Omega)$) the **covariance matrix** is defined: this is the matrix $C := [c_{ij}]$ where

$$c_{ij} := \mathrm{Cov}(X_i, X_j) \equiv \mathbb{E}\left[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])\right] \equiv \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j].$$

In general, the covariance matrix is a *symmetric and positive definite matrix* (notation $C \geqslant 0$), that is

$$Cv \cdot v \geqslant 0, \ \forall v \in \mathbb{R}^N$$

Indeed, clearly $c_{ij} = c_{ji}$. Moreover,

$$\begin{aligned} Cv \cdot v \quad &= \sum_{i,j} \mathbb{E}\left[(X_i - \mathbb{E}[X_i])v_i(X_j - \mathbb{E}[X_j])v_j\right] = \mathbb{E}\left[\sum_{i,j}(X_i - \mathbb{E}[X_i])v_i(X_j - \mathbb{E}[X_j])v_j\right] \\ &= \mathbb{E}\left[\left(\sum_i (X_i - \mathbb{E}[X_i])v_i\right)^2\right] \geqslant 0. \end{aligned}$$

It is sometimes useful to represent the entries of the covariance matrix in the form

$$c_{ij} = \mathrm{Cov}(X_i, X_j) = \rho_{ij}\sigma_i\sigma_j$$

where $\rho_{ij} = \rho(X_i, X_j)$ is the linear correlation of X_i and X_j, $\sigma_i = \mathbb{V}[X_i]^{1/2}$ is the standard deviation of X_i.

4.1.1. Law. Similarly to the scalar case, for random arrays we also have a definition of the law of X,

$$\mu_X(E) := \mathbb{P}(X \in E), \ \forall E \in \mathscr{B}_{\mathbb{R}^N}.$$

The well posedness of μ_X follows by an argument similar to the one-dimensional case. Given $\varphi : \mathbb{R}^N \longrightarrow \mathbb{R}$ we say that φ is **Borel-measurable** (or φ is a **Borel function**) if φ is measurable w.r.t. the Borel $\sigma-$algebra $\mathscr{B}_{\mathbb{R}^N}$. It turns out that

$$\varphi(X) \in L^1(\Omega, \mathscr{F}) \quad \Longleftrightarrow \quad \varphi \in L^1(\mathbb{R}^N, \mathscr{B}_{\mathbb{R}^N}, \lambda_N)$$

and the *change of variable formula* holds:

$$\mathbb{E}[\varphi(X)] = \int_{\mathbb{R}^N} \varphi(x_1,\ldots,x_N)\; d\mu_X(x_1,\ldots,x_N).$$

4.1.2. Cdf. The cdf of a r.vect. $X:\Omega \longrightarrow \mathbb{R}^N$ is a function $F_X:\mathbb{R}^N \longrightarrow [0,1]$ defined by

$$F_X(x_1,\ldots,x_N) := \mathbb{P}(X_1 \leqslant x_1,\ldots,X_N \leqslant x_N).$$

The cdf of a r.vect. fulfils properties similar to the cdfs of scalar (or **univariate**) r.vs. For example,

i) $\lim_{(x_1,\ldots,x_N)\longrightarrow(-\infty,\ldots,-\infty)} F_X(x_1,\ldots,x_N) = 0,, \lim_{(x_1,\ldots,x_N)\longrightarrow(+\infty,\ldots,+\infty)} F_X(x_1,\ldots,x_N) = 1.$
ii) F_X is monotonic increasing in each of its variables (the others remaining fixed).
iii) F_X is right continuous with left limits in each of its coordinates.

It is a straightforward exercise to prove these properties.

4.1.3. A.c. r.vects. We say that X is **absolutely continuous** (notation **a.c.**) if there exists $f_X(x_1,\ldots,x_N)$ such that

$$\mu_X(E) = \int_{\mathbb{R}^N} f_X(x_1,\ldots,x_N)\; dx_1\cdots dx_N,$$

where the last integral is w.r.t. the Lebesgue measure. In this case, $f_X \geqslant 0$ a.e. and

$$\int_{\mathbb{R}^N} f_X(x_1,\ldots,x_N)\; dx_1\cdots dx_N = 1,$$

and the change of variable formula takes the form

$$\mathbb{E}[\varphi(X)] = \int_{\mathbb{R}^N} \varphi(x_1,\ldots,x_N) f_X(x_1,\ldots,x_N)\; dx_1\cdots dx_N,$$

The density f_X is also called **joint density**. Having this, we automatically have that each of the components of X is an a.c. random variable. Indeed,

$$\mathbb{P}(X_j \in E_j) = \mathbb{P}(X \in \mathbb{R}^{j-1}\times E_j \times \mathbb{R}^{N-j-1}) = \int_{\mathbb{R}^{j-1}\times E_j\times\mathbb{R}^{N-j-1}} f_X(x_1,\ldots,x_N)\; dx_1\cdots dx_N$$

Since $f_X \in L^1(\mathbb{R}^N)$, Fubini-Tonelli theorem applies, so

$$\mathbb{P}(X_j \in E_j) = \int_{E_j} \underbrace{\left(\int_{\mathbb{R}^{N-1}} f_X(x_1,\ldots,x_N)\; dx_1\cdots dx_{j-1}dx_{j+1}\cdots dx_N\right)}_{=:f_{X_j}(x_j)} \; dx_j.$$

Functions

$$\boxed{f_{X_j}(x_j) = \int_{\mathbb{R}^{N-1}} f_X(x_1,\ldots,x_N)\; dx_1\cdots dx_{j-1}dx_{j+1}\cdots dx_N}$$

are called **marginal densities**. So, from the joint density it is possible to derive the marginal densities. The vice versa is more complex and we will return on later.

4.1.4. Multivariate Gaussian. A very important class of multivariate random variables are **multivariate Gaussian distributions**.

Definition 4.1.1

We say that X is Gaussian with **mean** $m \in \mathbb{R}^N$ and **covariance** C, with C a symmetric positive definite $N \times N$ matrix if X is absolutely continuous with density

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^N \det C}} e^{-\frac{1}{2}C^{-1}(x-m)\cdot(x-m)}, \ \forall x \in \mathbb{R}^N.$$

Remark 4.1.2

Since $C > 0$, it is invertible: indeed, C is injective because $Cx = 0$ implies $Cx \cdot x = 0$ which is possible iff $x = 0$. Since C is an $N \times N$ matrix, injectivity implies surjectivity, that is C is invertible. Thus, $\det C \neq 0$. Actually, since C is symmetric it is diagonalizable, that is there exists an orthogonal matrix T (that is, $T^{-1} = T^\top$ the transposed matrix of T) such that

$$TCT^\top = \operatorname{diag}(\sigma_1^2, \dots, \sigma_N^2) = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sigma_N^2 \end{bmatrix}$$

where the σ_j^2 are the eigenvalues of C, which are positive being $C > 0$. □

Example 4.1.3: $(**)$

Q. If $X \sim \mathcal{N}(m, C)$ is a multivariate Gaussian, then $\mathbb{E}[X] = m$ while the covariance matrix is C.

A. If $X \sim \mathcal{N}(m, C)$, we have

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X - m] + m = \int_{\mathbb{R}^N} (x-m)\frac{1}{\sqrt{(2\pi)^N \det C}} e^{-\frac{1}{2}C^{-1}(x-m)\cdot(x-m)}\, dx \\ &= m + \frac{1}{\sqrt{(2\pi)^N \det C}} \int_{\mathbb{R}^N} y e^{-\frac{1}{2}C^{-1}y\cdot y}\, dy = m\end{aligned}$$

being $y \longmapsto y e^{-\frac{1}{2}C^{-1}y\cdot y}$ even. We also notice that

$$\begin{aligned}\operatorname{Cov}(X_i, X_j) &= \mathbb{E}[(X_i - m_i)(X_j - m_j)] = \int_{\mathbb{R}^N} (x_i - m_i)(x_j - m_j)\frac{1}{\sqrt{(2\pi)^N \det C}} e^{-\frac{1}{2}C^{-1}(x-m)\cdot(x-m)}\, dx \\ &= \frac{1}{\sqrt{(2\pi)^N \det C}} \int_{\mathbb{R}^N} y_i y_j e^{-\frac{1}{2}C^{-1}y\cdot y}\, dy\end{aligned}$$

Since $C > 0$ is diagonalizable, $C = TDT^\top$ where $D = \mathrm{diag}(\sigma_1^2, \dots, \sigma_N^2)$, changing variable $u = T^\top y$, that is $y = Tu$, and since T orthogonal implies, in particular, that $|\det T| = 1$, we have

$$\begin{aligned}\mathrm{Cov}(X_i, X_j) &= \frac{1}{\sqrt{(2\pi)^N \det C}} \int_{\mathbb{R}^N} (Tu)_i (Tu)_j e^{-\frac{1}{2} C^{-1} Tu \cdot Tu}\, du \\ &= \frac{1}{\sqrt{(2\pi)^N \det C}} \int_{\mathbb{R}^N} (T_i \cdot u)(T_j \cdot u) e^{-\frac{1}{2} D^{-1} u \cdot u}\, du\end{aligned}$$

where T_i is the i−th line of the matrix T. Now,

$$\begin{aligned}\int_{\mathbb{R}^N} (T_i \cdot u)(T_j \cdot u) e^{-\frac{1}{2} D^{-1} u \cdot u}\, du &= \sum_{h,k} t_{ih} t_{jk} \int_{\mathbb{R}^N} u_h u_k e^{-\frac{1}{2} D^{-1} u \cdot u}\, du \\ &= \sum_{h,k} t_{ih} t_{jk} \prod_{m \neq h,k} \int_{\mathbb{R}} e^{-\frac{u_m^2}{2\sigma_m^2}}\, du_m \int_{\mathbb{R}^2} u_h u_k e^{-\frac{u_h^2}{2\sigma_h^2}} e^{-\frac{u_k^2}{2\sigma_k^2}}\, du_h du_k.\end{aligned}$$

Now,

$$\int_{\mathbb{R}} e^{-\frac{u_m^2}{2\sigma_m^2}}\, du_m = \sqrt{2\pi\sigma_m^2},$$

and, for $h \neq k$,

$$\int_{\mathbb{R}^2} u_h u_k e^{-\frac{u_h^2}{2\sigma_h^2}} e^{-\frac{u_k^2}{2\sigma_k^2}}\, du_h du_k = \int_{\mathbb{R}} u_h e^{-\frac{u_h^2}{2\sigma_h^2}}\, du_h \int_{\mathbb{R}} u_k e^{-\frac{u_k^2}{2\sigma_k^2}}\, du_k = 0,$$

while, for $h = k$,

$$\int_{\mathbb{R}^2} u_h^2 e^{-\frac{u_h^2}{2\sigma_h^2}}\, du_h = \sqrt{2\pi\sigma_h^2} \cdot \sigma_h^2.$$

Therefore, since $\det C = \det(TDT^\top) = \det(TT^\top) \det D = \det D$, we have

$$\mathrm{Cov}(X_i, X_j) = \frac{\sqrt{(2\pi)^N \det D}}{\sqrt{(2\pi)^N \det C}} \sum_k t_{ik} t_{jk} \sigma_k^2 = DT_i \cdot T_j = (T^\top DT)_{ij} = c_{ji} = c_{ij}.$$

4.2. Mapping multivariate random variables

Suppose X is a multivariate random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\Phi : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ be a map and $Y := \Phi(X)$. We notice that Y is a multivariate random variable iff

$$\{Y \in E\} = \{\Phi(X) \in E\} = \{X \in \Phi^{-1}(E)\} \in \mathcal{B}_{\mathbb{R}^N},\ \forall E \in \mathcal{B}_{\mathbb{R}^N}.$$

If X is a multivariate random variable, what we need is that Φ verifies the following definition:

Definition 4.2.1

A map $\Phi : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ is a **Borel map** if

$$\Phi^{-1}(E) \in \mathcal{B}_{\mathbb{R}^N},\quad \forall E \in \mathcal{B}_{\mathbb{R}^N}.$$

So, if Φ is a Borel map

$$\mu_Y(E) = \mu_X(\Phi^{-1}(E)), \ \forall E \in \mathscr{B}_{\mathbb{R}^N}.$$

When X is absolutely continuous, a natural question is whether $Y = \Phi(X)$ is also absolutely continuous and, in that case, what relation holds between the density of X and that of Y. Clearly, this is not true in general: if $\Phi(x) \equiv c$, then $Y = \Phi(X) \equiv c$, so $\mu_Y = \delta_c$, which does not admit a density. However, if Φ is regular enough, absolute continuity of Y may hold.

Definition 4.2.2

A map $\Phi = \Phi(x) : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ is a **diffeomorphism** if

i) Φ is a bijection;

ii) Both Φ, Φ^{-1} are differentiable with $\Phi', (\Phi^{-1})'$ continuous mappings.

Proposition 4.2.3

Let X be absolutely continuous, Φ be a diffeomeorphism on $\mathbb{R}^N$. Then, if $Y = \Phi(X)$, Y is absolutely continuous and

$$\boxed{f_Y(y) = f_X(\Phi^{-1}(y))|\det(\Phi^{-1})'(y)|}$$

Proof. Let $E \in \mathscr{B}_{\mathbb{R}^N}$. By the change of variable formula,

$$\mu_Y(E) = \mu_X(\Phi^{-1}(E)) = \int_{\Phi^{-1}(E)} f_X(x)\, dx \overset{y=\Phi(x),\ x=\Phi^{-1}(y)}{=} \int_E f_X(\Phi^{-1}(y))|\det(\Phi^{-1})'(y)|\, dy$$

from which the conclusion follows.

Example 4.2.4: $(**)$

Q. Let $X \sim \mathcal{N}(m, C)$, where $m \in \mathbb{R}^N$, and $C > 0$ is a symmetric $N \times N$ matrix. Show that there exists a matrix M such that $M^{-1}(X - m) \sim \mathcal{N}(0, \mathbb{I}_N)$.

A. Since $C > 0$ is symmetric, it can be diagonalized: there exists an orthogonal matrix T ($TT^\top = \mathbb{I}_N$) such that $TCT^\top = \mathrm{diag}(\sigma_1^2, \dots, \sigma_N^2) := D$. Then, the density of $Z = T(X - m) =: \Phi(X)$ ($\Phi^{-1}(z) = m + T^{-1}z = m + T^\top z$) is

$$f_Z(z) = \frac{1}{\sqrt{(2\pi)^N \det C}} e^{-\frac{1}{2}C^{-1}T^{-1}z\cdot T^{-1}z}|\det T^{-1}|,$$

and since T is orthogonal, $T^{-1} = T^\top$ and $\det(TT^\top) = det(\mathbb{I}_N) = 1$, from which $\det(T)^2 = 1$, that is $|\det T| = |\det T^{-1}| = 1$. Moreover,

$$C^{-1}T^{-1}z \cdot T^{-1}z = C^{-1}T^{-1}z \cdot T^\top z = TC^{-1}T^{-1}z \cdot z = (T^\top)^{-1}C^{-1}T^{-1}z \cdot z = (TCT^\top)^{-1}y \cdot z,$$

so

$$f_Z(z) = \frac{1}{\sqrt{(2\pi)^N \det C}} e^{-\frac{1}{2}D^{-1}z\cdot z} = \frac{1}{\sqrt{(2\pi)^N \det D}} e^{-\frac{1}{2}D^{-1}z\cdot z}$$

being $\det D = \det(TCT^\top) = \det(T^\top TC) = \det C$. Set now $\sqrt{D} := \mathrm{diag}(\sigma_1, \dots, \sigma_N)$, and set $z = \sqrt{D}y$, that is $y = (\sqrt{D})^{-1}z$. Then $Y = (\sqrt{D})^{-1}Z = (\sqrt{D})^{-1}T(X-m)$ has density

$$f_Y(y) = \frac{1}{\sqrt{(2\pi)^N \det D}} e^{-\frac{1}{2}D^{-1}\sqrt{D}y\cdot\sqrt{D}y} |\det\sqrt{D}|$$

Since D is diagonal as well as $D^{-1} = \mathrm{diag}\left(\frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_N^2}\right)$ and $\sqrt{D}$, we easily have

$$D^{-1}\sqrt{D}y \cdot \sqrt{D}y = \|y\|^2 = y_1^2 + \cdots + y_N^2,$$

and $\sqrt{\det D} = \det\sqrt{D}$, so

$$f_Y(y) = \frac{1}{\sqrt{(2\pi)^N}} e^{-\frac{1}{2}\|y\|^2}, \implies Y = (\sqrt{D})^{-1}T(X-m) \sim \mathscr{N}(0, \mathbb{I}_N).$$

4.3. Exercises

Exercise 4.3.1 ($**$)**.** *Let (X, Y) have density $f_{X,Y}(x, y) = 4xy1_{[0,1]^2}(x, y)$.*

i) *Determine the cdf $F_{X,Y}$.*
ii) *Compute $\mathbb{P}(X + Y < 1)$.*

Exercise 4.3.2 ($**$)**.** *Let (X, Y) have density $f_{X,Y}(x, y) = c(x^2 + \frac{xy}{2})1_{[0,1]\times[0,2]}(x, y)$.*

i) *Determine the value of the constant c in such a way that $f_{X,Y}$ be a probability density.*
ii) *Determine f_X.*
iii) *Compute $\mathbb{P}(X > Y)$.*

Exercise 4.3.3 ($**$)**.** *Let (X, Y) be a bivariate random variable with $f_{X,Y}(x, y) = e^{-(x+y)}1_{[0,+\infty[^2}(x, y)$. Determine the density of X/Y. (hint: start computing $F_{X/Y}$. . .)*

Exercise 4.3.4 ($**$)**.** *In a circular target with radius $R > 0$, the density of impact points (X, Y) is given by the formula*

$$f_{X,Y}(x, y) = c(R - \sqrt{x^2 + y^2})1_{B(0,R]}(x, y),$$

where $B(0, R] := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leqslant R^2\}$. Determine the value of c that makes f a probability density function and calculate the probability that the impact point falls in $B(0, a]$ with $a < R$. Additionally, compute the distribution of the distance from the center of the target and determine the mean distance of the impact point from the center.

Exercise 4.3.5 ($**$)**.** *Let (X, Y) be a bivariate random variable with density $f_{X,Y}(x, y) = \frac{1}{4}e^{-(x+y)/2}1_{[0,+\infty[^2}(x, y)$. Let $(Z, W) := (\frac{X-Y}{2}, Y)$. Determine $f_{Z,W}$, f_Z and f_W.*

Exercise 4.3.6 ($**$)**.** *Let (X, Y) be a bivariate random variable with joint density*

$$f_{XY}(x, y) = e^{-x-2|y|}1_{[0,+\infty[}(x).$$

i) *Check that f_{XY} is a true probability density.*
ii) *Define $Z := X^2 + Y$ and $W := 3X^2 - Y$. Show that (Z, W) is a.c. determining its density $f_{Z,W}$.*
iii) *Calculate $\mathbb{P}(Z + W \geqslant 0)$.*
iv) *Compute the marginal densities f_Z and f_W.*

Exercise 4.3.7 $(**)$**.** *Let (X, Y) be a random vector on $\mathbb{R}^2$ with joint density*

$$f_{X,Y}(x, y) := e^{-y} 1_{[0,1]}(x) 1_{[0,+\infty[}(y).$$

i) *Determine the joint density of $Z := XY$ and $W := \frac{X}{Y}$.*
ii) *Compute $\mathbb{P}(ZW > 1)$.*

Exercise 4.3.8 $(**)$**.** *A point (X, Y) is picked at random uniformly in the unit circle. This means that*

$$\mathbb{P}((X, Y) \in E) = \frac{1}{\pi} \lambda_2(E \cap B(0, 1]).$$

Find the joint density of (X, R) where $R = \sqrt{X^2 + Y^2}$.

Exercise 4.3.9. *Let*

$$f(x, y) := c e^{-\frac{x^2 - 2\rho x y + y^2}{2(1-\rho^2)}}, \ (x, y) \in \mathbb{R}^2.$$

i) *Determine the value of c in such a way f be a probability density. Is such f a Gaussian density?*
ii) *For the value c of i), let (X, Y) be such that $f_{X,Y} = f$. Determine the joint density of (X, Z) with $Z = \frac{Y - \rho X}{\sqrt{1-\rho^2}}$. Deduce the density of Z.*
iii) *Determine $\mathbb{P}(X > 0, \ Y > 0)$.*

5

Characteristic function

5.1. Fourier Transform of a Borel probability

A random variable X is characterized by its law, a probability measure on $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ (or $(\mathbb{R}^N, \mathscr{B}_{\mathbb{R}^N})$ for the multivariate case). All important quantities (probabilities, expectations) can be calculated in terms of the law of X and two random variable with same law identical form the probabilistic point of view. As a measure, the law of X is not an easy tool to handle. A more convenient tool is the cdf F_X or, even better, for absolutely continuous random variables its density f_X.

If X is a.c. with density f_X, being this an $L^1(\mathbb{R})$ function with $\int_{\mathbb{R}} f_X \, dx = 1$, its L^1 FT is well defined,

$$\widehat{f_X}(\xi) := \int_{\mathbb{R}} e^{-i\xi x} f_X(x) \, dx \equiv \int_{\mathbb{R}} e^{-i\xi x} \, d\mu_X(x).$$

This last integral makes sense whatever is X. This because, being μ_X a probability measure, $e^{-i\xi x}$ is an $L^1(\mathbb{R}, \mu_X)$ function:

$$\int_{\mathbb{R}} |e^{-i\xi x}| \, d\mu_X = \int_{\mathbb{R}} 1 \, d\mu_X = 1, \ \forall \xi \in \mathbb{R}.$$

This yields to the following extension of the FT to probability measures:

Definition 5.1.1

Let μ_X be a Borel-probability measure on $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$. We define FT of μ_X the function

$$\widehat{\mu_X}(\xi) := \int_{\mathbb{R}} e^{-i\xi x} \, d\mu_X(x), \ \xi \in \mathbb{R}.$$

Similarly, if μ_X is a Borel probability measure on $(\mathbb{R}^N, \mathscr{B}_{\mathbb{R}^N})$, we set

$$\widehat{\mu_X}(\xi) := \int_{\mathbb{R}^N} e^{-i\xi\cdot x} \, d\mu_X(x), \ \xi \in \mathbb{R}^N.$$

We notice that

$$\widehat{\mu_X}(\xi) = \mathbb{E}[e^{-i\xi X}],$$

for a random variable, and

$$\widehat{\mu_X}(\xi) = \mathbb{E}[e^{-i\xi\cdot X}],$$

for a multivariate random variable The function

$$\phi_X(\xi) := \mathbb{E}\left[e^{i\xi X}\right] = \widehat{\mu_X}(-\xi), \tag{5.1.1}$$

is called **characteristic function** of X.

Example 5.1.2: Gaussian distribution

If $X \sim \mathcal{N}(m, \sigma^2)$, $m \in \mathbb{R}$, $\sigma^2 > 0$, then

$$\phi_X(\xi) = e^{i\xi m - \frac{1}{2}\sigma^2\xi^2}, \ \forall \xi \in \mathbb{R}.$$

More in general, if $X \sim \mathcal{N}(m, C)$ with $m \in \mathbb{R}^N$ and $C > 0$ a symmetric matrix, then

$$\phi_X(\xi) = e^{i\xi\cdot m - \frac{1}{2}C\xi\cdot\xi}, \ \forall \xi \in \mathbb{R}^N.$$

Proof. We check the formula for the scalar case, the vector case being similar and left as exercise. We have

$$\begin{aligned}\phi_X(\xi) &= \int_{\mathbb{R}} e^{i\xi x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} \, dx \overset{y=x-m}{=} e^{i\xi m} \int_{\mathbb{R}} e^{-i\xi y} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} \, dy \\ &= e^{i\xi m} \widehat{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\sharp^2}{2\sigma^2}}}(-\xi) = e^{i\xi m} e^{-\frac{1}{2}\sigma^2(-\xi)^2} = e^{i\xi m - \frac{1}{2}\sigma^2\xi^2}. \quad \square\end{aligned}$$

Example 5.1.3: uniform distribution

Let $X \sim U(a, b)$. Then

$$\boxed{\phi_X(\xi) = \frac{e^{ib\xi} - e^{ia\xi}}{i\xi(b-a)}.} \quad \square$$

Proof. Here, $f_X(x) = \frac{1}{b-a} 1_{[a,b]}(x)$, so

$$\phi_X(\xi) = \frac{1}{b-a} \widehat{1_{[a,b]}}(-\xi).$$

Now, since $1_{[a,b]}(x) = 1_{[-\frac{b-a}{2}, \frac{b-a}{2}]}(x + \frac{a+b}{2})$ we have

$$\phi_X(\xi) = \frac{1}{b-a} \widehat{\text{rect}_{\,b-a/2}(\sharp + \frac{a+b}{2})}(-\xi) = e^{i\frac{a+b}{2}\xi} \frac{\sin\left(\frac{b-a}{2}\xi\right)}{\frac{b-a}{2}\xi},$$

and, by Euler formulas $\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}$, we get the conclusion.

Example 5.1.4: exponential

Let $X \sim \exp\lambda$. Then

$$\phi_X(\xi) = \frac{\lambda}{\lambda - i\xi}.$$

Proof. Here $f_X(x) = \lambda e^{-\lambda x} 1_{[0,+\infty[}(x)$, so

$$\phi_X(\xi) = \lambda \widehat{e^{-\lambda\sharp} 1_{[0,+\infty[}(\sharp)}(-\xi) = \lambda \int_0^{+\infty} e^{-\lambda x} e^{i\xi x} \, dx = \lambda \left[\frac{e^{(-\lambda + i\xi)x}}{-\lambda + i\xi} \right]_{x=0}^{x=+\infty} = \frac{\lambda}{\lambda - i\xi}.$$

In general, the characteristic function is continuous. This follows from the continuity of integrals depending on parameters because

- $\xi \longmapsto e^{i\xi x} \in \mathscr{C}(\mathbb{R})$, $\forall x \in \mathbb{R}$;
- $|e^{i\xi x}| = 1 \in L^1(\Omega)$, $\forall \xi \in \mathbb{R}$, $\forall x \in \mathbb{R}$.

Therefore $\phi_X(\xi) = \int_{\mathbb{R}} e^{i\xi x}\, d\mu_X(x) \in \mathscr{C}(\mathbb{R})$. Differently from usual properties of the FT, in general Riemann-Lebesgue's lemma does not hold.

Example 5.1.5: $(*)$

Let $X \sim x_0$ (constant random variable). Then $\mu_X = \delta_{x_0}$ and

$$\phi_X(\xi) = \int_{\mathbb{R}} e^{i\xi x}\, d\delta_{x_0}(x) = e^{i\xi x_0},$$

so in particular $|\phi_X(\xi)| \equiv 1$ so $\phi_X(\xi) \not\longrightarrow 0$ for $\xi \longrightarrow \pm\infty$.

Proposition 5.1.6

Let X be a random variable and assume that X has moment of order n, that is $\mathbb{E}[|X^n|] < +\infty$. Then $\phi_X \in \mathscr{C}^n(\mathbb{R})$ and

$$\partial_\xi^k \phi_X(\xi) = \mathbb{E}[(iX)^k e^{i\xi X}],\ \forall \xi \in \mathbb{R},\ k = 0, 1, \ldots, n.$$

In particular:

$$\partial_\xi^k \phi_X(0) = i^k \mathbb{E}[X^k],\ k = 0, 1, \ldots, n.$$

Proof. Notice that $X \in L^n(\Omega) \hookrightarrow L^k(\Omega)$ for every $k = 1, \ldots, n-1$. In particular, all moments $\mathbb{E}[X^k]$ of order k are finite for $k = 1, \ldots, n$. To compute the derivatives of ϕ_X, we apply the differentiation under integral sign theorem. We get

$$\partial_\xi^k \phi_X(\xi) = \mathbb{E}[(iX)^k e^{i\xi X}],$$

because $|(iX)^k e^{i\xi X}| = |X|^k \in L^1(\Omega)$ for every $\xi \in \mathbb{R}$. So, differentiation theorem applies and the conclusion follows.

In particular, if $\mathbb{V}[X] < +\infty$ then by the McLaurin formula we have

$$\phi_X(\xi) = \phi_X(0) + \partial_\xi \phi_X(0)\xi + \frac{1}{2}\partial_\xi^2 \phi_X(0)\xi^2 + o(\xi^2) = 1 + i\xi\mathbb{E}[X] - \frac{1}{2}\xi^2\mathbb{E}[X^2] + o(\xi^2).$$

5.2. Uniqueness of the characteristic function

The characteristic function *characterizes* uniquely a random variable X. This because

$$\phi_X = \phi_Y, \implies \mu_X = \mu_Y.$$

To show this is the goal of this section. Notice that, if X and Y are absolutely continuous, this is a consequence of injectivity of the L^1 FT: indeed

$$\phi_X \equiv \phi_Y, \iff \widehat{f_X}(-\sharp) \equiv \widehat{f_Y}(-\sharp), \iff \widehat{f_X} \equiv \widehat{f_Y}, \iff f_X = f_Y, \iff \mu_X = \mu_Y.$$

The general case is based on uniqueness for the FT of Borel probabilities. To show this we need a couple of auxiliary results. The first is the extension of the duality lemma:

Lemma 5.2.1: duality

$$\int_{\mathbb{R}} \widehat{\psi}\, d\mu = \int_{\mathbb{R}} \psi \widehat{\mu}\, d\xi,\ \forall \psi \in L^1(\mathbb{R}).$$

PROOF. First, notice that $\widehat{\psi} \in L^\infty(\mathbb{R})$, so $\widehat{\psi} \in L^1(\mathbb{R}, \mu)$ so $\int_{\mathbb{R}} \widehat{\psi}\, d\mu$ makes sense. We have

$$\int_{\mathbb{R}} \widehat{\psi}\, d\mu = \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-ix\xi} \psi(\xi)\, d\xi\, d\mu(x) \stackrel{Fubini}{=} \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-i\xi x}\, d\mu(x)\, \psi(\xi)\, d\xi = \int_{\mathbb{R}} \widehat{\mu}(\xi)\psi(\xi)\, d\xi.$$

So, if μ, ν are two Borel probabilities such that $\widehat{\mu} = \widehat{\nu}$, then

$$\int_{\mathbb{R}} \widehat{\psi}\, d\mu = \int_{\mathbb{R}} \widehat{\psi}\, d\nu,\ \forall \psi \in L^1(\mathbb{R}). \tag{5.2.1}$$

Now, if we could apply the previous identity with $\widehat{\psi} = 1_{[a,b]}$, we would have that $\mu([a,b]) = \nu([a,b])$ for every $[a,b]$, then easily for every intervals, whence $\mu = \nu$. Unfortunately, $1_{[a,b]}$ cannot be a FT, because it is discontinuous. The next proof shows how to circumvent this issue.

Theorem 5.2.2: uniqueness

If $\widehat{\mu} = \widehat{\nu}$ then $\mu = \nu$.

PROOF. Let $\varphi \in \mathcal{S}(\mathbb{R}) \subset L^1(\mathbb{R})$ (Schwarz's space). Then $\widehat{\varphi} \in \mathcal{S}(\mathbb{R}) \subset L^1(\mathbb{R})$, so, in particular, inversion formula applies, and

$$\varphi(x) = \frac{1}{2\pi} \widehat{\widehat{\varphi}}(-x) \implies \varphi = \widehat{\psi}.$$

Therefore, by identity (5.2.1), we get

$$\int_{\mathbb{R}} \varphi\, d\mu = \int_{\mathbb{R}} \varphi\, d\nu,\ \forall \varphi \in \mathcal{S}(\mathbb{R}). \tag{5.2.2}$$

We not build an $\mathcal{S}(\mathbb{R})$ approximation of $1_{[a,b]}$. We start with a particular case: define

$$\delta_n(x) := \begin{cases} 1, & -1 \leqslant x \leqslant 1, \\ \exp\left(\frac{1}{(1+\frac{1}{n})^2-1} - \frac{1}{(1+\frac{1}{n})^2 - x^2}\right), & 1 \leqslant |x| \leqslant 1 + \frac{1}{n}, \\ 0, & |x| \geqslant 1 + \frac{1}{n}. \end{cases}$$

We can check that

- $\delta_n \in \mathscr{C}^\infty(\mathbb{R})$
- $\delta_n \equiv 1$ on $[-1,1]$ and $\delta_n \equiv 0$ off $[-1-\frac{1}{n}, 1+\frac{1}{n}]$
- $0 \leqslant \delta_n(x) \leqslant 1$, $\forall x \in \mathbb{R}$,
- $\delta_n(x) \xrightarrow{pw} 1_{[-1,1]}(x)$

.In particular, $\delta_n \in \mathscr{S}(\mathbb{R})$. Define now

$$\delta_{n,[a,b]}(x) = \delta_n\left(\frac{2x-(a+b)}{b-a}\right).$$

Then $\delta_{n,[a,b]} \in \mathscr{S}(\mathbb{R})$, $0 \leqslant \delta_{n,[a,b]}(x) \leqslant 1$ and $\delta_{n,[a,b]}(x) \stackrel{pw}{\longrightarrow} 1_{[a,b]}(x)$. Therefore, from (5.2.2) we have

$$\int_{\mathbb{R}} \delta_{n,[a,b]}\, d\mu = \int_{\mathbb{R}} \delta_{n,[a,b]}\, d\nu,$$

and by dominated convergence we get

$$\int_{\mathbb{R}} 1_{[a,b]}\, d\mu = \int_{\mathbb{R}} 1_{[a,b]}\, d\nu, \quad \Longleftrightarrow \quad \mu([a,b]) = \nu([a,b]),$$

this for every $[a,b]$. The conclusion now follows.

So, for example,

$$X \sim \mathscr{N}(m,\sigma^2), \quad \Longleftrightarrow \quad \phi_X(\xi) = e^{i\xi m - \frac{1}{2}\sigma^2\xi^2}.$$

In certain circumstances, this is an important characterization that simplifies calculations.

Example 5.2.3: $(**)$

Q. Let $X \sim \mathscr{N}(m,C)$ be a multivariate Gaussian, $m \in \mathbb{R}^N$ and C symmetric and positive definite covariance matrix. Use the characteristic function to determine the distribution of $a \cdot X$, where $a \in \mathbb{R}^N$.

A. We have

$$\phi_{a\cdot X}(\xi) = \mathbb{E}\left[e^{i\xi a\cdot X}\right] = \mathbb{E}\left[e^{i(\xi a)\cdot X}\right] = e^{i(\xi a)\cdot m - \frac{1}{2}C(\xi a)\cdot(\xi a)} = e^{i\xi(a\cdot m) - \frac{1}{2}(Ca\cdot a)\xi^2},$$

from which we deduce that $a \cdot X \sim \mathscr{N}(a \cdot m, Ca \cdot a)$.

5.3. Exercises

Exercise 5.3.1 $(*)$**.** *Compute the characteristic functions of a Bernoulli, binomial and Poisson random variable*

Exercise 5.3.2 $(**)$**.** *Determine the characteristic function of a Cauchy random variable X, that is, such that $f_X(x) = \frac{1}{\pi}\frac{a}{a^2+(x-m)^2}$*

Exercise 5.3.3 $(**)$**.** *Determine the characteristic function of a Gamma random variable X, that is, such that $f_X(x) = \frac{\lambda^{\alpha-1}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}1_{[0,+\infty[}(x)$ (here $\alpha > 0$ and $\lambda > 0$).*

Exercise 5.3.4 $(**)$**.** *For each of the following functions say if i) they are characteristic functions of some random variable X, and (if yes), ii) what is the distribution of X.*

- $\phi(\xi) = (1-|\xi|)1_{[-1,1]}(\xi)$.
- $\phi(\xi) = \sin\xi$.
- $\phi(\xi) = \cos\xi$.
- $\phi(\xi) = \frac{1}{1+\xi^2}$.

- $\phi(\xi) = 1 - \sin\xi$.

Exercise 5.3.5 ($**$). *Let X, Y be absolutely continuous random variables. Prove the identity*

$$\int_{\mathbb{R}} \phi_X(\xi) f_Y(\xi) e^{-i\xi y}\, d\xi = \int_{\mathbb{R}} \phi_Y(x-y) f_X(x)\, dx.$$

Exercise 5.3.6 ($**+$). *Let X, Y be random variables on $(\Omega, \mathscr{F}, \mathbb{P})$ for which*

$$\mathbb{E}\left[e^{i\xi X} Y\right] = 0,\ \forall \xi \in \mathbb{R}.$$

Define $\mu(E) = \mathbb{E}[Y 1_E]$ for $E \in \mathscr{B}_{\mathbb{R}}$. Check that $\widehat{\mu} = 0$. Deduce that $Y = 0$ a.s.

Exercise 5.3.7 ($**+$). *Let μ be a probability on $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$. Then:*

i) $|\widehat{\mu}(\xi)| \leqslant 1 = \widehat{\mu}(0)$, *for every $\xi \in \mathbb{R}$.*

ii) $\widehat{\mu}(-\xi) = \overline{\widehat{\mu}(\xi)}$.

iii) $\widehat{\mu}$ *is positive definite in the sense that*

$$\sum_{j,k} \widehat{\mu}(\xi_j - \xi_k) z_j \overline{z_k} \geqslant 0,\quad \forall \xi_1, \ldots, \xi_n \in \mathbb{R},\ \forall z_1, \ldots, z_n \in \mathbb{C}.$$

iv) $\widehat{\mu} \in \mathscr{C}(\mathbb{R})$.

Exercise 5.3.8 ($**+$). *Let ϕ_X be the characteristic function of an absolutely continuous random variable X. Show that $|\phi_X|^2$ is still a characteristic function of a random variable Y, determining also its density f_Y.*

Exercise 5.3.9 ($***$). *Let $\widehat{\mu_X} \in L^1(\mathbb{R})$. Show that $d\mu_X = f_X(x)\, dx$ for some $f_X \in L^1(\mathbb{R})$, $f_X \geqslant 0$ and $\int_{\mathbb{R}} f_X(x)\, dx = 1$. (hint: use the duality Lemma with $\psi \in \mathcal{S}(\mathbb{R})$. . .)*

6

Independence

6.1. Independent Events

Independence is a key concept of Probability. Independence is a concept ranging from events, to $\sigma-$algebras to random variable, for a finite number of objects to infinitely many. We start by the simplest of the definitions: independence of events.

Definition 6.1.1

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Two events $E, F \in \mathscr{F}$ are said to be **independent** if

$$\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F).$$

More in general, n events $E_1, \ldots, E_n \in \mathscr{F}$ are independent iff

$$\mathbb{P}(E_1 \cap \cdots \cap E_n) = \mathbb{P}(E_1) \cdots \mathbb{P}(E_n).$$

Warning 6.1.2

Events might be pairwise independent but not jointly independent.

Proof. Consider the probability space of a rolling of two dices: $\Omega\{(i, j) \ : \ i, j \in \{1 \ldots, 6\}\}$, $\mathscr{F} = \mathscr{P}(\Omega)$, $p_{ij} = \frac{1}{36}$. Take the events $E := \{\text{first roll is } 1\}$, $F := \{\text{second roll is } 6\}$ and $G := \{\text{sum of rolls is } 7\}$. Notice that

$$\mathbb{P}(E) = \frac{1}{6}, \ \mathbb{P}(F) = \frac{1}{6}, \ \mathbb{P}(G) = \frac{6}{36} = \frac{1}{6}.$$

Clearly, $E \cap F = \{(1, 6)\}$ so $\mathbb{P}(E \cap F) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}(E)\mathbb{P}(F)$. Moreover, $E \cap G = \{(1, 6)\}$ so, again $\mathbb{P}(E \cap G) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}(E)\mathbb{P}(G)$ and, in the same manner $\mathbb{P}(F \cap G) = \mathbb{P}(F)\mathbb{P}(G)$. However, $E \cap F \cap G = \{(1, 6)\}$ so

$$\mathbb{P}(E \cap F \cap G) = \frac{1}{36} \neq \frac{1}{216} = \mathbb{P}(E)\mathbb{P}(F)\mathbb{P}(G). \quad \square$$

Remark 6.1.3

The previous example also shows that an event might be independent of two others, but not of their intersection.

PROOF. In notations of the previous example, E is independent of F and G. Notice that $\mathbb{P}(E \cap (F \cap G)) = \frac{1}{36}$ while, being $F \cap G = \{(1,6)\}$ so $\mathbb{P}(F \cap G) = \frac{1}{36}$ from which

$$\mathbb{P}(E \cap (F \cap G)) \neq \mathbb{P}(E) \cap \mathbb{P}(F \cap G). \quad \square$$

We now extend independence to $\sigma-$algebras.

Definition 6.1.4

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. We say that two sub $\sigma-$algebras $\mathscr{G}_1$ and $\mathscr{G}_2$ of $\mathscr{F}$ are **independent** if E_1 and E_2 are independent for every $E_1 \in \mathscr{G}_1$ and $E_2 \in \mathscr{G}_2$.

More in general, given a family of $\sigma-$algebras $(\mathscr{G}_j)_{j \in J} \subset \mathscr{F}$ we say that they are independent if

$$E_1, E_2, \ldots, E_n \text{ are independent } \forall E_1 \in \mathscr{G}_{j_1}, \ldots, \forall E_n \in \mathscr{G}_{j_n}, \ \forall j_1, \ldots, j_n \in J, \ \forall n \in \mathbb{N}.$$

In general, it is very difficult to characterize all the events of a $\sigma-$algebra. Fortunately, to check the independence of two $\sigma-$algebras it is sufficient to check that some generator families are independent:

Proposition 6.1.5

Let $\mathscr{G}_j := \sigma(\mathscr{A}_j)$ be the $\sigma-$algebra generated by a **multiplicative class** $\mathscr{A}_j \subset \mathscr{F}$, $j = 1, 2$ (that is, if $A, B \in \mathscr{A}_j$ then also $A \cap B \in \mathscr{A}_j$). The following facts are equivalent:

i) $\mathscr{G}_1$ and $\mathscr{G}_2$ are independent.
ii) E and F are independent, $\forall E \in \mathscr{A}_1, \forall F \in \mathscr{A}_2$.

6.2. Independent random variable

Independence extends in a natural way to random variable. Here, for simplicity we refer to the case of random variables, the definitions and properties for the multivariate case are similar.

Definition 6.2.1

Let $X, Y \in L(\Omega)$ be two random variables. We say that X and Y are **independent** if

$$\mathbb{P}(X \in E, \ Y \in F) = \mathbb{P}(X \in E)\mathbb{P}(Y \in F), \ \forall E, F \in \mathscr{B}_{\mathbb{R}}.$$

We call $\sigma-$algebra generated by a random variable X (sometimes also called **information generated by** X) the family

$$\sigma(X) := \{\{X \in E\} \ : \ E \in \mathscr{B}_{\mathbb{R}}\}.$$

It is easy to check that this is a $\sigma-$algebra (exercise). This $\sigma-$algebra represents the minimal family of events such that X is measurable.

Proposition 6.2.2

$$X, Y \text{ independent} \iff \sigma(X),\ \sigma(Y) \text{ independent.}$$

An extension of the definition if provided by the following

Proposition 6.2.3

Let $X, Y \in L(\Omega)$. Then, X and Y are independent if and only if

$$\mathbb{E}[\varphi(X)\psi(Y)] = \mathbb{E}[\varphi(X)]\mathbb{E}[\psi(Y)],\ \ \forall \varphi \in L^1(\mathbb{R}, \mu_X),\ \forall \psi \in L^1(\mathbb{R}, \mu_Y). \tag{6.2.1}$$

In particular: if $X, Y \in L^1(\Omega)$ are independent, then also $XY \in L^1(\Omega)$ and

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

PROOF. $\Longleftarrow$ X and Y are independent iff

$$\mathbb{P}(X \in E,\ Y \in F) = \mathbb{P}(X \in E)\mathbb{P}(Y \in F),\ \forall E, F \in \mathscr{B}_{\mathbb{R}}.$$

Now,

$$\mathbb{P}(X \in E,\ Y \in F) = \mathbb{P}((X, Y) \in E \times F) = \mathbb{E}[1_{E\times F}(X, Y)] = \mathbb{E}[1_E(X)1_F(Y)],$$

while,

$$\mathbb{P}(X \in E)\mathbb{P}(Y \in F) = \mathbb{E}[1_E(X)]\mathbb{E}[1_F(Y)].$$

So independence follows by the identity (6.2.1) taking $\varphi = 1_E$ and $\psi = 1_F$ we get In particular.

$\Longrightarrow$. The first part shows that independence is equivalent to (6.2.1) for $\varphi = 1_E$, $\psi = 1_F$, $E, F \in \mathscr{B}_{\mathbb{R}}$. By linearity we extend this to simple functions $s(X) = \sum_{j=1}^N 1_{E_j}(X)$ and similarly for $\widetilde{s}(Y)$. If now, φ, ψ are two positive Borel-measurable functions, there exist sequences (s_n), $(\widetilde{s}_n)$ of simple functions such that $s_n \uparrow \varphi$ and $\widetilde{s}_n \uparrow \psi$ point-wise everywhere. By monotone convergence, then,

$$\mathbb{E}[\varphi(X)\psi(Y)] \longleftarrow \mathbb{E}[s_n(X)\widetilde{s}_n(Y)] = \mathbb{E}[s_n(X)]\mathbb{E}[\widetilde{s}_n(Y)] \longrightarrow \mathbb{E}[\varphi(X)]\mathbb{E}[\psi(Y)].$$

Thus (6.2.1) now holds for $\varphi \in L_+(\mathbb{R}, \mu_X)$ and $\psi \in L_+(\mathbb{R}, \mu_Y)$. Finally, let $\varphi \in L^1(\mathbb{R}, \mu_X)$ and $\psi \in L^1(\mathbb{R}, \mu_Y)$. Writing $\varphi = \varphi_+ - \varphi_-$ and doing the same for ψ, we have

$$\varphi\psi = (\varphi_+ - \varphi_-)(\psi_+ - \psi_-) = \underbrace{(\varphi_+\psi_+ + \varphi_-\psi_-)}_{=(\varphi\psi)_+} - \underbrace{(\varphi_+\psi_- + \varphi_-\psi_+)}_{=(\varphi\psi)_-}$$

Therefore,

$$\mathbb{E}[(\varphi(X)\psi(Y))_+] = \mathbb{E}[\varphi_+(X)\psi_+(Y) + \varphi_-(X)\psi_-(Y)] = \mathbb{E}[\varphi_+(X)]\mathbb{E}[\psi_+(Y)] + \mathbb{E}[\varphi_-(X)]\mathbb{E}[\psi_-(Y)] < +\infty,$$

and, similarly,

$$\mathbb{E}[(\varphi(X)\psi(Y))_-] < +\infty,$$

from which we conclude that $\varphi(X)\psi(Y) \in L^1(\Omega)$ and, easily, formula (6.2.1) holds.

Remark 6.2.4

In particular, if X and Y are independent, then also $\varphi(X)$ and $\psi(Y)$ are independent, for any Borel functions φ, ψ.

Independence of random variables reflects on their cdf and densities (if any).

Proposition 6.2.5

Let $X, Y \in L(\Omega)$. The following properties are equivalent:

i) X and Y are independent.
ii) $F_{X,Y}(x,y) = F_X(x)F_Y(y)$, $\forall x, y \in \mathbb{R}$.
iii) If (X,Y) is absolutely continuous, then $f_{X,Y} = f_X f_Y$ a.e.

Proof. i) $\Longrightarrow$ ii). If X, Y are independent, then

$$F_{X,Y}(x,y) = \mathbb{P}(X \leqslant x,\ Y \leqslant y) = \mathbb{P}(X \leqslant x)\mathbb{P}(Y \leqslant y) = F_X(x)F_Y(y),\ \forall x, y \in \mathbb{R}.$$

ii) $\Longrightarrow$ i). Assume $F_{XY} \equiv F_X F_Y$. Consider a rectangle $R :=]a,b]\times]c,d]$. Since

$$R =]-\infty, b]\times]-\infty, d] \setminus \left(\underbrace{]-\infty, a]\times]-\infty, d]}_{=:R_1} \sqcup \underbrace{]a,b]\times]-\infty, c]}_{=:R_2} \right),$$

from which

$$\begin{aligned}
\mathbb{P}((X,Y) \in R) &= \mathbb{P}(X \leqslant b,\ Y \leqslant d) - \mathbb{P}(X \leqslant a,\ Y \leqslant d) - \underbrace{\mathbb{P}(a < X \leqslant b,\ Y \leqslant c)}_{=\mathbb{P}(X \leqslant b,\ Y \leqslant c) - \mathbb{P}(X \leqslant a,\ Y \leqslant c)} \\
&= F_{XY}(b,d) - F_{XY}(a,d) - (F_{XY}(b,c) - F_{XY}(a,c)) \\
&= F_X(b)F_Y(d) - F_X(a)F_Y(d) - (F_X(b)F_Y(c) - F_X(a)F_Y(c)) \\
&= \underbrace{(F_X(b) - F_X(a))}_{=\mathbb{P}(a < X \leqslant b)} F_Y(d) - (F_X(b) - F_X(a))\, F_Y(c) \\
&= \mathbb{P}(X \in]a,b])\mathbb{P}(Y \in]c,d]).
\end{aligned}$$

So, if $\mathscr{R}_X := \{X \in]a,b] \ :\ a \leqslant b\}$ and $\mathscr{R}_Y := \{Y \in]c,d] \ :\ c \leqslant d\}$, then $\mathscr{R}_X$ and $\mathscr{R}_Y$ are independent algebras of sets. Therefore $\sigma(\mathscr{R}_X) = \sigma(X)$ and $\sigma(\mathscr{R}_Y) = \sigma(Y)$ are independent, which is the conclusion.

iii) $\Longleftrightarrow$ i). Assume (X,Y) a.c., and let $f_{X,Y}$ be its density. Then

$$\mathbb{P}(X \in E,\ Y \in F) = \int_{E\times F} f_{X,Y}(x,y)\, dxdy$$

On the other hand,

$$\mathbb{P}(X \in E,\ Y \in F) = \mathbb{P}(X \in E)\mathbb{P}(Y \in F) = \int_E f_X(x)\, dx \int_F f_Y(y)\, dy \overset{Fub.}{=} \int_{E\times F} f_X(x)f_Y(y)\, dxdy.$$

So X, Y are independent iff

$$\int_{E\times F} f_{X,Y}(x,y)\, dxdy = \int_{E\times F} f_X(x)f_Y(y)\, dxdy, \iff \int_{E\times F} (f_{X,Y} - f_X f_Y) = 0,\ \forall E, F \in \mathscr{B}_{\mathbb{R}}.$$

Now, since $\sigma(\mathscr{B}_{\mathbb{R}} \times \mathscr{B}_{\mathbb{R}}) = \mathscr{B}_{\mathbb{R}^2}$ the previous relation holds for every Borel set of $\mathbb{R}^2$. Therefore, $f_{XY} = f_X f_Y$ a.e. (w.r.t. the Lebesgue measure).

Example 6.2.6

Q. Let X and Y be independent random variables with densities given by

$$f_X(x) := \frac{1}{\pi\sqrt{1-x^2}}\chi_{]-1,1[}(x), \quad f_Y(y) := \frac{y}{\sigma}e^{-\frac{y^2}{2\sigma}}\chi_{[0,+\infty[}(y).$$

Show that $XY \sim \mathcal{N}(0,\sigma)$.

A. First, we note that since $f_Y \equiv 0$ for $y \leqslant 0$, it follows that $\mathbb{P}(Y \leqslant 0) = 0$, or equivalently $\mathbb{P}(Y > 0) = 1$. Thus, letting $Z = XY$ and denoting the c.d.f. of Z by F_Z, we have:

$$F_Z(z) = \mathbb{P}(XY \leqslant z) = \mathbb{P}\left(X \leqslant \frac{z}{Y}\right) = \mathbb{P}\left((X,Y) \in \{(x,y) : x \leqslant \frac{z}{y}\}\right) = \int_{x\leqslant\frac{z}{y}} f_{X,Y}(x,y)\,dxdy,$$

By the independence of X and Y, we have $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, so

$$F_Z(z) = \int_0^{+\infty}\left(\int_{-\infty}^{z/y} f_X(x)f_Y(y)\,dx\right)dy = \int_0^{+\infty} f_Y(y)\int_{-\infty}^{z/y} f_X(x)\,dx\,dy.$$

Therefore,

$$f_Z(z) = F_Z'(z) = \int_0^{+\infty} f_Y(y)\frac{1}{y}f_X\left(\frac{z}{y}\right)dy.$$

In our case,

$$f_Z(z) = \int_0^{+\infty}\frac{y}{\sigma}e^{-\frac{y^2}{2\sigma}}\frac{1}{y}\frac{1}{\pi}\frac{1}{\sqrt{1-\frac{z^2}{y^2}}}\chi_{]-1,1[}\left(\frac{z}{y}\right)dy = \frac{1}{\pi\sigma}\int_0^{+\infty}e^{-\frac{y^2}{2\sigma}}\frac{1}{\sqrt{1-\frac{z^2}{y^2}}}\chi_{]-1,1[}\left(\frac{z}{y}\right)dy.$$

Now, note that $\chi_{]-1,1[}\left(\frac{z}{y}\right) = 1$ if and only if $\left|\frac{z}{y}\right| = \frac{|z|}{y} < 1$, which implies $y > |z|$; otherwise, it is 0. Thus,

$$f_Z(z) = \frac{1}{\pi\sigma}\int_{|z|}^{+\infty}e^{-\frac{y^2}{2\sigma}}\frac{1}{\sqrt{1-\frac{z^2}{y^2}}}dy = \frac{1}{\pi\sigma}\int_{|z|}^{+\infty}e^{-\frac{y^2}{2\sigma}}\frac{y}{\sqrt{y^2-z^2}}dy.$$

Setting $w = \sqrt{y^2-z^2}$, so that $dw = \frac{y}{\sqrt{y^2-z^2}}dy$, we have

$$f_Z(z) = \frac{1}{2\pi\sigma}\int_0^{+\infty}e^{-\frac{w^2+z^2}{2\sigma}}dw = \frac{1}{\pi\sigma}e^{-\frac{z^2}{2\sigma}}\int_0^{+\infty}e^{-\frac{w^2}{2\sigma}}dw = \frac{1}{\pi\sigma}e^{-\frac{z^2}{2\sigma}}\cdot\frac{1}{2}\int_{\mathbb{R}}e^{-\frac{w^2}{2\sigma}}dw.$$

The integral evaluates to $\sqrt{2\pi\sigma}$, so

$$f_Z(z) = \frac{1}{2\pi\sigma}e^{-\frac{z^2}{2\sigma}}\sqrt{2\pi\sigma} = \frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{z^2}{2\sigma}},$$

which is the conclusion. □

Independence of random variables reflects also on their characteristic functions.

Proposition 6.2.7: Kač theorem

Let $X, Y \in L(\Omega)$. The following properties are equivalent:

- X and Y are independent.

- $\phi_{X,Y}(\xi,\eta) = \phi_X(\xi)\phi_Y(\eta)$, $\forall \xi,\eta \in \mathbb{R}^2$.

PROOF. i) $\Longrightarrow$ ii). Let X,Y be independent. Then

$$\phi_{X,Y}(\xi,\eta) = \mathbb{E}[e^{i(\xi,\eta)\cdot(X,Y)}] = \mathbb{E}[e^{i\xi X+i\eta Y}] = \mathbb{E}[e^{i\xi X}e^{i\eta Y}] \overset{(6.2.1)}{=} \mathbb{E}[e^{i\xi X}]\mathbb{E}[e^{i\eta Y}] = \phi_X(\xi)\phi_Y(\eta).$$

ii) $\Longrightarrow$ i). Define the function

$$\nu(E\times F) = \mu_X(E)\mu_Y(F),\ \forall E,F \in \mathscr{B}_{\mathbb{R}}.$$

This ν is well defined on the product class $\mathscr{P} := \{E\times F\ :\ E,F\in\mathscr{B}_{\mathbb{R}}\} \subset \mathscr{B}_{\mathbb{R}^2}$ which is not a $\sigma-$algebra, nonetheless it contains rectangles $I\times J$ with I and J intervals. Therefore, $\sigma(\mathscr{P}) = \mathscr{B}_{\mathbb{R}^2}$. It is not difficult to check that ν is a pre-probability, so by Caratheodory's extension theorem, ν extends to a probability measure. By its definition, it is clear that

$$\int_{\mathbb{R}^2} \varphi(x)\psi(y)\,d\nu(x,y) = \int_{\mathbb{R}} \varphi(x)\,d\mu_X(x)\int_{\mathbb{R}}\psi(y)\,d\mu_Y(y).$$

From this,

$$\widehat{\nu}(\xi,\eta) = \int_{\mathbb{R}^2} e^{-i(\xi,\eta)\cdot(x,y)}\,d\nu(x,y) = \int_{\mathbb{R}} e^{-i\xi x}\,d\mu_X(x)\int_{\mathbb{R}} e^{-i\eta y}\,d\mu_Y(y) = \phi_X(x)\phi_Y(y).$$

Since $\widehat{\nu} = \phi_X\phi_Y = \phi_{X,Y} = \widehat{\mu}_{X,Y}$, by the uniqueness of FT of Borel measures (thm 5.2) we conclude that $\mu_{X,Y} = \nu$. In particular,

$$\mu_{X,Y}(E\times F) = \nu(E\times F) = \mu_X(E)\mu_Y(F),$$

that is

$$\mathbb{P}(X\in E,\ Y\in F) = \mathbb{P}(X\in E)\mathbb{P}(Y\in F),$$

which is the independence of X and Y.

Here is a nice (and important) application of the characterization of independent r.vs.

Proposition 6.2.8

Let X,Y be absolutely continuous, independent random variables with densities f_X and f_Y. Then $X+Y$ is absolutely continuous with density

$$f_{X+Y} = f_X * f_Y.$$

PROOF. We use the characteristic function: we have

$$\begin{aligned}\phi_{X+Y}(\xi) &= \mathbb{E}[e^{i\xi(X+Y)}] = \mathbb{E}[e^{i\xi X}e^{i\xi Y}] = \mathbb{E}[e^{i\xi X}]\mathbb{E}[e^{i\xi Y}] = \phi_X(\xi)\phi_Y(\xi)\\ &= \widehat{f_X}(-\xi)\widehat{f_Y}(-\xi) = \widehat{f_X * f_Y}(-\xi).\end{aligned}$$

Now, by the injectivity of FT, we conclude that

$$f_{X+Y} = f_X * f_Y.$$

These properties extend in a straightforward way to the case of any finite number of random variables.

6.3. i.i.d.

Modeling a random experiment repeated infinitely many times, we need to be able to work with *infinitely many independent random variables, all with the same distribution*.

Definition 6.3.1

Let $(X_n)_{n\in\mathbb{N}} \subset L(\Omega)$. We say that the (X_n) are **independent, identically distributed** if

- $(\sigma(X_n))$ is a family of independent $\sigma-$algebras
- $F_{X_n} \equiv F_{X_m}$ for all $n, m \in \mathbb{N}$.

We use the shortening (X_n) **i.i.d. random variables**.

We have seen that it is always possible to build a random variable with an assigned cdf (proposition 3.1). If, for example F is continuous, we take

$$(\Omega, \mathscr{F}, \mathbb{P}) = ([0,1], \mathscr{B}_{\mathbb{R}}, \lambda_1), \quad X(\omega) := F^{-1}(\omega), \quad \Longrightarrow \quad F_X \equiv F.$$

Extending this idea, we can build any finite number of i.i.d. random variables $X_1, \ldots, X_N$ all with a given cdf F. For example, if F are continuous, we take

$$\boxed{(\Omega, \mathscr{F}, \mathbb{P}) = ([0,1]^N, \mathscr{B}_{\mathbb{R}^N}, \lambda_N), \quad X_j(\omega_1, \ldots, \omega_N) = F^{-1}(\omega_j).}$$

The X_j are i.i.d. random variables. Indeed:

$$\mathbb{P}(X_j \leqslant x) = \lambda_N\left(\left\{\omega \in [0,1]^N \ : \ F^{-1}(\omega_j) \leqslant x\right\}\right) = \lambda_N\left(\left\{\omega \in [0,1]^N \ : \ \omega_j \leqslant F(x)\right\}\right) = F(x),$$

so $F_{X_j} \equiv F$, so the X_j have the same cdf F. Moreover, if $1 \leqslant i_1 < i_2 < \ldots < i_n \leqslant N$,

$$\begin{aligned}
F_{X_{i_1},\ldots,X_{i_n}}(x_{i_1}, \ldots, x_{i_n}) &= \mathbb{P}\left(X_{i_1} \leqslant x_{i_1}, \ldots, X_{i_n} \leqslant x_{i_n}\right) \\
&\leqslant \lambda_N\left(\left\{(\omega_1, \ldots, \omega_N) \in [0,1]^N \ : \ F^{-1}(\omega_{i_1}) \leqslant x_{i_1}, \ldots, F^{-1}(\omega_{i_n}) \leqslant x_{i_n}\right\}\right) \\
&\leqslant \lambda_N\left(\left\{(\omega_1, \ldots, \omega_N) \in [0,1]^N \ : \ \omega_{i_1} \leqslant F(x_{i_1}), \ldots, \omega_{i_n} \leqslant F(x_{i_n})\right\}\right) \\
&= F(x_{i_1}) \cdots F(x_{i_n} = F_{X_{i_1}}(x_{i_1}) \cdots F_{X_{i_n}}(x_{i_n}),
\end{aligned}$$

that is $X_{i_1}, \ldots, X_{i_n}$ are also independent.

This construction becomes complicate when we set $N = +\infty$. This because, we do not have an infinite dimensional version of the Lebesgue's measure. The proof of the existence of infinitely many i.i.d. random variables is more sophisticated.

Theorem 6.3.2

Given a cdf F, there exists a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a sequence $(X_n)_{n\in\mathbb{N}}$ of i.i.d. random variables such that $F_{X_n} \equiv F$ for every $n \in \mathbb{N}$.

Proof. We divide the proof in two steps. The first step proves the conclusion assuming F being the cdf of a uniform distribution on the interval $[0,1]$. In the second step we will remove this restriction.

.**First step.** Let $(\Omega, \mathscr{F}, \mathbb{P}) = ([0,1], \mathscr{B}_{\mathbb{R}}, \lambda_1)$. For $x \in [0,1]$, its binary expansion is uniquely defined as

$$x = \sum_{k=1}^{\infty} \frac{C_k(x)}{2^k},$$

choosing, by convention, the representation that eventually consists of the digit 1. We look at $C_k : [0,1] \longrightarrow \{0,1\} \subset \mathbb{R}$ as random variables. Notice that $\{C_k = 0\}$ and $\{C_k = 1\}$ are unions of 2^{k-1} intervals of length $\frac{1}{2^k}$ each, so they are Borel sets and C_k is a random variable for every k. They are also independent. Indeed, if $k < j$,

$$\{C_k = a,\ C_j = b\}$$

where $a, b \in \{0,1\}$ is made of $2^{j-1}/2 = 2^{j-2}$ intervals each of length $\frac{1}{2^j}$ so,

$$\mathbb{P}(C_k = a,\ C_j = b) = \lambda_1(C_k = a,\ C_j = b) = 2^{j-2}\frac{1}{2^j} = \frac{1}{4}$$

while

$$\mathbb{P}(C_k = a)\mathbb{P}(C_j = b) = \lambda_1(C_k = a)\lambda_1(C_j = b) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Let now $r : \mathbb{N}^2 \longrightarrow \mathbb{N}$ be a bijection, and define

$$C_{n,k} := C_{r(n,k)} \quad \text{and} \quad U_n := \sum_{k=1}^{\infty} \frac{C_{n,k}}{2^k}.$$

We notice that U_n is well-defined (the series converges) and measurable (as it is a pointwise limit of measurable functions). To show that U_n is uniformly distributed, we compute its cdf. Since $0 \leqslant U_n(x) \leqslant 1$ for all x, we have

$$\mathbb{P}(U_n \leqslant x) = 0,\ \forall x < 0,\ \mathbb{P}(U_n \leqslant x) = 1,\ \forall x \geqslant 1.$$

Let $0 \leqslant x < 1$. We have

$$\begin{aligned}\{U_n \leqslant x\} &= \left\{ y \in [0,1] : \sum_{k=1}^{\infty} \frac{C_{n,k}(y)}{2^k} \leqslant \sum_{k=1}^{\infty} \frac{C_k(x)}{2^k} \right\} \\ &= \bigsqcup_{k=1}^{\infty} \{C_{n,1} = c_1(x), \dots, C_{n,k-1} = c_{k-1}(x),\ C_{n,k}(y) < c_k(x)\} \sqcup \{C_{n,k} = c_k(x),\ \forall k\}.\end{aligned}$$

Noticed that

$$\mathbb{P}(C_{n,1} = c_1(x), \dots, C_{n,k} = c_k(x)) = \prod_j \mathbb{P}(C_{n,j} = c_j(x)) = \prod_j \frac{1}{2} = \frac{1}{2^k},$$

by the continuity from above we have

$$\mathbb{P}\left(\{C_{n,k} = c_k(x),\ \forall k\}\right) = \lim_{k \to \infty} \frac{1}{2^k} = 0.$$

Moreover,

$$\mathbb{P}(C_{n,1} = c_1(x), \dots, C_{n,k-1} = c_{k-1}(x), C_{n,k} < c_k(x)) = \frac{1}{2^{k-1}}\mathbb{P}(C_{n,k} < c_k(x)),$$

and since,

$$\mathbb{P}(C_{n,k} < c_k(x)) = \begin{cases} 0, & c_k(x) = 0, \\ \frac{1}{2}, & c_k(x) = 1, \end{cases} = \frac{c_k(x)}{2},$$

we conclude that

$$\mathbb{P}(U_n \leqslant x) = \sum_{k=1}^{\infty} \frac{c_k(x)}{2^k} = x.$$

Therefore,

$$\mathbb{P}(U_n \leqslant x) = \begin{cases} 0, & x \leqslant 0, \\ x, & 0 < x \leqslant 1, \\ 1, & x \geqslant 1, \end{cases}$$

and this shows that U_n is a uniform random variable. Finally, we verify that the U_n are independent. Now, denote by $\mathscr{F}_n := \sigma(C_{n,k} \;:\; k \in \mathbb{N})$. It is clear that U_n is $\mathscr{F}_n$ measurable and that $\mathscr{F}_n$ are independent $\sigma-$algebras, so also the U_n are independent.

Second step. Let F be a generic cfd. By the first step, $([0,1], \mathscr{B}_{\mathbb{R}}, \lambda_1)$ there is a sequence of i.i.d. (U_n), that is

$$\lambda_1(U_n \leqslant u) = u, \; u \in [0,1].$$

So, in particular,

$$F(x) = \lambda_1 \left(U_n \leqslant F(x)\right).$$

If F is continuous and strictly increasing, we can write previous relation as

$$F_n(x) = \lambda_1 \left(F^{-1}(U_n) \leqslant x\right),$$

so defining $X_n := F^{-1}(U_n)$ we have the desired sequence.
For a general cdf F, this is not necessarily continuous and strictly increasing. However, if we define

$$G : [0,1] \to [-\infty, +\infty], \quad G(y) := \inf\{x \in \mathbb{R} : F(x) \geqslant y\},$$

and we set $X_n := G(U_n)$. Then,

$$X_n = G(U_n) \leqslant x \iff U_n \leqslant F(x),$$

so

$$\mathbb{P}(X_n \leqslant x) = \lambda_1(U_n \leqslant F(x)) = F(x),$$

from which the conclusion follows.

6.4. Exercises

Exercise 6.4.1 (∗). *Let X and Y be random variables such that*

$$\mathbb{P}(X > x, \; Y > y) = \mathbb{P}(X > x)\mathbb{P}(Y > y), \; \forall x, y \in \mathbb{R}.$$

Does it follow from this that X and Y are independent?

Exercise 6.4.2 (∗∗). *Let X, Y be independent random variables.*

i) *Check that $\rho(X,Y) = 0$.*
ii) *Check that if (X,Y) is Gaussian, then also the vice versa of i) holds.*
iii) *Show, with an example, that it is possible to have $\rho(X,Y) = 0$ but X, Y are not independent.*

Exercise 6.4.3 (∗∗). *Random variables with the density $f(x) = \frac{\alpha}{2}e^{-\alpha|x|}$, where $\alpha > 0$, are called Laplace random variables with parameter α. Let X and Y be independent exponential random variables with parameter 2, and let $Z = X - Y$.*

i) *Find the density and the characteristic function of the random variable $-Y$.*

ii) *Prove that* Z *is a Laplace random variable, determine its parameter, and compute its characteristic function.*

Exercise 6.4.4 $(**)$**.** *Let* $X, Y \sim \exp\lambda$. *Show that if* X *and* Y *are independent, then the random variables* $X+Y$ *and* $\frac{X}{Y}$ *are also independent.*

Exercise 6.4.5 $(**)$**.** *Consider a rectangle* $R := [0,a] \times [0,b]$. *On each side of the rectangle, points* $(X,0)$ *and* $(0,Y)$ *are chosen randomly, uniformly in their respective intervals, and independently. Let* $T_{X,Y}$ *denote the triangle with vertices* $(0,0)$, $(X,0)$, *and* $(0,Y)$. *What is the probability that the area of* $T_{X,Y}$ *is less than one-quarter of the area of the rectangle* R*?*

Exercise 6.4.6 $(**)$**.** *Let* A *and* B *be independent random variables uniformly distributed on* $[0,1]$. *Consider the quadratic equation*

$$x^2 + 2Ax + B = 0.$$

What is the probability that its solutions are real?

Exercise 6.4.7 $(**)$**.** *Let* $X, Y \sim U([0,1])$ *be independent random variables and let*

$$U := \min(X,Y), \quad V := \max(X,Y).$$

Determine $\mathbb{E}[U], \mathbb{E}[V]$ *and* $\mathrm{Cov}(U,V)$.

Exercise 6.4.8 $(**)$**.** *We denote by* T_n *the best time recorded in the 100m of the n-th race. Since the temporal dimensions are not of interest to us, we assume* $T_n \sim U(0,1)$, *and we also assume that the* T_n *are independent random variables. Let* A_n *be the event "a new 100m record is set in the n-th race."*

i) *Compute the c.d.f. of the random variable* $S_n := \min\{T_1, \dots, T_{n-1}\}$.
ii) *Prove that* $\mathbb{P}(S_n > t,\ T_n > s) = \mathbb{P}(S_n > t)\mathbb{P}(T_n > s)$.
iii) *Describe* A_n *in terms of the random variables* T_k, $k = 1, \dots, n$, *and show that* $\mathbb{P}(A_n) = \frac{1}{n}$.
iv) *Assuming that the* A_n *are independent, what is the probability that a record remains unbroken forever?*

Exercise 6.4.9 $(**)$**.** *On the segment* $[a,b]$, *let* $c \in]a,b[$ *(i.e.,* $a < c < b$, *with* a, b, c *fixed). Two points* $X \in [a,c]$ *and* $Y \in [c,b]$ *are chosen randomly with a uniform distribution. Compute the probability that the lengths of the segments* $[a,X]$, $[X,Y]$, *and* $[Y,b]$ *can form the sides of a triangle. (Recall that* $\alpha, \beta, \gamma \geqslant 0$ *can be the lengths of the sides of a triangle if and only if* $\alpha \leqslant \beta + \gamma$, $\beta \leqslant \alpha + \gamma$, *and* $\gamma \leqslant \alpha + \beta$*).*

7

Conditioning

7.1. L^2 conditional expectation

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, $\mathscr{G}$ be a sub $\sigma-$algebra of events, that is $\mathscr{G} \subset \mathscr{F}$. A $\mathscr{G}-$measurable random variable Y is a random variable for which

$$\{Y \in E\} \in \mathscr{G}, \;\; \forall E \in \mathscr{B}_{\mathbb{R}}.$$

Equivalently,

$$Y \text{ is } \mathscr{G}-\text{measurable} \quad \Longleftrightarrow \quad \sigma(Y) \subset \mathscr{G}.$$

With a little abuse of notations, we will sometimes write $Y \in \mathscr{G}$ to represent this situation.

Given a random variable X, we consider the problem of determining the *"best approximation"* of X through a $\mathscr{G}-$measurable random variable. A natural setup for this problem is the following. Let $\mathcal{H} := L^2(\Omega, \mathscr{F}, \mathbb{P})$ be the Hilbert space of L^2 random variables equipped by the scalar product

$$\langle X, Y \rangle := \mathbb{E}[XY] \equiv \int_\Omega XY \, d\mathbb{P}.$$

Let also

$$\mathcal{G} := L^2(\Omega, \mathscr{G}, \mathbb{P}),$$

be the subspace of $\mathcal{H}$ made of $\mathscr{G}-$measurable random variables. Clearly, $\mathcal{G}$ is a closed subspace of $\mathcal{H}$. This because if $(Y_n) \subset \mathcal{G}$ is such that $Y_n \xrightarrow{L^2} Y$, then $Y \in L^2$ and since the limit of $\mathscr{G}-$measurable functions is a $\mathscr{G}-$measurable function, we conclude that $Y \in \mathcal{G}$. These facts suggest a proper set up of the approximation problem posed above: *determine* $Y \in \mathcal{G}$ *such that*

$$\|X - Y\|_2 = \min_{Z \in \mathcal{G}} \|X - Z\|_2.$$

In this setup, the solution is provided by the **orthogonal projection** of X on $\mathcal{G}$, that is

$$Y = \Pi_{\mathcal{G}} X.$$

It is convenient to recall that $\Pi_{\mathcal{G}} X$ is characterized to be the unique $Y \in \mathcal{G}$ such that

$$\langle X - Y, Z \rangle = 0, \; \forall Z \in \mathcal{G},$$

that is,

$$\mathbb{E}[XZ] = \mathbb{E}[YZ], \; \forall Z \in \mathcal{G}. \tag{7.1.1}$$

The orthogonal projection verifies some simple properties:

Proposition 7.1.1

The following properties hold:

i) (linearity): $\Pi_{\mathcal{G}}(\alpha X + \beta Y) = \alpha \Pi_{\mathcal{G}} X + \beta \Pi_{\mathcal{G}} Y$

ii) (monotonicity) $X \leqslant Y$ a.s., then $\Pi_{\mathcal{G}} X \leqslant \Pi_{\mathcal{G}} Y$ a.s.

iii) If X is $\mathscr{G}-$measurable, then $\Pi_{\mathcal{G}} X = X$.

iv) If X is independent of $\mathscr{G}$, then $\Pi_{\mathcal{G}} X = \mathbb{E}[X]$.

v) If $\mathscr{H} \subset \mathscr{G} \subset \mathscr{F}$ then $\Pi_{\mathcal{H}}(\Pi_{\mathcal{G}} X) = \Pi_{\mathcal{H}} X$.

vi) If $X \in L^2$ and $Y \in L^\infty$, with $Y \in \mathscr{G}$, then $\Pi_{\mathcal{G}}(XY) = Y \Pi_{\mathcal{G}} X$.

Proof. All the properties follow from the orthogonality characterization (7.1.1). For i) we have

$$\mathbb{E}[(\alpha X + \beta Y)Z] = \alpha \mathbb{E}[XZ] + \beta \mathbb{E}[YZ] = \alpha \mathbb{E}[(\Pi_{\mathcal{G}} X)Z] + \beta \mathbb{E}[(\Pi_{\mathcal{G}} Y)Z] = \mathbb{E}\left[(\alpha \Pi_{\mathcal{G}} X + \beta \Pi_{\mathcal{G}} Y) Z\right],$$

$\forall Z \in \mathcal{G}$. And since $\alpha \Pi_{\mathcal{G}} X + \beta \Pi_{\mathcal{G}} Y \in \mathcal{G}$ we conclude that

$$\Pi_{\mathcal{G}}(\alpha X + \beta Y) = \alpha \Pi_{\mathcal{G}} X + \beta \Pi_{\mathcal{G}} Y.$$

ii) Let $Z \in \mathcal{G}$, $Z \geqslant 0$. We have

$$\mathbb{E}[\Pi_{\mathcal{G}} X Z] = \mathbb{E}[XZ] \leqslant \mathbb{E}[YZ] = \mathbb{E}[\Pi_{\mathcal{G}} Y Z], \implies \mathbb{E}\left[(\Pi_{\mathcal{G}} Y - \Pi_{\mathcal{G}} X)Z\right] \geqslant 0.$$

Let $G := \{\Pi_{\mathcal{G}} Y - \Pi_{\mathcal{G}} X < -\varepsilon\}$ (with $\varepsilon > 0$) and $Z := 1_G \in \mathcal{G}$. Then, the previous says

$$0 \leqslant \mathbb{E}\left[(\Pi_{\mathcal{G}} Y - \Pi_{\mathcal{G}} X)1_G\right] \leqslant \mathbb{E}[-\varepsilon 1_G] = -\varepsilon \mathbb{P}(G), \implies \mathbb{P}(G) \leqslant 0,$$

which possible iff $\mathbb{P}(G) = 0$. Since ε is arbitrary, we conclude that $\mathbb{P}\left(\Pi_{\mathcal{G}} \widetilde{X} - \Pi_{\mathcal{G}} X < 0\right) = 0$.

iii) If $X \in \mathscr{G}$, then $X \in \mathcal{G}$, so $\Pi_{\mathcal{G}} X = X$.

iv) If X is independent of $\mathscr{G}$ we notice that

$$\mathbb{E}[XZ] = \mathbb{E}[X]\mathbb{E}[Z] = \mathbb{E}\left[\mathbb{E}[X]Z\right],$$

and since constants are $\mathscr{G}-$measurable we conclude.

v), vi) Straightforward.

Example 7.1.2

Let $\mathscr{G} = \sigma(E_1, \ldots, E_n)$ where (E_k) are a partition of Ω, that is $\Omega = \bigsqcup_{k=1}^n E_k$, with $0 < \mathbb{P}(E_k) < 1$, $k = 1, \ldots, n$. Then

$$\Pi_{\mathcal{G}} X = \sum_{k=1}^n \frac{1}{\mathbb{P}(E_k)} \mathbb{E}[X 1_{E_k}] 1_{E_k}.$$

Proof. It is easy to check that $\sigma(E_1, \ldots, E_n)$ is made of finite unions of the sets E_k. From this it follows that the $\mathscr{G}$ measurable functions are the simple functions with bases the E_k, that is function of type $\sum_{j=1}^h c_j 1_{E_{k_j}}$. Thus, $\mathcal{G} = \text{Span}(1_{E_1}, \ldots, 1_{E_n})$. In this case, setting $e_k := \frac{1_{E_k}}{\|1_{E_k}\|_2}$, (e_k) is an orthonormal basis for $\mathcal{G}$. we have

$$\Pi_{\mathcal{G}} X = \sum_{k=1}^n \langle X, e_k \rangle e_k = \sum_{k=1}^n \frac{1}{\|1_{E_k}\|_2^2} \mathbb{E}[X 1_{E_k}] 1_{E_k},$$

and since $\|1_{E_k}\|_2^2 = \mathbb{E}[1_{E_k}^2] = \mathbb{E}[1_{E_k}] = \mathbb{P}(E_k)$, the conclusion follows.

In previous example, we have that

$$\Pi_{\mathcal{G}} X(\omega) = \frac{1}{\mathbb{P}(E_k)} \mathbb{E}[X 1_{E_k}] = \frac{1}{\mathbb{P}(E_k)} \int_{E_k} X \, d\mathbb{P}, \ \omega \in E_k$$

This motivates the notation

$$\mathbb{E}[X \mid \mathscr{G}] := \Pi_{\mathcal{G}} X,$$

called the **conditional expectation of** X **given** $\mathscr{G}$. With this notation, the properties i),...,vi) of the Proposition 7.1.1 acquire a new flavor:

i) (linearity): $\mathbb{E}[(\alpha X + \beta Y) \mid \mathscr{G}] = \alpha \mathbb{E}[X \mid \mathscr{G}] + \beta \mathbb{E}[Y \mid \mathscr{G}]$, $\forall \alpha, \beta \in \mathbb{R}$.
ii) (monotonicity) $X \leqslant Y$ $\mathbb{P}-$a.s., then $\mathbb{E}[X \mid \mathscr{G}] \leqslant \mathbb{E}[Y \mid \mathscr{G}]$ $\mathbb{P}-$a.s.
iii) If X is $\mathscr{G}-$measurable, then $\mathbb{E}[X \mid \mathscr{G}] = X$.
iv) If X is independent of $\mathscr{G}$, then $\mathbb{E}[X \mid \mathscr{G}] = \mathbb{E}[X]$.
v) (sub-conditioning) If $\mathscr{H} \subset \mathscr{G} \subset \mathscr{F}$ then $\mathbb{E}\left[\mathbb{E}[X \mid \mathscr{G}] \mid \mathscr{H}\right] = \mathbb{E}[X \mid \mathscr{H}]$.
vi) If $X \in L^2$ and $Y \in L^\infty$, with $Y \in \mathscr{G}$, then $\mathbb{E}[XY \mid \mathscr{G}] = Y \mathbb{E}[X \mid \mathscr{G}]$.

This is the bridge to the next section topic.

7.2. L^1 conditional expectation

The properties of the L^2 conditional expectation enlighten the nature of an "expectation" of the orthogonal projection on $\mathcal{G} = L^2(\Omega, \mathscr{G}, \mathbb{P})$. As such, we could expect that

$$\mathbb{E}[X \mid \mathscr{G}],$$

should be well defined for $X \in L^1(\Omega, \mathscr{G}, \mathbb{P})$. However, since $L^1 \not\subset L^2$ (but rather, by the Cauchy-Schwarz's inequality, $L^2 \subset L^1$), and the definition of the L^2 conditional expectation is a typical Hilbert spaces story (something which is not L^1) the definition of this conditional expectation is not automatic. We will now show the way to do this.

Theorem 7.2.1

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a Probability space, $\mathscr{G} \subset \mathscr{F}$ a sub $\sigma-$algebra of $\mathscr{F}$. If $X \in L^1(\Omega, \mathscr{F}, \mathbb{P})$, there exists then a unique (modulo probability null sets) $Y \in L^1(\Omega, \mathscr{G}, \mathbb{P})$ such that

$$\mathbb{E}[XZ] = \mathbb{E}[YZ], \ \forall Z \in L^\infty(\Omega, \mathscr{G}, \mathbb{P}). \tag{7.2.1}$$

Y is called **conditional expectation of** X **given** $\mathscr{G}$ and we denote it by $\mathbb{E}[X \mid \mathscr{G}]$.

Proof. **Step 1.** Let $X \geqslant 0$ a.s.. Recall that there exists a sequence (S_n) of positive simple random variables such that

$$S_n \nearrow, \ S_n \xrightarrow{pw} X.$$

Since $(S_n) \subset L^\infty \subset L^2$, the conditional expectation $\mathbb{E}[S_n \mid \mathscr{G}]$ is well defined and, by monotonicity of the cond. exp., we have $\mathbb{E}[S_n \mid \mathscr{G}] \nearrow$. This authorizes to set

$$Y := \lim_n \mathbb{E}[S_n \mid \mathscr{G}].$$

Being Y the point wise limit of $\mathscr{G}$ measurable random variables, $Y \in \mathscr{G}$. Therefore, if $Z \in L^\infty(\Omega, \mathscr{G}, \mathbb{P})$ is positive, by monotone convergence,

$$\mathbb{E}[XZ] = \lim_n \mathbb{E}[S_n Z] \overset{(7.1.1)}{=} \lim_n \mathbb{E}\left[\mathbb{E}[S_n \mid \mathscr{G}]Z\right] = \mathbb{E}[YZ].$$

Now, for a generic $Z \in L^\infty(\Omega, \mathscr{G}, \mathbb{P})$, writing $Z = Z_+ - Z_-$ we have

$$\mathbb{E}[XZ_\pm] = \mathbb{E}[YZ_\pm], \implies \mathbb{E}[XZ] = \mathbb{E}[X(Z_+ - Z_-)] = \mathbb{E}[Y(Z_+ - Z_-)] = \mathbb{E}[YZ],$$

from which the conclusion follows.

Step 2. Let $X \in L^1$. Writing $X = X_+ - X_-$, we have

$$\mathbb{E}[X_\pm Z] = \mathbb{E}[Y_\pm Z], \ \forall Z \in L^\infty(\Omega, \mathscr{G}, \mathbb{P}). \tag{7.2.2}$$

Setting $Z_\pm = 1_{Y_\pm > 0} \in L^\infty(\Omega, \mathscr{G}, \mathbb{P})$ then

$$\mathbb{E}[Y_\pm] = \mathbb{E}[Y_\pm Z_\pm] = \mathbb{E}[X_\pm 1_{Y_\pm > 0}] \leqslant \mathbb{E}[X_\pm] < +\infty$$

because $X \in L^1$. Therefore, $Y_\pm \in L^1(\Omega, \mathscr{G}, \mathbb{P})$, then also $Y \in L^1(\Omega, \mathscr{G}, \mathbb{P})$. Now, by subtracting the two $\pm$ identities (7.2.2) we get

$$\mathbb{E}[XZ] = \mathbb{E}[(X_+ - X_-)Z] = \mathbb{E}[(Y_+ - Y_-)Z] = \mathbb{E}[YZ], \ \forall Z \in L^\infty(\Omega, \mathscr{G}, \mathbb{P}),$$

which is the (7.2.1).

Step 3. Uniqueness. If $Y, \widetilde{Y}$ veirfy the (7.2.1), then

$$\mathbb{E}[(Y - \widetilde{Y})Z] = 0, \ \forall Z \in L^\infty, \ Z \in \mathscr{G}.$$

Since $Y, \widetilde{Y} \in \mathscr{G}$, we have $\operatorname{sgn}(Y - \widetilde{Y}) \in \mathscr{G}$ and

$$0 = \mathbb{E}[(Y - \widetilde{Y})Z] = \mathbb{E}[|Y - \widetilde{Y}|]$$

from which $Y = \widetilde{Y}$ with probability 1.

L^1 conditional expectation verifies similar properties as for the L^2 conditional expectation.

Proposition 7.2.2

The following properties hold:

- i) (linearity): $\mathbb{E}[\alpha X + \beta Y \mid \mathscr{G}] = \alpha \mathbb{E}[X \mid \mathscr{G}] + \beta \mathbb{E}[Y \mid \mathscr{G}]$
- ii) (monotonicity) $X \leqslant Y$ $\mathbb{P}-$a.s., then $\mathbb{E}[X \mid \mathscr{G}] \leqslant \mathbb{E}[Y \mid \mathscr{G}]$ $\mathbb{P}-$a.s.
- iii) If X is $\mathscr{G}-$measurable, then $\mathbb{E}[X \mid \mathscr{G}] = X$.
- iv) If X is independent of $\mathscr{G}$, then $\mathbb{E}[X \mid \mathscr{G}] = \mathbb{E}[X]$.
- v) (sub-conditioning) If $\mathscr{H} \subset \mathscr{G} \subset \mathscr{F}$ then $\mathbb{E}\left[\mathbb{E}[X \mid \mathscr{G}] \mid \mathscr{H}\right] = \mathbb{E}[X \mid \mathscr{H}]$.
- vi) If $X \in L^1$ and $Y \in L^\infty$, with $Y \in \mathscr{G}$, then $\mathbb{E}[XY \mid \mathscr{G}] = Y\mathbb{E}[X \mid \mathscr{G}]$.

The proof is left as an exercise.

7.2.1. Conditional density. A particular case of conditional expectation is the following: given any two random variables X, Y, determine

$$\mathbb{E}[X \mid \sigma(Y)] =: \mathbb{E}[X \mid Y]$$

Proposition 7.2.3

Assume that (X,Y) is absolutely continuous bivariate random variable with density $f_{X,Y}$. Then,

$$\mathbb{E}[X \mid Y] = \varphi(Y),$$

where

$$\varphi(y) = \int_{\mathbb{R}} x f_{X|Y}(x|y)dx. \tag{7.2.3}$$

with

$$f_{X|Y}(x|y) := \frac{f_{X,Y}(x,y)}{f_Y(y)} \tag{7.2.4}$$

The function $f_{X|Y}$ is called **conditional density of X given Y**.

Proof. It is clear that $\varphi(Y) \in \sigma(Y)$. We verify the characterizing condition (7.2.1) for $\varphi(Y)$, that is

$$\mathbb{E}[XZ] = \mathbb{E}[\varphi(Y)Z], \ \forall Z \in L^\infty(\Omega, \mathscr{G}, \mathbb{P}).$$

Let $Z = 1_G$ where $G \in \sigma(Y) = \{\{Y \in E\} \ : \ E \in \mathscr{B}_{\mathbb{R}}\}$, that is, let us show that

$$\mathbb{E}[X1_{Y\in E}] = \mathbb{E}[\varphi(Y)1_{Y\in E}]$$

If this happens, we get (7.2.1) for every simple function, then the conclusion follows by a standard approximation argument. We have

$$\begin{aligned}
\mathbb{E}[\varphi(Y)1_{Y\in E}] &= \int_E \varphi(y) f_Y(y)\, dy = \int_E \left(\int_{\mathbb{R}} x f_{X|Y}(x|y)\, dx \right) f_Y(y)\, dy \\
&= \int_E \left(\int_{\mathbb{R}} x f(x,y)\, dx \right) dy = \int_{\mathbb{R}\times\mathbb{R}} x 1_E(y) f(x,y)\, dy\, dx \\
&= \mathbb{E}[X 1_E(Y)].
\end{aligned}$$

With this the conclusion follows.

Since

$$\mathbb{E}[X \mid Y] = \varphi(Y),$$

the notation

$$\varphi(y) = \mathbb{E}[X \mid Y = y]$$

is often used.

Example 7.2.4

Q. Let $X \sim \mathcal{N}(m, \sigma^2)$, and $Y \sim \mathcal{N}(0, \sigma^2)$ be independent random variables. Determine $\mathbb{E}[X+Y|X]$ and $\mathbb{E}[X|X+Y]$.

A. We have $\mathbb{E}[X+Y \mid X] = \mathbb{E}[X \mid X] + \mathbb{E}[Y \mid X] = X + \mathbb{E}[Y]$ because of the assumptions. More involved is the calculation of $\mathbb{E}[X \mid X+Y]$. We need first to determine the conditional density $f_{X|X+Y}(x|y)$, and this means that we need to determine first the joint density $f_{X,X+Y}$. Set $Z = X+Y$ in such a way that $(X, X+Y) = (X,Z) = T(X,Y)$ where

$$T(x,y) = (x, x+y) = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

As well known

$$f_{XZ}(x,z) = f_{XY}(T^{-1}(x,z)) \underbrace{|\det T^{-1}|}_{=1} = f_{XY}(x, z-x) \stackrel{indep.}{=} f_X(x) f_Y(z-x).$$

Therefore,

$$f_{X|Z}(x|z) = \frac{f_{XZ}(x,z)}{f_Z(z)} = \frac{f_X(x) f_Y(z-x)}{f_Z(z)}.$$

Since $Z = X+Y$ and X, Y are independent,

$$f_Z(z) = f_X * f_Y(z).$$

We have

$$\widehat{f_Z}(\xi) = \widehat{f_X}(\xi)\widehat{f_Y}(\xi) = e^{i\xi m - \frac{1}{2}\sigma^2\xi^2} e^{-\frac{1}{2}\sigma^2\xi^2} = e^{im\xi - \frac{1}{2}(2\sigma^2)\xi^2},$$

from which

$$f_Z(z) = \frac{1}{\sqrt{4\pi\sigma^2}} e^{-\frac{(z-m)^2}{4\sigma^2}}.$$

Therefore,

$$\mathbb{E}[X \mid Z = z] = \int_{\mathbb{R}} x f_{X|Z}(x|z)\, dx = \int_{\mathbb{R}} x \frac{f_X(x) f_Y(z-x)}{f_Z(z)}\, dx.$$

Now,

$$\begin{aligned} f_X(x) f_Y(z-x) &= \tfrac{1}{2\pi\sigma^2} e^{-\frac{(x-m)^2}{2\sigma^2}} e^{-\frac{(z-x)^2}{2\sigma^2}} = \tfrac{1}{2\pi\sigma^2} e^{-\frac{x^2-2mx+m^2+x^2-2xz+z^2}{2\sigma^2}} \\ &= \tfrac{1}{2\pi\sigma^2} e^{-\frac{x^2-(m+z)x+\left(\frac{m+z}{2}\right)^2}{\sigma^2}} e^{-\frac{-2\left(\frac{m+z}{2}\right)^2+m^2+z^2}{2\sigma^2}} \\ &= \tfrac{1}{2\pi\sigma^2} e^{-\frac{\left(x-\frac{m+z}{2}\right)^2}{\sigma^2}} e^{-\frac{(z-m)^2}{4\sigma^2}} = \tfrac{1}{\sqrt{\pi\sigma^2}} e^{-\frac{\left(x-\frac{m+z}{2}\right)^2}{\sigma^2}} f_Z(z), \end{aligned}$$

so

$$\mathbb{E}[X \mid Z = z] = \frac{1}{\sqrt{\pi\sigma^2}} \int_{\mathbb{R}} x e^{-\frac{\left(x-\frac{m+z}{2}\right)^2}{\sigma^2}}\, dx = \frac{1}{\sqrt{2}} \int_{\mathbb{R}} u e^{-\frac{\left(u-\frac{m+z}{\sqrt{2}}\right)^2}{2\sigma^2}} \frac{du}{\sqrt{2\pi\sigma^2}} = \frac{m+z}{2}.$$

We conclude that $\mathbb{E}[X \mid X+Y] = \frac{m+X+Y}{2}$.

7.3. Exercises

Exercise 7.3.1 ($**$)**.** *For each of the following cases, determine $\mathbb{E}[Y \mid X]$ known the joint distribution of (X,Y):*

i) $f_{X,Y}(x,y) = \lambda^2 e^{-\lambda y} 1_{[0,y]}(x)$.

ii) $f_{X,Y}(x,y) = x e^{-x(y+1)} 1_{[0,+\infty[^2}(x,y)$.

Exercise 7.3.2 ($**$)**.** *Let Y be a random variable with density $\frac{a}{y^2} 1_{[1,2]}(y)$, where $a > 0$ is a constant to be determined. Let also X be a random variable such that $f_{X|Y}(\sharp|y)$ is a Gaussian distribution $\mathcal{N}(0, y^2)$.*

i) *Compute the density of* XY, $\mathbb{E}[X \mid Y]$, $\mathbb{E}[X]$ *and* $\mathbb{P}(X > 0)$.
ii) *Compute* f_X *and* $\mathbb{V}[X]$. *Is* X *normal?*
iii) *Are* X *and* Y *independent?*

Exercise 7.3.3 $(**)$**.** *Let* X *and* Y *be independent random variables having Poisson distribution with parameter* λ_1 *and* λ_2 *respectively. Determine* $\mathbb{E}[X \mid X+Y]$.

Exercise 7.3.4 $(**)$**.** *Let* $(X,Y) \sim \mathcal{N}(m,C)$. *Determine* $\mathbb{E}[X \mid Y]$.

Exercise 7.3.5 $(**)$**.** *Let* X,Y *i.i.d. random variables with common density* f. *Determine* $\mathbb{E}[X-Y \mid X+Y]$.

Exercise 7.3.6 $(**)$**.** *Let* $X,Y \in L^1$. *Prove the formula*

$$\mathbb{E}\left[X \mid \mathbb{E}[X \mid Y]\right] = \mathbb{E}[X \mid Y].$$

Do the proof for both cases, assuming XY *is absolutely continuous and in general.*

Exercise 7.3.7 $(**+)$**.** *Prove the monotone convergence for the* L^1 *conditional expectation. That is: let* $(X_n) \subset L^1$ *be such that* $0 \leqslant X_n \leqslant X_{n+1}$ *a.s.,* $\forall n \in \mathbb{N}$. *Then,*

$$\lim_n \mathbb{E}[X_n \mid \mathscr{G}] = \mathbb{E}[\lim_n X_n \mid \mathscr{G}].$$

Exercise 7.3.8 $(**+)$**.** *Prove the dominated convergence property for the* L^1 *conditional expectation: let* $(X_n) \subset L(\Omega)$ *be such that:*

i) $X_n \xrightarrow{a.s.} X$.
ii) *there exists* $Y \in L^1$ *such that* $|X_n| \leqslant Y$ *a.s.* $\forall n \in \mathbb{N}$.

Then

$$\mathbb{E}[X_n|\mathscr{G}] \xrightarrow{a.s.} \mathbb{E}[X|\mathscr{G}].$$

Exercise 7.3.9 $(**+)$**.** *Let* $(\Omega, \mathscr{F}, \mathbb{P})$ *be a probability space,* $\mathscr{G} \subset \mathscr{F}$ *be a sub* $\sigma-$*algebra of* $\mathscr{F}$. *We define the conditional probability*

$$\mathbb{P}(E \mid \mathscr{G}) := \mathbb{E}\left[1_E \mid \mathscr{G}\right].$$

Check that:

i) $\mathbb{P}(\varnothing \mid \mathscr{G}) = 0$ *(a.s.),* $\mathbb{P}(\Omega \mid \mathscr{G}) = 1$ *(a.s.).*
ii) $0 \leqslant \mathbb{P}(E \mid \mathscr{G}) \leqslant 1$ *a.s.,* $\forall E \in \mathscr{F}$.
iii) $\mathbb{P}\left(\bigsqcup_n E_n \mid \mathscr{G}\right) = \sum_k \mathbb{P}(E_n \mid \mathscr{G})$ *(a.s.).*

8

Convergence

In this chapter we consider sequences of random variable (X_n) and discuss their convergence. Since the X_n are measurable functions, natural options for convergence are the L^p convergence and almost sure pointwise convergence or, as preferred in Probability Theory, convergence with probability 1. Other and weaker difinitions of convergence can be introduced, as *convergence in probability* and *weak convergence*. We introduce all these concepts exploring what are their relations.

8.1. L^p convergence

Let us recall that $L^p(\Omega)$, $1 \leqslant p \leqslant +\infty$ is a normed space equipped with $\|\cdot\|_p$ norm. This is defined as

$$\|X\|_p := \left(\int_\Omega |X|^p \, d\mathbb{P}\right)^{1/p} \equiv \mathbb{E}[|X|^p]^{1/p}, \ \ (1 \leqslant p < +\infty),$$

and

$$\|X\|_\infty := \text{ess sup}|X|, \ (p = +\infty).$$

We say that

$$X_n \xrightarrow{L^p} X, \quad \Longleftrightarrow \quad \|X_n - X\|_p \longrightarrow 0.$$

The $p-$norms are ordered in the sense, as we will prove now,

$$\|X\|_p \leqslant \|X\|_q, \ \ 1 \leqslant p < q \leqslant +\infty.$$

If $q = +\infty$ the inequality is trivial:

$$\|X\|_p = \mathbb{E}[|X|^p]^{1/p} \leqslant \mathbb{E}[\|X\|_\infty^p]^{1/p} = \|X\|_\infty \mathbb{E}[1]^{1/p} = \|X\|_\infty.$$

If $1 \leqslant p < q < +\infty$ the inequality is non trivial. It can be proved as a consequence of Hölder inequality or, in alternative, as a consequence of the following remarkable inequality:

Theorem 8.1.1: Jensen inequality

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Let $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ be a convex function, that is

$$\varphi(\lambda x + (1-\lambda)y) \leqslant \lambda\varphi(x) + (1-\lambda)\varphi(y), \ \forall \lambda \in [0,1], \ \forall x, y \in \mathbb{R},$$

Then, if $X \in L^1(\Omega)$, it holds

$$\varphi\left(\mathbb{E}[X]\right) \leqslant \mathbb{E}[\varphi(X)].$$

Proof. , For simplicity, we do the proof in the case of a differentiable function φ. In this case, being convex

$$\varphi(y) \geqslant \varphi(x) + \varphi'(x)(y-x), \ \forall x, y \in \mathbb{R},$$

so

$$\varphi(X) \geqslant \varphi(x) + \varphi'(x)(X-x),$$

and, taking the expectation, we would have

$$\mathbb{E}[\varphi(X)] \geqslant \varphi(x) + \varphi'(x)\left(\mathbb{E}[X]-x\right).$$

Choosing $X = \mathbb{E}[X]$ we get $\mathbb{E}[\varphi(X)] \geqslant \varphi(x) = \varphi(\mathbb{E}[X])$ which is the conclusion.

Corollary 8.1.2

$$\|X\|_p \leqslant \|X\|_q, \ \forall 1 \leqslant p < q < +\infty.$$

Proof. Let $\varphi(x) = x^{q/p}$. Since $\alpha > 1$, φ is convex. Therefore,

$$\|X\|_p^q = \mathbb{E}[|X|^p]^{q/p} = \varphi\left(\mathbb{E}[|X|^p]\right) \leqslant \mathbb{E}\left[\varphi(|X|^p)\right] = \mathbb{E}[|X|^q] = \|X\|_q^q,$$

from which the conclusion follows.

Thus, in particular, among the L^p convergences, the L^1 convergence is the weakest, the L^∞ the strongest.

8.2. Almost sure convergence

Almost sure convergence or, in probabilistic jargoon, *convergence with probability 1*:

$$X_n \stackrel{a.s.}{\longrightarrow} X, \quad \Longleftrightarrow \quad \mathbb{P}\left(\left\{\omega \in \Omega \ : \ \lim_n X_n(\omega) = X(\omega)\right\}\right) = 1.$$

We already know that this convergence is, in general, weaker than L^p convergence and, at the same time, it is not implied by L^p convergence when $p < +\infty$. The following fact provides a mild relation between the two convergences:

$$X_n \stackrel{L^p}{\longrightarrow} X, \implies \exists (X_{n_k}) \subset (X_n) \ : \ X_{n_k} \stackrel{a.s.}{\longrightarrow} X.$$

A possible strategy for proving $X_n \stackrel{a.s.}{\longrightarrow} X$ is to prove that the event of ω where convergence fails has probability 0. Let's describe this event. We may notice that

$$X_n(\omega) \longrightarrow X(\omega), \quad \Longleftrightarrow \quad \boxed{\forall \varepsilon > 0, \ \exists N = N(\varepsilon) \ : \ |X_n(\omega) - X(\omega)| \leqslant \varepsilon, \ \forall n \geqslant N.}$$

So

$$\{X_n \stackrel{a.s.}{\longrightarrow} X\} = \bigcap_{\varepsilon>0} \bigcup_N \bigcap_{n \geqslant N} \{|X_n - X| \leqslant \varepsilon\} \equiv \bigcap_k \bigcup_N \bigcap_{n \geqslant N} \left\{|X_n - X| \leqslant \frac{1}{k}\right\}$$

Therefore,

$$\{X_n \stackrel{a.s.}{\not\longrightarrow} X\} = \bigcup_k \bigcap_N \bigcup_{n \geqslant N} \left\{|X_n - X| \geqslant \frac{1}{k}\right\},$$

from which we obtain

$$X_n \xrightarrow{a.s.} X, \quad \iff \quad \mathbb{P}\left(\bigcap_N \bigcup_{n\geqslant N} \{|X_n - X| \geqslant \varepsilon\}\right) = 0, \ \forall \varepsilon > 0. \tag{8.2.1}$$

In general, given a sequence (X_n) we might not be able to identify a possible limit X to test convergence. Since $(X_n(\omega)) \subset \mathbb{R}$ and $\mathbb{R}$ is a *complete space* (that is convergence is the same of fulfilling the Cauchy property), we have

$$\{(X_n) \text{ Cauchy}\} = \bigcap_k \bigcup_N \bigcap_{n,m\geqslant N} \left\{|X_n - X_m| \leqslant \frac{1}{k}\right\},$$

so,

$$\{(X_n) \text{ not Cauchy}\} = \bigcup_k \bigcap_N \bigcup_{n,m\geqslant N} \left\{|X_n - X_m| \geqslant \frac{1}{k}\right\}.$$

Therefore,

$$(X_n) \text{ converges with } \mathbb{P} = 1, \quad \iff \quad \mathbb{P}\left(\bigcap_N \bigcup_{n,m\geqslant N} \{|X_n - X_m| \geqslant \varepsilon\}\right) = 0, \quad \forall \varepsilon > 0. \tag{8.2.2}$$

The two conditions (8.2.1) and (8.2.2) emphasize the role of the set

$$\limsup_n E_n := \bigcap_N \bigcup_{n\geqslant N} E_n,$$

which is the event of ω that belong to infinitely many E_n. The following result provides a condition to ensure that this is a probability $0-1$ event:

Lemma 8.2.1: Borel-Cantelli

Let $(E_n) \subset \mathscr{F}$. Then

$$\sum_n \mathbb{P}(E_n) < +\infty, \quad \Longrightarrow \quad \mathbb{P}(\limsup_n E_n) = 0. \tag{8.2.3}$$

Moreover, if the event E_n are independent,

$$\sum_n \mathbb{P}(E_n) = +\infty, \quad \Longrightarrow \quad \mathbb{P}(\limsup_n E_n) = 1. \tag{8.2.4}$$

Proof. For (8.2.3), notice that

$$\mathbb{P}\left(\bigcup_{n\geqslant N} E_n\right) \leqslant \sum_{n\geqslant N} \mathbb{P}(E_n),$$

and since $\bigcup_{n\geqslant N} E_n \downarrow \bigcap_N \bigcup_{n\geqslant N} E_n$, by the continuity from above,

$$\mathbb{P}\left(\limsup_n E_n\right) = \lim_N \mathbb{P}\left(\bigcup_{n\geqslant N} E_n\right) \leqslant \lim_N \sum_{n\geqslant N} \mathbb{P}(E_n) = 0.$$

For the (8.2.4) notice that

$$\mathbb{P}\left((\limsup_n E_n)^c\right) = \mathbb{P}\left(\bigcup_N \bigcap_{n\geqslant N} E_n^c\right) \leqslant \sum_n \mathbb{P}\left(\bigcap_{n\geqslant N} E_n^c\right).$$

Now, by independence

$$\mathbb{P}\left(\bigcap_{n\geqslant N} E_n^c\right) = \prod_{n\geqslant N} \mathbb{P}(E_n^c) = \prod_{n\geqslant N}(1-\mathbb{P}(E_n)) = \prod_{n\geqslant N} e^{\log(1-\mathbb{P}(E_n))},$$

and since $\log(1+x) \leqslant x$ for every $x > 0$, we have

$$\mathbb{P}\left(\bigcap_{n\geqslant N} E_n^c\right) \leqslant \prod_{n\geqslant N} e^{-\mathbb{P}(E_n)} = e^{-\sum_{n\geqslant N}\mathbb{P}(E_n)} = e^{-\infty} = 0,$$

from which the conclusion follows.

Warning 8.2.2

The (8.2.4) is false, in general, if the events E_n are not independent. Take $E_n \equiv E$ with $0 < \mathbb{P}(E) < 1$. Notice that, in order $E_n = E$ be independent of $E_m = E$ we must have $\mathbb{P}(E \cap E) = \mathbb{P}(E)\mathbb{P}(E)$ that is, $\mathbb{P}(E) = \mathbb{P}(E)^2$, which is true iff $\mathbb{P}(E) = 0, 1$. In this case

$$\sum_n \mathbb{P}(E_n) = \sum_n \mathbb{P}(E) = +\infty,$$

but

$$\mathbb{P}(\limsup_n E_n) = \mathbb{P}\left(\bigcap_N \bigcup_{n\geqslant N} E_n\right) = \mathbb{P}(E) < 1. \quad \square$$

Example 8.2.3

Q. Let (X_n) be a sequence of Bernoulli r.vs. with

$$\mathbb{P}(X_n = 1) = p_n, \quad \mathbb{P}(X_n = 0) = 1 - p_n.$$

Check that if $\sum_n p_n < +\infty$ then $X_n \xrightarrow{a.s.} 0$.

A. Notice that

$$\mathbb{P}(|X_n| \geqslant \varepsilon) = \mathbb{P}(X_n = 1) = p_n.$$

So, since $\sum_n p_n < +\infty$, by the Borel-Cantelli lemma se have

$$\mathbb{P}\left(\bigcap_N \bigcup_{n\geqslant N} \{|X_n| \geqslant \varepsilon\}\right) = 0, \ \forall \varepsilon > 0,$$

and, from (8.2.1), the conclusion follows.

Example 8.2.4: The monkey paradox

A monkey types randomly on a typewriter for an indefinite amount of time. As improbable as it may seem, given enough time this monkey should be able to reproduce any predefined text, such as the complete works of Shakespeare. The probability of this happening within a reasonable time frame is practically zero, but theoretically, given infinite time, the event becomes certain!

PROOF. We may represent the predefined text as a suitable binary sequence $x_1, \dots, x_N \in \{0,1\}^N$. So we assume the typewriter has just two keys, 0 and 1. Let X_n the $n-$th key typed. We assume $\mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 1) = \frac{1}{2}$.
The event "*the monkey reproduces the sequence $x_1, \dots, x_N$ at time n*" can be written as

$$E_n := \{X_n = x_1,\ X_{n+1} = x_2, \dots, X_{n+N-1} = x_N\}.$$

Notice that, since the X_j are independent

$$\mathbb{P}(E_n) = \frac{1}{2^N}.$$

In general, E_n and E_m are not independent, but $F_n := E_{nN}$ are independent and $\mathbb{P}(F_n) = \frac{1}{2^N}$ so, trivially,

$$\sum_n \mathbb{P}(F_n) = +\infty.$$

According to the second Borel-Cantelli's Lemma, $\mathbb{P}(\limsup_n F_n) = 1$, this meaning that the event "*the monkey types the sequence $x_1, \dots, x$ infinitely many times*" is a sure event!

Define now the "random time" T as the first time the monkey types the right sequence:

$$T(\omega) = n,\ \omega \in F_n \setminus \bigcup_{j=1}^{n-1} F_j.$$

Since

$$1 = \mathbb{P}(\limsup_n F_n) = \mathbb{P}\left(\bigcap_N \bigcup_{n \geqslant N} F_n\right),$$

the random time T is well defined and finite for almost every ω. We have

$$\mathbb{E}[T] = \sum_{n=0}^{\infty} n\mathbb{P}\left(F_n \setminus \bigcup_{j=1}^{n-1} F_j\right).$$

Now,

$$\mathbb{P}\left(F_n \setminus \bigcup_{j=1}^{n-1} F_j\right) = \mathbb{P}\left(F_n \cap \bigcap_{j=1}^{n-1} F_j^c\right) = \mathbb{P}(F_n) \prod_{j=1}^{n-1} \mathbb{P}(F_j^c) = \frac{1}{2^N} \prod_{j=1}^{n-1} \left(1 - \frac{1}{2^N}\right) = \frac{1}{2^N}\left(1 - \frac{1}{2^N}\right)^{n-1}.$$

Therefore

$$\mathbb{E}[T] = \sum_{n=1}^{\infty} n \frac{1}{2^N}\left(1 - \frac{1}{2^N}\right)^{n-1} = \frac{1}{2^N} \sum_{n=1}^{\infty} nq^{n-1},$$

where $q = 1 - \frac{1}{2^N}$. Now, recall that $\sum_{n=0}^{\infty} q^n = \frac{1}{1-q}$. Differentiating w.r.t q we get

$$\sum_{n=1}^{\infty} nq^{n-1} = \frac{1}{(1-q)^2}.$$

(this formula holds for $|q| < 1$). Therefore,

$$\mathbb{E}[T] = \frac{1}{2^N} q \sum_{n=1}^{\infty} n q^{n-1} = \left(1 - \frac{1}{2^N}\right) \frac{1}{2^N} \frac{1}{\left(1 - \left(1 - \frac{1}{2^N}\right)\right)^2} = 2^N \left(1 - \frac{1}{2^N}\right) = 2^N - 1.$$

So imagine that the donkey has to write a text made of $N = 60$ binary digits, typing 1 key each a second. According to the previous calculation, it will take an expected time $2^N - 1 = 2^{60} - 1$ seconds for the monkey to reproduce the sequence, a time far beyond the time life of the Universe. . . □

8.3. Convergence in Probability

Definition 8.3.1

Let $(X_n) \subset L(\Omega)$. We say that

$$X_n \xrightarrow{\mathbb{P}} X, \iff \lim_{n \to +\infty} \mathbb{P}\left(|X_n - X| \geqslant \varepsilon\right) = 0, \ \forall \varepsilon > 0.$$

The convergence in probability is weaker than both the L^p convergence and the convergence with probability 1.

Proposition 8.3.2

$$X_n \xrightarrow{L^p} X, \implies X_n \xrightarrow{\mathbb{P}} X.$$

Proof. By Chebishev's inequality,

$$\mathbb{P}(|X_n - X| \geqslant \varepsilon) \leqslant \frac{1}{\varepsilon^p} \mathbb{E}\left[|X_n - X|^p 1_{|X_n - X| \geqslant \varepsilon}\right] \leqslant \mathbb{E}\left[|X_n - X|^p\right] = \frac{1}{\alpha^p} \|X_n - X\|_p^p \longrightarrow 0.$$ □

Warning 8.3.3

The vice versa is false. Take $(\Omega, \mathscr{F}, \mathbb{P}) = ([0,1], \mathscr{B}_{[0,1]}, \lambda_1)$ and define

$$X_n(\omega) := n^2 1_{[0,1/n]}(\omega).$$

Then,

$$\|X_n\|_1 = n^2 \frac{1}{n} = n \longrightarrow +\infty,$$

so in particular (X_n) cannot converge in L^1. However, for $\varepsilon > 0$ fixed

$$\mathbb{P}(|X_n| \geqslant \varepsilon) = \mathbb{P}(X_n > 0) = \frac{1}{n} \longrightarrow 0.$$

From this it follows that $X_n \xrightarrow{\mathbb{P}} 0$. □

Proposition 8.3.4

$$X_n \xrightarrow{a.s.} X, \implies X_n \xrightarrow{\mathbb{P}} X.$$

PROOF. Recalling of (8.2.1), we have that

$$\mathbb{P}(|X_n - X| \geqslant \varepsilon) \leqslant \mathbb{P}\left(\bigcup_{k \geqslant n} |X_k - X| \geqslant \varepsilon\right).$$

Therefore

$$0 \leqslant \lim_n \mathbb{P}(|X_n - X| \geqslant \varepsilon) \leqslant \lim_n \mathbb{P}\left(\bigcup_{k \geqslant n} |X_n - X| \geqslant \varepsilon\right) = \mathbb{P}\left(\bigcap_n \bigcup_{k \geqslant n} |X_n - X| \geqslant \varepsilon\right) = 0.$$

Warning 8.3.5

The vice versa is false. Take $(\Omega, \mathscr{F}, \mathbb{P}) = ([0,1], \mathscr{B}_{[0,1]}, \lambda_1)$. We already shown that there exists a sequence $(X_n) \subset L^1([0,1])$ such that $X_n \xrightarrow{L^1} 0$ (whence $X_n \xrightarrow{\mathbb{P}} 0$) but $X_n(\omega)$ is pw convergent for no $\omega \in [0,1]$.

As we can see from the examples, convergence in probability is a very weak form of convergence.

8.4. Convergence in distribution

All types of convergence examined so far, namely, L^p convergence, convergence in probability, and almost sure convergence, involve directly the random variables X_n as functions on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$. As we know, random variables are perfectly known through their laws or, more practically, through their associated functions like cdfs and characteristic functions. For example, one could say that $X_n \longrightarrow X$ in some weak sense if

$$\mu_{X_n}(E) \longrightarrow \mu_X(E), \quad \forall E \in \mathscr{B}_{\mathbb{R}}, \tag{8.4.1}$$

or

$$\int_{\mathbb{R}} \varphi \, d\mu_{X_n} \longrightarrow \int_{\mathbb{R}} \varphi \, d\mu_X, \ \forall \varphi \in L^1. \tag{8.4.2}$$

The problem with such a definition is that, even a trivial sequence, as $X_n \equiv x_n$ with $x_n \longrightarrow x^*$ in $\mathbb{R}$, wouldn't be convergent to $X^* \equiv x^*$: assuming $x^* \neq x_n$ for every n, and taking $E = \{x^*\}$, we would have

$$\mu_{X_n}(\{x^*\}) = \mathbb{P}(X_n = x^*) = 0 \longrightarrow 0, \text{ but } \mu_{X^*}(\{x^*\}) = 1.$$

Similarly, if $x < x^*$, then being $x_n \longrightarrow x^*$, $x \leqslant x_n$ for n large, so $F_{X_n}(x) = 1 \longrightarrow 1$ but $F_{X^*}(x) = 0$ so

$$F_{X_n}(x) \not\longrightarrow F_{X^*}(x).$$

However, restricting the class of Borel sets in (8.4.1) or the class of functions φ in (8.4.2), we obtain an interesting definition:

Definition 8.4.1

Let $(X_n) \subset L(\Omega)$. We say that (X_n) **converges in distribution to** X, and we write $X_n \xrightarrow{d} X$, if

$$\int_{\mathbb{R}} \varphi \, d\mu_{X_n} \longrightarrow \int_{\mathbb{R}} \varphi \, d\mu_X, \; \forall \varphi \in \mathscr{C}_b(\mathbb{R}) \tag{8.4.3}$$

($\mathscr{C}_b(\mathbb{R})$ stands for the space of bounded and continuous function of $\mathbb{R}$).

Remark 8.4.2

Equivalently

$$X_n \xrightarrow{d} X, \quad \Longleftrightarrow \quad \mathbb{E}[\varphi(X_n)] \longrightarrow \mathbb{E}[\varphi(X)], \; \forall \varphi \in \mathscr{C}_b(\mathbb{R}).$$

Proposition 8.4.3

The following properties are equivalent:

i) $X_n \xrightarrow{d} X$.
ii) $F_{X_n}(x) \longrightarrow F_X(x)$, $\forall x \in \mathbb{R}$ where F_X is continuous.

Proof. i) $\Longrightarrow$ ii). Let x be a continuity point for F_X:

$$\lim_{y \to x-} F_X(y) = F_X(x) = \lim_{y \to x+} F_X(y).$$

Notice that

$$F_{X_n}(x) = \mu_{X_n}(]-\infty, x]) = \int_{\mathbb{R}} 1_{]-\infty,x]} \, d\mu_{X_n}.$$

Let now $\varphi_\varepsilon, \psi_\varphi \in \mathscr{C}_b(\mathbb{R})$ piecewise linear defined as

$$\varphi_\varepsilon(y) := \begin{cases} 1, & y \leqslant x - \varepsilon, \\ -\frac{1}{\varepsilon}(y - x + \varepsilon)) + 1, & x - \varepsilon \leqslant y \leqslant x, \\ 0, & y \geqslant x. \end{cases} \quad \psi_\varepsilon(y) := \begin{cases} 1, & y \leqslant x, \\ -\frac{1}{\varepsilon}(y - x) + 1, & x \leqslant y \leqslant x + \varepsilon, \\ 0, & y \geqslant x + \varepsilon. \end{cases}$$

Then

$$F_{X_n}(x) \leqslant \int_{\mathbb{R}} \psi_\varepsilon \, d\mu_{X_n}.$$

Since $X_n \xrightarrow{d} X$, and $\psi_\varepsilon \in \mathscr{C}_b$, we have

$$\int_{\mathbb{R}} \psi_\varepsilon \, d\mu_{X_n} \longrightarrow \int_{\mathbb{R}} \psi_\varepsilon \, d\mu_X \leqslant \int_{\mathbb{R}} 1_{]-\infty,x+\varepsilon]} \, d\mu_X = F_X(x+\varepsilon),$$

so, there exists $N = N(\varepsilon)$ such that

$$F_{X_n}(x) \leqslant F_X(x+\varepsilon) + \varepsilon, \; \forall n \geqslant N.$$

Similarly,

$$F_{X_n}(x) \geqslant \int_{\mathbb{R}} \varphi_\varepsilon \, d\mu_{X_n} \longrightarrow \int_{\mathbb{R}} \varphi_\varepsilon \, d\mu_X \geqslant \int_{\mathbb{R}} 1_{]-\infty,x-\varepsilon]} \, d\mu_X = F_X(x-\varepsilon).$$

So, for n large (we can always say $n \geqslant N$) we have

$$F_{X_n}(x) \geqslant F_X(x-\varepsilon) - \varepsilon, \; \forall n \geqslant N.$$

Therefore

$$F_X(x-\varepsilon) - F_X(x) - \varepsilon \leqslant F_{X_n}(x) - F_X(x) \leqslant F_X(x+\varepsilon) - F_X(x) + \varepsilon,\ \forall n \geqslant N,$$

so if $\ell := \lim_n (F_{X_n}(x) - F_X(x))$ we have

$$F_X(x-\varepsilon) - F_X(x) - \varepsilon \leqslant \ell \leqslant F_X(x+\varepsilon) - F_X(x) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, letting $\varepsilon \to 0$, and recalling that x is a continuity point $F_X(x \pm \varepsilon) \longrightarrow F_X(x)$, we get $0 \leqslant \ell \leqslant 0$, so $\ell = 0$ which is the conclusion.

ii) $\Longrightarrow$ i). The proof is a bit technical and omitted here.

An useful equivalent characterization is provided by the following result.

Theorem 8.4.4: continuity theorem

$$X_n \xrightarrow{d} X, \quad \Longleftrightarrow \quad \phi_{X_n}(\xi) \longrightarrow \phi_X(\xi),\ \forall \xi \in \mathbb{R}.$$

PROOF. $\Longrightarrow$ Take $\varphi(x) = e^{i\xi x}$ in (8.4.3) and we have the conclusion.
$\Longleftarrow$ The argument is similar to that one used in the proof of injectivity of FT for Borel probability 5.2. Let $\psi \in L^1(\mathbb{R})$. By duality

$$\int_{\mathbb{R}} \widehat{\psi}\, d\mu_{X_n} = \int_{\mathbb{R}} \psi(\xi)\phi_{X_n}(\xi)\, d\xi \longrightarrow \int_{\mathbb{R}} \psi(\xi)\phi_X(\xi)\, d\xi = \int_{\mathbb{R}} \widehat{\psi}\, d\mu,$$

by dominated convergence because i) $\psi(\xi)\phi_{X_n}(\xi) \longrightarrow \psi(\xi)\phi_X(\xi)$ a.e. $\xi \in \mathbb{R}$ and ii) $|\psi(\xi)\phi_{X_n}(\xi)| \leqslant |\psi(\xi)|$ a.e. $\xi \in \mathbb{R}$. So (8.4.3) holds for $\varphi = \widehat{\psi}$, with $\psi \in L^1(\mathbb{R})$. Arguing as in the proof of Theorem 5.2, we get that from this (8.4.3) extends to every $\varphi \in \mathscr{C}_b(\mathbb{R})$.

Example 8.4.5

Q. Let $X_n \sim \mathcal{N}(0, 1/n)$. *Then* $X_n \xrightarrow{d} 0$.

A. We have

$$\phi_{X_n}(\xi) = e^{-\frac{1}{2}\frac{1}{n}\xi^2} \longrightarrow 1 = \phi_0(\xi). \quad \square$$

Convergence in probability implies convergence in distribution.

Proposition 8.4.6

$$X_n \xrightarrow{\mathbb{P}} X, \quad \Longrightarrow \quad X_n \xrightarrow{d} X.$$

PROOF. We know

$$\lim_n \mathbb{P}(|X_n - X| \geqslant \varepsilon) = 0,\ \forall \varepsilon > 0.$$

For simplicity, we prove (8.4.3) for $\varphi \in \mathscr{C}_b^1(\mathbb{R})$ that is $\varphi, \varphi' \in \mathscr{C}_b(\mathbb{R})$. We write

$$\mathbb{E}[\varphi(X_n)] - \mathbb{E}[\varphi(X)] = \mathbb{E}\left[\varphi(X_n) - \varphi(X)\right] = \mathbb{E}\left[(\varphi(X_n) - \varphi(X))\,1_{|X_n - X| < \varepsilon} + (\varphi(X_n) - \varphi(X))\,1_{|X_n - X| \geqslant \varepsilon}\right].$$

Now,

$$\left|\mathbb{E}\left[(\varphi(X_n) - \varphi(X))\,1_{|X_n - X| \geqslant \varepsilon}\right]\right| \leqslant \mathbb{E}\left[|\varphi(X_n) - \varphi(X)|\,1_{|X_n - X| \geqslant \varepsilon}\right] \leqslant 2\|\varphi\|_\infty \mathbb{P}\left(|X_n - X| \geqslant \varepsilon\right).$$

Notice also that $|\varphi(x) - \varphi(y)| \leqslant \|\varphi'\|_\infty |x - y|$ so

$$\mathbb{E}\left[(\varphi(X_n) - \varphi(X))\,1_{|X_n - X| < \varepsilon}\right] \leqslant \mathbb{E}\left[\|\varphi'\|_\infty |X_n - X| 1_{|X_n - X| < \varepsilon}\right] \leqslant \|\varphi'\|_\infty \varepsilon.$$

Therefore,

$$|\mathbb{E}[\varphi(X_n)] - \mathbb{E}[\varphi(X)]| \leqslant \varepsilon\|\varphi'\|_\infty + 2\|\varphi\|_\infty \mathbb{P}\left(|X_n - X| \geqslant \varepsilon\right).$$

Since $\mathbb{P}(|X_n - X| \geqslant \varepsilon) \longrightarrow 0$, there exists N such that $\mathbb{P}(|X_n - X| \geqslant \varepsilon) \leqslant \varepsilon$, $\forall n \geqslant N$, so

$$|\mathbb{E}[\varphi(X_n)] - \mathbb{E}[\varphi(X)]| \leqslant \varepsilon\left(\|\varphi'\|_\infty + 2\|\varphi\|_\infty\right), \ \forall n \geqslant N,$$

and this means that

$$\mathbb{E}[\varphi(X_n)] \longrightarrow \mathbb{E}[\varphi(X)],$$

that is, $X_n \xrightarrow{d} X$.

Warning 8.4.7

The vice versa is false. Take, as usual, $(\Omega, \mathscr{F}, \mathbb{P}) = ([0,1], \mathscr{B}_{[0,1]}, \lambda_1)$ and define

$$X_n(\omega) = \begin{cases} (-1)^n, & \omega \in [0,1/2], \\ -(-1)^n, & \omega \in]1/2, 1]. \end{cases}$$

It is clear that $\mu_{X_n} \equiv \frac{1}{2}(\delta_{-1} + \delta_1)$, therefore $X_n \xrightarrow{d} X \sim B(-1,1,1/2)$. However, (X_n) is not convergent in probability. Indeed: $X_{2k} \equiv X = 1_{[0,1/2]} - 1_{]1/2,1]} \xrightarrow{\mathbb{P}} X$ while $X_{2k+1} \equiv Y = -1_{[0,1/2]} + 1_{]1/2,1]} \xrightarrow{\mathbb{P}} Y$. However, if $X_n \xrightarrow{\mathbb{P}} Z$, then, necessarily, $Z = X = Y$ but $\mathbb{P}(X = Y) = 0$. □

It is convenient to keep in mind the logical relations between various convergences:

$$\begin{array}{ccccc} X_n \xrightarrow{L^p} X & & & & \\ & \searrow & & & \\ \Downarrow\ (p = +\infty) & & X_n \xrightarrow{\mathbb{P}} X & \Longrightarrow & X_n \xrightarrow{d} X. \\ & \nearrow & & & \\ X_n \xrightarrow{a.s.} X & & & & \end{array}$$

8.5. Exercises

Exercise 8.5.1 (**). *Use Jensen's inequality to prove the inequality*

$$\exp\left(\int_\Omega \log X \, d\mathbb{P}\right) \leqslant \int_\Omega X \, d\mathbb{P},$$

for $X \geqslant 0$ $\mathbb{P}-$a.s and deduce the classical inequality between geometric and arithmetic means:

$$\sqrt[N]{x_1 \cdots x_N} \leqslant \frac{1}{N} \sum_{n=1}^{N} x_k, \quad \forall x_1, \ldots, x_N \in [0, +\infty[.$$

Exercise 8.5.2 ($*$)**.** *Let $X_n \sim \exp(n)$, $n \in \mathbb{N}$. Show that $X_n \xrightarrow{\mathbb{P}} 0$ for $n \to +\infty$.*

Exercise 8.5.3 ($**$)**.** *Let (X_n) be independent r.vs. with $X_n \sim B(0, 1, 1 - \frac{1}{n})$. Discuss L^p, a.s., $\mathbb{P}$ and d convergence of (X_n).*

Exercise 8.5.4 ($**$)**.** *Suppose that $\varepsilon_n > 0$ are such that $\sum_n \varepsilon_n < +\infty$ and $\mathbb{P}(|X_n| \geqslant \varepsilon_n) \leqslant \varepsilon_n$. Show that the series $\sum_n X_n$ is absolutely convergent with probability 1.*

Exercise 8.5.5 ($**+$)**.** *Let $X_n \xrightarrow{\mathbb{P}} X$ and $Y_n \xrightarrow{\mathbb{P}} Y$. Show that also $X_n + Y_n \xrightarrow{\mathbb{P}} X + Y$.*

Exercise 8.5.6 ($**$)**.** *Let (X_n) be a sequence of random variables with densities*

$$f_{X_n}(x) = \frac{1}{\pi} \frac{n}{1 + n^2 x^2}, \; x \in \mathbb{R}.$$

i) *Is $X_n \xrightarrow{d} 0$?*
ii) *Is $X_n \xrightarrow{\mathbb{P}} 0$?*
iii) *Assuming the X_n independent, is $X_n \xrightarrow{a.s.} 0$?*

Exercise 8.5.7 ($**+$)**.** *Let $(U_n) \sim U([0, 1])$ and $X_n := \min(U_1, \ldots, U_n)$.*

i) *Determine F_{X_n}.*
ii) *Discuss convergence of (nX_n).*

Exercise 8.5.8 ($**+$)**.** *Let X_n be i.i.d. random variables with*

$$\mathbb{P}(X_n > x) = \frac{1}{\sqrt{x}}, \; \forall x \geqslant 1.$$

Let $M_n := \max(X_1, \ldots, X_n)$.

i) *Determine the cdf of M_n.*
ii) *Discuss convergence in distribution of (M_n), identifying also its limit (if any).*

9

Limit Theorems

The law of large numbers (LLN) provides a mathematical foundation for the intuitive idea that, over a large number of repeated independent experiments X_k, the average outcome approximates the common expected value:

$$\overline{X}_n := \frac{1}{n}\sum_{k=1}^{n} X_k \longrightarrow m \equiv \mathbb{E}[X_k].$$

For example, in repeated coin tosses, the proportion of heads approaches 50% as the number of tosses increases. The LLN is a cornerstone of statistics, underpinning concepts such as sampling and estimation, and is widely applied in fields ranging from finance to physics and beyond.

If the X_k are i.i.d. random variables with common mean $m \equiv \mathbb{E}[X_k]$ and variance $\sigma^2 := \mathbb{V}[X_k]$, we have that $\overline{X}_n - m$ has mean 0 and variance

$$\mathbb{V}[\overline{X}_n - m] = \mathbb{E}\left[\left(\overline{X}_n - m\right)^2\right] = \frac{1}{n^2}\mathbb{E}\left[\left(\sum_{k=1}^{n}(X_k - m)\right)^2\right] = \frac{1}{n^2}\left(\sum_{k=1}^{n}\mathbb{V}[X_k] + \sum_{k\neq j}\mathrm{Cov}(X_k, X_j)\right).$$

Since the X_k are independent, $\mathrm{Cov}(X_k, X_j) = 0$ for $k \neq j$. So

$$\mathbb{V}[\overline{X}_n - m] = \frac{1}{n^2}\sum_{k=1}^{n}\sigma^2 = \frac{\sigma^2}{n}, \quad \Longleftrightarrow \quad \mathbb{V}\left[\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}(X_k - m)\right] = 1.$$

Therefore $\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}(X_k - m)$ is a r.v. with mean 0 and variance 1. It turns out that, for n large, no matter how the X_k are distributed, $\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}(X_k - m)$ takes more and more the shape of a standard Gaussian $\mathcal{N}(0,1)$. This happens, in general, in a very weak form as the convergence in distribution, that is

$$\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}(X_k - m) \xrightarrow{d} \mathcal{N}(0,1),$$

and this is known as the *Central Limit Theorem* (CLT), originally discovered by Bernoulli.

9.1. Weak Laws

There are many versions of the LLN, which differ in the way the sample average converges to the mean m. Broadly speaking, these results fall into two categories: *strong laws* (SLLN), where the convergence is almost sure, and *weak laws* (WLLN), where the convergence is weaker, typically convergence in L^1 or in probability, or even just in distribution. In general, the stronger the mode of convergence, the harder the proof. Here, for illustrative purposes, we will restrict attention to the proofs of the simplest cases.

9.1.1. Chebishev's WLLN. Let $(X_k) \subset L^1(\Omega)$ be a sequence of independent random variables with $\mathbb{E}[X_k] \equiv m$. We notice that, by replacing X_k with $X_k - m$ we may always assume that $m = 0$ because

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - m) = \frac{1}{n}\sum_{k=1}^{n} X_k - m.$$

The Chebishev WLLN is a simple result that assumes, one one hand, the more restrictive requirement $(X_n) \subset L^2$ but, on the other hand, it does not requires the (X_n) are identically distributed.

Proposition 9.1.1: Chebyshev

Let $(X_k) \subset L^2(\Omega)$ be independent r.vs. such that

- $\mathbb{E}[X_n] \equiv m$;
- $\mathbb{V}[X_n] \leqslant M$.

Then

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\mathbb{P}} m. \tag{9.1.1}$$

In particular, **Chebyshev's bound** holds:

$$\mathbb{P}\left(|\overline{X}_n - m| \geqslant \varepsilon\right) \leqslant \frac{1}{\varepsilon^2 n^2}\sum_{k=1}^{n}\mathbb{V}[X_k]. \tag{9.1.2}$$

Proof. Assuming $m = 0$ and setting

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k,$$

by Chebishev's inequality,

$$\mathbb{P}\left(|\overline{X}_n| \geqslant \varepsilon\right) \leqslant \frac{1}{\varepsilon^2}\mathbb{E}[|\overline{X}_n|^2].$$

Now,

$$\begin{aligned}\mathbb{E}[|\overline{X}_n|^2] &= \mathbb{E}\left[\left|\frac{1}{n}\sum_{k=1}^{n} X_k\right|^2\right] = \frac{1}{n^2}\sum_{k,j}\mathbb{E}\left[X_k X_j\right] = \frac{1}{n^2}\left(\sum_{k=1}^{n}\mathbb{E}[X_k^2] + \sum_{k\neq j}\mathbb{E}[X_k]\mathbb{E}[X_j]\right)\\ &= \frac{1}{n^2}\sum_{k=1}^{n}\mathbb{V}[X_k] \leqslant \frac{M}{n}.\end{aligned}$$

Thus,

$$\mathbb{P}\left(|\overline{X}_n| \geqslant \varepsilon\right) \leqslant \frac{M}{\varepsilon^2 n} \longrightarrow 0, \quad \Longleftrightarrow \quad \overline{X}_n \xrightarrow{\mathbb{P}} 0.$$

Chebyshev's bound is sometimes used to determine the number n of observations of a random variable such that the mismatch of the average $\overline{X}_n$ to the mean value m by an error ε has sufficiently small probability.

Example 9.1.2

Q. *Let* $(X_k) \subset L^2(\Omega)$ *with* $\mathbb{E}[X_k] \equiv m \geqslant 10$ *(unknown) and* $\mathbb{V}[X_k] \equiv 2$. *Determine n in such a way that the probability that* $\overline{X}_n$ *mismatch m by more than* 1% *of m be less than* 1%.

A. We have to determine n in such a way that

$$\mathbb{P}\left(|\overline{X}_n - m| \geqslant \frac{m}{100}\right) \leqslant \frac{1}{100}.$$

By Chebyshev's bound,

$$\mathbb{P}\left(|\overline{X}_n - m| \geqslant \frac{m}{100}\right) \leqslant \frac{1}{\left(\frac{m}{100}\right)^2 n^2} \sum_{k=1}^n \mathbb{V}[X_k] = \frac{10^4}{m^2 n^2} n\sigma^2 = \frac{2 \times 10^4}{m^2 n} \overset{m \geqslant 10}{\leqslant} \frac{2 \times 10^4}{10^2 n} = \frac{2 \times 10^2}{n}$$

Imposing

$$\frac{2 \times 10^2}{n} \leqslant \frac{1}{10^2}, \quad \Longleftrightarrow \quad n \geqslant 2 \times 10^4,$$

we get that for $n = 2 \times 10^4 = 20.000$ we have the desired bound.

Despite its weak form and simplicity, Chebishev L^2-WLLN has some remarkable applications.

9.1.2. Monte-Carlo approximation method. Consider the problem of computing an integral as

$$\int_0^1 f(x)\, dx,$$

for a continuous function f. We know that the definition is based on the idea that

$$\int_0^1 f(x)\, dx \approx \sum_{k=1}^n f(x_k) d\, dx_k,$$

where $\{x_k\} \subset [0,1]$ and $dx_k = x_{k+1} - x_k$. If points x_k divide $[0,1]$ in n equal parts, $x_k = \frac{k}{n}$ we have

$$\int_0^1 f(x)\, dx \approx \sum_{k=1}^n f(x_k)\frac{1}{n} = \frac{1}{n}\sum_k f(x_k).$$

Imagine now that x_k are outcomes of n independent random variables U_k, $k = 1, \dots, n$ with uniform distribution in $[0,1]$. It seems reasonable that, for n large,

$$\frac{1}{n}\sum_{k=1}^n f(U_k) \approx \int_0^1 f(x)\, dx.$$

This is a consequence of Čebishev weak law. Indeed, if U_n are independent, also $X_n := f(U_n)$ are independent. Moreover,

$$\mathbb{E}[X_n] = \mathbb{E}[f(U_n)] = \int_{\mathbb{R}} f(u) 1_{[0,1]}(u)\, du = \int_0^1 f(u)\, du,$$

while

$$\mathbb{V}[X_n] = \int_0^1 f(u)^2\, du - \left(\int_0^1 f(u)\, du\right)^2 \leqslant \|f\|_\infty^2.$$

Thus, weak law applies and

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n} f(U_k) - \int_0^1 f(x)\,du\right| \geqslant \varepsilon\right) \leqslant \frac{\|f\|_\infty}{\varepsilon^2 n}$$

9.1.3. Weierstrass–Bernstein's theorem. Another nice application of Chebishev WLLN is an original proof of Weierstrass' polynomial approximation theorem due to Bernstein.

Theorem 9.1.3: Weierstrass–Bernstein

Let $f \in \mathscr{C}([0,1])$ and set

$$p_n(x) := \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

Then

$$p_n \xrightarrow{\|\cdot\|_\infty} f.$$

Polynomials p_n are called **Bernstein's polynomials** of f.

PROOF. Let $X_n \sim B(1,0,x)$ be i.i.d. Bernoulli random variables that is

$$\mathbb{P}(X_n = 0) = 1-x, \;\; \mathbb{P}(X_n = 1) = x.$$

If $S_n := \sum_{j=1}^n X_j$ then,

$$\mathbb{P}(S_n = k) = \binom{n}{k} x^k (1-x)^{n-k}, \;\; k = 0, \dots, n.$$

Notice that

$$\mathbb{E}[f(\overline{X}_n)] = \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\mathbb{P}(S_n = k) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k(1-x)^{n-k} =: p_n(x).$$

Since $\mathbb{E}[X_n] = x$ and $\mathbb{V}[X_n] = x(1-x)$, the assumptions of Chebishev's WLLN are verified. Therefore

$$\forall \delta > 0, \;\; \mathbb{P}\left(|\overline{X}_n - x| \geqslant \delta\right) \leqslant \frac{x(1-x)}{\delta^2 n} \leqslant \frac{1}{4\delta^2 n}, \;\; \forall x \in [0,1]. \tag{9.1.3}$$

We now assess $\|f - p_n\|_\infty$. Let $x \in [0,1]$ and notice that

$$\begin{aligned} f(x) - p_n(x) &= f(x) - \mathbb{E}[f(\overline{X}_n)] = \mathbb{E}\left[f(x) - f(\overline{X}_n)\right] \\ &= \mathbb{E}\left[(f(x) - f(\overline{X}_n))1_{|\overline{X}_n - x| < \delta}\right] + \mathbb{E}\left[(f(x) - f(\overline{X}_n))1_{|\overline{X}_n - x| \geqslant \delta}\right]. \end{aligned}$$

We now need a remarkable property of any $f \in \mathscr{C}([0,1])$ (Heine–Cantor's theorem): *f is uniformly continuous*, that is

$$\forall \varepsilon > 0, \; \exists \delta = \delta(\varepsilon) > 0, \; : \; |f(\xi) - f(\eta)| \leqslant \varepsilon, \;\; \forall \xi, \eta \in [0,1], \; : \; |\xi - \eta| \leqslant \delta.$$

From this,

$$\left|\mathbb{E}\left[(f(x) - f(\overline{X}_n))1_{|\overline{X}_n - x| < \delta}\right]\right| \leqslant \varepsilon \mathbb{E}\left[1_{|\overline{X}_n - x| < \delta}\right] \leqslant \varepsilon.$$

while

$$\left|\mathbb{E}\left[(f(x)-f(\overline{X}_n))1_{|\overline{X}_n-x|\geqslant\delta}\right]\right| \leqslant 2\|f\|_\infty \mathbb{E}\left[1_{|\overline{X}_n-x|\geqslant\delta}\right] = 2\|f\|_\infty \mathbb{P}\left(|\overline{X}_n-x|\geqslant\delta\right) \leqslant \frac{\|f\|_\infty}{2\delta^2 n}.$$

Therefore,

$$\|f-p_n\|_\infty = \max_{x\in[0,1]} |f(x)-p_n(x)| \leqslant \varepsilon + \frac{\|f\|_\infty}{2\delta^2 n}.$$

Now, since $\frac{\|f\|_\infty}{2\delta^2 n} \to 0$, there exists $N = N(\delta) = N(\varepsilon)$ such that $\frac{\|f\|_\infty}{2\delta^2 n} \leqslant \varepsilon$ for $n \geqslant N$, so

$$\|f-p_n\|_\infty \leqslant 2\varepsilon, \ \forall n \geqslant N,$$

and this precisely means the conclusion.

9.1.4. L^1 WLLN. The natural hypothesis we might expect under which

$$\frac{1}{n}\sum_{k=1}^n X_k \longrightarrow m, \quad \text{with } \mathbb{E}[X_k] \equiv m,$$

is $(X_n) \subset L^1$. In this case, the variance $\mathbb{V}[X_n]$ is generally not defined, so the Chebishev's WLLN does not apply. However, we have the following theorem:

Theorem 9.1.4: WLLN

Let $(X_n) \subset L^1(\Omega)$ be i.i.d. r.vs. If $m := \mathbb{E}[X_k]$ (constant in k), then

$$\frac{1}{n}\sum_{k=1}^n X_k \xrightarrow{L^1} m.$$

9.2. Strong laws

Let $(X_k) \subset L^1(\Omega)$ be independent with $\mathbb{E}[X_k] \equiv m$. Replacing X_k by $X_k - m$ we can always assume that $m = 0$. To ensure $\overline{X}_n \xrightarrow{a.s.} 0$, by (8.2.1) we need to check that

$$\mathbb{P}\left(\limsup_n \{|\overline{X}_n| \geqslant \varepsilon\}\right) = \mathbb{P}\left(\bigcap_N \bigcup_{n\geqslant N} |\overline{X}_n| \geqslant \varepsilon\right) = 0, \ \forall \varepsilon > 0. \tag{9.2.1}$$

A sufficient condition to make this true is provided by Borel–Cantelli's Lemma 8.2.1: if

$$\sum_n \mathbb{P}\left(|\overline{X}_n| \geqslant \varepsilon\right) < +\infty,$$

then (9.2.1) holds true. Under the extra assumption $(X_k) \subset L^2(\Omega)$, from Chebishev's bound (9.1.2) we have

$$\mathbb{P}(|\overline{X}_n| \geqslant \varepsilon) \leqslant \frac{1}{\varepsilon^2 n^2}\sum_{k=1}^n \mathbb{V}[X_k] = \frac{n\mathbb{V}[X_1]}{n^2\varepsilon} \equiv \frac{C}{n},$$

a bound which is essentially useless to prove convergence for $\sum_n \mathbb{P}(|\overline{X}_n| \geqslant \varepsilon)$.

9.2.1. L^4 SLLN. An appropriate bound for $\mathbb{P}(|\overline{X}_n| \geqslant \varepsilon)$ can be obtained under stronger integrability for the X_k:

Proposition 9.2.1

Let $(X_k) \subset L^4(\Omega)$ be independent random variables with $\mathbb{E}[X_n] \equiv m$ and $\mathbb{E}[X_k^4] \leqslant K$. Then

$$\frac{1}{n}\sum_{k=1}^n X_k \xrightarrow{a.s.} m.$$

PROOF. We assume $m = 0$. By Chebishev's inequality

$$\mathbb{P}(|\overline{X}_n| \geqslant \varepsilon) = \mathbb{P}\left(\left|\sum_{k=1}^n X_k\right| \geqslant n\varepsilon\right) \leqslant \frac{1}{n^4\varepsilon^4}\mathbb{E}\left[\left|\sum_{k=1}^n X_k\right|^4\right] = \frac{1}{n^4\varepsilon^4}\sum_{k_1,k_2,k_3,k_4=1}^n \mathbb{E}[X_{k_1}X_{k_2}X_{k_3}X_{k_4}].$$

Now, if one of the indexes k is different from the other three, say $k_1 \neq k_2, k_3, k_4$ then, by independence,

$$\mathbb{E}[X_{k_1}X_{k_2}X_{k_3}X_{k_4}] = \mathbb{E}[X_{k_1}]\cdot\mathbb{E}[X_{k_2}X_{k_3}X_{k_4}] = 0.$$

So,

$$\begin{aligned}\sum_{k_1,k_2,k_3,k_4=1}^n \mathbb{E}[X_{k_1}X_{k_2}X_{k_3}X_{k_4}] &= \sum_{k=1}^n \mathbb{E}[X_k^4] + 3\sum_{k_1,k_2=1}^n \mathbb{E}[X_{k_1}^2X_{k_2}^2] \\ &\overset{CS}{\leqslant} \sum_{k=1}^n \mathbb{E}[X_k^4] + 3\sum_{k_1,k_2=1}^n \mathbb{E}[X_{k_1}^4]^{1/2}\mathbb{E}[X_{k_2}^4]^{1/2} \leqslant nK + 3n^2K.\end{aligned}$$

Therefore,

$$\mathbb{P}\left(|\overline{X}_n| \geqslant \varepsilon\right) \leqslant \frac{1}{n^4\varepsilon^4}(3n^2+n)K \leqslant \frac{4K}{\varepsilon^4 n^2}.$$

From this, $\sum_n \mathbb{P}\left(|\overline{X}_n| \geqslant \varepsilon\right) < +\infty$, so the Borel-Cantelli Lemma applies, and we conclude.

Example 9.2.2: empirical probability

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $E \in \mathscr{F}$ be an event. Let $(X_k)_{k\geqslant 1}$ be a sequence of independent Bernoulli random variables with parameter $p = \mathbb{P}(E)$, i.e.

$$X_k \sim \text{Bernoulli}(p), \qquad p = \mathbb{P}(E).$$

Interpret $X_k(\omega)$ as the outcome of an experiment that equals 1 if $\omega \in E$ and 0 otherwise.

Clearly, $X_k \in L^4(\Omega)$, therefore by the L^4 SLLN we have

$$\frac{1}{n}\sum_{k=1}^n X_k \xrightarrow{\text{a.s.}} \mathbb{E}[X_1] = p = \mathbb{P}(E).$$

In words: *over a long sequence of independent trials, the empirical frequency of the event E converges almost surely to its theoretical probability.*

9.2.2. Kolmogorov's maximal inequality. The key ingredient to get a.s. convergence is to get a good bound for

$$\mathbb{P}\left(|\overline{X}_n| \geqslant \varepsilon\right) = \mathbb{P}\left(\left|\sum_{k=1}^{n} X_k\right| \geqslant n\varepsilon\right).$$

Let $S_n := \sum_{k=1}^{n} X_k$

Theorem 9.2.3: Kolmogorov's maximal inequality

Let $(X_k) \subset L^2(\Omega)$ be independent random variables with $\mathbb{E}[X_k] = 0$ for every k. Then

$$\mathbb{P}\left(\max_{j\in\{1,\dots,n\}} |S_j| \geqslant \alpha\right) \leqslant \frac{1}{\alpha^2}\mathbb{E}\left[1_{\max_{j=1,\dots,n}|S_j|\geqslant\alpha} S_n^2\right] \left(\leqslant \frac{1}{\alpha^2}\mathbb{E}[|S_n|^2]\right). \tag{9.2.2}$$

Proof. Notice that

$$\begin{aligned}\{\max_{j=1,\dots,n}|S_j| \geqslant \alpha\} &= \{|S_1| \geqslant \alpha\} \sqcup \{|S_1| < \alpha,\ |S_2| \geqslant \alpha\} \sqcup \{|S_1|, |S_2| < \alpha,\ |S_3| \geqslant \alpha\} \sqcup \dots \\ &= \bigsqcup_{j=1}^{n} \left\{|S_k| < \alpha,\ \forall k = 0,\dots,j-1,\ |S_j| \geqslant \alpha\right\} =: \bigsqcup_{j=1}^{n} E_j,\end{aligned}$$

where we defined $S_0 := 0$. Therefore,

$$\mathbb{P}\left(\max_{j=1,\dots,n}|S_j| \geqslant \alpha\right) = \sum_{j=1}^{n} \mathbb{P}(E_j).$$

Now, on E_j we have $1 \leqslant \frac{|S_j|^2}{\alpha^2}$, so

$$\mathbb{P}(E_j) = \mathbb{E}\left[1_{E_j}\right] \leqslant \frac{1}{\alpha^2}\mathbb{E}\left[1_{E_j} S_j^2\right].$$

We show that

$$\mathbb{E}\left[1_{E_j} S_j^2\right] = \mathbb{E}\left[1_{E_j} S_n^2\right],\ \forall j = 1,\dots,n.$$

Indeed,

$$S_n^2 = \left(S_j + \sum_{k=j+1}^{n} X_k\right)^2 = S_j^2 + 2S_j \sum_{k=j+1}^{n} X_k + \left(\sum_{k=j+1}^{n} X_k\right)^2 =: S_j^2 + 2S_jT_j + T_j^2.$$

Now, since T_j is independent of both S_j and 1_{E_j}, we have

$$\begin{aligned}\mathbb{E}\left[1_{E_j} S_n^2\right] &= \mathbb{E}\left[1_{E_j} S_j^2\right] + 2\mathbb{E}\left[1_{E_j} S_j T_j\right] + \mathbb{E}\left[1_{E_j} T_j^2\right] \\ &= \mathbb{E}\left[1_{E_j} S_j^2\right] + 2\mathbb{E}\left[1_{E_j} S_j\right] \underbrace{\mathbb{E}\left[T_j\right]}_{=0} + \mathbb{E}\left[1_{E_j} T_j^2\right] \\ &= \mathbb{E}\left[1_{E_j} S_j^2\right] + \mathbb{E}\left[1_{E_j} T_j^2\right] \\ &\geqslant \mathbb{E}\left[1_{E_j} S_j^2\right].\end{aligned}$$

Therefore,

$$\mathbb{P}\left(\max_{j=1,\dots,n}|S_j|\geqslant\alpha\right)\leqslant\frac{1}{\alpha^2}\sum_{j=1}^{n}\mathbb{E}\left[1_{E_j}S_n^2\right]=\frac{1}{\alpha^2}\mathbb{E}\left[\sum_{j=1}^{n}1_{E_j}S_n^2\right]=\frac{1}{\alpha^2}\mathbb{E}\left[1_{\max_j|S_j|\geqslant\alpha}S_n^2\right].$$

Corollary 9.2.4: Kolmogorov SLLN

Let $(X_n)\subset L^2(\Omega)$ be i.i.d. random variables, and let $m:=\mathbb{E}[X_n]$. Then

$$\frac{1}{n}\sum_{k=1}^{n}\xrightarrow{a.s.}m.$$

PROOF. As usual, we assume $m=0$. Let also $\sigma^2:=\mathbb{V}[X_k]$ (constant in k because the X_k are i.i.d.). For $k\in\mathbb{N}$ we define

$$A_k:=\left\{\max_{n=2^k,\dots,2^{k+1}}|\overline{X}_n|\geqslant\varepsilon\right\}\subset\left\{\max_{n=2^k,\dots,2^{k+1}}|S_n|\geqslant\varepsilon 2^k\right\}.$$

By Kolomogorov's maximal inequality (9.2.2), we have

$$\mathbb{P}(A_k)\leqslant\frac{1}{\varepsilon 2^{2k}}\mathbb{E}[|S_{2^{k+1}}|^2]=\frac{1}{\varepsilon 2^{2k}}\sum_{j=1}^{2^{k+1}}\mathbb{V}[X_j]=2\frac{\sigma^2}{\varepsilon^2}\frac{1}{2^k}.$$

Therefore, by Borel-Cantelli's Lemma 8.2.3

$$\mathbb{P}\left(\limsup A_k\right)=0,$$

that is

$$\mathbb{P}\left(\bigcup_N\bigcap_{k\geqslant N}\left\{\max_{n=2^k,\dots,2^{k+1}}|\overline{X}_n|\leqslant\varepsilon\right\}\right)=1,\ \forall\varepsilon>0.$$

Now, take $\varepsilon_m\downarrow 0$. For each m there exists $\mathbb{P}(E_m)=0$ event such that

$$\bigcup_N\bigcap_{k\geqslant N}\left\{\max_{n=2^k,\dots,2^{k+1}}|\overline{X}_n|\leqslant\varepsilon_m\right\}=\Omega\backslash E_m.$$

So, if $E:=\bigcup_m E_m$, $\mathbb{P}(E)=0$ and

$$\bigcap_m\bigcup_N\bigcap_{k\geqslant N}\left\{\max_{n=2^k,\dots,2^{k+1}}|\overline{X}_n|\leqslant\varepsilon_m\right\}=\Omega\backslash E.$$

From this it follows that $\overline{X}_n\xrightarrow{a.s.}0$. Indeed: let $\omega\in\Omega\backslash E$. For $\varepsilon>0$, there exists $\varepsilon_m\leqslant\varepsilon$. Since

$$\omega\in\bigcup_N\bigcap_{k\geqslant N}\left\{\max_{n=2^k,\dots,2^{k+1}}|\overline{X}_n|\leqslant\varepsilon_m\right\},\implies\exists N,\ |\overline{X}_n|\leqslant\varepsilon_m\leqslant\varepsilon,\ \forall n\geqslant 2^N.$$

From this the conclusion follows.

Refining further the proof of Kolmnogorov's SLLN, it is possible to prove the

Theorem 9.2.5: Khinchin SLLN

Let $(X_n) \subset L^1(\Omega)$ be i.i.d. random variables, and let $m := \mathbb{E}[X_n]$. Then

$$\overline{X}_n \xrightarrow{a.s.} m.$$

9.3. Central Limit Theorem

We already noticed that if $(X_k) \subset L^2(\Omega)$ are i.i.d. with mean m and variance σ^2, then

$$\frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} (X_k - m)$$

is a mean 0 and variance 1 random variable for every $n \in \mathbb{N}$. For large n and independently from the specific distribution of the X_k, the previous random variable approximates a standard Gaussian. This it the celebrated *Central Limit Theorem*:

Theorem 9.3.1

Let $(X_k) \subset L^2(\Omega)$ be i.i.d. random variables with $\mathbb{E}[X_k] \equiv m$ and $\mathbb{V}[X_k] = \sigma^2$. Then,

$$\frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} (X_k - m) \xrightarrow{d} \mathcal{N}(0,1).$$

Proof. Let $Z_n := \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} (X_k - m)$, and let $\phi(\xi) = \phi_{X_k - m}(\xi)$ (independent of n because of the i.i.d. assumption). Then

$$\phi_{Z_n}(\xi) = \mathbb{E}\left[e^{i\xi \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n}(X_k - m)}\right] = \mathbb{E}\left[\prod_{k=1}^{n} e^{i\frac{\xi}{\sigma\sqrt{n}}(X_k - m)}\right] \overset{indep.}{=} \prod_{k=1}^{n} \mathbb{E}\left[e^{i\frac{\xi}{\sigma\sqrt{n}}(X_k - m)}\right] = \phi\left(\frac{\xi}{\sigma\sqrt{n}}\right)^n.$$

Now,

$$\phi(\eta) = \phi(0) + \phi'(0)\eta + \frac{\phi''(0)}{2}\eta^2 + o(\eta^2),$$

and since $\phi(0) = 1$ and

$$\phi'(\eta) = \partial_\eta \mathbb{E}[e^{i\eta(X_1 - m)}] = \mathbb{E}\left[i(X_1 - m)e^{i\eta(X_1 - m)}\right], \quad \Longrightarrow \quad \phi'(0) = i\mathbb{E}[X_1 - m] = 0,$$

$$\phi''(\eta) = -\mathbb{E}[(X_1 - m)^2 e^{i\eta(X_1 - m)}], \quad \Longrightarrow \quad \phi''(0) = -\mathbb{E}[(X_1 - m)^2] = -\sigma^2,$$

we have

$$\phi(\eta) = 1 - \frac{\sigma^2}{2}\eta^2 + o(\eta^2).$$

Therefore

$$\phi_{Z_n}(\xi) = \left(1 - \frac{\sigma^2}{2}\frac{\xi^2}{n\sigma^2} + o\left(\frac{\xi^2}{n\sigma^2}\right)\right)^n = \left(1 - \frac{\xi^2}{2n} + o\left(\frac{\xi^2}{n}\right)\right)^n \longrightarrow e^{-\frac{\xi^2}{2}} = \phi_{\mathcal{N}}(\xi).$$

By the continuity theorem 8.4.4 the conclusion now follows.

Remark 9.3.2

In particular, since the standard normal cdf is continuous on $\mathbb{R}$, we have that

$$\lim_{n\to+\infty}\mathbb{P}\left(a\leqslant\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}(X_k-m)\leqslant b\right)=\int_a^b e^{-\frac{x^2}{2}}\frac{dx}{\sqrt{2\pi}},\ \forall a<b.$$

Example 9.3.3

A polling company needs to estimate the probability p of success for a certain candidate in an election where there are, besides him, two other candidates. The task is to determine how many people need to be surveyed so that, with at least 95% probability, the estimated percentage from the sample differs from p by less than 1%:

i) assuming p is completely unknown;
ii) knowing that $p < 30\%$.

Describe the survey using random variables X_n, which are identically distributed, with $X_n = 0$ if the preference is not for the candidate, and $X_n = 1$ if it is and use the Čebishev bound.

9.4. Exercises

Exercise 9.4.1 ($**$). *Let $(X_n) \subset L^1(\Omega)$ be independent random variables with $\mathbb{E}[X_n] \equiv m$. Use characteristic functions to show that*

$$\overline{X}_n \xrightarrow{d} m.$$

Exercise 9.4.2 ($**$). *Let (X_n) be i.i.d. random variables, uniformly distributed on $[-1,1]$. Show that*

$$\lim_n \mathbb{P}\left((1-\varepsilon)\sqrt{\frac{n}{3}}\leqslant|(X_1,\dots,X_n)|\leqslant(1+\varepsilon)\sqrt{\frac{n}{3}}\right)=1,\ \forall\varepsilon>0.$$

(here $|(x_1,\dots,x_n)| = \sqrt{x_1^2+\cdots+x_n^2}$ is the Euclidean norm of $\mathbb{R}^n$).

Exercise 9.4.3 ($**$). *Let (X_n) be independent and such that*

$$\mathbb{P}(X_n=\pm n)=\frac{1}{2n\log n},\ \ \mathbb{P}(X_n=0)=1-\frac{1}{2n\log n}.$$

Show that

i) $\overline{X}_n \xrightarrow{\mathbb{P}} 0$ *(use Chebishev's lemma)*
ii) $\overline{X}_n$ *does not converge a.s. (use the second Borel-Cantelli statement).*

Exercise 9.4.4 ($**+$). *Let $(X_n) \subset L^1(\Omega)$ be i.i.d. random variables with $m = \mathbb{E}[X_n]$. Discuss a.s. convergence for*

$$\frac{1}{n}\sum_{k=1}^{n}X_kX_{k+1}.$$

Exercise 9.4.5 ($**+$). *Compute*

$$\lim_{n\to+\infty}\int_{[0,1]^n}\frac{x_1^2+\cdots+x_n^2}{x_1+\cdots+x_n}\,dx_1\cdots dx_n.$$

(hint: $dx_1\cdots dx_n = \mu_{(X_1,\dots,X_n)}$ with X_k i.i.d., $X_k \sim U([0,1])\dots$)

Exercise 9.4.6 $(**)$. *Let $(X_n) \subset L(\Omega)$ be i.i.d. random variables with common Cauchy distribution with density*

$$f(x) = \frac{1}{\pi}\frac{a}{a^2+x^2},$$

for some $a \neq 0$. Notice that

i) *Is $(X_n) \subset L^1$?*
ii) *By using the characteristic function, compute the distribution of $\overline{X}_n$.*
iii) *Discuss convergence of $\overline{X}_n$ in distribution.*

Exercise 9.4.7 $(**+)$. *Let (X_n) be i.i.d. random variables, $X_n \sim U[0,1]$. Use Central Limit Theorem to determine the limit*

$$\lim_{n\to\infty} \mathbb{P}\left(a \leqslant \left(\frac{e^{-n}}{X_1 \cdot X_2 \cdots X_n}\right)^{1/\sqrt{n}} \leqslant b \right), \quad 0 < a < b$$

Exercise 9.4.8 $(**+)$. *Let (X_n) be i.i.d. with $\mathbb{E}[X_k] \equiv 0$ and $\mathbb{E}[X_k^2] \equiv 1$. Show that*

$$\frac{\sum_{k=1}^n X_k}{\sqrt{\sum_{k=1}^n X_k^2}} \stackrel{d}{\longrightarrow} \mathcal{N}(0,1).$$

To conclude you need to use the following fact: if $Y_n \stackrel{d}{\longrightarrow} Y$ and $Z_n \stackrel{a.s.}{\longrightarrow} c \in \mathbb{R}$ (constant) then $Z_nY_n \stackrel{d}{\longrightarrow} Y$. (you can prove this as non trivial exercise).

Exercise 9.4.9 $(**+)$. *Let (X_n) be i.i.d. random variables, each with Poisson distribution*

$$\mathbb{P}(X_n = j) = \frac{1}{j!}e^{-1}.$$

i) *Determine the distribution of $S_n := \sum_{k=1}^n X_k$. In particular, determine $\mathbb{P}(S_n \leqslant n)$.*
ii) *Use the CLT to show that*

$$\lim_n e^{-n} \sum_{j=0}^n \frac{n^j}{j!} = \frac{1}{2}.$$

10

Martingales

10.1. Definitions

A martingale is a "time dependent" random variable for which *the best prediction of the future value is the present value*. Here, time can be either a *discrete time*, represented by $n \in \mathbb{N}$ or a *continuous time* $t \in \mathbb{R}$. For simplicity, here we will focus on the discrete time case.

Before we can dive into the definition of martingale, we have to introduce the definition of **filtration**. Informally, this is a time dependent family of $\sigma-$algebras that represent the "information" available up to a certain time. As time goes by, the information increases, this meaning that the $\sigma-$algebras of the filtration are nested:

Definition 10.1.1: filtration

A family $(\mathscr{F}_n)_{n\in\mathbb{N}}$ of $\sigma-$algebras of Ω is called **filtration** if

$$\mathscr{F}_m \subset \mathscr{F}_n, \ \forall m \leqslant n.$$

A fundamental case of filtration is that one generated by a one-parameter family of random variables:

Definition 10.1.2: natural filtration

Let $(X_n)n \geqslant 0$ be a one-parameter family of random variables on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. The **natural filtration** generated by (X_n) is

$$\mathscr{F}_n := \sigma\left(X_m \ : \ m \leqslant n\right).$$

We are now ready to introduce the main definition of this Chapter:

Definition 10.1.3: martingale

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, $(M_n)_{n\geqslant 0} \subset L^1(\Omega)$ be a one parameter family of L^1 random variables, and $(\mathscr{F}_n)_{n\geqslant 0}$ a filtration. We say that (M_n) is a **martingale** w.r.t. $(\mathscr{F}_n)$ if

$$\mathbb{E}\left[M_n \mid \mathscr{F}_m\right] = M_m, \ \forall n \geqslant m. \tag{10.1.1}$$

Proposition 10.1.4

The condition (10.1.1) is equivalent to

$$\mathbb{E}\left[M_{n+1} \mid \mathscr{F}_n\right] = M_n, \ \forall n. \tag{10.1.2}$$

PROOF. Indeed, it is clear that (10.1.2) is a particular case of (10.1.1). Vice versa, if (10.1.2) holds then, by the sub-conditioning property of the conditional expectation (property v) of Prop. 7.2.2), for $n > m$ we have

$$\mathbb{E}\left[M_n \mid \mathscr{F}_m\right] = \mathbb{E}\Big[\mathbb{E}[M_n \mid \mathscr{F}_{n-1}] \mid \mathscr{F}_m\Big] = \mathbb{E}\left[M_{n-1} \mid \mathscr{F}_m\right] = \cdots = \mathbb{E}\left[M_{m+1} \mid \mathscr{F}_m\right] = M_m.$$

Example 10.1.5: Doob martingale

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, $X \in L^1(\Omega)$. Let $(\mathscr{F}_n)_n$ be a filtration and define

$$M_n := \mathbb{E}[X \mid \mathscr{F}_n].$$

Then $(M_n)_n$ is a martingale w.r.t. $(\mathscr{F}_n)$.

PROOF. Since $\mathscr{F}_n \subset \mathscr{F}_{n+1}$, by sub-conditioning we have

$$\mathbb{E}[M_{n+1} \mid \mathscr{F}_n] = \mathbb{E}\big[\mathbb{E}[X \mid \mathscr{F}_{n+1}] \mid \mathscr{F}_n\big] = \mathbb{E}[X \mid \mathscr{F}_n] = M_n.$$

Example 10.1.6

Let $(\Omega, \mathscr{F}, \mathbb{P})$, $(X_k) \subset L^1(\Omega)$ be independent random variables. Define

$$M_n := \sum_{k=0}^{n} \left(X_k - \mathbb{E}[X_k]\right).$$

Then (M_n) is a martingale w.r.t. the natural filtration of (X_n).

PROOF. Let $\mathscr{F}_n := \sigma(X_1, \ldots, X_n)$. Since

$$M_{n+1} = \underbrace{M_n}_{\in \mathscr{F}_n} + (X_{n+1} - \mathbb{E}[X_{n+1}]),$$

we have

$$\mathbb{E}[M_{n+1} \mid \mathscr{F}_n] = M_n + \mathbb{E}\left[(X_{n+1} - \mathbb{E}[X_{n+1}]) \mid \mathscr{F}_n\right].$$

Since the X_k are independent, in particular $X_{n+1} - \mathbb{E}[X_{n+1}]$ is independent of $\mathscr{F}_n$ so

$$\mathbb{E}\left[(X_{n+1} - \mathbb{E}[X_{n+1}]) \mid \mathscr{F}_n\right] = \mathbb{E}\left[(X_{n+1} - \mathbb{E}[X_{n+1}])\right] = 0,$$

from which $\mathbb{E}[M_{n+1} \mid \mathscr{F}_n] = M_n$.

10.2. Super and sub martingales

We start with the following

Example 10.2.1: Gambler's wins process

We model a gambler who puts wages on an hazardous game. At $k-$th game, X_k represents the payoff for a unitary bet. For simplicity, we assume that X_k is a Bernoulli random variable with $\mathbb{P}(X_k = 1) = p$ (win), $\mathbb{P}(X_k = -1) = 1 - p$ (loss). We assume also that $k-$th game win is independent of the previous wins, that is X_k is independent of X_j $j = 1, \dots, k-1$. We call $Y_k \geqslant 0$ the *wage* on the $k-$th game. This is assumed to be random and non anticipative, that is Y_k, which is the wage put for the $k-$th game depends on what happened until the $k-1$-th game. In other words $Y_k \in \sigma(X_1, \dots, X_{k-1})$. If w is the gambler's initial fortune, the total win after n games is

$$G_n := w + Y_1 X_1 + \ldots + Y_n X_n.$$

Let us consider an example that illustrates why it is convenient to allow the wager process to be random. We start by betting the entire initial wealth w on the first game. After the first game we may either win w (in which case $G_1 = 2w$) or lose everything (so $G_1 = 0$). In the latter case, $Y_2 = 0$ (and consequently $Y_3 = Y_4 = \cdots = 0$), meaning that we effectively stop playing. If instead we win the first game, we may choose to bet $Y_2 = 2w$ in the second round. In other words,

$$Y_2 = \begin{cases} 0, & \text{if } X_1 = -1, \\ 2w, & \text{if } X_1 = 1. \end{cases}$$

After the second game, we might have $G_2 = 4w$ if we win again, or $G_2 = 0$ if we lose. For the next game, we could then decide:

$$Y_3 = \begin{cases} 0, & \text{if } X_1 = -1, \text{ or } (X_1 = 1,\ X_2 = -1), \\ 4w, & \text{if } X_1 = X_2 = 1. \end{cases}$$

We observe that, in this construction, Y_k depends on $X_1, \dots, X_{k-1}$.
Then

$$\mathbb{E}[G_{n+1} \mid \mathscr{F}_n] = G_n + \mathbb{E}\left[\underbrace{Y_{n+1}}_{\in \mathscr{F}_n} X_{n+1} \mid \mathscr{F}_n \right] = G_n + Y_{n+1}\mathbb{E}[X_{n+1} \mid \mathscr{F}_n]$$

Since X_{n+1} is independent of $X_1, \dots, X_n$ we have

$$\mathbb{E}[X_{n+1} \mid \mathscr{F}_n] = \mathbb{E}[X_{n+1}] = +1 \cdot \mathbb{P}(X_{n+1} = 1) + (-1) \cdot \mathbb{P}(X_{n+1} = -1) = p - (1-p) = 2p - 1.$$

In particular, for a *fair game* $p = 1/2$ (same probability to win and to loose) we have that

$$\mathbb{E}[X_{n+1} \mid \mathscr{F}_n] = 0,$$

so (G_n) is a martingale. In the more realistic case of an *unfair game*, that is $p < \frac{1}{2}$, we have

$$\mathbb{E}[G_{n+1} \mid \mathscr{F}_n] = G_n + \underbrace{Y_{n+1}}_{\geqslant 0}(2p-1) \leqslant G_n,$$

that is the best prediction on future wins is always worst than actual win.

The last case of the example yields to the

Definition 10.2.2

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, $(M_n)_{n\geqslant 0} \subset L^1(\Omega)$ be a one parameter family of L^1 random variables, and $(\mathscr{F}_n)_{n\geqslant 0}$ a filtration. We say that (M_n) is

- a **super-martingale** w.r.t. $(\mathscr{F}_n)$ if

$$\mathbb{E}\left[M_{n+1} \mid \mathscr{F}_n\right] \leqslant M_n, \ \forall n.$$

- a **sub-martingale** w.r.t. $(\mathscr{F}_n)$ if

$$\mathbb{E}\left[M_{n+1} \mid \mathscr{F}_n\right] \geqslant M_n, \ \forall n.$$

Proposition 10.2.3: Jensen's inequality

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, $\mathscr{G} \subset \mathscr{F}$ a sub-$\sigma-$algebra of $\mathscr{F}$. Then, for $X \in L^1$,

- if φ is convex, we have

$$\varphi\left(\mathbb{E}[X \mid \mathscr{G}]\right) \leqslant \mathbb{E}\left[\varphi(X) \mid \mathscr{G}\right]$$

- if φ is concave, we have

$$\varphi\left(\mathbb{E}[X \mid \mathscr{G}]\right) \geqslant \mathbb{E}\left[\varphi(X) \mid \mathscr{G}\right]$$

PROOF. For simplicity we assume $\varphi \in \mathscr{C}^1(\mathbb{R})$. If φ is convex we have

$$\varphi(y) \geqslant \varphi(x) + \varphi'(x)(y-x), \ \forall x, y \in \mathbb{R}.$$

Applying this with $x = \mathbb{E}[X \mid \mathscr{G}]$ and $y = X$ we get

$$\varphi(X) \geqslant \varphi(\mathbb{E}[X \mid \mathscr{G}]) + \varphi'(\mathbb{E}[X \mid \mathscr{G}])\left(X - \mathbb{E}[X \mid \mathscr{G}]\right).$$

and taking the conditional expectation we have

$$\begin{aligned} \mathbb{E}[\varphi(X) \mid \mathscr{G}] &\geqslant \mathbb{E}\left[\underbrace{\varphi\left(\mathbb{E}[X \mid \mathscr{G}]\right)}_{\in \mathscr{G}} \mid \mathscr{G}\right] + \mathbb{E}\left[\underbrace{\varphi'(\mathbb{E}[X \mid \mathscr{G}])}_{\in \mathscr{G}}(X - \mathbb{E}[X \mid \mathscr{G}]) \mid \mathscr{G}\right]. \\ &= \varphi\left(\mathbb{E}[X \mid \mathscr{G}]\right) + \varphi'\left(\mathbb{E}[X \mid \mathscr{G}]\right) \underbrace{\mathbb{E}\left[X - \mathbb{E}[X \mid \mathscr{G}] \mid \mathscr{G}\right]}_{=0} \\ &= \varphi\left(\mathbb{E}[X \mid \mathscr{G}]\right). \end{aligned}$$

For the case of φ concave, φ is convex and the conclusion follows.

Corollary 10.2.4

Let (M_n) be a martingale w.r.t. $(\mathscr{F}_n)$. Then

- if φ is convex, then $(\varphi(M_n))$ is a sub-martingale.
- if φ is concave, then $(\varphi(M_n))$ is a super-martingale.

PROOF. If φ is convex we have

$$\mathbb{E}\left[\varphi(M_{n+1}) \mid \mathscr{F}_n\right] \overset{Jensen}{\geqslant} \varphi\left(\mathbb{E}[M_{n+1} \mid \mathscr{F}_n]\right) = \varphi(M_n),$$

from which we deduce that $(\varphi(M_n))$ is a sub-martingale.

So, for instance, if (M_n) is a martingale, (M_n^2), (e^{M_n}) are sub-martingales (provided they make sense).

The concept of martingale can be extended to the case of continuous time dependent random variables. Here, we will limit to few definitions.

Definition 10.2.5: filtration

A family $(\mathscr{F}_t)_{t\geqslant 0}$ of $\sigma-$algebras of Ω is called **filtration** if

$$\mathscr{F}_s \subset \mathscr{F}_t, \ \forall s \leqslant t.$$

Definition 10.2.6: natural filtration

Let $(X_t)t \geqslant 0$ be a one-parameter family of random variables on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. The **natural filtration** generated by (X_t) is

$$\mathscr{F}_t := \sigma\left(X_s \ : \ s \leqslant t\right).$$

Definition 10.2.7: (super/sub)martingale

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, $(M_t)_{t\geqslant 0} \subset L^1(\Omega)$ be a one parameter family of L^1 random variables, and $(\mathscr{F}_t)_{t\geqslant 0}$ a filtration. We say that (M_t) is a

- **martingale** w.r.t. $(\mathscr{F}_t)$ if
$$\mathbb{E}\left[M_t \mid \mathscr{F}_s\right] = M_s, \ \forall t \geqslant s.$$
- **super-martingale** w.r.t. $(\mathscr{F}_t)$ if
$$\mathbb{E}\left[M_t \mid \mathscr{F}_s\right] \leqslant M_s, \ \forall t \geqslant s.$$
- **sub-martingale** w.r.t. $(\mathscr{F}_t)$ if
$$\mathbb{E}\left[M_t \mid \mathscr{F}_s\right] \geqslant M_s, \ \forall t \geqslant s.$$

10.3. Martingale transform

The example of gambler's wins process can be extended as follows. We start from the following

Definition 10.3.1

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, and $(\mathscr{F}_n)$ be a filtration. We say that $(X_n) \subset L(\Omega)$ is **non-anticipative** (or **adapted**) w.r.t. (to) $(\mathscr{F}_n)$ if $X_n \in \mathscr{F}_n$ for every $n \in \mathbb{N}$.

If we think $(\mathscr{F}_n)$ as the *information* available at time n, saying that (X_n) is non anticipative means that events like $\{X_n \in E\}$ do not depend on the future. We now introduce a general way to generate martingales based on a given one.

Proposition 10.3.2

Let $(M_n) \subset L^2$ is be a martingale w.r.t. the filtration $\mathscr{F}_n$ and $(X_n) \subset L^2$ be non anticipative w.r.t. $(\mathscr{F}_n)$. Let

$$Y_n := \sum_{k=0}^{n-1} X_k dM_k, \quad dM_k = M_{k+1} - M_k.$$

Then, $(Y_n) \subset L^1$ is a martingale w.r.t. $(\mathscr{F}_n)$, also called **martingale transform of X w.r.t. M.**

PROOF. Since $X_k, dM_k \in L^2$, by the Cauchy-Schwarz inequality $X_k dM_k \in L^1$, so $Y_n \in L^1$. We check that (Y_n) is a martingale w.r.t. $(\mathscr{F}_n)$. Since $X_n \in \mathscr{F}_n$, we have

$$\mathbb{E}[Y_{n+1} \mid \mathscr{F}_n] = \mathbb{E}[Y_n + X_n dM_n \mid \mathscr{F}_n] = Y_n + X_n \mathbb{E}[dM_n \mid \mathscr{F}_n],$$

and since (M_n) is a martingale w.r.t. $(\mathscr{F}_n)$,

$$\mathbb{E}[dM_n \mid \mathscr{F}_n] = \mathbb{E}[M_{n+1} - M_n \mid \mathscr{F}_n] = \mathbb{E}[M_{n+1} \mid \mathscr{F}_n] - M_n = 0,$$

so $\mathbb{E}[Y_{n+1} \mid \mathscr{F}_n] = Y_n$, that is (Y_n) is a martingale w.r.t. $(\mathscr{F}_n)$.

Under suitable conditions, a vice-versa also holds. Let ε_k be i.i.d. Bernoulli random variables with $\mathbb{P}(\varepsilon_k = 1) = p$, $\mathbb{P}(\varepsilon_k = -1) = 1 - p$ (here $0 < p < 1$). Let $\eta_k := \varepsilon_k - 2p + 1$, in such a way that $\mathbb{E}[\eta_k = 0$. We define $M_0 = 0$,

$$M_n := \sum_{k=1}^{n} \eta_k.$$

Let $\mathscr{F}_n := \sigma(\varepsilon_1, \ldots, \varepsilon_n) = \sigma(\eta_1, \ldots, \eta_n)$. Since the η_k are independent, (M_n) is a martingale.

Theorem 10.3.3

Let $(Y_n) \subset L^1$ be a martingale w.r.t. $(\mathscr{F}_n)$. Then, there exists a non anticipative $(X_n) \subset L^1$ such that

$$Y_n := Y_0 + \sum_{k=0}^{n-1} X_k dM_k.$$

PROOF. Since $Y_k \in \mathscr{F}_k$, we have

$$Y_k = \varphi_k(\varepsilon_1, \ldots, \varepsilon_k),$$

for some Borel function φ_k. Since $Y_k = \mathbb{E}[Y_{k+1} \mid \mathscr{F}_k]$, we have

$$\varphi_k(\varepsilon_1, \dots, \varepsilon_k) = \mathbb{E}[\varphi_{k+1}(\varepsilon_1, \dots, \varepsilon_{k+1}) \mid \varepsilon_1, \dots \varepsilon_k]$$

$$= \mathbb{E}[\varphi_{k+1}(\varepsilon_1, \dots, \varepsilon_k, 1) 1_{\varepsilon_{k+1}=1} \mid \varepsilon_1, \dots, \varepsilon_k] + \mathbb{E}[\varphi_{k+1}(\varepsilon_1, \dots, \varepsilon_k, -1) 1_{\varepsilon_{k+1}=-1} \mid \varepsilon_1, \dots, \varepsilon_k]$$

$$= \varphi_{k+1}(\varepsilon_1, \dots, \varepsilon_k, 1) \underbrace{\mathbb{E}[1_{\varepsilon_{k+1}=1} \mid \varepsilon_1, \dots, \varepsilon_k]}_{\overset{indep}{=} \mathbb{E}[1_{\varepsilon_{k+1}=1}] = \mathbb{P}(\varepsilon_{k+1}=1) = p} + \varphi_{k+1}(\varepsilon_1, \dots, \varepsilon_k, -1) \underbrace{\mathbb{E}[1_{\varepsilon_{k+1}=-1} \mid \varepsilon_1, \dots, \varepsilon_k]}_{= \mathbb{E}[1_{\varepsilon_{k+1}=-1}] = 1-p}$$

$$= p \varphi_{k+1}(\varepsilon_1, \dots, \varepsilon_k, 1) + (1-p) \varphi_{k+1}(\varepsilon_1, \dots, \varepsilon_k, -1).$$

Now,

$$Y_{k+1} - Y_k = \varphi_{k+1}(\varepsilon_1, \dots, \varepsilon_{k+1}) - \varphi_k(\varepsilon_1, \dots, \varepsilon_k).$$

For $\varepsilon_{k+1} = 1$,

$$dY_k = \varphi_{k+1}(\varepsilon_1, \dots, \varepsilon_k, 1) - \varphi_k(\varepsilon_1, \dots, \varepsilon_k) = (1-p) \left(\varphi_{k+1}(\varepsilon_1, \dots, \varepsilon_k, 1) - \varphi_{k+1}(\varepsilon_1, \dots, \varepsilon_k, -1) \right)$$

while, for $\varepsilon_{k+1} = -1$,

$$dY_k = \varphi_{k+1}(\varepsilon_1, \dots, \varepsilon_k, -1) - \varphi_k(\varepsilon_1, \dots, \varepsilon_k) = -p \left(\varphi_{k+1}(\varepsilon_1, \dots, \varepsilon_k, 1) - \varphi_{k+1}(\varepsilon_1, \dots, \varepsilon_k, -1) \right).$$

Therefore, defining

$$X_k := \frac{\varphi_{k+1}(\varepsilon_1, \dots, \varepsilon_k, 1) - \varphi_{k+1}(\varepsilon_1, \dots, \varepsilon_k, -1)}{2} \in \mathscr{F}_k,$$

and recalling that $\eta_{k+1} = \varepsilon_{k+1} - 2p + 1 = 2(1-p)$ if $\varepsilon_{k+1} = 1$ and $= -2p$ if $\varepsilon_{k+1} = -1$, we just have

$$dY_k = X_k \eta_{k+1} = X_k dM_k.$$

From this,

$$Y_n - Y_0 = \sum_{k=0}^{n-1} dY_k = \sum_{k=0}^{n-1} X_k dM_k,$$

as stated.

10.4. Exercises

Exercise 10.4.1 ($*$)**.** *Let $(X_k) \subset L^1$ be independent with $\mathbb{E}[X_k] = 1$. Let $\mathscr{F}_n := \sigma(X_1, \dots, X_n)$, $\mathscr{F}_0 = \{\varnothing, \Omega\}$ and define*

$$M_0 := 1, \;\; M_n := \prod_{k=1}^n X_k.$$

i) *Check that $(M_n) \subset L^1$ is a martingale w.r.t. $(\mathscr{F}_n)$.*
ii) *Is it true that if $\mathbb{E}[X_k] > 1$, then (M_n) is a sub-martingale?*

Exercise 10.4.2 ($*$)**.** *Let (X_n) be a sub-martingale w.r.t. $(\mathscr{F}_n)$. Define*

$$Y_n := \max(X_n, a).$$

Show that also (Y_n) is a sub-martingale w.r.t $(\mathscr{F}_n)$.

Exercise 10.4.3 ($**$)**.** *Let $(X_k) \subset L^2(\Omega)$ be such that (S_n), $S_n = \sum_{k=1}^n X_k$ is a martingale w.r.t. $\mathscr{F}_n := \sigma(X_1, \dots, X_n)$. Show that*

$$\mathbb{E}[X_i X_j] = 0, \; \forall i \neq j.$$

Exercise 10.4.4 (∗∗). *Let* $(X_n) \subset L^1(\Omega)$ *be such that*

$$\mathbb{E}[X_{n+1} \mid \mathscr{F}_n] = aX_n + bX_{n-1},$$

with $0 < a, b < 1$ *and* $a+b = 1$ *and* $\mathscr{F}_n = \sigma(X_1, \dots, X_n)$. *Determine* α *in such a way that* $(\alpha X_n + X_{n-1})$ *be a martingale*

Exercise 10.4.5 (∗∗). *Let* X_n *be Bernoulli with* $\mathbb{P}(X_n = 1) = p$, $\mathbb{P}(X_n = -1) = 1 - p$ *and* $p \neq \frac{1}{2}$. *Define*

$$Y_n := \left(\frac{p}{1-p}\right)^{X_n}$$

Check that (Y_n) *is a martingale w.r.t. the natural filtration* $\mathscr{F}_n := \sigma(X_1, \dots, X_n)$.

Exercise 10.4.6 (∗∗). *Let* (X_k) *be i.i.d. random variables with* $\mathbb{E}[X_k] \equiv 0$ *and* $\mathbb{V}[X_k] \equiv \sigma^2$. *Check that*

$$M_n := \left(\sum_{k=1}^{n} X_k\right)^2 - n\sigma^2$$

is a martingale w.r.t. the natural filtration of (X_k).

Exercise 10.4.7 (∗∗+). *Let* $(M_n) \subset L^2$ *be a martingale w.r.t.* $(\mathscr{F}_n)$.

i) *Check that if* $k < m < n$ *then* $\mathbb{E}[(M_n - M_m)M_k] = 0$.
ii) *Check that* $\mathbb{E}[(M_n - M_n)^2 \mid \mathscr{F}_k] = \mathbb{E}[M_n^2 \mid \mathscr{F}_k] - \mathbb{E}[M_m^2 \mid \mathscr{F}_k]$.
iii) *Check that* $\exists \lim_n \mathbb{E}[M_n^2] \leqslant +\infty$.
iv) *Show that if* $\mathbb{E}[M_n^2] \leqslant K < +\infty$ *for every* $n \in \mathbb{N}$, *then necessarily* (M_n) *is convergent in* L^2 *norm when* $n \to +\infty$.

Exercise 10.4.8 (∗∗+). *Let* (Z_k) *be independent random variables with*

$$\mathbb{P}(Z_n = \pm a_n) = \frac{1}{2n^2}, \;\; \mathbb{P}(Z_n = 0) = 1 - \frac{1}{n^2},$$

where $a_1 = 2$, $a_n = 4\sum_{k=1}^{n-1} a_k$, $n \geqslant 2$.

i) *Check that* $M_n := \sum_{k=1}^{n} Z_k$ *is a martingale w.r.t.* $\mathscr{F}_n := \sigma(Z_1, \dots, Z_n)$.
ii) *Discuss a.s. limit of* (M_n).
iii) *What about* L^1 *convergence of* M_n*?*

Exercise 10.4.9 (∗∗). *At time* $n = 0$, *a nonempty urn contains b black and w white balls. On each subsequent day, a ball is chosen at random from the urn (each ball in the urn has the same probability of being picked) and then put back together with another ball of the same color. Therefore, at the end of day n, here are* $n + b + w$ *balls in the urn. Let* B_n *denote the number of black balls in the urn at day n, and let*

$$X_n := \frac{B_n}{b + w + n}.$$

Check that (X_n) *is a martingale w.r.t. to its natural filtration.*

11

Brownian Motion

11.1. Definition

The **Brownian Motion** (hereafter just BM) – equivalently, the **Wiener process** –, arises to describe the irregular movement of small particles suspended in a fluid, caused by incessant collisions with the fluid's molecules. Despite continuity, typical trajectories are highly irregular, with apparently random changes of direction, which makes them non differentiable. Empirical observations of the BM enlighten a number of main features:

- trajectories $\gamma = \gamma(t)$ are continuous function of t;
- increments $\gamma(t) - \gamma(s)$ are Gaussian, with mean 0 and variance proportional to $t - s$;
- consecutive increments, that is $\gamma(t) - \gamma(s)$ and $\gamma(s) - \gamma(r)$ with $r < s < t$ are independent.

A natural model is to look at trajectories as outcomes of some time-dependent random variable,

$$W = W(t, \omega) : [0, +\infty[\times \Omega \longrightarrow \mathbb{R}^d.$$

with the agreement that

- for $\omega \in \Omega$ fixed, $t \longmapsto W(t, \omega)$ is the *trajectory*;
- for $t \in [0, +\infty[$ fixed, $\omega \longmapsto W(t, \omega)$ is the *time t position*. The notation W_t or $W(t)$ is used for the random variable $W(t, \sharp)$.

Such type of functions, depending on a scalar (usually interpreted as "time") and on a random parameter ω are called *stochastic processes*. For technical simplicity, we will focus on $d = 1$, the one-dimensional BM.

Definition 11.1.1

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. A function $W : [0, +\infty[\times \Omega \longrightarrow \mathbb{R}$ is called BM if

i) $W_0 = 0$ $\mathbb{P}-$a.s.

ii) if $0 < t_1 < \ldots < t_n$,

$$(W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}}) \sim \mathscr{N}(0, \mathrm{diag}(t_1, t_2 - t_1, \ldots, t_n - t_{n-1}))$$

iii) $W_\sharp(\omega) \in \mathscr{C}([0, \infty[)$ $\mathbb{P}-$a.s.

11.2. Lévy-Ciesielski construction

The original Wiener's construction of the BM was based on a Fourier representation for the time derivative of W_t:

$$\partial_t W_t = \sum_n \langle \partial_t W_t, e_n \rangle e_n,$$

with (e_n) the classic trigonometric basis. The proof is complicate, but Lévy and Ciesielsky found a much easier way to do this by using the Haar basis. Let us recall that this is a basis for $L^2([0,1])$ made of functions

$$e_0(t) \equiv 1, \quad e_{k,n}(t) = \begin{cases} 2^{\frac{n-1}{2}}, & \frac{k-1}{2^n} \leqslant t < \frac{k}{2^n}, \\ -2^{\frac{n-1}{2}}, & \frac{k}{2^n} \leqslant t < \frac{k+1}{2^n}, \\ 0, & \text{otherwise.} \end{cases} \quad k = 1, \dots, 2^n - 1, \ k \text{ odd}, \ n \in \mathbb{N},$$

We set $\mathscr{I} := \{(k,n) \ : \ n \in \mathbb{N}, \ k = 1, \dots, 2^n - 1 \text{ odd}\}$.

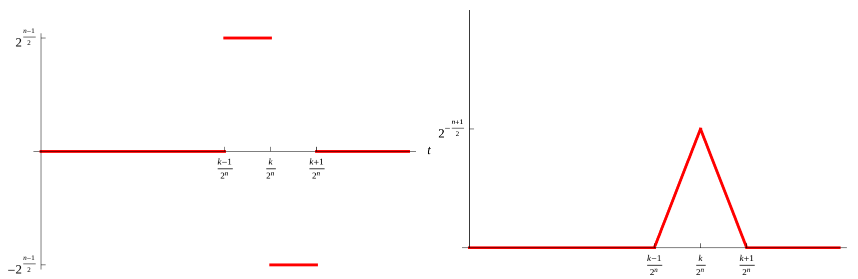


FIGURE 1. Haar's functions (left), Schauder's functions (right)

Now, if

$$\partial_t W_t = X_0 e_0(t) + \sum_{(k,n)\in\mathscr{I}} X_{k,n} e_{k,n}(t) \tag{11.2.1}$$

the Fourier coefficients X_0 and $X_{k,n}$ are

$$\begin{aligned} X_0 &= \langle \partial_t W_t, e_0 \rangle_2 = \int_0^1 \partial_t W \, dt \ = \ W_1 - W_0 = W_1, \\ X_{k,n} &= \langle \partial_t W_t, e_{k,n} \rangle_2 = 2^{\frac{n-1}{2}} \left(\int_{\frac{k-1}{2^n}}^{\frac{k}{2^n}} \partial_t W_t \, dt - \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} \partial_t W_t \, dt \right) \\ &= 2^{\frac{n-1}{2}} \left[W_{k/2^n} - W_{(k-1)/2^n} - \left(W_{(k+1)/2^n} - W_{k/2^n} \right) \right]. \end{aligned}$$

Now, assuming W_t is already defined, it is not difficult to check that the $X_{k,n}$ are independent r.vs. $X_{k,n} \sim \mathscr{N}(0,1)$. So, integrating the (11.2.1), we may expect that

$$W_t(\omega) = X_0(\omega) \int_0^t e_0(s) \, ds + \sum_{(k,n)\in\mathscr{I}} X_{k,n}(\omega) \int_0^t e_{k,n}(s) \, ds \tag{11.2.2}$$

It is conventient to introduce the so-called *Schauder functions*

$$s_0(t) := \int_0^t e_0 = \int_0^t 1 = t, \quad s_{k,n}(t) := \int_0^t e_{k,n} = \begin{cases} 0, & t \notin \left[\frac{k-1}{2^n}, \frac{k+1}{2^n}\right], \\ 2^{\frac{n-1}{2}} \left(t - \frac{k-1}{2^n}\right), & t \in \left[\frac{k-1}{2^n}, \frac{k}{2^n}\right], \\ -2^{\frac{n-1}{2}} \left(t - \frac{k}{2^n}\right) + \frac{1}{2^{\frac{n+1}{2}}}, & t \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right]. \end{cases}$$

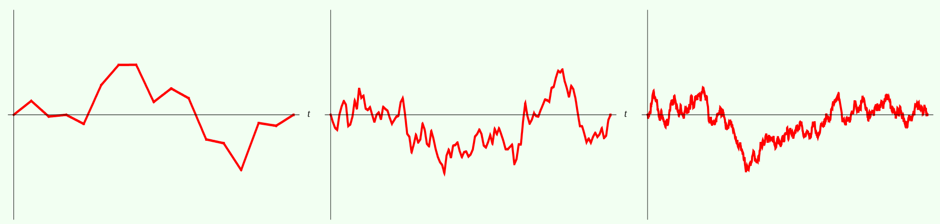
Clearly, Schauder-s functions are continuous on $[0, 1]$ so the series in (11.2.2) is made of continuous functions. We prove now that the series is convergent in uniform norm (of $\mathscr{C}([0, 1])$) with probability 1.

Theorem 11.2.1: Lévy–Ciesielski, 1961

Let $X_0, X_{k,n}$, $(k, n) \in \mathscr{I}$ be i.i.d. random variables $\mathscr{N}(0, 1)$ on some $(\Omega, \mathscr{F}, \mathbb{P})$. Let $s_0, s_{k,n}$, $(k, n) \in \mathscr{I}$ be the Schauder functions on $[0, 1]$. Then,

$$W_t(\omega) := X_0(\omega)s_0(t) + \sum_{(k,n)\in\mathscr{I}} X_{k,n}(\omega)s_{k,n}(t),\ t \in [0, 1],\ \omega \in \Omega, \tag{11.2.3}$$

is uniformly convergent with probability 1 and $(W_t)_{0\leqslant t\leqslant 1}$ fulfills the definition 11.1 in $[0, 1]$.



Proof. Let's start introducing the notation

$$S_n := \sum_{k<2^n,\ k\ odd} X_{k,n}s_{k,n}, \implies W = X_0 s_0 + \sum_{n=0}^{\infty} S_n.$$

Notice that every Schauder function is $\mathscr{C}$, therefore $S_n \in \mathscr{C}$. Our goal is to prove uniform convergence with probability 1, that is

$$\sum_n \|S_n\|_\infty < +\infty,\ \ \mathbb{P} - a.s.$$

The idea is to prove that

$$\mathbb{P}\left(\exists N\ :\ \|S_n\|_\infty \leqslant \frac{1}{n^2},\ \forall n \geqslant N\right) = 1,$$

or, equivalently

$$\mathbb{P}\left(\bigcap_N \bigcup_{n\geqslant N} \|S_n\|_\infty > \frac{1}{n^2}\right) = 0.$$

Applying Borel-Cantelli's Lemma, we are led to estimate $\mathbb{P}(\|S_n\|_\infty > \alpha)$. Now since $s_{h,n}$ and $s_{k,n}$ have disjoint supports for $h \neq k$ (h, k odd), we have

$$\|S_n\|_\infty = \frac{1}{2^{\frac{n+1}{2}}} \max_{k<2^n,\ k\ odd} |X_{k,n}|.$$

Hence

$$\mathbb{P}(\|S_n\|_\infty > \alpha) = \mathbb{P}\left(\max_{k<2^n,\ k\ odd} |X_{k,n}| > 2^{\frac{n+1}{2}}\alpha\right) \leqslant \sum_k \mathbb{P}\left(|X_{k,n}| > 2^{\frac{n+1}{2}}\alpha\right).$$

Since $X_{k,n}$ are all $\mathcal{N}(0,1)$, we have

$$\mathbb{P}(\|S_n\|_\infty > \alpha) \leqslant 2 \cdot 2^{n-1} \int_{2^{\frac{n+1}{2}}\alpha}^{+\infty} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}\, dy.$$

Notice now that,

$$\int_a^{+\infty} e^{-\frac{y^2}{2}}\, dy \leqslant \int_a^{+\infty} \frac{y}{a} e^{-\frac{y^2}{2}}\, dy = \frac{1}{a}\left[-e^{-\frac{y^2}{2}}\right]_{y=a}^{y=+\infty} = \frac{e^{-\frac{a^2}{2}}}{a}, \quad \forall a > 0.$$

so,

$$\mathbb{P}(\|S_n\|_\infty > \alpha) \leqslant \frac{2^n}{\sqrt{2\pi}} \frac{1}{2^{\frac{n+1}{2}}\alpha} e^{-\frac{2^{n+1}\alpha^2}{2}} = \frac{1}{\sqrt{2\pi}} \frac{2^{\frac{n}{2}}}{\alpha} e^{-2^n\alpha^2}.$$

We can now conclude: taking $\alpha = \frac{1}{n^2}$ in the previous estimate

$$\sum_n \mathbb{P}\left(\|S_n\|_\infty > \frac{1}{n^2}\right) \leqslant \frac{1}{\sqrt{2\pi}} \sum_n n^2 2^{n/2} e^{-\frac{2^n}{n^4}}$$

being the series clearly convergent (for instance, by root test we have $\left(n^2 2^{n/2} e^{-\frac{2^n}{n^4}}\right)^{1/n} = n^{2/n}\sqrt{2} e^{-\frac{2^n}{n^5}} \longrightarrow 0$). Therefore, Borel-Cantelli Lemma applies and the conclusion follows.

It remains to prove that W fulfills Def. 11.1 for $t \in [0,1]$. i) it is evident, ii) it follows by uniform convergence. For simplicity, we just prove that $W_t \sim \mathcal{N}(0,t)$. To check this we compute the characteristic function of W_t:

$$\mathbb{E}[e^{i\xi W(t)}] = e^{-\frac{\xi^2}{2t}}.$$

By construction

$$\mathbb{E}[e^{i\xi W(t)}] = \mathbb{E}\left[\lim_N e^{i\xi(X_0 s_0 + \sum_{n=0}^N S_n)}\right] \overset{Leb.}{=} \lim_N \mathbb{E}\left[e^{i\xi(X_0 s_0(t) + \sum_{n=0}^N S_n(t))}\right].$$

By independence

$$\mathbb{E}\left[e^{i\xi(X_0 s_0(t) + \sum_{n=0}^N S_n(t))}\right] = \mathbb{E}[e^{i\xi s_0(t) X_0}] \prod_{n=0}^N \prod_k \mathbb{E}[e^{i\xi s_{k,n}(t) X_{k,n}}],$$

and because every X_0 and $X_{k,n}$ is a standard gaussian $\mathcal{N}(0,1)$ we have

$$\mathbb{E}[e^{i\xi s_0(t) X_0}] = e^{-\frac{\xi^2 s_0(t)^2}{2}}, \quad \mathbb{E}[e^{i\xi s_{k,n}(t) X_{k,n}}] = e^{-\frac{\xi^2 s_{k,n}(t)^2}{2}},$$

so

$$\mathbb{E}[e^{i\xi W(t)}] = \lim_N e^{-\frac{\xi^2}{2}\left(s_0(t)^2 + \sum_{n=0}^N \sum_k s_{k,n}(t)^2\right)}.$$

To finish just notice that

$$s_{k,n}(t)^2 = \left(\int_0^t e_{k,n}(s)\, ds\right)^2 = \langle \chi_{[0,t]}, e_{k,n}\rangle_2^2, \overset{Parseval}{\Longrightarrow} \lim_N \left(s_0(t)^2 + \sum_{n=0}^N \sum_k s_{k,n}(t)^2\right) = \|\chi_{[0,t]}\|_2^2 = t,$$

and by this, finally, we get $\mathbb{E}[e^{i\xi W(t)}] = e^{-\frac{t\xi^2}{2}}$, that is $W(t) \sim \mathcal{N}(0,t)$.

To complete the construction of the BM we need to show that it can be defined on $t \geqslant 0$ and not only for $t \in [0,1]$. This can be done in the following way. Let (W_n) independent BMs on $[0,1]$ (this can be done by choosing countable copies of coefficients X for the series (11.2.3) in such a way they are independent). We define

$$W(t) := \begin{cases} W_0(t), & t \in [0,1], \\ W_0(1) + W_1(t-1), & t \in [1,2], \\ W_0(1) + W_1(1) + W_2(t-2), & t \in [2,3], \\ \vdots \end{cases}$$

It is easy to check that W is now a BM.

11.3. Exercises

Exercise 11.3.1 ($*$). *Let W be a BM. Show that, for $\lambda \neq 0$, $\widetilde{W}(t) := \lambda W(t/\lambda^2)$ is a BM.*

Exercise 11.3.2 ($**$). *Let W and $\widetilde{W}$ be two independent BMs and $\rho \in]-1,1[$ a constant. Define*

$$X_t = \rho W_t + \sqrt{1-\rho^2}\widetilde{W}_t,\ t \geqslant 0.$$

i) *Check that X_t is a BM.*
ii) *More in general, for which values a, b is $X_t = aW_t + b\widetilde{W}_t$ a BM?*

Exercise 11.3.3 ($**$). *Let (W_t) be a BM. Which of the following processes are still BM?*

i) $-W_t$
ii) $\sqrt{t}W_1$
iii) $W_{2t} - W_t$

Exercise 11.3.4 ($**$). *Let (W_t) be a BM, $\mathscr{F}_t$ its natural filtration.*

i) *Check that (W_t) is a martingale w.r.t. $(\mathscr{F}_t)$.*
ii) *Determine if (W_t^2) is a martingale/sub-martingale/super-martingale.*
iii) *Determine $f(t)$ in such a way that $W_t^2 - f(t)$ be a martingale.*

Exercise 11.3.5 ($**+$). *Show that, for $0 < s < t$,*

$$\mathbb{P}(W_s > 0, W_t > 0) = \frac{1}{4} + \frac{1}{2\pi}\arcsin\sqrt{\frac{s}{t}}.$$

Exercise 11.3.6 ($**$). *Let $X := \int_0^b W(t)^2\,dt$. Compute $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$. (hint: you can use Fubini's theorem to exchange $\mathbb{E}$ with $\int$ if required).*

Exercise 11.3.7 ($**+$). *Let $X = \int_0^b W(t)\,dt$. Determine the distribution of X. (hint: compute the characteristic function of X; you can use the approximation*

$$\int_0^b f(t)\,dt = \lim_n \frac{1}{n}\sum_k f\left(k\frac{b}{n}\right).$$

Exercise 11.3.8 ($**+$). *Let W be a BM. Show that $\widetilde{W}(t) := tW(1/t)$ if $t > 0$ and $\widetilde{W}(0) = 0$ is still a BM (the difficult part is the continuity at $t = 0+$).*

12

Brownian Paths

We explore some properties of brownian paths that emphasize their irregular character.

12.1. Length

An important feature of Brownian paths is that they have infinite lengths. We start recalling the concept of *length* of a curve $\gamma = \gamma_t : [a,b] \longrightarrow \mathbb{R}^d$:

$$\mathscr{L}_{[a,b]}(\gamma) := \sup_{\pi} \sum_k |\gamma_{t_{k+1}} - \gamma_{t_k}| =: \sup_{\pi} S_1(\gamma;\pi),$$

where $\pi = \{t_0 = a < t_1 < \ldots < t_n = b\}$ is a subdivision of $[a,b]$. We define

$$|\pi| := \max_k \{|t_{k+1} - t_k|\}.$$

It is an easy exercise to prove that if $\gamma \in \mathscr{C}([a,b])$ then

$$\mathscr{L}_{[a,b]}(\gamma) = \lim_{|\pi| \to 0} S_1(\gamma;\pi).$$

It is convenient to introduce also the *quadratic variation*

$$S_2(\gamma;\pi) := \sum_k |\gamma_{t_{k+1}} - \gamma_{t_k}|^2.$$

We start by proving the following

Lemma 12.1.1

$$S_2(W,\pi) \overset{L^2(\Omega)}{\longrightarrow} b - a \quad (|\pi| \longrightarrow 0)$$

Proof. Notice that

$$\begin{aligned} \|S_2(W;\pi) - (b-a)\|_2^2 &= \mathbb{E}\left[(S_2(W;\pi) - (b-a))^2\right] \\ &= \mathbb{E}\left[S_2(W;\pi)^2 - 2(b-a)S_2(W;\pi) + (b-a)^2\right] \\ &= \mathbb{E}[S_2(W;\pi)^2] - 2(b-a)\mathbb{E}[S_2(W;\pi)] + (b-a)^2. \end{aligned}$$

Now, by definition

$$\begin{aligned}
\mathbb{E}[S_2(W;\pi)^2] &= \mathbb{E}\left[\left(\sum_k (W_{t_{k+1}} - W_{t_k})^2\right)^2\right] \\
&= \sum_k \mathbb{E}\left[(W_{t_{k+1}} - W_{t_k})^4\right] + \sum_{h\neq k} \mathbb{E}\left[(W_{t_{k+1}} - W_{t_k})^2 (W_{t_{h+1}} - W_{t_h})^2\right] \\
&= 3\sum_k (t_{k+1} - t_k)^2 + \sum_{h\neq k} \mathbb{E}[(W_{t_{k+1}} - W_{t_k})^2]\mathbb{E}[(W_{t_{h+1}} - W_{t_h})^2] \\
&= 3\sum_k (t_{k+1} - t_k)^2 + \sum_{h\neq k} (t_{k+1} - t_k)(t_{h+1} - t_h) \\
&= 3\sum_k (t_{k+1} - t_k)^2 + \sum_h (T - (t_{h+1} - t_h))(t_{h+1} - t_h) \\
&= 2\sum_k (t_{k+1} - t_k)^2 + (b-a)^2.
\end{aligned}$$

Moreover

$$\mathbb{E}[S_2(W;\pi)] = \mathbb{E}\left[\sum_k (W_{t_{k+1}} - W_{t_k})^2\right] = \sum_k (t_{k+1} - t_k) = b - a,$$

so

$$\begin{aligned}
\mathbb{E}\left[(S_2(W;\pi) - (b-a))^2\right] &= 2\sum_k (t_{k+1} - t_k)^2 + (b-a)^2 - 2(b-a)^2 + (b-a)^2 = 2\sum_k (t_{k+1} - t_k)^2 \\
&\leqslant |\pi| \sum_k (t_{k+1} - t_k) = (b-a)|\pi|.
\end{aligned}$$

Therefore, if $|\pi| \longrightarrow 0$ then

$$\mathbb{E}\left[(S_2(\pi) - (b-a))^2\right] \leqslant (b-a)|\pi| \longrightarrow 0,$$

from which the conclusion follows. □

Proposition 12.1.2

$$\mathbb{P}\left(\mathscr{L}_{[a,b]}(W) = +\infty\right) = 1. \tag{12.1.1}$$

Proof. Notice that

$$\begin{aligned}
S_2(W;\pi) &= \sum_{k=1}^n (W_{t_{k+1}} - W_{t_k})^2 \leqslant \max_k |W_{t_{k+1}} - W_{t_k}| \sum_{k=1}^n |W_{t_{k+1}} - W_{t_k}| \\
&\leqslant \max_k |W_{t_{k+1}} - W_{t_k}| \mathscr{L}_{[a,b]}(W).
\end{aligned}$$

Since $W_\sharp \in \mathscr{C}([a,b])$ it is uniformly continuous (Heine-Cantor theorem), it means that

$$\forall \varepsilon > 0,\ \exists \delta(\varepsilon) > 0,\ :\ |t-s| \leqslant \delta(\varepsilon),\ \Longrightarrow\ |W_t - W_s| \leqslant \varepsilon,$$

so, in particular,

$$|\pi| \leqslant \delta(\varepsilon), \implies \max_k |W_{t_{k+1}} - W_{t_k}| \leqslant \varepsilon.$$

Thus, if $\mathscr{L}_{[a,b]}(W) < +\infty$, we would deduce that

$$S_2(W;\pi) \leqslant \mathscr{L}_{[a,b]}(W)\varepsilon,$$

and in particular,

$$\lim_{|\pi|\to 0} S_2(W;\pi) = 0.$$

However, from the Lemma we proved that $S_2(W;\pi) \xrightarrow{L^2} b-a$, hence, extracting a subsequence, $S_2(W;\pi) \longrightarrow b-a > 0$ $\mathbb{P}$−a.s.: in particular $S_2(W;\pi) \longrightarrow 0$ with $\mathbb{P}=0$, hence, necessarily $\mathscr{L}(W) < +\infty$ with $\mathbb{P}=0$. □

12.2. Regularity

We know that $W_t - W_s \sim \mathscr{N}(0, t-s)$, so, in particular

$$\mathbb{E}[(W_t - W_s)^2] = t - s.$$

This could suggest that $(W_t - W_s)^2 \approx t - s$, that is $|W_t - W_s| \approx |t-s|^{1/2}$. So, brownian paths would be more than continuous, but still non differentiable.

Definition 12.2.1

We say that $f \in \mathscr{C}([a,b])$ is $0 < \alpha \leqslant 1$ Hölder continuous (and we write $f \in \mathscr{C}^\alpha([a,b])$) on $[a,b]$ if

$$[f]_{\alpha,[a,b]} = \sup_{t\neq s\in[a,b]} \frac{|f_t - f_s|}{|t-s|^\alpha} < +\infty.$$

Case $\alpha = 1$ corresponds to Lipschitz continuous functions that, as known, are almost everywhere differentiable. If we expect that brownian paths are $1/2$ Hölder continuous, the following results won't be surprising.

Proposition 12.2.2

$$\mathbb{P}(W \in \mathscr{C}^\alpha) = 0, \quad \forall\, \alpha > \frac{1}{2}.$$

Proof. We notice that

$$\begin{aligned} S_2(W;\pi) &= \sum_k \left(\frac{|W_{t_{k+1}} - W_{t_k}|}{|t_{k+1}-t_k|^\alpha}\right)^2 |t_{k+1}-t_k|^{2\alpha} \leqslant [W_\sharp]^2_{\alpha,[a,b]} \sum_k |t_{k+1}-t_k|^{2\alpha} \\ &= [W_\sharp]^2_{\alpha,[a,b]} \sum_k |t_{k+1}-t_k||t_{k+1}-t_k|^{2\alpha-1} \overset{\alpha>1/2}{\leqslant} [W_\sharp]^2_{\alpha,[a,b]} |\pi|^{2\alpha-1} \sum_k |t_{k+1}-t_k| \\ &\leqslant [W_\sharp]^2_{\alpha,[a,b]} |\pi|^{2\alpha-1} |b-a|. \end{aligned}$$

So, if $[W_\sharp]_{\alpha,[a,b]} < +\infty$ then $S_2(W;\pi) \longrightarrow 0$. But we know that this happens almost never, so we conclude that $\mathbb{P}(W \in \mathscr{C}^\alpha) = \mathbb{P}([W(\cdot)]_{\alpha,[a,b]} < +\infty) = 0$. □

The same conclusion holds for $\alpha = \frac{1}{2}$ (see exercises) but the previous proof does not work. It is however true that paths are $\alpha < \frac{1}{2}$ Hölder continuous. To achieve this is much more complex. We will limit to sketch the argument.

The starting point is a remarkable inequality:

Lemma 12.2.3: Besov's inequality

Let $f \in \mathscr{C}([a,b])$, $p \geqslant 1$ and $\beta > \frac{1}{p}$. Then, there exists a constant $C = C(a,b,\beta,p)$ such that

$$|f(t) - f(s)| \leqslant C|t-s|^{\beta-1/p} \left(\int_a^b \int_a^b \frac{|f(u)-f(v)|^p}{|u-v|^{1+\beta p}}\, du\, dv \right)^{1/p}. \tag{12.2.1}$$

Accepting this inequality we have the

Theorem 12.2.4

$$\mathbb{P}\left(\mathscr{C}^\alpha([a,b])\right) = 1,\ \forall \alpha < \frac{1}{2}.$$

Proof. Let $\alpha < \frac{1}{2}$. The goal is to prove that

$$\mathbb{P}\left([W]_{\alpha,[a,b]} < +\infty\right) = 1.$$

To this aim notice that, from (12.2.1), we have

$$\frac{|W_t - W_s|}{|t-s|^{\beta-1/p}} \leqslant C \left(\int_a^b \int_a^b \frac{|W_u - W_v|^p}{|u-v|^{1+\beta p}}\, dudv \right)^{1/p}$$

so,

$$[W]_{\beta-1/p} \leqslant C \left(\int_a^b \int_a^b \frac{|W_u - W_v|^p}{|u-v|^{1+\beta p}}\, dudv \right)^{1/p}$$

Now, take $p = 2n$ (here $n \in \mathbb{N}$, $n \geqslant 1$): we have

$$[W]_{\beta-1/2n}^{2n} \leqslant C \int_a^b \int_a^b \frac{|W_u - W_v|^{2n}}{|u-v|^{1+\beta 2n}}\, dudv.$$

Taking expectations, and recalling that $W_u - W_v \sim \mathscr{N}(0, u-v)$ so, in particular,

$$\mathbb{E}[|W_u - W_v|^{2n}] = K_n |u-v|^n,$$

for some constant K_n, we have

$$\begin{aligned}\mathbb{E}\left[[W]^{2n}_{\beta-1/2n}\right] &\leqslant C\mathbb{E}\left[\int_a^b\int_a^b \frac{|W_u-W_v|^{2n}}{|u-v|^{1+2\beta n}}\,dudv\right] = C\int_a^b\int_a^b \mathbb{E}\left[\frac{|W_u-W_v|^{2n}}{|u-v|^{1+2\beta n}}\right]dudv\\ &= C\int_a^b\int_a^b \frac{\mathbb{E}\left[|W_u-W_v|^{2n}\right]}{|u-v|^{1+2\beta n}}\,dudv = C\int_a^b\int_a^b \frac{K_n|u-v|^n}{|u-v|^{1+2\beta n}}\,dudv\\ &= CK_n\int_a^b\int_a^b \frac{1}{|u-v|^{1+2\beta n-n}}\,dudv < +\infty \iff 1+2\beta n-n<1, \iff \beta<\frac{1}{2}\end{aligned}$$

So, in particular,

$$\mathbb{P}\left([W]_{\beta-1/2n} < +\infty\right) = 1,\ \forall \beta<\frac{1}{2},\ n\geqslant 1.$$

In conclusion, if $\alpha<\frac{1}{2}$ is fixed, piking β in such a way that $\alpha<\beta<\frac{1}{2}$ and n large enough such that $\beta-\frac{1}{2n}>\alpha$ (well possible because $\frac{1}{2n}\longrightarrow 0$), we have $[W]_\alpha \leqslant [W]_{\beta-1/2n}$, so

$$\mathbb{P}\left([W]_\alpha<+\infty\right) \geqslant \mathbb{P}\left([W]_{\beta-1/2n}<+\infty\right) = 1,$$

from which the conclusion finally follows. □

12.3. Differentiability

Since $\mathscr{C}^1$ functions (that is, continuous function together with their derivative) are easily $\mathscr{C}^\alpha$ functions for every $\alpha<1$, it follows that

$$\mathbb{P}\left(W_\sharp\in\mathscr{C}^1([a,b])\right) \leqslant \mathbb{P}\left(W_\sharp\in\mathscr{C}^\alpha([a,b])\right) = 0,\ \forall\frac{1}{2}<\alpha<1.$$

A slightly more general result can be easily achieved concerning the regularity of paths: paths are never differentiable with probability 1!

Proposition 12.3.1

$$\mathbb{P}\left(\{\omega\in\Omega : \exists\partial_t W_t(\omega)\}\right) = 0,\quad \forall t\geqslant 0.$$

Proof. We start recalling that

$$\partial_t W_t(\omega) = \lim_{h\to 0}\frac{W_{t+h}(\omega)-W_t(\omega)}{h}\in\mathbb{R},$$

so, in particular,

$$\exists\partial_t W(\omega),\ \Longrightarrow\ \exists L=L(\omega),\ \exists\delta_0=\delta_0(\omega)\ :\ \left|\frac{W_{t+h}(\omega)-W_t(\omega)}{h}\right|\leqslant L,\ \forall|h|\leqslant\delta_0.$$

In other words,

$$\{\exists\partial_t W_t\}\subset\bigcup_{L>0}\bigcup_{\delta_0>0}\bigcap_{|h|\leqslant\delta_0}\left\{\left|\frac{W_{t+h}-W_t}{h}\right|\leqslant L\right\}$$

The goal is to prove that, for $L > 0$ and $\delta_0 > 0$ fixed,

$$\mathbb{P}\left(\bigcap_{|h|\leqslant\delta_0}\left\{\left|\frac{W_{t+h}-W_t}{h}\right|\leqslant L\right\}\right)=0.$$

To make this a countable calculation, without loss of generality, we will actually show

$$\mathbb{P}\left(\bigcap_{n\geqslant N}\left\{\left|\frac{W_{t+1/n}-W_t}{1/n}\right|\leqslant L\right\}\right)=\mathbb{P}\left(\bigcap_{n\geqslant N}\left\{\left|W_{t+1/n}-W_t\right|\leqslant\frac{L}{n}\right\}\right)=0.$$

Now, $W_{t+1/n}-W_t\sim\mathcal{N}(0,\frac{1}{n})$ so

$$\begin{aligned}\mathbb{P}\left(\left|W_{t+1/n}-W_t\right|\leqslant\frac{L}{n}\right) &= \frac{\sqrt{2n}}{\sqrt{\pi}}\int_0^{\frac{L}{n}}e^{-\frac{y^2}{2\frac{1}{n}}}\,dy\\ &= \sqrt{\frac{2n}{\pi}}\int_0^{\frac{L}{n}}e^{-\frac{(\sqrt{n}y)^2}{2}}\,dy \overset{\sqrt{n}y=z}{=} \sqrt{\frac{2}{\pi}}\int_0^{\frac{L}{\sqrt{n}}}e^{-\frac{z^2}{2}}\,dz,\end{aligned}$$

so

$$\mathbb{P}\left(\bigcap_{n\geqslant N}\left\{\left|W_{t+1/n}-W_t\right|\leqslant\frac{L}{n}\right\}\right)\leqslant\sqrt{\frac{2}{\pi}}\int_0^{\frac{L}{\sqrt{n}}}e^{-\frac{z^2}{2}}\,dz,\ \forall n\geqslant N.$$

Letting $n\to+\infty$ we have the conclusion. □

Remark 12.3.2. What actually the previous proof shows is that the event $\{\omega\in\Omega\ :\ \exists\partial_t W_t(\omega)\}$ is a subset of a probability 0 set. If the probability $\mathbb{P}$ is complete (that is, a subset of a null event is a (null) event), then we conclude. □

12.4. Exercises

Exercise 12.4.1 ($**+$)**.** *Let $\pi_n := \{\frac{k}{n}\ :\ k=0,\dots,n\}$ be the subdivision on $[0,1]$ in n equal parts. Check that*

$$\sum_{k=0}^{n-1}W_{k/n}(W_{(k+1)/n}-W_{k/n})\xrightarrow{L^2}\frac{1}{2}\left(W_1^2-1\right).$$

Exercise 12.4.2 ($***$)**.** *Let π_n be a dyadic subdivision of $[a,b]$, that is*

$$\pi_n := \left\{a+\frac{k}{2^n}(b-a)\ :\ k=0,\dots 2^n\right\}.$$

Then

$$S_2(W;\pi_n)\xrightarrow{a.s.}b-a.$$

Warning: we proved that that $S_2(W;\pi)\xrightarrow{L^2(\omega)}b-a$, so there is an a.s. convergent subsequence. Here, one has to prove directly that $(S_2(W;\pi_n))$ converges pointwise a.s. (hint: try to express the set where $S_2(W;\pi_n)\not\longrightarrow b-a$). □

Exercise 12.4.3. *The goal is to prove that*

$$\mathbb{P}\left(W\in\mathscr{C}^{1/2}\right)=0.$$

Let π *be a subdivision of* $[a,b]$ *and set* $X_k := \frac{W(t_k)-W(t_{k-1})}{\sqrt{t_k-t_{k-1}}}$. *What kind of r.vs are the* X_k*? Hence, noticed that* $[W]_{\frac{1}{2},[a,b]} \geqslant \max_k |X_k|$, *show that*

$$\mathbb{P}\left([W]_{\frac{1}{2},[a,b]} \leqslant \lambda\right) = 0, \ \ \forall \lambda > 0$$

and conclude. □