TÉCNICO
LISBOA

Probability theory: an introduction

Patrícia Gonçalves

November 2017

TÉCNICO LISBOA

TÉCNICO LISBOA

Contents

Chapter 1

Measurable spaces

1.1 Set theory

Let Ω be an abstract space and we shall denote its elements by ω.

Definition 1.1.1 (Algebra).

An non-empty collection $\mathscr{F}$ of subsets of Ω is an algebra if and only if:

- $E \in \mathscr{F} \Rightarrow E^c \in \mathscr{F}$
- $E_1, E_2 \in \mathscr{F} \Rightarrow E_1 \cup E_2 \in \mathscr{F}$

Note that we shall refer to the first (resp. second) property above as saying that an algebra is stable for the complementary (resp. for finite unions).

Definition 1.1.2 (Monotone class).

An non-empty collection $\mathscr{F}$ of subsets of Ω is a monotone class if and only if:

- $E_j \in \mathscr{F}\,, E_j \subset E_{j+1}\,, \forall j \Rightarrow \cup_{j\geq 1} E_j \in \mathscr{F}$
- $E_j \in \mathscr{F}\,, E_j \supset E_{j+1}\,, \forall j \Rightarrow \cap_{j\geq 1} E_j \in \mathscr{F}$

Remark 1.1.3. *We will use the notation $E_j \uparrow$ for $E_j \subset E_{j+1}\,, \forall j$ and $E_j \downarrow$ for $E_j \supset E_{j+1}$.*

Definition 1.1.4 (σ-algebra).

An non-empty collection $\mathscr{F}$ of subsets of Ω is a σ-algebra if and only if:

- $E \in \mathscr{F} \Rightarrow E^c \in \mathscr{F}$
- $E_j \in \mathscr{F}$, $\forall j \Rightarrow \cup_{j\geq 1} E_j \in \mathscr{F}$

Note that we shall refer to the first (resp. second) property above as saying that a σ-algebra is stable for the complementary (resp. for countable unions).

Theorem 1.1.5. *An algebra is a σ-algebra if and only if it is a monotone class.*

Proof. Let us first prove that if $\mathscr{F}$ is an algebra and a monotone class, then $\mathscr{F}$ is a σ-algebra. For that it is enough to show that $\mathscr{F}$ is stable for the countable union. Take a collection $\{E_j\}_{j\geq 1} \in \mathscr{F}$ and for each $n \geq 1$ let $F_n = \cup_{j=1}^n E_j$. Since $\mathscr{F}$ is an algebra then $F_n \in \mathscr{F}$ and on the other hand $F_n \subset F_{n+1}$ and since $\mathscr{F}$ is a monotone class, we have that $\cup_{n\geq 1} F_n = \cup_{j\geq 1} E_j \in \mathscr{F}$.
Let us now prove that if $\mathscr{F}$ is an algebra and a σ-algebra, then $\mathscr{F}$ is a monotone class. For any collection $\{E_j\}_{j\geq 1} \in \mathscr{F}$ it holds that $\cup_{j\geq 1} E_j \in \mathscr{F}$ because $\mathscr{F}$ is a σ-algebra. Now for a collection $\{E_j\}_{j\geq 1} \in \mathscr{F}$ with $E_j \downarrow$ we have that

$$\cap_{j\geq 1} E_j = \Big(\cup_{j\geq 1} E_j^c \Big)^c.$$

Since $E_j \in \mathscr{F}$, then $E_j^c \in \mathscr{F}$, and since $\mathscr{F}$ is a σ-algebra we conclude that $\cup_{j\geq 1} E_j^c \in \mathscr{F}$. From this we get that $\Big(\cup_{j\geq 1} E_j^c \Big)^c \in \mathscr{F}$ and we are done. □

Example 1. *The collection $\mathscr{S}$ of all subsets of Ω is a σ-algebra and it is called the total σ-algebra. The collection $\{\emptyset, \Omega\}$ is a σ-algebra and is it called the trivial σ-algebra.*

Remark 1.1.6.

1. *If A is an index set and if for $\alpha \in A$, $\mathscr{F}_\alpha$ is a σ-algebra (or a monotone class), then $\cap_{\alpha\in A} \mathscr{F}_\alpha$ is a σ-algebra (or a monotone class).*

2. *Given a non empty collection of sets $\mathscr{C}$, there exists a minimal σ-algebra (or algebra or monotone class) containing $\mathscr{C}$, which consists in the intersection of all σ-algebras (or algebras or monotone classes) containing $\mathscr{C}$. There is at least one, namely the total σ-algebra $\mathscr{S}$. This σ-algebra (or algebra or monotone class) is called the σ-algebra generated by $\mathscr{C}$.*

Theorem 1.1.7. *Let $\mathscr{F}_0$ be an algebra, $\mathscr{C}$ the minimal monotone class containing $\mathscr{F}_0$ and $\mathscr{F}$ the minimal σ-algebra containing $\mathscr{F}_0$. Then $\mathscr{F} = \mathscr{C}$.*

Proof. To show the equality we will show first that $\mathscr{C} \subset \mathscr{F}$ and second that $\mathscr{F} \in \mathscr{C}$. Since a σ-algebra is an algebra and a monotone class, then $\mathscr{F}$ is a monotone class which contains $\mathscr{F}_0$ and since $\mathscr{C}$ is the minimal monotone class, we conclude that $\mathscr{C} \subset \mathscr{F}$. Now we will prove that $\mathscr{F} \subset \mathscr{C}$. For that purpose, it is enough to show that $\mathscr{C}$ is a σ-algebra and from the previous theorem it is enough to show that $\mathscr{C}$ is an algebra. We need to prove two things: $\mathscr{C}$ is stable for finite intersections and for the complementary. Let us define the following subsets of $\mathscr{C}$:

$$\begin{aligned} \mathscr{C}_1 &:= \{E \in \mathscr{C} : E \cap F \in \mathscr{C}, \quad \forall F \in \mathscr{F}_0\} \\ \mathscr{C}_2 &:= \{E \in \mathscr{C} : E \cap F \in \mathscr{C}, \quad \forall F \in \mathscr{C}\}. \end{aligned} \tag{1.1.1}$$

Note that from the definition of $\mathscr{C}_1$ and $\mathscr{C}_2$ above, we have that $\mathscr{C}_i \subset \mathscr{C}$ for $i = 1, 2$. Let us show first that $\mathscr{C}_1$ and $\mathscr{C}_2$ are monotone classes. We start with $\mathscr{C}_1$. Therefore, we need to show that for a collection $\{E_j\}_{j\geq 1} \in \mathscr{C}_1$ with $E_j \uparrow$ we have that $\cap_{j\geq 1} E_j \in \mathscr{C}_1$. This means that $\forall F \in \mathscr{F}_0$ we need to have

$$\Big(\cup_{j\geq 1} E_j\Big) \cap F \in \mathscr{C}.$$

Let us check that this is indeed true. Well,

$$\Big(\cup_{j\geq 1} E_j\Big) \cap F = \cup_{j\geq 1}(E_j \cap F).$$

Since $E_j \in \mathscr{C}_1$, then $E_j \cap F \in \mathscr{C}, \forall F \in \mathscr{F}_0$. But since $E_j \subset E_{j+1}$, this implies that $E_j \cap F \subset E_{j+1} \cap F$ and since $\mathscr{C}$ is a monotone class, then $\cup_{j\geq 1}(E_j \cap F) \in \mathscr{C}$. From this we conclude that $\mathscr{C}_1$ fulfils the first property for being a monotone class. The second property is proved in a completely similar way, just by taking into account that $F \cap \Big(\cap_{j\geq 1} E_j\Big) = \cap_{j\geq 1}(F \cap E_j)$. We can also do a similar argument to show that $\mathscr{C}_2$ is a monotone class. Now recall that $\mathscr{F}_0$ is an algebra so that $\mathscr{F}_0 \subset \mathscr{C}_1 \subset \mathscr{C}$. But then $\mathscr{C} \subset \mathscr{C}_1$ since $\mathscr{C}$ is the smallest monotone class that contains $\mathscr{F}_0$. From here we conclude that

$$\mathscr{C}_1 = \mathscr{C}.$$

This means that $\forall F \in \mathscr{F}_0$ and $E \in \mathscr{C}$ we have that $F \cap E \in \mathscr{C}$, which means that $\mathscr{F}_0 \subset \mathscr{C}_2$. But since $\mathscr{C}_2$ is a monotone class containing $\mathscr{F}_0$ we have that $\mathscr{C}_2 = \mathscr{C}$. This implies that $\forall E \in \mathscr{C}$ and $F \in \mathscr{C}$ it holds that $E \cap F \in \mathscr{C}$, so that $\mathscr{C}$ is stable for the intersection. Now let us prove that $\mathscr{C}$ is stable for the complementary. Let

$$\mathscr{C}_3 := \{E \in \mathscr{C} : E^c \in \mathscr{C}\}.$$

By definition $\mathscr{C}_3 \subset \mathscr{C}$. Let us now prove that $\mathscr{C}_3$ is a monotone class. For that purpose, let $\{E_j\}_{j\geq 1} \in \mathscr{C}_3$ with $E_j \uparrow$ and we have to show that $\cap_{j\geq 1} E_j \in \mathscr{C}_3$. For that purpose note that $\Big(\cup_{j\geq 1} E_j \Big)^c = \cap_{j\geq 1} E_j^c$. Now, since $E_j \subset E_{j+1}$, then $E_{j+1}^c \subset E_j^c$ and $\cap_{j\geq 1} E_j^c \in \mathscr{C}$, since $\mathscr{C}$ is a monotone class. Analogously, we have for $\{E_j\}_{j\geq 1} \in \mathscr{C}_3$ with $E_j \downarrow$ that $\Big(\cap_{j\geq 1} E_j \Big)^c = \cup_{j\geq 1} E_j^c \in \mathscr{C}$. From this we conclude that $\mathscr{C}_3$ is a monotone class. But since $\mathscr{F}_0 \in \mathscr{C}_3$ (because $\mathscr{F}_0$ is an algebra - so that it is stable for the complementary) and $\mathscr{F}_0 \subset \mathscr{C}$, we conclude that $\mathscr{C} \subset \mathscr{C}_3$, from where it follows that $\mathscr{C} = \mathscr{C}_3$. This means that $\forall E \in \mathscr{C}$ we have that $E^c \in \mathscr{C}$, so that $\mathscr{C}$ is stable for the complementary. This shows that $\mathscr{C}$ is an algebra and we are done.

□

1.2 Probability measure

Definition 1.2.1. *Let Ω be an abstract space and $\mathscr{F}$ a σ-algebra of subsets of Ω. A probability measure $\mathbb{P}(\cdot)$ in $\mathscr{F}$ is a function $\mathbb{P} : \mathscr{F} \longrightarrow [0,1]$ which satisfies the following properties:*

1. $\forall E \in \mathscr{F}, \mathbb{P}(E) \geq 0.$

2. *If $\{E_j\}_{j\geq 1}$ is a countable collection of disjoint sets of $\mathscr{F}$, then*

$$\mathbb{P}(\cup_{j\geq 1} E_j) = \sum_{j\geq 1} \mathbb{P}(E_j) \quad \textit{(countable additivity)}.$$

3. $\mathbb{P}(\Omega) = 1.$

The triple $(\Omega, \mathscr{F}, \mathbb{P})$ is called a probability space, Ω is called the sample space and its elements ω are called the sample points.

Exercise:

Prove, as a consequence of the previous definition, that

1. $\forall E \in \mathscr{F}, \mathbb{P}(E) \leq 1$.
2. $\mathbb{P}(\varnothing) = 0$.
3. $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$.
4. $\mathbb{P}(E \cup F) + \mathbb{P}(E \cap F) = \mathbb{P}(E) + \mathbb{P}(F)$.
5. $E \subset F \Rightarrow \mathbb{P}(E) = \mathbb{P}(F) - \mathbb{P}(F \setminus E) \leq \mathbb{P}(F)$.
6. Monotone property: If $E_j \uparrow E$ or $E_j \downarrow E$, then $\mathbb{P}(E_j) \to \mathbb{P}(E)$.
7. Boole's inequality $\mathbb{P}(\cup_{j\geq 1} E_j) \leq \sum_{j\geq 1} \mathbb{P}(E_j)$.

Recall that

1. If $\{E_j\}_{j\geq 1}$ is a countable collection of disjoint sets of $\mathscr{F}$, then
$$\mathbb{P}(\cup_{j\geq 1} E_j) = \sum_{j\geq 1} \mathbb{P}(E_j) \quad \text{(countable additivity)}.$$
2. When above we have a finite collection we say it is the finite additivity property.
3. If $E_j \downarrow \varnothing$, then $\mathbb{P}(E_j) \to 0 \quad$ (continuity).

Theorem 1.2.2. *The finite additivity and the continuity together are equivalent to countable additivity.*

Proof. Let us first show that countable additivity implies finite additivity (which is trivial) and continuity. Let $\{E_j\}_{j\geq 1} \in \mathscr{F}$ such that $E_j \downarrow \varnothing$. We have the following equality:
$$E_n = \cup_{k\geq n}(E_k | E_{k+1}) \cup \cap_{k\geq 1} E_k.$$
If $E_j \downarrow \varnothing$ then $\cap_{k\geq 1} E_k = \varnothing$, from where we get that
$$\mathbb{P}(E_n) = \mathbb{P}(\cup_{k\geq n} E_k | E_{k+1}) = \sum_{k\geq n} \mathbb{P}(E_k | E_{k+1}).$$

Note that in the last equality we used the countable additivity and the fact that the sets $\{E_k|E_{k+1}\}_{k\geq 1}$ are disjoint. Since the series above is convergent (since it equals $\mathbb{P}(E_n)$) we have that

$$\lim_{n\to\infty} \mathbb{P}(E_n) = 0$$

and continuity holds. Let us now prove that continuity and finite additive imply countable additivity. Let $\{E_k\}_{k\geq 1}$ be a collection of disjoint sets of $\mathscr{F}$. Then

$$F_{n+1} = \cup_{k\geq n+1} E_k \downarrow \emptyset,$$

since ... $\subset F_{n+3} \subset F_{n+2} \subset F_{n+1}$ and from the continuity we have that $\lim_{n\to\infty} \mathbb{P}(F_{n+1}) = 0$. Moreover, if finite additivity also holds, then

$$\begin{aligned} \mathbb{P}(\cup_{k\geq 1} E_k) &= \mathbb{P}(\cup_{k=1}^{n} E_k) + \mathbb{P}(\cup_{k\geq n+1} E_k) \\ &= \sum_{k=1}^{n} \mathbb{P}(E_k) + \mathbb{P}(F_{n+1}) \\ &\geq \sum_{k=1}^{n} \mathbb{P}(E_k). \end{aligned} \tag{1.2.1}$$

From this we conclude that the series $\sum_{k\geq 1} \mathbb{P}(E_k)$ is convergent since it is bounded from above by $\mathbb{P}(\cup_{k\geq 1} E_k)$. Sending $n \to \infty$ we have that $\mathbb{P}(\cup_{k\geq 1} E_k) \geq \sum_{k\geq 1} \mathbb{P}(E_k)$ and from Boole's inequality we also have that $\mathbb{P}(\cup_{k\geq 1} E_k) \leq \sum_{k\geq 1} \mathbb{P}(E_k)$, from where we obtain that

$$\mathbb{P}(\cup_{k\geq 1} E_k) = \sum_{k\geq 1} \mathbb{P}(E_k)$$

and countable additivity is proved. □

Let $\Lambda \in \Omega$. The trace of the σ-algebra $\mathscr{F}$ in Λ is the collection of all the sets of the form $\Lambda \cap F$, where $F \in \mathscr{F}$. It is easy to see that this is a σ-algebra that we denote by $\Lambda \cap \mathscr{F}$. Suppose now that $\Lambda \in \mathscr{F}$ and $\mathbb{P}(\Lambda) > 0$. Then we can define $\mathbb{P}_\Lambda$ in $\Lambda \cap \mathscr{F}$ in the following way: for any $E \in \Lambda \cap \mathscr{F}$:

$$\mathbb{P}_\Lambda(E) = \frac{\mathbb{P}(E)}{\mathbb{P}(\Lambda)}.$$

And $\mathbb{P}_\Lambda$ is a probability measure in $\Lambda \cap \mathscr{F}$. The triple $(\Lambda, \Lambda \cap \mathscr{F}, P_\Lambda)$ is called the trace of $(\Omega, \mathscr{F}, \mathbb{P})$ in Λ.

Example 2. *Discrete sample space*

Let $\Omega = \{w_j, j \in \mathbb{N}\}$ *and let* $\mathscr{F}$ *be the total* σ*-algebra in* $\mathscr{F}$*. Choose a sequence of numbers* $\{p_j, j \in \mathbb{N}\}$ *such that for all* $j \in \mathbb{N}$*,* $p_j \geq 0$ *and* $\sum_{j \in \mathbb{N}} p_j = 1$ *and let* $\mathbb{P} : \mathscr{F} \to [0,1]$ *defined on* $E \in \mathscr{F}$ *by*

$$\mathbb{P}(E) = \sum_{w_j \in E} p_j.$$

Show that $\mathbb{P}$ *is a probability measure and that all the probability measures on* $(\Omega, \mathscr{F})$ *are of the form above.*

Example 3. *Continuous sample spaces*

Let $\mathscr{U} = (0,1]$ *and let* $\mathscr{C} := \{(a,b] : 0 < a < b \leq 1\}$*,* $\mathscr{B}$ *the minimal* σ*-algebra containing* $\mathscr{C}$*, m the Lebesgue measure on* $\mathscr{B}$*. Then* $(\mathscr{U}, \mathscr{B}, m)$ *is a probability space. Analogously, consider in* $\mathbb{R}$ *the collection* $\mathscr{C}$ *of intervals of the form* $(a,b]$*,* $-\infty < a < b < +\infty$*. The algebra* $\mathscr{B}_0$ *generated by* $\mathscr{C}$ *consists of finite unions of disjoint sets of the form* $(a,b]$*,* $(-\infty, a]$ *or* $(b, +\infty)$*. The Borel* σ*-algebra is the* σ*-algebra, denoted hereafter by* $\mathscr{B}$*, generated by* $\mathscr{B}_0$ *or by* $\mathscr{C}$*. Not that the Lebesgue measure m in* $\mathbb{R}$ *is NOT a probability measure.*

1.3 Distribution function

Definition 1.3.1. *A distribution function* $F : \mathbb{R} \longrightarrow \mathbb{R}$ *is increasing, right continuous and satisfies* $\lim_{x \to -\infty} F(x) = 0$ *and* $\lim_{x \to \infty} F(x) = 1$ *.*

Example 4.

$F_1(x) = \mathbf{1}_{[0,+\infty)}(x)$.

$F_2(x) = \frac{1}{2}\mathbf{1}_{[0,\frac{1}{2})}(x) + \mathbf{1}_{[\frac{1}{2},+\infty)}$.

$F_3(x) = x\mathbf{1}_{[0,1)}(x) + \mathbf{1}_{[1,+\infty)}$.

$F_4(x) = x\mathbf{1}_{[0,\frac{1}{2})}(x) + \mathbf{1}_{[\frac{1}{2},+\infty)}$.

Pay attention to the graph of the functions.

Lemma 1.3.2. *Each probability measure μ in $\mathscr{B}$ defines a distribution function F through the following correspondence:*

$$\forall x \in \mathbb{R}, \mu((-\infty, x]) = F(x). \tag{1.3.1}$$

Proof. First note that $(-\infty, x] \in \mathscr{B}$, since $(-\infty, x] = \cap_{n\geq 1}(-\infty, x+1/n]$. Now let us prove that F is increasing. take $x_1 \leq x_2$. Then $(-\infty, x_1] \subset (-\infty, x_2]$. From property 5. of the probability measure μ given in the previous exercise, we have that

$$\mu((-\infty, x_1]) \leq \mu((-\infty, x_2]),$$

and this means that $F(x_1) \leq F(x_2)$. Let us now prove that F is right continuous. Let $\{x_n\}_{n\geq 1}$ be a sequence of real numbers such that $x_n \downarrow x$. Then $(-\infty, x_n] \downarrow (-\infty, x]$ and from the monotone property 6. of a probability measure given in the previous exercise, we have that

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

Analogously, if $x \downarrow -\infty$ (resp. $x \uparrow +\infty$), then $(-\infty, x] \downarrow \varnothing$ (resp. $(-\infty, x] \uparrow \mathbb{R}$) and again it follows that

$$\begin{aligned} &\lim_{x\to-\infty} F(x) = \lim_{x\to-\infty} \mu((-\infty, x]) = \mu(\varnothing) = 0 \\ &(\text{resp.} \quad \lim_{x\to+\infty} F(x) = \lim_{x\to+\infty} \mu((-\infty, x]) = \mu(\mathbb{R}) = 1.) \end{aligned} \tag{1.3.2}$$

□

Remark 1.3.3. *As a consequence of the previous lemma we have for $-\infty < a < b < \infty$ that*

- $\mu((a,b]) = F(b) - F(a);$ $\quad \mu([a,b)) = F(b^-) - F(a^-);$
- $\mu((a,b)) = F(b^-) - F(a);$ $\quad \mu([a,b]) = F(b) - F(a^-);$

To show the equalities above we need to write the sets that appear in each case in terms of sets of the form $(-\infty, x]$. For example,

$$\mu((a,b]) = \mu((-\infty, b] | (-\infty, a]) = \mu((-\infty, b]) - \mu((-\infty, a]) = F(b) - F(a).$$

For a dense subset D of $\mathbb{R}$ the correspondence given in (1.3.1) is determined for $x \in D$ or if in the previous equalities we take $a, b \in D$.

Theorem 1.3.4. *Each distribution function F determines a probability measure μ in $\mathscr{B}$ through one of the correspondences given above.*

Proof. Let us just give a sketch of the proof of this result. Let F be a given distribution function and let us define for $a, b \in \mathbb{R}$ with $a < b$ the weight $\mu((a,b]) = F(b) - F(a)$. Let us see that μ defined in this way is countable additive. For that purpose take $E_j = (a_j, b_j]$ disjoint and check that $\mu(\cup_{j\geq 1} E_j) = \sum_{j\geq 1} \mu(E_j)$. Now increase the domain of μ preserving the countable additivity, that is if $S = \cup_{j\geq 1}(a_j, b_j]$ then define $\mu(S) = \sum_{j\geq 1} \mu((a_j, b_j]) = \sum_{j\geq 1} F(b_j) - F(a_j)$. Here we have to be careful because S can have several representations and we have to check that the definition of $\mu(S)$ does not depend on the chosen representation. We also note at this point that any open interval (a,b) is in the extended domain, since $(a,b) = \cup_{j\geq 1}(a, b - 1/n]$. Now, since we are defining a measure in $\mathscr{B}$ plus the fact that any open set O can be written as the union of a countable collection of disjoint open intervals, that is $O = \cup_{j\geq 1}(a_j, b_j)$ and this representation is unique, we define $\mu(O) = \sum_{j\geq 1} \mu((a_j, b_j)) = \sum_{j\geq 1} F(b_j^-) - F(a_j)$. The notation $F(b^{\pm})$ denotes the lateral limits from the right $(+)$ or left $(-)$ of b. Up to now the measure μ is defined on open subsets of $\mathbb{R}$. For closed sets, we use the definition with the complementary, that is, if C is a closed subset of $\mathbb{R}$, then $\mu(C) = 1 - \mu(C^c)$ and C^c is open, so that $\mu(C^c)$ is well defined. Now, for $a \in \mathbb{R}$, we define $\mu(\{a\}) = \mu((-\infty, a]) - \mu((-\infty, a)) = F(a) - F(a^-)$. So, at this point we also know the value of μ in countable sets. But we still have work to do to characterize the measure in $\mathscr{B}$. Now we make use of the exterior measure μ^* and the interior measure μ_* defined on $S \in \mathscr{B}$ by

$$\mu^*(S) = \inf_{\substack{O \text{ open} \\ S \subset O}} \mu(O)$$

$$\mu_*(S) = \inf_{\substack{C \text{ closed} \\ C \subset S}} \mu(C).$$

Note that $\mu^*(S) \geq \mu_*(S)$, but the equality is not always true. When the values coincide, we denote it by $\mu(S)$ and say that S is measurable for F. We need to check that this definition agrees with the previous one, where μ had already been defined. The next tasks are: we need to verify that the measurable sets form a σ-algebra let us say $\mathscr{L}$ and that in $\mathscr{L}$ the measure μ is a probability measure. If we have this, then since $\mathscr{L}$ is a σ-algebra that contains all the sets of the form $(a, b]$, then $\mathscr{B} \subset \mathscr{L}$. Note that $\mathscr{L}$ can be bigger than $\mathscr{B}$ but this is fine, since then the restriction of μ to $\mathscr{B}$ is a probability measure and this is what we want to show. □

The question now is: Is this probability measure μ unique?

Theorem 1.3.5. *Let μ and ν be two probability measures defined in the same σ-algebra $\mathscr{F}$ generated by the algebra $\mathscr{F}_0$. If $\mu(E) = \nu(E)$ for any $E \in \mathscr{F}_0$ then $\mu = \nu$.*

Proof. Let $\mathscr{C} = \{E \in \mathscr{F} : \mu(E) = \nu(E)\}$. By hypothesis we have that $\mathscr{F}_0 \subset \mathscr{C}$. On the other hand we claim that $\mathscr{C}$ is a monotone class. To see that, take a collection $\{E_j\}_{j\geq 1}$ with $E_j \in \mathscr{C}$ such that $E_j \uparrow$ or $E_j \downarrow$, the proof being the same in each case. Now note that

$$\mu(\cup_{\geq 1} E_j) = \lim_{j\to\infty} \mu(E_j) = \lim_{j\to\infty} \nu(E_j) = \nu(\cup_{\geq 1} E_j).$$

Above we used in the first and third equalities the monotone property of probability measures and in the second equality we used the fact that $E_j \in \mathscr{C}$. From this we conclude that $\cup_{\geq 1} E_j \in \mathscr{C}$ and from a similar computation we also conclude that $\cap_{\geq 1} E_j \in \mathscr{C}$. Since $\mathscr{C}$ is also a algebra, from Theorem 1.1.5 we conclude that $\mathscr{C}$ is a σ-algebra form where it follows that $\mathscr{C} = \mathscr{F}$ and therefore μ and ν coincide in $\mathscr{F}$ as we wanted to prove. □

Now we are able to conclude the following result:

Theorem 1.3.6. *Given a probability measure μ in $\mathscr{B}$ there exists a unique distribution function F which satisfies $\mu((-\infty, x]) = F(x)\ \forall x \in \mathbb{R}$. Conversely, given a distribution function F, there exists a unique probability measure μ in $\mathscr{B}$ satisfying $\mu((-\infty, x]) = F(x)\ \forall x \in \mathbb{R}$.*

We shall call μ the probability measure of F and F the distribution function of μ.

Example 5. *Instead of* $(\mathbb{R}, \mathscr{B})$ *we can consider a restriction to a fixed interval* $[a, b]$. *As example take* $\mathscr{U} = [0, 1]$. *Let us see how to define the distribution function* F.

Let F *be a distribution function such that* $F(x) = 0$*, if* $x < 0$ *and* $F(x) = 1$*, if* $x \geq 1$. *The probability measure* μ *will have support* $[0, 1]$ *since* $\mu(-\infty, 0) = 0 = \mu(1, +\infty)$. *The trace of* $(\mathbb{R}, \mathscr{B}, \mu)$ *in* $\mathscr{U}$ *can be denoted by* $(\mathscr{U}, \mathscr{B}_{\mathscr{U}}, m)$*, where* $\mathscr{B}_{\mathscr{U}}$ *is the trace of* $\mathscr{B}$ *in* $\mathscr{U}$ *and any probability measure in* $\mathscr{B}_{\mathscr{U}}$ *can be seen as that trace. As example, we have the uniform distribution given by* F_3 *above.*

Definition 1.3.7. *An atom of a measure* μ *defined in* $\mathscr{B}$ *is a singleton* $\{x\}$ *such that* $\mu(\{x\}) > 0$.

Definition 1.3.8. *A measure is said to be atomic if and only if* μ *is zero on any set not containing any atom.*

Exercise:

Prove that if F is the distribution function of μ then

$$\mu(\{x\}) = F(x) - F(x^-).$$

Prove that μ is atomless (that is μ does not have atoms) if and only if F is continuous.

Let us now go for a small digression in monotone functions. For that purpose, let f be an increasing function defined on $\mathbb{R}$. This means that for all $x \leq y$ it holds $f(x) \leq f(y)$. Let us see some properties of these kind of functions.

1. Both lateral limits exist and are finite for any $x \in \mathbb{R}$:

$$\lim_{y \downarrow x} f(y) = f(x^+) \quad \text{and} \quad \lim_{y \uparrow x} f(y) = f(x^-).$$

2. When $x = \pm\infty$ the limits above exist but can be equal to $\pm\infty$.

3. The function is continuous (resp. right-continuous) at x if and only if the limits above are both (resp. $f(x^+)$ is) equal to $f(x)$.

4. We say that the function has a jump at x if the limits above exist but are different. The value $f(x)$ has to satisfy

$$f(x^-) \leq f(x) \leq f(x^+.)$$

5. When there is a jump at x, we say that x is a point of jump of f and $f(x^+) - f(x^-)$ is the size of the jump.

Lemma 1.3.9. *The set of jumps of f is countable (can be finite).*

Proof. To prove this, first associate to each point of jump x, the interval

$$I_x = (f(x^-), f(x^+)).$$

Then, if x' is another point of jump of f and $x < x'$, then there exists $\tilde{x}$ such that $x < \tilde{x} < x'$ and

$$f(x^+) \leq f(\tilde{x}) \leq f(x'^-).$$

As a consequence the intervals I_x and $I_{x'}$ are disjoint and can be consecutive if $f(x^+) = f(x'^-)$. Therefore we associate to the set of points of jump of f a collection of disjoint intervals in the range of f. Now, this collection is, at most, countable since each interval contains a rational number, so that the collection of intervals is in one-to-one correspondence with a certain subset of the rational numbers, being the latter countable. Since the set of points of jump of f is in one-to-one correspondence with the set of intervals associated with it, then the proof ends. □

Example 6. *Let $\{a_n\}_{n\geq 1}$ be any given enumeration of the rational numbers and let $\{b_n\}_{n\geq 1}$ be a sequence of non-negative real numbers such that*

$$\sum_{n\geq 1} b_n < +\infty.$$

Consider

$$f(x) = \sum_{n\geq 1} b_n \delta_{a_n}(x)$$

where for each $n \geq 1$ we have $\delta_{a_n}(x) = \mathbf{1}_{[a_n,+\infty)}(x)$, namely, the Heaviside function at a_n. Since $0 \leq \delta_{a_n}(x) \leq 1$, the series above is absolutely and uniformly convergent. Since $\delta_{a_n}(x)$ is increasing, then if $x_1 \leq x_2$ we have that

$$f(x_2) - f(x_1) = \sum_{n\geq 1} b_n(\delta_{a_n}(x_2) - \delta_{a_n}(x_1)) \geq 0,$$

so that f is increasing. Then

$$f(x^+) - f(x^-) = \sum_{n\geq 1} b_n(\delta_{a_n}(x^+) - \delta_{a_n}(x^-)).$$

But for each $n \geq 1$, $\delta_{a_n}(x^+) - \delta_{a_n}(x^-)$ is zero or one if $x \neq a_n$ or $x = a_n$.

From this we conclude that f is discontinuous (jumps) in the rational numbers and nowhere else.

The previous example shows that the set of points of jump of an increasing function may be dense.

1.4 Random variable

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, $\mathbb{R}^* := [-\infty, \infty]$ and $\mathscr{B}^*$ be the extended Borel σ-algebra, that is, its elements are sets in $\mathscr{B}$ with one or both $+\infty, -\infty$.

Definition 1.4.1. *A function X with domain $\Lambda \in \mathscr{F}$ taking values in $\mathbb{R}^*$ is a random variable if: $\forall B \in \mathscr{B}^*$ we have that*

$$X^{-1}(B) \in \Lambda \cap \mathscr{F}, \tag{1.4.1}$$

where $\Lambda \cap \mathscr{F}$ is the trace of $\mathscr{F}$ in Λ, $X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in \mathscr{B}\}$.

Remark 1.4.2. *A random variable that takes values in the complex numbers is a function from $\Lambda \in \mathscr{F}$ to the complex plane whose real and imaginary parts are random variables taking finite values.*

From now on we assume that $\Lambda = \Omega$ and that X is real and takes finite values with probability one. The general case can be reduced to this one, considering the trace of $(\Omega, \mathscr{F}, \mathbb{P})$ in the set

$$\Lambda_0 := \{\omega \in \Omega : |X(\omega)| < \infty\}$$

and taking the real and imaginary parts of X.

Consider now the inverse application $X^{-1} : \mathbb{R} \longrightarrow \Omega$ defined on $A \subset \mathbb{R}$, by $X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$. The condition (1.4.1) tells us that X^{-1} takes elements of $\mathscr{B}$ into elements of $\mathscr{F}$: $X^{-1}(\mathscr{B}) \in \mathscr{F}$. A function which satisfies this property is said to be measurable wrt $\mathscr{F}$. Therefore a random variable is a measurable function from Ω to $\mathbb{R}$ (or $\mathbb{R}^*$).

Theorem 1.4.3. *For each function $X : \Omega \longrightarrow \mathbb{R}$ (or $\mathbb{R}^*$), the inverse application X^{-1} satisfies the following properties:*

- $X^{-1}(A^c) = (X^{-1}(A))^c$,
- $X^{-1}(\cup_\alpha A_\alpha) = \cup_\alpha X^{-1}(A_\alpha)$,
- $X^{-1}(\cap_\alpha A_\alpha) = \cap_\alpha X^{-1}(A_\alpha)$

where α belongs to an index set not necessarily countable.

Exercise:
Prove last theorem.

Theorem 1.4.4. *X is a random variable if and only if $\forall x \in \mathbb{R}$ (or x in a dense subset of $\mathbb{R}$) we have $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathscr{F}$.*

Proof. We note that last condition above, namely $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathscr{F}$ can be written as $X^{-1}((-\infty, x]) \in \mathscr{F}$. Let us also note that since X is a r.v. and since $(-\infty, x] \in \mathscr{B}$, then trivially we have that $X^{-1}((-\infty, x]) \in \mathscr{F}$. To prove the theorem it then sufficient to show that the condition $X^{-1}((-\infty, x]) \in \mathscr{F}$, with $x \in \mathbb{R}$ implies that for any Borelian $B \in \mathscr{B}$ we have that $X^{-1}(B) \in \mathscr{F}$. For that purpose let

$$\mathscr{A} := \{S \subset \mathbb{R} : X^{-1}(S) \in \mathscr{F}\}.$$

Let us check that $\mathscr{A}$ is a σ-algebra. We start by showing the stability for the complementary. For that purpose, let $S \in \mathscr{A}$. Note that from the previous theorem we have that $X^{-1}(S^c) = \left(X^{-1}(S)\right)^c$, and since X is a r.v. then $X^{-1}(S) \in \mathscr{F}$. Now, since $\mathscr{F}$ is a σ-algebra we have that $\left(X^{-1}(S)\right)^c \in \mathscr{F}$. Now we prove the stability for the countable union. Take a collection $\{S_j\}_{j\geq 1} \in \mathscr{A}$ and note that from the previous theorem we have that $X^{-1}(\cup_{j\geq 1} S_j) = \cup_{j\geq 1} X^{-1}(S_j)$. Since X is a r.v. we have that $X^{-1}(S_j) \in \mathscr{F}$ and since $\mathscr{F}$ is a σ-algebra we conclude that $\cup_{j\geq 1} X^{-1}(S_j) \in \mathscr{F}$. From this we conclude that $\mathscr{A}$ is a σ-algebra which by hypothesis contains the intervals of the form $(-\infty, x]$ which generate $\mathscr{B}$ (even in the case where x is in a dense subset of $\mathbb{R}$). Therefore $\mathscr{B} \subset \mathscr{A}$, which means that $\forall B \in \mathscr{B}$ it holds that $X^{-1}(B) \in \mathscr{F}$.

□

In this case since $\mathbb{P}$ is defined in $\mathscr{F}$ we denote the probability wrt $\mathbb{P}$ of the set $\{\omega \in \Omega : X(\omega) \in B\}$ simply by $\mathbb{P}(X \in B)$, for $B \in \mathscr{B}$.

Theorem 1.4.5. *Each random variable X defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ induces a probability space $(\mathbb{R}, \mathscr{B}, \mu)$ through the following correspondence*

$$\forall B \in \mathscr{B}, \mu(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B).$$

Proof. Let us now prove that μ defined above is a probability measure. First note that for $B \in \mathscr{B}$ we have that $\mu(B) = \mathbb{P}(X \in B) \geq 0$, since $\mathbb{P}$ is a probability measure. Now let $\{B_j\}_{j\geq 1}$ be a collection of disjoint sets in $\mathscr{B}$. Then $\{X^{-1}(B_n)\}_{j\geq 1}$ are also disjoint. If not, suppose that there exists n, m such that $X^{-1}(B_n) \cap X^{-1}(B_m) \neq \varnothing$. This means that there exists $\omega \in X^{-1}(B_n)$ and $\omega \in X^{-1}(B_m)$ so that $X(\omega) \in B_n \cap B_m$, which is absurd since $B_n \cap B_m = \varnothing$. Therefore,

$$\begin{aligned} \mu\Big(\cup_{n\geq 1} B_n \Big) &= \mathbb{P}\Big(X^{-1}\Big(\cup_{n\geq 1} B_n\Big)\Big) = \mathbb{P}\Big(\cup_{n\geq 1} X^{-1}(B_n)\Big) \\ &= \sum_{n\geq 1} \mathbb{P}(X^{-1}(B_n)) = \sum_{n\geq 1} \mu(B_n). \end{aligned}$$

Finally we note that

$$\mu(\mathbb{R}) = \mathbb{P}(X^{-1}(\mathbb{R})) = 1,$$

since $\mathbb{P}$ is a probability measure. This ends the proof. □

Remark 1.4.6.

1. *The collection of sets $\{X^{-1}(S) ; S \in \mathbb{R}\}$ is a σ-algebra for any function X.*

2. In case X is a random variable, the collection $\{X^{-1}(B)\,;\, B \in \mathscr{B}\}$ is the σ-algebra generated by X, which consists in the smallest sub σ-algebra of $\mathscr{F}$ which contains all the sets of the form $\{\omega \in \Omega : X(\omega) \leq x\}$ with $x \in \mathbb{R}$.

3. The measure μ is going to be denoted by $\mu := \mathbb{P} o X^{-1}$ and it is called the probability distribution measure of X and its associated F is the distribution function of X: $F(x) = \mu((-\infty, x]) = \mathbb{P}(X \leq x)$.

Note that X determines μ and μ determines F, the converse is false. Two random variables which have the same distribution are said to be identically distributed.

Example 7. *Consider the probability space $(\mathscr{U}, \mathscr{B}, m)$, $\mathscr{U} = [0,1]$, $\mathscr{B}$ is the Borel σ-algebra in $\mathscr{U}$ and m is the Lebesgue measure; and the random variables $X_i : \mathscr{U} \longrightarrow \mathscr{U}$ given by $X_1(\omega) = \omega$ and $X_2(\omega) = 1 - \omega$.*

We observe that $X_1 \neq X_2$ but they are identically distributed since:

$$m(\omega \in \mathscr{U} : X_1(\omega) \leq x) = m(\omega \in \mathscr{U} : \omega \leq x) = m([0,x]) = x$$

$$\begin{aligned} m(\omega \in \mathscr{U} : X_2(\omega) \leq x) &= m(\omega \in \mathscr{U} : 1 - \omega \leq x) \\ &= m(\omega \in \mathscr{U} : 1 - x \leq \omega) \\ &= 1 - m(\omega < 1 - x) \\ &= 1 - m([0, 1-x]) = 1 - (1-x) = x. \end{aligned}$$

Example 8. *Let us now consider a r.v. X with Bernoulli distribution with parameter $p \in (0,1)$. For that purpose consider $\Omega := \{\omega_1, \omega_2\}$ and the probability measure given by $\mathbb{P}(\{\omega_1\}) = p = 1 - \mathbb{P}(\{\omega_2\})$. The random variable $X : \Omega \to \mathbb{R}$ is given by $X(\omega_1) = 1$ and $X(\omega_2) = 0$. Then $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = \omega_2$. On the other hand, the induced measure μ_X is atomic, with atoms $\{0\}$ and $\{1\}$ since:*

$$\mu(\{0\}) = p > 0 \qquad \mu(\{1\}) = 1 - p > 0.$$

A simple computation also shows that the distribution function is given by

$$F_X(x) = (p - q)\mathbf{1}_{[0,1)} + \mathbf{1}_{[1,+\infty)}.$$

Exercise:
Find μ_X and F_X for the r.v. $X:\Omega\to\mathbb{R}$, where $\Omega:=\{\omega_1,\ldots,\omega_n\}$ and for each $k=1,\ldots,n$ $\mathbb{P}(X(\omega_k)=k)=p_k$ with $p_k=\binom{n}{k}p^k(1-p)^{n-k}$.

Example 9. *Given a distribution function F there exists a r.v. X such that $F_X=F$? The answer is yes. We already know how to define the measure μ associated to F, through the following relation: $F(x)=\mu((-\infty,x])$ for all $x\in\mathbb{R}$. Therefore, defining $X:\mathbb{R}\to\mathbb{R}$ by $X(\omega)=\omega$ we conclude that $F_x(x)=P(X(\omega)\leq x)=\mathbb{P}(\omega\leq x)=\mu((-\infty,x])=F(x)$. Note that we already know that such measure μ exists.*

Now we give a way to construct random variables.

Theorem 1.4.7 (Constructing random variables)**.**
If X is a random variable and $f:\mathbb{R}\longrightarrow\mathbb{R}$ is a Borel measurable function (that is $f^{-1}(\mathscr{B})\in\mathscr{B}$), then $f(X)$ is a random variable.

Proof. To prove the theorem is is enough to note that for $B\in\mathscr{B}$ it holds that $(f o X)^{-1}(B)=X^{-1}(f^{-1}(B))$, since f is Borel measurable then $f^{-1}(B)\in\mathscr{B}$, and since X is a r.v. we conclude that $X^{-1}(f^{-1}(B))\in\mathscr{F}$ and we are done. $\square$

We note that according to the previous theorem if we compose a r.v. X with any continuous function $f:\mathbb{R}\to\mathbb{R}$ then $f(X)$ is a r.v.

1.5 Types of distribution functions

Recall the definition of a distribution function F. Let $\{a_j\}_{j\geq1}$ be the countable set of points of jump of F and let b_j be the size of the jump at a_j:

$$F(a_j^+)-F(a_j^-)=F(a_j)-F(a_j^-)=b_j.$$

Let

$$F_d(x)=\sum_{j\geq1}b_j\delta_{a_j}(x),$$

where $\delta_{a_j}(x)$ is the Heaviside function at a_j. The function $F_d(x)$ represents all the jumps of F in $(-\infty, x]$. Note that F_d is increasing, right-continuous, $F_d(-\infty) = 0$ and

$$F_d(+\infty) = \sum_{j\geq 1} b_j \leq 1.$$

The function F_d is the jumping part of F.

Theorem 1.5.1. *The function $F_c(x) = F(x) - F_d(x)$ is positive, increasing and continuous.*

Proof. Let us first prove that F is increasing. Let $x < x'$. Then, by the definition of F_d we have that

$$F_d(x') - F_d(x) = \sum_j b_j(\delta_{a_j}(x') - \delta_{a_j}(x)) = \sum_{j:x<a_j<x'} F(a_j) - F(a_j') \leq F(x') - F(x).$$

Now since both F_d and F are increasing we conclude that $0 \leq F_d(x') - F_d(x) \leq F(x') - F(x)$ which is equivalent to saying that $0 \leq F_c(x') - F_c(x)$, so that F_c is increasing. Note that taking $x = -\infty$ in the first display above we can conclude that $F_d(x') \leq F(x')$, so that $F_c(x') \geq 0$. Let us now prove that F_c is continuous. Note that F_d is right continuous since each δ_{a_j} is also right continuous and by the Weierstrass test the series which defines F_d is uniformly converging in x. Since F is also right continuous, we conclude that F_c is right continuous. Moreover, $F_d(x) - F_d(x^-) = b_j \mathbf{1}_{\{x=a_j\}}$ and the same holds for F, that is $F(x) - F(x^-) = b_j \mathbf{1}_{\{x=a_j\}}$, by the definition of b_j and a_j. Then

$$F_c(x) - F_c(x') = F(x) - F(x') - (F_d(x) - F_d(x')) = 0.$$

From this we conclude that F is left continuous from where the continuity follows. □

Theorem 1.5.2. *Let F be a distribution function. Suppose that there exists a continuous function G_c and a function G_d of the form $G_d(x) = \sum_{j\geq 1} b_j' \delta_{a_j'}(x)$, where $\{a_j'\}_{j\geq 1}$ is a countable set of real numbers and $\sum_{j\geq 1} b_j' < \infty$, such that $F = G_c + G_d$. Then $G_c = F_c$ and $G_d = F_d$ where F_c and F_d were defined above.*

Proof. Let us suppose that $F_d \neq G_d$. Then, or the sets $\{a_j\}_{j\geq 1}$ and $\{a'_j\}_{j\geq 1}$ are not equal or they are equal and then we can have $a_j = a'_j$ for all $j \geq 1$ but $b_j \neq b'_j$ for some $j \geq 1$. In any case, we should have, for at least one j and one $\tilde{a}_j$, that $\tilde{a}_j = a_j$ or $\tilde{a}_j = a'_j$. Then, for that $\tilde{a}_j = a_j$ we have

$$F_d(\tilde{a}_j) - F_d(\tilde{a}_j^-) = b_j \neq b'_j = G_d(\tilde{a}_j) - G_d(\tilde{a}_j^-)$$

(or the other way around). Since $F = F_c + F_d$ and $F = G_c + G_d$, then

$$F_c(\tilde{a}_j) + F_d(\tilde{a}_j) = G_c(\tilde{a}_j) + G_d(\tilde{a}_j)$$

$$F_c(\tilde{a}_j^-) + F_d(\tilde{a}_j^-) = G_c(\tilde{a}_j^-) + G_d(\tilde{a}_j^-).$$

From the previous equalities we conclude that

$$F_c(\tilde{a}_j) - F_c(\tilde{a}_j^-) - G_c(\tilde{a}_j) + G_c(\tilde{a}_j^-) = G_d(\tilde{a}_j) - G_d(\tilde{a}_j^-) - F_d(\tilde{a}_j) + F_d(\tilde{a}_j^-) \neq 0.$$

But then $F_c - G_c$ would not be a continuous function, which is absurd. From this it follows that $F_d = G_d$ so that $F_c = G_c$ and the proof ends. □

Definition 1.5.3. *A distribution function that can be represented in the form* $F = \sum_{j\geq 1} b_j \delta_{a_j}$, *where* $\{a_j\}_{j\geq 1}$ *is a countable (or finite) set of real numbers* $b_j > 0$ *for every* j *and* $\sum_{j\geq 1} b_j = 1$ *is called a* discrete distribution function. *A distribution function that is continuous everywhere is called a* continuous distribution function.

Suppose that for a distribution function F we have that $F_c \neq 0$ and $F_d \neq 0$. Let $\alpha = F_d(+\infty)$ such that $0 < \alpha < 1$ and let

$$F_1 = \frac{1}{\alpha} F_d \quad \text{and} \quad F_2 = \frac{1}{1-\alpha} F_c.$$

Then

$$F = F_d + F_c = \alpha F_1 + (1-\alpha) F_2, \tag{1.5.1}$$

where F_1 is a discrete distribution function and F_2 is a continuous distribution function and F is a convex combination of F_1 and F_2.

Remark 1.5.4. *If* $F_c = 0$ *then* F *is discrete and we take* $\alpha = 1$, *so that* $F_1 = F$ *and* $F_2 = 0$; *and if* $F_d = 0$, *then* F *is continuous and we take* $\alpha = 0$ *and* $F_1 = 0$ *and* $F_2 = F$ *and in both cases* (1.5.1) *holds.*

The two previous theorems can be combined in one:

Theorem 1.5.5 (Convex combination of distribution functions)**.**

Every distribution function can be written as the convex combination of a discrete and a continuous distribution function. Such decomposition is unique.

Now let us see another type of distribution function.

Definition 1.5.6. *A function f is in $\mathbb{L}^1(\mathbb{R})$ iff $\int_{\mathbb{R}} |f(y)|\,dy < \infty$.*

Definition 1.5.7. *A function F is said to be absolutely continuous (in $\mathbb{R}$ wrt the Lebesgue measure) iff there exists a function $f \in \mathbb{L}^1$ such that $\forall x < x'$ we have that*

$$F(x') - F(x) = \int_x^{x'} f(y)\,dy.$$

There is a result in measure theory that says that such a function F has a derivative equal to f almost everywhere (a.e.). This means that the derivative is equal to zero on a set of full Lebesgue measure. In particular if F is a distribution function then

$$f \geq 0 \quad \text{a.e. and} \quad \int_{\mathbb{R}} f(y)\,dy = 1. \tag{1.5.2}$$

Such function above is called a density.

Conversely, given any $f \in \mathbb{L}^1$ satisfying the previous conditions in (1.5.2), the function F defined for all $x \in \mathbb{R}$ as

$$F(x) = \int_{-\infty}^{x} f(y)\,dy$$

is a distribution function that is absolutely continuous.

Theorem 1.5.8. *Let $F : \mathbb{R} \to \mathbb{R}$ such that for $x \leq x'$ and f a density function, we have that*

$$F(x') - F(x) = \int_x^{x'} f(t)dt.$$

Then, F is a.e. differentiable and $F' = f$ a.e.

Definition 1.5.9. *A function F is called singular if and only if it is not identically zero and F' exists and is equal to zero a.e.*

The next theorem can be seen in any book of measure theory and for that reason the proof is omitted.

Theorem 1.5.10. *Let F be bounded increasing with $F(-\infty) = 0$ and let F' denote its derivative whenever it exists. Then:*

1. *If S is the set of points x for which $F'(x)$ exists with $0 \leq F'(x) < +\infty$, then $m(S^c) = 0$.*

2. *The function F' belongs to $\mathbb{L}^1$ and we have for every $x < x'$ that*

$$\int_x^{x'} F'(y)\,dy \leq F(x') - F(x).$$

3. *If for all $x \in \mathbb{R}$*

$$F_{ac}(x) = \int_{-\infty}^{x} F'(y)\,dy \quad \text{and} \quad F_s(x) = F(x) - F_{ac}(x),$$

then $F'_{ac} = F'$ a.e., so that $F'_s = F' - F'_{ac} = 0$ a.e. and consequently F_s is singular if it is not identically zero.

Definition 1.5.11. *Any positive function f that is equal to F' a.e. is called a density of F. F_{ac} is the absolutely continuous part of F and F_s is its singular part.*

Remark 1.5.12. *Note that:*

1. the discrete part F_d defined above is part of the singular part F_s defined above;

2. F_{ac} is increasing and $F_{ac} \leq F$. (Check it!)

Moreover, if $x < x'$ then $F_s(x') - F_s(x) = F(x') - F(x) - \int_x^{x'} f(y)\,dy \geq 0$, (from (2) of the previous theorem) therefore F_s is also increasing and $F_s \leq F$. (Check it!)

Theorem 1.5.13. *Every distribution function F can be written as the convex combination of a discrete, a singular and an absolutely continuous distribution function and such decomposition is unique.*

Proof. Note that if

$$\tilde{\beta} := \int_{\mathbb{R}} F'(t)dt = 0,$$

then $F'(t) = 0$ for t a.e. so that $F_{ac}(x) = 0$ for all x. Therefore, $F_c = F_s$, where F_c is the continuous part of F. If $\tilde{\beta} = 1$, then $F_{ac}(+\infty) = 1$ and $F_{ac} = F_c$ and F_c is absolutely continuous. Now, if $\tilde{\beta} \in (0,1)$, then if $\alpha \in (0,1)$ and $\beta \in (0,1)$, then F can be written as

$$F = \alpha F_1 + \beta F_2 + \gamma F_3,$$

where

$$\beta = (1-\alpha)\tilde{\beta}, \qquad \gamma = (1-\alpha)(1-\tilde{\beta})$$
$$F_1 = F_d, \quad F_2 = \frac{1}{\tilde{\beta}} F_{ac}, \quad F_3 = \frac{1}{(1-\tilde{\beta})} F_s.$$

□

In the next section we are going to construct a singular distribution function. It is called the Cantor distribution function.

1.5.1 The Cantor distribution function

Let us construct the ternary Cantor set. This is a construction which is done by induction. It goes like this. From the closed interval $[0,1]$ remove the central interval $(\frac{1}{3}, \frac{2}{3})$. Then in the two remaining intervals remove the central intervals $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$. After the 1st step we remain with two intervals of size $\frac{1}{3}$. In the 2nd step we remain with four intervals of size $\frac{1}{3^2}$ and so on. After n steps we have removed $1+2+4+8+\cdots+2^{n-1} = 2^n - 1$ disjoint intervals and remain 2^n closed intervals of size $\frac{1}{3^n}$. Let us order these intervals, by order from left to right and denote them by $J_{n,k}$, where $1 \leq k \leq 2^n - 1$ and denote their union by U_n. Note that

$$m(U_n) = 1 - \left(\frac{2}{3}\right)^n.$$

As n increases the set U_n increases to an open set U and let $\mathscr{C} := U^c$ (the complementary wrt $[0,1]$) be the Cantor set. Then

$$m(\mathscr{C}) = 1 - m(U) = 1 - \lim_{n\to\infty} m(U_n) = 1 - 1 = 0.$$

Now we define the Cantor distribution function. For each n,k, with $n \geq 1$ and $k = 1,\cdots,2^n - 1$ let $c_{n,k} = \frac{k}{2^n}$ and let us define F in U in the following way: if $x \in J_{n,k}$ then $F(x) = c_{n,k}$. In each $J_{n,k}$ the function F is constant and it is strictly greater on any $J_{n,k'}$ at the right of $J_{n,k}$. Therefore, F is increasing and $F(0^+) = 0$ and $F(1^-) = 1$. Now we complete the definition by setting $F(x) = 0$ for $x \leq 0$ and $F(x) = 1$ for $x \geq 1$. Up to here the function F is defined on the domain

$$\mathscr{D} = (-\infty, 0] \cup U \cup [1, +\infty)$$

and is increasing.

Now, since each $J_{n,k}$ is at a distance which is greater or equal than $1/3^n$ from any other $J_{n,k'}$ and since the total variation of F over each of the 2^n disjoint intervals that remain after removing $J_{n,k}$ is $\frac{1}{2^n}$, it follows that

$$0 \leq x' - x \leq \frac{1}{3^n} \Rightarrow 0 \leq F(x') - F(x) \leq \frac{1}{2^n}.$$

Then, the function F is uniformly continuous on $\mathscr{D}$. Note that it is known that $\mathscr{D}$ is dense in $\mathbb{R}$. Now, to define the function in the full space $\mathbb{R}$ we need the following result.

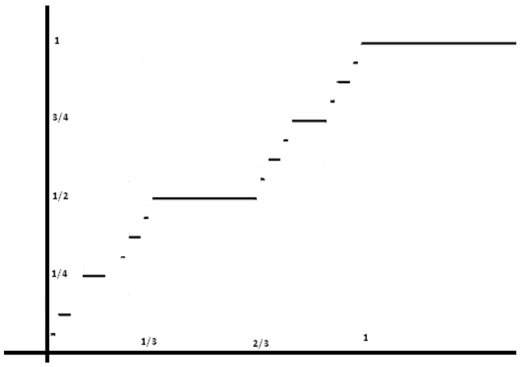
Lemma 1.5.14. *Let f be increasing on a dense subset $\mathscr{D}$ of $\mathbb{R}$. If for any $x \in \mathbb{R}$*

$$\tilde{f}(x) = \inf_{x < t \in \mathscr{D}} f(t),$$

then $\tilde{f}$ is increasing and right continuous everywhere. If f uniformly continuous, then $\tilde{f}$ is uniformly continuous.

By Lemma 1.5.14 there exists a continuous and increasing function $\tilde{F}$ defined on $\mathbb{R}$ that coincides with F on $\mathscr{D}$. This function $\tilde{F}$ is a continuous distribution function that is constant on each $J_{n,k}$ so that $\tilde{F}' = 0$ on U and also on $\mathbb{R} \backslash \mathscr{C}$, which means that $\tilde{F}$ is singular. Below we see the graph of F after some steps of the induction procedure.

Definition 1.5.15. *A random variable X is said to be discrete if it takes values in a finite or countable set, that is, if there exists a finite or countable set $B \in \mathbb{R}$ such that $\mathbb{P}(X \in B) = 1$.*



1.5.2 Types of random variables

Definition 1.5.16. *A random variable X whose distribution function F has a density f is said to be absolutely continuous.*

Note that,

1. If X is discrete, then $\mathbb{P}(X \in B) = \sum_{i:x_i \in B} \mathbb{P}(X = x_i)$;
2. If X is absolutely continuous with density f, then $\mathbb{P}(X \in A) = \int_A f(y)\,dy$, for any $A \in \mathscr{B}$.

Exercise:

Let X be a r.v. with density given by

$$f(x) = \frac{1}{(1+x)^2}\mathbf{1}_{(0,+\infty)}(x).$$

Let $Y = max(X, c)$, where c is a strictly positive constant.

a) Find the distribution of X and Y and do the graphical representation.

b) Decompose the distribution function of Y in its discrete, absolutely continuous and singular parts.

Note that the distribution function of X is given by $F_X(x) = \int_{-\infty}^{x} f(t)dt$, that is

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0, \\ \frac{x}{1+x}, & \text{otherwise.} \end{cases}$$

The distribution function of Y is given by:

$$F_Y(y) = \begin{cases} 0, & \text{if } y < c\,, \\ \frac{y}{1+y}, & \text{otherwise.} \end{cases}$$

The distribution function of Y decomposes into its discrete part

$$F_d(y) = \begin{cases} 0, & \text{if } y < c\,, \\ \frac{c}{1+c}, & \text{otherwise.} \end{cases}$$

and the absolutely continuous part:

$$F_{ac}(y) = \begin{cases} 0, & \text{if } y < c\,, \\ \frac{-1}{1+y} + \frac{1}{1+c}, & \text{otherwise.} \end{cases}$$

Since for all $y \in \mathbb{R}$ we have that $F_d(y)+F_{ac}(y) = F_Y(y)$, then the singular part fo F is null, that is $F_s(y) = 0$. The random variable Y is of mixed type.

1.6 Random vectors

A random vector is just a vector whose components are random variables. We focus on the case $d = 2$. Basically here we just rewrite what we have seen before in a 2-dimensional setting. Note that the Borel σ-algebra in $\mathbb{R}^2$ is the σ-algebra generated by rectangles of the form

$$\{(x, y) : a < x \leq b\,; c < y \leq d\}$$

and it is also generated by products sets of the form

$$B_1 \times B_2 = \{(x, y) : x \in B_1\,; y \in B_2\},$$

where $B_1, B_2 \in \mathscr{B}$. A function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is Borel measurable iff $f^{-1}(\mathscr{B}) \in \mathscr{B}^2$.

Definition 1.6.1. *Let X and Y be two random variables defined on the same probability space $(\Omega, \mathscr{F}, \mathbb{P})$. The random vector (X, Y) induces a probability measure*

$\nu \in \mathscr{B}^2$ such that for $A \in \mathscr{B}^2$

$$\nu(A) = \mathbb{P}((X,Y) \in A) = \mathbb{P}(\omega \in \Omega : (X(\omega), Y(\omega)) \in A).$$

The measure ν is called the distribution measure of (X,Y).

We also define the inverse application $(X,Y)^{-1}$ in the following way:

$$\forall A \in \mathscr{B}^2 \, : \, (X,Y)^{-1}(A) = \{\omega \in \Omega \, : \, (X(\omega), Y(\omega)) \in A\}.$$

We note that the results that we have seen above for X^{-1} are also true for $(X,Y)^{-1}$.

Theorem 1.6.2. *If X and Y are random variables and if $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is Borel measurable, then $f(X,Y)$ is a random variable.*

The proof of last result is analogous to the proof of Theorem 1.4.7.

Example 10.

1. *If X is a random variable and if $f : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous, then $f(X)$ is a random variable. Therefore:*

   • *X^r; $|X|^r$ for positive real r; $e^{-\lambda X}$, for real λ, e^{itX}, for real t are random variables;*

2. *If X and Y are random variables then all these are random variables:*

   • *$X \pm Y$; $X.Y$; X/Y; $X \wedge Y := min(X,Y)$; $X \vee Y := max(X,Y)$;*

Theorem 1.6.3. *If $\{X_j\}_{j\geq 1}$ is a sequence of random variables, then*

$$\inf_j X_j; \; \sup_j X_j; \; \lim_j \inf X_j; \; \lim_j \sup X_j;$$

are random variables not necessarily finite but a.e. defined and $\lim_{j\to+\infty} X_j$ is a random variable on the set where there is convergence or divergence to $\pm\infty$.

Proof. Note that for all $x \in \mathbb{R}$ we have that

$$\{\sup_j X_j \leq x\} = \cup_j \{X_j \leq x\}.$$

Since $\{X_j \leq x\} \in \mathscr{F}$, because X is a r.v. then $\{\sup_j X_j \leq x\} \in \mathscr{F}$. From Theorem 1.4.4 we conclude that $\sup_j X_j$ is a r.v. Analogously, since all $x \in \mathbb{R}$ we have that

$$\{\inf_j X_j > x\} = \cap_j \{X_j > x\}.$$

Since $\{X_j > x\} \in \mathscr{F}$, because X is a r.v. then $\{\inf_j X_j > x\} \in \mathscr{F}$. From Theorem 1.4.4 we conclude that $\inf_j X_j$ is a r.v. Now for the $\limsup X_j$ note that

$$\limsup_j = \inf_n \Big(sup_{j\geq n} X_j \Big).$$

From the previous arguments we know that $sup_{j\geq n} X_j$ is a r.v. and also that $\inf_n \Big(sup_{j\geq n} X_j \Big)$ is a r.v., from where the proof ends. □

Definition 1.6.4. *The distribution function of a random vector (X,Y) is defined on $(x,y) \in \mathbb{R}^2$ by*

$$F(x,y) = \mathbb{P}(X \leq x, Y \leq y).$$

F is also called the joint distribution function of the r.v. X and Y.

The distribution function just defined satisfies the following properties:

1. F is increasing in each variable.

2. F is right-continuous in each variable.

3. $\lim_{x\to-\infty} F(x,y) = 0$ and $\lim_{y\to-\infty} F(x,y) = 0$.

4. $\lim_{x\to+\infty, y\to+\infty} F(x,y) = 1$

Exercise:
Prove the previous properties.

Note that the distribution function of X (resp. Y) is obtained from the joint distribution function by taking the limit:

$$\lim_{y\to\infty} F(x,y) = F(x)$$

(resp. $\lim_{x\to\infty} F(x,y) = F(y)$).

The properties above are not sufficient to guarantee that a function $F : \mathbb{R}^2 \to \mathbb{R}$ is the distribution function of a random vector. Let us see an example.

Example 11. *Let*

$$F(x,y) = \mathbf{1}_{\{x\geq 0, y\geq 0, x+y\geq 1\}}.$$

It is easy to see that F satisfies the properties above, nevertheless it is not the distribution function of a random vector. Suppose it is. Then we would have, for example, that: $\mathbb{P}(X \in (0,1], Y \in (0,1]) = -1$, which cannot happen since $\mathbb{P}$ is a probability measure.

We need to introduce some extra condition, in order to avoid what we have seen in the previous example. That condition is the following: • For any $a_1 < b_1$ and $a_2 < b_2$ we have

$$\mathbb{P}(X \in (a_1, b_1], Y \in (a_2, b_2]) \geq 0.$$

A function F satisfying the properties above is the distribution function of a random vector.

Definition 1.6.5. *A random vector (X,Y) is discrete iff it takes a finite or countable number of values.*

Definition 1.6.6. *Let (X,Y) be a random vector and let F be its distribution function. If there exists a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that $f(x,y) \geq 0$ and $\int\int_{R^2} f(x,y)\,dx\,dy = 1$ and if for any $(x,y) \in \mathbb{R}^2$*

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv,$$

then f is called the density function of the random vector (X,Y) or the joint density of the r.v. X and Y. In this case, we say that the random vector is absolutely continuous.

1.7 Stochastic Independence

Definition 1.7.1. *The collection of random variables $\{X_j\}_{j=1,\cdots,n}$ are said to be independent iff for any $\{B_j\}_{j=1,\cdots,n}$ with $B_j \in \mathscr{B}$, for any $j = 1, \cdots, n$, we have that*

$$\mathbb{P}\Big(\cap_{j=1}^n (X_j \in B_j)\Big) = \Pi_{j=1}^n \mathbb{P}(X_j \in B_j). \tag{1.7.1}$$

Remark 1.7.2.

1. The r.v. of an infinite family are said to be independent iff the r.v. in any finite subfamily are independent.

2. The r.v. are said to be pairwise independent iff every two of them are independent.

3. Note that (1.7.1) implies that any of its subfamilies is independent, since

$$\mathbb{P}\Big(\cap_{j=1}^k(X_j \in B_j)\Big)=\mathbb{P}\Big(\cap_{j=1}^n(X_j \in B_j)\Big)=\Pi_{j=1}^n\mathbb{P}(X_j \in B_j)=\Pi_{j=1}^k\mathbb{P}(X_j \in B_j)$$

Remark 1.7.3. *We note that (1.7.1) is equivalent to*

$$\mathbb{P}\Big(\cap_{j=1}^n (X_j \leq x_j)\Big) = \Pi_{j=1}^n \mathbb{P}(X_j \leq x_j), \tag{1.7.2}$$

for every set of real numbers $\{x_j\}_{j=1}^n$. To prove this it is enough to check that the set

$$\mathscr{C} = \{B : \mathbb{P}((X_1, X_2, \ldots, X_n) \in B) = \prod_{i=1}^n \mathbb{P}(X_i \in B)\}$$

where $B = B_1 \times B_2 \times \cdots \times B_n$ and $B_i \in \mathscr{B}$ for each $i = 1, \ldots, n$, forms a σ-algebra that contains the sets of the form $(-\infty, x_1] \times (-\infty, x_2] \times \ldots (-\infty, x_n]$. This is left as an exercise to the reader.

We can rewrite (1.7.1) in terms of the probability measure $\mu_{(X_1,\cdots,X_n)}$ induced by the random vector $(X_1, \cdots, X_n)$ on $(\mathbb{R}^n, \mathscr{B}^n)$ as

$$\mu_{(X_1,\cdots,X_n)}(B_1 \times \cdots \times B_n) = \Pi_{j=1}^n \mu_j(B_j) = \mu_1(B_1) \times \cdots \times \mu_n(B_n),$$

where $\mu_j := \mu_{X_j}$ is the probability measure induced by each random variable X_j in $(\mathbb{R}, \mathscr{B})$. Note that the induced measure in this case is the product measure!

Remark 1.7.4. *We can define the n-dimensional distribution function* $F_{(X_1,\cdots,X_n)}$ *as*

$$\begin{aligned} F_{(X_1,\cdots,X_n)}(x_1,\cdots,x_n) =& \mathbb{P}(X_1 \leq x_1,\cdots,X_n \leq x_n) \\ =& \mu_{(X_1,\cdots,X_n)}((-\infty,x_1]\times\cdots\times(-\infty,x_n]). \end{aligned}$$

and the condition (1.7.2) *is rewritten as* $F(x_1,\cdots,x_n) = \Pi_{j=1}^{n} F_j(x_j)$.

Example 12. *Let* X_1 *and* X_2 *be independent r.v. given by*

$$X_1 = \begin{cases} 1, & 1/2 \\ -1, & 1/2 \end{cases} \quad \text{and} \quad X_2 = \begin{cases} 1, & 1/2 \\ -1, & 1/2. \end{cases}$$

Then, the three r.v. $\{X_1, X_2, X_1X_2\}$ *are pairwise independent but they are not totally independent. To prove the assertion check that the r.v.* X_1X_2 *satisfies:*

$$X_1X_2 = \begin{cases} 1, & 1/2 \\ -1, & 1/2. \end{cases}$$

Then

$$\mathbb{P}(X_1 = 1, X_1X_2 = 1) = \mathbb{P}(X_1 = 1, X_2 = 1) = \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1) = 1/4$$

and

$$\mathbb{P}(X_1 = 1)\mathbb{P}(X_1X_2 = 1) = 1/4.$$

Doing some similar computation we conclude that X_1 *and* X_1X_2 *are independent. Analogously we can conclude that* X_2 *and* X_1X_2 *are independent. Now note that* X_1, X_2, X_1X_2 *are not independent, since*

$$\mathbb{P}(X_1 = 1, X_2 = -1, X_1X_2 = 1) = \mathbb{P}(\emptyset) = 0$$

but

$$\mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = -1)\mathbb{P}(X_1X_2 = 1) \neq 0.$$

Whenever a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is fixed, the sets in $\mathscr{F}$ will be called events. We have seen above the notion of independent r.v. but what about independent events?

Definition 1.7.5. *We say that the events* $\{E_j\}_{j=1}^n$ *are independent iff their indicators are independent, that is, for any subset* $\{j_1, \cdots, j_\ell\}$ *of* $\{1, \cdots, n\}$ *we have that*

$$\mathbb{P}\Big(\cap_{k=1}^{\ell} E_{j_k} \Big) = \Pi_{k=1}^{\ell} \mathbb{P}(E_{j_k}).$$

Theorem 1.7.6. *If* $\{X_j\}_{j=1}^n$ *are independent r.v. and* $\{f_j\}_{j=1}^n$ *are Borel measurable functions, then* $\{f_j(X_j)\}_{j=1}^n$ *are independent r.v.*

Proof. <for $j = 1, \cdots, n$ let $B_j \in \mathscr{B}$. Then $F_j^{-1}(B_j) \in \mathscr{B}$. Therefore

$$\cup_{j=1}^n \{f_j(X_j) \in B_j\} = \cup_{j=1}^n \{X_j \in f_j^{-1}(B_j)\}$$

and

$$\begin{aligned}\mathbb{P}(\cup_{j=1}^n \{f_j(X_j) \in B_j\}) = \mathbb{P}(\cup_{j=1}^n \{X_j \in f_j^{-1}(B_j)\}) &= \prod_{j=1}^n \mathbb{P}(X_j \in f_j^{-1})(B_j)) \\ &= \prod_{j=1}^n \mathbb{P}(f_j(X_j) \in B_j)\end{aligned}$$

and we are done. □

We have seen above that if $X_1, \cdots, X_n$ are independent r.v. then

$$F_{(X_1, \cdots, X_n)}(x_1, \cdots, x_n) = \Pi_{j=1}^n F_{X_j}(x_j).$$

Now let us see the reciprocal.

Proposition 1.7.7. *If there exist functions* $F_1, \cdots, F_n$ *such that*

$$\lim_{x_j \to \infty} F_j(x_j) = 1$$

for all $j = 1, \cdots, n$ *and if for all* $(x_1, \cdots, x_n) \in \mathbb{R}^n$

$$F_{(X_1, \cdots, X_n)}(x_1, \cdots, x_n) = \Pi_{j=1}^n F_j(x_j),$$

then $\{X_j\}_{j=1}^n$ *are independent and* $F_j := F_{X_j}$ *for all* $j = 1, \cdots, n$.

Proof. To prove the proposition it is enough to see that $F_{X_i} = F_i$, then it follows from the definition. Note that

$$\begin{aligned} F_{X_i}(x_i) &= \lim_{x\to+\infty} F_{X_1,\cdots,X_n}(x,\cdots,x,x_i,x\cdots) \\ &= \lim_{x\to+\infty} F_1(x)\cdots F_i(x_i)\cdots F_n(x) \\ &= \lim_{x\to+\infty} F_1(x)\cdots F_i(x_i)\cdots \lim_{x\to+\infty} F_n(x) \\ &= F_i(x_i). \end{aligned}$$

Finally the reader can check that the r.v. are independent. This ends the proof. □

Proposition 1.7.8.

- *If* $\{X_j\}_{j=1}^n$ *are independent r.v. with densities* $f_{X_1},\cdots,f_{X_n}$*, then the function*

$$f(x_1,\cdots,x_n) = \Pi_{j=1}^n f_{X_j}(x_j)$$

is the joint density of $\{X_j\}_{j=1}^n$ *or the density of the random vector* $(X_1,\cdots,X_n)$.

- *On the other hand, if* $X_1,\cdots,X_n$ *has a joint density* f *which satisfies*

$$f(x_1,\cdots,x_n) = \Pi_{j=1}^n f_j(x_j)$$

for all $(x_1,\cdots,x_n)\in\mathbb{R}^n$ *with* $f_j(x)\geq 0$ *and* $\int_{\mathbb{R}} f_j(x)dx = 1$*, then* $X_1,\cdots,X_n$ *are independent and* f_j *is the density of* X_j*.*

Proof. Since $X_1,\cdots,X_n$ are independent, then

$$\begin{aligned} F_{(X_1,\cdots,X_n)}(x_1,\cdots,x_n) &= \prod_{i=1}^n F_{X_i}(x_i) = \prod_{i=1}^n \int_{-\infty}^{x_i} f_{X_i}(t_i)dt_i \\ &= \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} f_{X_1}(t_1)\cdots f_{X_n}(t_n)dt_1\cdots dt_n \end{aligned}$$

so that $f_{X_1}(t_1)\cdots f_{X_n}(t_n)$ is the joint density function of $X_1\cdots,X_n$.

Now note that

$$\begin{aligned}F_{(X_1,\cdots,X_n)}(x_1,\cdots,x_n) &= \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} f(t_1,\cdots,t_n)dt_1\cdots dt_n\\ &= \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} f_1(t_1)\cdots f_n(t_n)dt_1\cdots dt_n\\ &= \prod_{i=1}^{n}\int_{x_i}^{+\infty} f_i(t_i)dt_i.\end{aligned}$$

taking $F_i(x_i) = \int_{x_i}^{+\infty} f_i(t_i)dt_i$ we have that $\lim_{x_i\to\infty} F_i(x_i) = 1$, so that by the previous proposition $X_1,\cdots,X_n$ are independent and $F_i = F_{X_i}$ and f_i is the density of X_i. □

Constructing independent r.v.

Let $(\Omega_1,\mathscr{F}_1,\mathbb{P}_1)$ and $(\Omega_2,\mathscr{F}_2,\mathbb{P}_2)$ be discrete probability spaces and where $\mathscr{F}_j$ is the total σ-algebra. We define the product space $\Omega^2 := \Omega_1\times\Omega_2$ as the space of points $\omega = (\omega_1,\omega_2)$ with $\omega_1\in\Omega_1$ and $\omega_2\in\Omega_2$. The product σ-algebra $\mathscr{F}^2$ is the collection of all the subsets of Ω^2. We know from the beginning of the course that the probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ are determined by their values in ω_1, ω_2 respectively. Since Ω^2 is also countable we can define a probability measure $\mathbb{P}^2$ in $\mathscr{F}^2$ as

$$\mathbb{P}^2(\{(\omega_1,\omega_2)\}) = \mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\})$$

which is the product measure of $\mathbb{P}_1$ and $\mathbb{P}_2$. Check that it is a probability measure! It has the property that if $S_1\in\mathscr{F}_1$ and $S_2\in\mathscr{F}_2$, then

$$\mathbb{P}^2(S_1\times S_2) = \mathbb{P}_1(S_1)\mathbb{P}_2(S_2).$$

Now, let X_1 be a r.v. on Ω_1 and X_2 a r.v. on Ω_2; B_1 and B_2 Borel sets and $S_1 = X_1^{-1}(B_1) := \{\omega_1\in\Omega_1 : X_1\in B_1\}$ and $S_2 = X_2^{-1}(B_2)$. Note that $S_1\in\mathscr{F}_1$ and $S_2\in\mathscr{F}_2$. Then

$$\begin{aligned}&\mathbb{P}^2(X_1\in B_1\times X_2\in B_2)\\ &\quad=\mathbb{P}^2(S_1\times S_2) = \mathbb{P}_1(S_1)\mathbb{P}_2(S_2) = \mathbb{P}_1(X_1\in B_1)\mathbb{P}_2(X_2\in B_2).\end{aligned}$$

To X_1 on Ω_1 and X_2 in Ω_2, we associate the function $\tilde{X}_1$ and $\tilde{X}_2$ defined on $\omega \in \Omega^2$ as $\tilde{X}_1(\omega) = X_1(\omega_1)$ and $X_2(\omega) = X_2(\omega_2)$. Now we have

$$\begin{aligned}&\cap_{j=1}^{2}\{\omega \in \Omega^2 : \tilde{X}_j(\omega) \in B_j\}\\ &\qquad = \Omega_1 \times \{\omega_2 \in \Omega_2 : X_2(\omega_2) \in B_2\} \cap \{\omega_1 \in \Omega_1 : X_1(\omega_1) \in B_1\} \times \Omega_2\\ &\qquad = \{\omega_1 \in \Omega_1 : X_1(\omega_1) \in B_1\} \times \{\omega_2 \in \Omega_2 : X_2(\omega_2) \in B_2\}.\end{aligned}$$

From where we conclude that

$$\mathbb{P}^2(\cap_{j=1}^{2}\{\tilde{X}_j \in B_j\}) = \mathbb{P}^2(\tilde{X}_1 \in B_1)\mathbb{P}^2(\tilde{X}_2 \in B_2),$$

so that the random variables $\tilde{X}_1$ and $\tilde{X}_2$ are independent!

Now we extend the construction to n discrete probability spaces. Let $n \geq 2$ and $(\Omega_j, \mathscr{F}_j, \mathbb{P}_j)$ be n discrete probability spaces where $\mathscr{F}_j$ is the total σ-algebra. We define the product space $\Omega^n := \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$ as the space of points $\omega = (\omega_1, \cdots, \omega_n)$ with $\omega_j \in \Omega_j$. The product σ-algebra $\mathscr{F}^n$ is the collection of all the subsets of Ω^n. We know from the beginning of the course that for each j, the probability measure $\mathbb{P}_j$ is determined by its value in ω_j. Since Ω^n is also countable we can define a probability measure $\mathbb{P}^n$ in $\mathscr{F}^n$ as

$$\mathbb{P}^n(\{(\omega_1, \cdots, \omega_n)\}) = \Pi_{j=1}^{n}\mathbb{P}_j(\{\omega_j\})$$

which is the product measure of the $\{\mathbb{P}_j\}_{j=1}^{n}$. Check that it is a probability measure. It has the property that if $S_j \in \mathscr{F}_j$, then

$$\mathbb{P}^n(S_1 \times \cdots \times S_n) = \Pi_{j=1}^{n}\mathbb{P}_j(S_j).$$

Now, let X_j be a r.v. on Ω_j, B_j a Borel set and $S_j = X_j^{-1}(B_j) := \{\omega_j \in \Omega_j : X_j \in B_j\}$. Note that $S_j \in \mathscr{F}_j$. Then

$$\begin{aligned}&\mathbb{P}^n(X_1 \in B_1 \times \cdots \times X_n \in B_n)\\ &\qquad = \mathbb{P}^n(S_1 \times \cdots \times S_n) = \Pi_{j=1}^{n}\mathbb{P}_j(S_j) = \Pi_{j=1}^{n}\mathbb{P}_j(X_j \in B_j).\end{aligned}$$

To each function X_j on Ω_j we associate the function $\tilde{X}_j$ on Ω^n defined on $\omega \in \Omega$ as $\tilde{X}_j(\omega) = X_j(\omega_j)$. Now we have

$$\begin{aligned}&\cap_{j=1}^{n}\{\omega \in \Omega^n : \tilde{X}_j(\omega) \in B_j\}\\ &\qquad = \cap_{j=1}^{n}\Omega_1 \times \cdots \times \Omega_{j-1} \times \{\omega_j \in \Omega_j : X_j(\omega_j) \in B_j\} \times \Omega_{j+1} \times \cdots \times \Omega_n\\ &\qquad = \Pi_{j=1}^{n}\{\omega_j \in \Omega_j : X_j(\omega_j) \in B_j\}\end{aligned}$$

From where we conclude that

$$\mathbb{P}^n(\cap_{j=1}^n\{\tilde{X}_j \in B_j\}) = \Pi_{j=1}^n \mathbb{P}^n(\tilde{X}_j \in B_j),$$

so that the random variables $\{\tilde{X}_j\}_{j=1}^n$ are independent! Let $\mathscr{U}^n = \{(x_1, \cdots, x_n) : 0 \le x_j \le 1, 1 \le j \le n\}$. The trace on $\mathscr{U}^n$ of $(\mathbb{R}^n, \mathscr{B}^n, m^n)$ is a probability space. For $j = 1, \cdots, n$, let $f_j : \mathbb{R} \to \mathbb{R}$ be a Borel measurable function and let $X_j(x_1, \cdots, x_n) = f_j(x_j)$. Then, the r.v. $\{X_j\}_{j=1}^n$ are independent. If $f_j(x_j) = x_j$ then we get the n-coordinate variables in the cube.

Theorem 1.7.9 (Existence of product measures)**.**

Let $\{\mu_j\}_j$ be a finite or infinite sequence of probability measures on $(\mathbb{R}, \mathscr{B})$ or equivalently, let their distribution functions be given. There exists a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a sequence of independent r.v. $\{X_j\}_j$ defined on it such that for each j, the measure μ_j is the probability measure of X_j.

The proof of this theorem is omitted since it can be found in any book on measure theory.

1.8 Mathematical Expectation

Mathematical expectation is integration on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with respect to the probability measure $\mathbb{P}$. To avoid complications we assume that the r.v. are finite everywhere.

Definition 1.8.1. *A countable partition of Ω is a countable family of disjoint sets A_j with $A_j \in \mathscr{F}$ and such that $\Omega = \cup_{j\ge} A_j$. In this case we have that $1 = \mathbf{1}_\Omega = \sum_j \mathbf{1}_{A_j}$.*

Definition 1.8.2. *A r.v. X is said to belong to the weighted partition $\{A_j, bj\}$ is for all $\omega \in \Omega$ we have that $X(\omega) = \sum_j b_j \mathbf{1}_{A_j}(\omega)$. Note that X is a discrete r.v.*

Remark 1.8.3. *Every discrete r.v. belongs to a weighted partition: take $\{b_j\}_j$ as the countable set of the possible values of X and $A_j = \{\omega \in \Omega; X(\omega) = b_j\}$. If j ranges over a finite set the r.v. is said to be simple.*

If X is a positive discrete r.v. belonging to the weighted partition $\{A_j, bj\}$, then its expectation is defined as

$$\mathbb{E}[X] = \sum_j b_j \mathbb{P}(A_j).$$

Note that $\mathbb{E}[X]$ is a number, in this case, since $b_j \geq 0$, positive or $+\infty$. Suppose now that X is a positive random variable and for each positive integers m, n, let

$$A_{mn} = \left\{\omega : \frac{n}{2^m} \leq X(\omega) \leq \frac{n+1}{2^m}\right\} = X^{-1}\left(\left[\frac{n}{2^m}, \frac{n+1}{2^m}\right]\right),$$

so that $A_{mn} \in \mathscr{F}$. For each m, let X_m be the random variable that takes the value $\frac{n}{2^m}$ in A_{mn}, that is

$$X_m(\omega) = \frac{n}{2^m} \quad \text{iff} \quad \frac{n}{2^m} \leq X(\omega) \leq \frac{n+1}{2^m}.$$

It is easy to see that for each m we have that for all $\omega \in \Omega$, $X_m(\omega) \leq X_{m+1}(\omega)$. Now let $\omega \in \Omega$ and note that if $\frac{n}{2^m} \leq X(\omega) \leq \frac{n+1}{2^m}$, then $X_m(\omega) = \frac{n}{2^m}$, so that

$$0 \leq X(\omega) - X_m(\omega) < \frac{1}{2^m},$$

from where we get that $\lim_{m\to\infty} X_m(\omega) = X(\omega)$. So the sequence of r.v. $\{X_m\}_m$ is increasing and converges pointwisely to X.

Note that

$$\mathbb{E}[X_m] = \sum_{n=0}^{\infty} \frac{n}{2^m} \mathbb{P}\left(\frac{n}{2^m} \leq X < \frac{n+1}{2^m}\right).$$

If $\mathbb{E}[X_m] = +\infty$ then we define $\mathbb{E}[X] = +\infty$, otherwise, we define $\mathbb{E}[X] = \lim_{n\to\infty} \mathbb{E}[X_m]$. Note that the limit can be infinite.

For a general r.v. X we take $X = X^+ - X^-$, where $X^+ = X \vee 0$ and $X^- = (-X) \vee 0$. Both X^+, X^- are positive, so their expectation is defined and unless both expectations are $+\infty$ we define $\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$. We set that X has finite or infinite expectation according to $\mathbb{E}[X]$ is finite or infinite.

When the expectation of X exists we use the notation

$$\mathbb{E}[X] = \int_\Omega X(\omega)\mathbb{P}(d\omega) = \int_\Omega X(\omega)d\mathbb{P}.$$

For $\Lambda \in \mathscr{F}$ we have

$$\mathbb{E}[X\mathbf{1}_\Lambda] = \int_\Lambda X(\omega)\mathbb{P}(d\omega) = \int_\Omega \mathbf{1}_\Lambda(\omega)X(\omega)d\mathbb{P}$$

and it is called the integral of X wrt $\mathbb{P}$ over the set Λ. When the integral above exists and is finite we say that X is integrable in Λ wrt $\mathbb{P}$.

Example 13 (Lebesgue-Stietjes integral).
For $(\mathbb{R}, \mathscr{B}, \mu)$ *and* $X = f$ *and* $\omega = x$ *we have*

$$\int_\Lambda X(\omega)\mathbb{P}(d\omega) = \int_\Lambda f(x)\mu(dx) = \int_\Lambda f(x)d\mu.$$

When F *is the distribution function of* μ *we also write (for* $\Lambda = (a, b]$*)*

$$\int_{(a,b]} f(x)dF(x).$$

To distinguish the intervals $(a, b], [a, b], (a, b)$ *and* $[a, b)$ *we use the notation*

$$\int_{a+0}^{b+0}, \int_{a-0}^{b+0}, \int_{a+0}^{b-0}, \int_{a-0}^{b-0}.$$

For $(\mathscr{U}, \mathscr{B}, m)$ *the integral is* $\int_a^b f(x)m(dx) = \int_a^b f(x)dx$*. Since* μ *is atomless we do not need to distinguish the intervals.*

Let us now see some properties of the mathematical expectation. We prove some of them but the rest are left to the reader. In what follows X and Y are r.v. and $a, b \in \mathbb{R}$ and $\Lambda \in \mathscr{F}$.

(1) Absolute integrability: $\int_\Lambda X d\mathbb{P}$ is finite iff $\int_\Lambda |X| d\mathbb{P}$ is finite.

Note that $|X| = X^+ + X^-$. Suppose that $\int_\Lambda |X| d\mathbb{P} < \infty$. Then $\int_\Lambda X^\pm d\mathbb{P} < \infty$. Therefore, $\int_\Lambda X d\mathbb{P} = \int_\Lambda X^+ d\mathbb{P} - \int_\Lambda X^- d\mathbb{P} < \infty$. On the other hand if $\int_\Lambda X d\mathbb{P} < \infty$, then $\int_\Lambda X^\pm d\mathbb{P} < \infty$ and this implies that $\int_\Lambda |X| d\mathbb{P} < \infty$.

(2) Linearity: $\int_\Lambda (aX + bY) d\mathbb{P} = a\int_\Lambda X d\mathbb{P} + b\int_\Lambda Y d\mathbb{P}$, as long as the right hand side makes sense, that is, it is not $+\infty - \infty$ nor $-\infty + \infty$.

(3) Set additivity: If $\{\Lambda_n\}_{n\geq 1}$ are disjoint, then

$$\int_{\cup_{n\geq 1}\Lambda_n} X d\mathbb{P} = \sum_{n\geq 1}\int_{\Lambda_n} X d\mathbb{P}.$$

(4) Positivity: If $X \geq 0$ a.e. on Λ (this means there is a subset of Λ with weight one wrt $\mathbb{P}$ where X is positive), then

$$\int_{\Lambda} X d\mathbb{P} \geq 0.$$

(5) Monotonicity: If $X_1 \leq X \leq X_2$ a.e. in Λ, then

$$\int_{\Lambda} X_1 d\mathbb{P} \leq \int_{\Lambda} X d\mathbb{P} \leq \int_{\Lambda} X_2 d\mathbb{P}.$$

To prove it, apply the previous item for $X - X_1$ and for $X_2 - X$.

(6) Mean value Theorem: If $a \leq X \leq b$ a.e. in Λ, then

$$a\mathbb{P}(\Lambda) \leq \int_{\Lambda} X d\mathbb{P} \leq b\mathbb{P}(\Lambda).$$

To prove it, apply the previous item for $X_1 = a$ and for $X_2 = b$.

(7) Modulus inequality: $\left|\int_{\Lambda} X d\mathbb{P}\right| \leq \int_{\Lambda} |X| d\mathbb{P}$.

To prove the result note that $\int_{\Lambda} X d\mathbb{P} = \int_{\Lambda} X^+ - X^- d\mathbb{P} = \int_{\Lambda} X^+ d\mathbb{P} - \int_{\Lambda} X^- d\mathbb{P}$. Then
$\left|\int_{\Lambda} X d\mathbb{P}\right| = \left|\int_{\Lambda} X^+ d\mathbb{P}\right| + \left|\int_{\Lambda} X^- d\mathbb{P}\right| = \int_{\Lambda} X^+ d\mathbb{P} + \int_{\Lambda} X^- d\mathbb{P} = \int_{\Lambda} |X| d\mathbb{P}$.

(8) Dominated convergence Theorem: If $\lim_{n\to\infty} X_n = X$ a.e. on Λ and if for $n \geq 1$ $|X_n| \leq Y$ a.e. on Λ and $\int_{\Lambda} Y d\mathbb{P} < \infty$, then

$$\lim_{n\to\infty}\int_{\Lambda} X_n d\mathbb{P} = \int_{\Lambda} X d\mathbb{P} = \int_{\Lambda} \lim_{n\to\infty} X_n d\mathbb{P}.$$

To prove the result note that $|X_n| \leq Y$ implies that if $X_n \geq 0$, then $X_n \leq Y$ and if $X_n < 0$, then $-X_n \leq Y$. Then, from Fatou's lemma we have that

$$\int_\Lambda Y\,d\mathbb{P} + \int_\Lambda X\,d\mathbb{P} = \int_\Lambda Y + X\,d\mathbb{P} \leq \liminf_n \int_\Lambda Y + X_n d\mathbb{P}$$
$$= \int_\Lambda Y\,d\mathbb{P} + \liminf \int_\Lambda X_n d\mathbb{P}$$

From here we conclude that

$$\int_\Lambda X\,d\mathbb{P} \leq \liminf \int_\Lambda X_n d\mathbb{P}$$

Now we repeat the argument with $-X_n$and we conclude that

$$\int_\Lambda -X\,d\mathbb{P} \leq \liminf \int_\Lambda -X_n d\mathbb{P}$$

so that

$$\int_\Lambda X\,d\mathbb{P} \geq \limsup \int_\Lambda X_n d\mathbb{P}$$

from where the equality follows.

(9) Bounded convergence Theorem: If $\lim_{n\to\infty} X_n = X$ a.e. on Λ and there exists a constant M such that $n \geq 1$ $|X_n| \leq M$ a.e. on Λ, then the result of (8) is true.

(10) Monotone convergence Theorem: If $X_n \geq 0$ and $X_n \uparrow X$ a.e. on Λ, then the previous equality is true if we allow $+\infty$ as a value.

To prove the theorem we note at first that since $X_n \uparrow X$, then the limit $\lim_{n\to\infty} X_n(\omega)$ exists for each ω, being possibly equal to infinity. Note that, by a previous Theorem X is a r.v. On the other hand, by the monotonicity property we have $\int_\Lambda X_n d\mathbb{P} \leq \int_\Lambda X\,d\mathbb{P}, \forall n$ and note that $\{\int_\Lambda X_n d\mathbb{P}\}_{n\in\mathbb{N}}$ is an increasing sequence. Let $\Phi := \lim_n \int_\Lambda X_n d\mathbb{P}$, which exists and note that $\Phi \leq \int_\Lambda X\,d\mathbb{P}$. Now we have to prove the reversed inequality. Let $\alpha \in (0,1)$ and let φ be a simple function which is positive with $\varphi \leq X$. Let $E_n = \{\omega : X_n(\omega) \geq \alpha\varphi(\omega)\}$. Note that $\{E_n\}_{n\in\mathbb{N}}$ is an increasing sequence of sets in $\mathscr{F}$, whose union is Ω. Moreover

$$\int X_n d\mathbb{P} \geq \int_{E_n} X_n d\mathbb{P} \geq \alpha \int_{E_n} \varphi\,d\mathbb{P}.$$

Note that if $I(\cdot)=\int_{\cdot}\varphi d\mathbb{P}$, then $I(\cdot)$ is a measure and the limit $\lim_n I(E_n)$ exists and

$$\lim_n I(E_n)=\int \varphi d\mathbb{P}.$$

Then $\Phi \geq \alpha\int \varphi d\mathbb{P}$ for all $\alpha\in(0,1)$. Sending α to 1 we obtain that $\Phi\geq\int\varphi d\mathbb{P}$ for all φ simple and positive with $\varphi\leq X$ and this implies that $\Phi\geq\int X d\mathbb{P}$, since

$$\int X d\mathbb{P}=\sup\left\{\int \varphi d\mathbb{P}\ :\ 0\leq\varphi\leq X;\ \varphi \text{ simple and positive}\right\};$$

so that the proof ends.

(11) Integration term by term: If $\sum_{n\geq 1}\int_\Lambda |X_n| d\mathbb{P}<\infty$, then $\sum_{n\geq 1}|X_n|<\infty$ a.e. on Λ, so that $\sum_{n\geq 1}X_n$ converges a.e. on Λ and

$$\int_\Lambda \sum_{n\geq 1} X_n d\mathbb{P}=\sum_{n\geq 1}\int_\Lambda X_n d\mathbb{P}.$$

(12) Fatou's Lemma: If $X_n\geq 0$ a.e. on Λ, then

$$\int_\Lambda (\liminf_{n\to\infty} X_n) d\mathbb{P}\leq \liminf_{n\to\infty}\int_\Lambda X_n d\mathbb{P}.$$

To prove Fatou's lemma we do the following. For $k\in\mathbb{Z}^+$ if $j\geq k$ then $\inf_{n\geq k}X_n\leq X_j$. Then, from the monotonicity property we have for all $j\geq k$ that

$$\int_\Lambda \inf_{n\geq k} X_n d\mathbb{P}\leq\int_\Lambda X_j d\mathbb{P},$$

which implies that

$$\int_\Lambda \inf_{n\geq k} X_n d\mathbb{P}\leq\inf_{j\geq k}\int_\Lambda X_j d\mathbb{P}.$$

Since $\inf_{n\geq k}X_n\uparrow \liminf X_n$ when $k\to\infty$, then taking $k\to\infty$ we have that

$$\begin{aligned}\int_\Lambda \liminf_{n\to\infty} X_n d\mathbb{P}&=\int_\Lambda \lim_{k\to\infty}\inf_{n\geq k} X_n d\mathbb{P}=\lim_{k\to\infty}\int_\Lambda \inf_{n\geq k}X_n d\mathbb{P}\\ &\leq \lim_{k\to\infty}\inf_{n\geq k}\int_\Lambda X_n d\mathbb{P}=\liminf_{n\to\infty}\int_\Lambda X_n d\mathbb{P}.\end{aligned}$$

Note that in the previous inequality we used the Monotone Convergence Theorem.

- Digression on the Riemann-Stieltjes integral

Let f be a continuous function defined on $[a,b]$ and let F be a distribution function. The Riemann-Stieltjes integral of f on $[a,b]$ wrt F is defined as the limit of the Riemann sums of the form

$$\sum_{i=1}^{n} f(\tilde{x}_i)(F(x_{i+1})-F(x_i)), \tag{1.8.1}$$

where $x_1 = a, x_n = b$, $x_i < x_{i+1}$ and $\tilde{x}_i$ is an arbitrary point in $[x_i, x_{i+1}]$. The limit is takes by making the norm of the partition $\{x_i\}_i$ tending to 0, that is $\max_{i=1,\cdots,n}(x_{i+1}-x_i) \to 0$. The limit exists, when f is continuous, and it is denoted by $\int_a^b f(x)dF(x)$. Note that

$$\int_{\mathbb{R}} f(x)dF(x) = \lim_{a\to-\infty, b\to+\infty} \int_a^b f(x)dF(x).$$

Example 14. *Compute $\int_{\mathbb{R}} F_0(x)dF_0(x)$, for $F_0(x)=\delta_0(x)$, the Heaviside function at* 0.

Note that the integral above does not exist since the limit in (1.8.1) *does not exist. This is because if we take $x_i < 0 < x_{i+1}$ for some i, then $F_0(x_{i+1})-F_0(x_i) = 1$ and as a consequence, the value of the sum:*

$$\sum_{i=1}^{n} F_0(\tilde{x}_i)(F(x_{i+1})-F(x_i)),$$

depends on whether $\tilde{x}_i \in [x_i, x_{i+1}]$ is such that $F_0(\tilde{x}_i)$ is 0 *or* 1.

To avoid the previous cases, in order to extend the definition to discontinuous functions we do it like this. Let f be a Borel measurable function $f:\mathbb{R}\to\mathbb{R}$. We want to define $\int_{\mathbb{R}} f(x)dF(x)$ for a distribution function F. First we define it for $f(x) = \mathbf{1}_{[a,b]}(x)$ as $\int_{\mathbb{R}} f(x)dF(x) = F(b)-F(a)$. Then we extend the definition as we did before.

Remark 1.8.4.

- *When F is the distribution function of a discrete random variable X taking values $\{x_i\}_{i\geq 1}$ then*

$$\int_{\mathbb{R}} f(x)dF(x) = \sum_{i\geq 1} f(x_i)\mathbb{P}(X = x_i)$$

and

$$\int_{(a,b]} f(x)dF(x) = \sum_{i:a<x_i\leq b} f(x_i)\mathbb{P}(X = x_i).$$

• *When F is the distribution function of an absolutely continuous random variable X with density f_X, then*

$$\int_{\mathbb{R}} f(x)dF(x) = \int_{\mathbb{R}} f(x)f_X(x)dx$$

and

$$\int_a^b f(x)dF(x) = \int_a^b f(x)f_X(x)dx.$$

• *When $F = \alpha F_d + \beta F_{ac} + \gamma F_s$, then*

$$\int_{\mathbb{R}} f(x)dF(x) = \alpha \int_{\mathbb{R}} f(x)dF_d + \beta \int_{\mathbb{R}} f(x)dF_{ac} + \gamma \int_{\mathbb{R}} f(x)dF_s.$$

• *When F does not have singular part we have*

$$\int_a^b f(x)dF(x) = \sum_{i:a<x_i\leq b} f(x_i)\mathbb{P}(X = x_i) + \int_a^b f(x)f_X(x)dx.$$

Proposition 1.8.5. *For a r.v. X with distribution function F we have that*

$$\mathbb{E}[X] = \int_0^{+\infty} (1 - F(x))dx - \int_{-\infty}^0 F(x)dx.$$

Proof. Note that $\mathbb{E}[X] = \int_{\mathbb{R}} xdF(x)$.

First we claim that $\int_0^{+\infty} xdF(x) = \int_0^{+\infty}(1 - F(x))dx$. Since $d(xF(x)) = xdF(x) + F(x)dx$ we have, for $b > 0$, that

$$\int_0^b d(xF(x)) = \int_0^b xdF(x) + \int_0^b F(x)dx.$$

The term on the left hand side of last expression is equal to $bF(b)$ so that we obtain

$$\int_0^b xdF(x) = \int_0^b F(b) - F(x)dx \leq \int_0^b (1 - F(x))dx \leq \int_0^{+\infty} (1 - F(x))dx.$$

From here it follows that

$$\int_0^{+\infty} x dF(x) = \lim_{b\to+\infty} \int_0^b x dF(x) \le \int_0^{+\infty} (1-F(x))dx.$$

On the other hand let $a > 0$ and with $b > a$. Then

$$\int_0^b F(b) - F(x)dx \ge \int_0^a F(b) - F(x)dx = \int_0^a F(b) - 1dx + \int_0^a 1 - F(x)dx$$
$$= a(F(b)-1) + \int_0^a 1 - F(x)dx$$

From here it follows that

$$\int_0^{+\infty} x dF(x) = \lim_{b\to\infty} \int_0^b x dF(x) = \lim_{b\to+\infty} \int_0^b F(b) - F(x)dx$$
$$\ge a(F(b)-1) + \int_0^a 1 - F(x)dx = \int_0^a 1 - F(x)dx.$$

Since last inequality holds for any $a > 0$ we get that

$$\int_0^{+\infty} x dF(x) \ge \lim_{a\to+\infty} \int_0^a 1 - F(x)dx = \int_0^{+\infty} 1 - F(x)dx.$$

Putting together the previous two inequalities we prove the claim. Analogously we can show that $\int_{-\infty}^0 x dF(x) = -\int_{-\infty}^0 F(x)dx$. This is left to the reader. An alternative proof consists in first showing that for any non-negative r.v. we have that $\mathbb{E}[X] = \int_0^{+\infty} 1 - F(x)dx$. Once this is proved we take $X = X^+ - X^-$ and use the fact that both X^+ and X^- are positive to conclude that $\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-] = \int_0^{+\infty} 1 - F_{X^+}(x)dx - \int_0^{+\infty} 1 - F_{X^-}(x)dx$. Then we need to relate $F_{X^+}(x)$ and $F_{X^-}(x)$. This is left to the reader. □

Corollary 1.8.6. *For a non-negative r.v. X, we have that*

$$\mathbb{E}[X] = \int_0^{+\infty} \mathbb{P}(X > x)dx.$$

Proof. It is enough to note that r.v. X we have that $F(x) = 0$ for $x < 0$, since X is non-negative. From this observation the result follows. □

Theorem 1.8.7. *Let $k \in \mathbb{N}$, then*

$$\mathbb{E}[X^k] = k\Big[\int_0^{+\infty} (1 - F_X(x))x^{k-1}dx - \int_{-\infty}^0 F_X(x)x^{k-1}dx\Big].$$

Proof. We do the proof for k even but we note that for k odd it is completely analogous. Since k is even we note that X^k is a non-negative r.v., therefore from last corollary it follows that

$$\begin{aligned}\mathbb{E}[X^k] = \int_0^{+\infty} \mathbb{P}(X^k > x)dx &= \int_0^{+\infty} \mathbb{P}(X > x^{1/k})dx + \int_0^{+\infty} \mathbb{P}(X < -x^{1/k})dx \\ &= \int_0^{+\infty} 1 - F_X(x^{1/k})dx + \int_0^{+\infty} F_X((-x^{1/k})^-)dx.\end{aligned}$$

Now doing the change of variables $x = y^k$, last expression writes as

$$\int_0^{+\infty} 1 - F_X(y)ky^{k-1}dy + \int_0^{+\infty} F_X((-y)^-)ky^{k-1}dy.$$

Now note that $F_X(-y)$ and $F_X((-y)^-)$ are monotone functions which are equal except in a set of points which are the jump points, therefore they are equal except at most on a countable set of points, so that the last expression coincides with

$$\int_0^{+\infty} 1 - F_X(y)ky^{k-1}dy + \int_0^{+\infty} F_X(-y)ky^{k-1}dy,$$

and doing the change of variables $-y = u$ last expression equals to

$$\int_0^{+\infty} 1 - F_X(y)ky^{k-1}dy - \int_{-\infty}^0 F_X(u)ku^{k-1}du.$$

This ends the proof. □

Theorem 1.8.8 (Integrability criterion)**.** *For a r.v. X we have that*

$$\sum_{n\geq 1} \mathbb{P}(|X| \geq n) \leq \mathbb{E}[|X|] \leq 1 + \sum_{n\geq 1} \mathbb{P}(|X| \geq n),$$

so that $\mathbb{E}[|X|] < \infty$ iff the series above converges.

Proof. Let $\Lambda_n := \{\omega : X(\omega) \in [n, n+1)\}$. Note that Λ_n are disjoint sets so that

$$\mathbb{E}[|X|] = \sum_{n=0}^{+\infty} \int_{\Lambda_n} |X| d\mathbb{P}.$$

Applying the mean value theorem to each integral we obtain that

$$\sum_{n=0}^{+\infty} n\mathbb{P}(\Lambda_n) \le \mathbb{E}[|X|] \le \sum_{n=0}^{+\infty} (n+1)\mathbb{P}(\Lambda_n) = 1 + \sum_{n=0}^{+\infty} n\mathbb{P}(\Lambda_n). \tag{1.8.2}$$

To finish the proof it is enough to show that

$$\sum_{n=0}^{+\infty} n\mathbb{P}(\Lambda_n) = \sum_{n=0}^{+\infty} \mathbb{P}(|X| \ge n) \tag{1.8.3}$$

being its value finite or infinite. For that purpose let us fix m. Then, truncating the series at m we have that

$$\begin{aligned}
\sum_{n=0}^{m} n\mathbb{P}(\Lambda_n) &= \sum_{n=0}^{m} n\Big(\mathbb{P}(|X| \ge n) - \mathbb{P}(|X| \ge n+1)\Big) \\
&= \sum_{n=0}^{m} n\mathbb{P}(|X| \ge n) - \mathbb{P}(|X| \ge n) - \sum_{n=1}^{m+1} (n-1)\mathbb{P}(|X| \ge n) \\
&= \sum_{n=1}^{m} \mathbb{P}(|X| \ge n) - m\mathbb{P}(|X| \ge m+1)
\end{aligned}$$

Then, since $m\mathbb{P}(|X| \ge m+1) \ge 0$, we have that

$$\sum_{n=1}^{m} n\mathbb{P}(\Lambda_n) \le \sum_{n=1}^{m} \mathbb{P}(|X| \ge n) = \sum_{n=1}^{m} n\mathbb{P}(\Lambda_n) + m\mathbb{P}(|X| \ge m+1).$$

On the other hand $m\mathbb{P}(|X| \ge m+1) \le \int_{\{|X| \ge m+1\}} |X| \mathbb{P}$, so that if $\mathbb{E}[X] < +\infty$, then the term on the right hand side of last equality vanishes as $n \to \infty$. Then (1.8.3) follows with both series finite. Now, if $\mathbb{E}[|X|] = +\infty$, since (1.8.2) holds, then

$$\sum_{n=1}^{+\infty} n\mathbb{P}(\Lambda_n) = +\infty.$$

From here it follows that $\sum_{n=0}^{m} n\mathbb{P}(|X| \ge n) = +\infty$ diverges with m since

$$\sum_{n=1}^{m} n\mathbb{P}(\Lambda_n) \le \sum_{n=1}^{m} \mathbb{P}(|X| \ge n) \le \sum_{n=1}^{m} n\mathbb{P}(\Lambda_n) + m\mathbb{P}(|X| \ge m+1).$$

This ends the proof. □

Lemma 1.8.9. *If X is a non-negative r.v. which takes only integer values, then*

$$\mathbb{E}[X] = \sum_{n\geq 1} \mathbb{P}(X \geq n).$$

Proof. From Corollary 1.8.6 we have that

$$\mathbb{E}[X] = \int_0^{+\infty} \mathbb{P}(X > x) = \sum_{n=0}^{+\infty}(1 - F(n)) = \sum_{n=0}^{+\infty} \mathbb{P}(X > n) = \sum_{n=1}^{+\infty} \mathbb{P}(X \geq n).$$

This result can also be seen as a corollary to the previous Theorem. To do that note

$$\mathbb{E}[X] = \mathbb{E}[|X|] \leq 1 + \sum_{n=1}^{+\infty} \mathbb{P}(|X| \geq n) = \sum_{n=0}^{+\infty} \mathbb{P}(X \geq n).$$

Now we write the term at the right hand side of last expression as

$$\sum_{n=0}^{+\infty} \sum_{j=n}^{+\infty} \mathbb{P}(X = j)$$

and we use Fubini's Theorem to get that

$$\mathbb{E}[X] \leq \sum_{j=0}^{+\infty} \sum_{n=0}^{j} \mathbb{P}(X = j) = \sum_{j=0}^{+\infty} j\mathbb{P}(X = j) = \sum_{j=1}^{+\infty} j\mathbb{P}(X = j).$$

Repeating these arguments and noting that

$$\sum_{j=1}^{+\infty} j\mathbb{P}(X = j) = \sum_{n=1}^{+\infty} \mathbb{P}(X \geq n),$$

we obtain from (1.8.2) that

$$\sum_{n=1}^{+\infty} \mathbb{P}(X \geq n) \leq \mathbb{E}[X] \leq \sum_{n=1}^{+\infty} \mathbb{P}(X \geq n),$$

from where the result follows. □

There is a basic relation between an integral wrt $\mathbb{P}$ over sets of $\mathscr{F}$ and the Lebesgue-Stieltjes integral wrt to μ over sets of $\mathscr{B}$.

Theorem 1.8.10. *Let X be a r.v. defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ which induces the probability space $(\mathbb{R}, \mathscr{B}, \mu)$ and let $f : \mathbb{R} \to \mathbb{R}$ be a Borel measurable function. Then,*

$$\int_\Omega f(X(\omega))\mathbb{P}(d\omega) = \int_\mathbb{R} f(x)\mu(dx)$$

as long as both sides exist.

Proof. To prove the result we use the classical argument as we used in the construction of the integral. First we prove the result for a function $f = \mathbf{1}_B$ with $B \in \mathscr{B}$. In this case we have that

$$\int_\mathbb{R} f(x)\mu(dx) = \mu(B) \quad \text{and} \quad \int_\Omega f(X(\omega))\mathbb{P}(d\omega) = \mathbb{P}(X \in B)$$

and the equality holds from the definition of μ which was defined as the push-forward of $\mathbb{P}$. Now, from linearity the equality is going to be true for functions of the form $f = \sum_j b_j \mathbf{1}_{B_j}$ where $b_j \in \mathbb{R}$ and $B_j \in \mathscr{B}$. In the case where f is a general positive Borel-measurable function we take a sequence $\{f_m\}_{m\in\mathbb{N}}$ of functions of the form $f_m = \sum_j b_j^m \mathbf{1}_{B_j^m}$ as above, in such a way that $f_m \uparrow f$ and for each f_m we have that

$$\int_\Omega f_m(X(\omega))\mathbb{P}(d\omega) = \int_\mathbb{R} f_m(x)\mu(dx)$$

and from the monotone convergence Theorem we conclude the result for positive functions. To prove it for general f we use the decomposition $f = f^+ - f^-$ and the equality follows. □

In higher dimensions the result is the same. We state it for $d = 2$ as

Theorem 1.8.11. *Let (X, Y) be a random vector defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ which induces the probability space $(\mathbb{R}^2, \mathscr{B}^2, \nu)$ and let $f : \mathbb{R}^2 \to \mathbb{R}$ be a Borel measurable function. Then,*

$$\int_\Omega f(X(\omega), Y(\omega))\mathbb{P}(d\omega) = \iint_{\mathbb{R}^2} f(x, y)\nu(dx, dy).$$

We do not show the previous theorem here since it is exactly the same proof as in the one-dimensional case.

From the previous theorem, for a r.v. X with distribution function F_X and distribution measure μ_X, it holds that

$$\mathbb{E}[X] = \int_{\mathbb{R}} x\mu_X(dx) = \int_{\mathbb{R}} x dF_X(x)$$

and more generally $\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x)\mu_X(dx) = \int_{\mathbb{R}} f(x)dF_X(x)$.

Remark 1.8.12. *An important consequence of the previous theorem is that for $f(x,y) = x + y$ we obtain that (linearity of the integral)*

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

To prove the previous equality we note that for $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x,y) = x + y$ we have that

$$\mathbb{E}[X+Y] = \int_{\Omega} X + Y\mathbb{P}(d\omega) = \iint_{\mathbb{R}^2} (x+y)\nu(dx,dy).$$

The integral at the right hand side of last equality is equal to

$$\iint_{\mathbb{R}^2} x\,\nu(dx,dy) + \iint_{\mathbb{R}^2} y\,\nu(dx,dy).$$

On the other hand if $f(x,y) = x$ we obtain that

$$\mathbb{E}[X] = \int_{\Omega} X\mathbb{P}(d\omega) = \iint_{\mathbb{R}^2} x\,\nu(dx,dy)$$

and the same is true when we take $f(x,y) = y$ from where the result follows.

Definition 1.8.13. *Let $a \in \mathbb{R}$ and $r \geq 0$. The absolute moment of a r.v. X of order r about a is defined as $\mathbb{E}[|X-a|^r]$.*

Remark 1.8.14. *If μ_X and F_X are the distribution measure and the distribution function of X, then*

$$\mathbb{E}[|X-a|^r] = \int_{\mathbb{R}} |x-a|^r \mu(dx) = \int_{\mathbb{R}} |x-a|^r dF_X(x),$$

$$\mathbb{E}[(X-a)^r] = \int_{\mathbb{R}} (x-a)^r \mu(dx) = \int_{\mathbb{R}} (x-a)^r dF_X(x).$$

When $r = 1$ and $a = 0$, the previous moment is $\mathbb{E}[X]$. The moments about $a = \mathbb{E}[X]$ are called central moments and the one of order 2 is called the variance:

$$Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Definition 1.8.15 (The space $\mathbb{L}^p = \mathbb{L}^p(\Omega, \mathscr{F}, \mathbb{P})$)**.** *For a positive number p, we say that $X \in \mathbb{L}^p$ iff $\mathbb{E}[|X|^p] < \infty$.*

Theorem 1.8.16. *Let X and Y be random variables and p, q such that $1 < p < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$ (p and q are said to be conjugate). Then*

1. *(Holder's Inequality)*

$$|\mathbb{E}[XY]| \leq \mathbb{E}[|XY|] \leq (\mathbb{E}[|X|^p])^{1/p} \mathbb{E}[|Y|^q])^{1/q} \tag{1.8.4}$$

2. *(Minkowski's inequality)*

$$(\mathbb{E}[|X+Y|^p])^{1/p} \leq (\mathbb{E}[|X|^p])^{1/p} + (\mathbb{E}[|Y|^p])^{1/p}$$

Remark 1.8.17. *When $p = 2$,* (1.8.4) *is called the Cauchy-Schwarz's inequality.*

Do the proof of the previous result.

Theorem 1.8.18 (Jensen's inequality)**.** *If $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function and X and $\varphi(X)$ are integrable r.v. then*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)].$$

Proof. To prove the theorem first note that since φ is convex then it is continuous. Then, we know that $\varphi(X)$ is a r.v. Now convexity means that there exist $\{\lambda_j\}_{j=1,\cdots,n}$ and $\{y_j\}_{j=1,\cdots,n}$ such that

$$\varphi\Big(\sum_{j=1}^{n} \lambda_j y_j\Big) \leq \sum_{j=1}^{n} \lambda_j \varphi(y_j).$$

We prove this theorem by using the classical argument that we used in the construction of the integral. First, we prove it for a simple r.v. X, then we approximate a positive r.v. X by a sequence of simple functions and use Monotone's convergence theorem and finally we use the equality $X = X^+ - X^-$. To prove it for a simple r.v. suppose that X is a r.v. taking the values y_j with probability λ_j with $j = 1, \cdots, n$. Since $\mathbb{E}[X] = \sum_{j=1}^n y_j \lambda_j$ and since $\mathbb{E}[\varphi(X)] = \sum_{j=1}^n \varphi(y_j)\lambda_j$, then we result follows by convexity of φ. □

Example 15.

1. $\varphi(x) = |x|$;
2. $\varphi(x) = x^2$;
3. $\varphi(x) = |x|^p$, $p \geq 1$.

Theorem 1.8.19 (Tchebychev's Basic inequality)**.** *Let X be a non-negative r.v. For any $\lambda > 0$,*

$$\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}[X]}{\lambda}.$$

Proof. To prove the result note that

$$\mathbb{E}[X] = \int_\Omega X d\mathbb{P} = \int_{\{X \geq \lambda\}} X d\mathbb{P} + \int_{\{X < \lambda\}} X d\mathbb{P}.$$

Since X is non-negative we obtain that

$$\mathbb{E}[X] \geq \int_{\{X \geq \lambda\}} X d\mathbb{P} \geq \lambda \mathbb{P}(X \geq \lambda).$$

□

Theorem 1.8.20 (Tchebychev's Classic inequality)**.** *Let X be a r.v. with finite variance. For any $\lambda > 0$,*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \lambda) \leq \frac{Var(X)}{\lambda^2}.$$

Proof. To prove the result it is enough to note that $|X - \mathbb{E}[X]| \geq \lambda$ implies that $|X - \mathbb{E}[X]|^2 \geq \lambda^2$ and use the previous Theorem. □

Theorem 1.8.21 (Markov's inequality)**.** *Let X be a r.v. with $\mathbb{E}[|X|]^t < \infty$. For any $\lambda > 0$,*

$$\mathbb{P}(|X| \geq \lambda) \leq \frac{\mathbb{E}[|X|^t])}{\lambda^t}.$$

Proof. To prove the result follow the same argument as in the previous proof.

□

General Tchebychev's inequality

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a strictly positive and increasing function in $(0, +\infty)$, such that $\varphi(u) = \varphi(-u)$ and let X be a r.v. such that $\mathbb{E}[\varphi(X)] < +\infty$. then, for each $u > 0$ it holds that

$$\mathbb{P}(|X| \geq u) \leq \frac{\mathbb{E}[\varphi(X)]}{\varphi(u)}.$$

Theorem 1.8.22. *If X and Y are two independent r.v. with finite expectation, then*

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

Proof. We prove again this theorem using the classical argument starting by showing it for discrete r.v. For that purpose, let X, Y be discrete such that

$$\Lambda_j = \{\omega : X(\omega) = b_j\}$$

$$\tilde{\Lambda}_k = \{\omega : Y(\omega) = a_k\}.$$

Then $\mathbb{E}[X] = \sum_j b_j \mathbb{P}(\Lambda_j)$ and $\mathbb{E}[Y] = \sum_k a_k \mathbb{P}(\tilde{\Lambda}_k)$. Now note that XY is a discrete r.v. and $XY(\omega) = a_k b_j$ for $\omega \in \Lambda_j \cap \tilde{\Lambda}_k$. Then,

$$\mathbb{E}[XY] = \sum_{j,k} a_k b_j \mathbb{P}(\Lambda_j \cap \tilde{\Lambda}_k).$$

Since $\mathbb{P}(\Lambda_j \cap \tilde{\Lambda}_k) = \mathbb{P}(X = b_j, Y = a_k)$ and by independence we obtain that

$$\begin{aligned}\mathbb{E}[XY] = \sum_{j,k} a_k b_j \mathbb{P}(X = b_j)\mathbb{P}(Y = a_k) &= \sum_j b_j \mathbb{P}(X = b_j) \sum_k a_k \mathbb{P}(Y = a_k) \\ &= \mathbb{E}[X]\mathbb{E}[Y].\end{aligned} \tag{1.8.5}$$

Now we extend the result to positive r.v. with finite mean. Then, we know (recall the argument that we used when constructing the integral) that there exist X_m and Y_m discrete r.v. such that $\mathbb{E}[X_m]\uparrow\mathbb{E}[Y]$ and $\mathbb{E}[Y_m]\uparrow\mathbb{E}[Y]$. To see that they are also independent note that

$$\begin{aligned}\mathbb{P}\Big(X_m=\tfrac{j}{2^m},Y_m=\tfrac{k}{2^m}\Big)&=\mathbb{P}\Big(\tfrac{j}{2^m}\leq X<\tfrac{j+1}{2^m},\tfrac{k}{2^m}\leq Y<\tfrac{k+1}{2^m}\Big)\\&=\mathbb{P}\Big(\tfrac{j}{2^m}\leq X<\tfrac{j+1}{2^m}\Big)\mathbb{P}\Big(\tfrac{k}{2^m}\leq Y<\tfrac{k+1}{2^m}\Big)\\&=\mathbb{P}\Big(X_m=\tfrac{j}{2^m}\Big)\mathbb{P}\Big(Y_m=\tfrac{k}{2^m}\Big).\end{aligned}\tag{1.8.6}$$

In the second equality above we used the independence of X and Y. Another way to show that X_m and Y_m are independent is to see that $X_m=\frac{[2^mX]}{2^m}$ and $Y_m=\frac{[2^mY]}{2^m}$ and since they are functions of X and Y, their independence follows. Finally, we have that $0\leq XY-X_mY_m=X(Y-Y_m)+Y_m(X-X_m)$ and by the Monotone Convergence Theorem we conclude that

$$\mathbb{E}[XY]=\lim_{m\to+\infty}\mathbb{E}[X_mY_m]=\lim_{m\to+\infty}\mathbb{E}[X_m]\mathbb{E}[Y_m]=\mathbb{E}[X]\mathbb{E}[Y].$$

So far the result is true for positive r.v. For the general case we take $X=X^+-X^-$ and $Y=Y^+-Y^-$. Since X and Y are independent we get that X^+ and Y^+ are independent and also X^- and Y^-. To conclude note that

$$\mathbb{E}[XY]=\mathbb{E}[(X^+-X^-)(Y^+-Y^-)]$$

and expand the product and use the independence. □

Remark 1.8.23. *We note that a short proof of the previous result can be derived by using Fubini's Theorem. For that purpose, use that*

$$\mathbb{E}[XY]=\int_\Omega XY\,d\mathbb{P}=\iint_{\mathbb{R}^2}xy\,\nu(dx,dy).$$

Since $\nu=\mu_X\times\mu_Y$ *the term at the right hand side of last equality is equal to*

$$\iint_{\mathbb{R}^2}xy\mu_X(dx)\mu_Y(dy)=\int_{\mathbb{R}}x\mu_X(dx)\int_{\mathbb{R}}y\mu_Y(dy)=\mathbb{E}[X]\mathbb{E}[Y].$$

Corollary 1.8.24. *If* $\{X_j\}_{j=1,\cdots,n}$ *are independent r.v. with finite mean, then*

$$\mathbb{E}\Big[\prod_{j=1}^n X_j\Big]=\prod_{j=1}^n\mathbb{E}[X_j].$$

We leave the proof of the previous result to the reader. Now we introduce the notion of correlation which measure how r.v. can affect each other.

Definition 1.8.25. *Let X and Y be r.v. with finite expectation. The covariance between X and Y is defined as*

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

When $Cov(X,Y) = 0$, we say that X and Y are uncorrelated.

Remark 1.8.26. *Be careful: uncorrelation does NOT imply independence.*

Example 16. *Analyze the case when (X,Y) has density given by*

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\Big(\big(\frac{x-\mu_1}{\sigma_1}\big)^2 - 2\rho\frac{x-\mu_1}{\sigma_1}\frac{y-\mu_2}{\sigma_2} + \big(\frac{y-\mu_2}{\sigma_2}\big)^2\Big)}$$

and take $\rho = 0$.

Proposition 1.8.27. *Let $X_1, \cdots, X_n$ be integrable r.v. such that $Cov(X_i, X_j) = 0$ for $i \neq j$. Then*

$$Var(X_1 + \cdots + X_n) = \sum_{i=1}^{n} Var(X_i).$$

Proof. First let us suppose that the X_j's have zero mean. Then

$$Var\Big(\sum_{j=1}^{n} X_j\Big) = \mathbb{E}\Big[\Big(\sum_{j=1}^{n} X_j\Big)^2\Big] = \mathbb{E}\Big[\sum_{j=1}^{n} X_j^2\Big] + \mathbb{E}\Big[\sum_{i\neq j=1}^{n} X_i X_j\Big].$$

Since the r.v. have $Cov(X_i, X_j) = 0$ for $i \neq j$ then the term at the right hand side of last equality is equal to zero so that we conclude the proof. Now for general X_j we take $Y_j = X_j - \mathbb{E}[X_j]$ and note that Y_j's are mean zero. So, apply the first part of the proof to Y_j and conclude. □

Example 17. *Let X and Y be r.v. with finite variance: show that if X and Y are independent, then*

$$Var(XY) = Var(X)Var(Y) + (\mathbb{E}[X])^2 Var(Y) + (\mathbb{E}[Y])^2 Var(X).$$

TÉCNICO LISBOA

1.9 Exercises

1.9.1 Exercises on set theory

Exercise 1:

Show that, if $\mathscr{A}$ and $\mathscr{B}$ are two σ-algebras, then $\mathscr{A} \bigcap \mathscr{B}$ is also a σ-algebra.

Exercise 2:

Let $\Omega := \{\omega_1, \omega_2, \omega_3\}$ be a sample space.

1. Exhibit all the σ-algebras of Ω.
2. Compute $\sigma(\{\omega_1\})$. Check that it is a σ-algebra.

Exercise 3:

Recall that, for a topological space S the Borel σ-algebra $\mathscr{B}(S)$ is generated by the family of open subsets of S. Prove that the Borel σ-algebra of $\mathbb{R}$ is generated by $\pi(\mathbb{R}) = \{(-\infty, x] : x \in \mathbb{R}\}$.

Exercise 4:

Let X be a random variable defined on a sample space Ω. Compute $\sigma(X)$, that is the σ-algebra generated by X, when

1. $\Omega := \{\omega_1, \omega_2, \omega_3\}$ and $X(\omega_1) = X(\omega_2) = X(\omega_3) = 1$.
2. $\Omega := \{\omega_1, \omega_2, \omega_3\}$ and $X(\omega_1) = 0$, $X(\omega_2) = 1$ and $X(\omega_3) = 2$.
3. $\Omega := \{\omega_1, \omega_2, \omega_3\}$ and $X(\omega_1) = 0$, $X(\omega_2) = 0$ and $X(\omega_3) = 1$.
4. $\Omega := \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and $X(\omega_1) = 0$, $X(\omega_2) = 0$, $X(\omega_3) = 1$ and $X(\omega_4) = 2$.

Exercise 5:

Let Ω be a sample space, $\mathscr{F}$ be a σ-algebra of subsets of Ω.

Assume that $\mu(\cdot)$ is a set map defined on Ω satisfying the following conditions:

1. $\forall E \in \mathscr{F}$, $\mu(E) \geq 0$;

2. If $\{E_j\}_{j\geq 1}$ is a countable collection of disjoint sets in $\mathscr{F}$, then

$$\mu\Big(\bigcup_{j\geq 1} E_j\Big) = \sum_{j\geq 1} \mu(E_j);$$

3. $\mu(\Omega) = 1$.

Prove that

1. $\forall E \in \mathscr{F}$, $\mu(E) \leq 1$;

2. $\forall E \in \mathscr{F}$, $\mu(\emptyset) = 0$;

3. $\forall E \in \mathscr{F}$, $\mu(E) = 1 - \mu(E^c)$;

4. $\forall E, F \in \mathscr{F}$, $\mu(E \bigcup F) + \mu(E \bigcap F) = \mu(E) + \mu(F)$;

5. $\forall E, F \in \mathscr{F}$ such that $E \subseteq F$, $\mu(E) = \mu(F) - \mu(F \setminus E) \leq \mu(F)$;

6. Let $\{E_j\}_{j\geq 1}$ be an increasing (decreasing) sequence of sets in $\mathscr{F}$ that is $E_j \subseteq E_{j+1}$ ($E_j \supseteq E_{j+1}$) for all $j \geq 1$. Prove that, if $\{E_j\}_{j\geq 1}$ is an increasing (decreasing) sequence of sets in $\mathscr{F}$ such that $E_j \uparrow E$ ($E_j \downarrow E$), that is $E = \bigcup_{j\geq 1} E_j$ ($E = \bigcap_{j\geq 1} E_j$), then $\lim_{j\to+\infty} \mu(E_j) = \mu(E)$;

7. (Boole's inequality): $\mu\Big(\bigcup_{j\geq 1} E_j\Big) \leq \sum_{j\geq 1} \mu(E_j)$.

Exercise 6:

Let $\{E_j\}_{j\geq 1}$ be random events belonging to $\mathscr{F}$, a σ-field of events of a sample space Ω.

Let $\mu(\cdot)$ be a probability measure defined on $\mathscr{F}$. Show that for all $n \geq 1$

1. $\mu\Big(\bigcap_{j=1}^{n} E_j\Big) \geq 1 - \sum_{j=1}^{n} \mu(E_j^c)$;

2. If $\mu(E_j) \geq 1 - \varepsilon$, for $j \in \{1, \cdots, n\}$, then $\mu\Big(\bigcap_{j=1}^{n} E_j\Big) \geq 1 - n\varepsilon$;

3. $\mu\Big(\bigcap_{j\geq 1} E_j\Big) \geq 1 - \sum_{j\geq 1} \mu(E_j^c)$;

Exercise 7:

Prove the following properties:

1. If $\mu(E_j) = 0$ for all $j \geq 1$, then $\mu\Big(\bigcup_{j\geq 1} E_j\Big) = 0$;
2. If $\mu(E_j) = 1$ for all $j \geq 1$, then $\mu\Big(\bigcap_{j\geq 1} E_j\Big) = 1$;

Exercise 8:

Take $\{E_j\}_{j\geq 1}$ and $\{F_j\}_{j\geq 1}$ belonging to the same probability space $(\Omega, \mathscr{F}, \mu)$. Suppose that $\lim_{j\to+\infty} \mu(E_j) = 1$ and $\lim_{j\to+\infty} \mu(F_j) = p$, with $p \in [0,1]$. Show that $\lim_{j\to+\infty} \mu(E_j \bigcap F_j) = p$.

Exercise 9:

Let

$$\limsup_n E_n = \bigcap_{n\geq 1} \bigcup_{k\geq n} E_k, \tag{1.9.1}$$

$$\liminf_n E_n = \bigcup_{n\geq 1} \bigcap_{k\geq n} E_k. \tag{1.9.2}$$

If (1.9.2) and (1.9.1) are equal we write

$$\lim_n E_n = \liminf_n E_n = \limsup_n E_n.$$

Let $\{E_n\}_{n\geq 1}$ belong to a probability space $(\Omega, \mathscr{F}, \mu)$. Show that

1. $$\mu\Big(\liminf_n E_n\Big) \leq \liminf_n \mu(E_n) \leq \limsup_n \mu(E_n) \leq \mu\Big(\limsup_n E_n\Big).$$
2. If $\lim_{n\to+\infty} E_n = E$, then $\lim_{n\to+\infty} \mu(E_n) = \mu(E)$.

1.9.2 Exercises on Random variables and distribution functions

Exercise 1:

Specify the distribution function and the distribution measure of the random variable X.

(a) If X has probability function defined on $k \in \{0,1\}$ and given by

$$\mathbb{P}(X = k) = p^k(1-p)^{1-k}.$$

That is X has Bernoulli distribution of parameter p.

(b) If X has probability function defined in $k \in \{0, \cdots, n\}$ and given by

$$\mathbb{P}(X = k) = C_k^n p^k(1-p)^{n-k}.$$

That is X has Binomial distribution of parameter n and p.

(c) If X has probability function defined in $k \in \{0, 1, \cdots\}$ and given by

$$\mathbb{P}(X = k) = \frac{e^{-\alpha}\alpha^k}{k!},$$

$\alpha > 0$. That is X has Poisson distribution of parameter α.

(d) If X has probability function defined in $k \in \{0, 1, \cdots\}$ and given by

$$\mathbb{P}(X = k) = p(1-p)^k.$$

That is X has Geometric distribution of parameter p.

(e) If X has probability density function given by

$$f(x) = \alpha e^{-\alpha x}\mathbf{1}_{[0,+\infty)}(x),$$

with $\alpha > 0$. That is X has Exponential distribution with parameter α.

(f) If X has probability density function given by

$$f(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x)$$

for $a, b \in \mathbb{R}$ with $a < b$. That is X has Uniform distribution in $[a, b]$.

(g) If X has probability density function given by

$$f(x) = \frac{1}{\pi(1+x^2)},$$

$x \in \mathbb{R}$. That is X has Cauchy distribution.

(h) If X has probability density function given by

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2},$$

$x \in \mathbb{R}$. That is X has Gaussian distribution.

Exercise 2:

Let $\sigma > 0$. Let X be a random variable with probability density function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$.

(a) Prove that $f(\cdot)$ is indeed a probability density function. How does the graph of f look like when σ is very small?

(b) Compute $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$.

Exercise 3:

Let X be a random variable with probability density function given by $f(x) = cx^2\mathbf{1}_{[-1,1]}(x)$.

(a) Determine the value of the constant c.

(b) Exhibit the distribution function $F_X(\cdot)$ and find x_1 such that $F_X(x_1) = 1/4$.

Exercise 4:

Let X be a random variable with distribution function given by $F_X(x) = x^3\mathbf{1}_{[0,1]}(x) + \mathbf{1}_{(1,\infty]}(x)$.

(a) Find the probability density function of X.

(b) Prove that it is indeed a probability density function.

Exercise 5:

A random variable X is said to be symmetric around μ if $\mathbb{P}(X \geq \mu + x) = \mathbb{P}(X \leq \mu - x)$ for all $x \in \mathbb{R}$. If $\mu = 0$ we simply say that X is symmetric.

Let X be a random variable symmetric around the point $b \in \mathbb{R}$ and suppose that X takes the values a, b and $2b - a$, with $a < 0$ and $b > 0$.

(a) Show that $\mathbb{E}[X] = b$.

(b) Suppose that $\mathbb{E}[X] = 1$, $a = -1$, $Var(X) = 3$ and determine the distribution function of X and its induced measure μ_X.

(c) Compute $\mu_X((-\infty, -1])$, $\mu_X((-\infty, 3/2])$ and $\mu_X(\{1\})$.

Exercise 6:

Let X be a symmetric random variable that takes the values $a \neq b \neq c$.
Suppose that $\mathbb{P}(X = 0) = 1/5$.
Give the results in terms of $a \neq 0$.

(a) Exhibit the distribution function and the distribution measure of X.

(b) Compute $\mathbb{E}[X]$ and $Var(X)$.

Exercise 7:

Let X be a random variable with probability density function $f_X(\cdot)$ and for $b > 0$ and $c \in \mathbb{R}$ let $Y = bX + c$.

(a) Prove that the probability density function of Y is given by $f_Y(y) = \frac{1}{b} f_X\left(\frac{y-c}{b}\right)$.

(b) Let X be a random variable with Cauchy distribution.
Compute the probability density function of $Y = bX + M$, where $b > 0$ and $M \in \mathbb{R}$.

(c) Let X be a random variable with standard Normal distribution.
Compute the probability density function of $Y = \sigma X + \mu$, where $\sigma > 0$ and $\mu \in \mathbb{R}$.

(d) Let X be a random variable with Gamma distribution with parameter α and 1.
Compute the probability density function of $Y = \frac{X}{\beta}$.
What is the distribution of Y when $\alpha = 1$?

Exercise 8:

Let X be a random variable with density function given by $f(x) = (1 + x)^{-2}\mathbf{1}_{(0,+\infty)}(x)$.
Let $Y = \max(X, c)$, where c is a positive constant $c > 0$.

(a) Show that $f(\cdot)$ is a probability density function.

(b) Exhibit the distribution function of X and Y. Justify that F_X is in fact a distribution function.

(c) Decompose $F_Y(\cdot)$ in its discrete, absolutely continuous and singular parts.

(d) Compute $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

Exercise 9:

Let X be a random variable uniformly distributed on the interval $[0,1]$.
Let Y be the random variable defined as $Y = min(1/2, X)$.

(a) Determine the distribution function of X and Y and represent their graph.

(b) Decompose $F_Y(\cdot)$ in its discrete, absolutely continuous and singular parts.

(c) Compute $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

Exercise 10:

Let X be a random variable with exponential distribution with parameter $\lambda > 0$. Let $Y = \max(X, \lambda)$.

(a) Determine the distribution function of X and Y and represent their graph.

(b) Decompose $F_Y(\cdot)$ in its discrete, absolutely continuous and singular parts.

Exercise 11:

Let X be a random variable uniformly distributed on $[0,2]$.
Let Y be the random variable defined by $Y = min(1, X)$.

(a) Determine the distribution functions of X and Y and represent their graph.

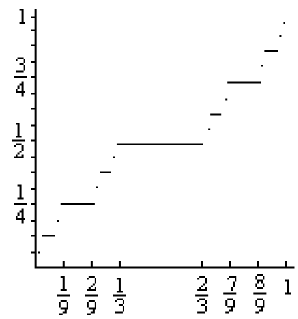
(b) Decompose $F_Y(\cdot)$ in its discrete, absolutely continuous and singular parts.

Exercise 12:

Let X be a random variable with Cantor distribution:

(a) Describe the construction of its distribution function $F_X(\cdot)$.

(b) Justify that X is a singular random variable.



(c) Compute $\mathbb{P}\left(X = \frac{1}{3}\right)$. Justify.

(d) Compute $\mathbb{P}\left(\frac{1}{3} < X < \frac{2}{3}\right)$, $\mathbb{P}(X \le \frac{2}{3})$ and $\mathbb{P}(\frac{1}{9} < X \le \frac{8}{9})$.

(e) Compute $\mathbb{E}[X]$. Justify.

Exercise 13: Let U be a random variable uniformly distributed in $[0,1]$.

(a) Find a function $f : [0,1] \to \mathbb{R}$ such that $f(U)$ is a random variable uniform in $[0,2]$.

(b) Find a function $f : [0,1] \to \mathbb{R}$ such that $f(U)$ is a random variable with Bernoulli distribution of parameter p, where $p \in (0,1)$.

(c) Find a function $f : [0,1] \to \mathbb{R}$ such that $f(U)$ is a random variable with exponential distribution of parameter $\lambda > 0$.

(d) Let $0 < p < q < 1$. Construct a random vector (X,Y) such that X has distribution Bernoulli with parameter p, Y has distribution Bernoulli with parameter q and $X \le Y$ almost surely.

(e) Let $0 < \lambda_1 < \lambda_2$. Construct a random vector (X,Y) such that X has exponential distribution with parameter λ_1, Y has exponential distribution with parameter λ_2 and $X \ge Y$ almost surely.

1.9.3 Exercises on Random vectores and Stochastic Independence.

Exercise 1:

Select a point uniformly in the unitary circle $\mathscr{C} = \{(x, y) : x^2 + y^2 \leq 1\}$. Let X and Y be the coordinates of the selected point.

(a) Determine the joint density of X and Y.

(b) Determine $\mathbb{P}(X < Y)$, $\mathbb{P}(X > Y)$ and $\mathbb{P}(X = Y)$.

(c) What is probability of finding the point in the first quadrant? Justify.

Exercise 2:

Suppose that X and Y are random variables identically distributed with symmetric distribution around zero and with joint distribution given by

$X \setminus Y$	-1	0	...
-1	...	0	...
0	0	...	0
...	θ	0	θ

(a) If $\mathbb{P}(X = -1) = 2/5$, complete the table.

(b) Compute $\mathbb{E}[X]$, $\mathbb{E}[Y]$ and $Var(X)$.

(c) Are the random variables X and Y independent? Justify.

(d) Find the probability functions of the random variables $X + Y$ and XY. Justify if $X + Y$ and XY are symmetric random variables around zero.

(e) Represent the graph of the distribution function of the random variable $X + Y$.

(f) Explicit the measure μ_{X+Y}.

(g) Compute $\mu_{X+Y}(\{0\})$ and $\mu_{X+Y}((-\infty, 0])$.

Exercise 3:

Suppose that X and Y are random variables with joint distribution given by:

$X \setminus Y$	1	2	3
1	0	1/5	0
2	1/5	1/5	1/5
3	0	1/5	0

(a) Compute the marginal probability functions of X and Y.

(b) Compute $\mathbb{E}[X]$, $\mathbb{E}[Y]$ and $Var(X)$.

(c) Are the random variables X and Y independent? Justify.

(d) If Z and W are independent random variables, then $\mathbb{E}[ZW] = \mathbb{E}[Z]\mathbb{E}[W]$.

Is the opposite true? Prove or exhibit a counter example.

(e) Find the distribution function of X and represent its graph.

(f) Exhibit the distribution measure μ_X of X.

(g) Compute the distribution function of $X + Y$.

(h) Compute the distribution function of $X - Y$.

Exercise 4:

Suppose that X and Y are random variables with joint distribution given by:

$X \setminus Y$	1	0	-1
1	0	a	0
0	b	c	b
-1	0	a	0

where $a, b, c > 0$.

(a) Compute the marginal probability functions of X and Y.

Justify that $2a + 2b + c = 1$.

(b) Compute $\mathbb{E}[X]$, $\mathbb{E}[Y]$ and $Var(X)$.

(c) Verify that the random variable XY is such that $XY = 0$ almost surely.

(d) Are the random variables X and Y independent? Justify.

(e) If Z and W are independent random variables, then $\mathbb{E}[ZW] = \mathbb{E}[Z]\mathbb{E}[W]$.

Is the opposite true? Prove or exhibit a counter example.

(f) Take $c = 1/4$ and a, b such that $a = 2b$.

(f_1) Find the distribution function of X and represent its graph.

(f_2) Exhibit the distribution measure μ_X of X.

Exercise 5:

Let X be a random variable such that $X \sim \mathcal{U}[0,1]$. Compute the distribution of $Y = -\log(X)$.

Exercise 6:

Let X and Y be i.i.d. random variables with $X \sim \mathcal{U}[0,1]$. Compute the distribution of $Z = X/Y$.

Exercise 7:

Let X and Y have joint density given by $f(x,y)$. Show that

$$f_{X+Y}(u) = \int_{\mathbb{R}} f(u-t,t)dt.$$

Moreover, if X and Y are independent with densities f_X and f_Y, respectively, then

$$f_{X+Y}(u) = \int_{\mathbb{R}} f_X(t) f_Y(u-t)dt.$$

Exercise 8:

Let X be a r.v. with density $f(x) = \frac{1}{4}e^{-|x|/2}$, for $x \in \mathbb{R}$. Compute the distribution of $Y = |X|$.

Exercise 9:

Show that the function

$$F(x,y) = \begin{cases} 1 - e^{-(x+y)}, & x \geq 0 \quad \text{and} \quad y \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

is not the distribution function of a random vector.

Exercise 10:

Show that the function

$$F(x,y) = \begin{cases} (1-e^{-x})(1-e^{-y}), & x \geq 0 \quad \text{and} \quad y \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

is the distribution function of a random vector.

Exercise 11:

Let X and Y be i.i.d. random variables with uniform distribution on $[\theta - 1/2, \theta + 1/2]$, with $\theta \in \mathbb{R}$. Compute the distribution of $X - Y$.

Exercise 12:

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with Rayleigh distribution with parameter θ, that is, the density of X_1 is given by

$$f(x) = \begin{cases} \dfrac{x}{\theta^2} e^{-\dfrac{x^2}{2\theta^2}}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

(a) Compute the joint density of $Y_1, \ldots, Y_n$, where for each $i = 1, \ldots, n$ it holds that $Y_i = X_i^2$.

(b) Compute the distribution of $U = \min_{1 \le i \le n} X_i$.

(c) Compute the distribution of $Z = X_1 / X_2$.

Exercise 13:

Let $X_1, X_2, \ldots, X_n$ be independent random variables with exponential distribution with parameter $\alpha_1, \ldots, \alpha_n$, respectively.

(a) Compute the distribution of $Y = \min_{1 \le i \le n} X_i$ and $Z = \max_{1 \le i \le n} X_i$.

(b) Show that for each $p = 1, \ldots, n$ it holds that

$$\mathbb{P}(X_p = \min_{1 \le i \le n} X_i) = \frac{\alpha_p}{\alpha_1 + \cdots + \alpha_n}.$$

(Hint: Consider the event $\{X_p < \min_{i \ne p} X_i\}$).

Exercise 14:

Let $X_1, X_2, \ldots, X_n$ be independent random variables with distribution functions $F_1, F_2, \cdots, F_n$ respectively. Find the distribution functions of the random variables $\min_{1 \le i \le n} X_i$ and $\max_{1 \le i \le n} X_i$.

Exercise 15:

Let X and Y be independent random variables each assuming the values 1 and -1 with probability $1/2$. Show that $\{X, Y, XY\}$ are pairwise independent but not totally independent.

1.9.4 Exercises on Mathematical Expectation

Exercise 1:

In each case, compute $\mathbb{E}(X)$ and $Var(X)$, if they exist:

(a) If X has probability function defined on $k \in \{0,1\}$ and given by $\mathbb{P}(X = k) = p^k(1-p)^{1-k}$.
That is X has Bernoulli distribution of parameter p.

(b) If X has probability function defined in $k \in \{0,\cdots,n\}$ and given by $\mathbb{P}(X = k) = C_k^n p^k(1-p)^{n-k}$.
That is X has Binomial distribution of parameter n and p.

(c) If X has probability function defined in $k \in \{0,1,\cdots\}$ and given by $\mathbb{P}(X = k) = \frac{e^{-\alpha}\alpha^k}{k!}$, $\alpha > 0$.
That is X has Poisson distribution of parameter α.

(d) If X has probability function defined in $k \in \{0,1,\cdots\}$ and given by $\mathbb{P}(X = k) = p(1-p)^k$.
That is X has Geometric distribution of parameter p.

(e) If X has probability density function given by $f(x) = \alpha e^{-\alpha x}\mathbf{1}_{[0,+\infty)}(x)$, with $\alpha > 0$.
That is X has Exponential distribution with parameter α.

(f) If X has probability density function given by $f(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x)$ for $a, b \in \mathbb{R}$ with $a < b$.
That is X has Uniform distribution in $[a, b]$.

(g) If X has probability density function given by $f(x) = \frac{1}{\pi(1+x^2)}$, $x \in \mathbb{R}$.
That is X has Cauchy distribution.

(h) If X has probability density function given by $f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, $x \in$

$\mathbb{R}$.
That is X has Normal distribution.

Exercise 2:

Prove that:

(a) For any random variable X with distribution function F_X, it holds that

$$E[X] = \int_0^{+\infty} 1 - F_X(x)dx - \int_{-\infty}^0 F_X(x)dx$$

(b) and for any $k \in \mathbb{N}$

$$E[X^k] = k\int_0^{+\infty} (1 - F_X(x))x^{k-1}dx - k\int_{-\infty}^0 F_X(x)x^{k-1}dx.$$

(c) If X is non-negative, then

$$E[X] = \int_0^{+\infty} 1 - F_X(x)dx.$$

(d) If X is discrete and takes non-negative integer values, then

$$E[X] = \sum_{n=1}^{+\infty} P(X \geq n).$$

(e) If X has Exponential distribution with parameter $\lambda > 0$, then $E[X^k] = k!/\lambda^k$, for any $k \in \mathbb{N}$.

(f) Let X and Y be random variables, such that Y is stochastically dominated by X, that is for all $x \in \mathbb{R}$ it holds that $F_X(x) \leq F_Y(x)$. Show that $E[X] \geq E[Y]$, if both expectations exist.

Exercise 3:

Show that:

(a) if X is a constant random variable, then $Var(X) = 0$.

(b) if $a \in \mathbb{R}$ then $Var(X + a) = Var(X)$.

(c) if $a, b \in \mathbb{R}$ then $Var(aX+b) = a^2 Var(X)$.

Exercise 4:

Prove:

(a) Basic Tchebychev's inequality:
If X is a non-negative random variable (that is $X \geq 0$), then for all $\lambda > 0$:

$$P(X \geq \lambda) \leq \frac{1}{\lambda} E(X).$$

(b) Classical Tchebychev's inequality:
If X is an integrable random variable, then for all $\lambda > 0$:

$$P(|X - E(X)| \geq \lambda) \leq \frac{1}{\lambda^2} Var(X).$$

(b) Markov's inequality:
If X is a random variable, then for all $t > 0$ and $\lambda > 0$:

$$P(|X| \geq \lambda) \leq \frac{1}{\lambda^t} E(|X|^t).$$

Exercise 5:

(a) Let X be a non-negative random variable, that is $X \geq 0$, such that $E(X) = 0$.

Show that $P(X = 0) = 1$, that is, $X = 0$ almost surely.

(b) Let X be a random variable independent of itself.

Show that X is constant with probability 1 (that is, there exists a constant c such that $P(X = c) = 1$).

Exercise 6:

Let $X_1, \cdots, X_n$ be integrable random variables, such that for $i \neq j$,

$$Cov(X_i, X_j) := E[X_i X_j] - E[X_i]E[X_j] = 0.$$

Show that

$$Var(X_1 + \cdots + X_n) = Var(X_1) + \cdots + Var(X_n).$$

Exercise 7:

Let $X_1, \cdots, X_n$ be independent random variables with distribution function $F_{X_1}, \cdots, F_{X_n}$, respectively.

(a) Find the distribution function of $\max_{1\le j\le n} X_j$ and $\min_{1\le j\le n} X_j$.

(b) Suppose that the random variables are identically distributed with finite mean. Show that

$$\lim_{n\to+\infty} \frac{1}{n} E\Big[\max_{1\le j\le n} |X_j|\Big] = 0.$$

Exercise 8:

Let X and Y be random variables defined on a probability space $(\Omega, \mathscr{F}, P)$, both with finite expectation. Show that

(a) $E[X+Y] = E[X] + E[Y]$.

(b) if X and Y are independent, then $E[XY] = E[X]E[Y]$.

Exercise 9:

Let (X, Y) be a random vector with density function given by

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\Big\{\frac{-1}{2(1-\rho^2)}\Big[\Big(\frac{x-\mu_1}{\sigma_1}\Big)^2 - 2\rho\Big(\frac{x-\mu_1}{\sigma_1}\Big)\Big(\frac{y-\mu_2}{\sigma_2}\Big) + \Big(\frac{y-\mu_2}{\sigma_2}\Big)^2\Big]\Big\}.$$

(a) Find the marginal distributions of X and Y.

(b) Assume that X and Y are independent. Compute the distribution of $X+Y$.

(c) Show that X and Y are independent if and only if $\rho = 0$.

Exercise 10:

Let X and Y be random variables taking only the values 0 and 1. Show that, if $E[XY] = E[X]E[Y]$ then X and Y are independent.

Exercise 11:

Let X and Y be random variables with finite variance. Show that, if $Var(X) \neq Var(Y)$ then $X+Y$ and $X-Y$ are not independent.

Exercise 12:

Let X and Y be i.i.d. random variables with Uniform distribution in $[0,1]$. Compute the expectation of $\min(X,Y)$ and $\max(X,Y)$.

Exercise 13:

Prove Wald's equation, that is, show that $E[S_t] = E[N_t]E[X_1]$, where $S(t)$ is a compound stochastic process, or else, $S(t) := \sum_{i=1}^{N_t} X_i$, where N_t is a counting process (i.e. N_t takes values in $\mathbb{N}$) and $\{X_i\}_{i\geq 1}$ is a sequence of i.i.d. random variables and independent of N_t for all t.

Exercise 14:

Let X be a random variable and $F_X(\cdot)$ its distribution function. Prove that, for any $a \geq 0$, we have

$$\int_{\mathbb{R}} \Big(F_X(x+a) - F_X(x)\Big) dx = a.$$

Exercise 15:

Show that if $Cov(X,Y) = \sqrt{Var(X)}\sqrt{Var(Y)}$, then there exist constants a and b such that

$$P(Y = aX + b) = 1.$$

TÉCNICO
LISBOA

Chapter 2

Convergence of sequences of r.v.

2.1 Convergence a.e., $\mathbb{L}^p$ and in probability

Recall that we have seen that if $\{X_n\}_{n\geq 1}$ is a sequence of r.v. then $\lim_{n\to+\infty} X_n$ is a r.v. The notion of convergence we use is of convergence to a finite limit: if we say $\{X_n\}_{n\in\mathbb{N}}$ converges in $\Lambda \in \mathscr{F}$, this means that for all $\omega \in \Lambda$ we have that the sequence $\{X_n(\omega)\}_{n\in\mathbb{N}}$ converges. When $\Lambda = \Omega$ we say the convergence holds everywhere.

Definition 2.1.1 (Almost everywhere convergence). *The sequence $\{X_n\}_{n\in\mathbb{N}}$ is said to converge almost everywhere to X iff there exists a null set N such that*

$$\forall \omega \in \Omega \setminus N : \lim_{n\to\infty} X_n(\omega) = X(\omega) \quad \textit{finite}.$$

Theorem 2.1.2. *A sequence of r.v. $\{X_n\}_{n\in\mathbb{N}}$ converges a.e. to X iff for every $\epsilon > 0$ we have that*

$$\lim_{m\to\infty} \mathbb{P}(|X_n - X| \leq \epsilon \quad \textit{for all} \quad n \geq m) = 1$$

or equivalently

$$\lim_{m\to\infty} \mathbb{P}(|X_n - X| > \epsilon \quad \textit{for some} \quad n \geq m) = 0 \tag{2.1.1}$$

Proof. Let us suppose that there is convergence a.e. and let $\Omega_0 = \Omega \setminus N$ where N is the set where the convergence does not hold. For $\varepsilon > 0$ and $m \geq 1$, let $A_m(\varepsilon)$ be the set inside the first probability above:

$$A_m(\varepsilon) = \cup_{n=m}^{+\infty} \{|X_n - X| \leq \varepsilon\}. \tag{2.1.2}$$

Then, $A_m(\varepsilon) \subset A_{m+1}(\varepsilon)$, so that $A_m(\epsilon) \uparrow$. Fix $\omega_0 \in \Omega_0$ and note that the convergence of $X(\omega_0)$ to $X(\omega_0)$ implies that given $\varepsilon > 0$, there exists an order $m(\omega_0, \varepsilon)$ such that for any $n \geq m(\omega_0, \varepsilon)$ it holds that $|X_n(\omega_0) - X(\omega_0)| \leq \varepsilon$. Then, ω_0 belongs to some $A_m(\varepsilon)$. Since this property holds for any ω_0 we have that

$$\Omega_0 \subset \cup_{m=1}^{+\infty} A_m(\varepsilon).$$

By the monotone property of the measure $\mathbb{P}$ it holds that

$$\lim_{m\to+\infty} \mathbb{P}(A_m(\varepsilon)) \geq \mathbb{P}(\Omega_0) = 1.$$

Since $\mathbb{P}$ is a probability measure it follows that

$$\lim_{m\to+\infty} \mathbb{P}(A_m(\varepsilon)) = 1.$$

This proves the first result. Reciprocally, suppose that

$$1 = \lim_{m\to\infty} \mathbb{P}(A_m(\varepsilon)) = \mathbb{P}(\cup_{m=1}^{+\infty} A_m(\varepsilon)) = \mathbb{P}(A(\varepsilon)).$$

For $\varepsilon > 0$ and $\omega_0 \in A(\varepsilon)$ we have that there exists an order m such that for all $n \geq m$ it holds that $|X_n(\omega_0) - X(\omega_0)| \leq \varepsilon$. Let $\varepsilon = 1/n$ and let $A = \cap_{n=1}^{+\infty} A(1/n)$ and note that $\mathbb{P}(A) = \mathbb{P}(\cap_{n=1}^{+\infty} A(1/n)) = 1$ for all $n \geq 1$. If $\omega_0 \in A$ then the property: there exists an order m such that for all $n \geq m$ it holds that $|X_n(\omega_0) - X(\omega_0)| \leq \varepsilon$ holds for any $\varepsilon = 1/n$. Now we prove that in fact it holds for any $\varepsilon > 0$. Fix $\varepsilon > 0$ not necessarily of the form $1/n$. Then take n such that $1/n < \varepsilon$. Then since there exists an order m such that for all $n \geq m$ it holds that $|X_n(\omega_0) - X(\omega_0)| \leq \frac{1}{n} < \varepsilon$. Since the property holds for any $\varepsilon > 0$ and for all $\omega_0 \in A$ with $\mathbb{P}(A) = 1$, it follows the a.e. convergence. □

Definition 2.1.3 (Convergence in probability)**.** *The sequence $\{X_n\}_{n\in\mathbb{N}}$ is said to converge in probability to X, iff for every $\epsilon > 0$ it holds that*

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0.$$

Theorem 2.1.4 (Convergence a.e. implies convergence in probability)**.** *Convergence a.e. to X implies convergence in probability to X.*

Proof. Note that (2.1.1) implies the previous limit. To see that note that (2.1.1) means that for all $\varepsilon > 0$

$$\lim_{m\to+\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

But

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) \leq \mathbb{P}(\cup_{n\geq m}|X_n - X| > \varepsilon) = 0$$

so that the proof ends. □

Definition 2.1.5 (Convergence in $\mathbb{L}^p$, $0 < p < \infty$)**.** *The sequence $\{X_n\}_{n\in\mathbb{N}}$ is said to converge in $\mathbb{L}^p$ to X, iff $X_n \in \mathbb{L}^p$, $X \in \mathbb{L}^p$ and*

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0. \tag{2.1.3}$$

Definition 2.1.6. *We say that X is dominated by Y if $|X| \leq Y$ a.e. and that the sequence is dominated by Y, if this is true for any n with the same Y. Moreover, if above Y is constant we say that X or X_n is uniformly bounded.*

Above we can suppose $X = 0$ since the definitions hold for $X_n - X$.

Theorem 2.1.7 (Convergence in $\mathbb{L}^p$ implies convergence in probability)**.** *Convergence in $\mathbb{L}^p$ implies convergence in probability. The converse is true if the sequence is dominated by some $Y \in \mathbb{L}^p$.*

Note that in the two previous sentences we can take $X = 0$ since $X_n \to X$ in $\mathbb{L}^p$ is such that $\{X_n\}_{n\geq 1}$ is dominated by Y, then $\{X_n - X\}_{n\geq 1}$ is dominated by $|X_n| + |X| \leq Y + |X| \in \mathbb{L}^p$.

Proof. To prove the first affirmation we use the general Tchebychev's inequality with $\varphi(u) = |u|^p$. Then,

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}[|X|^p]}{\varepsilon^p} \to_{n\to+\infty} 0.$$

Now, suppose that $|X_n| \leq Y$ a.e. and that $\mathbb{E}[Y^p] < +\infty$. Then,

$$\begin{aligned} \mathbb{E}[|X_n|^p] &= \int_{\{|X_n|\leq\varepsilon\}} |X_n|^p d\mathbb{P} + \int_{\{|X_n|>\varepsilon\}} |X_n|^p d\mathbb{P} \\ &\leq \varepsilon^p + \int_{\{|X_n|>\varepsilon\}} Y^p d\mathbb{P}. \end{aligned} \tag{2.1.4}$$

Now, since $\mathbb{P}(|X_n| > \varepsilon) \to 0$ as $n \to +\infty$ and since Y^p is integrable, we have, from a result from measure theory that the term at the right hand side of last expression vanishes as $n \to +\infty$. Then we take $\varepsilon \to 0$ and the result follows.

□

Remark 2.1.8.

- *Convergence in probability does not imply convergence in $\mathbb{L}^p$ and convergence in $\mathbb{L}^p$ does not imply convergence a.e.*
- *Convergence a.e. does not imply convergence in $\mathbb{L}^p$.*

Theorem 2.1.9 (Scheffé's Theorem)**.** *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of r.v. with densities $f_1, f_2, \cdots$ and let X be a r.v. with density f. If $\lim_n f_n = f$, holds a.e. then $\lim_{n\to\infty} \int_{\mathbb{R}} |f_n - f| dx = 0$.*

Exercise: do the proof of the previous theorem.

Definition 2.1.10 ($\limsup_n$ and $\liminf_n$)**.** *Let $\{E_n\}_{n\in\mathbb{N}}$ be a sequence of subsets of Ω. The $\limsup_n E_n$ and the $\liminf_n E_n$ are defined by*

$$\limsup_n E_n = \cap_{m=1}^{\infty} \cup_{n=m}^{\infty} E_n \quad \textit{and} \quad \liminf_n E_n = \cup_{m=1}^{\infty} \cap_{n=m}^{\infty} E_n \tag{2.1.5}$$

Remark 2.1.11.

- *Note that a point belongs to $\limsup_n E_n$ iff it belongs to infinitely many terms of the sequence $\{E_n\}_{n\in\mathbb{N}}$ and belongs to $\liminf_n E_n$ iff it belongs to all the terms of the sequence from a certain point on. To see this note that a point belongs to an infinite number of sets E_n iff that point does not belong to all the E_n^c from a certain order on. Then, the second affirmation is a consequence of the first. Let ω be a point that belongs to infinitely many E_n, then ω belongs to $F_m = \cup_{n=m}^{+\infty} E_n$ for all $m \geq 1$ and so ω belongs to $\cap_{m=1}^{+\infty} F_m = \limsup_n E_n$. Reciprocally, let $\omega \in \limsup_n E_n$ i.e. suppose that $\omega \in \cap_{m=1}^{+\infty} F_m$. Then $\omega \in F_m$ for all $m \geq 1$. If ω belongs only to a finite* *number of* E_n*'s then there would exist an* $m \geq 1$ *such that* $\omega \notin E_n$ *for* $n \geq m$ *and we would have that* $\omega \notin \cup_{n=m}^{+\infty} E_n = F_m$ *which is absurd. Therefore,* ω *belongs to infinitely many* E_n*'s.*

- *Also note that* $(\limsup_n E_n^c)^c = \liminf_n E_n$.
- *The event* $\limsup_n E_n$ *occurs iff the events* E_n *occur i.o.*
- *If each* $E_n \in \mathscr{F}$*, then* $\mathbb{P}(\limsup_n E_n) = \lim_{m\to\infty} \mathbb{P}(\cup_{n=m}^{\infty} E_n)$*. To prove this result note that if* $F_m = \cup_{n=m}^{+\infty} E_n$ *then* $F_m \downarrow$*. Therefore,*

$$\mathbb{P}(\cap_{m=1}^{+\infty} F_m) = \lim_{m\to+\infty} \mathbb{P}(F_m).$$

As an exercise show that $\mathbb{P}(\liminf_n E_n) = \lim_{m\to\infty} \mathbb{P}(\cap_{n=m}^{\infty} E_n)$.

Lemma 2.1.12 (Borel-Cantelli - the convergent part)**.** *For* $\{E_n\}_{n\in\mathbb{N}}$ *arbitrary events, if* $\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty$*, then* $\mathbb{P}(E_n \ i.o.) = 0$.

Proof. Let $F_m = \cup_{n=m}^{+\infty} E_n$. From Boole's inequality we have that

$$\mathbb{P}(F_m) \leq \sum_{n=m}^{+\infty} \mathbb{P}(E_n).$$

Now note that the hypothesis of the theorem implies that $\lim_{m\to+\infty} \mathbb{P}(F_m) = 0$, since the series $\sum_{n=1}^{\infty} \mathbb{P}(E_n)$ is converging. Then

$$0 \leq \mathbb{P}(\limsup_n E_n) \leq \lim_{m\to+\infty} \mathbb{P}(F_m) = 0,$$

from where the result follows. □

We can rephrase Theorem 2.1.2:

Theorem 2.1.13. *A sequence of r.v.* $\{X_n\}_{n\geq1}$ *converges a.e. to* 0 *iff* $\forall \epsilon > 0$ *we have that*

$$\mathbb{P}(\{|X_n| > \epsilon\} \ i.o.) = 0.$$

Proof. Let us denote $A_m = \cup_{n=m}^{+\infty} \{|X_n| \leq \varepsilon\}$. Then

$$\{\{|X_n| > \epsilon\} \ i.o.\} = \cap_{m=1}^{+\infty} \cup_{n=m}^{+\infty} \{|X_n| > \varepsilon\} = \cap_{m=1}^{+\infty} A_m^c.$$

From Theorem 2.1.2 we know that $X_n \to_{n\to+\infty} 0$ iff for all $\varepsilon > 0$ we have that $\mathbb{P}(A^c_m) \to_{n+\infty} 0$. Since $A^c_m \downarrow$, since $A^c_m = \cup_{n=m}^{+\infty}\{|X_n| > \varepsilon\}$ and last limit is equivalent to (2.1.1). □

Theorem 2.1.14. *If $\{X_n\}_{n\geq 1}$ converges in probability to X, then there exists a sequence $\{n_k\}$ of integers growing to ∞ such that $X_n \to X$ a.e. This means that convergence in probability implies converges a.e. along a subsequence.*

Proof. Let us take $X = 0$. Then, as a particular case of the convergence in probability, it follows that for all $k \in \mathbb{N}$ we have that

$$\lim_{n\to+\infty} \mathbb{P}(|X_n| > 1/2^k) = 0.$$

Then for all $k \in \mathbb{N}$ there exists n_k such that $\lim_{n\to+\infty} \mathbb{P}(|X_{n_k}| > \frac{1}{2^k}) < \frac{1}{2^k}$. So we have that

$$\sum_{k\in\mathbb{N}} \mathbb{P}\Big(|X_{n_k}| > \frac{1}{2^k}\Big) \leq \sum_{k\in\mathbb{N}} \frac{1}{2^k} < +\infty.$$

Now, having n_k fixed, let $\tau_k := \Big\{|X_{n_k}| > \frac{1}{2^k}\Big\}$ and from the Borel-Cantelli's Lemma (converging part) we have that $\mathbb{P}\Big(\Big\{|X_{n_k}| > \frac{1}{2^k}\Big\} i.o.\Big) = 0$ and this implies the a.e. convergence of X_{n_k} to 0 as $k \to +\infty$. To see this we do it as we have already done before, we fix $\varepsilon > 0$ and we choose k such that $\frac{1}{2^k} < \varepsilon$. □

If we add independence to the Borel-Cantelli's Lemma, then we have

Lemma 2.1.15 (Borel-Cantelli - the divergent part)**.** *For independent events $\{E_n\}_{n\in\mathbb{N}}$, if $\sum_{n=1}^{\infty} \mathbb{P}(E_n) = +\infty$, then $\mathbb{P}(E_n\ i.o.) = 1$.*

Proof. Note that

$$\mathbb{P}(\liminf_n E^c_n) = \mathbb{P}(\cap_{n=m}^{+\infty} E^c_n).$$

Moreover, since the events $\{E_n\}_{n\in\mathbb{N}}$ are independent, then the events $\{E^c_n\}_{n\in\mathbb{N}}$ are also independent. Therefore, if $m' > m$, then

$$\mathbb{P}\Big(\cap_{n=m}^{m'} E^c_n\Big) = \prod_{n=m}^{m'} \mathbb{P}(E^c_n) = \prod_{n=m}^{m'} (1 - \mathbb{P}(E_n)).$$

Now, we use the fact, that for all $x \geq 0$ it holds that $1-x \leq e^{-x}$. As a consequence last probability is bounded from above by

$$\prod_{n=m}^{m'} e^{-\mathbb{P}(E_n)} = e^{-\sum_{n=m}^{m'} \mathbb{P}(E_n)}$$

and taking the limit $m' \to +\infty$ we have that $e^{-\sum_{n=m}^{m'} \mathbb{P}(E_n)} \to 0$, since the series $\sum_{n=1}^{\infty} \mathbb{P}(E_n)$ is diverging. Now, from the monotonicity property of $\mathbb{P}$ we have that

$$\mathbb{P}\Big(\cap_{n=m}^{+\infty} E_n^c\Big) = \lim_{m' \to +\infty} \mathbb{P}\Big(\cap_{n=m}^{m'} E_n^c\Big) = 0.$$

Then $\mathbb{P}(\liminf_n E_n^c) = 0 \iff \mathbb{P}(\limsup_n E_n) = 1$. □

Remark 2.1.16. *Removing the independence assumption, the result is false. To see that, take $E_n = A$ for all $n \geq 1$ with $0 < \mathbb{P}(A) < 1$. Then $\sum_{n\geq 1} \mathbb{P}(A_n) = \sum_{n\geq 1} \mathbb{P}(A) = +\infty$. But the event $\{A_n \; i.o.\} = A$ and $P(A) < 1$.*

We observe that the previous result also holds with pairwise independence.

Lemma 2.1.17. *For events $\{E_n\}_{n\in\mathbb{N}}$ which are pairwise independent, if $\sum_{n=1}^{\infty} \mathbb{P}(E_n) = +\infty$, then $\mathbb{P}(E_n \; i.o.) = 1$.*

Proof. Let $I_n = \mathbf{1}_{E_n}$ and in that case the pairwise independence hypothesis can be written as

$$\mathbb{E}[I_m I_n] = \mathbb{E}[I_m]\mathbb{E}[I_n]$$

for $m \neq n$. Consider the series $\sum_{n\geq 1} I_n(\omega)$. This series diverges iff an infinite number of terms is equal to 1, that it if ω belongs to an infinite number of E_n's. Then, the conclusion of the theorem can be written as

$$\mathbb{P}\Big(\sum_{n\geq 1} I_n = +\infty\Big) = 1.$$

The other hypothesis can be written as $\sum_{n\geq 1} \mathbb{E}[I_n] = +\infty$. Consider the partial sum $J_k = \sum_{n=1}^{k} I_n$. From Tchebychev's inequality we have that

$$\mathbb{P}\Big(|J_k - \mathbb{E}[J_k]| > A\sigma(J_k)\Big) \leq \frac{\sigma^2(J_k)}{A^2\sigma^2(J_k)} = \frac{1}{A^2}.$$

From here it follows that

$$\mathbb{P}\Big(|J_k - \mathbb{E}[J_k]| \leq A\sigma(J_k)\Big) \geq 1 - \frac{1}{A^2}.$$

Above $\sigma^2(J_k)$ denotes the variance of J_k. Now, let $p_n = \mathbb{E}[I_n] = \mathbb{P}(E_n)$. Then

$$\begin{aligned}
\mathbb{E}[J_k^2] &= \mathbb{E}\Big[\sum_{n=1}^{k} I_n^2\Big] + 2\mathbb{E}\Big[\sum_{1\leq m<n\leq k} I_m I_n\Big] \\
&= \sum_{n=1}^{k} \mathbb{E}[I_n^2] + 2\sum_{1\leq m<n\leq k} \mathbb{E}[I_m]\mathbb{E}[I_n] \\
&= \sum_{n=1}^{k} \Big(\mathbb{E}[I_n]\Big)^2 + 2\sum_{1\leq m<n\leq k} \mathbb{E}[I_m]\mathbb{E}[I_n] + \sum_{k=1}^{n}(\mathbb{E}[I_n^2] - (\mathbb{E}[I_n])^2) \\
&= \Big(\sum_{n=1}^{k} p_n\Big)^2 + \sum_{n=1}^{k}(p_n - p_n^2).
\end{aligned} \tag{2.1.6}$$

Therefore, $\sigma^2(J_k) = \sum_{n=1}^{k} \sigma^2(I_n)$. Since

$$\sum_{n=1}^{k} p_n = \sum_{n=1}^{k} \mathbb{P}(E_n) = \sum_{n=1}^{k} \mathbb{E}[I_n] = \mathbb{E}[J_k] \to_{k\to+\infty} +\infty,$$

then $\sigma^2(J_k) = \sum_{n=1}^{k}(p_n - p_n^2) \leq \sum_{n=1}^{k} p_n$, so that $\sigma(J_k) \leq \Big(\mathbb{E}[J_k]\Big)^{1/2} = o(\mathbb{E}[J_k])$. Now, if $k > k_0(A)$ we have that $\frac{\sigma(J_k)}{\mathbb{E}[J_k] \leq \frac{1}{2A}}$. Then

$$1 - \frac{1}{A^2} \leq \mathbb{P}\Big(|J_k - \mathbb{E}[J_k]| \leq A\sigma(J_k)\Big) \leq \mathbb{P}\Big(J_k \geq -A\sigma(J_k) + \mathbb{E}[J_k]\Big) \tag{2.1.7}$$

and this implies that

$$1 - \frac{1}{A^2} \leq \mathbb{P}\Big(J_k \geq \frac{\mathbb{E}[J_k]}{2}\Big). \tag{2.1.8}$$

Now, observe that J_k increases with K. Since the inequality above holds for $k \geq K_0(A)$, then we can replace J_k by $\lim_k J_k$ and then we get that

$$\mathbb{P}\Big(\lim_{k\to+\infty} J_k = +\infty\Big) \geq 1 - \frac{1}{A^2}. \tag{2.1.9}$$

Since the constant A is arbitrary, we can take the limit as $A \to 0$ to conclude that

$$\mathbb{P}\Big(\lim_{k\to+\infty} J_k = +\infty\Big) = 1. \tag{2.1.10}$$

This concludes the proof. □

Putting together the previous results we have the following statement:

Corollary 2.1.18 (Zero-One law)**.** *For independent events* $\{E_n\}_{n\in\mathbb{N}}$*, then*

$$\mathbb{P}(E_n\ i.o.) = 0 \quad or \quad 1$$

if

$$\sum_{n=1}^{\infty}\mathbb{P}(E_n) < \infty \quad or \quad \sum_{n=1}^{\infty}\mathbb{P}(E_n) = \infty.$$

2.2 Weak convergence

If a sequence of r.v. $\{X_n\}_{n\geq 1}$ converges to some limit, does the sequence of probability distribution measures $\{\mu_n\}_{n\in\mathbb{N}}$ converges in some sense? Is is true that $\lim_n \mu_n(A)$ exists for any $A \in \mathscr{B}$? The answer to the questions above is no. Let us see an example. For each $n \in \mathbb{N}$, take $X_n = c_n$, where c_n is a constant such that $\lim_{n\to+\infty} c_n = 0$. Then $X_{n\to+\infty}X_n = 0$ deterministically. Let μ_n be the measure induced by X_n and let μ be the measure induced by the limit r.v. which is equal to 0. Let I be an interval of $\mathbb{R}$ such that $0 \notin \bar{I}$ where $\bar{I}$ is the closure of the set I. Then $\lim_{n\to+\infty}\mu_n(I) = 0 = \mu(I)$. On the other hand if I is an interval such that $0 \in I^0$, where I^0 is the interior of the set I, then $\lim_{n\to+\infty}\mu_n(I) = 1 = \mu(I)$. From this we see that if c_n oscillates between strictly positive and negative numbers and if $I = (a, 0)$ or $I = (0, b)$, then $\mu_n(I)$ oscillates between 0 and 1 and $\mu(I) = 0$. Nevertheless, if $I = (a, 0]$ or $I = [0, b)$, then $\mu_n(I)$ oscillates between 0 and 1 but $\mu(I) = 1$. Note that $\mu = \delta_{\{0\}}$.

Another tricky example is to take X_n with uniform distribution in (c_n, c_n') with $c_n < 0 < c_n'$ and both sequences $\{c_n\}_{n\in\mathbb{N}}$ and $\{c_n'\}_{n\in\mathbb{N}}$ converge to 0 as $n \to +\infty$. Analyse this case.

Now the relevant question is : and if $\{\mu_n\}_{n\in\mathbb{N}}$ converges in some sense, is the limit necessarily a probability measure? The answer is again no. Take $X_n = c_n$ but now with $c_n \to +\infty$ and note that $X_n \to +\infty$ and if $I = (a, b)$, then $\lim_{n\to+\infty}\mu_n(I) = 0$ for any $a, b \in \mathbb{R}$. So, if there is a limiting measure it gives weight 0 to any finite interval so that the limit should be equal to zero.

Definition 2.2.1. *A probability measure* μ *in* $(\mathbb{R}, \mathscr{B})$ *with* $\mu(\mathbb{R}) \leq 1$ *is a called a subprobability measure.*

Definition 2.2.2 (Weak convergence)**.** *A sequence of subprobability measures* $\{\mu_n\}_{n\in\mathbb{N}}$ *in* $(\mathbb{R}, \mathscr{B})$ *is said to converge weakly to a subprobability measure* μ *iff there exists a dense subset* D *of* $\mathbb{R}$ *such that* $\forall a, b \in D, a < b$*,*

$$\lim_{n\to\infty} \mu_n((a,b]) = \mu((a,b]).$$

We will use the notation $\mu_n \to^v \mu$ *and* μ *is said to be the weak limit of* $\{\mu_n\}_{n\in\mathbb{N}}$*.*

Definition 2.2.3. *An interval* (a,b) *is said to be a continuity interval of* μ *is* a, b *are not atoms of* μ *(in other words this means that* $\mu((a,b)) = \mu([a,b])$*).*

Lemma 2.2.4. *Let* $\{\mu_n\}_{n\in\mathbb{N}}$ *and* μ *be subprobability measures. The following propositions are equivalent:*

1. *For every finite interval* (a,b) *and* $\epsilon > 0$*, there exists an* $n_0(a,b,\epsilon)$ *such that if* $n \geq n_0$*, then*

$$\mu((a+\epsilon, b-\epsilon)) - \epsilon \leq \mu_n((a,b)) \leq \mu((a-\epsilon, b+\epsilon)) + \epsilon. \qquad (2.2.1)$$

2. *for every continuity interval* $(a,b]$ *of* μ *we have that*

$$\lim_{n\to\infty} \mu_n((a,b]) = \mu((a,b]).$$

3. $\mu_n \to^v \mu$*.*

Proof. Let us first prove that *1.* implies *2.* Let (a,b) be an interval of continuity of μ. From the monotonicity of μ is follows that

$$\lim_{\varepsilon\to 0} \mu(a+\varepsilon, b-\varepsilon) = \mu(a,b) \leq m\mu[a,b] = \lim_{\varepsilon\to 0} \mu(a-\varepsilon, b+\varepsilon).$$

Taking the limit as $n \to +\infty$ and then the limit when $\varepsilon \to 0$ in (2.2.1) we obtain that

$$\mu((a,b)) \leq \liminf_{n\to+\infty} \mu_n((a,b)) \leq \limsup_{n\to+\infty} \mu_n([a,b]) \leq \mu([a,b]) = \mu((a,b)).$$

Now we want to prove that Let us first prove that *2.* implies *3.* This means that we want to prove that there exists D a dense subset of $\mathbb{R}$ such that for any $a < b \in D$ it holds that $\mu_n((a,b])_{n\to+\infty} \mu((a,b])$. Since the set which contains the atoms of μ is at most numerable, then its complementary, let us call it D is dense. Therefore, if $a < b \in D$, then (a, b) is a continuity interval for μ and from *2.* it holds that

$$\lim_{n\to\infty} \mu_n((a,b]) = \mu((a,b]). \tag{2.2.2}$$

Now we prove that *3.* implies *1.* Given (a, b) and $\varepsilon > 0$, there exist $a_1, a_2, b_1, b_2 \in D$ satisfying

$$a - \varepsilon < a_1 < a < a_2 < a + \varepsilon \quad \text{and} \quad b - \varepsilon < b_1 < b < b_2 < b + \varepsilon.$$

Now, note that from the notion of weak convergence we have that there exists $n_0 \in \mathbb{N}$ such that for any $n \geq n_0$ and for $i = 1, 2$ and $j = 1, 2$

$$|\mu_n((a_i, b_j]) - \mu((a_i, b_j])| \leq \varepsilon.$$

Therefore,

$$\begin{aligned} &\mu((a+\epsilon, b-\epsilon)) - \varepsilon \leq \mu((a_2, b_1]) - \varepsilon \leq \mu_n((a_2, b_1]) \\ &\leq \mu_n((a,b)) \leq \mu_n((a, b_2]) \leq \mu((a_1, b_2]) + \varepsilon \leq \mu((a-\epsilon, b+\epsilon)) + \varepsilon. \end{aligned} \tag{2.2.3}$$

And this proves *1.* □

An immediate consequence of the theorem is that the weak limit is *unique*. Let us check it. Suppose that besides (2.2.2) we also have for $a < b \in D'$ where D' is a dense subset of $\mathbb{R}$ that

$$\lim_{n\to\infty} \mu_n((a,b]) = \tilde{\mu}((a,b]).$$

What we want is to show that $\mu = \tilde{\mu}$. Let $\mathscr{A}$ be the set of the common atoms of μ and $\tilde{\mu}$. Then, from item *2.* of the previous theorem we know that

$$\mu((a,b]) = \lim_{n\to\infty} \mu_n((a,b]) = \tilde{\mu}((a,b]).$$

So that $\mu((a,b]) = \tilde{\mu}((a,b])$ for all $a, b \in \mathscr{A}^c$. Since $\mathscr{A}^c$ is a dense subset of $\mathbb{R}$, then we know that the measure coincide in all the intervals whose extreme points are in a dense subset of $\mathbb{R}$ and from a Theorem that we have seen in the beginning of the course, we conclude that the two measures μ and $\tilde{\mu}$ are equal.

Recall that given any sequence of real numbers in a subset of $[0,1]$, there is a subsequence which converges and the limit is an element of that set. This means that $[0,1]$ is sequentially compact. Now we prove that the set of subprobability measures is sequentially compact with respect to the weak convergence.

Theorem 2.2.5 (Helly's extraction theorem)**.** *Given any sequence of subprobability measures, there exists a subsequence that converges weakly to a subprobability measure.*

Proof. Let $x \in \mathbb{R}$ and define the subdistribution function as $F(x) = \mu((-\infty, x])$. The function F has the same properties as the distribution function that we defined for distribution functions, that is, F is increasing, continuous from the right, $\lim_{x\to-\infty} F(x) = 0$ but $\lim_{x\to+\infty} F(x) \leq 1$. let D be a countable dense subset of $\mathbb{R}$ and let $\{r_k\}_{k\geq 1}$ be an enumeration of D. Note that the sequence of real numbers $\{F_n(r_1)\}_{n\geq 1}$ is bounded and by the Bolzano-Weierstrass theorem we know that there exists a subsequence $\{F_{1k}\}_{k\geq 1}$ of that sequence such that the limit $\lim_{k\to+\infty} F_{1k}(r_1) = \ell_1$ exists and let us denote it by ℓ_1. Clearly $0 \leq \ell_1 \leq 1$. Now we repeat the procedure. Note that the sequence of real numbers $\{F_{1k}(r_2)\}_{k\geq 1}$ is bounded, so that there exists a subsequence $\{F_{2k}\}_{k\geq 1}$ of $\{F_{1k}\}_{k\geq 1}$ such that $\lim_{k\to+\infty} F_{2k}(r_2) = \ell_2$ and again $0 \leq \ell_2 \leq 1$. Since $\{F_{2k}\}_{k\geq 1}$ is a subsequence of $\{F_{1k}\}_{k\geq 1}$, then it also holds that $\lim_{k\to+\infty} F_{2k}(r_1) = \ell_1$. Now we repeat the argument and at the m-th step we have a sequence $\{F_{mk}\}_{k\geq 1}$ such that $\lim_{k\to+\infty} F_{mk}(r_i) = \ell_i$ for all $i = 1, \cdots, m$. Now we consider the sequence $\{F_{kk}\}_{k\geq 1}$ which converges in every point r_m for $m \geq 1$. For that purpose, note that for r_m fixed, ignoring the first $m-1$ terms, the sequence $\{F_{kk}\}_{k\geq 1}$ is a subsequence of $\{F_{mk}\}_{k\geq 1}$ which converges in r_m to ℓ_m, so that $\{F_{kk}\}_{k\geq 1}$ also converges in r_m to ℓ_m. Up to now we have proved the existence of a subsequence $\{n_k\}_{k\geq 1}$ and of a function G defined on D, which is increasing and such that

$$\forall r \in D, \lim_{k\to+\infty} F_{n_k}(r) = G(r).$$

Now we need to extend the function to $\mathbb{R}$. For that purpose, let $F : \mathbb{R} \to \mathbb{R}$ be defined on $x \in \mathbb{R}$ by $F(x) = \inf_{x<r\in D} G(r)$. From Lemma 1.5.14, we know that F is increasing and right continuous everywhere. Let $\mathscr{C}$ be the points of continuity of F. Then $\mathscr{C}$ is dense and now we have to prove that for all $x \in \mathscr{C}$ it holds that

$\lim_{k\to+\infty} F_{n_k}(x) = F(x)$. We leave this as an exercise for the reader. To this function F corresponds a unique subprobability μ through the correspondence $F(x) = \mu((-\infty, x])$. The proof of this result is quite similar to the one of Theorem 1.3.6 and is left to the reader. Now we see that for all $a < b \in \mathscr{C}$, it holds that $\lim_{k\to+\infty} \mu_{n_k}((a,b]) = \lim_{k\to+\infty} F_{n_k}(b) - F_{n_k}(a) = F(b) - F(a) = \mu((a,b])$. This means that $\mu_{n_k} \to^v \mu$. □

Definition 2.2.6. *Given F_n and F subdistribution functions, we say that F_n converges weakly to F and we write $F_n \to^v F$ if $\mu_n \to^v \mu$, where μ_n and μ are the subprobability measures of F_n and F, respectively.*

Theorem 2.2.7. *If every weakly converging subsequence of a sequence $\{\mu_n\}_{n\in\mathbb{N}}$ of subprobability measures converges to the same μ, then $\mu_n \to^v \mu$.*

Proof. Let us suppose that μ_n does not converge weakly to μ. Then, from item *2.* of Lemma 2.2.4, there exists a continuity interval (a,b) of μ such that $\mu((a,b))$ does not converge to $\mu((a,b))$. Then, by the Bolzano-Weierstrass theorem, there exists a subsequence n_k going to $+\infty$ such that $\mu_{n_k}((a,b))$ converges to a limit, that we denote by L and we know that $L \neq \mu((a,b))$. From Helly's extraction theorem, we can extract from $\{\mu_{n_k}\}_{k\geq 1}$ a subsequence $\{\mu_{n'_k}\}_{k'\geq 1}$ such that it converges weakly to μ, by the hypothesis of the theorem. Then, from item *2.* of Lemma 2.2.4, we have that $\mu_{n'_k}((a,b)) \to_{k'\to+\infty} \mu((a,b))$ for any continuity interval (a,b) of μ. But $\mu_{n'_k}((a,b)) \to_{k'\to+\infty} L \neq \mu((a,b))$ and this is an absurd. □

Now we want to give another characterization of weak converge by integration the measures with certain spaces of test functions. For that purpose we need to introduce some notation. Let us define the following subsets of the set of continuous functions from $\mathbb{R}$ to $\mathbb{R}$.

Let

1. C_k be the space of functions $f : \mathbb{R} \to \mathbb{R}$ which are continuous and with compact support.

2. C_0 be the space of functions $f : \mathbb{R} \to \mathbb{R}$ which are continuous and go to zero at infinity: $\lim_{|x|\to\infty} f(x) = 0$.

3. C_B be the space of functions $f : \mathbb{R} \to \mathbb{R}$ which are continuous and bounded.

4. C be the space of functions $f : \mathbb{R} \to \mathbb{R}$ which are continuous.

Note that we have the following inclusion:

$$C_k \subset C_0 \subset C_B \subset C.$$

We recall now a lemma from real analysis which will be useful for our purposes.

Lemma 2.2.8. *Suppose that $f \in C_k$ has support in the interval $[a,b]$ (recall that the compact subsets of $\mathbb{R}$ are closed intervals). Given a dense subset $\mathscr{A}$ of $\mathbb{R}$ and $\varepsilon > 0$, there exists a simple function f_ε defined in (a,b) such that $\sup_{x\in\mathbb{R}} |f(x) - f_\varepsilon(x)| \le \varepsilon$. If we take $f \in C_0$ then the same results is true if (a,b) is replaced by $\mathbb{R}$.*

We have the following criterion for the weak convergence.

Theorem 2.2.9. *Let μ_n and μ be subprobability measures. Then $\mu_n \to^v \mu$ iff for all $f \in C_k$ (or C_0) we have that*

$$\int_{\mathbb{R}} f(x)\mu_n(dx) \to_{n\to\infty} \int_{\mathbb{R}} f(x)\mu(dx). \tag{2.2.4}$$

Proof. Suppose that $\mu_n \to^v \mu$. By definition we know that $\mu_n((a,b]) \to_{n\to+\infty} \mu((a,b])$ for $a,b \in D$, where D is a dense subset of $\mathbb{R}$. This means that (2.2.4) holds for $f = \mathbf{1}_{(a,b]}$. By linearity of the integral, it also holds for simple functions that take values in D. Now let $f \in C_0$ and let $\varepsilon > 0$. By Lemma 2.2.8 there exists a simple function f_ε which takes values in D and such that

$$\sup_{x\in\mathbb{R}} |f(x) - f_\varepsilon(x)| \le \varepsilon.$$

Then

$$\begin{aligned} \Big| \int_{\mathbb{R}} f(x)\mu_n(dx) - \int_{\mathbb{R}} f(x)\mu(dx) \Big| \le & \Big| \int (f - f_\varepsilon)(x)\mu_n(dx) \Big| \\ & + \Big| \int f_\varepsilon(x)\mu_n(dx) - \int f_\varepsilon(x)\mu(dx) \Big| \\ & + \Big| \int (f - f_\varepsilon)(x)\mu_(dx) \Big|. \end{aligned} \tag{2.2.5}$$

Now note that from Lemma 2.2.8 we have that $\left|\int (f - f_\varepsilon)(x)\mu_n(dx)\right| \le \varepsilon$ and the same bound is true for the same integral but with respect to the measure μ. On the other hand,

$$\lim_{n\to+\infty}\left|\int f_\varepsilon(x)\mu_n(dx) - \int f_\varepsilon(x)\mu(dx)\right| = 0 \tag{2.2.6}$$

since f_ε is a simple function. Therefore

$$\lim_{n\to+\infty}\left|\int_{\mathbb{R}} f(x)\mu_n(dx) - \int_{\mathbb{R}} f(x)\mu(dx)\right| \le 2\varepsilon \tag{2.2.7}$$

and since ε is arbitrary we can take it to 0 and we are done. Now we suppose that (2.2.4) is true for $f \in C_k$. Let $\mathscr{A}$ be the set of atoms of μ and let $D = \mathscr{A}^c$. We shall prove the weak convergence on the set D. For that purpose, let $g = \mathbf{1}_{(a,b]}$ with $a, b \in D$. Given $\varepsilon > 0$, there exists $\delta > 0$ such that $a + \delta < b - \delta$ and for $U = (a-\delta, a+\delta) \cup (b-\delta, b+\delta)$, we have $\mu(U) < \varepsilon$. Note that this is true since a and b are not atoms of μ. Now define the continuous function g_1 which is equal to 1 in $(a+\delta, b-\delta)$, equal to 0 in $(a, b)^c$, and in $(a, a+\delta)$ and in $(b-\delta, b)$ it is linear. Analogously define the continuous function g_2 which is equal to 1 in (a, b), equal to 0 in $(a-\delta, b+\delta)^c$, and in $(a-\delta, a)$ and in $(b, b+\delta)$ it is linear. From this we have that $g_1 \le g \le g_2 \le g_1 + 1$ and as a consequence

$$\int g_1(x)\mu_n(dx) \le \int g(x)\mu_n(dx) \le \int g_2(x)\mu_n(dx).$$

Since g_1 and g_2 are functions of compact support, by hypothesis we have that

$$\lim_{n\to+\infty}\int g_i(x)\mu_n(dx) = \int g_i(x)\mu(dx),$$

for $i = 1, 2$. On the other hand, we also have that

$$\int g_1(x)\mu(dx) \le \int g(x)\mu(dx) \le \int g_2(x)\mu(dx).$$

Since $\int g_2(x)\mu(dx) - \int g_1(x)\mu(dx) \le \int_U \mu(dx) = \mu(U) < \varepsilon$ and since ε is arbitrary we conclude that

$$\lim_{n\to+\infty}\int g(x)\mu_n(dx) = \int g(x)\mu(dx),$$

and we are done. $\square$

Corollary 2.2.10. *If $\{\mu_n\}_{n\geq 1}$ is a sequence of subprobability measures such that for any $f \in C_k$ the limit*

$$\lim_{n\to+\infty} \int_{\mathbb{R}} f(x)\mu_n(dx)$$

exists, then $\{\mu_n\}_{n\geq 1}$ weakly converges.

Proof. From Helly's extraction theorem, we know that there exists a subsequence $\{\mu_{n_k}\}_{k\geq 1}$ such that $\mu_{n_k} \to^v \mu$, where μ is a subprobability measure. From the previous theorem we also know that

$$\int_{\mathbb{R}} f(x)\mu_{n_k}(dx) \to_{n\to\infty} \int_{\mathbb{R}} f(x)\mu(dx).$$

From the uniqueness theorem, namely Theorem 2.2.7, if we prove that all the subsequence of $\{\mu_n\}_{n\geq 1}$ converges weakly to this measure μ, where μ is a subprobability measure, then we conclude that $\mu_n \to^v \mu$. Let $\{\mu_{n_j}\}_{j\geq 1}$ be a subsequence of $\{\mu_n\}_{n\geq 1}$ such that $\mu_{n_j} \to^v \nu$. We want to prove that $\mu = \nu$. Now we also know from the previous theorem that for any $f \in C_k$ we have that

$$\int_{\mathbb{R}} f(x)\mu_{n_j}(dx) \to_{n\to\infty} \int_{\mathbb{R}} f(x)\nu(dx).$$

From the hypothesis of the theorem we conclude that

$$\int_{\mathbb{R}} f(x)\mu(dx) = \int_{\mathbb{R}} f(x)\nu(dx),$$

for any $f \in C_k$. We leave the reader prove that the previous identity implies that $\mu = \nu$, from where the proof ends. □

Definition 2.2.11 (Convergence in distribution)**.** *A sequence of r.v. $\{X_n\}_{n\in\mathbb{N}}$ is said to converge in distribution to F iff the corresponding sequence of distribution functions $\{F_n\}_{n\in\mathbb{N}}$ converges weakly to the distribution function F.*

If X is a distribution function which has distribution function F, we will say that $\{X_n\}_{n\in\mathbb{N}}$ converges in distribution to X.

Theorem 2.2.12 (Convergence in probability implies convergence in distribution)**.** *Let F_n and F be the distribution functions of the r.v. X_n and X. If $\{X_n\}_{n\in\mathbb{N}}$ converges to X in probability, then $F_n \to^n F$.*

Proof. We start the proof by saying that if $X_n \to_{n\to+\infty} X$ in probability, then for any $f \in C_k$ it holds that $f(X_n) \to_{n\to+\infty} f(X)$ in probability. On the other hand also note that since $f \in C_k$ is also bounded, the previous convergence also holds in $\mathbb{L}^1$. This means that

$$\Big|\mathbb{E}[f(X_n) - f(x)]\Big| \leq \mathbb{E}[|f(X_n) - f(X)|] \to_{n\to+\infty} 0.$$

Last identity is equivalent to

$$\int_{\mathbb{R}} f(x)\mu_n(dx) \to_{n\to\infty} \int_{\mathbb{R}} f(x)\mu(dx)$$

for any function $f \in C_k$. From Theorem 2.2.9 this is equivalent to $\mu_n \to^{v} \mu$.

□

In the next lemma we prove that convergence in probability and convergence in distribution are equivalent when the limit is a constant.

Lemma 2.2.13. *Let $c \in \mathbb{R}$. Then $\{X_n\}_{n\in\mathbb{N}}$ converges to c in probability iff $\{X_n\}_{n\in\mathbb{N}}$ converges to c in distribution.*

Proof. From the previous theorem it is enough to show that convergence in distribution to a constant implies convergence in probability to the same constant c. Let μ_n be the measure induced by X_n and let μ be the measure induced by $X = c$. Recall that we want to prove that $\mathbb{P}(|X_n - c| > \varepsilon) \to_{n\to+\infty} 0$. Note that the previous probability is equal to

$$\mathbb{P}(|X_n - c| > \varepsilon) = \mathbb{P}(X_n \in (c-\varepsilon, c+\varepsilon)^c) = \mu_n((c-\varepsilon, c+\varepsilon)^c).$$

Let $I = (c-\varepsilon, c+\varepsilon)^c$. Then I is a continuity interval for μ for any $\varepsilon > 0$. By hypothesis, we know that

$$\lim_{n\to+\infty} \mu_n(I) = \mu(I) = 1 - \mu((c-\varepsilon, c+\varepsilon)) = 1 - \mathbb{P}(X \in (c-\varepsilon, c+\varepsilon)) = 0.$$

With this the proof ends. □

Note that it is not true that if $\{X_n\}_{n\in\mathbb{N}}$ converges in distribution to X and $\{Y_n\}_{n\in\mathbb{N}}$ converges in distribution to Y, the sum $\{X_n+Y_n\}_{n\in\mathbb{N}}$ converges in distribution to $X+Y$. We will see in the next chapter that last sentence is true when we add the hypothesis that X_n and Y_n are independent. For now we see the special case when $Y=0$.

Theorem 2.2.14. *If $\{X_n\}_{n\in\mathbb{N}}$ converges in distribution to X and $\{Y_n\}_{n\in\mathbb{N}}$ converges in distribution to* 0*, then:*

- $\{X_n+Y_n\}_{n\in\mathbb{N}}$ *converges in distribution to X*
- $\{X_nY_n\}_{n\in\mathbb{N}}$ *converges in distribution to* 0

Proof. Let us prove the first item. Take $f\in C_k$ and suppose that M is a constant such that $|f|\leq M$. Since f is continuous of compact support, it is bounded by M and it is uniformly continuous. Then given $\varepsilon>0$ there exists δ such that $|x-y|<\delta$ then $|f(x)-f(y)|<\varepsilon$. As a consequence

$$\begin{aligned}\mathbb{E}[|f(X_n+Y_n)-f(X_n)|] &\leq \int_{\{|f(X_n+Y_n)-f(X_n)|\leq\varepsilon\}}\varepsilon d\mathbb{P}+2M\int_{\{|f(X_n+Y_n)-f(X_n)|>\varepsilon\}}d\mathbb{P}\\ &\leq \varepsilon\mathbb{P}(f(X_n+Y_n)-f(X_n)|\leq\varepsilon)+2M\mathbb{P}(f(X_n+Y_n)-f(x)|>\varepsilon)\\ &\leq \varepsilon+2M\mathbb{P}(|Y_n|>\delta),\end{aligned}$$

and since Y_n converges in distribution to 0, from the previous lemma, it converges to 0 in probability, so that when we take the limit as $n\to+\infty$ in last inequality we obtain that

$$\lim_{n\to+\infty}\mathbb{E}[|f(X_n+Y_n)-f(X_n)|]\leq\varepsilon. \tag{2.2.8}$$

Since ε is arbitrary, taking it to 0, we conclude that

$$\lim_{n\to+\infty}\mathbb{E}[|f(X_n+Y_n)-f(X_n)|]=\lim_{n\to+\infty}\mathbb{E}[f(X_n)]=\mathbb{E}[f(X)].$$

In the last equality we used Theorem 2.2.9. From this we conclude that for any $f\in C_k$ it holds that

$$\int f(X_n+Y_n)d\mathbb{P}\to_{n\to\infty}\int f(X)d\mathbb{P},$$

which means that $X_n + Y_n$ converges in distribution to X.

Now we prove the second item. Given $\varepsilon > 0$ let us choose A such that $\pm A$ are both continuity points of the distribution function of the r.v. X and sufficiently big such that $\lim_{n\to+\infty} \mathbb{P}(|X_n| > A) = \mathbb{P}(|X| > A) < \varepsilon$. Note that the first limit is true since X_n converges to X in distribution. The inequality above is a consequence of the fact that $\mathbb{P}(|X| > A) = \mu((-A,A)^c)$, μ is a probability measure and A is quite big. Then $\mathbb{P}(|X_n| > A) < \varepsilon$ for all $n \geq n(\varepsilon)$. But for $n \geq n(\varepsilon)$ it holds that

$$\begin{aligned}
\mathbb{P}(|X_nY_n| > \varepsilon) &= \mathbb{P}(|X_nY_n| > \varepsilon, |X_n| > A) + \mathbb{P}(|X_nY_n| > \varepsilon, |X_n| \leq A) \\
&\leq \mathbb{P}(||X_n| > A) + \mathbb{P}\Big(|Y_n| > \frac{\varepsilon}{A}\Big) \qquad (2.2.9) \\
&\leq \varepsilon + \mathbb{P}\Big(|Y_n| > \frac{\varepsilon}{A}\Big).
\end{aligned}$$

Now for n sufficiently the last inequality becomes $\mathbb{P}(|X_nY_n| > \varepsilon) \leq \varepsilon$, and since ε is arbitrary we can send it to 0 and we proved that $\{X_nY_n\}_{n\in\mathbb{N}}$ converges in probability to 0 which implies the convergence in distribution. □

We finish now with the following corollary whose proof we leave for the reader.

Corollary 2.2.15. *If X_n converges to X in distribution , if the sequences of real numbers α_n and β_n converges, respectively, to α and β, then $\alpha X_n + \beta_n$ converges in distribution to $\alpha X + \beta$.*

Exercise:

Analyse if X_n converges in probability and in distribution to X, where $\Omega = \{0,1\}$, $X_n(0) = 0$ and $X_n(1) = 1$ each with probability $1/2$ and $X(0) = 1$ and $X(1) = 0$ each with probability $1/2$.

The first thing to compute is the distribution function of X_n and of X and we find out that they are the same. Therefore, the convergence in distribution is true. But $\mathbb{P}(|X_n(\omega) - X(\omega)| > \varepsilon)$ does not vanish as $n \to +\infty$ since $|X_n(\omega) - X(\omega)| = 1$ for all n and for all $\omega \in \Omega$. Then the convergence in probability does not hold.

Now we note that from item *2.* of Lemma 2.2.4 we know that $\mu_n \to^v \mu$ iff for every continuity interval $(a,b]$ of μ we have that

$$\lim_{n\to\infty} \mu_n((a,b]) = \mu((a,b]).$$

this can be translated into (check!) saying that X_n converges in distribution to X iff $F_n(x) \to_{n\to+\infty} F(x)$ for all x continuity point of F.

Exercise:
Take $X_n = 1/n$ and $X = 0$ and analyse the convergence in distribution of X_n to 0.

2.3 Law of Large Numbers

2.3.1 Weak Law of Large Numbers

The law of large numbers has to do with the partial sums of a sequence of r.v. $\{X_n\}_{n\geq 1}$.

$$S_n := X_1 + \cdots + X_n$$

The notion of weak or strong law of large numbers depends on whether

$$\frac{S_n - \mathbb{E}[S_n]}{n} \to_{n\to\infty} 0,$$

in probability or a.e. (note that it needs that $\mathbb{E}[S_n]$ to be finite!) We have seen in the previous chapters that if a sequence converges to 0 in $\mathbb{L}^2$ then it converges to 0 in probability and then it converges a.e. to 0 along a subsequence. Note that

$$\mathbb{E}[S_n^2] = \mathbb{E}\Big[\sum_{j=1}^n X_j^2 + \sum_{i\neq j} X_i X_j\Big] = O(n^2).$$

But if $\{X_j\}_{j\geq 1}$ are uncorrelated and mean zero, we have that

$$\mathbb{E}[S_n^2] = \mathbb{E}\Big[\sum_{j=1}^n X_j^2\Big] = O(n).$$

From where we conclude that $\frac{S_n}{n}$ converges to 0 in $\mathbb{L}^2$.

Theorem 2.3.1 (Tchebychev)**.**

If $\{X_n\}_{n\in\mathbb{N}}$ *is a sequence of uncorrelated r.v. whose second moments have a common bound, then* $\frac{S_n - \mathbb{E}[S_n]}{n} \rightarrow_{n\rightarrow\infty} 0$ *in* $\mathbb{L}^2$ *and therefore also in probability.*

Proof. It is enough to suppose first that X_n has mean zero for all $n \geq 1$ and then use the computations above to conclude the result for X_n. After that take $Y_n = X_n - \mathbb{E}[X_n]$ to conclude the result in the case where the r.v. are not mean zero. □

In fact the previous result also holds with convergence a.e. This is the content of the next theorem.

Theorem 2.3.2 (Rajchman)**.**

If $\{X_n\}_{n\in\mathbb{N}}$ *is a sequence of uncorrelated r.v. whose second moments have a common bound, then*

$$\frac{S_n - \mathbb{E}[S_n]}{n} \rightarrow_{n\rightarrow\infty} 0 \quad a.e.$$

Proof. Let us start by supposing that $\mathbb{E}[X_n] = 0$ for all $n \geq 1$ and then we repeat the argument of the previous proof. At this point we know that $\mathbb{E}[S_n^2] \leq Mn$ where M is a positive constant. From this is follows that $\mathbb{P}(|S_n| > n\varepsilon) \leq \frac{M}{n\varepsilon^2}$. We want to prove that $S_n/n \rightarrow 0$ a.e. that is $\mathbb{P}(\limsup_n \{|S_n| > n\varepsilon\}) = 0$ and from Borel-Cantelli's Lemma (converging part) it is enough to prove that

$$\sum_{n\geq} \mathbb{P}(|S_n| > n\varepsilon) < +\infty.$$

From the computations above, we conclude that

$$\sum_{n\geq} \mathbb{P}(|S_{n^2}| > n^2\varepsilon) < \sum_{n\geq 1} \frac{M}{n^2\varepsilon^2} < +\infty,$$

and as consequence $S_{n^2}/n^2 \rightarrow_{n\rightarrow+\infty} 0$ a.e. Up to now we have proved the result for a subsequence and we want to prove it to the whole sequence. For that purpose, for $n \geq 1$ let $D_n = \max_{n^2 \leq k < (n+1)^2} |S_k - S_{n^2}|$. We have that

$$\frac{|S_k|}{k} = \frac{|S_k - S_{n^2} + S_{n^2}|}{k} \leq \frac{|S_k - S_{n^2}|}{n^2} + \frac{|S_{n^2}|}{n^2} \leq \frac{D_n}{n^2} + \frac{|S_{n^2}|}{n^2}.$$

So the proof ends as long as we show that $D_n/n^2 \to_{n\to+\infty} 0$ a.e. Note that

$$\mathbb{P}\Big(\frac{D_n}{n^2} > \varepsilon\Big) = \mathbb{P}(D_n > n^2\varepsilon) \leq \frac{\mathbb{E}[(D_n)^2]}{\varepsilon^2 n^4}.$$

We want to show that $\sum_{n\geq 1} \frac{\mathbb{E}[(D_n)^2]}{\varepsilon^2 n^4} < +\infty$ and then from Borel-Cantelli's Lemma we are done. Note that

$$\begin{aligned} D_n^2 &= \max_{n^2\leq k<(n+1)^2} |S_k - S_{n^2}|^2 = \max_{n^2\leq k<(n+1)^2} \Big|\Big(\sum_{i=n^2+1}^{k} X_i\Big)^2\Big| \\ &\leq \max_{n^2\leq k<(n+1)^2} (k-(n^2+1)) \sum_{i=n^2+1}^{k} X_i^2 \leq ((n+1)^2-(n^2+1)) \sum_{i=n^2+1}^{k} X_i^2. \end{aligned} \tag{2.3.1}$$

From this it follows that

$$\frac{\mathbb{E}[D_n^2]}{\varepsilon^2 n^4} \leq \frac{2n}{\varepsilon^2 n^4} \sum_{i=n^2+1}^{k} \mathbb{E}[X_i^2] \leq \frac{4M}{\varepsilon^2 n^2}$$

which implies that $\sum_{n\geq 1} \frac{\mathbb{E}[(D_n)^2]}{\varepsilon^2 n^4} < +\infty$ and we are done. □

Up to now we have seen the convergence of $\frac{S_n - \mathbb{E}[S_n]}{n}$ a.e., $\mathbb{L}^2$ and in probability but we assumed that the second moments of X_j are finite for all j. Now we want to weak that hypothesis in order to prove the law of large numbers. We start with the notion of equivalent sequences.

Definition 2.3.3 (Equivalent sequences (Kintchine))**.**

Two sequences of r.v. $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\in\mathbb{N}}$ are said to be equivalent iff

$$\sum_{n\geq 1} \mathbb{P}(X_n \neq Y_n) < \infty.$$

Theorem 2.3.4. *If $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\in\mathbb{N}}$ are equivalent then*

$$\sum_{n\geq 1} (X_n - Y_n)$$

converges a.e. Moreover, if $a_n \to +\infty$, then

$$\frac{1}{a_n} \sum_{j=1}^{n} (X_j - Y_j)$$

converges a.e. to 0.

Proof. From Borel-Cantelli's Lemma, we know that since the sequences are equivalent, then

$$\mathbb{P}(\limsup_n \{X_n \neq Y_n\}) = 0 \iff \mathbb{P}(\liminf_n \{X_n \neq Y_n\}) = 1.$$

This means that there exists a set Ω_0 with $\mathbb{P}(\Omega_0) = 1$ such that for all $\omega \in \Omega_0$ there exists an order $n_0(\omega)$ such that for all $n \geq n_0(\omega)$, by definition of lim inf, it holds that $X_n(\omega) = Y_n(\omega)$. Then, $\sum_{n\geq 1}(X_n - Y_n) = \sum_{n=1}^{n_0}(X_n - Y_n)$ and this is finite. This proves the first result. Now the second follows from $\frac{1}{a_n}\sum_{j\geq 1}(X_j - Y_j) = \frac{1}{a_n}\sum_{j=1}^{n_0}(X_j - Y_j) \to_{n\to+\infty} 0$.

□

Corollary 2.3.5. *In the same conditions as in the previous theorem, With probability* 1*, the expression* $\sum_{n\geq 1} X_n$ *or* $\frac{1}{a_n}\sum_{j=1}^{n} X_j$ *converges or diverges to* $\pm\infty$ *in the same way as* $\sum_{n\geq 1} Y_n$ *or* $\frac{1}{a_n}\sum_{j=1}^{n} Y_j$*. In particular, if* $\frac{1}{a_n}\sum_{j=1}^{n} X_j$ *converges to* X *in probability, then* $\frac{1}{a_n}\sum_{j=1}^{n} Y_j$ *also does.*

Proof. Let us prove the last sentence, the other is left to the reader. From the previous theorem we have that $\frac{1}{a_n}\sum_{j=1}^{n}(X_j - Y_j)$ converges to 0 a.e. and therefore it also converges in probability. If $\frac{1}{a_n}\sum_{j=1}^{n} X_j$ converges to X in probability then

$$\frac{1}{a_n}\sum_{j=1}^{n} Y_j = \frac{1}{a_n}\sum_{j=1}^{n} X_j + \frac{1}{a_n}\sum_{j=1}^{n}(Y_j - X_j$$

converges to X in probability and we are done. □

Theorem 2.3.6 (Weak Law of Large Numbers of Kintchine)**.**

If $\{X_n\}_{n\geq 1}$ *is a sequence of pairwise independent and identically distributed r.v. with finite mean* m*, then* $\frac{S_n}{n} \to m$ *in probability.*

Proof. Let F be the distribution function of X_n for all $n \geq 1$. Then $m = \mathbb{E}[X_n] = \int_{\mathbb{R}} x dF(x)$ and $\mathbb{E}[|X_n|] = \int_{\mathbb{R}} |x| dF(x) < +\infty$ We have already seen in Theorem 1.8.8 that

$$\mathbb{E}[|X_n|] < +\infty \iff \sum_{n\geq 1}\mathbb{P}(|X_n| > n) < +\infty.$$

Now let $Y_n = X_n \mathbf{1}_{|X_n|\leq n}$. A simple computation shows that the sequences $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ are equivalent. Now let $T_n = \sum_{j=1}^n Y_j$ and note that by the previous corollary the proof ends as long as we show that $\frac{T_n}{n} \to m$ in probability. But now the advantage is that T_n is the sum of r.v. which are pairwise independent and with finite second moments (since they are bounded). Note that

$$\sigma^2(T_n) = \sum_{j=1}^n \sigma^2(Y_j) \leq \sum_{j=1}^n \mathbb{E}[Y_j^2] = \sum_{j=1}^n \int_{\{|x|\leq j\}} |x|^2 dF(x) \leq \sum_{j=1}^n j \int_{\{|x|\leq j\}} |x| dF(x)$$
$$\leq \sum_{j=1}^n j \int_{\mathbb{R}} |x| dF(x) \leq \frac{n(n+1)}{2} \mathbb{E}[|X_1|] = O(n^2). \tag{2.3.2}$$

In the first equality we used the fact that the r.v. Y_j are uncorrelated. As we have seen above, last bound is not good for our purposes, we need something better. So let us consider a sequence of integer numbers $\{a_n\}_{n\geq 1}$ such that $\lim_n a_n = +\infty$, but with $a(n) = o(n)$, for example, $a_n = \sqrt{n}$. Then we have

$$\begin{aligned}\sum_{j=1}^n \int_{\{|x|\leq j\}} |x|^2 dF(x) &= \Big(\sum_{j=1}^{a_n} + \sum_{j=a_n+1}^{n} \Big) \int_{\{|x|\leq j\}} |x|^2 dF(x) \\ &\leq \sum_{j\leq a_n} a_n \int_{\{|x|\leq a_n\}} |x| dF(x) + \sum_{j=a_n+1}^{n} a_n \int_{\{|x|\leq a_n\}} |x| dF(x) \\ &\quad + \sum_{j=a_n+1}^{n} n \int_{\{a_n<|x|\leq n\}} |x| dF(x) \\ &\leq \sum_{j=1}^{n} a_n \int_{\{|x|\leq a_n\}} |x| dF(x) + n \sum_{j=a_n+1}^{n} \int_{\{a_n<|x|\leq n\}} |x| dF(x) \\ &\leq n a_n \mathbb{E}[|X_1|] + n^2 \int_{\{|x|>a_n\}} |x| dF(x).\end{aligned} \tag{2.3.3}$$

From this we conclude that

$$\frac{1}{n^2} \sum_{j=1}^n \int_{\{|x|\leq j\}} |x|^2 dF(x) \leq \frac{a_n}{n} \mathbb{E}[|X_1|] + \int_{\{|x|>a_n\}} |x| dF(x). \tag{2.3.4}$$

The first term at the right hand side of last expression vanishes as $n \to +\infty$ since $a_n = o(n)$ and the second term also vanishes as $n \to +\infty$ since the r.v.

X_j are integrable (since they have finite mean) and the probability of set in the integral is vanishing as $n \to +\infty$. Therefore we conclude that $\sigma^2(T_n) = o(n^2)$. Now we use Tchebychev's inequality to conclude that

$$\mathbb{P}(|T_n - \mathbb{E}[T_n]| > \varepsilon n|) \leq \frac{\sigma^2(T_n)}{\varepsilon^2 n^2} \to_{n\to+\infty} 0.$$

From this we conclude that

$$\frac{T_n - \mathbb{E}[T_n]}{n} = \frac{\sum_{j=1}^n (Y_j - \mathbb{E}[Y_j])}{n} \to_{n\to+\infty} 0.$$

Now we just have to argue that $\mathbb{E}[Y_j] = \mathbb{E}[X_j \mathbf{1}_{\{X_j \leq j\}}] \to_{j+\infty} \mathbb{E}[X_1] = m$ from where we conclude that $T_n/n \to_{n\to+\infty} m$. This ends the proof. □

2.3.2 Convergence of Series

Theorem 2.3.7 (Kolmogorov's Inequality)**.**

Let $\{X_n\}_{n\geq 1}$ *be independent r.v. with* $\mathbb{E}[X_n] = 0$ *for every* n *and* $\mathbb{E}[X_n^2] = \sigma^2(X_n) < \infty$*. Then, for every* $\varepsilon > 0$ *it holds that*

$$\mathbb{P}\Big(\max_{1\leq j\leq n} |S_j| > \varepsilon\Big) \leq \frac{\sigma^2(S_n)}{\varepsilon^2}.$$

Proof. Fix $\varepsilon > 0$ and for $\omega \in \Lambda$ with $\Lambda = \{\omega : \max_{1\leq j\leq n} |S_j(\omega)| > \varepsilon\}$ we define the r.v.

$$\nu(\omega) = \min\{j : 1 \leq j \leq n; |S_j(\omega)| > \varepsilon\}.$$

Let

$$\Lambda_k = \{\omega : \nu(\omega) = k\} = \{\omega : \max_{1\leq j\leq k-1} |S_j(\omega)| \leq \varepsilon, |S_k(\omega)| > \varepsilon\}.$$

Note that the Λ_k tells us that when we sum k r.v. X_j then $|S_k|$ is, for the first time, bigger that ε which means that the previous sums have absolute value less than ε. Also note that for $k = 1$, above in Λ_k we should convention that $\max_{1\leq j\leq 0} |S_j(\omega)|$ is fixed as being equal to 0. So ν is the first time the max of S_j exceeds ε and Λ_k is the event where that happens for the first time in the k-th step. Note that the Λ_k are disjoint and $\Lambda = \cup_{k=1}^n \Lambda_k$. Then,

$$\begin{aligned}\int_\Lambda S_n^2 d\mathbb{P} &= \int_{\cup_{k=1}^n \Lambda_k} S_n^2 d\mathbb{P} = \sum_{k=1}^n \int_{\Lambda_k} S_n^2 d\mathbb{P} \\ &= \sum_{k=1}^n \int_{\Lambda_k} (S_k + S_n - S_k)^2 d\mathbb{P} \\ &= \sum_{k=1}^n \int_{\Lambda_k} S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2 d\mathbb{P}\end{aligned}$$

Now note that for $\varphi_k = \mathbf{1}_{\Lambda_k}$, the r.v. $\varphi_k S_k$ and $S_n - S_k$ are independent since $\varphi_k S_k$ is a function of the r.v. $\{X_1, \cdots, X_k\}$ and $S_n - S_k$ is a function of the r.v. $\{X_{k+1}, \cdots, X_n\}$. From this observation it follows that $\int_{\Lambda_k} S_k(S_n - S_k) d\mathbb{P} = 0$. Then

$$\sigma^2(S_n) = \int S_n^2 d\mathbb{P} \geq \int_\Lambda S_n^2 d\mathbb{P} \geq \sum_{k=1}^n \int_{\Lambda_k} S_n^2 d\mathbb{P} \geq \varepsilon^2 \sum_{k=1}^n \mathbb{P}(\Lambda_k) = \varepsilon^2 \mathbb{P}(\Lambda)$$

from where we conclude that

$$P(\Lambda) = \mathbb{P}\Big(\max_{1\leq j\leq n} |S_j| > \varepsilon \Big) \leq \frac{\sigma^2(S_n)}{\varepsilon^2}.$$

□

The next result does not impose any condition on the second moments of X_n. We leave the proof for the interested reader.

Theorem 2.3.8. *Let $\{X_n\}_{n\geq 1}$ be independent r.v. with $\mathbb{E}[X_n] < \infty$ and suppose that $\exists A > 0$ s.t. $|X_n - \mathbb{E}[X_n]| \leq A < \infty$, $\forall n \in \mathbb{N}$. Then, $\forall \varepsilon > 0$:*

$$\mathbb{P}\Big(\max_{1\leq j\leq n} |S_j| \leq \varepsilon \Big) \leq \frac{(2A + 4\varepsilon)^2}{\sigma^2(S_n)}.$$

Now we also state a theorem that we will not prove but that will be needed below. We leave the proof for the interested reader.

Theorem 2.3.9 (Three series of Kolmogorov)**.**

Let $\{X_n\}_{n\geq 1}$ be a sequence of independent r.v. and for a positive constant $A > 0$ let $Y_n = X_n \mathbf{1}_{\{|X_n|\leq A\}}$. Then the series $\sum_{n\geq 1} X_n$ converges a.e. iff the following three series converge:

1. $\sum_{n\geq 1} \mathbb{P}(|X_n| > A) = \sum_{n\geq 1} \mathbb{P}(X_n \neq Y_n)$
2. $\sum_{n\geq 1} \mathbb{E}[Y_n]$
3. $\sum_{n\geq 1} \sigma^2(Y_n)$.

Exercise:
Prove the previous theorem.

2.3.3 Strong Law of Large Numbers

Now our interest is focused in showing the strong law of large numbers. We start with the next lemma which will be useful in what follows.

Lemma 2.3.10 (Kronecker's Lemma).
Let $\{x_k\}_{k\geq 1}$ a sequence of real numbers, $\{a_k\}_{k\geq 1}$ a sequence of strictly positive real numbers $\uparrow \infty$. If $\sum_{n\geq 1} \frac{x_n}{a_n} < \infty$, then $\frac{1}{a_n}\sum_{j=1}^n x_j \to 0$.

Proof. For $n \geq 1$ let $b_n = \sum_{j=1}^n \frac{x_j}{a_j}$ and note that b_∞ exists since it is equal to the sum of the series. Let $a_0 = b_0 = 0$. Then $\frac{x_n}{a_n} = b_n - b_{n-1}$ which means that $x_n = a_n(b_n - b_{n-1})$. Therefore,

$$\frac{1}{a_n}\sum_{j=1}^n x_j = \frac{1}{a_n}\sum_{j=1}^n a_j(b_j - b_{j-1}) = \frac{1}{a_n}\sum_{j=1}^n a_j b_j - \frac{1}{a_n}\sum_{j=0}^{n-1} a_{j+1} b_j$$
$$= b_n - \frac{1}{a_n}\sum_{j=0}^{n-1}(a_{j+1} - a_j)b_j.$$

Now note that

$$\frac{1}{a_n}\sum_{j=0}^{n-1}(a_{j+1} - a_j) = \frac{a_n - a_0}{a_n} = 1.$$

From here is follows that

$$\lim_{n\to+\infty}\frac{1}{a_n}\sum_{j=1}^n x_j = \lim_{n\to+\infty}(b_n - \frac{1}{a_n}\sum_{j=0}^{n-1}(a_{j+1} - a_j)b_j)$$
$$= b_\infty - \lim_{n\to+\infty}\frac{1}{a_n}\sum_{j=0}^{n-1}(a_{j+1} - a_j)b_j,$$

and

$$\frac{1}{a_n}\sum_{j=0}^{n-1}(a_{j+1}-a_j)b_j = \frac{1}{a_n}\sum_{j=0}^{n-1}(a_{j+1}-a_j)(b_j-b_\infty)+\frac{1}{a_n}\sum_{j=0}^{n-1}(a_{j+1}-a_j)b_\infty,$$

and the term at the left hand side of last expression vanishes as $n \to +\infty$ and the term at the right hand side of last expression is equal to b_∞. From this the proof ends. □

Theorem 2.3.11. *Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a positive, even and continuous function, such that as $|x|$ increases: $\frac{\varphi(x)}{|x|} \uparrow$ and $\frac{\varphi(x)}{x^2} \downarrow$. Let $\{X_n\}_{n\geq 1}$ be a sequence of independent r.v. with $\mathbb{E}[X_n] = 0$ for every n and let $0 < a_n \uparrow +\infty$. If $\sum_{n\geq 1} \frac{\mathbb{E}[\varphi(X_n)]}{\varphi(a_n)} < \infty$, then $\sum_{n\geq 1} \frac{X_n}{a_n}$ converges a.e.*

Proof. Let F_n denote the distribution function of X_n and define for each $n \geq 1$ the r.v. $Y_n = X_n \mathbf{1}_{|X_n|\leq a_n}$. Then

$$\sum_{n\geq 1}\mathbb{E}\Big[\frac{Y_n^2}{a_n^2}\Big] = \sum_{n\geq 1}\int_{\{|x|\leq a_n\}}\frac{x^2}{a_n^2}dF_n(x).$$

Note that by the hypothesis in φ we have that if $|x| \leq a_n$ then $\frac{\varphi(x)}{x^2} \geq \frac{\varphi(a_n)}{a_n^2}$. Then

$$\sum_{n\geq 1}\sigma^2\Big(\frac{Y_n}{a_n}\Big) \leq \sum_{n\geq 1}\mathbb{E}\Big[\frac{Y_n^2}{a_n^2}\Big] \leq \sum_{n\geq 1}\int_{\{|x|\leq a_n\}}\frac{\varphi(x)}{\varphi(a_n)}dF_n(x) \leq \sum_{n\geq 1}\mathbb{E}\Big[\frac{\varphi(x)}{\varphi(a_n)}\Big]$$
$$= \sum_{n\geq 1}\frac{\mathbb{E}[\varphi(x)]}{\varphi(a_n)} < +\infty.$$

Taking the sequence of r.v. $\Big\{\frac{Y_n-\mathbb{E}[Y_n]}{a_n}\Big\}_{n\geq 1}$, then $\mathbb{E}\Big[\frac{Y_n-\mathbb{E}[Y_n]}{a_n}\Big] = 0$, $\Big|\frac{Y_n-\mathbb{E}[Y_n]}{a_n}\Big| \leq 2$ and finally $\sum_{n\geq 1}\sigma^2(\frac{Y_n}{a_n}) < +\infty$. Then from the Theorem of Three series of Kolmogorov, namely Theorem 2.3.9, we have that $\sum_{n\geq 1}\frac{Y_n-\mathbb{E}[Y_n]}{a_n}$ converges a.e. On the other hand

$$\sum_{n\geq 1}\frac{|\mathbb{E}[Y_n]|}{a_n} = \sum_{n\geq 1}\frac{1}{a_n}\Big|\int_{\{|x|\leq a_n\}}x\,dF_n(x)\Big| = \sum_{n\geq 1}\frac{1}{a_n}\Big|\int_{\{|x|>a_n\}}x\,dF_n(x)\Big|$$
$$\leq \sum_{n\geq 1}\frac{1}{a_n}\int_{\{|x|>a_n\}}|x|\,dF_n(x).$$

Since for $|x| > a_n$ we have that $\frac{|x|}{a_n} \leq \frac{\varphi(x)}{\varphi(a_n)}$, then

$$\sum_{n\geq 1} \frac{|\mathbb{E}[Y_n]|}{a_n} \leq \sum_{n\geq 1} \int_{\{|x|>a_n\}} \frac{\varphi(x)}{\varphi(a_n)} dF_n(x) \leq \sum_{n\geq 1} \frac{\mathbb{E}[\varphi(X_n)]}{\varphi(a_n)} < +\infty$$

From this it follows that $\sum_{n\geq 1} \frac{Y_n}{a_n}$ converges a.e. To finish the proof it remains to check that the sequences $\{X_n\}_{n\ geq1}$ and $\{Y_n\}_{n\geq 1}$ are equivalent. To prove it note that:

$$\sum_{n\geq 1} \mathbb{P}(X_n \neq Y_n) = \sum_{n\geq 1} \mathbb{P}(|X_n| > a_n) = \sum_{n\geq 1} \int_{\{|x|>a_n\}} dF_n(x)$$
$$\leq \sum_{n\geq 1} \int_{\{|x|>a_n\}} \frac{\varphi(x)}{\varphi(a_n)} dF_n(x) \leq \sum_{n\geq 1} \frac{\mathbb{E}[\varphi(X_n)]}{\varphi(a_n)} < +\infty.$$

Since the sequences are equivalent we conclude that $\sum_{n\geq 1} \frac{X_n}{a_n}$ converges a.e., which ends the proof. □

We note that being in the hypothesis of the previous theorem, from Kronecker's lemma we can conclude that $\frac{1}{a_n} \sum_{j=1}^{n} X_j$ converges a.e., as $n \to +\infty$, to 0.

Now we state the Strong Law of Large Numbers due to Kolmogorov.

Theorem 2.3.12 (Strong Law of Large Numbers of Kolmogorov)**.**

Let $\{X_n\}_{n\geq 1}$ be a sequence of independent and identically distributed r.v., then

1. $\mathbb{E}[|X_1|] < \infty \Rightarrow \frac{S_n}{n} \to \mathbb{E}[X_1]$ *a.e.*
2. $E[|X_1|] = \infty \Rightarrow \limsup_n \frac{|S_n|}{n} = +\infty$ *a.e.*

Proof. Let us start with the proof of the first item. For each $n \geq 1$ the r.v. $Y_n = X_n \mathbf{1}_{|X_n|\leq n}$. Then

$$\sum_{n\geq 1} \mathbb{P}(X_n \neq Y_n) = \sum_{n\geq 1} \mathbb{P}(|X_n| > n) = \sum_{n\geq 1} \mathbb{P}(|X_1| > n) < +\infty,$$

since X_1 is integrable. Now we apply the previous corollary with $\varphi(x) = x^2$ to the sequence $\{Y_n - \mathbb{E}[Y_n]\}_{n\geq 1}$. Then,

$$\sum_{n\geq 1} \frac{\sigma^2(Y_n)}{n^2} \leq \sum_{n\geq 1} \frac{\mathbb{E}[Y_n^2]}{n^2} = \sum_{n\geq 1} \frac{1}{n^2} \int_{\{|x|\leq n\}} x^2 dF(x).$$

In this case we do not have any information about the second moments so that we proceed as follows. Last term is equal to:

$$\sum_{n\geq 1}\frac{1}{n^2}\sum_{j=1}^{n}\int_{\{j-1<|x|\leq j\}}x^2dF(x)=\sum_{j\geq 1}\int_{\{j-1<|x|\leq j\}}x^2dF(x)\sum_{n\geq j}\frac{1}{n^2}.$$

Since last series is convergent (take the integral test for example) and since $sum_{n\geq j}\frac{1}{n^2}\leq C/j$, then the previous expression is bounded from above by

$$\sum_{j\geq 1}\int_{\{j-1<|x|\leq j\}}x^2dF(x)\sum_{n\geq j}\frac{1}{n^2}\frac{C}{j}=C\mathbb{E}[|X_1|]<+\infty.$$

From the previous observation we conclude that $\frac{1}{n}\sum_{j=1}^{n}(Y_j-\mathbb{E}[Y_j])$ converges a.e., as $n\to+\infty$, to 0. On the other hand $\mathbb{E}[Y_n]\to_{n\to+\infty}\mathbb{E}[X_1]$ and $\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}[Y_j]$ converges, as $n\to+\infty$, to $\mathbb{E}[X_1]$, from where we conclude that $\frac{1}{n}\sum_{j=1}^{n}Y_j$ converges a.e., as $n\to+\infty$, to $\mathbb{E}[X_1]$. Since the sequences are equivalent the proof of the first item ends. Now we prove the second item. Let $A>0$. Then by hypothesis we have that $\frac{\mathbb{E}[|X_1|]}{A}=+\infty$. From the integrability criterion, namely Theorem 1.8.8 we have that $\sum_{n\geq 1}\mathbb{P}(|X_1|>An)=+\infty$. Then $|S_n-S_n-1|=|X_n|>An$ implies that $|S_n|>An/2$ or $|S_{n-1}|>An/2$. Since Borel-Cantelli's Lemma (the divergent part) implies that

$$\mathbb{P}(\{|X_n|>An\}\,i.o.)=1,$$

we can now conclude that

$$\mathbb{P}\Big(\Big\{|S_n|>\frac{An}{2}\Big\}\,i.o.\Big)=1.$$

This means that for each $A>0$, there exists a set $N(A)$ with $\mathbb{P}(N(A))=0$, such that for each $\omega\in N(A)^c$ it holds that

$$\limsup_{n\geq 1}\frac{S_n(\omega)}{n}\geq\frac{A}{2}.$$

Now let $N=\cup_{m=1}^{+\infty}N(m)$, then $\mathbb{P}(N)=0$ and for $\omega\in N^c$ the previous inequality is true. Since it holds for any $A=m>0$, then the upper limit is $+\infty$. From this the proof ends. □

We end this chapter with a generalization of the previous result in the case where the mean is infinite.We leave the proof as an exercise to the interested reader.

Theorem 2.3.13 (Strong Law of Large Numbers of Feller)**.**

Let $\{X_n\}_{n\geq 1}$ *be a sequence of independent and identically distributed r.v. with* $E[|X_1|] = \infty$ *and let* $\{a_n\}_{n\geq 1}$ *be a sequence of positive real numbers such that* $\frac{a_n}{n}\uparrow$ *. Then*

$$\limsup_{n\geq 1}\frac{|S_n|}{a_n}=0 \quad a.e. \text{ or } \limsup_{n\geq 1}\frac{|S_n|}{a_n}=+\infty$$

depending on whether

$$\sum_{n\geq 1}\mathbb{P}(|X_n|\geq a_n) \quad \text{is finite or infinite.}$$

Exercise:
Prove the previous theorem.

2.4 Exercises

Exercise 1:

Let $(\mathscr{E}_n)_{n\geq 1}$ be random events on a probability space $(\Omega,\mathscr{F},P)$. Show that

$$\mathbb{P}(\mathscr{E}_n)\xrightarrow[n\to+\infty]{}0 \Leftrightarrow \mathbf{1}_{\mathscr{E}_n}\xrightarrow[n\to+\infty]{}0, \quad \text{in probability.}$$

Exercise 2:

Let $(X_n)_{n\geq 1}$ be a sequence of random variables.
Show that if $\mathbb{E}(X_n)\xrightarrow[n\to+\infty]{}\alpha$ and $Var(X_n)\xrightarrow[n\to+\infty]{}0$, then $X_n\xrightarrow[n\to+\infty]{}\alpha$, in probability.

Exercise 3:

(a) Let $(X_n)_{n\geq 1}$ be a sequence of random variables such that for each $n \geq 1$ it holds that

$$\mathbb{P}(X_n = 1) = \frac{1}{n} \quad \text{and} \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n}.$$

Show that

$$X_n \xrightarrow[n\to+\infty]{} 0, \quad \text{in probability.}$$

(b) Now suppose that for each $n \geq 1$ we have that $\mathbb{P}(X_n = 1) = p_n$ and $\mathbb{P}(X_n = 0) = 1 - p_n$, and suppose that $(X_n)_{n\geq 1}$ are independent. Show that:

$$\begin{aligned}
&(1) \quad X_n \xrightarrow[n\to+\infty]{} 0, \quad \text{in probability} \Leftrightarrow p_n \xrightarrow[n\to+\infty]{} 0.\\
&(2) \quad X_n \xrightarrow[n\to+\infty]{} 0, \quad \text{in } \mathbb{L}^p \Leftrightarrow p_n \xrightarrow[n\to+\infty]{} 0.\\
&(3) \quad X_n \xrightarrow[n\to+\infty]{} 0, \quad \text{almost everywhere} \Leftrightarrow \sum_{n\geq 1} p_n < +\infty.
\end{aligned}$$

(c) Justify if in (a) the sequence $(X_n)_{n\geq 1}$ converges almost everywhere to 0.

Exercise 4:

Prove the Tchebychev's weak law:

Let $(X_n)_{n\geq 1}$ be a sequence of random variables pairwise independent, with finite variance and uniformly bounded, i.e. there exists a constant $c < +\infty$ such that $Var(X_n) \leq c$ for all $n \geq 1$. Then,

$$\frac{S_n - \mathbb{E}(S_n)}{n} \to_{n\to+\infty} 0, \quad \text{in probability,}$$

where $S_n = \sum_{j=1}^{n} X_j$ is the sequence of the partial sums of $(X_n)_{n\geq 1}$.

Exercise 5:

Prove the Bernoulli's Law of Large Numbers:

Consider a sequence of independent Binomial experiments, with the same probability p of success in each experiment. Let S_n be the number of successes in the first n experiments. Then,

$$\frac{S_n}{n} \to_{n\to+\infty} p, \quad \text{in probability.}$$

Exercise 6:

Consider a sequence of independent Binomial experiments with probability p_n of success in the n-th trial. For $n \geq 1$, let $X_n = 1$ if the n-trial is a success, and $X_n = 0$ otherwise. Show that

(a) If $\sum_{n\geq 1} p_n = +\infty$, then $\mathbb{P}(\sum_{n\geq 1} X_n = +\infty) = 1$, (there are an infinite number of successes a.e.).

(b) If $\sum_{n\geq 1} p_n < +\infty$, then $\mathbb{P}(\sum_{n\geq 1} X_n < \infty) = 1$, (there are a finite number of successes a.e.).

Exercise 7:

Let $(X_n)_{n\geq 1}$ be a sequence of independent random variables such that for each $n \geq 1$ it holds that

$$\mathbb{P}(X_n = e^n) = \frac{1}{n+1} \quad \text{and} \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n+1}.$$

Analyze the convergence of $(X_n)_{n\geq 1}$ to $X = 0$ in the case of

(a) convergence in probability.

(b) convergence in $\mathbb{L}^p$, for $p > 0$.

(c) convergence almost everywhere.

(d) convergence in distribution.

Exercise 8:

Let $(X_n)_{n\geq 1}$ be a sequence of independent random variables such that for each $n \geq 1$ it holds that

$$\mathbb{P}(X_n = 1) = \frac{1}{2^n} \quad \text{and} \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{2^n}.$$

Show that $X_n \xrightarrow[n\to+\infty]{} 0$,

(a) in probability.

(b) in $\mathbb{L}^p$, for $p > 0$.

(c) almost everywhere.

(d) in distribution.

Exercise 9:

Let X and Y be random variables defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. The covariance between X and Y is defined by

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

Let $X_1, \cdots, X_n$ be uncorrelated random variables, i.e. such that $Cov(X_i, X_j) = 0$, for $i \neq j$, with $\mathbb{E}[X_i] = \mu$ and $Var(X_i) \leq C < +\infty$, for all $i \geq 1$, where C is a constant. If $S_n := X_1 + \cdots + X_n$, show that

(a) $\mathbb{E}[S_n] = n\mu$ and $Cov(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$.

(b) $Var(S_n) = Var(X_1) + \cdots + Var(X_n)$.

(c) $\frac{S_n}{n} \xrightarrow[n \to +\infty]{} \mu$, in $\mathbb{L}^2$ and in probability.

Exercise 10:

Let $(X_n)_{n\geq 2}$be a sequence of independent and identically distributed random variables such that X_1 has exponential distribution with parameter 1. For each $n \geq 2$ let $Y_n = X_n / \log(n)$. Analyze the convergence of $(Y_n)_{n\geq 2}$ to $Y = 0$ in the case of

(a) convergence in probability.

(b) convergence in $\mathbb{L}^1$.

(c) convergence almost everywhere.

(d) convergence in distribution.

Exercise 11:

Let $X_1, X_2, X_3 \ldots$ be independent random variables with $X_n \sim \mathscr{U}[0, a_n]$, with $a_n > 0$. Show that

(a) If $a_n = n^2$, then, with probability 1, only a finite number of X_n takes values less than 1.

(b) If $a_n = n$, then, with probability 1, an infinite number of X_n takes values less than 1.

Exercise 12:

Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. random variables such that $X_1 \sim \mathscr{U}[0,1]$. Show that n^{-X_n} converges to 0 in probability but it does not converge to 0 almost surely.

Exercise 13:

Let $(X_n)_{n\geq 1}$ be a sequence of random variables such that for $n \in \mathbb{N}$ it holds that

$$\mathbb{P}(X_n = n^2) = \frac{1}{n}^2 \quad \text{and} \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n}^2.$$

Show that X_n converges almost surely (find the limit X) but $\mathbb{E}[X_n^m]$ does not converge to $\mathbb{E}[X^m]$, for all $m \in \mathbb{N}$.

Exercise 14:

Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. random variables such that $X_1 \sim \mathscr{U}[0,1]$. Find the limit in probability of $\left(\prod_{k=1}^n X_k\right)^{1/n}$.

Exercise 15:

Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. random variables such that $\mathbb{E}[X_1] = 1$ and $Var(X_1) = 1$. Show that

$$\frac{\sum_{k=1}^n X_k}{\sqrt{n\sum_{k=1}^n X_k^2}} \to_{n\to+\infty} \frac{1}{\sqrt{2}}$$

in probability.

Exercise 16:

Let $(X_n)_{n\geq 1}$ be a sequence of independent random variables such that $\mathbb{E}[X_n] = 0$ and $\mathbb{E}[X_n^2] = 1$ for all $n \in \mathbb{N}$. Let $S_n := X_1 + \cdots + X_n$ and for all $x \in \mathbb{R}$ let $\varphi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$. If $\mathbb{P}(S_n \leq \sqrt{n}x) \to \varphi(x)$ for all $x \in \mathbb{R}$, show that $\limsup_{n\to+\infty} \frac{S_n}{\sqrt{n}} = +\infty$ almost everywhere.

Exercise 17:

Show that if X_n converges to X in probability, as $n \to +\infty$, and if $g : \mathbb{R} \to \mathbb{R}$ is a continuous function, then $g(X_n)$ converges to $g(X)$ in probability, as $n \to +\infty$.

Exercise 18:

Let $(X_n)_{n\geq 1}$ be a sequence of independent random variables with distribution function F_n. Prove that, $\mathbb{P}(\lim_n X_n = 0) = 1$ if and only if $\forall \varepsilon > 0$,

$$\sum_{n\geq 1}\{1 - F_n(\varepsilon) + F_n(-\varepsilon^-)\} < +\infty.$$

Exercise 19:

If $\sum_{n\geq 1}\mathbb{P}(|X_n| > n) < \infty$, then $\limsup_n \frac{|X_n|}{n} \leq 1$ almost everywhere.

Exercise 20:

(a) Let X and Y be independent random variables with laws $X \sim$Poisson(λ_1) and $Y \sim$Poisson(λ_2). What is the law of $X+Y$?

(b) Let Z be a random variable with law Poisson(λ), and let $\xi_1, \xi_2, \dots$ be i.i.d. Bernoulli(p) random variables, independent of Z. Define $X := \sum_{j=1}^{Z} \xi_i$. Show that X has law Poisson$(p\lambda)$.

Remark: Item (b) is know as the *Poisson coloring theorem*. You can think you have a Poisson number of balls, and color each ball either red (with probability p) or blue (with probability $1-p$), independently. Then the number of red balls is also Poisson distributed. This is one of the basic results in the theory of *Poisson Point Process*.

Exercise 21:

(a) Let X be a random variable with law Exp(λ), and let $t, s > 0$. Prove that

$$\mathbb{P}(X > t + s | X > s) = P(X > t).$$

This property is called "lack of memory of the exponential distribution".

(b) Let Y_n be a geometric random variable with success probability $\frac{\lambda}{n}$ (assume n large enough, so that $\frac{\lambda}{n} < 1$). Show that $\frac{Y_n}{n}$ converges weakly to an Exp(λ) distribution.

Chapter 3

Characteristic functions

3.1 Definitions and properties

In this chapter we introduce the notion of characteristic functions which is going to be a very useful tool in order to prove weak convergence results.

Definition 3.1.1. *For any r.v. X with probability measure μ and distribution function F, the characteristic function of X is defined as the function $\varphi : \mathbb{R} \to \mathbb{C}$ given by*

$$\varphi(t) = \mathbb{E}[e^{itX}] = \int_\Omega e^{itX} d\mathbb{P} = \int_\mathbb{R} e^{itx}\mu(dx) = \int_\mathbb{R} e^{itx} dF(x).$$

Note that before we have discussed the notion of the integrals above when the r.v. involved are real-valued and here we need it for a complex valued function. Then, we observe that the real and imaginary parts of φ_X are given, respectively by

$$Re\varphi(t) = \int_\mathbb{R} \cos(tx)\mu(dx) \quad \text{and} \quad Im\varphi(t) = \int_\mathbb{R} \sin(tx)\mu(dx),$$

so that the definitions make sense.

Now we enumerate some properties of the characteristic function:

- $\forall t \in \mathbb{R}$: $|\varphi(t)| \le 1 = \varphi(0)$.

To prove this item note that

$$|\varphi(t)|\sqrt{(\mathbb{E}[\cos(tX)])^2+(\mathbb{E}[\sin(tX)])^2}=F(\mathbb{E}[\cos(tX)],\mathbb{E}[\sin(tX)]),$$

where $F(a,b)=\sqrt{a^2+b^2}$. Note that the function F is convex, so that by Jensen's inequality we conclude that

$$F(\mathbb{E}[\cos(tX)],\mathbb{E}[\sin(tX)])\leq\mathbb{E}[F(\cos(tX),\sin(tX))],$$

from where the inequality follows.

- $\forall t\in\mathbb{R}$: $\varphi(t)=\overline{\varphi(-t)}$.

- φ is uniformly continuous.

  Let $h>0$. Then

$$\begin{aligned}\Big|\varphi(t+h)-\varphi(t)\Big| &= \Big|\int_{\mathbb{R}}(e^{i(t+h)x}-e^{itx})\mu(dx)\Big| = \Big|\int_{\mathbb{R}}(e^{ihx}-1)e^{itx}\mu(dx)\Big|\\ &\leq \int_{\mathbb{R}}|e^{ihx}-1|\mu(dx)\end{aligned}$$

  Now note that $|e^{ihx}-1|\leq 2$ so that $\int_{\mathbb{R}}|e^{ihx}-1|\mu(dx)\leq 2$ and since $\lim_{h\to 0}e^{ihx}=1$, we conclude, from the Dominated Convergence Theorem, that $\int_{\mathbb{R}}|e^{ihx}-1|\mu(dx)$ vanishes as $h\to 0$. Note that there is not dependence on t, so the the convergence is uniform.

- If φ_X is the c.f. for a r.v. X, then

$$\varphi_{aX+b}(t)=\varphi_X(at)e^{itb} \quad\text{and}\quad \varphi_{-X}(t)=\overline{\varphi_X(t)}.$$

- If $\{\varphi_n\}_{n\geq 1}$ is a sequence of characteristic functions, $\lambda_n\geq 0$ with $\sum_{n\geq 1}\lambda_n=1$, then $\sum_{n\geq 1}\lambda_n\varphi_n$ is a characteristic function. For each $n\geq 1$, let μ_n be the probability measure corresponding to φ_n. Then, observe (prove it!) $\sum_{n\geq 1}\lambda_n\mu_n$ is again a probability measure. Therefore , defining $\psi(t)=\int_{\mathbb{R}}e^{itx}\sum_{n\geq 1}\lambda_n\mu_n(dx)$ then, a simple computations shows that $\psi(t)$ is equal to $\sum_{n\geq 1}\lambda_n\varphi_n(t)$

- If $\{\varphi_n\}_{n\geq 1}$ is a sequence of characteristic functions, then $\prod_{j=1}^n \varphi_j$ is a characteristic function.

  We know that given probability measures $\{\mu_j\}_{j=1,\cdots,n}$ where μ_j is corresponding to φ_j, then there exist independent r.v. $\{X_j\}$ all defined in the same probability space $\Omega, \mathscr{F}, \mathbb{P})$ whose induced measure is μ_j. Then for $S_n = \sum_{j=1}^n X_j$ we have that

$$\mathbb{E}[e^{itS_n}] = \prod_{j=1}^n \mathbb{E}[e^{itX_j}] = \prod_{j=1}^n \varphi_j(t).$$

Exercise:
Do the missing proofs of the properties above.

Let $S_n = X_1 + \cdots + X_n$, where X_j are independent. Then, from the previous property we know that $\varphi_{S_n}(t) = \prod_{j=1}^n \varphi_{X_j}(t)$. But, what can we say about the distribution of S_n? Let us now go for a small digression on the convolution.

Definition 3.1.2 (Convolution of distribution functions)**.**

The convolution of two distribution functions F_1 and F_2 is the distribution function F defined on $x \in \mathbb{R}$ as $F(x) = \int_{\mathbb{R}} F_1(x-y)dF_2(y)$. In this case we use the notation $F = F_1 * F_2$.

Exercise:
Check that F given above is in fact a distribution function.

Theorem 3.1.3. *Let X and Y be two independent r.v. with distribution functions F_X and F_Y respectively. Then $X+Y$ has distribution function $F_X * F_Y$.*

Proof. Note that we want to prove that for $x \in \mathbb{R}$ we have that

$$\mathbb{P}(X+Y \leq z) = F_X * F_Y(z).$$

Let $f(x,y) = \mathbf{1}_{\{x+y\leq z\}}$ and note that f is $\mathscr{B}^2$-measurable. Then

$$\int_\Omega f(X,Y)d\mathbb{P} = \iint_{\mathbb{R}^2} f(x,y)\mu^2(dx,dy),$$

where μ^2 is the measure induced by the r.v. (X,Y) and since the r.v. X and Y are independent we know that $\mu^2 = \mu_X \times \mu_Y$, that is μ^2 is the product measure between μ_x and μ_Y. Then, by Fubini's theorem, the previous integral equals to

$$\begin{aligned}\int_{\mathbb{R}}\int_{\mathbb{R}} f(x,y)\mu_X(dx)\mu_Y(dy) &= \int_{\mathbb{R}} \mu_X((-\infty, z-y])\mu_Y(dy)\\ &= \int_{\mathbb{R}} F_X(z-y)\mu_Y(dy)\\ &= \int_{\mathbb{R}} F_X(z-y)F_Y(dy) = F_X * F_Y(z).\end{aligned}$$

□

Definition 3.1.4 (Convolution of density functions).

The convolution of two probability density functions f_1 and f_2 is the probability density function f defined on $x \in \mathbb{R}$ as $f(x) = \int_{\mathbb{R}} f_1(x-y)f_2(y)dy$. In this case we also use the notation $f = f_1 * f_2$.

Exercise:
Check that f given above is in fact a density function.

Theorem 3.1.5. *The convolution of two absolutely continuous distribution functions F_1 and F_2 with densities f_1 and f_2, is absolutely continuous with density $f = f_1 * f_2$.*

Proof. Let $p = f_1 * f_2$, which we know to be a density from the previous exercise. Then

$$\begin{aligned}\int_{-\infty}^{x} p(y)dy = \int_{-\infty}^{x} f_1 * f_2(y)dy &= \int_{-\infty}^{x}\int_{\mathbb{R}} f_1(y-z)f_2(z)\,dz\,dy\\ &= \int_{\mathbb{R}}\int_{-\infty}^{x} F_1(x-z)f_2(z)\,dz\\ &= \int_{\mathbb{R}}\int_{-\infty}^{x} F_1(x-z)F_2(dz) = F_1 * F_2(x).\end{aligned}$$

Then $p = f$ is the density of $F_1 * F_2$. □

And what can we say about the probability measure corresponding to $F_1 * F_2$? We shall denote this measure by $\mu_1 * \mu_2$. We introduce the notation

$$A \pm B = \{x \pm y : x \in A, y \in B\},$$

for A and B subsets of $\mathbb{R}$.

Theorem 3.1.6. *For each $B \in \mathscr{B}$ we have that*

$$(\mu_1 * \mu_2)(B) = \int_{\mathbb{R}} \mu_1(B - y)\mu_2(dy).$$

Moreover, for each $\mathscr{B}$- measurable funtion g integrable with respect to $\mu_1 * \mu_2$, we have that

$$\int_{\mathbb{R}} g(u)\mu_1 * \mu_2(du) = \iint_{R^2} g(x+y)\mu_1(dx)\mu_2(dy).$$

Proof. First note that $\mu_1 * \mu_2$ is a probability measure, we leave this as an exercise to the reader. To show that the corresponding distribution function is $F_1 * F_2$ we have to compute $\mu_1 * \mu_2((-\infty, x])$ and show that it coincides with

$$F(x) = \int_{\mathbb{R}} F_1(x-y)dF_2(y).$$

Now,

$$\begin{aligned}\mu_1 * \mu_2((-\infty, x]) &= \int_{\mathbb{R}} \mu_1((-\infty, x] - y)\mu_2(dy) = \int_{\mathbb{R}} \mu_1((-\infty, x-y])\mu_2(dy) \\ &= \int_{\mathbb{R}} F_1(x-y)F_2(dy) = F_1 * F_2(x).\end{aligned}$$

This shows the first affirmation. Now we prove the second one. Let $g = \mathbf{1}_B$. Then, for each y, we have that $g_y(x) = g(x+y) = \mathbf{1}_{\{B-y\}}$. Now

$$\int_{\mathbb{R}} g(x+y)\mu_1(dx) = \mu_1(B-y).$$

And

$$\begin{aligned}\iint_{\mathbb{R}^2} g(x+y)\mu_1(dx)\mu_2(dy) &= \int_{\mathbb{R}} \mu_1(B-y)\mu_2(dy) = \mu_1 * \mu_2(B-y) \\ &= \int_{\mathbb{R}} g(u)\mu_1 * \mu_2(du).\end{aligned}$$

This ends the proof. □

Let us now compute the characteristic function of the convolution $\mu_1 * \mu_2$. We have that the sum of two r.v. with probability measures μ_1 and μ_2 has induced measure given by $\mu_1 * \mu_2$ and its characteristic function is given by

$$\iint e^{itu}\mu_1 * \mu_2(du) = \iint e^{ity}e^{itx}\mu_1(dx)\mu_2(dy) = \int_{\mathbb{R}} e^{itx}\mu_1(dx)\int_{\mathbb{R}} e^{ity}\mu_2(dy)$$

which is equal to $\varphi_1(t)\varphi_2(t)$.

Then we conclude the next result.

Theorem 3.1.7. *The sum of a finite number of independent r.v. corresponds to the convolution of their distribution functions and to the product of their characteristic functions.*

Lemma 3.1.8. *If φ is a characteristic function, then $|\varphi|^2$ is a characteristic function.*

Proof. We know that given a r.v. X with characteristic function φ, then there exists a r.v. Y with the same distribution of X (and therefore the same characteristic function) which is independent of X. Then the characteristic function of $X-Y$ is given by $\varphi_{X-Y}(t) = \varphi_X(t)\varphi_Y(-t) = \varphi_X(t)\varphi_X(-t) = \varphi_X(t)\overline{\varphi_X(t)} = |\varphi_X(t)|^2$. □

Example 18.

1. *$X \sim Ber(p)$ we have that $\varphi_X(t) = e^{it}p + (1-p)$.*

2. *$X \sim U[-a,a]$ we have that $\varphi_X(t) = \frac{sin(at)}{at}$, if $t \neq 0$ and $\varphi_X(0) = 1$.*

3. *$X \sim N(\mu,\sigma^2)$, we have that $\varphi_X(t) = e^{it\mu}e^{-\frac{\sigma^2 t^2}{2}}$.*

3.2 Inversion formula

The question now is: Given a characteristic function how can we find the correspondent distribution function or the distribution measure?

Theorem 3.2.1 (The characteristic function determines the distribution)**.**
If $x_1 < x_2$ then

$$\mu((x_1,x_2)) + \frac{1}{2}\mu(\{x_1\}) + \frac{1}{2}\mu(\{x_2\}) = \frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T}\frac{e^{-itx_1}-e^{itx_2}}{it}\varphi(t)\,dt.$$

**Note that the integrand function is defined by continuity at $t=0$.*

Proof. Note that

$$\int_{-T}^{T}\frac{e^{-itx_1}-e^{itx_2}}{it}\varphi(t)\,dt = \int_{-T}^{T}\int_{\mathbb{R}} e^{itx}\mu(dx)\Big[\frac{e^{-itx_1}-e^{itx_2}}{it}\Big]dt$$

Note that the function inside square brackets in the expression above is bounded since $\frac{e^{itx-1}}{it}\sim x$ when t is close to 0. Then from Fubini's Theorem, the last integral is equal to

$$\int_{\mathbb{R}}\mu(dx)\int_{-T}^{T}\Big[\frac{e^{-it(x-x_1)}-e^{it(x-x_2)}}{it}\Big]dt \tag{3.2.1}$$

Above we used Fubini's Theorem since

$$\Big|\frac{e^{-it(x-x_1)}-e^{it(x-x_2)}}{it}\Big| = \Big|\int_{x_1}^{x_2}e^{-itu}du\Big|$$

and

$$\int_{\mathbb{R}}\int_{-T}^{T}|x_2-x_1|\,dt\mu(dx)\le 2T|x_2-x_1|.$$

So the integrand is dominated by an integrable function with respect to the product measure $dt\mu(dx)$ in $[-T,T]\times\mathbb{R}$. Now note that

$$e^{-it(x-x_1)}-e^{it(x-x_2)} = \cos(t(x-x_1))-\cos(t(x-x_2))+i(\sin(t(x-x_1))-\sin(t(x-x_2)))$$

and since the integral above, with respect to t is in a symmetric domain and the function $\cos(t(x-x_1))-\cos(t(x-x_2))$ is even and $i\sin(t(x-x_1))-i\sin(t(x-x_2))$ is odd, we have that (3.2.1) is equal to

$$2\int_{\mathbb{R}}\mu(dx)\Big(\int_0^T\frac{\sin(t(x-x_1))}{t}dt-\int_0^T\frac{\sin(t(x-x_1))}{t}dt\Big).$$

By a change of variables last expression equals to

$$2\int_{\mathbb{R}} \mu(dx)\Big(\int_0^{T(x-x_1)} \frac{\sin(sx)}{s}ds - \int_0^{T(x-x_2)} \frac{\sin(sx)}{s}ds\Big).$$

Now we note that $\int_0^T \frac{\sin(s)}{s}ds \to_{T\to+\infty} \frac{\pi}{2}$. A simple way to note this is to argue as follows:

$$\int_0^{+\infty} \frac{\sin(s)}{s}ds = \int_0^{+\infty} \sin(s)\int_0^{+\infty} e^{-xu}\,du\,ds = \int_0^{+\infty}\int_0^{+\infty} e^{-su}\sin(s)\,du\,ds$$
$$= \int_0^{+\infty} \frac{1}{1+u^2}du = \frac{\pi}{2}.$$

Now we take the limit as $T \to +\infty$ in $2\Big(\int_0^{T(x-x_1)} \frac{\sin(sx)}{s}ds - \int_0^{T(x-x_2)} \frac{\sin(sx)}{s}ds\Big)$ and it equals to

1. $-2\Big(\int_{-\infty}^0 \frac{\sin(sx)}{s}ds - \int_{-\infty}^0 \frac{\sin(sx)}{s}ds\Big) = 0$, if $x < x_1 < x_2$;
2. $2\Big(\int_{-\infty}^0 \frac{\sin(sx)}{s}ds = \pi$, if $x = x_1 < x_2$;
3. 2π, if $x_1 < x < x_2$;
4. π, if $x_1 < x_2 = x$;
5. 0, if $x_1 < x_2 < x$;

From the previous equalities we obtain the result. □

Remark 3.2.2. *Note that if (x_1, x_2) is a continuity interval for μ, then the previous theorem says that*

$$F(x_2) - F(x_1) = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-itx_1} - e^{itx_2}}{it}\varphi(t)\,dt. \tag{3.2.2}$$

3.3 Uniqueness of distribution

Theorem 3.3.1. *If two probability measures (or two distribution functions) have the same characteristic function, then the probability measures (or the distribution functions) are the same.*

Proof. If x_1 and x_2 are not atoms of μ (or F) then (3.2.2) gives us the value of $\mu((x_1, x_2))$ which is determined by the characteristic function. Therefore, given μ_1 and μ_2 with the same characteristic function we have that $\mu_1((a,b)) = \mu_2((a,b))$, where a and b are not atoms of μ_1 nor μ_2. Since the set of atoms of a probability measure is at most countable, the points in $\mathbb{R}$ which are not atoms for both the measures μ_1 and μ_2 is dense in $\mathbb{R}$. Now, from Theorem 1.3.5 it follows that $\mu_1 = \mu_2$. □

Theorem 3.3.2. *If $\varphi \in L^1(\mathbb{R})$, then $F \in C^1(\mathbb{R})$ and*

$$F'(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-ixt}\varphi(t)dt,$$

that is φ is the characteristic function of an absolutely continuous r.v.

Proof. To prove the result, we apply the previous theorem for $x = x_2$ and $x_1 = x - h$ where $h > 0$. Then the theorem says that

$$\mu((x-h,x)) + \frac{1}{2}\mu(\{x\}) + \frac{1}{2}\mu(\{x-h\}) = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{e^{ith}-1}{it} e^{-itx}\varphi(t)\,dt.$$

The term on the left hand side of last equality is equal to

$$\frac{F(x)+F(x^-)}{2} - \frac{F(x-h)+F((x-h)^-)}{2}.$$

Note that since $\varphi \in L^1(\mathbb{R})$ the previous integral exists since the integrand function is bounded by $|h\varphi(t)|$. From the Dominated Convergence Theorem, we can send $h \to 0$ and we conclude that the integral is equal to 0. Therefore we obtain that

$$\frac{F(x)+F(x^-)}{2} = \lim_{h\to 0} \frac{F(x-h)+F((x-h)^-)}{2},$$

from where we conclude that F is left continuous. Since F is a distribution function, it is continuous at the right and then F is continuous. Going back we can rewrite

$$\frac{F(x)-F(x-h)}{h} = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{e^{ith}-1}{ith} e^{-itx}\varphi(t)\,dt,$$

and the limit exists when $h \to 0$, so that F has a derivative from the left at x and

$$F'(x^-) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi(t)\, dt.$$

Analogously we can show that F has a derivative at x from the right and

$$F'(x^+) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi(t)\, dt.$$

We conclude that F' exists and it is a continuous function, since the right hand side of the previous equality is continuous. Since F' is continuous we conclude that for all $x \in \mathbb{R}$ $F(x) = \int_{-\infty}^{x} F'(u)\, du$, so that F' is a probability density. □

Corollary 3.3.3. *If $\varphi \in L^1(\mathbb{R})$, then $p(x) \in L^1(\mathbb{R})$ where*

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ixt} \varphi(t) dt$$

and

$$\varphi(x) = \int_{-\infty}^{\infty} e^{itx} p(x)\, dx.$$

Exercise: do the proof of the corollary.

Theorem 3.3.4 (Atoms of μ)**.**

- *For each x_0 we have that*

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-itX_0} \varphi(t)\, dt = \mu(\{x_0\}). \tag{3.3.1}$$

- *It holds that*

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |\varphi(t)|^2\, dt = \sum_{x\in\mathbb{R}} (\mu(\{x\}))^2.$$

Proof. To prove the first affirmation we repeat the proof of Theorem 3.2.1 and we obtain that the left hand side of (3.3.1) is equal to

$$\int_{\mathbb{R}\setminus\{x_0\}} \frac{\sin(T(x-x_0))}{T(x-x_0)}\mu(dx) + \int_{\{x_0\}} \mu(dx). \tag{3.3.2}$$

The integrand function at the left hand side in the previous expression is bounded by 1 and goes to 0 when $T \to +\infty$, then by the Dominated Convergence Theorem the integral vanishes as $T \to +\infty$, from where the result follows.

To prove the second affirmation we note that since the number of atoms of μ is at most countable, all the terms (except at most a countable number of them) are equal to 0 so that the series above makes sense.

Also note that $|\varphi(t)|^2$ is a characteristic function. We have seen above that it is the characteristic function of the r.v. $X - Y$ where X and Y are i.i.d.. The distribution measure of $|\varphi(t)|^2$ is $\mu * \mu'$ where $\mu'(B) = \mu(-B)$ for each $B \in \mathscr{B}$. Applying the first affirmation with $x_0 = 0$ and with the characteristic function $|\varphi(t)|^2$ we get that

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} |\varphi(t)|^2\, dt = \mu * \mu'(\{0\}) = \int_{\mathbb{R}} \mu'(x)\mu(dx) = \sum_{x\in\mathbb{R}} \mu(\{x\})\mu(\{x\})$$

and the proof ends. Above we used the fact that the integrand is non-zero when x is an atom of μ. □

Corollary 3.3.5. *μ is atomless (F is continuous) iff*

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} |\varphi(t)|^2\, dt = 0.$$

Definition 3.3.6 (Symmetric random variable)**.**

We say that a r.v. X is symmetric around 0 iff X and $-X$ have the same distribution.

Remark 3.3.7. *For a symmetric r.v. its distribution μ has the following property $\mu(B) = \mu(-B)$ for any $B \in \mathscr{B}$. Such probability measure is said to be symmetric around 0. Equivalently, for the distribution function, we have that for any $x \in \mathbb{R}$, $F(x) = 1 - F(x^-)$.*

Theorem 3.3.8. *A r.v. X or a p.m. μ is symmetric iff its characteristic function is real-valued (for all t.)*

Proof. If X and $-X$ have the same distribution, then they determine the same characteristic function. Therefore, $\varphi_X(t) = \varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)}$. Reciprocally, if φ_X is real, then from the previous equalities we conclude that $\varphi_X(t) = \varphi_{-X}(t)$. From (3.3.1) we conclude that X and $-X$ have the same distribution and therefore, X is symmetric. □

Theorem 3.3.9 (Lévy's converging Theorem)**.**

Let $\{\mu_n\}_{n\geq 1}$ be probability measures on $\mathbb{R}$ with characteristic function $\{\varphi_n\}_{n\geq 1}$.

- *If μ_∞ is a probability measure on $\mathbb{R}$ and $\mu_n \to^v \mu_\infty$, then $\varphi_n(t) \to_{n\to\infty} \varphi_\infty(t)$, where φ_∞ is the characteristic function of μ_∞.*
- *If $\varphi_n(t) \to_{n\to\infty} \varphi_\infty(t)$ for all $t \in \mathbb{R}$, and $\varphi_\infty(t)$ is continuous at $t = 0$, then*
  - *$\mu_n \to^v \mu_\infty$ where μ_∞ is a probability measure,*
  - *φ_∞ is a characteristic function of μ_∞.*

Proof. Let us prove the first affirmation. Note that

$$\varphi_n(t) = \mathbb{E}[e^{itX_n}] = \mathbb{E}[\cos(tX_n)] + i\mathbb{E}[\sin(tX_n)].$$

From Theorem 2.2.9 which in fact holds if we take functions in C_B (prove it!) and since the functions $\sin(\cdot)$ and $\cos(\cdot)$ are continuous and bounded, we have that

$$\lim_{n\to+\infty} \mathbb{E}[\cos(tX_n)] = \mathbb{E}[\cos(tX)], \quad \lim_{n\to+\infty} \mathbb{E}[\sin(tX_n)] = \mathbb{E}[\sin(tX)]$$

and we are done.

Now let us suppose that $\varphi_n(t) \to_{n\to+\infty} \varphi_\infty(t)$ for all $t \in \mathbb{R}$. Fix $\varepsilon > 0$. Since φ_∞ is a continuous function we know that there exists an $a > 0$ such that $\frac{1}{a}\int_{-a}^{a}(1-\varphi(t))dt \leq \varepsilon$. Then, since $|\varphi_n(t)| \leq 1$ for all $n \geq 1$, by the Dominated Convergence Theorem, we have that

$$\lim_{n\to+\infty} \int_{-a}^{a}(1-\varphi_n(t))dt = \int_{-a}^{a} \lim_{n\to+\infty}(1-\varphi_n(t))dt = \int_{-a}^{a}(1-\varphi(t))dt \leq \varepsilon.$$

Therefore, there exists an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ it holds that

$$\frac{1}{a}\int_{-a}^{a}(1-\varphi_n(t))dt \leq \varepsilon.$$

From the next lemma, we conclude that

$$\mu_n\Big(\Big[-\frac{2}{a},\frac{2}{a}\Big]^c\Big) \leq \varepsilon$$

for all $n \geq n_0$. Since for the values of $n = 1, \cdots, n_0 - 1$ the measure μ_n is a probability measure, we also conclude that

$$\mu_n\Big(\Big[-\frac{2}{a},\frac{2}{a}\Big]^c\Big) \leq \varepsilon$$

for all $n \geq 1$ by changing the interval if necessary. Then the sequence $\{\mu_n\}_{n\geq 1}$ is tight, that is, any subsequence $\{\mu_{n_k}\}_{k\geq 1}$ has a converging subsequence. To show that the whole sequence converges we need to show that the limit point is a probability measure. Let us suppose that $\{\mu_{n_k}\}_{k\geq 1}$ converges weakly to μ_∞ as $k \to +\infty$. The previous measure μ_∞ is a subprobability measure. We will show that it is a probability measure. Note that, for δ such that $-2/\delta, 2/\delta$ are not atoms of μ_∞, then $\mu_\infty(\mathbb{R}) \geq \mu_\infty([-2/\delta, 2/\delta]) = \lim_{n\to+\infty} \mu_n([-2/\delta, 2/\delta]) \geq 1-\varepsilon$. Since ε is arbitrary we conclude that μ_∞ is a probability measure. Now let φ be the characteristic function of μ_∞. Now, from the first part of the theorem we know that $\varphi_{n_k}(t) \to_{k\to+\infty} \varphi_\infty(t)$, from where it follows that every weak limit of μ_{n_k} has characteristic function φ_∞. Them from the uniqueness theorem it follows that $\mu_{n_k} \to_{k\to+\infty} \mu_\infty$ and since all subsequence converges weakly to the same measure, we are done. □

Lemma 3.3.10. *For each $a > 0$ if holds that*

$$\mu\Big(\Big[-\frac{2}{a},\frac{2}{a}\Big]^c\Big) \leq \frac{1}{a}\int_{-a}^{a}(1-\varphi(t))dt.$$

Proof. Note that

$$\frac{1}{a}\int_{-a}^{a}(1-\varphi(t))dt = \frac{1}{a}\int_{-a}^{a}\Big(1-\int_{\mathbb{R}} e^{itx}d\mu\Big)dt = \frac{1}{a}\int_{-a}^{a}\int_{\mathbb{R}}(1-e^{itx})d\mu\, dt. \tag{3.3.3}$$

From Fubini's theorem last integral writes as

$$\frac{1}{a}\int_{\mathbb{R}}\int_{-a}^{a}(1-\cos(tx)-i\sin(tx))dt\,d\mu = \frac{1}{a}\int_{\mathbb{R}}\int_{-a}^{a}(1-\cos(tx))dt\,d\mu. \tag{3.3.4}$$

Above we used the fact that the $\sin(x)$ is an odd function and the domain of integration is symmetric. Computing the time integral above, last expression equals to

$$2\int_{\mathbb{R}}\Big(1-\frac{\sin(ax)}{x}\Big)d\mu \geq 2\int_{|x|\geq 2/a}\Big(1-\frac{\sin(ax)}{x}\Big)d\mu. \tag{3.3.5}$$

Since $|ax| \geq 2$ then $\sin(ax) \leq ax$ and from this we bound from below the previous expression by

$$2\int_{|x|\geq 2/a}\frac{1}{2}d\mu \geq \mu\Big(\Big[-\frac{2}{a},\frac{2}{a}\Big]^c\Big). \tag{3.3.6}$$

□

Corollary 3.3.11. *If $\{\mu_n\}\_n \geq 1$ and μ are probability measures with characteristic functions $\{\varphi_n\}_{n\geq 1}$ and φ, then $\mu_n \to^{v} \mu_\infty$ iff $\varphi_n(t) \to_{n\to\infty} \varphi(t)$, for all $t \in \mathbb{R}$.*

Example 19. *Exercises:*

1) *Take μ_n which gives mass $1/2$ to 0 and to n and analyze it.*

2) *Take μ_n as Uniform in $[-n,n]$ and analyze it.*

Theorem 3.3.12. *If F has finite absolute moment of order k, with $k \geq 1$, then φ has a continuous k-th derivative which is given by:*

$$\varphi^k(t) = \int_{\mathbb{R}}(ix)^k e^{itx}dF(x).$$

Proof. We do the proof for $k = 1$. Note that

$$\frac{\varphi(t+h)-\varphi(t)}{h} = \int_{\mathbb{R}}\frac{e^{i(t+h)x}-e^{itx}}{h}dF(x) = \int_{\mathbb{R}}e^{itx}\Big(\frac{e^{ihx}-1}{h}\Big)dF(x)$$

Since

$$\frac{e^{ihx}-1}{h} \to_{h\to 0} ix$$

and since

$$\left|\frac{e^{ihx}-1}{h}\right| \le |x|,$$

and by the hypothesis of the theorem we can use the Dominated Convergence Theorem to conclude that

$$\varphi'(t) = \lim_{h\to 0}\frac{\varphi(t+h)-\varphi(t)}{h} = \lim_{h\to 0}\int_{\mathbb{R}} e^{itx}\Big(\frac{e^{ihx}-1}{h}\Big)dF(x) = \int_{\mathbb{R}} ixe^{itx}dF(x).$$

Now the proof goes by induction. We leave this exercise to the reader. □

Theorem 3.3.13. *If F has finite absolute moment of order k, with $k \ge 1$, then φ has the following expansion around a neighbourhood of $t = 0$:*

$$\varphi(t) = \sum_{j=0}^{k}\frac{i^j}{j!}m^j t^j + o(|t|^k)$$

$$\varphi(t) = \sum_{j=0}^{k-1}\frac{i^j}{j!}m^j t^j + \frac{\theta_k}{k!}\mu^k|t|^k,$$

where m^j is the moment of order j, μ^k is the absolute moment of order k and $\theta_k \le 1$.

Exercise: do the proof of the result above.

In what follows $\{X_n\}_{n\ge 1}$ is a sequence of i.i.d.r.v. with distribution function F and $S_n = \sum_{j=1}^n X_j$. Now we are going to reprove the weak law of large numbers by using the powerful tool of the characteristic function.

Theorem 3.3.14 (The weak law of large numbers)**.**

If F has finite mean $m < \infty$, then $\frac{S_n}{n} \to m$ in probability.

Proof. Note that since the limit is a constant m, then the convergence in probability is equivalent to the convergence in distribution. Therefore, from Levy's converging theorem we just have to show that the corresponding characteristic functions converge, that is

$$\varphi_{\frac{S_n}{n}}(t) \to_{n\to+\infty} e^{itm}. \tag{3.3.7}$$

Note that e^{itm} is the characteristic function corresponding to the r.v. $X = m$ or the measure $\mu = \delta_m(\cdot)$, the Dirac supported on the set $\{m\}$. But, by the i.i.d. hypothesis we have that

$$\varphi_{\frac{S_n}{n}}(t) = (\varphi(\tfrac{t}{n}))^n \tag{3.3.8}$$

and from the previous theorem the last expression equals to

$$\Big(1 + \frac{itm}{n} + o(|\tfrac{t}{n}|)\Big)^n$$

and by the next lemma with $c_n = itm + o(\frac{t}{n})n$, last expression converges, as $n \to +\infty$, to e^{itm} and we are done. □

Lemma 3.3.15. *If $\{c_n\}_{n\geq 1}$ is a sequence of complex numbers with*

$$\lim_{n\to+\infty} c_n = c \in \mathbb{C},$$

then

$$\lim_{n\to+\infty} \Big(1 + \frac{c_n}{n}\Big)^n = e^c$$

Theorem 3.3.16 (The central limit theorem)**.**

If F has finite mean $m < \infty$ and variance σ^2 such that $0 < \sigma^2 < +\infty$, then

$$\frac{S_n - mn}{\sigma\sqrt{n}} \to \Phi$$

in distribution, where Φ is the distribution function of $\mathcal{N}(0,1)$.

Proof. Let us suppose that $m = 0$ and at the end we can simply consider $Y_j = X_j - m$. From Levy's converging theorem it is enough to show the convergence of the corresponding characteristic functions. Note that

$$\varphi_{\frac{S_n}{\sigma\sqrt{n}}}(t) = (\varphi(\tfrac{t}{\sigma\sqrt{n}}))^n \tag{3.3.9}$$

and from Theorem 3.3.13 last expression equals to

$$\Big(1+\frac{i^2\sigma^2}{2}\Big(\frac{t}{\sigma\sqrt{n}}\Big)^2+o\Big(\Big|\frac{t}{\sigma\sqrt{n}}\Big|^2\Big)\Big)^n$$

and by the previous lemma with $c_n=-\frac{t^2}{2}+o(\frac{t}{\sigma\sqrt{n}})n$, last expression converges, as $n\to+\infty$, to $e^{-\frac{t^2}{2}}$ and we are done since $e^{-\frac{t^2}{2}}$ corresponds to the characteristic function of $\mathscr{N}(0,1)$ and by the uniqueness theorem. □

3.4 Exercises

Exercise 1:

Compute the characteristic function of each one of the following random variables:

(a) X such that $\mathbb{P}(X=a)=1$ and $\mathbb{P}(X\neq a)=0$.

(b) X such that $\mathbb{P}(X=1)=1/2$ and $\mathbb{P}(X=-1)=1/2$.

(c) X with Bernoulli distribution with parameter p.

(d) X with Binomial distribution with parameter n and p.

(e) X with Geometric distribution with parameter p.

(f) X with Poisson distribution with parameter λ.

(g) X with exponential distribution with parameter λ.

(h) X with uniform distribution on $[-a,a]$, with $a>0$.

(i) X with triangular distribution on $[-a,a]$, with $a>0$.

(j) X with Gaussian distribution with mean μ and variance σ^2.

Exercise 2:

(a) Show that for X and Y independent random variables it holds that $\varphi_{X+Y}=\varphi_X\varphi_Y$.

(b) Show that if φ is a characteristic function, then $|\varphi|^2$ is also a characteristic function.

Exercise 3:

Let φ be a characteristic function. Show that $\psi(t) = e^{\lambda(\varphi(t)-1)}$ with $\lambda > 0$ is also a characteristic function.
Suggestion: Let $N, X_1, X_2, \cdots$ be independent random variables with $N \sim$Poisson(λ) and $(X_n)_{n\geq 1}$ identically distributed with $\varphi_{X_n} = \varphi$ for all $n \geq 1$. Let $Y := S_N$, with $S_n = X_1 + \cdots + X_n$. Then $\varphi_Y = \psi$.

Exercise 4:

Let φ_X be a characteristic function of a random variable X with Binomial distribution with parameter n and p. Find φ_X and $\mathbb{E}[X]$ and verify that $i^{-1}\varphi_X'(0) = \mathbb{E}[X] = np$.

Exercise 5:

Let $(X_n)_{n\geq 1}$ be a sequence of random variables with Uniform distribution $\mathscr{U}[-n, n]$. Find φ such that

$$\varphi_n(t) \xrightarrow[n\to+\infty]{} \varphi(t),$$

for all $t \in \mathbb{R}$ where for each $n \geq 1$, φ_n is the characteristic function of X_n. Verify if φ is a characteristic function.

Exercise 6:

(a) Show that if $Y := aX + b$ for $a, b \in \mathbb{R}$ and $a \neq 0$ then $\varphi_Y(t) := e^{itb}\varphi_X(at)$.

(b) Is $\varphi(t) := \mathbf{1}_{[0,\infty)}(t)$ a characteristic function? Justify.

(c) Is $\varphi(t) := t\mathbf{1}_{[0,1]}(t) + \mathbf{1}_{[1,\infty)}(t)$ a characteristic function? Justify.

(d) Show that X is a symmetric if and only if its characteristic function φ_X, takes values in $\mathbb{R}$.

(e) Let $\varphi(t) = \frac{1+\cos(3t)}{2}$. Find X such that φ is its characteristic function.

Exercise 7:

(a) Using characteristic functions show that if X and Y are independent and identically distributed random variables and if $X \sim \mathcal{N}(0,1)$ then $X + Y \sim \mathcal{N}(0,2)$.

(b) Obtain the previous result using convolutions. Justify.

(c) Compute the 3-rd centered moment of the random variable $X+Y$, i.e. compute $\mathbb{E}[(X+Y)^3]$. Suggestion: use characteristic functions.

(d) Let $X_1, \cdots, X_n$ be independent and identically distributed random variables such that $X_1 \sim \mathcal{N}(0,1)$. Using characteristic functions, show that

$$\frac{S_n}{n} \xrightarrow[n\to+\infty]{} 0,$$

in **probability**, where $S_n := X_1 + \cdots + X_n$.

Exercise 8:

Let $X_1, \cdots, X_n$ be independent random variables with Poisson distribution with parameter $\lambda_1, \cdots, \lambda_n$, respectively, where $\lambda_i > 0$, for all $i \geq 1$.

(a) Verify that $\mathbb{E}[X_1] = \lambda_1$.

(b) Compute the characteristic function φ_{X_1} of X_1.

(c) Verify that $d_t \log(\varphi_{X_1}(t)) = \lambda_1 i e^{it}$ and conclude that $i^{-1}\varphi'_{X_1}(0) = \mathbb{E}[X_1]$.

(d) Compute the characteristic function of $S_n = X_1 + \cdots + X_n$.

Exercise 9:

(a) Let X be a constant random variable and let φ_X be its characteristic function.

Show that $|\varphi_X(t)|^2 = 1$ for all $t \in \mathbb{R}$.

(b) Let X be a random variable independent of itself. Show that X is constant a.e.

(c) Let X be a symmetric random variable that takes only two values θ and $-\theta$, with $\theta > 0$. Show that there is no $\theta \in \mathbb{R}$ such that $\varphi_X(t) = 1$ for all $t \in \mathbb{R}$ where φ_X denotes the characteristic function of X. Show that $\varphi''_X(0) = -\theta^2$. Conclude that $Var(X) = \theta^2$.

Exercise 10:

Find the distribution of the random variable $X+Y+Z$, knowing that X,Y and Z are independent and identically distributed random variables and such that X has Bernoulli distribution with parameter p, i.e. X induces the measure $\mu_X := p\delta_{\{1\}} + (1-p)\delta_{\{0\}}$.
Solve the exercise in two different ways: using the convolution and characteristic functions.

Exercise 11:

(a) Let X be a symmetric random variable that takes the values $a \neq b \neq c$. Knowing that $\mathbb{P}(X=0) = 1/5$, compute φ_X i.e. the characteristic function of X.

(b) Verify that there is no $a \in \mathbb{R}$ such that $\varphi_X(t) = 1$ for all $t \in \mathbb{R}$.

(c) Compute $\varphi'_X(t)$ and verify that $i^{-1}\varphi'_X(0) = \mathbb{E}[X]$.

(d) Find a such that $\varphi''_X(0) = -1$. Conclude that $Var(X) = 1$.

Exercise 12:

Justify if $\varphi(t) := \frac{e^{ita}+1}{2}$ is the characteristic functions of a symmetric random variable?

Find the random variable whose characteristic function is φ.

Exercise 13:

Find the distribution of the random variable $X+Y$, knowing that X has Poisson distribution of parameter λ_1 and Y is independent of X and has Poisson distribution of parameter λ_2. Solve in two different ways: using the convolution and characteristic functions.

Exercise 14:

Let X and Y be independent and identically distributed random variables such that X induces the measure $\mu_X := p\delta_{\{1\}} + q\delta_{\{-1\}}$ where $p+q=1$.

(a) Compute the characteristic function of X.

(b) Show that X is symmetric if and only if $p = 1/2$.

(c) Take $p = 1/2$. Let φ_{X+Y} be the characteristic function of the random variable $X+Y$. Verify that $\varphi_{X+Y}(t) := cos^2(t)$, for all $t \in \mathbb{R}$.

(d) Using the convolution, determine the distribution function of the random variable $X+Y$. Show that $X+Y$ is symmetric if and only if $p = 1/2$. In this case, compute again the characteristic function of the random variable $X+Y$ and conclude that for all $t \in \mathbb{R}$

$$cos^2(t) := \frac{1+cos(2t)}{2}.$$

Exercise 15:

Let X and Y be independent and identically distributed random variables with $X \sim \mathcal{N}(0,1)$.

(a) Using characteristic functions and the convolution, show that $X+Y \sim \mathcal{N}(0,2)$.

(b) Show, using characteristic functions, that if $Z := \sigma X + \mu$ then $Z \sim \mathcal{N}(\mu, \sigma^2)$.

(c) Let φ_Z be the characteristic function of Z. Compute $|\varphi_Z|^2$ and verify that $|\varphi_Z|^2 \leq 1$. Is the random variable Z symmetric?

(d) Show that $i^{-1}\varphi_Z'(0) := \mu$ and that $-\varphi_Z''(0) = \mu^2 + \sigma^2$. Conclude that $Var(Z) = \sigma^2$.

Exercise 16:

(a) Let X be a random variable with exponential distribution with parameter $a > 0$. Compute $\varphi_X'(t)$, where φ_X is the characteristic function of X and verify that $i^{-1}\varphi_X'(0) = \mathbb{E}[X]$.

(b) Find a such that $\varphi_X''(0) = -1/8$. Compute $Var(X)$.

Exercise 17:

(a) Find the random variable X such that $\varphi(t) := \cos(t)$ is its characteristic function. Justify.

(b) Show that a symmetric random variable has all its odd moments equal to zero.

(c) Is $\varphi(t) := \mathbf{1}_{[-1,1]}(t)$ a characteristic function?

(d) Justify if $\varphi(t) := \frac{e^{it}+1}{2}$ is the characteristic function of a symmetric random variable? Find the random variable whose characteristic function is φ. Compute $|\varphi|^2$.

Exercise 18:

Using characteristic functions, show that for $g : \mathbb{R} \to \mathbb{R}$ a continuous function, if

$$X_n \xrightarrow[n\to+\infty]{} X, \quad \text{weakly}$$

then

$$g(X_n) \xrightarrow[n\to+\infty]{} g(X), \quad \text{weakly.}$$

Exercise 19:

Using characteristic functions prove Slutsky's Theorem:

Let $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$ be two sequences of random variables and let X be a random variable. Suppose that

$$X_n \xrightarrow[n\to+\infty]{} X, \quad \text{weakly} \quad \textit{and} \quad Y_n \xrightarrow[n\to+\infty]{} c, \quad \text{in} \quad \text{probability},$$

where c is a constant. Then

(a)

$$X_n + Y_n \xrightarrow[n\to+\infty]{} X + c, \quad \text{weakly.}$$

(b)

$$X_n - Y_n \xrightarrow[n\to+\infty]{} X - c, \quad \text{weakly.}$$

(c)

$$X_n Y_n \xrightarrow[n\to+\infty]{} Xc, \quad \text{weakly.}$$

(d) if $c \neq 0$ and $\mathbb{P}(Y_n \neq 0) = 1$, for all $n \geq 1$, then $\dfrac{X_n}{Y_n} \xrightarrow[n\to+\infty]{} \dfrac{X}{c}$, weakly.

Exercise 20:

Show, using characteristic functions that if $(X_n)_{n\geq 1}$ is a sequence of i.i.d.r.v. with $\mathbb{E}(X_1) = \mu < \infty$, then $\frac{S_n}{n} \xrightarrow[n\to+\infty]{} \mu$, in probability, where $S_n = \sum_{j=1}^n X_j$.

Exercise 21:

(a) Show, using characteristic functions that if $X \sim B(m,p)$ and $Y \sim B(n,p)$, and if X and Y are independent then $X+Y \sim B(n+m,p)$.

(b) Show that if X has standard Cauchy distribution, then $\varphi_{2X} = (\varphi_X)^2$. Use (without showing) that

$$\frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{\cos(tx)}{1+x^2}\,dx = e^{-|t|}.$$

(c) It is true that if X and Y are independent random variables then $\varphi_{X+Y} = \varphi_X \varphi_Y$. And the reciprocal, is it true? Prove and present a counter-example.

Exercise 22:

(a) Let $\varphi(t) = \cos(at)$ with $a > 0$. Show that φ is a characteristic function.

(b) Let $\varphi(t) = \cos^2(t)$. Find X such that φ is its characteristic function.

Exercise 23:

Let X and Y be i.i.d.r.v. with $\mathbb{E}(X) = 0$ and $Var(X) = 1$. Show that if $X+Y$ and $X-Y$ are independent then $X, Y \sim \mathcal{N}(0,1)$.

Chapter 4

Discrete time Martingales

4.1 Conditional expectation

Definition 4.1.1 (Conditional probability).

Given a set $A \in \mathscr{F}$ with $\mathbb{P}(A) > 0$ we define $\mathbb{P}_A(\cdot)$ in the following way:

$$\mathbb{P}_A(E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(A)}.$$

$\mathbb{P}_A$ is a probability measure and it is called the conditional probability with respect to A. The expectation with respect to this probability is called the conditional expectation wrt A:

$$\mathbb{E}_A[X] = \int_\Omega X(\omega)\mathbb{P}_A(d\omega) = \frac{1}{\mathbb{P}(A)} \int_A X(\omega)\mathbb{P}(d\omega).$$

Definition 4.1.2. *If we take now a partition of Ω that is $(A_n)_{n\geq 1}$ with $\Omega = \cup_{n\geq 1} A_n$, $A_n \in \mathscr{F}$ and $A_n \cap A_m = \varnothing$ if $m \neq n$, then given a set $E \in \mathscr{F}$ we have that*

$$\mathbb{P}(E) = \sum_{n\geq 1} \mathbb{P}(E \cap A_n) = \sum_{n\geq 1} \mathbb{P}_{A_n}(E)\mathbb{P}(A_n).$$

Definition 4.1.3. *As above we have that (if $\mathbb{E}[X]$ is finite)*

$$\mathbb{E}[X] = \int_\Omega X(\omega)\mathbb{P}(d\omega) = \int \cup_{n\geq 1} A_n X(\omega)\mathbb{P}(d\omega)$$
$$= \sum_{n\geq 1} \int_{A_n} X(\omega)\mathbb{P}(d\omega) = \sum_{n\geq 1} \mathbb{P}(A_n)\mathbb{E}_{A_n}[X].$$

Example 20. *Suppose that we have a card deck with 52 cards and that we take one out and it is spades. What is the probability of taking another card of the deck and that it is also spades?*

Theorem 4.1.4 (Wald's equation)**.**

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of i.i.d.r.v. with finite mean. For $k \geq 1$ let $\mathscr{F}_k$ be the σ-algebra generated by X_j with $j = 1, \cdots, k$. Suppose that N is a random variable taking positive integer values such that for all $k \geq 1$ we have that $\{N \leq k\} \in \mathscr{F}_k$ and $\mathbb{E}[N] < \infty$. Then $\mathbb{E}[S_N] = \mathbb{E}[X_1]\mathbb{E}[N]$.

Proof. To prove it note that

$$\begin{aligned}\mathbb{E}[S_N] &= \int_\Omega S_N \mathbb{P}(d\omega) = \int_{\{N\geq 1\}} S_N \mathbb{P}(d\omega) = \sum_{k\geq 1}\int_{\{N=k\}} S_N \mathbb{P}(d\omega)\\ &= \sum_{k\geq 1}\sum_{j=1}^{k}\int_{\{N=k\}} X_j \mathbb{P}(d\omega) = \sum_{j\geq 1}\sum_{k\geq j}\int_{\{N=k\}} X_j \mathbb{P}(d\omega)\\ &= \sum_{j\geq 1}\int_{\{N\geq j\}} X_j \mathbb{P}(d\omega) = \sum_{j\geq 1}\Big(\mathbb{E}[X_j] - \int_{\{N\leq j-1\}} X_j \mathbb{P}(d\omega)\Big).\end{aligned}$$

Now we note that the set $\{N \leq j-1\}$ and the r.v. X_j are independent (remember that $\{N \leq j-1\} \in \mathscr{F}_{j-1}$ and note the definition of $\mathscr{F}_{j-1}$), therefore we get

$$\mathbb{E}[S_N] = \sum_{j\geq 1}\mathbb{E}[X_j]\mathbb{P}(N \geq j) = \mathbb{E}[X_1]\sum_{j\geq 1}\mathbb{P}(N \geq j) = \mathbb{E}[X_1]\mathbb{E}[N].$$

To justify that we can interchange summations we have to repeat the computations taking $|X_j|$ and we will see that we get the result $\mathbb{E}[|X_1|]\mathbb{E}[N]$ which is finite by hypothesis. □

Now, let X be a discrete r.v. and let $A_n = \{X = a_n\}$. Given an integrable r.v. Y we define the function $\mathbb{E}[Y|\mathscr{G}]$ in Ω as

$$\mathbb{E}[Y|\mathscr{G}] = \sum_{n\geq 1}\mathbf{1}_{A_n}(\cdot)\mathbb{E}[Y|A_n],$$

this means that $\mathbb{E}[Y|\mathscr{G}]$ is a discrete r.v. that takes the value $\mathbb{E}[Y|A_n]$ on the set A_n.

We can rewrite the expression above as

$$\mathbb{E}[Y] = \sum_{n\geq 1} \int_{A_n} \mathbb{E}[Y|\mathcal{G}]\mathbb{P}(d\omega) = \int_{\Omega} \mathbb{E}[Y|\mathcal{G}]\mathbb{P}(d\omega)$$

Analogously for any $A \in \mathcal{G}$, A is a union of subcolletion of the A_n's, so that, for every $A \in \mathcal{G}$ we have that

$$\int_A Y\mathbb{P}(d\omega) = \int_A \mathbb{E}[Y|\mathcal{G}]\mathbb{P}(d\omega)$$

Attention to the measurability of the functions involved.

Now, we suppose that we have two functions φ_1 and φ_2 both $\mathcal{G}$ measurable and such that

$$\int_A Y\mathbb{P}(d\omega) = \int_A \varphi_1\mathbb{P}(d\omega) = \int_A \varphi_2\mathbb{P}(d\omega).$$

If we take the set $A = \{\omega \in \Omega : \varphi_1(\omega) > \varphi_2(\omega)\}$, then $A \in \mathcal{G}$ and we conclude that $\mathbb{P}(A) = 0$. Repeating the argument exchanging φ_1 with φ_2 we conclude that $\varphi_1 = \varphi_2$ a.e.

This means that $\mathbb{E}[Y|\mathcal{G}]$ is unique up to a equivalence and we are going to denote $\mathbb{E}_{\mathcal{G}}[Y]$ or $\mathbb{E}[Y|\mathcal{G}]$ to denote that class. The results holds for any σ-algebra.

Theorem 4.1.5. *If $\mathbb{E}[|Y|] < \infty$ and $\mathcal{G}$ is a σ-algebra contained in $\mathcal{F}$, then, there exists a unique equivalence class of integrable r.v. $\mathbb{E}[Y|\mathcal{G}]$ belonging to $\mathcal{G}$ such that for any $A \in \mathcal{G}$ it holds that $\int_A Y\mathbb{P}(d\omega) = \int_A \mathbb{E}[Y|\mathcal{G}]\mathbb{P}(d\omega)$.*

Definition 4.1.6 (Conditional expectation)**.**

Given an integrable r.v. Y and a σ-algebra $\mathcal{G}$, the conditional expectation $\mathbb{E}_{\mathcal{G}}[Y]$ of Y with respect to $\mathcal{G}$ is any one of the equivalence class of r.v. on Ω such that:

1. *it belongs to $\mathcal{G}$;*

2. *it has the same integral as Y over any set in $\mathcal{G}$.*

Note that for $Y = \mathbf{1}_\Lambda$ with $\Lambda \in \mathscr{F}$ we write $\mathbb{P}(\Lambda|\mathscr{G}) = \mathbb{E}[\mathbf{1}_\Lambda|\mathscr{G}]$ and this is the conditional probability of Λ relatively to $\mathscr{G}$. This is any one of the equivalence class of r.v. belonging to $\mathscr{G}$ and satisfying

$$\forall B \in \mathscr{G} : \mathbb{P}(B \cap \Lambda) = \int_B \mathbb{P}(\Lambda|\mathscr{G})\mathbb{P}(d\omega).$$

Theorem 4.1.7. *Let Y and ZY be integrable r.v. and let $Z \in \mathscr{G}$. Then*

$$\mathbb{E}[YZ|\mathscr{G}] = Z\mathbb{E}[Y|\mathscr{G}], \quad a.e.$$

Exercise: do the proof of the theorem.

Let us note that $\mathbb{E}[X|\mathscr{T}] = \mathbb{E}[X]$, where $\mathscr{T}$ is the trivial σ-algebra, that is $\mathscr{T} := \{\emptyset, \Omega\}$.

4.2 Properties of the conditional expectation

Let X and X_n be integrable r.v.

1. If $X \in \mathscr{G}$, then $\mathbb{E}[X|\mathscr{G}] = X$ a.e., this is true also if $X = a$ a.e.,
2. $\mathbb{E}[X_1 + X_2|\mathscr{G}] = \mathbb{E}[X_1|\mathscr{G}] + \mathbb{E}[X_2|\mathscr{G}]$,
3. If $X_1 \leq X_2$ then $\mathbb{E}[X_1|\mathscr{G}] \leq \mathbb{E}[X_2|\mathscr{G}]$,
4. $|\mathbb{E}[X|\mathscr{G}]| \leq \mathbb{E}[|X||\mathscr{G}]$,
5. If $X_n \uparrow X$, then $\mathbb{E}[X_n|\mathscr{G}] \uparrow \mathbb{E}[X|\mathscr{G}]$,
6. If $X_n \downarrow X$, then $\mathbb{E}[X_n|\mathscr{G}] \downarrow \mathbb{E}[X|\mathscr{G}]$,
7. If $|X_n| \leq Y$, Y is integrable and $X_n \to X$, then $\mathbb{E}[X_n|\mathscr{G}] \to \mathbb{E}[X|\mathscr{G}]$,
8. $\mathbb{E}[|XY||\mathscr{G}]^2 \leq \mathbb{E}[X^2|\mathscr{G}]\mathbb{E}[Y^2|\mathscr{G}]$. (Cauchy-Schwarz inequality)

Exercise: do the proof.

Theorem 4.2.1 (Jensen's inequality). *If φ is a convex function on $\mathbb{R}$ and X and $\varphi(X)$ are integrable r.v., then for each $\mathscr{G}$:*

$$\varphi(\mathbb{E}[X|\mathscr{G}]) \leq \mathbb{E}[\varphi(X)|\mathscr{G}].$$

Exercise: do the proof.

Note that when $\Lambda = \Omega$, the defining relation for the conditional expectation says that

$$\mathbb{E}[\mathbb{E}[Y|\mathscr{G}]|\mathscr{T}] = \mathbb{E}[Y|\mathscr{T}] = \mathbb{E}[\mathbb{E}[Y|\mathscr{T}]|\mathscr{G}]$$

This can be generalized and it is called the tower law.

Theorem 4.2.2 (Tower law). *If Y is integrable and $\mathscr{F}_1 \subset \mathscr{F}_2$, then:*

- $\mathbb{E}[Y|\mathscr{F}_1] = \mathbb{E}[Y|\mathscr{F}_2]$ *iff* $\mathbb{E}[Y|\mathscr{F}_2] \in \mathscr{F}_1$.
- $\mathbb{E}[\mathbb{E}[Y|\mathscr{F}_2]|\mathscr{F}_1] = \mathbb{E}[Y|\mathscr{F}_1] = \mathbb{E}[\mathbb{E}[Y|\mathscr{F}_1]|\mathscr{F}_2]$

As a particular case we note that

$$\mathbb{E}[\mathbb{E}[Y|X_1,X_2]|X_1] = \mathbb{E}[Y|X_1] = \mathbb{E}[\mathbb{E}[Y|X_1]|X_1,X_2].$$

Proof. We start with the first assertion. Let start by assuming that $\mathbb{E}[Y|\mathscr{F}_1] = \mathbb{E}[Y|\mathscr{F}_2]$, then by 1) in page 128 we have that $\mathbb{E}[Y|\mathscr{F}_2] \in \mathscr{F}_1$.

Now let us assume that $\mathbb{E}[Y|\mathscr{F}_2] \in \mathscr{F}_1$. Then, for $A \in \mathscr{F}_1$, 2) in page 128 holds, from where the result follows.

Now let us prove the second assertion. Note that $\mathbb{E}[Y|\mathscr{F}_1] \in \mathscr{F}_2$, and from the first assertion applied to $\mathbb{E}[Y|\mathscr{F}_1]$ we conclude the second equality. Let us prove the first equality now. For that purpose note that if $\Lambda \in \mathscr{F}_1$ then $\Lambda \in \mathscr{F}_2$, so that

$$\int_\Lambda \mathbb{E}[\mathbb{E}[Y|\mathscr{F}_2]|\mathscr{F}_1]\mathbb{P}(d\omega) = \int_\Lambda \mathbb{E}[Y|\mathscr{F}_2]\mathbb{P}(d\omega) = \int_\Lambda Y\mathbb{P}(d\omega).$$

Moreover, $\mathbb{E}[\mathbb{E}[Y|\mathscr{F}_2]|\mathscr{F}_1] \in \mathscr{F}_1$ so that, both properties defining the conditional expectation are verified and we are done. □

4.3 Conditional independence

Let $\mathscr{F}$ be a σ-algebra and let $\{\mathscr{F}_\alpha\}_{\alpha\in A}$, where A is a index set, be contained in $\mathscr{F}$.

Definition 4.3.1. *The collection $\{\mathscr{F}_\alpha\}_{\alpha\in A}$ is said to be conditionally independent to a σ-algebra $\mathscr{G}$ iff for any finite collection of sets $A_1, \cdots, A_n$ with $A_j \in \mathscr{F}_j$ and with α_j's distinct indices of A we have*

$$\mathbb{P}\Big(\cap_{j=1}^n A_j|\mathscr{G}\Big) = \prod_{j=1}^n \mathbb{P}(A_j|\mathscr{G}).$$

Note that if $\mathscr{G} = \mathscr{T}$ then the previous condition is just the usual independence.

Theorem 4.3.2. *For each $\alpha \in A$, let $\mathscr{F}^{(\alpha)}$ be the smallest σ-algebra containing all $\mathscr{F}_\beta$ with $\beta \in A\backslash\{\alpha\}$. Then, the $\mathscr{F}_\alpha$'s are conditionally independent relatively to a σ-algebra $\mathscr{G}$ iff for each α and $A_\alpha \in \mathscr{F}_\alpha$ we have*

$$\mathbb{P}\Big(A_\alpha|\mathscr{F}^{(\alpha)} \vee \mathscr{G}\Big) = \mathbb{P}(A_\alpha|\mathscr{G}),$$

where $\mathscr{F}^{(\alpha)} \vee \mathscr{G}$ denotes the smallest σ-algebra containing $\mathscr{F}^{(\alpha)}$ and $\mathscr{G}$.

Note that if in the previous theorem $\mathscr{G} = \mathscr{T}$ and $\mathscr{F}_\alpha$ is generated by a r.v. say X_α then we have

Corollary 4.3.3. *Let $(X_\alpha)_{\alpha\in A}$ be a collection of r.v. and for each α let $\mathscr{F}^{(\alpha)}$ be the σ-algebra generated by all the r.v. except by X_α. Then, the r.v. X_α's are independent iff for each α and $B \in \mathscr{B}$ we have*

$$\mathbb{P}(X_\alpha \in B|\mathscr{F}^{(\alpha)}) = \mathbb{P}(X_\alpha \in B) \quad a.e.$$

Now, let X_1 and X_2 be two independent r.v. What happens if we condition X_1+X_2 by X_1?

Theorem 4.3.4. *Let X_1 and X_2 be two independent r.v. with probability measures μ_1 and μ_2, respectively. Then, for each $B \in \mathscr{B}$:*

$$\mathbb{P}(X_1+X_2 \in B|X_1) = \mathbb{P}(X_1+X_2 \in B|\mathscr{F}_1) = \mu_2(B-X_1) \quad a.e.$$

where $\mathscr{F}_1$ is the $\sigma-$algebra generated by X_1.

More generally, let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent r.v. with probability measures $(\mu_n)_{n\in\mathbb{N}}$ and let $S_n = X_1+\cdots+X_n$. Then, for each $B \in \mathscr{B}$:

$$\mathbb{P}(S_n \in B|S_1,\cdots,S_{n-1}) = \mu_n(B-S_{n_1}) = \mathbb{P}(S_n \in B|S_{n-1}) \quad a.e.$$

Exercise: Prove all the results above.

Let us look quickly at the proof of the previous theorem.

Remember that $\mathbb{P}(X_1+X_2 \in B|X_1) = \mathbb{E}[\mathbf{1}_{\{X_1+X_2\in B\}}|X_1]$. Now using the Theorem of page 76 we have that, for $\Lambda \in \mathscr{F}_1$ (note that this set is such that $\Lambda = X_1^{-1}(A)$, where $A \in \mathscr{B}$, to prove this use the trick with monotone classes, see the Theorem in page 4)

$$\begin{aligned}
&\int_\Lambda \mu_2(B-X_1)\mathbb{P}(d\omega) = \int_A \mu_2(B-x_1)\mu_1(dx_1)\\
&= \int_A \mu_1(dx_1)\int_\Omega \mathbf{1}_{\{x_1+x_2\in B\}}\mu_2(dx_2) = \int\int_{\{x_1\in A, x_1+x_2\in B\}} \mu_1\times\mu_2(dx_1,dx_2)\\
&= \int\int_{\{X_1\in A, X_1+X_2\in B\}} \mathbb{P}(d\omega) = \mathbb{P}(X_1\in A, X_1+X_2\in B)\\
&= \int_\Lambda \mathbf{1}_{\{X_1+X_2\in B\}}\mathbb{P}(d\omega).
\end{aligned}$$

Since $\mu_2(B-X_1)\in\mathscr{F}_1$ and since the previous relation is true for any $\Lambda\in\mathscr{F}_1$, the result follows. As an exercise, prove the second assertion of the theorem.

4.3.1 Conditional distribution of X given a set A.

Given a r.v. X in a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and for an event A with $\mathbb{P}(A) > 0$ we define the conditional distribution of X given A as:

$$\mathbb{P}(X \in B|A) = \frac{\mathbb{P}((X \in B) \cap A)}{\mathbb{P}(A)}.$$

Exercise: Check that this gives a probability measure on the Borel σ-algebra.

Now, we can define the conditional distribution function of X given the set A on $x \in \mathbb{R}$ as

$$F_X(x|A) = \mathbb{P}(X \leq x|A) = \frac{\mathbb{P}((X \leq x) \cap A)}{\mathbb{P}(A)}$$

The conditional expectation of X given the set A is the expectation of the conditional distribution given by

$$\mathbb{E}[X|A] = \int x \, dF_X(x|A)$$

if it exists. As above, if we take now a partition of Ω that is $(A_n)_{n\geq 1}$ with $\Omega = \cup_{n\geq 1} A_n$, $A_n \in \mathscr{F}$ and $A_n \cap A_m = \varnothing$ if $m \neq n$, then

$$\mathbb{P}(X \in B) = \sum_{n\geq 1} \mathbb{P}(X \in B|A_n)\mathbb{P}(A_n).$$

Also for any x, $F_X(x) = \mathbb{P}(X \leq x) = \sum_{n\geq 1} \mathbb{P}(X \leq x|A_n)\mathbb{P}(A_n) = \sum_{n\geq 1} F_X(x|A_n)\mathbb{P}(A_n)$ and analogously

$$\mathbb{E}[X] = \int x dF_X(x) = \sum_{n\geq 1} \mathbb{P}(A_n)\mathbb{E}[X|A_n].$$

1) Let $X \sim U[-1,1]$ and let $A = \{X \geq 0\}$. What is the conditional distribution of X given A?

4.3.2 Conditional distribution of X given a discrete r.v. Y

Let us suppose now that the partition is generated by a discrete r.v. Let Y be a discrete r.v. defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ taking the values $(a_n)_{n\in\mathbb{N}}$. Then the events $\{Y = a_n\}$ form a partition of Ω. In this case $\mathbb{P}(X \in B|Y = a_n)$ is called the conditional distribution of X given $Y = a_n$ and we have that

$$\mathbb{P}(X \in B|Y = a_n) = \sum_{n\geq 1} \mathbb{P}(X \in B|Y = a_n)\mathbb{P}(Y = a_n).$$

Also for any x,

$$\begin{aligned} F_X(x) = \mathbb{P}(X \leq x) &= \sum_{n\geq 1} \mathbb{P}(X \leq x|Y = a_n)\mathbb{P}(Y = a_n) \\ &= \sum_{n\geq 1} F_X(x|Y = a_n)\mathbb{P}(X = a_n) \end{aligned}$$

and analogously $\mathbb{E}[X] = \int x dF_X(x) = \sum_{n\geq 1} \mathbb{P}(Y = a_n)\mathbb{E}[X|Y = a_n]$.

Note that for B fixed we have that $\mathbb{P}(X \in B|Y = a_n)$ is a function of a_n let us say $g(a_n)$. Defining $g(y) = \mathbb{P}(X \in B|Y = y)$ we have that $\mathbb{P}(X \in B) = \int \mathbb{P}(X \in B|Y = y)dF_Y(y) = \int g(y)dF_Y(y)$. Moreover,

$$F_X(x) = \int F_X(x|Y = y)dF_Y(y) \qquad \mathbb{E}[X] = \int \mathbb{E}[X|Y = y]dF_Y(y).$$

When X is integrable the function $\varphi(y) = \mathbb{E}[X|Y = y]$ is finite. In this case, the r.v. $\varphi(Y)$ is called the conditional expectation of X given Y: $\varphi(Y) = \mathbb{E}[X|Y]$. We note that $\mathbb{E}[X|Y = y]$ is the value of the random variable $\mathbb{E}[X|Y]$ when $Y = y$. The last formula can be interpreted as

$$\mathbb{E}[X] = \mathbb{E}[\varphi(Y)] = \mathbb{E}[\mathbb{E}[X|Y]].$$

2) Consider the following experience: a player tosses a fair coin n times obtaining k heads with $0 \leq k \leq n$. After that a second player tosses the same coin k times. Let X be the number of heads obtained by the second player. What is the expectation of X supposing that all the events are independent?

4.3.3 Conditional distribution: general case

Let us define now the conditional expectation for general r.v. X and Y. Before we defined the conditional distribution of X when Y was discrete, so that $\mathbb{P}(Y = y) = 0$ for all $y \neq a_n$. But now we want to extend this to the continuous case in which the probability above is null for all $y \in \mathbb{R}$. How to do it? We define by approximation. Take I an interval containing y with size Δy and define

$$\mathbb{P}(X \in B|Y = y) \sim \mathbb{P}(X \in B|Y \in I) = \frac{\mathbb{P}(X \in B, Y \in I)}{\mathbb{P}(Y \in I)}.$$

If $\mathbb{P}(X \in B|Y \in I)$ has a limit when $\Delta y \to 0$ we call to the limit $\mathbb{P}(X \in B|Y = y)$:

$$\lim_{\Delta y \to 0} \mathbb{P}(X \in B|Y \in I) = \mathbb{P}(X \in B|Y = y).$$

Let us go back to the case in which X is discrete.

Then we have

$$\begin{aligned} F_{(X,Y)}(x, y) = \mathbb{P}(X \leq x, Y \leq y) &= \sum_{n:a_n \leq y} \mathbb{P}(X \leq x, Y = a_n) \\ &= \sum_{n:a_n \leq y} \mathbb{P}(X \leq x|Y = a_n)\mathbb{P}(Y = a_n) \\ &= \sum_{n:a_n \leq y} F_X(x|Y = a_n)\mathbb{P}(Y = a_n) \\ &= \int_{-\infty}^{y} F_X(x|Y = a)dF_Y(a). \end{aligned}$$

Note that in the discrete case, the joint distribution is like a composition of the marginal distribution of Y with the conditional distribution of X given Y. Let use then the last equality!

Definition 4.3.5. *Let X and Y be two r.v. defined on the same probability space. A function $\mathbb{P}(X \in B|Y = y)$ defined for each borelian B and $y \in \mathbb{R}$ is a (regular) conditional distribution for X given Y if:*

1. *for each y fixed, $\mathbb{P}(X \in B|Y = y)$ defines a probability measure in $\mathscr{B}$,*

2. *for any $B \in \mathscr{B}$ fixed, $\mathbb{P}(X \in B|Y = y)$ is a measurable function of y,*

3. *for any $(x, y) \in \mathbb{R}^2$ it holds that*

$$\mathbb{P}(X \leq x, Y \leq y) = \int_{-\infty}^{y} F_X(x|Y = a)dF_Y(a).$$

$\mathbb{P}(X \in B|Y = y)$ is called the conditional probability of X belonging to B given that $Y = y$ and $F_X(\cdot|Y = y) = \mathbb{P}(X \leq \cdot|Y = y)$ is the conditional distribution of X given $Y = y$.

Theorem 4.3.6. *Let X and Y be two r.v. defined on the same probability space. There exists a (regular) conditional distribution for X given Y. In fact there exists only one in the sense that they are equal a.e.: that is, if $\mathbb{P}_1(X \in B|Y = y)$ and $\mathbb{P}_2(X \in B|Y = y)$ are conditional distributions for X given Y, then there exists a borelian B_0 such that $\mathbb{P}(Y \in B_0) = 1$ and $\mathbb{P}_1(X \in B|Y = y) = \mathbb{P}_2(X \in B|Y = y)$ for all $B \in \mathscr{B}$ and $y \in B_0$.*

Theorem 4.3.7. *For each $B \in \mathscr{B}$ fixed, the limit*

$$\lim_{\Delta a \to 0} \mathbb{P}(X \in B|Y \in I) = \mathbb{P}(X \in B|Y = a)$$

exists a.e. Moreover, for each $B \in \mathscr{B}$ fixed, the limit is equal to $\mathbb{P}(X \in B|X = y)$ as given in the definition above, a.e.

1) What is the conditional distribution of Y given Y? Let us guess it. If it is given that $Y = y$, then $Y = y$! So the candidate is $\mathbb{P}(Y = y|Y = y) = 1$ the distribution which gives weight 1 to the point y. Check that for $B = (q_1, q_2)$ with $q_i \in \mathbb{Q}$ it holds that

$$\mathbb{P}(Y \in B|Y = y) = \lim_{\Delta a \to 0} \mathbb{P}(Y \in B|Y \in I),$$

which proves the result.

Note however that if we take $B = \{y_0\}$ then

$$\mathbb{P}(Y = y_0|Y = y) = \lim_{\Delta a \to 0} \mathbb{P}(Y = y_0|Y \in I) = 0!$$

This does not contradict our result but contradicts our intuition!

2) Given $Y = y$ what is the conditional distribution of $Z = g(Y)$? Recall that above we have seen that if $Y = y$, then $\mathbb{P}(Y = y|Y = y) = 1$. Here it is analogous. In this case we have that $\mathbb{P}(g(Y) = g(y)|Y = y) = 1$.

3) Let X be a symmetric r.v. around 0. What is the conditional distribution of X given the r.v. $|X|$? Given that $|X| = y > 0$, then $X = y$ or $-y$, there are no other possibilities and from symmetry we have that:

$$\mathbb{P}(X = y||X| = y) = \frac{1}{2} = \mathbb{P}(X = -y||X| = y), \quad y > 0,$$

and $\mathbb{P}(X = 0||X| = 0) = 1$.

Let us do it now in a different way. Suppose $y > 0$. Take $B = (q_1, a_2)$ with $q_i \in \mathbb{Q}$ and take $I \subset B$. Then

$$\mathbb{P}(X \in B||X| \in I) = \mathbb{P}(X \in I) = \frac{1}{2}\Big(\mathbb{P}(X \in I) + \mathbb{P}(X \in -I)\Big) = \frac{1}{2}\mathbb{P}(|X| \in I).$$

And

$$\mathbb{P}(X \in -B||X| \in I) = \mathbb{P}(X \in -I) = \frac{1}{2}\mathbb{P}(|X| \in I).$$

Since $I \subset B$ we have that

$$\mathbb{P}(X \in B||X| \in I) = \frac{1}{2} = \mathbb{P}(X \in -B||X| \in I).$$

Therefore,

$$\mathbb{P}(X \in B||X| = y) = \lim_{\Delta y \to 0} \mathbb{P}(X \in B||X| \in I) = \frac{1}{2},$$
$$\mathbb{P}(X \in -B||X| = y) = \lim_{\Delta y \to 0} \mathbb{P}(X \in -B||X| \in I) = \frac{1}{2}.$$

Taking B decreasing to $\{y\}$ we see that the conditional probability gives weight $1/2$ to each one of the points y and $-y$. The proof that $\mathbb{P}(X = 0||X| = 0) = 1$ can be reached by taking $B = (q_1, q_2)$ as above with $q_1 < 0 < q_2$.

4) Let X and Y be independent r.v. each one with law $N(0, \sigma^2)$ with $\sigma^2 > 0$. What is the conditional distribution of (X, Y) given $\sqrt{X^2 + Y^2}$?

For $z > 0$, $\sqrt{X^2 + Y^2} = z$ iff (X, Y) is in the circle centered at $(0, 0)$ with radius z. Therefore the conditional distribution is concentrated in that circle, that is, in the set of points of $\mathbb{R}^2$ given by $\mathscr{C} := \{(x, y) : x^2 + y^2 = z\}$.

Note that the joint density function of (X,Y) is given by

$$f(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)^2}{2\sigma^2}}.$$

Note that the density is constant on the circle $\mathscr{C}$. Therefore, before the experience all the points in the circle $\mathscr{C}$ had the same "chance" and our guess for the distribution is the uniform distribution on the circle, that is, for $B \in \mathscr{B}^2$ and $z > 0$:

$$\mathbb{P}((X,Y) \in B | \sqrt{X^2+Y^2} = z) = \frac{\text{"size of"}(B \cap \mathscr{C})}{2\pi z}.$$

Prove it!

4.4 Discrete time Martingales

Let $(X_n)_{n\in\mathbb{N}}$ be independent r.v. with mean zero and let $S_n = \sum_{j=1}^n X_j$. Then

$$\begin{aligned}\mathbb{E}[S_{n+1}|X_1,\cdots,X_n] &= \mathbb{E}[X_1+\cdots+X_n+X_{n+1}|X_1,\cdots,X_n]\\ &= S_n + \mathbb{E}[X_{n+1}|X_1,\cdots,X_n] = S_n + \mathbb{E}[X_{n+1}]\\ &= S_n.\end{aligned}$$

Historically, the equation above gave rise to consider dependent r.v. which satisfy $\mathbb{E}[X_{n+1}|X_1,\cdots,X_n] = 0$ and this opened a way to define a class of stochastic processes which are extremely useful - **the martingales**.

Definition 4.4.1 (Smartingale: martingale, submartingale, supermartingale)**.**

The sequence of r.v. and $\sigma-$algebras $(X_n, \mathscr{F}_n)_{n\in\mathbb{N}}$ is said to be a martingale iff for each $n \in \mathbb{N}$ we have that

1. *$\mathscr{F}_n \subset \mathscr{F}_{n+1}$ and $X_n \in \mathscr{F}_n$, (this means that X_n is adapted to $\mathscr{F}_n$)*
2. *$\mathbb{E}[|X_n|] < \infty$ for each $n \in \mathbb{N}$, (this means that X_n is integrable)*
3. *for each $n \in \mathbb{N}$, we have that*

$$\begin{aligned} X_n &= \mathbb{E}[X_{n+1}|\mathscr{F}_n] \quad a.e. \quad (martingale)\\ X_n &\le \mathbb{E}[X_{n+1}|\mathscr{F}_n] \quad a.e. \quad (submartingale)\\ X_n &\ge \mathbb{E}[X_{n+1}|\mathscr{F}_n] \quad a.e. \quad (supermartingale)\end{aligned}$$

Example 21. *Check that* $(Y_n, \mathscr{F}_n)_{n\in\mathbb{N}}$ *is a (sub)martingale in each case below:*

1. *let* $(X_n)_n$ *be a sequence of independent r.v. with mean zero,* $\mathscr{F}_n = \sigma(X_1, \cdots, X_n)$ *and* $Y_n = S_n$*,*
2. *let* $(X_n)_n$ *be a sequence of independent r.v. with mean one,* $\mathscr{F}_n = \sigma(X_1, \cdots, X_n)$ *and* $Y_n = \prod_{k=1}^n X_k$*,*
3. *let* X *be an integrable r.v. and let* $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \cdots \subset \mathscr{F}$*,* $Y_n = \mathbb{E}[X|\mathscr{F}_n]$*, (GOOD FOR CREATING MARTINGALES!)*
4. *let* $(X_n)_n$ *be a sequence of non-negative integrable r.v.,* $\mathscr{F}_n = \sigma(X_1, \cdots, X_n)$ *and* $Y_n = S_n$*, (sub)*

Note that the condition for martingale implies that for $n < m$ we have that

$$X_n = \mathbb{E}[X_m|\mathscr{F}_n] \quad \text{a.e.}$$

Theorem 4.4.2 (Jensen's inequality)**.**
Let $(X_n, \mathscr{F}_n)_{n\in\mathbb{N}}$ *be a submartingal and let* φ *be an increasing convex function defined on* $\mathbb{R}$*. If* $\varphi(X_n)$ *is integrable for any* n*, then* $(\varphi(X_n), \mathscr{F}_n)_{n\in\mathbb{N}}$ *is also a submartingal.*

Corollary 4.4.3. *If* $(X_n, \mathscr{F}_n)_{n\in\mathbb{N}}$ *is a submartingal then* $(X_n^+, \mathscr{F}_n)_{n\in\mathbb{N}}$ *is a submartingale. If* $(X_n, \mathscr{F}_n)_{n\in\mathbb{N}}$ *is a martingal, then* $(|X_n|, \mathscr{F}_n)_{n\in\mathbb{N}}$ *and* $(|X_n|^p, \mathscr{F}_n)_{n\in\mathbb{N}}$ *for* $1 < p < \infty$ *if* $X_n \in \mathbb{L}^p$ *are also submartingales.*

Exercise: Prove the theorem.

4.4.1 Martingales in Game theory

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d.r.v. taking the value 1 with probability p and -1 with probability $1-p$. The interpretation is that $X_n = 1$ represents a success while $X_n = -1$ represents a failure of a player at the n-th time he is playing a game. Let us suppose that the player can win or lose a certain amount V_n at the n-th time he plays the game, that is, V_n is the amount of the bet at time n. Then, at time n the player possesses

$$Y_n = \sum_{i=1}^{n} V_i X_i = Y_{n-1} + V_n X_n.$$

It is quite natural to assume that the amount V_n may depend on the previous amounts, that is, of $V_1, \cdots, V_{n-1}$ and also of $X_1, \cdots, X_{n-1}$. In other words, let $\mathscr{F}_0 = \{\emptyset, \Omega\}$ and $\mathscr{F}_n = \sigma(X_1, \cdots, X_n)$. Then, V_n is a function $\mathscr{F}_{n-1}$ measurable, that is, the sequence that determines the player's strategy is said to be predictable.

Let $S_n = X_1 + \cdots + X_n$. Then

$$Y_n = \sum_{i=1}^{n} V_i \Delta S_i,$$

where $\Delta S_i = S_i - S_{i-1}$. Then, the sequence $(Y_n, \mathscr{F}_n)_{n\in\mathbb{N}}$ is said to be the transform of S by V.

From the player's point of view, the game is said to be fair (favorable or unfavorable) if at each step if $\mathbb{E}[Y_{n+1} - Y_n | \mathscr{F}_n] = 0 \quad (\geq 0 \quad \text{or} \quad \leq 0)$

We want to analyze in which conditions the game is fair? A simple computation shows that :

1. The game is fair if $p = 1 - p = 1/2$. $(Y_n, \mathscr{F}_n)_{n\in\mathbb{N}}$ is a martingale.

2. The game is favorable if $p > 1 - p$. $(Y_n, \mathscr{F}_n)_{n\in\mathbb{N}}$ is a submartingale.

3. The game is fair if $p < 1 - p$. $(Y_n, \mathscr{F}_n)_{n\in\mathbb{N}}$ is a supermartingale.

Let us now consider another strategy. Take $(V_n, \mathscr{F}_{n-1})_{n\geq 1}$ with $V_1 = 1$ and for $n \geq 1$ we have that $V_n = 2^{n-1}$ if $X_1 = -1, \cdots, X_{n-1} = -1$ and 0 otherwise.

Under this strategy, a player starts to bet 1 euro and doubles the bet in the next play if he had lost or leaves immediately the game in case he had won.

If $X_1 = -1, \cdots X_n = -1$, then the total loss after n plays is $\sum_{i=1}^{n} 2^{i-1} = 2^n - 1$.

Therefore, if $X_{n+1} = 1$ then $Y_{n+1} = Y_n + X_{n+1} V_{n+1} = -(2^n - 1) + 2^n = 1$.

Let $\tau := \inf\{n \geq 1 : Y_n = 1\}$, that is the first time that $Y_n = 1$. If $p = \frac{1}{2}$, then the game is fair and

$$\mathbb{P}(\tau = n) = \mathbb{P}(Y_n = 1, Y_k \neq 1, \forall k = 1, \cdots, n-1) = \left(\frac{1}{2}\right)^n.$$

From where we conclude that

$$\mathbb{P}(\tau < \infty) = \mathbb{P}(\cup_{n\geq 1} \tau = n) = \sum_{n\geq 1} \left(\frac{1}{2}\right)^n = 1.$$

Moreover, $\mathbb{P}(Y_\tau = 1) = 1$ and $\mathbb{E}[Y_\tau] = 1$.

Therefore, even in a fair game, applying the strategy described above, a player can, in finite time, complete the game with success, that is, increase his capital in one unity: $\mathbb{E}[Y_\tau] = 1 > Y_0 = 0$. In game theory this type of system - double the bet after a loss and leave the game immediately after a win - is called a martingale.

We note however that $p = 1/2$, so that $(Y_n, \mathscr{F}_n)_{n\in\mathbb{N}}$ is a martingale and $\mathbb{E}[Y_n] = \mathbb{E}[Y_0] = 0$ for all $n \geq 1$. Above the same is not true for a random time (above we took the random time τ.)

Definition 4.4.4 (Markov time)**.**

A r.v. τ which takes values in the set $\{0, 1, \cdots, \infty\}$ is said to be a Markov time wrt a σ-algebra $\mathscr{F}_n$ if for each $n \geq 0$ we have that $\{\tau = n\} \in \mathscr{F}_n$. When $\mathbb{P}(\tau < \infty) = 1$, the Markov time is said to be a stopping time.

If $(X_n, \mathscr{F}_n)_{n\in\mathbb{N}}$ is a sequence of r.v. and σ-algebras with $\mathscr{F}_n \in \mathscr{F}_{n+1}$, and if τ is a Markov time wrt $\mathscr{F}_n$, then we write $X_\tau = \sum_{n=0}^{\infty} X_n \mathbf{1}_{\{\tau=n\}}$. Note that since $\mathbb{P}(\tau < \infty) = 1$ we have that $X_\tau = 0$ in the set $\tau = \infty$. Prove that X_τ is a r.v.

Example 22 (Prove it!)**.** *Let $(X_n, \mathscr{F}_n)_{n\in\mathbb{N}}$ be a martingale (or submartingale) and τ a Markov time wrt $\mathscr{F}_n$. Then the stopping process $X^\tau = (X_{n\wedge\tau}, \mathscr{F}_n)$ is also a martingale (or submartingale).*

4.5 Exercises

Exercise 1: Show that:

(a) if $(X_n)_{n\geq 1}$ is a sequence of independent r.v. with $\mathbb{E}[X_n] = 0$ for all $n \geq 1$, then $(S_n, \mathscr{F}_n)_{n\geq 1}$ where $S_n = \sum_{j=1}^{n} X_j$ and $\mathscr{F}_n = \sigma(X_1, \cdots, X_n)$ is a martingale

(b) if $(X_n)_{n\geq 1}$ is a sequence of independent r.v. with $\mathbb{E}[X_n] = 1$ for all $n \geq 1$, then $(\tilde{X}_n, \mathscr{F}_n)_{n\geq 1}$ where $\tilde{X}_n = \prod_{j=1}^n X_j$ and $\mathscr{F}_n = \sigma(X_1, \cdots, X_n)$, is a martingale.

(c) given an integrable r.v. X, that is with $\mathbb{E}[|X_n|] < +\infty$ and a set of σ-algebras $\mathscr{F}_0 \subsetneqq \mathscr{F}_1 \subsetneqq \cdots \subsetneqq \mathscr{F}_n$, then $(X_n, \mathscr{F}_n)_{n\geq 1}$ where $X_n = \mathbb{E}[X|\mathscr{F}_n]$ is a martingale.

Exercise 2: Show that:

(a) if $(X_n)_{n\geq 1}$ is a sequence of non-negative integrable r.v., then $(S_n, \mathscr{F}_n)_{n\geq 1}$ where $S_n = \sum_{j=1}^n X_j$ and $\mathscr{F}_n = \sigma(X_1, \cdots, X_n)$ is a submartingale.

(b) if $(X_n, \mathscr{F}_n)_{n\geq 1}$ i a martingale and $g : \mathbb{R} \to \mathbb{R}$ is a convex function with $\mathbb{E}[|g(X_n)|] < +\infty$ for all $n \geq 1$, then $(g(X_n), \mathscr{F}_n)_{n\geq 1}$ is a submartingale.

Exercise 3: Let $(X_n)_{n\geq 1}$ be i.i.d. r.v. with $\mathbb{P}(X_1 = 1) = p$ and $\mathbb{P}(X_1 = -1) = q$ with $p + q = 1$. If $p \neq q$, show that if $S_n = \sum_{j=1}^n X_j$ and $\mathscr{F}_n = \sigma(X_1, \cdots, X_n)$, then

(a) $(Y_n, \mathscr{F}_n)_{n\geq 1}$ is a martingale, where $Y_n = \left(\frac{q}{p}\right)^{S_n}$.

(b) $(Z_n, \mathscr{F}_n)_{n\geq 1}$ is a martingale, where $Z_n = S_n - n(p - q)$.

Exercise 4: Show that if $(X_n)_{n\geq 1}$ is a sequence of i.i.d. r.v. with $\mathbb{E}[X_n] = 0$ and $Var(X_n) = \sigma^2$ for all $n \geq 1$, then $(\mathscr{W}_n, \mathscr{F}_n)_{n\geq 1}$ is a martingale, where $\mathscr{F}_n = \sigma(X_1, \cdots, X_n)$ and

(a)

$$\mathscr{W}_n = \Big(\sum_{j=1}^n X_j\Big)^2 - n\sigma^2.$$

(b)

$$\mathscr{W}_n = \frac{e^{\lambda \sum_{j=1}^n X_j}}{(E[e^{\lambda X_1}])^n}.$$

Exercise 5: Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. r.v. that take values on a finite set $\mathscr{I}$. For each $y \in \mathscr{I}$, let $f_0(y) = \mathbb{P}(X_1 = y)$ and let $f_1 : \mathscr{I} \to [0, 1]$ be a non-negative function such that $\sum_{y\in\mathscr{I}} f_1(y) = 1$. Show that $(\mathscr{W}_n, \mathscr{F}_n)_{n\geq 1}$ is a martingale, where $\mathscr{F}_n = \sigma(X_1, \cdots, X_n)$ and

$$\mathscr{W}_n = \frac{f_1(X_1)\cdots f_1(X_n)}{f_0(X_1)\cdots f_0(X_n)}.$$

The r.v. $\mathscr{W}_n$ are known as likelihood ratios.

Exercise 6: Let $(X_n, \mathscr{F}_n)_{n\geq 1}$ be a martingale.

(a) Show that, for all $n < m$ it holds that $X_n = E[X_m|\mathscr{F}_n]$.

(b) Conclude that $\mathbb{E}[X_1] = \mathbb{E}[X_n]$ for all $n \geq 1$.

(c) For each $n \geq 2$ let $Y_n = X_n - X_{n-1}$ and take $Y_1 = X_1$. We observe that Y_n is called the increment of the martingale. Show that $\mathbb{E}[Y_n] = 0$ for all $n \geq 0$.

(d) Assume that $\mathbb{E}[X_n^2] < +\infty$ for all $n \geq 1$. Show that the increments of the martingale are non correlated.

(e) Show that $Var(X_n) = \sum_{j=1}^n Var(Y_j)$.

Exercise 7: Let $(X_n, \mathscr{F}_n)_{n\geq 1}$ and $(Y_n, \mathscr{F}_n)_{n\geq 1}$ be two martingales with $X_1 = Y_1 = 0$. Show that

$$\mathbb{E}[X_n Y_n] = \sum_{k=2}^{n} \mathbb{E}[(X_k - X_{k-1})(Y_k - Y_{k-1})].$$

Exercise 8: Let $(X_n, \mathscr{F}_n)_{n\geq 1}$ be a martingale (or submartingale) and τ a Markov time (with respect to $\mathscr{F}_n$). Then, the stopping time

$$X^{\tau} = (X_{\min\{n,\tau\}}, \mathscr{F}_n)$$

is also a martingale (or a submartingale).

Exercise 9:

(a) Prove Wald's inequality. Let $(X_n)_{n\geq 1}$ be a sequence of integrable i.i.d. r.v. and let τ be a stopping time with respect to $\mathscr{F}_n = \sigma(X_1, \cdots, X_n)$ and $\mathbb{E}[\tau] < \infty$. Then, $\mathbb{E}[X_1 + \cdots + X_\tau] = \mathbb{E}[X_1]\mathbb{E}[\tau]$.

(b) Analyze the case in which $\mathbb{P}(X_1 = 1) = 1/2 = \mathbb{P}(X_1 = -1)$ and $\tau = \inf\{n \geq 1 : X_1 + \cdots + X_\tau = 1\}$. What do you conclude about $\mathbb{E}[\tau]$?

Exercise 10: Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d.r.v. such that $\mathbb{P}(X_1 = 1) = p = 1 - \mathbb{P}(X_1 = -1)$. Interpret $X_n = 1$ as a success and $X_n = -1$ as the lost of a player in its n-th play. Assume that the player can win or lose in the n-th play the amount V_n (so that V_n is the amount of the bet in the n-th play). The total amount of the player at the n-th play is given by $Y_n = \sum_{i=1}^{n} X_i V_i$. Assume that V_i is predictable with respect to $\mathscr{F}_n = \sigma(X_1, \cdots, X_n)$.

a) Verify in which conditions the game is fair, favorable or unfavorable. In each case, verify if $(Y_n, \mathscr{F}_n)_n$ is a martingale, sub-martingale or supermartingale.

b) Now consider the following strategy $V_1 = 1$ and

$$V_n = 2^{n-1}\mathbf{1}_{\{X_1=-1,\cdots,X_{n-1}=-1\}}.$$

Say by words what means that strategy. Is $(V_n)_n$ predictable with respect to $\mathscr{F}_n$? Let

$$\tau = \inf\{n \geq 1 : Y_n = 1\}.$$

Take $p = 1/2$, compute the probability function of τ and express $\mathbb{P}(\tau < \infty)$. Compute $\mathbb{E}[Y_\tau]$. What can you say about the game?

TÉCNICO LISBOA

Bibliography

[1] Kai Lai Chung: "A course in probability theory", Academic Press.

[2] Richard Durrett: "Probability: Theory and Examples", Duxbury Resource Center

[3] Barry James: "Probabilidade: um curso em nível intermediário", Publications from IMPA.

[4] Patrick Billingsley: "Probability and measure", Wiley editors.

[5] A. Shiryayev: "Probability", Springer-Verlag.