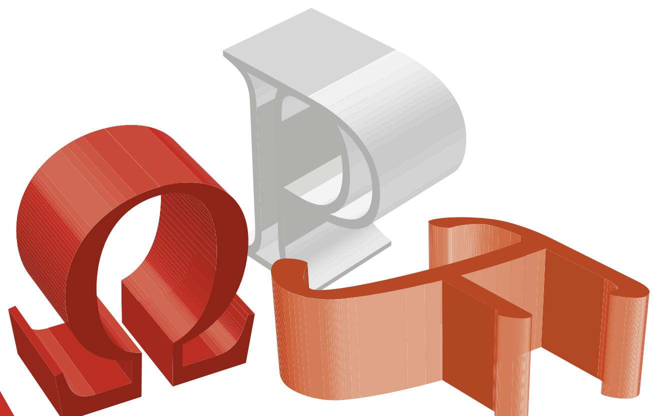
académica

An Introduction to Random Variables, Random Vectors and Stochastic Processes

Julia Calatayud Gregori | Juan Carlos Cortés López | Marc Jornet Sanz | Rafael Jacinto Villanueva Micó



Editorial
Universitat Politècnica de València

Julia Calatayud Gregori
Juan Carlos Cortés López
Marc Jornet Sanz
Rafael Jacinto Villanueva Micó

An Introduction to Random Variables, Random Vectors and Stochastic Processes

Editorial
Universitat Politècnica
de València

Collection *Académica*

To cite this publication please use the following reference: Calatayud Gregori, J.; Cortés López, J.C; Jornet Sanz, M.; Villanueva Micó, R.J. (2019). *An Introduction to Random Variables, Random Vectors and Stochastic Processes,* Valencia: Universitat Politècnica de València

sale: www.lalibreria.upv.es / Ref.: 0552_05_01_01

Printed: Byprint Percom, sl

ISBN: 978-84-9048-824-9
Print on demand

If the reader detects a mistake in the book or wishes to contact the authors, he can send an email to edicion@editorial.upv.es

Printed in Spain

Abstract

This book is aimed at covering the bases on random variables, random vectors and stochastic processes, necessary to be able to address the study of stochastic models based mainly on random and stochastic differential equations. The approach of the text is fundamentally practical. The theoretical results, including demonstrations of a more constructive nature, are combined with numerous examples and exercises chosen with the aim of instructing in fundamental ideas and interpretations. At the end of each chapter two appendices have been included. The first appendix contains a collection of carefully chosen problems for the reader to work on the main contents of the chapter, and also to study, through the proposed problems, some theoretical extensions that have not been dealt with throughout the corresponding chapter. Therefore, solving these proposed exercises is an excellent opportunity to go beyond the contents developed in each chapter. In the second appendix, we have included the resolution of some exercises using the *Mathematica*® software. These exercises have been selected just to illustrate how to carry out basic computations.

The reader can find abundant bibliography to expand the contents of this book. Some texts, on which we have based part of the contents of the chapters, are: for random variables and vectors, Casella and Berger 2006; De Groot 1988; Quesada and García 1988, and for stochastic processes, Allen 2010; Çinlar 1975; Durret 2016; Karlin 1966; Karlin and Taylor 1981; Prabhu 2007. The following texts can also be very useful, if the interested reader wants to connect the contents of this book with the theory and applications of random differential equations (Soong 1973) and Itô stochastic differential equations (Allen 2007; Kloeden and Platen 1992).

Finally, we want to express our deepest thanks to Elena López Navarro for her help in solving and preparing technical aspects of this book.

Valencia, fall 2019

Julia Calatayud
jucagre@doctor.upv.es

Juan Carlos Cortés
jccortes@imm.upv.es

Marc Jornet
marjorsa@doctor.upv.es

Rafael Villanueva
rjvillan@imm.upv.es

Contents

Chapter 1

Random Variables

In this chapter we will introduce the main definitions, concepts, properties and results related to univariate random variables. We will start by defining probability spaces, which consist of a sample space, a σ–algebra of events and a probability measure. Having an underlying probability space, one can define the concept of random variable as a Borel measurable map from the sample space. Each random variable has an associated probability distribution, which is described through the distribution function, probability density or mass function, moment generating function, characteristic function, etc. The main statistical properties of a random variable are obtained via probabilistic operators, such as the expectation and the variance. These probabilistic operators satisfy certain important inequalities that will be both stated and proved. We will study the Hilbert space $\mathrm{L}_2(\Omega)$ of random variables with well-defined expectation and variance. Finally, we will introduce the different types of convergence of a sequence of random variables.

1.1 Preliminaries on Probability Spaces and Univariate Random Variables

Roughly speaking, Probability Theory studies random experiments. The set of all possible outcomes of the random experiment is called sample space. Each event (collection of outcomes) has a certain probability of occurrence. A real univariate random variable associates each outcome with a real value. Let us see some examples of these mathematical objects, and afterwards we will proceed with the formal definitions.

EXAMPLE 1.1.1 *Tossing coins.*

Down below, we describe the relevant objects of the random experiment of tossing two fair coins from a probabilistic standpoint.

- *Random experiment: Tossing a pair of fair coins (or equivalently, flipping a coin twice), which may give as a result heads-heads* (h,h)*, heads-tails* (h,t)*, tails-heads* (t,h) *or tails-tails* (t,t)*.*
- *Sample space: The set of possible outcomes, that is,*

$$\Omega = \{\omega_1 = (h,h);\ \omega_2 = (h,t);\ \omega_3 = (t,h);\ \omega_4 = (t,t)\}.$$

- *Outcomes:* ω_1*,* ω_2*,* ω_3 *and* ω_4*.*
- *Elementary events:*

$$E_1 = \{\omega_1\}, \quad E_2 = \{\omega_2\}, \quad E_3 = \{\omega_3\}, \quad E_4 = \{\omega_4\}.$$

- *Compounded events: In Table 1.1, we describe some compounded events. Notice that, to express events mathematically, knowing Set Theory becomes very useful. In particular, being proficient in the handling of unions* ($\cup$)*, intersections* ($\cap$) *or complementaries* (^{c}) *of sets is very important.*

Events	***Mathematical description***
$A =$ *to get at least a head*	$A = E_1 \cup E_2 \cup E_3 = (E_4)^c$
$B =$ *to get just a head*	$B = E_2 \cup E_3 = (E_1 \cup E_4)^c$
$C =$ *to get 3 heads*	$C = \emptyset = \Omega^c$
$D =$ *to get at least a head or a tail*	$D = \Omega = \emptyset^c$

Table 1.1: Some compounded events expressed using operations sets. Example 1.1.1.

- *Random variables: We exhibit two examples of univariate random variables* X *and* Y*. The first one counts the number of heads when tossing two coins. The second one computes the difference in absolute value between the number of heads and tails. Notice that both* X *and* Y *are maps from* Ω *to* $\mathbb{R}$*, which are evaluated at each outcome* $\omega \in \Omega$*.*

$$X = \textit{number of heads} \Rightarrow \begin{array}{rccl} X: & \Omega & \to & \mathbb{R} \\ & \omega_i & \to & X(\omega_i) \in \{0,1,2\} \end{array}$$

$$X(\omega_1) = 2, \quad X(\omega_2) = X(\omega_3) = 1, \quad X(\omega_4) = 0.$$

$$Y = \textit{difference in absolute value between heads and tails}$$

$$\begin{array}{rccl} Y: & \Omega & \to & \mathbb{R} \\ & \omega_i & \to & Y(\omega_i) \in \{0, 2\} \end{array}$$

$$Y(\omega_1) = Y(\omega_4) = 2, \quad Y(\omega_2) = Y(\omega_3) = 0.$$

Hence, we say, informally, that a real random variable is a function with domain Ω and codomain or support ($\mathcal{S}$) the set of the real numbers $\mathbb{R}$. In the previous example:

$$\mathcal{S}(X) = \{0, 1, 2\}, \qquad \mathcal{S}(Y) = \{0, 2\}.$$

Example 1.1.2 *Trading assets.*

Down below, we describe the random experience concerning stocks from a probabilistic point of view.

- *Random experiment*[a]*: Stock traded in the Spanish index IBEX-35.*
- *Sample space:*

$$\Omega = \{\textit{every social, economic outcome that determines the stock prices}\}.$$

- *Outcomes:* $\omega \in \Omega$ *(any social, economic outcome).*
- *Elementary events:*

$$E_\omega = \{\omega\}, \ \omega \in \Omega.$$

- *Compounded events:* $A = \bigcup_{j \in J} \omega_j$ = *a set of social, economic outcomes.*
- *Random variable:*

$$X = \textit{value of a share of ACS tomorrow.}$$

We will be interested in evaluating probabilities like:

$$\mathbb{P}[\{\omega \in \Omega : \ 36 < X(\omega) \leq 40\}], \quad \mathbb{P}[\{\omega \in \Omega : \ X(\omega) \geq 36.9\}].$$

These probabilities will depend on the probability distribution associated to X*.*

Observe that, in contrast with the random experiment of flipping two coins, now we do not know the sample space Ω *explicitly. This is not important, as we are just interested in the codomain of the random variable* X*, which is a subset of* $\mathbb{R}$*.*

[a] If we knew how the variables that determine the value of the shares of this Spanish financial index change, it would not be a random experience!

To formalize and better understand these concepts, we need to introduce important results belonging to the realm of Probability Theory. Both experiments described in Examples 1.1.1 and 1.1.2 are not predictable, in the sense that they appear according to a random mechanism that is too complex to be understood by using deterministic tools.

To study problems associated with the random variable X, one first collects relevant subsets of Ω, the events, in a class $\mathcal{F}_\Omega$ called a σ–field or σ–algebra. In order for $\mathcal{F}_\Omega$ to contain all those relevant events, it is natural to include all the ω in the sample space Ω and also the union, difference, and intersection of any events in $\mathcal{F}_\Omega$, the set Ω and its complement, the empty set $\emptyset$.

Coming back to Example 1.1.2 of the market share, if we consider the price X of a stock, not only should the events $\{\omega \in \Omega : X(\omega) = c\}$ belong to $\mathcal{F}_\Omega$, but also

$$\mathbb{P}[\{\omega \in \Omega : a < X(\omega) < b\}],\ \mathbb{P}[\{\omega \in \Omega : b < X(\omega)\}],\ \mathbb{P}[\{\omega \in \Omega : X(\omega) \leq a\}],$$

and many more events that may be relevant. So, it is natural to require that elementary operations such as $\cup$ (union), $\cap$ (intersection), ^{c} (complementary) on the events of $\mathcal{F}_\Omega$ will not land outside the class $\mathcal{F}_\Omega$. This is the intuitive meaning of a σ–algebra.

Definition 1.1.1 *σ–field (or σ–algebra) and measurable space.*

Given a set Ω, a σ–field $\mathcal{F}_\Omega$ is a collection of subsets of Ω satisfying the following conditions:

i) $\emptyset, \Omega \in \mathcal{F}_\Omega$ (we really only need to impose that either $\emptyset \in \mathcal{F}_\Omega$ or $\Omega \in \mathcal{F}_\Omega$, because of the next condition ii)).

ii) If $A \in \mathcal{F}_\Omega$, then $A^c \in \mathcal{F}_\Omega$.

iii) If $A_1, A_2, \ldots \in \mathcal{F}_\Omega$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}_\Omega$.

The pair $(\Omega, \mathcal{F}_\Omega)$ is called a measurable space (or probabilizable space). The elements of $\mathcal{F}_\Omega$ are referred to as measurable sets (or events in the Probability Theory).

Using the previous conditions adequately one can deduce that many other sets lie in $\mathcal{F}_\Omega$:

- Finite or countable infinite intersections:

$$A_1 \cap A_2 = ((A_1)^c \cup (A_2)^c)^c \in \mathcal{F}_\Omega, \qquad \bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} (A_i)^c \right)^c \in \mathcal{F}_\Omega.$$

- Differences: $A \setminus B = A \cap B^c$.

To summarize, a σ–algebra is closed under finite and countable infinite unions, intersections, differences, etc.

Example 1.1.3 *Some elementary σ–fields.*

We show the most elementary σ–fields, say the trivial one, $\mathcal{F}_1$, the one associated to a partition, $\mathcal{F}_2$, and the whole set of parts of Ω (its power set), $\mathcal{F}_3$:

- $\mathcal{F}_1 = \{\emptyset, \Omega\}$, *the smallest σ–field (the trivial one).*
- $\mathcal{F}_2 = \{\emptyset, \Omega, A, A^c\}$, *where $A \subset \Omega$.*
- $\mathcal{F}_3 = 2^\Omega = \{A : A \subset \Omega\}$, *the biggest σ–field (called the power set of Ω).*

In general the power set 2^Ω is unnecessarily too big. This motivates the concept of σ–field generated by a collection of sets. One can prove that, given a collection $\mathfrak{C}$ of subsets of Ω, there exists the smallest σ–field $\sigma(\mathfrak{C})$ on Ω containing $\mathfrak{C}$. We call $\sigma(\mathfrak{C})$ the σ–field generated by $\mathfrak{C}$. Notice that

$$\sigma(\mathfrak{C}) = \bigcap \{\mathcal{G} : \mathfrak{C} \subset \mathcal{G}, \mathcal{G} \text{ is } \sigma\text{–algebra}\}.$$

If $\mathfrak{C}$ were a σ–algebra, then it would coincide with $\sigma(\mathfrak{C})$.

Example 1.1.4 *σ–field generated by a collection of sets.*

All the basic σ–algebras presented in Example 1.1.3 are actually σ–algebras generated by a collection $\mathfrak{C}$ of subsets of Ω:

- $\mathcal{F}_1 = \{\emptyset, \Omega\} = \sigma(\emptyset)$, *i.e.,* $\mathfrak{C} = \emptyset$.
- $\mathcal{F}_2 = \{\emptyset, \Omega, A, A^c\} = \sigma(\{A\})$, *i.e.,* $\mathfrak{C} = \{A\}$.
- $\mathcal{F}_3 = 2^\Omega = \sigma(\mathcal{F}_3)$, *i.e.,* $\mathfrak{C} = \mathcal{F}_3$.

Furthermore, if we consider a finite partition, say $\mathcal{P} = \{A_1, \dots, A_n\}$ of Ω, the σ–field generated by $\mathcal{P}$ is made up of all the unions $A_{i_1} \cup \dots A_{i_n}$, where $\{i_1, \dots, i_n\}$ is an arbitrary subset of $\{1, \dots, n\}$. Therefore, the σ–field $\sigma(\mathcal{P})$ has 2^n elements.

EXERCISE 1.1.1 *Computing some σ–algebras.*

Consider the random experience of flipping a coin twice (see the Example 1.1.1 for the notation). Compute:

a) The smallest σ–algebra containing both $\{\omega_1\}$ and $\{\omega_2\}$, which is denoted by $\sigma(\{\omega_1, \omega_2\})$.
b) The biggest σ–algebra containing $\{\omega_1\}$ and $\{\omega_2\}$.

Solution: Recall the definition of the sample space as

$$\Omega = \{\omega_1 = (h, h),\ \omega_2 = (h, t),\ \omega_3 = (t, h),\ \omega_4 = (t, t)\}.$$

It contains the set of possible outcomes.

a) We have

$$\sigma(\{\omega_1, \omega_2\}) = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \{\omega_1, \omega_3, \omega_4\}, \{\omega_2, \omega_3, \omega_4\}, \Omega\}.$$

b) Recall that the biggest σ–algebra is always the power set 2^Ω.

An important σ–algebra on $\mathbb{R}$, called the σ–algebra of Borel, is presented in the following example. In the context of random variables, this is in fact the most important σ–algebra on $\mathbb{R}$.

EXAMPLE 1.1.5 *σ–field of Borel.*

The σ–field of Borel is defined as the σ–algebra generated by the collection of semi-open intervals $(a, b]$ in $\Omega = \mathbb{R}$:

$$\mathcal{F}_{\mathbb{R}} = \mathcal{B}_{\mathbb{R}} = \sigma(\{(a, b] : -\infty < a < b < \infty\}).$$

All the intervals of the form $(a, b]$, unions of such intervals and intersections and complements of all the resulting sets belong to this σ–algebra. This results in a wide variety of intervals (open, semi-open and closed) as well as other sets that belong to the σ–field $\mathcal{B}_{\mathbb{R}}$. For instance:

- $(a, b) = \bigcup_{n=1}^{\infty} (a, b - \frac{1}{n}] \in \mathcal{B}_{\mathbb{R}}$.
- $\{x\} = \bigcap_{n=1}^{\infty} (x - \frac{1}{n}, x] \in \mathcal{B}_{\mathbb{R}}$.

Although it is possible to construct odd sets that do not belong to $\mathcal{B}_{\mathbb{R}}$, in practice nearly any event that we may imagine belongs to $\mathcal{B}_{\mathbb{R}}$.

It is important to emphasize that alternative but equivalent definitions for $\mathcal{B}_{\mathbb{R}}$ exist:

$$\begin{aligned}\mathcal{B}_{\mathbb{R}} &= \sigma(\{[a,b] : -\infty < a < b < \infty\}),\\ \mathcal{B}_{\mathbb{R}} &= \sigma(\{[a,b) : -\infty < a < b < \infty\}),\\ \mathcal{B}_{\mathbb{R}} &= \sigma(\{(a,\infty) : -\infty \le a < \infty\}),\\ \mathcal{B}_{\mathbb{R}} &= \sigma(\{(-\infty,b) : -\infty < b \le \infty\}),\end{aligned}$$

etc. These equivalences may be proved by establishing the double inclusion, $\subset$ and $\supset$.

The σ–algebra $\mathcal{B}_{\mathbb{R}}$ may also be defined as the one generated by the open subsets of $\mathbb{R}$.

Given $A \subset \mathbb{R}$, we may define the σ–algebra of Borel on A as

$$\mathcal{B}_A = \{A \cap B : B \in \mathcal{B}_{\mathbb{R}}\}.$$

This is the σ–algebra generated by the open sets in the relative topology of A, or by the relative intervals in A (i.e., $(a,b] \cap A$).

Definition 1.1.2 *Probability measure.*

Given Ω a set and a σ–algebra $\mathcal{F}_\Omega$ of Ω, a probability measure is a map $\mathbb{P} : \mathcal{F}_\Omega \to [0,1]$ satisfying the following conditions:

i) The probability of the whole set Ω is 1: $\mathbb{P}[\Omega] = 1$.

ii) Additivity property: $\mathbb{P}\left[\bigcup_{n=1}^{\infty} A_n\right] = \sum_{n=1}^{\infty} \mathbb{P}[A_n]$ if $A_n \cap A_m = \emptyset$, $n \neq m$: $n, m \geq 1$.

From these conditions many others can be derived:

Proposition 1.1.1 *Properties of the probability measure.*

The probability measure $\mathbb{P}$ satisfies the following properties:

i) The probability of the empty set is zero, $\mathbb{P}[\emptyset] = 0$.

ii) For each $A \in \mathcal{F}_\Omega$, $\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$.

iii) *For each* $A, B \in \mathcal{F}_\Omega$,

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B].$$

iv) *The measure* $\mathbb{P}$ *is monotonic: If* $A \subset B$, *then* $\mathbb{P}[A] \leq \mathbb{P}[B]$.

v) *Subadditivity property (Boole's inequality):*

$$\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] \leq \sum_{i=1}^{\infty} \mathbb{P}[A_i], \quad \forall A_1, A_2, \ldots \in \mathcal{F}_\Omega.$$

vi) *Continuity property: If* $\{A_n : n \geq 1\} \subset \mathcal{F}_\Omega$ *is an increasing sequence of events (i.e.,* $A_n \subset A_{n+1}$, *for each* $n \geq 1$*) with* $A = \cup_{n=1}^{\infty} A_n$, *then* $\mathbb{P}[A] = \lim_{n\to\infty} \mathbb{P}[A_n]$.

Analogously, if $\{A_n : n \geq 1\} \subset \mathcal{F}_\Omega$ *is a decreasing sequence of events (i.e.,* $A_n \supset A_{n+1}$, *for each* $n \geq 1$*) with* $A = \cap_{n=1}^{\infty} A_n$, *then* $\mathbb{P}[A] = \lim_{n\to\infty} \mathbb{P}[A_n]$.

Proof:

i) *Take* $A_1, A_2, \ldots = \emptyset$ *and apply the additivity property:* $\mathbb{P}[\emptyset] = \sum_{n=1}^{\infty} \mathbb{P}[\emptyset]$. *This is only possible if* $\mathbb{P}[\emptyset] = 0$.

ii) *In the additivity property, take* $A_1 = A$, $A_2 = A^c$, *and* $A_3 = A_4 = \ldots = \emptyset$. *Then* $1 = \mathbb{P}[\Omega] = \mathbb{P}[A] + \mathbb{P}[A^c] + \sum_{i=3}^{\infty} \mathbb{P}[\emptyset] = \mathbb{P}[A] + \mathbb{P}[A^c]$.

iii) *Let* $A_1 = A \cap B^c$, $A_2 = B \cap A^c$, $A_3 = A \cap B$, *and* $A_4 = A_5 = \ldots = \emptyset$. *Then* $A_i \cap A_j = \emptyset$ *for* $i \neq j$ *and* $A \cup B = A_1 \cup A_2 \cup A_3$. *By the additivity property,* $\mathbb{P}[A \cup B] = \mathbb{P}[A_1] + \mathbb{P}[A_2] + \mathbb{P}[A_3]$. *Notice that* $A = A_1 \cup A_3$ *and* $B = A_2 \cup A_3$, *with* $A_1 \cap A_3 = A_2 \cap A_3 = \emptyset$. *Then* $\mathbb{P}[A] = \mathbb{P}[A_1] + \mathbb{P}[A_3]$ *and* $\mathbb{P}[B] = \mathbb{P}[A_2] + \mathbb{P}[A_3]$. *Thus, we derive* $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$.

iv) *If* $A \subset B$, *then* $B = A \cup (B \backslash A)$, *therefore* $\mathbb{P}[B] = \mathbb{P}[A] + \mathbb{P}[B \backslash A] \geq \mathbb{P}[A]$.

v) *Take* $B_1 = A_1$, $B_i = A_i \backslash \cup_{j=1}^{i-1} A_j$, *for* $i \geq 2$. *Notice that* $\mathbb{P}[B_i] \leq \mathbb{P}[A_i]$, $i \geq 1$, *by the monotony of the probability measure. The sets* $B_1, B_2, \ldots$ *are pairwise disjoint and* $\cup_{i=1}^{\infty} B_i = \cup_{i=1}^{\infty} A_i$. *Then, by the additivity property,*

$$\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] = \mathbb{P}\left[\bigcup_{i=1}^{\infty} B_i\right] = \sum_{i=1}^{\infty} \mathbb{P}[B_i] \leq \sum_{i=1}^{\infty} \mathbb{P}[A_i].$$

vi) *Suppose that* $\{A_n : n \geq 1\} \subset \mathcal{F}_\Omega$ *is an increasing sequence of events. Take* $B_1 = A_1$, $B_i = A_i \backslash \cup_{j=1}^{i-1} A_j = A_i \backslash A_{i-1}$, *for* $i \geq 2$. *Notice that* $\mathbb{P}[B_i] = \mathbb{P}[A_i] - \mathbb{P}[A_{i-1}]$. *Then*

$$\mathbb{P}[A] = \mathbb{P}\left[\bigcup_{i=1}^{\infty} B_i\right] = \sum_{i=1}^{\infty} \mathbb{P}[B_i] = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}[B_i] = \lim_{n\to\infty} \mathbb{P}[A_n].$$

The case in which $\{A_n : n \geq 1\} \subset \mathcal{F}_\Omega$ *is a decreasing sequence of events is analogous, by working with the complementary* A_n^c *instead.* ☐

Example 1.1.6 *Bonferroni's inequality.*

In Proposition 1.1.1–iii) we have shown that $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$. *Since* $\mathbb{P}[A \cup B] \leq 1$ *and after some rearranging,*

$$\mathbb{P}[A \cap B] \geq \mathbb{P}[A] + \mathbb{P}[B] - 1.$$

This inequality is a special case of what is known as Bonferroni's inequality (see Exercises 3 and 4 of Appendix 1.A). Bonferroni's inequality allows us to bound the probability of a simultaneous event (the intersection) in terms of the probabilities of the individual events. This inequality is particularly useful when it is difficult (or even impossible) to calculate the intersection probability, but some idea of the size of this probability is desired. Let us suppose that A *and* B *are two events and each has probability* 0.95. *Then the probability that both will occur is bounded below by*

$$\mathbb{P}[A \cap B] \geq \mathbb{P}[A] + \mathbb{P}[B] - 1 = 0.95 + 0.95 - 1 = 0.90.$$

Observe that unless the probabilities of the individual events are sufficiently large, Bonferroni's bound is a useless (but correct!) negative number.

Definition 1.1.3 *Probability space.*

The triplet $(\Omega, \mathcal{F}_\Omega, \mathbb{P})$ *where:*

i) Ω *is the sample space,*

ii) $\mathcal{F} \subset 2^\Omega$ *is the* σ*–field of* Ω,

iii) $\mathbb{P} : \mathcal{F}_\Omega \to [0, 1]$ *is the probability measure,*

is called a probability space.

So far we have introduced the concept of random variable in a rough sense. Next we formalize its definition.

Definition 1.1.4 *Real random variable.*

We say that $X : \Omega \to \mathbb{R}$ is a real random variable if

$$X^{-1}(A) \in \mathcal{F}_\Omega, \quad \forall A \in \mathcal{B}_\mathbb{R},$$

being $\mathcal{B}_\mathbb{R}$ the σ–algebra of Borel. In words of measure theory, X is Borel measurable.

According to this definition, given $B \in \mathcal{B}_\mathbb{R}$, the set $\{\omega \in \Omega : X(\omega) \in B\} = \{X \in B\}$ is an event of $\mathcal{F}_\Omega$, whose probability can be computed:

$$\mathbb{P}\left[\{\omega \in \Omega : X(\omega) \in B\}\right] = \mathbb{P}\left[X \in B\right].$$

Example 1.1.7 *Probability measure in the random experiment of tossing a pair of fair coins.*

In the first random experiment of tossing a pair of fair coins, see Example 1.1.1, we define $\mathbb{P}[\{\omega_i\}] = \frac{1}{4}$, for each $i \in \{1, 2, 3, 4\}$. Then

$$\mathbb{P}\left[\{\omega \in \Omega : X(\omega_1) = 2\}\right] = \mathbb{P}\left[\{\omega \in \Omega : X(\omega_4) = 0\}\right] = 1/4.$$

Example 1.1.8 *Flipping a coin until a tail shows.*

For this random experiment one has the following objects:

- $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots\}$, $\omega_i =$ *the outcome where $i-1$ tosses are heads (H) and the i–th toss is a tail (T): $\omega_1 = T$, $\omega_2 = (H, T)$, $\omega_3 = (H, H, T)$, etc.*
- $\mathcal{F}_\Omega = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \ldots, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \ldots\}$ *is a σ–algebra of Ω such that $\{\omega_i\} \in \mathcal{F}_\Omega$ (it is the power set of Ω).*
- *Event: $B = \{$the first tail occurs after an odd number of tosses$\}$, i.e., $B = \{\omega_1, \omega_3, \omega_5, \ldots\} \in \mathcal{F}_\Omega$.*
- $\mathbb{P}\left[\{\omega_i\}\right] = \left(\frac{1}{2}\right)^i$ *defines a probability measure in Ω.*

On account of the previous theory, one gets:

$$\mathbb{P}\left[B\right] = \sum_{i=1}^{\infty} \mathbb{P}\left[\{\omega_{2i-1}\}\right] = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{2i-1} = \frac{2}{3}.$$

1.2 Conditional Probability

So far all the probabilities have been calculated with reference to the sample space Ω, however often we have additional information about random experiments that allow us to update the sample space. In such a case we want to be able to update probabilities taking into account the new information. This leads to the following concept.

Definition 1.2.1 *Conditional probability.*

If $A, B \in \mathcal{F}_\Omega$ and $\mathbb{P}[B] > 0$, then the conditional probability of A given B is defined as

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

Remark 1.2.1 *Some observations about conditional probability.*

- *Analogously to Definition 1.2.1, we can define*

  $$\mathbb{P}[B|A] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[A]}, \qquad A, B \in \mathcal{F}_\Omega,$$

  provided $\mathbb{P}[A] > 0$.

- *Since $\mathbb{P}[B|B] = 1$, intuitively we can think that the original sample space has been updated to B. In particular, if $A \cap B = \emptyset$, i.e., A and B are disjoint (also called incompatible or mutually disjoint) events, $\mathbb{P}[A|B] = \mathbb{P}[B|A] = 0$ as expected.*

Example 1.2.1 *Computing conditional probabilities.*

Let us suppose that a hat contains ten cards numbered from 1 to 10. The sample space of the random experiment of drawing a card from the hat is $\Omega = \{1, 2, \ldots, 10\}$, and the probability of the event $A = \{$the number of the drawn card is 10$\}$ is $\mathbb{P}[A] = 1/10$. Now, let us assume that one of the cards is drawn and that we are told that its number is at least 5. The new sample space, with the updating information, is $B = \{5, 6, 7, 8, 9, 10\}$. With this information the updated (conditional) *probability of event A is*

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} = \frac{\frac{1}{10}}{\frac{6}{10}} = \frac{1}{6} > \frac{1}{10} = \mathbb{P}[A].$$

Observe that $A \cap B = \{10\}$.

From the definitions of conditional probabilities $\mathbb{P}[A|B]$ and $\mathbb{P}[B|A]$, one derives the following 'turning around' conditional probability relationship termed *Bayes' Rule*:

$$\mathbb{P}[A|B] = \mathbb{P}[B|A]\frac{\mathbb{P}[A]}{\mathbb{P}[B]}.$$

The following result generalizes the previous formula for the case that the sample space is partitioned into a set of events.

Proposition 1.2.1 *Bayes' Rule.*

Let $A_1, A_2, \ldots$ be a partition of the sample space Ω of a probability space $(\Omega, \mathcal{F}_\Omega, \Omega)$ and let B be any event of $\mathcal{F}_\Omega$. Then, for each $i = 1, 2, \ldots$,

$$\mathbb{P}[A_i|B] = \frac{\mathbb{P}[B|A_i]}{\sum_{j=1}^{\infty} \mathbb{P}[B|A_j]\mathbb{P}[A_j]}.$$

Proof: *Since $A_1, A_2, \ldots$ is a partition of the sample space,*

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup \cdots,$$

being each $E_i = A_i \cap B$, $i = 1, 2 \ldots$, disjoint, i.e., $E_i \cap E_j = \emptyset$, $i, j = 1, 2, \ldots$, $i \neq j$ (notice that E_i may be the empty set). Then applying firstly the additivity property of the probability measure (see Definition 1.1.2) and secondly the definition of conditional probability, one gets

$$\mathbb{P}[B] = \sum_{j=1}^{\infty} \mathbb{P}[A_j \cap B] = \sum_{j=1}^{\infty} \mathbb{P}[B|A_j]\mathbb{P}[A_j].$$

This relationship is sometimes referred to as Total or Partition Probability Rule.

Now, using again the definition of conditional probability and this last relationship, one obtains

$$\mathbb{P}[A_i|B] = \frac{\mathbb{P}[B|A_i]\mathbb{P}[A_i]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B|A_i]\mathbb{P}[A_i]}{\sum_{j=1}^{\infty}\mathbb{P}[B|A_j]\mathbb{P}[A_j]}.$$

$\square$

Example 1.2.2 *Applying Bayes' Rule.*

Morse code uses 'dots' and 'dashes' in the following proportion: $3:4$*. Let us suppose that there exist interferences on the transmission line so that dots and dashes can be mistakenly received with probability* $1/5$*. Using this information and Bayes' Rule, we can compute the probability of correctly receiving a dot. Indeed, let us define the events*

$$A_1 = \{\text{dot sent}\}, \qquad A_2 = \{\text{dash sent}\},$$

which define a partition in the sample space. Observe that

$$\mathbb{P}[A_1] = \frac{3}{7}, \qquad \mathbb{P}[A_2] = \frac{4}{7}.$$

Then, if we define the event $B = \{\text{receive a dot}\}$*, then by the Total Probability Rule,*

$$\begin{aligned}\mathbb{P}[B] &= \mathbb{P}[B|A_1] \times \mathbb{P}[A_1] + \mathbb{P}[B|A_2] \times \mathbb{P}[A_2] \\ &= \frac{4}{5} \times \frac{3}{7} + \frac{1}{5} \times \frac{4}{7} = \frac{16}{35}.\end{aligned}$$

Finally, we compute the probability of correctly receiving a dot using the Bayes' Rule:

$$\mathbb{P}[A_1|B] = \frac{\frac{4}{5} \times \frac{3}{7}}{\frac{16}{35}} = \frac{3}{4}.$$

Observe that often the occurrence of an event, say B, does not make any effect on the occurrence of another event A. So,

$$\mathbb{P}[A|B] = \mathbb{P}[A].$$

If this holds, then using this relationship along with Bayes' Rule, one derives

$$\mathbb{P}[B|A] = \mathbb{P}[A|B]\frac{\mathbb{P}[B]}{\mathbb{P}[A]} = \mathbb{P}[A]\frac{\mathbb{P}[B]}{\mathbb{P}[A]} = \mathbb{P}[B].$$

As a consequence, neither A nor B affect each other. On the other hand, using the definition of conditional independence one gets

$$\mathbb{P}\left[A\right] = \mathbb{P}\left[A|B\right] = \frac{\mathbb{P}\left[A \cap B\right]}{\mathbb{P}\left[B\right]} \Rightarrow \mathbb{P}\left[A \cap B\right] = \mathbb{P}\left[A\right]\mathbb{P}\left[B\right].$$

These facts motivate the following definition:

Definition 1.2.2 *Pair of independent events.*

We say that $A, B \in \mathcal{F}_\Omega$ are independent events if

$$\mathbb{P}\left[A \cap B\right] = \mathbb{P}\left[A\right]\mathbb{P}\left[B\right].$$

The following result is very intuitive in terms of Set Theory.

Proposition 1.2.2 *Some properties about independent and complementaries events.*

Let A, $B \in \mathcal{F}_\Omega$ independent events. Then, the following pairs of events are also independent:

i) A and B^c.

i) A^c and B.

i) A^c and B^c.

Proof: Because of their similarity, we only prove the first statement.

i) According to Definition 1.2.2, it is enough to prove that $\mathbb{P}[A \cap B^c] = \mathbb{P}[A]\mathbb{P}[B^c]$. Observe that using the definition of conditional probability, A and B are inde-

pendent and by Proposition 1.1.1–ii), one obtains

$$\begin{aligned}\mathbb{P}[A \cap B^c] &= \mathbb{P}[A] - \mathbb{P}[A \cap B] &&= \mathbb{P}[A] - \mathbb{P}[A|B]\mathbb{P}[B] \\ &= \mathbb{P}[A] - \mathbb{P}[A]\mathbb{P}[B] &&= \mathbb{P}[A](1 - \mathbb{P}[B]) \\ &= \mathbb{P}[A]\mathbb{P}[B^c].\end{aligned}$$

□

So far we have introduced the concept of independence for two events, and we have obtained some consequences. Now, we extend the concept of independence for a collection of events.

Definition 1.2.3 *Mutually/simultaneous independent events and pairwise independent events.*

We say that $A_1, \ldots, A_n \in \mathcal{F}_\Omega$ are mutually or simultaneous independent events if

$$\mathbb{P}\left[A_{i_1} \cap \ldots \cap A_{i_k}\right] = \mathbb{P}\left[A_{i_1}\right] \cdots \mathbb{P}\left[A_{i_k}\right],$$

for all $1 \leq i_1 < i_2 < \ldots < i_k \leq n$, $1 \leq k \leq n$.

We say that $A_1, \ldots, A_n \in \mathcal{F}_\Omega$ are pairwise independent events if

$$\mathbb{P}[A_i \cap A_j] = \mathbb{P}[A_i]\mathbb{P}[A_j], \quad \forall\, 1 \leq i < j \leq n.$$

The following example shows why mutually or simultaneous independence has required an extremely strong condition.

Example 1.2.3 *Pairwise independence does not imply mutually independence.*

Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, with $\mathcal{F}_\Omega = 2^\Omega$ and $\mathbb{P}[\{\omega_i\}] = \frac{1}{4}$, $1 \leq i \leq 4$. Consider the events

$$A = \{\omega_1, \omega_2\}, \quad B = \{\omega_1, \omega_3\}, \quad C = \{\omega_1, \omega_4\}.$$

Then

$$A \cap B = B \cap C = A \cap C = A \cap B \cap C = \{\omega_1\},$$

so that

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B],\ \mathbb{P}[A \cap C] = \mathbb{P}[A]\mathbb{P}[C],\ \mathbb{P}[B \cap C] = \mathbb{P}[B]\mathbb{P}[C],$$

which implies pairwise independence, but

$$\mathbb{P}[A \cap B \cap C] = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}[A]\mathbb{P}[B]\mathbb{P}[C],$$

so there is no mutually independence.

Remark 1.2.2 *About the term independence.*

In general, the word "independence" stands for "mutually/simultaneous independence". Sometimes, this is also referred to as "complete independence". Recall that this is stronger than "pairwise independence". Therefore, the term "independence' in Probability Theory must be carefully used for the sake of accuracy.

1.3 Probability Distribution of a Univariate Random Variable

Definition 1.3.1 *Distribution function (d.f.).*

The function defined as

$$F_X(x) = \mathbb{P}[X \leq x] = \mathbb{P}[\{\omega \in \Omega :\ X(\omega) \leq x\}] \in [0, 1], \quad \forall x \in \mathbb{R},$$

is the d.f. $F_X(x)$ of the random variable X.

Proposition 1.3.1 *Properties of the d.f.*

The d.f. F_X of any arbitrary random variable, say X, possesses the following key properties:

- *i) It is nonnegative.*
- *ii) It is monotonically increasing.*
- *iii) It is right-continuous.*
- *iv)* $\lim_{x \to -\infty} F_X(x) = 0$ *and* $\lim_{x \to \infty} F_X(x) = 1$.

<u>*Proof:*</u>

i) *The nonnegativity of F_X follows from the nonnegativity of the probability measure.*

ii) *The fact that F_X is monotonically increasing follows from the monotony of $\mathbb{P}$, see Proposition 1.1.1–iv).*

iii) *The function F_X is right-continuous because of the continuity of $\mathbb{P}$, see Proposition 1.1.1–vi). Indeed, let $x \in \mathbb{R}$ and consider a sequence $x_n > x$ such that $x_n \to x$ as $n \to \infty$. Then $(-\infty, x] = \cap_{n=1}^{\infty}(-\infty, x_n]$, therefore*

$$F_X(x) = \mathbb{P}[X \leq x] = \lim_{n\to\infty} \mathbb{P}[X \leq x_n] = \lim_{n\to\infty} F_X(x_n).$$

iv) *If $x_n \to -\infty$ as $n \to \infty$, then $\cap_{n=1}^{\infty}(-\infty, x_n] = \emptyset$, so*

$$0 = \mathbb{P}[\emptyset] = \lim_{n\to\infty} \mathbb{P}[X \leq x_n] = \lim_{n\to\infty} F_X(x_n),$$

by the continuity of $\mathbb{P}$, see Proposition 1.1.1–vi). The analysis for the case $x_n \to \infty$ can be performed analogously. □

Notice that the probability of X lying in an interval $(a, b]$ or the probability of X being a point x may be calculated as follows:

$$\mathbb{P}\left[\{\omega \in \Omega :\ a < X(\omega) \leq b\}\right] = F_X(b) - F_X(a),$$

$$\mathbb{P}\left[\{\omega \in \Omega :\ X(\omega) = x\}\right] = F_X(x) - \lim_{\epsilon\to 0^+} F_X(x - \epsilon).$$

With these probabilities, we can approximate the probability of the event $\{\omega \in \Omega :\ X(\omega) \in B\}$ for very complicated subsets B of $\mathbb{R}$. Notice that F_X is a continuous function at x if and only if $\mathbb{P}[X = x] = 0$. It can be proved that, in general, the number of discontinuities of F_X is at most countable (this is because of its monotony).

Definition 1.3.2 *Distribution of a random variable.*

The probability

$$\mathbb{P}_X[B] = \mathbb{P}\left[X \in B\right] = \mathbb{P}\left[\{\omega \in \Omega :\ X(\omega) \in B\}\right],$$

for $B \in \mathcal{B}_{\mathbb{R}}$, is the probability distribution or probability law of the random variable X. The probability law of X is fully determined by its d.f., F_X.

DEFINITION 1.3.3 *Discrete distributions and probability mass function (p.m.f.).*

We say that X has a discrete probability distribution if its d.f. is expressed as

$$F_X(x) = \sum_{k:\, x_k \le x} p_k,\ x \in \mathbb{R}, \quad \textit{where} \quad p_k = \mathbb{P}\left[X = x_k\right], \begin{cases} 0 \le p_k \le 1, \\ \displaystyle\sum_{k=1}^{\infty} p_k = 1. \end{cases}$$

The set of values p_k are usually termed the p.m.f. of X.

REMARK 1.3.1 *Some important observations about p.m.f.'s and d.f.'s*

For a discretely distributed random variable X, we would like to emphasize the following facts:

- *The d.f. $F_X(x)$ can have jumps (it is right-continuous). Its plot is similar to an upstairs. See Figure 1.1 (left) in the context of Example 1.3.1.*
- *The d.f. and the corresponding distribution are discrete. A random variable with such a distribution is a discrete random variable.*
- *A discrete random variable assumes only a finite or countably infinitely many values: $x_1, x_2, \ldots$ with probabilities $p_k = \mathbb{P}\left[X = x_k\right]$, respectively. As previously indicated, p_k is referred to as the p.m.f. See Figure 1.1 (right) in the context of Example 1.3.1.*
- *The probability that X lies in a Borel set $B \in \mathcal{B}_{\mathbb{R}}$ is given by $\mathbb{P}_X[B] = \mathbb{P}[X \in B] = \sum_{k:\, x_k \in B} p_k$.*

EXAMPLE 1.3.1 *Binomial and Poisson distributions.*

In this example we show two important discrete distributions. In Table 1.2, we give the explicit expressions for the p.m.f. of the Binomial and Poisson distributions.

Random Variable	***Distribution***
$X \sim \mathrm{Bi}(n;p)$, $n \in \mathbb{N}$, $0 \le p \le 1$	$\mathbb{P}\left[X = k\right] = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, \ldots, n$
$X \sim \mathrm{Po}(\lambda)$, $\lambda > 0$	$\mathbb{P}\left[X = k\right] = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}$, $k = 0, 1, \ldots$

Table 1.2: P.m.f. for the Binomial and Poisson distributions. Example 1.3.1.

Let us see two practical problems in which the Binomial and Poisson distributions arise:

- *Binomial: Let $X =$ the number of ill patients living in separate rooms of a hospital. Assume that* 30% *are ill and the hospital has 50 patients. Calculate the following probability* $\mathbb{P}[X = 15]$.

  In this case, $X \sim \mathrm{Bi}(n = 50; p = 0.3)$, *therefore*

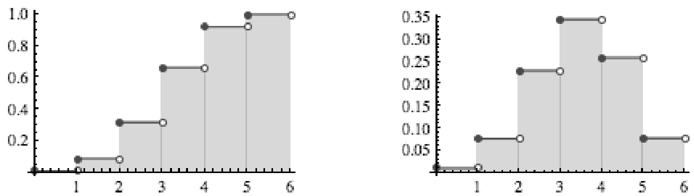
$$\mathbb{P}[X = 15] = \binom{50}{15} 0.3^{15}(1 - 0.3)^{50-15} = 0.122347.$$

- *Poisson: Let $X =$ the number of cars arriving at a petrol station every minute. Assume that, on average, 3 cars per minute usually arrive there. Calculate the following probability* $\mathbb{P}[X = 5]$.

  In this case, $X \sim \mathrm{Po}(\lambda = 3)$ *(the average of a Poisson distribution is its parameter λ, see Exercise 1.4.1). Then*

$$\mathbb{P}[X = 5] = \mathrm{e}^{-3} \cdot \frac{3^5}{5!} = 0.100819.$$

In Figure 1.1, we show the d.f. and the p.m.f. of a Binomial random variable with fixed parameters.



(a) Binomial d.f. **(b)** Binomial p.m.f.

Figure 1.1: Left: D.f. of $X \sim \mathrm{Bi}(n = 5; p = 0.6)$. Right: P.m.f. of $X \sim \mathrm{Bi}(n = 5; p = 0.6)$. Example 1.3.1.

EXAMPLE 1.3.2 *Geometric distribution.*

In this example we introduce an important discrete distribution, referred to as Geometric distribution, that appears when tossing for a head. Suppose we do an experiment that consists of tossing a coin until a head appears. Observe that this example is strongly related to the random experiment described in Example 1.1.8. Let p the probability of a head on any given toss, and define

Para seguir leyendo haga click aquí