Chapter 1

Measure theory and Probability

1.1 Set sequences

In this section Ω is a set and $\mathcal{P}(\Omega)$ is the class of all subsets of Ω.

Definition 1.1 (Set sequence)
A *set sequence* is a map

$$\begin{gathered} I\!N \to \mathcal{P}(\Omega) \\ n \rightsquigarrow A_n \end{gathered}$$

We represent it by $\{A_n\}_{n\in I\!N} \in \mathcal{P}(\Omega)$.

Theorem 1.1 (The De Morgan laws)
It holds that

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

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

Definition 1.2 (Monotone set sequence)
A set sequence $\{A_n\}_{n\in I\!N} \in \mathcal{P}(\Omega)$ is said to be *monotone increasing* if and only if $A_n \subseteq A_{n+1}$, $\forall n \in I\!N$. We represent it by $\{A_n\} \uparrow$.

When $A_n \supseteq A_{n+1}$, $\forall n \in I\!N$, the sequence is said to be *monotone decreasing*, and we represent it by $\{A_n\} \downarrow$.

Example 1.1 Consider the sequences defined by:

(i) $A_n = (-n, n)$, $\forall n \in I\!N$. This sequence is monotone increasing, since $\forall n \in I\!N$,

$$A_n = (-n, n) \subset (-(n+1), n+1) = A_{n+1}.$$

(ii) $B_n = (-1/n, 1 + 1/n)$, $\forall n \in I\!N$. This sequence is monotone decreasing, since $\forall n \in I\!N$,

$$B_n = (-1/n, 1+1/n) \supset (-1/(n+1), 1+1/(n+1)) = B_{n+1}.$$

Definition 1.3 (Limit of a set sequence)

(i) We call *lower limit* of $\{A_n\}$, and we denote it $\underline{\lim}\, A_n$, to the set of points of Ω that belong to all A_ns except for a finite number of them.

(ii) We call *upper limit* of $\{A_n\}$, and we denote it $\overline{\lim}\, A_n$, to the set of points of Ω that belong to infinite number of A_ns. It is also said that A_n occurs *infinitely often* (i.o.), and it is denoted also $\overline{\lim}\, A_n = A_n$ i.o.

Example 1.2 If $\omega \in A_{2n}$, $\forall n \in I\!N$, then $\omega \in \overline{\lim}\, A_n$ but $\omega \notin \underline{\lim}\, A_n$ since there is an infinite number of A_ns to which ω does not belong, $\{A_{2n-1}\}_{n \in I\!N}$.

Proposition 1.1 (Another characterization of limit of a set sequence)

(i) $\underline{\lim}\, A_n = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n$

(ii) $\overline{\lim} A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$

Proposition 1.2 (Relation between lower and upper limits)
The lower and upper limits of a set sequence $\{A_n\}$ satisfy

$$\underline{\lim} A_n \subseteq \overline{\lim} A_n$$

Definition 1.4 (Convergence)
A set sequence $\{A_n\}$ *converges* if and only if

$$\underline{\lim} A_n = \overline{\lim} A_n.$$

Then, we call limit of $\{A_n\}$ to

$$\lim_{n\to\infty} A_n = \underline{\lim} A_n = \overline{\lim} A_n.$$

Definition 1.5 (Inferior/Superior limit of a sequence of real numbers)
Let $\{a_n\}_{n\in\mathbb{N}} \in \mathbb{R}$ be a sequence. We define:

(i) $\liminf_{n\to\infty} a_n = \sup_k \inf_{n\geq k} a_n$;

(ii) $\limsup_{n\to\infty} a_n = \inf_k \sup_{n\geq k} a_n$.

Proposition 1.3 (Convergence of monotone set sequences)
Any monotone (increasing of decreasing) set sequence converges, and it holds:

(i) If $\{A_n\} \uparrow$, then $\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$.

(ii) If $\{A_n\} \downarrow$, then $\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n$.

Example 1.3 Obtain the limits of the following set sequences:

(i) $\{A_n\}$, where $A_n = (-n, n)$, $\forall n \in I\!N$.

(ii) $\{B_n\}$, where $B_n = (-1/n, 1 + 1/n)$, $\forall n \in I\!N$.

(i) By the previous proposition, since $\{A_n\} \uparrow$, then

$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} (-n, n) = I\!R.$$

(ii) Again, using the previous proposition, since $\{A_n\} \downarrow$, then

$$\lim_{n\to\infty} B_n = \bigcap_{n=1}^{\infty} B_n = \bigcap_{n=1}^{\infty} \left(-\frac{1}{n}, 1 + \frac{1}{n}\right) = [0, 1].$$

Problems

1. Prove Proposition 1.1.
2. Define sets of real numbers as follows. Let $A_n = (-1/n, 1]$ if n is odd, and $A_n = (-1, 1/n]$ if n is even. Find $\underline{\lim}\, A_n$ and $\overline{\lim}\, A_n$.
3. Prove Proposition 1.2.
4. Prove Proposition 1.3.
5. Let $\Omega = I\!R^2$ and A_n the interior of the circle with center at the point $((-1)^n/n, 0)$ and radius 1. Find $\underline{\lim}\, A_n$ and $\overline{\lim}\, A_n$.
6. Prove that $(\underline{\lim}\, A_n)^c = \overline{\lim}\, A_n^c$ and $(\overline{\lim}\, A_n)^c = \underline{\lim}\, A_n^c$.
7. Using the De Morgan laws and Proposition 1.3, prove that if $\{A_n\} \uparrow A$, then $\{A_n^c\} \downarrow A^c$ while if $\{A_n\} \downarrow A$, then $\{A_n^c\} \uparrow A^c$.

8. Let $\{x_n\}$ be a sequence of real numbers and let $A_n = (-\infty, x_n)$. What is the connection between $\liminf_{n\to\infty} x_n$ and $\underline{\lim}\, A_n$? Similarly between $\limsup_{n\to\infty} x_n$ and $\overline{\lim}\, A_n$.

1.2 Structures of subsets

A probability function will be a function defined over events or subsets of a sample space Ω. It is convenient to provide a "good" structure to these subsets, which in turn will provide "good" properties to the probability function. In this section we study collections of subsets of a set Ω with a good structure.

Definition 1.6 (Algebra)
An *algebra* (also called *field*) $\mathcal{A}$ over a set Ω is a collection of subsets of Ω that has the following properties:

(i) $\Omega \in \mathcal{A}$;

(ii) If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$;

(iii) If $A_1, A_2, \ldots, A_n \in \mathcal{A}$, then $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$.

An algebra over Ω contains both Ω and $\emptyset$. It also contains all finite unions and intersections of sets from $\mathcal{A}$. We say that $\mathcal{A}$ is closed under complementation, finite union and finite intersection. Extending property (iii) to an infinite sequence of elements of $\mathcal{A}$ we obtain a σ-algebra.

Definition 1.7 (σ-algebra)
A σ-algebra (or σ-field) $\mathcal{A}$ over a set Ω is a collection of subsets of Ω that has the following properties:

(i) $\Omega \in \mathcal{A}$;

(ii) If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$;

(iii) If $A_1, A_2 \ldots$ is a sequence of elements of $\mathcal{A}$, then $\cup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Thus, a σ-algebra is closed under countable union. It is also closed under countable intersection. Moreover, if $\mathcal{A}$ is an algebra, a countable union of sets in $\mathcal{A}$ can be expressed as the limit of an increasing sequence of sets, the finite unions $\bigcup_{i=1}^{n} A_i$. Thus, a σ-algebra is an algebra that is closed under limits of increasing sequences.

Example 1.4 The smallest σ-algebra is $\{\emptyset, \Omega\}$. The smallest σ-algebra that contains a subset $A \subset \Omega$ is $\{\emptyset, A, A^c, \Omega\}$. It is contained in any other σ-algebra containing A. The collection of all subsets of Ω, $\mathcal{P}(\Omega)$, is a well known algebra called the *algebra of the parts* of Ω.

Definition 1.8 (σ-algebra spanned by a collection $\mathcal{C}$ of events)
Given a collection of sets $\mathcal{C} \subset \mathcal{P}(\Omega)$, we define the σ-algebra spanned by $\mathcal{C}$, and we denote it by $\sigma(\mathcal{C})$, as the smallest σ-algebra that contains $\mathcal{C}$.

Remark 1.1 For each $\mathcal{C}$, the σ-algebra spanned by $\mathcal{C}$, $\sigma(\mathcal{C})$, always exists, since $\mathcal{A}_{\mathcal{C}}$ is the intersection of all σ-algebras that contain $\mathcal{C}$ and at least $\mathcal{P}(\Omega) \supset \mathcal{C}$ is a σ-algebra.

When Ω is finite or countable, it is common to work with the σ-algebra $\mathcal{P}(\Omega)$, so we will use this one unless otherwise stated. In the case $\Omega = I\!R$, later we will consider probability measures and we will want to obtain probabilities of intervals. Thus, we need a

σ-algebra containing all intervals. The Borel σ-algebra is based on this idea, and it will be used by default when $\Omega = I\!\!R$.

Definition 1.9 (Borel σ-algebra)
Consider the sample space $\Omega = I\!\!R$ and the collection of intervals of the form

$$I = \{(-\infty, a] : \; a \in I\!\!R\}.$$

We define the *Borel σ-algebra* over $I\!\!R$, represented by $\mathcal{B}$, as the σ-algebra spanned by I.

The Borel σ-algebra $\mathcal{B}$ contains all complements, countable intersections and unions of elements of I. In particular, $\mathcal{B}$ contains all types of intervals and isolated points of $I\!\!R$, although $\mathcal{B}$ is not equal to $\mathcal{P}(I\!\!R)$. For example,

- $(a, \infty) \in \mathcal{B}$, since $(a, \infty) = (-\infty, a]^c$, and $(-\infty, a] \in I\!\!R$.
- $(a, b] \in I\!\!R$, $\forall a < b$, since this interval can be expressed as $(a, b] = (-\infty, b] \cap (a, \infty)$, where $(-\infty, b] \in \mathcal{B}$ and $(a, \infty) \in \mathcal{B}$.
- $\{a\} \in \mathcal{B}, \forall a \in I\!\!R$, since $\{a\} = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, a\right]$, which belongs to $\mathcal{B}$.

When the sample space Ω is continuous but is a subset of $I\!\!R$, we need a σ-algebra restricted to subsets of Ω.

Definition 1.10 (Restricted Borel σ-algebra)
Let $A \subset I\!\!R$. We define the *Borel σ-algebra restricted to A* as the collection

$$\mathcal{B}_A = \{B \cap A : \; B \in \mathcal{B}\}.$$

In the following we define the space over which measures, including probability measures, will be defined. This space will be the one whose elements will be suitable to "measure".

Definition 1.11 (Measure space)
The pair $(\Omega, \mathcal{A})$, where Ω is a sample space and $\mathcal{A}$ is an algebra over Ω, is called *measurable space*.

Problems

1. Let $A, B \in \Omega$ with $A \cap B = \emptyset$. Construct the smallest σ-algebra that contains A and B.

2. Prove that an algebra over Ω contains all finite intersections of sets from $\mathcal{A}$.

3. Prove that a σ-algebra over Ω contains all countable intersections of sets from $\mathcal{A}$.

4. Prove that a σ-algebra is an algebra that is closed under limits of increasing sequences.

5. Let $\Omega = I\!R$. Let $\mathcal{A}$ be the class of all finite unions of disjoint elements from the set

$$\mathcal{C} = \{(a, b], (-\infty, a], (b, \infty); a \leq b\}$$

Prove that $\mathcal{A}$ is an algebra.

1.3 Set functions

In the following $\mathcal{A}$ is an algebra and we consider the extended real line given by $\overline{I\!R} = I\!R \cup \{-\infty, +\infty\}$.

Definition 1.12 (Additive set function)
A set function $\phi : \mathcal{A} \to \overline{I\!R}$ is *additive* if it satisfies:

$$\text{For all } \{A_i\}_{i=1}^{n} \in \mathcal{A} \text{ with } A_i \cap A_j = \emptyset,\ i \neq j,\ \phi\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \phi(A_i).$$

We will assume that $+\infty$ and $-\infty$ cannot both belong to the range of ϕ. We will exclude the cases $\phi(A) = +\infty$ for all $A \in \mathcal{A}$ and $\phi(A) = -\infty$ for all $A \in \mathcal{A}$. Extending the definition to an infinite sequence, we obtain a σ-additive set function.

Definition 1.13 (σ-additive set function)
A set function $\phi : \mathcal{A} \to \overline{I\!R}$ is σ-*additive* if it satisfies:

$$\text{For all } \{A_n\}_{n \in I\!N} \in \mathcal{A} \text{ with } A_i \cap A_j = \emptyset,\ i \neq j,$$

$$\phi\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \phi(A_n).$$

Observe that a σ-additive set function is well defined, since the infinite union of sets of $\mathcal{A}$ belongs to $\mathcal{A}$ because $\mathcal{A}$ is a σ-algebra. It is easy to see that an additive function satisfies $\mu(\emptyset) = 0$. Moreover, countable additivity implies finite additivity.

Definition 1.14 (Measure)
A set function $\phi : \mathcal{A} \to \overline{I\!R}$ is a *measure* if

(a) ϕ is σ-additive;

(b) $\phi(A) \geq 0, \forall A \in \mathcal{A}$.

Definition 1.15 (Probability measure)
A measure μ with $\mu(\Omega) = 1$ is called a *probability measure*.

Example 1.5 (Counting measure)
Let Ω be any set and consider the σ-algebra of the parts of Ω, $\mathcal{P}(\Omega)$. Define $\mu(A)$ as the number of points of A. The set function μ is a measure known as the *counting measure*.

Example 1.6 (Probability measure)
Let $\Omega = \{x_1, x_2, \ldots\}$ be a finite or countably infinite set, and let $p_1, p_2, \ldots,$ be nonnegative numbers. Consider the σ-algebra of the parts of Ω, $\mathcal{P}(\Omega)$, and define

$$\mu(A) = \sum_{x_i \in A} p_i.$$

The set function μ is a probability measure if and only if $\sum_{i=1}^{\infty} p_i = 1$.

Example 1.7 (Lebesgue measure)
A well known measure defined over $(\mathbb{R}, \mathcal{B})$, which assigns to each element of $\mathcal{B}$ its length, is the *Lebesgue measure*, denoted here as λ. For an interval, either open, close or semiclosed, the Lebesgue measure is the length of the interval. For a single point, the Lebesgue measure is zero.

Definition 1.16 (σ-finite set function)
A set function $\phi : \mathcal{A} \to \mathbb{R}$ is *σ-finite* if $\forall A \in \mathcal{A}$, there exists a sequence $\{A_n\}_{n \in \mathbb{N}}$ of disjoint elements of $\mathcal{A}$ with $\phi(A_n) < \infty \ \forall n$, whose union covers A, that is,

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

Definition 1.17 (Measure space)
The triplet $(\Omega, \mathcal{A}, \mu)$, where Ω is a sample space, $\mathcal{A}$ is an algebra and μ is a measure defined over $(\Omega, \mathcal{A})$, is called *measure space*.

Definition 1.18 (Absolutely continuous measure with respect to another)
A measure μ on Borel subsets of the real line is *absolutely continuous with respect to* another measure λ if $\lambda(A) = 0$ implies that $\mu(A) = 0$. It is also said that μ is dominated by λ, and written as $\mu << \lambda$.

If a measure on the real line is simply said to be absolutely continuous, this typically means absolute continuity with respect to Lebesgue measure.

Although a probability function is simply a measure μ satisfying $\mu(\Omega) = 1$ as mentioned above, in Section 1.4 we give the classical definition of probability through the axioms of Kolmogoroff.

Problems

1. Prove that for any finitely additive set function μ defined on an algebra $\mathcal{A}$,
$$\mu(\emptyset) = 0.$$
2. Prove that for any finitely additive set function μ defined on an algebra $\mathcal{A}$,
$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B).$$
3. Prove that for any finitely additive set function μ defined on an algebra $\mathcal{A}$, if $B \subseteq A$, then
$$\mu(A) = \mu(B) + \mu(A - B).$$
4. Prove that for any nonnegative finitely additive set function μ

defined on an algebra $\mathcal{A}$, then for all $A_1, \ldots, A_n \in \mathcal{A}$,

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

5. Prove that for any measure μ defined on an algebra $\mathcal{A}$, then for all $A_1, \ldots, A_n \in \mathcal{A}$ such that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$,

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

1.4 Probability measures

Definition 1.19 (Random experiment)
A *random experiment* is a process for which:

- the set of possible results is known;
- its result cannot be predicted without error;
- if we repeat it in identical conditions, the result can be different.

Definition 1.20 (Elementary event, sample space, event)
The possible results of the random experiment that are indivisible are called *elementary events*. The set of elementary events is known as *sample space*, and it will be denoted Ω. An *event* A is a subset of Ω, such that once the random experiment is carried out, we can say that A "has occurred" if the result of the experiment is contained in A.

Example 1.8 Examples of random experiments are:

(a) Tossing a coin. The sample space is $\Omega = \{$"*head*", "*tail*"$\}$. Events are: $\emptyset$, $\{$"*head*"$\}$, $\{$"*tail*"$\}$, Ω.

(b) Observing the number of traffic accidents in a minute in Spain. The sample space is $\Omega = I\!N \cup \{0\}$.

(c) Drawing a Spanish woman aged between 20 and 40 and measuring her weight (in kgs.). The sample space is $\Omega = [m, \infty)$, were m is the minimum possible weight.

We will require that the collection of events has a structure of σ-algebra. This will make possible to obtain the probability of all complements, unions and intersections of events. The probabilities will be set functions defined over a measurable space composed by the sample space Ω and a σ-algebra of subsets of Ω.

Example 1.9 For the experiment (a) described in Example 1.8, a measurable space is:

$$\Omega = \{\text{“}head\text{”}, \text{“}tail\text{”}\}, \quad \mathcal{A} = \{\emptyset, \{\text{“}head\text{”}\}, \{\text{“}tail\text{”}\}, \Omega\}.$$

For the experiment (b), the sample space is $\Omega = I\!N \cup \{0\}$. If we take the σ-algebra $\mathcal{P}(\Omega)$, then $(\Omega, \mathcal{P}(\Omega))$ is a measurable space. Finally, for the experiment (c), with sample space $\Omega = [m, \infty) \subset I\!R$, a suitable σ-algebra is the Borel σ-algebra restricted to Ω, $\mathcal{B}_{[m,\infty)}$.

Definition 1.21 (Axiomatic definition of probability by Kolmogoroff)
Let $(\Omega, \mathcal{A})$ be a measurable space, where Ω is a sample space and $\mathcal{A}$ is a σ-algebra over Ω. A *probability function* is a set function $P : \mathcal{A} \to [0, 1]$ that satisfies the following axioms:

(i) $P(\Omega) = 1$;

(ii) For any sequence $A_1, A_2, \ldots$ of disjoint elements of $\mathcal{A}$, it holds

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

By axiom (ii), a probability function is a measure for which the measure of the sample space Ω is 1. The triplet $(\Omega, \mathcal{A}, P)$, where P is a probability P, is called *probability space*.

Example 1.10 For the experiment (a) described in Example 1.8, with the measurable space $(\Omega, \mathcal{A})$, where

$$\Omega = \{\text{``}head\text{''}, \text{``}tail\text{''}\}, \quad \mathcal{A} = \{\emptyset, \{\text{``}head\text{''}\}, \{\text{``}tail\text{''}\}, \Omega\},$$

define

$$P_1(\emptyset) = 0,\ P_1(\{\text{``}head\text{''}\}) = p,\ P_1(\{\text{``}tail\text{''}\}) = 1 - p,\ P_1(\Omega) = 1,$$

where $p \in [0, 1]$. This function verifies the axioms of Kolmogoroff.

Example 1.11 For the experiment (b) described is Example 1.8, with the measurable space $(\Omega, \mathcal{P}(\Omega))$, define:

- For the elementary events, the probability is

$$\begin{array}{llll} P(\{0\}) = 0.131, & P(\{1\}) = 0.272, & P(\{2\}) = 0.27, & P(\{3\}) = 0.183, \\ P(\{4\}) = 0.09, & P(\{5\}) = 0.012, & P(\{6\}) = 0.00095, & P(\{7\}) = 0.00005, \end{array}$$

$$P(\emptyset) = 0, \quad P(\{i\}) = 0,\ \forall i \geq 8.$$

- For other events, the probability is defined as the sum of the probabilities of the elementary events contains in that event, that is, if $A = \{a_1, \ldots, a_n\}$, where $a_i \in \Omega$ are the elementary events, the probability of A is

$$P(A) = \sum_{i=1}^{n} P(\{a_i\}).$$

This function verifies the axioms of Kolmogoroff.

Proposition 1.4 (Properties of the probability)
The following properties are consequence of the axioms of Kolmogoroff:

(i) $P(\emptyset) = 0$;

(ii) Let $A_1, A_2, \ldots, A_n \in \mathcal{A}$ with $A_i \cap A_j = \emptyset$, $i \neq j$. Then,

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

(iii) $\forall A \in \mathcal{A}$, $P(A) \leq 1$.

(iv) $\forall A \in \mathcal{A}$, $P(A^c) = 1 - P(A)$.

(v) For $A, B \in \mathcal{A}$ with $A \subseteq B$, it holds $P(A) \leq P(B)$.

(vi) Let $A, B \in \mathcal{A}$ be two events. Then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

(vii) Let $A_1, A_2, \ldots, A_n \in \mathcal{A}$ be events. Then

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{i=1}^{n} P(A_i) - \sum_{\substack{i_1,i_2=1 \\ i_1<i_2}}^{n} P(A_{i_1} \cap A_{i_2}) \\ &+ \sum_{\substack{i_1,i_2,i_3=1 \\ i_1<i_2<i_3}}^{n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) + \cdots \\ &+ (-1)^{n-1} P(A_1 \cap A_2 \cap \cdots \cap A_n). \end{aligned}$$

In particular, for $n = 2$ we get property (vi).

Proposition 1.5 (Bool's inequality)
For any sequence $\{A_n\} \subset \mathcal{A}$ it holds

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n).$$

Proposition 1.6 (Sequential continuity of the probability)
Let $(\Omega, \mathcal{A}, P)$ be a probability space. Then, for any sequence $\{A_n\}$ of events from $\mathcal{A}$ it holds

$$P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n).$$

Example 1.12 Consider the random experiment of selecting randomly a number from $[0,1]$. Then the sample space is $\Omega = [0,1]$. Consider also the Borel σ-algebra restricted to $[0,1]$ and define the function

$$g(x) = P([0,x]), \quad x \in (0,1).$$

Proof that g is always right continuous, for each probability measure P that we choose.

Example 1.13 (Construction of a probability measure for countable Ω)
If Ω is finite or countable, the σ-algebra that is typically chosen is $\mathcal{P}(\Omega)$. In this case, in order to define a probability function, it suffices to define the probabilities of the elementary events $\{a_i\}$ as $P(\{a_i\}) = p_i$, $\forall a_i \in \Omega$ with the condition that $\sum_i p_i = 1$, $p_i \geq 0$, $\forall i$. Then, $\forall A \subset \Omega$,

$$P(A) = \sum_{a_i \in A} P(\{a_i\}) = \sum_{a_i \in A} p_i.$$

Example 1.14 (Construction of a probability measure in $(I\!\!R, \mathcal{B})$)
How can we construct a probability measure in $(I\!\!R, \mathcal{B})$? In general, it is not possible to define a probability measure by assigning directly a numerical value to each $A \in \mathcal{B}$, since then probably the axioms of Kolmogoroff will not be satisfied.

For this, we will first consider the collection of intervals

$$\mathcal{C} = \{(a,b], (-\infty, b], (a, +\infty) : a < b\}. \tag{1.1}$$

We start by assigning values of P to intervals from $\mathcal{C}$, by ensuring that P is σ-additive on $\mathcal{C}$. Then, we consider the algebra $\mathcal{F}$ obtained by doing finite unions of disjoint intervals from $\mathcal{C}$ and we extend P to $\mathcal{F}$. The extended function will be a probability measure on $(\mathbb{R}, \mathcal{F})$. Finally, there is a unique extension of a probability measure from $\mathcal{F}$ to $\sigma(\mathcal{F}) = \mathcal{B}$, see the following propositions.

Proposition 1.7 Consider the collection of all finite unions of disjoint intervals from $\mathcal{C}$ in (1.1),

$$\mathcal{F} = \left\{ \bigcup_{i=1}^{n} A_i : A_i \in \mathcal{C},\ A_i \text{ disjoint} \right\}.$$

Then $\mathcal{F}$ is an algebra.

Next we extend P from $\mathcal{C}$ to $\mathcal{F}$ as follows.

Proposition 1.8 (Extension of the probability function)

(a) For all $A \in \mathcal{F}$, since $A = \bigcup_{i=1}^{n} (a_i, b_i]$, with $a_i, b_i \in \mathbb{R} \cup \{-\infty, +\infty\}$, let us define

$$P_1(A) = \sum_{i=1}^{n} P(a_i, b_i].$$

Then, P_1 is a probability measure over $(\mathbb{R}, \mathcal{F})$.

(b) For all $A \in \mathcal{C}$, it holds that $P(A) = P_1(A)$.

Observe that $\mathcal{B} = \sigma(\mathcal{F})$. Finally, can we extend P from $\mathcal{F}$ to $\mathcal{B} = \sigma(\mathcal{F})$? If the answer is positive, is the extension unique? The next theorem gives the answer to these two questions.

Theorem 1.2 (Caratheodory's Extension Theorem) Let $(\Omega, \mathcal{A}, P)$ be a probability space, where $\mathcal{A}$ is an algebra. Then, P can be extended from $\mathcal{A}$ to $\sigma(\mathcal{A})$, and the extension is unique (i.e., there exists a unique probability measure $\hat{P}$ over $\sigma(\mathcal{A})$ with $\hat{P}(A) = P(A)$, $\forall A \in \mathcal{A}$).

The extension of P from $\mathcal{F}$ to $\sigma(\mathcal{F}) = \mathcal{B}$ is done by steps. First, P is extended to the collection of the limits of increasing sequences of events in $\mathcal{F}$, denoted $\mathcal{C}$. It holds that $\mathcal{C} \supset \mathcal{F}$ and $\mathcal{C} \supset \sigma(\mathcal{F}) = \mathcal{B}$ (Monotone Class Theorem). The probability of each event A from $\mathcal{C}$ is defined as the limit of the probabilities of the sequences of events from $\mathcal{F}$ that converge to A. Afterwards, P is extended to the σ-algebra of the parts of $I\!R$. For each subset $A \in \mathcal{P}(I\!R)$, the probability is defined as the infimum of the probabilities of the events in $\mathcal{C}$ that contain A. This extension is not countably additive on $\mathcal{P}(I\!R)$, only on a smaller σ-algebra, so P it is not a probability measure on $(I\!R, \mathcal{P}(I\!R))$. Finally, a σ-algebra in which P is a probability measure is defined as the collection $\mathcal{H}$ of subsets $H \subset \Omega$, for which $P(H) + P(H^c) = 1$. This collection is indeed a σ-algebra that contains $\mathcal{C}$ and P is a probability measure on it. It holds that $\sigma(\mathcal{F}) = \mathcal{B} \subset \mathcal{H}$ and P restricted to $\sigma(\mathcal{F}) = \mathcal{B}$ is also a probability measure on $(I\!R, \mathcal{B})$.

Problems

1. Prove the properties of the probability measures in Proposition 1.4.
2. Prove Bool's inequality in Proposition 1.4.
3. Prove the Sequential Continuity of the probability in Proposition 1.6.

1.5 Other definitions of probability

When Ω is finite, say $\Omega = \{a_1, \ldots, a_k\}$, many times the elementary events are equiprobable, that is, $P(\{a_1\}) = \cdots = P(\{a_k\}) = 1/k$. Then, for $A \subset \Omega$, say $A = \{a_{i_1}, \ldots, a_{i_m}\}$, then

$$P(A) = \sum_{j=1}^{m} P(\{a_{i_j}\}) = \frac{m}{k}.$$

This is the definition of probability given by Laplace, which is useful only for experiments with a finite number of possible results and whose results are, a priori, equally frequent.

Definition 1.22 (Laplace rule of probability)
The *Laplace probability* of an event $A \subseteq \Omega$ is the proportion of results favorable to A; that is, if k is the number of possible results or cardinal of Ω and $k(A)$ is the number of results contained in A or cardinal of A, then

$$P(A) = \frac{k(A)}{k}.$$

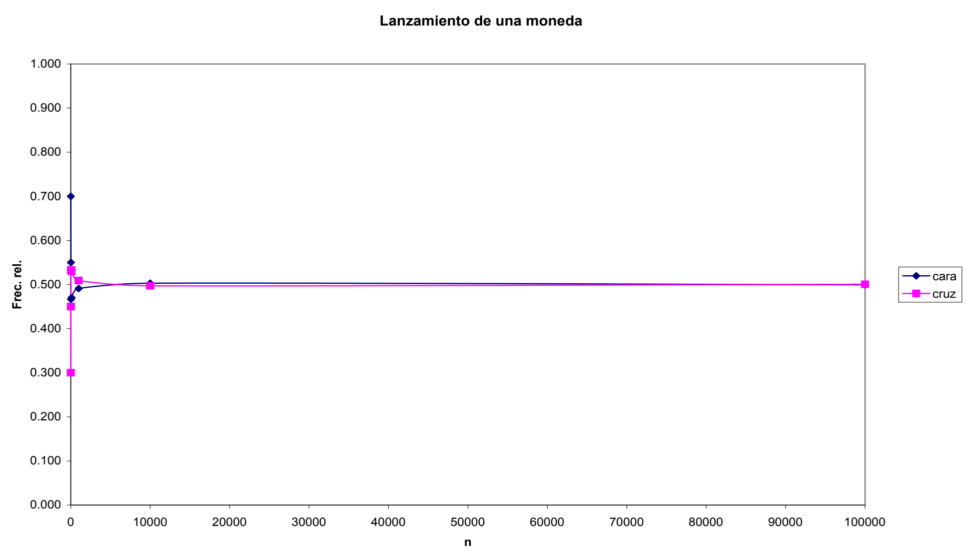
In order to apply the Laplace rule, we need to learn to count. The counting techniques are comprised in the area of Combinatorial Analysis.

The following examples show intuitively the frequentist definition of probability.

Example 1.15 (Frequentist definition of probability)
The following tables report the relative frequencies of the results of the experiments described in Example 1.8, when each of these experiments are repeated n times.

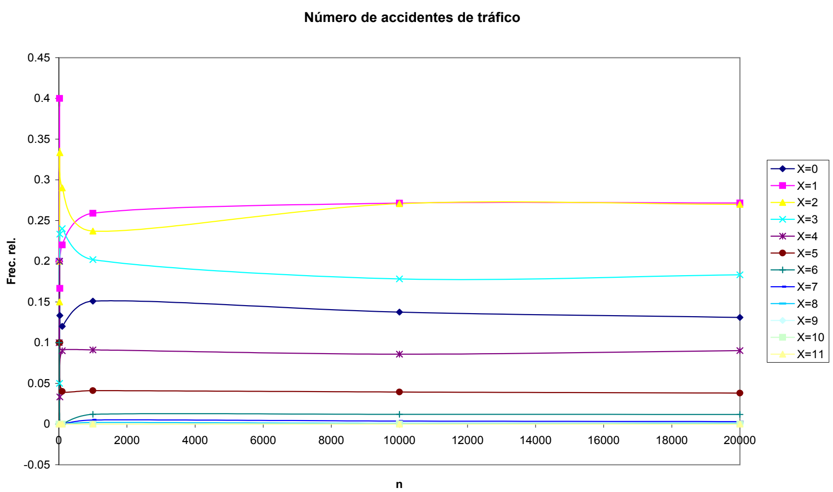
(a) Tossing a coin n times. The table shows that both frequencies of “head” and “tail” converge to 0.5.

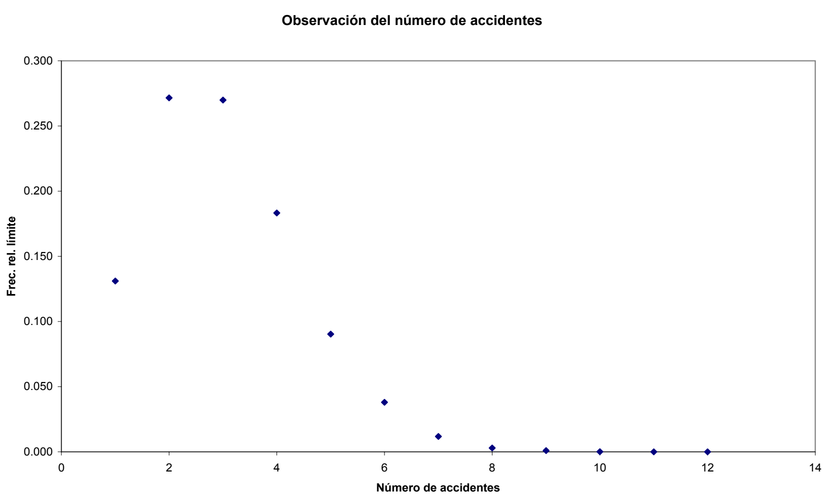
n	Result "head"	"tail"
10	0.700	0.300
20	0.550	0.450
30	0.467	0.533
100	0.470	0.530
1000	0.491	0.509
10000	0.503	0.497
100000	0.500	0.500



(b) Observation of the number of traffic accidents in n minutes. We can observe in the table below that the frequencies of the possible results of the experiment seem to converge.

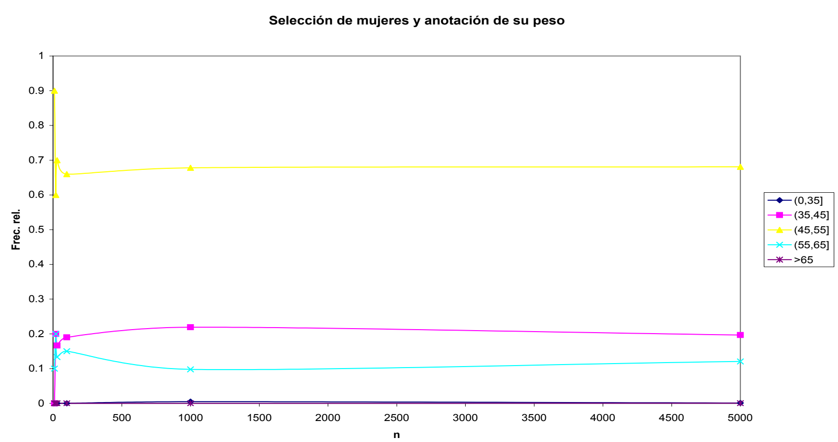
	Result								
n	0	1	2	3	4	5	6	7	8
10	0.1	0.2	0.2	0.2	0.1	0.1	0.1	0	0
20	0.2	0.4	0.15	0.05	0.2	0	0	0	0
30	0.13	0.17	0.33	0.23	0.03	0	0	0	0
100	0.12	0.22	0.29	0.24	0.09	0	0	0	0
1000	0.151	0.259	0.237	0.202	0.091	0.012	0.002	0	0
10000	0.138	0.271	0.271	0.178	0.086	0.012	0.0008	0.0001	0
20000	0.131	0.272	0.270	0.183	0.090	0.012	0.00095	0.00005	0

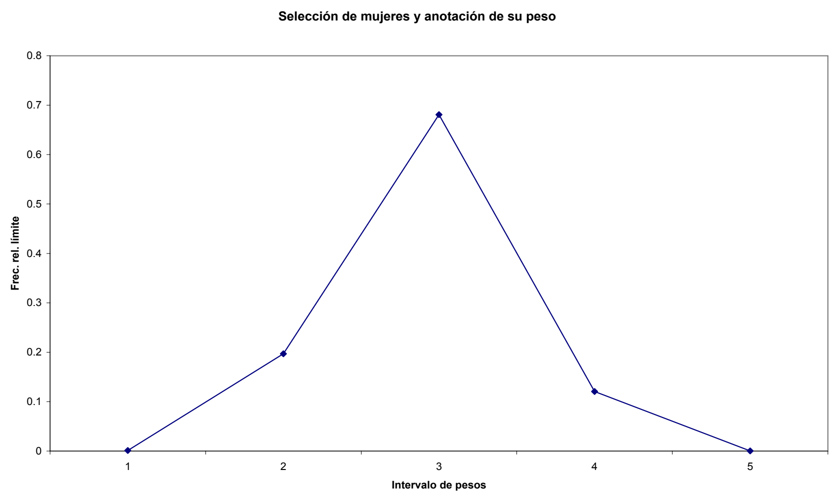




(c) Independient drawing of n women aged between 20 and 40 and measuring their weight (in kgs.). Again, we observe that the relative frequencies of the given weight intervals seem to converge.

n	Weight intervals $(0,35]$	$(35,45]$	$(45,55]$	$(55,65]$	$(65,\infty)$
10	0	0	0.9	0.1	0
20	0	0.2	0.6	0.2	0
30	0	0.17	0.7	0.13	0
100	0	0.19	0.66	0.15	0
1000	0.005	0.219	0.678	0.098	0
5000	0.0012	0.197	0.681	0.121	0.0004





Definition 1.23 (Frequentist probability)
The *frequentist definition of probability* of an event A is the limit of the relative frequency of this event, when we let the number of repetitions of the random experiment grow to infinity.

If the experiment is repeated n times, and n_A is the number of repetitions in which A occurs, then the probability of A is

$$P(A) = \lim_{n\to\infty} \frac{n_A}{n}\,.$$

Problems

1. Check if the Laplace definition of probability satisfies the axioms of Kolmogoroff.

2. Check if the frequentist definition of probability satisfies the axioms of Kolmogoroff.

1.6 Measurability and Lebesgue integral

A measurable function relates two measurable spaces, preserving the structure of the events.

Definition 1.24 Let $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ be two measurable spaces. A function $f : \Omega_1 \to \Omega_2$ is said to be *measurable* if and only if $\forall B \in \mathcal{A}_2$, $f^{-1}(B) \in \mathcal{A}_1$, where $f^{-1}(B) = \{\omega \in \Omega_1 : f(\omega) \in B\}$.

The sum, product, quotient (when the function in the denominator is different from zero), maximum, minimum and composition of two measurable functions is a measurable function. Moreover, if $\{f_n\}_{n\in I\!N}$ is a sequence of measurable functions, then

$$\sup_{n\in I\!N}\{f_n\}, \quad \inf_{n\in I\!N}\{f_n\}, \quad \liminf_{n\in I\!N} f_n, \quad \limsup_{n\in I\!N} f_n, \quad \lim_{n\to\infty} f_n,$$

assuming that they exist, are also measurable. If they are infinite, we can consider $\bar{I\!R}$ instead of $I\!R$.

The following result will gives us a tool useful to check if a function f from $(\Omega_1, \mathcal{A}_1)$ into $(\Omega_2, \mathcal{A}_2)$ is measurable.

Theorem 1.3 Let $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ be measure spaces and let $f : \Omega_1 \to \Omega_2$. Let $C_2 \subset \mathcal{P}(\Omega_2)$ be a collection of subsets that generates A_2, i.e, such that $\sigma(\mathcal{C}_2) = \mathcal{A}_2$. Then f is measurable if and only of $f^{-1}(A) \in \mathcal{A}_1$, $\forall A \in \mathcal{C}_2$.

Corolary 1.1 Let $(\Omega, \mathcal{A})$ be a measure space and $f : \Omega \to I\!R$ a function. Then f is measurable if and only if $f^{-1}(-\infty, a] \in \mathcal{A}$, $\forall a \in I\!R$.

The Lebesgue integral is restricted to measurable functions. We are going to define the integral by steps. We consider measurable functions defined from a measurable space $(\Omega, \mathcal{A})$ on the measurable space $(I\!R, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra. We consider also a σ-finite measure μ.

Definition 1.25 (Indicator function)
Given $S \in \mathcal{A}$, an indicator function, $\mathbf{1}_S : \Omega \to I\!R$, gives value 1 to elements of S and 0 to the rest of elements:

$$\mathbf{1}_S(\omega) = \begin{cases} 1, & \omega \in S; \\ 0, & \omega \notin S. \end{cases}$$

Next we define simple functions, which are linear combinations of indicator functions.

Definition 1.26 (Simple function)
Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. Let a_i be real numbers and $\{S_i\}_{i=1}^n$ disjoint elements of $\mathcal{A}$. A *simple function* has the form

$$\phi = \sum_{i=1}^{n} a_i \mathbf{1}_{S_i}.$$

Proposition 1.9 Indicators and simple functions are measurable.

Definition 1.27 (Lebesgue integral for simple functions)

(i) The *Lebesgue integral* of a simple function ϕ with respect to a σ-finite measure μ is defined as

$$\int_\Omega \phi \, d\mu = \sum_{i=1}^{n} a_i \, \mu(S_i).$$

(ii) The *Lebesgue integral* of ϕ with respect to μ over a subset $A \in \mathcal{A}$ is

$$\int_A \phi \, d\mu = \int_A \phi \cdot \mathbf{1}_A \, d\mu = \sum_{i=1}^{n} a_i \, \mu(A \cap S_i).$$

The next theorem says that for any measurable function f on $I\!R$, we can always find a sequence of measurable functions that converge to f. This will allow the definition of the Lebesgue integral.

Theorem 1.4 Let $f : \Omega \to I\!R$. It holds:

(a) f is a positive measurable function if and only if $f = \lim_{n\to\infty} f_n$, where $\{f_n\}_{n\in I\!N}$ is an increasing sequence of non negative simple functions.

(b) f is a measurable function if and only if $f = \lim_{n\to\infty} f_n$, where $\{f_n\}_{n\in I\!N}$ is an increasing sequence of simple functions.

Definition 1.28 (Lebesgue integral for non-negative functions)
Let f be a non-negative measurable function defined over $(\Omega, \mathcal{A}, \mu)$ and $\{f_n\}_{n\in I\!N}$ be an increasing sequence of simple functions that converge pointwise to f (this sequence can be always constructed). The *Lebesgue integral* of f with respect to the σ-finite measure μ is defined as

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

The previous definition is correct due to the following uniqueness theorem.

Theorem 1.5 (Uniqueness of Lebesgue integral)
Let f be a non-negative measurable function. Let $\{f_n\}_{n\in I\!N}$ and $\{g_n\}_{n\in I\!N}$ be two increasing sequences of non-negative simple functions that converge pointwise to f. Then

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

Definition 1.29 (Lebesgue integral for general functions) For a measurable function f that can take negative values, we can write it as the sum of two non-negative functions in the form:

$$f = f^+ - f^-,$$

where $f^+(\omega) = \max\{f(\omega), 0\}$ is the *positive part* of f and $f^-(\omega) = \max\{-f(\omega), 0\}$ is the *negative part* of f. If the integrals of f^+ and f^- are finite, then the *Lebesgue integral* of f is

$$\int_\Omega f\, d\mu = \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu,$$

assuming that at least one of the integrals on the right is finite.

Definition 1.30 The *Lebesgue integral* of a measurable function f over a subset $A \in \mathcal{A}$ is defined as

$$\int_A f\, d\mu = \int_\Omega f \mathbf{1}_A d\, \mu.$$

Definition 1.31 A function is said to be *Lebesgue integrable* if and only if

$$\left| \int_\Omega f\, d\mu \right| < \infty.$$

Moreover, if instead of doing the decomposition $f = f^+ - f^-$ we do another decomposition, the result is the same.

Theorem 1.6 Let f_1, f_2, g_1, g_2 be non-negative measurable functions and let $f = f_1 - f_2 = g_1 - g_2$. Then,

$$\int_\Omega f_1\, d\mu - \int_\Omega f_2\, d\mu = \int_\Omega g_1\, d\mu - \int_\Omega g_2\, d\mu.$$

Proposition 1.10 A measurable function f is Lebesgue integrable if and only if $|f|$ is Lebesgue integrable.

Remark 1.2 The Lebesgue integral of a measurable function f defined from a measurable space $(\Omega, \mathcal{A})$ on $(\mathbb{R}, \mathcal{B})$, over a Borel set $I = (a, b) \in \mathcal{B}$, will also be expressed as

$$\int_I f \, d\mu = \int_a^b f(x) \, d\mu(x).$$

Proposition 1.11 If a function $f : \mathbb{R} \to \mathbb{R}_+ = [0, \infty)$ is Riemann integrable, then it is also Lebesgue integrable (with respect to the Lebesgue measure λ) and the two integrals coincide, i.e.,

$$\int_A f(x) \, d\mu(x) = \int_A f(x) \, dx.$$

Example 1.16 The *Dirichlet function* $\mathbf{1}_{\mathbb{Q}}$ is not continuous in any point of its domain.

- This function is not Riemann integrable in $[0, 1]$ because each subinterval will contain at least a rational number and and irrational number, since both sets are dense in $\mathbb{R}$. Then, each superior sum is 1 and also the infimum of this superior sums, whereas each lowersum is 0, the same as the suppremum of all lowersums. Since the suppresum and the infimum are different, then the Riemann integral does not exist.

- However, it is Lebesgue integrable on $[0, 1]$ with respect to the Lebesgue measure λ, since by definition

  $$\int_{[0,1]} \mathbf{1}_{\mathbb{Q}} \, d\lambda = \lambda(\mathbb{Q} \cap [0, 1]) = 0,$$

  since $\mathbb{Q}$ is numerable.

Proposition 1.12 (Properties of Lebesgue integral)

(a) If $P(A) = 0$, then $\int_A f \, d\mu = 0$.

(b) If $\{A_n\}_{n\in I\!N}$ is a sequence of disjoint sets with $A = \bigcup_{n=1}^{\infty} A_n$, then $\int_A f\, d\mu = \sum_{n=1}^{\infty} \int_{A_n} f\, d\mu$.

(c) If two measurable functions f and g are equal in all parts of their domain except for a subset with measure μ zero and f is Lebesgue integrable, then g is also Lebesgue integrable and their Lebesgue integral is the same, that is,

$$\text{If } \mu\left(\{\omega \in \Omega : f(\omega) \neq g(\omega)\}\right) = 0, \text{ then } \int_\Omega f\, d\mu = \int_\Omega g\, d\mu.$$

(d) Linearity: If f and g are Lebesgue integrable functions and a and b are real numbers, then

$$\int_A (a\, f + b\, g)\, d\mu = a \int_A f\, d\mu + b \int_A g\, d\mu.$$

(e) Monotonocity: If f and g are Lebesgue integrable and $f < g$, then

$$\int f\, d\mu \leq \int g\, d\mu.$$

Theorem 1.7 (Monotone convergence theorem) Consider a point-wise non-decreasing sequence of $[0, \infty]$-valued measurable functions $\{f_n\}_{n\in I\!N}$ (i.e., $0 \leq f_n(x) \leq f_{n+1}(x)$, $\forall x \in I\!R$, $\forall n > 1$) with $\lim_{n\to\infty} f_n = f$. Then,

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

Theorem 1.8 (Dominated convergence theorem) Consider a sequence of real-valued measurable functions $\{f_n\}_{n\in I\!N}$ with $\lim_{n\to\infty} f_n = f$. Assume that the sequence is dominated by an integrable function g (i.e., $|f_n(x)| \leq g(x)$, $\forall x \in I\!R$, with $\int_\Omega g(x)\,d\mu < \infty$). Then,

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

Theorem 1.9 (Hölder's inequality)
Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. Let $p, q \in I\!R$ such that $p > 1$ and $1/p + 1/q = 1$. Let f and g be measurable functions with $|f|^p$ and $|g|^q$ μ-integrable (i.e., $\int |f|^p d\mu < \infty$ and $\int |g|^q d\mu < \infty$.). Then, $|fg|$ is also μ-integrable (i.e., $\int |fg| d\mu < \infty$) and

$$\int_\Omega |fg| d\mu \leq \left(\int_\Omega |f|^p\,d\mu\right)^{1/p} \left(\int_\Omega |g|^q\,d\mu\right)^{1/q}.$$

The particular case with $p = q = 2$ is known as *Schwartz's inequality*.

Theorem 1.10 (Minkowski's inequality)
Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. Let $p \geq 1$. Let f and g be measurable functions with $|f|^p$ and $|g|^p$ μ-integrable. Then, $|f+g|^p$ is also μ-integrable and

$$\left(\int_\Omega |f+g|^p d\mu\right)^{1/p} \leq \left(\int_\Omega |f|^p\,d\mu\right)^{1/p} + \left(\int_\Omega |g|^p\,d\mu\right)^{1/p}.$$

Definition 1.32 (L^p space)
Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. Let $p \neq 0$. We define the $L^p(\mu)$ space as the set of measurable functions f with $|f|^p$ μ-integrable, that is,

$$L^p(\mu) = L^p(\Omega, \mathcal{A}, \mu) = \left\{ f: \ f \text{ measurable and } \int_\Omega |f|^p\,d\mu < \infty \right\}.$$

By the Minkowski's inequality, the $L^p(\mu)$ space with $1 \leq p < \infty$ is a vector space in $I\!R$, in which we can define a norm and the corresponding metric associated with that norm.

Proposition 1.13 (Norm in L^p space)
The function $\phi : L^p(\mu) \to I\!R$ that assigns to each function $f \in L^p(\mu)$ the value $\phi(f) = \left(\int_\Omega |f|^p \, d\mu\right)^{1/p}$ is a norm in the vector space $L^p(\mu)$ and it is denoted as

$$\|f\|_p = \left(\int_\Omega |f|^p \, d\mu\right)^{1/p}.$$

Now we can introduce a *metric* in $L^p(\mu)$ as

$$d(f,g) = \|f - g\|_p.$$

A vector space with a metric obtained from a norm is called a *metric space.*

Problems

1. Proof that if f and g are measurable, then $\max\{f,g\}$ and $\min\{f,g\}$ are measurable.

1.7 Distribution function

We will consider the probability space $(I\!R, \mathcal{B}, P)$. The distribution function will be a very important tool since it will summarize the probabilities over Borel subsets.

Definition 1.33 (Distribucion function)
Let $(I\!R, \mathcal{B}, P)$ a probability space. The *distribution function* (d.f.) associated with the probability function P is defined as

$$\begin{aligned} F:\ & I\!R \to [0,1] \\ & x \rightsquigarrow F(x) = P(-\infty, x]. \end{aligned}$$

We can also define $F(-\infty) = \lim_{x\downarrow -\infty} F(x)$ and $F(+\infty) = \lim_{x\uparrow +\infty} F(x)$. Then, the distribution function is $F : [-\infty, +\infty] \to [0,1]$.

Proposition 1.14 (Properties of the d.f.)

(i) The d.f. is monotone increasing, that is,

$$x < y \Rightarrow F(x) \leq F(y).$$

(ii) $F(-\infty) = 0$ and $F(+\infty) = 1$.

(iii) F is right continuous for all $x \in I\!R$.

Remark 1.3 If the d.f. was defined as $F(x) = P(-\infty, x)$, then it would be left continuous.

We can speak about a d.f. without reference to the probability measure P that is used to define the d.f.

Definition 1.34 A function $F : [-\infty, +\infty] \to [0, 1]$ is a d.f. if and only if satisfies Properties (i)-(iii).

Now, given a d.f. F verifying (i)-(iii), is there a unique probability function over $(I\!R, \mathcal{B})$ whose d.f. is exactly F?

Proposition 1.15 Let $F : [-\infty, +\infty] \to [0, 1]$ be a function that satisfies properties (i)-(iii). Then, there is a unique probability function P_F defined over $(I\!R, \mathcal{B})$ such that the distribution function associated with P_F is exactly F.

Remark 1.4 Let a, b be real numbers with $a < b$. Then $P(a, b] = F(b) - F(a)$.

Theorem 1.11 The set $D(F)$ of discontinuity points of F is finite or countable.

Definition 1.35 (Discrete d.f.)
A d.f. F is discrete if there exist a finite or countable set $\{a_1, \ldots, a_n, \ldots\} \subset I\!R$ such that $P_F(\{a_i\}) > 0, \forall i$ and $\sum_{i=1}^{\infty} P_F(\{a_i\}) = 1$, where P_F is the probability function associated with F.

Definition 1.36 (Probability mass function)
The collection of numbers $P_F(\{a_1\}), \ldots, P_F(\{a_n\}), \ldots$, such that $P_F(\{a_i\}) > 0$, $\forall i$ and $\sum_{i=1}^{\infty} P_F(\{a_i\}) = 1$, is called *probability mass function*.

Remark 1.5 Observe that

$$F(x) = P_F(-\infty, x] = \sum_{a_i \leq x} P_F(\{a_i\}).$$

Thus, $F(x)$ is a step function and the length of the step at a_n is exactly the probability of a_n, that is,

$$\begin{aligned} P_F(\{a_i\}) &= P(-\infty, a_n] - P(-\infty, a_n) = F(a_n) - \lim_{x \downarrow a_n} F(x) \\ &= F(a_n) - F(a_n-). \end{aligned}$$

Theorem 1.12 (Radon-Nykodym Theorem)
Given a measurable space $(\Omega, \mathcal{A})$, if a σ-finite measure μ on $(\Omega, \mathcal{A})$ is absolutely continuous with respect to a σ-finite measure λ on $(\Omega, \mathcal{A})$, then there is a measurable function $f : \Omega \to [0, \infty)$, such that for any measurable set A,

$$\mu(A) = \int_A f \, d\lambda.$$

The function f satisfying the above equality is uniquely defined up to a set with measure μ zero, that is, if g is another function which satisfies the same property, then $f = g$ except in a set with measure μ zero. f is commonly written $d\mu/d\lambda$ and is called the *Radon–Nikodym derivative*. The choice of notation and the name of the function reflects the fact that the function is analogous to a derivative in calculus in the sense that it describes the rate of change of density of one measure with respect to another.

Theorem 1.13 A finite measure μ on Borel subsets of the real line is absolutely continuous with respect to Lebesgue measure if and only if the point function

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

is a locally and absolutely continuous real function.

If μ is absolutely continuous, then the Radon-Nikodym derivative of μ is equal almost everywhere to the derivative of F. Thus, the absolutely continuous measures on $I\!R^n$ are precisely those that have densities; as a special case, the absolutely continuous d.f.'s are precisely the ones that have probability density functions.

Definition 1.37 (Absolutely continuous d.f.)
A d.f. is *absolutely continuous* if and only if there is a non-negative Lebesgue integrable function f such that

$$\forall x \in I\!R, F(x) = \int_{(-\infty,x]} f d\lambda,$$

where λ is the Lebesgue measure. The function f is called *probability density function*, p.d.f.

Proposition 1.16 Let $f : I\!R \to I\!R_+ = [0, \infty)$ be a Riemann integrable function such that $\int_{-\infty}^{+\infty} f(t)dt = 1$. Then, $F(x) = \int_{-\infty}^{x} f(t)dt$ is an absolutely continuous d.f. whose associated p.d.f is f.

All the p.d.f.'s that we are going to see are Riemann integrable.

Proposition 1.17 Let F be an absolutely continuous d.f. Then it holds:

(a) F is continuous.

(b) If f is continuous in the point x, then F is differentiable in x and $F'(x) = f(x)$.

(c) $P_F(\{x\}) = 0$, $\forall x \in I\!R$.

(d) $P_F(a,b) = P_F(a,b] = P_F[a,b) = P_F[a,b] = \int_a^b f(t)dt$, $\forall a, b$ with $a < b$.

(e) $P_F(B) = \int_B f(t)dt, \forall B \in \mathcal{B}$.

Remark 1.6 Note that:

(1) Not all continuous d.f.'s are absolutely continuous.

(2) Another type of d.f's are those called *singular* d.f.'s, which are continuous. We will not study them.

Proposition 1.18 Let F_1, F_2 be d.f.'s and $\lambda \in [0,1]$. Then, $F = \lambda F_1 + (1-\lambda)F_2$ is a d.f.

Definition 1.38 (Mixed d.f.)
A d.f. is said to be *mixed* if and only if there is a discrete d.f. F_1, an absolutely continuous d.f. F_2 and $\lambda \in [0,1]$ such that $F = \lambda F_1 + (1-\lambda)F_2$.

1.8 Random variables

A random variable transforms the elements of the sample space Ω into real numbers (elements from $I\!R$), preserving the σ-algebra structure of the initial events.

Definition 1.39 Let $(\Omega, \mathcal{A})$ be a measurable space. Consider also the measurable space $(I\!R, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra over $I\!R$. A random variable (r.v.) is a function $X : \Omega \to I\!R$ that is measurable, that is,

$$\forall B \in \mathcal{B}, \quad X^{-1}(B) \in \mathcal{A},$$

where $X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\}$.

Remark 1.7 Observe that:

(a) a r.v. X is simply a measurable function in $I\!R$. The name random variable stems from the fact the result of the random experiment $\omega \in \Omega$ is random, and then the observed value of the r.v., $X(\omega)$, is also random.

(b) the measurability property of the r.v. will allow transferring probabilities of events $A \in \mathcal{A}$ to probabilities of Borel sets $I \in \mathcal{B}$, where I is the image of A through X.

Example 1.17 For the experiments introduced in Example 1.8, the following are random variables:

(a) For the measurable space $(\Omega, \mathcal{A})$ with sample space $\Omega = \{\text{“}\textit{head}\text{”}, \text{“}\textit{tail}\text{”}\}$ and σ-algebra $\mathcal{A} = \{\emptyset, \{\text{“}\textit{head}\text{”}\}, \{\text{“}\textit{tail}\text{”}\}, \Omega\}$, a random variable is:

$$X(\omega) = \begin{cases} 1 & \text{if } \omega = \text{“}\textit{head}\text{”}, \\ 0 & \text{if } \omega = \text{“}\textit{tail}\text{”}; \end{cases}$$

This variable counts the number of heads when tossing a coin. In fact, it is a random variable, since for any event from the final space $B \in \mathcal{B}$, we have

✓ If $0, 1 \in B$, then $X^{-1}(B) = \Omega \in \mathcal{A}$.

✓ If $0 \in B$ but $1 \notin B$, then $X^{-1}(B) = \{\text{“tail”}\} \in \mathcal{A}$.

✓ If $1 \in B$ but $0 \notin B$, then $X^{-1}(B) = \{\text{“head”}\} \in \mathcal{A}$.

✓ If $0, 1 \notin B$, then $X^{-1}(B) = \emptyset \in \mathcal{A}$.

(b) For the measurable space $(\Omega, \mathcal{P}(\Omega))$, where $\Omega = \mathbb{N} \cup \{0\}$, since $\Omega \subset \mathbb{R}$, a trivial r.v. is $X_1(\omega) = \omega$. It is a r.v. since for any $B \in \mathcal{B}$,

$$X_1^{-1}(B) = \{\omega \in \mathbb{N} \cup \{0\} : \; X_1(\omega) = \omega \in B\}$$

is the set of natural numbers (including zero) that are contained in B. But any countable set of natural numbers belongs to $\mathcal{P}(\Omega)$, since this σ-algebra contains all subsets of $\mathbb{N} \cup \{0\}$. Therefore, X_1="Number of traffic accidents in a minute in Spain" is a r.v.

Another r.v. could be

$$X_2(\omega) = \begin{cases} 1 & \text{if } \omega \in \mathbb{N}; \\ 0 & \text{if } \omega = 0. \end{cases}$$

Again, X_2 is a r.v. since for each $B \in \mathcal{B}$,

- ✓ If $0, 1 \in B$, then $X_2^{-1}(B) = \Omega \in \mathcal{P}(\Omega)$.
- ✓ If $1 \in B$ but $0 \notin B$, then $X_2^{-1}(B) = \mathbb{N} \in \mathcal{P}(\Omega)$.
- ✓ If $0 \in B$ but $1 \notin B$, then $X_2^{-1}(B) = \{0\} \in \mathcal{P}(\Omega)$.
- ✓ If $0, 1 \notin B$, then $X_2^{-1}(B) = \emptyset \in \mathcal{P}(\Omega)$.

(c) As in previous example, for the measurable space $(\Omega, \mathcal{B}_\Omega)$, where $\Omega = [m, \infty)$, a possible r.v. is $X_1(\omega) = \omega$, since for each $B \in \mathcal{B}$, we have

$$X_1^{-1}(B) = \{\omega \in [a, \infty) : \; X_1(\omega) = \omega \in B\} = [a, \infty) \cap B \in \mathcal{B}_\Omega.$$

Another r.v. would be the indicator of less than 65 kgs., given by

$$X_2(\omega) = \begin{cases} 1 & \text{if } \omega \geq 65, \\ 0 & \text{if } \omega < 65. \end{cases}$$

Theorem 1.14 Any function X from $(I\!R, \mathcal{B})$ in $(I\!R, \mathcal{B})$ that is continuous is a r.v.

The probability of an event from $I\!R$ induced by a r.v. is going to be defined as the probability of the "original" events from Ω, that is, the probability of a r.v. preserves the probabilities of the original measurable space. This definition requires the measurability property, since the "original" events must be in the initial σ-algebra so that they have a probability.

Definition 1.40 (Probability induced by a r.v.)
Let $(\Omega, \mathcal{A}, P)$ be a measure space and let $\mathcal{B}$ be the Borel σ-algebra over $I\!R$. The *probability induced* by the r.v. X is a function $P_X : \mathcal{B} \to I\!R$, defined as

$$P_X(B) = P(X^{-1}(B)), \ \forall B \in \mathcal{B}.$$

Theorem 1.15 The probability induced by a r.v. X is a probability function in $(I\!R, \mathcal{B})$.

Example 1.18 For the probability function P_1 defined in Example 1.10 and the r.v. defined in Example 1.17 (a), the probability induced by a r.v. X is described as follows. Let $B \in \mathcal{B}$.

✓ If $0, 1 \in B$, then $P_{1X}(B) = P_1(X^{-1}(B)) = P_1(\Omega) = 1$.

✓ If $0 \in B$ but $1 \notin B$, then $P_{1X}(B) = P_1(X^{-1}(B)) = P_1(\{\text{"tail"}\}) = 1/2$.

✓ If $1 \in B$ but $0 \notin B$, then $P_{1X}(B) = P_1(X^{-1}(B)) = P_1(\{\text{"head"}\}) = 1/2$.

✓ If $0, 1 \notin B$, then $P_{1X}(B) = P_1(X^{-1}(B)) = P_1(\emptyset) = 0$.

Summarizing, the probability induced by X is

$$P_{1X}(B) = \begin{cases} 0, & \text{if } 0,1 \notin B; \\ 1/2, & \text{if } 0 \text{ or } 1 \text{ are in } B; \\ 1, & \text{if } 0,1 \in B. \end{cases}$$

In particular, we obtain the following probabilities

✓ $P_{1X}(\{0\}) = P_1(X = 0) = 1/2.$

✓ $P_{1X}((-\infty, 0]) = P_1(X \leq 0) = 1/2.$

✓ $P_{1X}((0, 1]) = P_1(0 < X \leq 1) = 1/2.$

Example 1.19 For the probability function P introduced in Example 1.11 and the r.v. X_1 defined in Example 1.17 (b), the probability induced by the r.v. X_1 is described as follows. Let $B \in \mathcal{B}$ such that $I\!N \cap B = \{a_1, a_2, \ldots, a_p\}$.

$$P_{X_1}(B) = P(X_1^{-1}(B)) = P_1((I\!N \cup \{0\}) \cap B) = P(\{a_1, a_2, \ldots, a_p\}) = \sum_{i=1}^{p} P(\{a_i$$

Definition 1.41 (Degenerate r.v.)
A r.v. X is said to be *degenerate* at a point $c \in I\!R$ if and only if

$$P(X = c) = P_X(\{c\}) = 1.$$

Since P_X is a probability function, there is a distribution function that summarizes its values.

Definition 1.42 (Distribucion function)
The *distribucion function* (d.f.) of a r.v. X is defined as the function $F_X : I\!R \to [0, 1]$ with

$$F_X(x) = P_X(-\infty, x], \ \forall x \in I\!R.$$

Definition 1.43 A r.v. X is said to be *discrete* (*absolutely continuous*) if and only if its d.f. F_X is discrete (absolutely continuous).

Remark 1.8 It holds that:

(a) a discrete r.v. takes a finite or countable number of values.

(b) a continuous r.v. takes infinite number of values, and the probability of single values are zero.

Definition 1.44 (Support of a r.v.)

(a) If X is a discrete r.v., we define the *support* of X as

$$D_X := \{x \in I\!R : P_X\{x\} > 0\}.$$

(b) If X continuous, the *support* is defined as

$$D_X := \{x \in I\!R : f_X(x) > 0\}.$$

Observe that for a discrete r.v., $\sum_{x \in D_X} P_X\{x\} = 1$ and D_X is finite or countable.

Example 1.20 From the random variables introduced in Example 1.17, those defined in (a) and (b) are discrete, along with X_2 form (c).

For a discrete r.v. X, we can define a function the gives the probabilities of single points.

Definition 1.45 The *probability mass function* (p.m.f) of a discrete r.v. X is the function $p_X : I\!R \to I\!R$ such that

$$p_X(x) = P_X(\{x\}), \ \forall x \in I\!R.$$

We will also use the notation $p(X = x) = p_X(x)$.

The probability function induced by a discrete r.v. X, P_X, is completely determined by the distribution function F_X or by the mass function p_X. Thus, in the following, when we speak about the "distribution" of a discrete r.v. X, we could be referring either to the probability function induced by X, P_X, the distribution function F_X, or the mass function p_X.

Example 1.21 The r.v. X_1 : "Weight of a randomly selected Spanish woman aged within 20 and 40", defined in Example 1.17 (c), is continuous.

Definition 1.46 The *probability density function* (p.d.f.) of X is a function $f_X : \mathbb{R} \to \mathbb{R}$ defined as

$$f_X(x) = \begin{cases} 0 & \text{if } x \in S; \\ F'_X(x) & \text{if } x \notin S. \end{cases}$$

It is named probability density function of x because it gives the density of probability of an infinitesimal interval centered in x.

The same as in the discrete case, the probabilities of a continuous r.v. X are determined either by P_X, F_X or the p.d.f. f_X. Again, the "distribution" of a r.v., could be referring to any of these functions.

Random variables, as measurable functions, inherit all the properties of measurable functions. Furthermore, we will be able to calculate Lebesgue integrals of measurable functions of r.v.'s using as measure their induced probability functions. This will be possible due to the following theorem.

Theorem 1.16 (Theorem of change of integration space) Let X be a r.v. from $(\Omega, \mathcal{A}, P)$ in $(\mathbb{R}, \mathcal{B})$ an g another r.v. from

$(I\!R, \mathcal{B})$ in $(I\!R, \mathcal{B})$. Then,

$$\int_\Omega (g \circ X)\, dP = \int_{I\!R} g\, dP_X.$$

Remark 1.9 Let F_X be the d.f. associated with the probability measure P_X. The integral

$$\int_{I\!R} g\, dP_X = \int_{-\infty}^{+\infty} g(x)\, dP_X(x)$$

will also be denoted as

$$\int_{I\!R} g\, dF_x = \int_{-\infty}^{+\infty} g(x)\, dF_X(x).$$

Proposition 1.19 If X is an absolutely continuous r.v. with d.f. F_X and p.d.f. with respect to the Lebesgue measure $f_X = dF_X/d\mu$ and if g is any function for which $\int_{I\!R} |g|\, dP_X < \infty$, then

$$\int_{I\!R} g\, dP_X = \int_{I\!R} g \cdot f_X\, d\mu.$$

In the following we will see how to calculate these integrals for the most interesting cases of d.f. F_X.

(a) F_X discrete: The probability is concentrated in a finite or numerable set $D_X = \{a_1, \ldots, a_n, \ldots\}$, with probabilities $P\{a_1\}, \ldots, P\{a_n\}, .$ Then, using properties (a) and (b) of the Lebesgue integral,

$$\begin{aligned}\int_{I\!R} g\, dP_X &= \int_{D_X} g\, dP_X + \int_{D_X^c} g(x)\, dP_X \\ &= \int_{D_X} g\, dP_X \\ &= \sum_{n=1}^{\infty} \int_{\{a_n\}} g\, dP_X \\ &= \sum_{n=1}^{\infty} g(a_n) P\{a_n\}.\end{aligned}$$

(b) F_X absolutely continuous: In this case,

$$\int_{\mathbb{R}} g\, dP_X = \int_{\mathbb{R}} g \cdot f_X\, d\mu$$

and if $g \cdot f_X$ is Riemann integrable, then it is also Lebesgue integrable and the two integrals coincide, i.e.,

$$\int_{\mathbb{R}} g\, dP_X = \int_{\mathbb{R}} g \cdot f_X\, d\mu = \int_{-\infty}^{+\infty} g(x) f_X(x) dx.$$

Definition 1.47 The *expectation* of the r.v. X is defined as

$$\mu = E(X) = \int_{\mathbb{R}} X\, dP.$$

Corolary 1.2 The expectation of X can be calculated as

$$E(X) = \int_{\mathbb{R}} x\, dF_X(x),$$

and provided that X is absolutely continuous with p.d.f. $f_X(x)$, then

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

Definition 1.48 The k-th *moment* of X with respect to $a \in \mathbb{R}$ is defined as

$$\alpha_{k,a} = E(g_{k,a} \circ X),$$

where $g_{k,a}(x) = (x-a)^k$, provided that the expectation exists.

Remark 1.10 It holds that

$$\alpha_{k,a} = \int_{\mathbb{R}} g_{k,a} \circ X\, dP = \int_{\mathbb{R}} g_{k,a}(x)\, dF_X(x) = \int_{\mathbb{R}} (x-a)^k\, dF_X(x).$$

Observe that for the calculation of the moments of a r.v. X we only require is its d.f.

Definition 1.49 The k-th *moment* of X with respect to the mean μ is

$$\mu_k := \alpha_{k,\mu} = \int_{I\!R} (x-\mu)^k \, dF_X(x).$$

In particular, second moment with respect to the mean is called *variance*,

$$\sigma_X^2 = V(X) = \mu_2 = \int_{I\!R} (x-\mu)^2 \, dF_X(x).$$

The standard deviation is $\sigma_X = \sqrt{V(X)}$.

Definition 1.50 The k-th *moment* of X with respect to the origin μ is

$$\alpha_k := \alpha_{k,0} = E(X^n) = \int_{I\!R} x^k \, dF_X(x).$$

Proposition 1.20 It holds

$$\mu_k = \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} \mu^{k-i} \alpha_i.$$

Lemma 1.1 If $\alpha_k = E(X^k)$ exists and is finite, then there exists α_m and is finite, $\forall m \leq k$.

One way of obtaining information about the distribution of a random variable is to calculate the probability of intervals of the type $(E(X)-\epsilon, E(X)+\epsilon)$. If we do not know the theoretical distribution of the random variable but we do know its expectation and variance, the Tchebychev's inequality gives a lower bound of this probability. This inequality is a straightforward consequence of the following one.

Theorem 1.17 (Markov's inequality)
Let X be a r.v. from $(\Omega, \mathcal{A}, P)$ in $(I\!R, \mathcal{B})$ and g be a non-negative r.v. from $(I\!R, \mathcal{B}, P_X)$ in $(I\!R, \mathcal{B})$ and let $k > 0$. Then, it holds

$$P\left(\{\omega \in \Omega : g(X(\omega)) \geq k\}\right) \leq \frac{E[g(X)]}{k}.$$

Theorem 1.18 (Tchebychev's inequality)
Let X be a r.v. with finite mean μ and finite standard deviation σ. Then

$$P\left(\{\omega \in \Omega : |X(\omega) - \mu| \geq k\sigma\}\right) \leq \frac{1}{k^2}.$$

Corolary 1.3 Let X be a r.v. with mean μ and standard deviation $\sigma = 0$. Then, $P(X = \mu) = 1$.

Problems

1. Prove Markov's inequality.
2. Prove Tchebychev's inequality.
3. Prove Corollary 1.3.

1.9 The characteristic function

We are going to define the characteristic function associated with a distribution function (or with a random variable). This function is pretty useful due to its close relation with the d.f. and the moments of a r.v.

Definition 1.51 (Characteristic function)
Let X be a r.v. defined from the measure space $(\Omega, \mathcal{A}, P)$ into

$(\mathbb{R}, \mathcal{B})$. The characteristic function (c.f.) of X is

$$\varphi(t) = E\left[e^{itX}\right] = \int_\Omega e^{itX} dP, \quad t \in \mathbb{R}.$$

Remark 1.11 (The c.f. is determined by the d.f.)
The function $Y_t = g_t(X) = e^{itX}$ is a composition of two measurable functions, $X(\omega)$ and $g_t(x)$, that is, $Y_t = g_t \circ X$. Then, Y_t is measurable and by the Theorem of change of integration space, the c.f. is calculated as

$$\begin{aligned} \varphi(t) &= \int_\Omega e^{itX} dP = \int_\Omega (g_t \circ X) dP \\ &= \int_\Omega g_t dP_X = \int_{\mathbb{R}} g_t(x) dF(x) = \int_{\mathbb{R}} e^{itx} dF(x), \end{aligned}$$

where P_X is the probability induced by X and F is the d.f. associated with P_X. Observe that the only thing that we need to obtain φ is the d.f., F, that is, φ is uniquely determined by F.

Remark 1.12 Observe that:

- $\int_{\mathbb{R}} e^{itx} dF(x) = \int_{\mathbb{R}} \cos(tx) dF(x) + i \int_{\mathbb{R}} \sin(tx) dF(x)$.
- Since $|\cos(tx)| \le 1$ and $|\sin(tx)| \le 1$, then it holds:

  $$\int_{\mathbb{R}} |\cos(tx)| \le 1, \quad \int_{\mathbb{R}} |\sin(tx)| \le 1$$

  and therefore, $|\cos(tx)|$ and $|\sin(tx)|$ are integrable. This means that $\varphi(t)$ exists $\forall t \in \mathbb{R}$.

- Many properties of the integral of real functions can be translated to the integral of the complex function e^{itx}. In practically all cases, the result is a straightforward consequence of the fact that to integrate a complex values function is equivalent to integrate separately the real and imaginary parts.

Proposition 1.21 (Properties of the c.f.)
Let $\varphi(t)$ be the characteristic function associated with the d.f. F. Then

(a) $\varphi(0) = 1$ (φ is non-vanishing at $t = 0$);

(b) $|\varphi(t)| \leq 1$ (φ is bounded);

(c) $\varphi(-t) = \overline{\varphi(t)}$, $\forall t \in I\!R$, where $\overline{\varphi(t)}$ denotes the conjugate complex of $\varphi(t)$;

(d) $\varphi(t)$ is uniformly continuous in $I\!R$, that is,

$$\lim_{h \downarrow 0} |\varphi(t+h) - \varphi(t)| = 0, \quad \forall t \in I\!R.$$

Theorem 1.19 (c.f. of a linear transformation)
Let X be a r.v. with c.f. $\varphi_X(t)$. Then, the c.f. of $Y = aX + b$, where $a, b \in I\!R$, is $\varphi_Y(t) = e^{itb}\varphi_X(at)$.

Example 1.22 (c.f. for some r.v.s)
Here we give the c.f. of some well known random variables:

(i) For the Binomial distribution, $\text{Bin}(n, p)$, the c.f. is given by

$$\varphi(t) = (q + pe^{it})^n.$$

(ii) For the Poisson distribution, $\text{Pois}(\lambda)$, the c.f. is given by

$$\varphi(t) = \exp\left\{\lambda(e^{it} - 1)\right\}.$$

(iii) For the Normal distribution, $N(\mu, \sigma^2)$, the c.f. is given by

$$\varphi(t) = \exp\left\{i\mu t - \frac{\sigma^2 t^2}{2}\right\}.$$

Lemma 1.2 $\forall x \in I\!R$, $|e^{ix} - 1| \leq |x|$.

Remark 1.13 (Proposition 1.21 does not determine a c.f.)
If $\varphi(t)$ is a c.f., then Properties (a)-(d) in Proposition 1.21 hold but the reciprocal is not true, see Example 1.23.

Theorem 1.20 (Moments are determined by the c.f.)
If the n-th moment of F, $\alpha_n = \int_{I\!R} x^n dF(x)$, is finite, then

(a) The n-th derivative of $\varphi(t)$ at $t = 0$ exists and satisfies $\varphi^{n)}(0) = i^n \alpha_n$.

(b) $\varphi^{n)}(t) = i^n \int_{I\!R} e^{itx} x^n dF(x)$.

Corolary 1.4 (Series expansion of the c.f.)
If $\alpha_n = E(X^n)$ exists $\forall n \in I\!N$, then it holds that

$$\varphi_X(t) = \sum_{n=0}^{\infty} \alpha_n \frac{(it)^n}{n!}, \quad \forall t \in (-r, r),$$

where $(-r, r)$ is the radius of convergence of the series.

Example 1.23 (Proposition 1.21 does not determine a c.f.)
Consider the function $\varphi(t) = \frac{1}{1+t^4}$, $t \in I\!R$. This function verifies properties (a)-(d) in Proposition 1.21. However, observe that the first derivative evaluated at zero is

$$\varphi'(0) = \left| \frac{-4t^3}{(1+t^4)^2} \right|_{t=0} = 0.$$

The second derivative at zero is

$$\varphi'(0) = \left| \frac{-12t^2(1+t^4)^2 + 4t^3\, 2(1+t^4)\, 4t^3}{(1+t^4)^4} \right|_{t=0} = 0.$$

Then, if $\varphi(t)$ is the c.f. of a r.v. X, the mean and variance are equal to

$$E(X) = \alpha_1 = \frac{\varphi'(0)}{i} = 0, \quad V(X) = \alpha_2 - (E(X))^2 = \alpha_2 = \frac{\varphi''(0)}{i^2} = 0.$$

But a random variable with mean and variance equal to zero is a degenerate variable at zero, that is, $P(X = 0) = 1$, and then its c.f. is

$$\varphi(t) = E\left[e^{it0}\right] = 1, \quad \forall t \in I\!R,$$

which is a contradiction.

We have seen already that the d.f. determines the c.f. The following theorem gives an expression of the d.f. in terms of the c.f. for an interval. This result will imply that the c.f. determines a unique d.f.

Theorem 1.21 (Inversion Theorem)
Let $\varphi(t)$ be the c.f. corresponding to the d.f. $F(x)$. Let a, b be two points of continuity of F, that is, $a, b \in \mathcal{C}(F)$. Then,

$$F(b) - F(a) = \frac{1}{2\pi} \lim_{T\to\infty} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt.$$

As a consequence of the Inversion Theorem, we obtain the following result.

Theorem 1.22 (The c.f. determines a unique d.f.)
If $\varphi(t)$ is the c.f. of a d.f. F, then it is not the c.f. of any other d.f.

Remark 1.14 (c.f. for an absolutely continuous r.v.)
If F is absolutely continuous with p.d.f. f, then the c.f. is

$$\varphi(t) = E\left[e^{itX}\right] = \int_{I\!R} e^{itx} dF(x) = \int_{I\!R} e^{itx} f(x) dx.$$

We have seen that for absolutely continuous F, the c.f. $\varphi(t)$ can be expressed in terms of the p.d.f. f. However, it is possible to express the p.d.f. f in terms of the c.f. $\varphi(t)$? The next theorem is the answer.

Theorem 1.23 (Fourier transform of the c.f.)
If F is absolutely continuous and $\varphi(t)$ is Riemann integrable in $I\!R$, that is, $\int_{-\infty}^{\infty} |\varphi(t)|\, dt < \infty$, then $\varphi(t)$ is the c.f. corresponding to an absolutely continuous r.v. with p.d.f. given by

$$f(x) = F'(x) = \frac{1}{2\pi} \int_{I\!R} e^{-itx} \varphi(t) dt,$$

where the last term is called the Fourier transform of φ.

In the following, we are going to study the c.f. of random variables that share the probability symmetrically in $I\!R_+$ and $I\!R_-$.

Definition 1.52 (Symmetric r.v.)
A r.v. X is symmetric if and only if $X \stackrel{d}{=} -X$, that is, iff $F_X(x) = F_{-X}(x)$, $\forall x \in I\!R$.

Remark 1.15 (Symmetric r.v.)
Since φ is determined by F, X is symmetric iff $\varphi_X(t) = \varphi_{-X}(t)$, $\forall t \in I\!R$.

Corolary 1.5 (Another characterization of a symmetric r.v.)
X is symmetric iff $F_X(-x) = 1 - F_X(x^-)$, where $F_X(x^-) = P(X < x)$.

Corolary 1.6 (Another characterization of a symmetric r.v.)
X is symmetric iff $\varphi_X(t) = \varphi_X(-t)$, $\forall t \in I\!R$.

Corolary 1.7 (Properties of a symmetric r.v.)
Let X be a symmetric r.v. Then,

(a) If F_X is absolutely continuous, then $f_X(x) = f_X(-x)$, $\forall x \in \mathbb{R}$.

(b) If F_X is discrete, then $P_X(x) = P_X(-x)$, $\forall x \in \mathbb{R}$.

Theorem 1.24 (c.f. of a symmetric r.v.)
The c.f. φ of the d.f. F is real iff F is symmetric.

Remark 1.16 We know that $\varphi(t)$, $\forall t \in \mathbb{R}$ determines completely $F(x)$, $\forall x \in \mathbb{R}$. However, if we only know $\varphi(t)$ for t in a finite interval, then do we know completely $F(x)$, $\forall x \in \mathbb{R}$? The answer is no, since we can find two different d.f.s with the same c.f. in a finite interval, see Example 1.24.

Example 1.24 (The c.f. in a finite interval does not determine the d.f.)
Consider the r.v. X taking the values $\mp(2n+1)$, $n = 0, 1, 2, \cdots$, with probabilities

$$P(X = 2n+1) = P(X = -(2n+1)) = \frac{4}{\pi^2(2n+1)^2}, \quad n = 0, 1, 2, \cdots$$

Consider also the r.v. Y taking values $0, \mp(4n+2)$, $n = 0, 1, 2, \ldots$, with probabilities

$$P(Y = 0) = 1/2$$
$$P(Y = 4n+2) = P(X = -(4n+2)) = \frac{2}{\pi^2(2n+1)^2}, \quad n = 0, 1, 2, \cdots$$

Using the formulas

$$\sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} = \frac{\pi^2}{8};$$
$$\frac{8}{\pi^2} \sum_{n=0}^{\infty} \frac{\cos(2n+1)t}{(2n+1)^2} = 1 - \frac{2|t|}{\pi}, \quad |t| \leq \pi,$$

prove:

(a) P_X is a probability function;

(b) P_Y is a probability function;

(c) $\varphi_X(t) = \varphi_Y(t)$, for $|t| \leq \pi/2$.

Remark 1.17 From the series expansion of the c.f. in Corollary 1.4, one is tempted to conclude that the c.f., and therefore also the d.f., of a r.v. is completely determined by all of its moments, provided that they exist. This is false, see Example 1.25.

Example 1.25 (Moments do not always determine the c.f.)
For $a \in [0,1]$, consider the p.d.f. defines by

$$f_a(x) = \frac{1}{24} e^{-x^{1/4}}(1 - a\sin(x^{1/4}), \quad x \geq 0.$$

Using the following formulas:

$$\int_0^{\infty} x^n e^{-x^{1/4}} \sin(x^{1/4}) dx = 0, \quad \forall n \in I\!N \cup \{0\};$$
$$\int_0^{\infty} x^n e^{-x^{1/4}} = 4(4n+3)! \forall n \in I\!N \cup \{0\},$$

prove:

(a) f_a is a p.d.f. $\forall a \in [0,1]$.

(b) The moments α_n are the same $\forall a \in [0,1]$.

(c) The series $\sum_{n=0}^{\infty} \alpha_n \frac{(it)^n}{n!}$ diverges for all $t \neq 0$.

Definition 1.53 (Moment generating function)
We define the moment generating function (m.g.f.) of a r.v. X as

$$M(t) = E\left[e^{tX}\right], \quad t \in (-r, r) \subset I\!R,\ r > 0,$$

assuming that there exists $r > 0$ such that the integral exists for all $t \in (-r, r)$. If such $r > 0$ does not exist, then we say that the m.g.f. of X does not exist.

Remark 1.18 Remember that the c.f. always exists unlike the m.g.f.

Proposition 1.22 (Properties of the m.g.f.)
If there exists $r > 0$ such that the series

$$\sum_{n=0}^{\infty} \frac{(tx)^n}{n!}$$

is uniformly convergent in $(-r, r)$, where r is called the radius of convergence of the series, then it holds that

(a) The n-th moment α_n exists and if finite, $\forall n \in I\!N$;

(b) The n-th derivative of $M(t)$, evaluated at $t = 0$, exists and it satisfies $M^{n)}(0) = \alpha_n$, $\forall n \in I\!N$;

(c) $M(t)$ can be expressed as

$$M(t) = \sum_{n=0}^{\infty} \frac{\alpha_n}{n!} t^n, \quad t \in (-r, r).$$

Remark 1.19 Under the assumption of Proposition 1.22, the moments $\{\alpha_n\}_{n=0}^{\infty}$ determine the d.f. F.

Remark 1.20 It might happen that $M(t)$ exists for t outside $(-r, r)$, but that it cannot be expressed as the series $\sum_{n=0}^{\infty} \frac{\alpha_n}{n!} t^n$.