Francesca Biagini
Massimo Campanino

Elements of Probability and Statistics

An Introduction to Probability with de Finetti's Approach and to Bayesian Statistics

UNITEXT

Springer

UNITEXT - La Matematica per il 3+2

Volume 98

More information about this series at http://www.springer.com/series/5418

Francesca Biagini · Massimo Campanino

Elements of Probability and Statistics

An Introduction to Probability with de Finetti's Approach and to Bayesian Statistics

Springer

Francesca Biagini
Department of Mathematics
Ludwig-Maximilians-Universität
Munich
Germany

Massimo Campanino
Department of Mathematics
Università di Bologna
Bologna
Italy

ISSN 2038-5722 ISSN 2038-5757 (electronic)
UNITEXT - La Matematica per il 3+2
ISBN 978-3-319-07253-1 ISBN 978-3-319-07254-8 (eBook)
DOI 10.1007/978-3-319-07254-8

Library of Congress Control Number: 2015958841

Translation from the Italian language edition: Elementi di Probabilità e Statistica di Francesca Biagini e Massimo Campanino, © Springer-Verlag Italia, Milano 2006. All rights reserved.

Cover design: Simona Colombo, Giochi di Grafica, Milano, Italy

Printed on acid-free paper

This Springer imprint is published by SpringerNature
The registered company is Springer International Publishing AG Switzerland

Nous ne possédons une ligne, un surface,
un volume que si notre amour l'occupe.

M. Proust

To Thilo and Oskar

Francesca Biagini

To my brother Vittorio

Massimo Campanino

Preface

This book is based on the lectures notes for the course, Probability and Mathematical Statistics, taught for many years by one of the authors (M.C.) and then, divided into two sections, by both authors at the University of Bologna (Italy).

We follow the approach of de Finetti, see de Finetti [1] for a complete detailed exposition. Although de Finetti [1] was conceived as a textbook of probability for mathematics students, it was also meant to illustrate the point of view of the author on the foundations of probability and mathematical statistics and discuss it in relation to prevalent approaches, resulting often of difficult access for beginners. This was the main reason that prompted us to arrange the lectures notes of our courses into a more organic way and to write a textbook for an initial class on probability and mathematical statistics.

The first five chapters are devoted to elementary probability. After that in the next three chapters we develop some elements of Markov chains in discrete and continuous time also in connection with queueing processes, and introduce basic concepts in mathematical statistics in the Bayesian approach. Then we propose six chapters of exercises, which cover most of the topics treated in the theoretical part. In the appendices we have inserted summary schemes and complementary topics (two proofs of Stirling formula). We also informally recall some elements of calculus, as this has often proved useful for the students.

This book offers a comprehensive but concise introduction to probability and mathematical statistics without requiring notions of measure theory; hence it can be used in basic classes on probability for mathematics students and is particularly suitable for computer science, physics and engineering students.

We are grateful to Springer for allowing us to publish the English version of the book. We wish to thank Elisa Canova, Alessandra Cretarola, Nicola Mezzetti and Quirin Vogel for their fundamental help with latex, for both the Italian and the English version.

Munich
Bologna
June 2015

Francesca Biagini
Massimo Campanino

Contents

Part I
Probability

Chapter 1
Random Numbers

1.1 Introduction

Probability Theory deals with the quantification of our degree of uncertainty. Its main object of interest are random entities and, in particular, random numbers. What is meant by *random number*?

A random number is a well defined number, whose value is not necessarily known. For example we can use random numbers to describe the result of a determined experiment, or the value of an option at a prefixed time, or the value of a meteorological magnitude at a given time. All these quantities have a well defined value, but may not be known either because they refer to the future and there are no means to predict their values with certainty or, even if they refer to the past, there is no available information at the moment.

We shall denote random numbers with capital letters. Even if the value of a random number is in general not known, we can speak about *the set of its possible values*, that will be denoted by $I(X)$. Certain numbers can be considered as particular cases of random numbers, whose set of possible values consists of a single element.

Example 1.1.1 Let the random numbers X, Y represent respectively the results of throwing a coin and a die. If we denote head and tail by 0 and 1 and the sides of the die with the numbers from 1 to 6, we have:

$$I(X) = \{0, 1\},$$
$$I(Y) = \{1, 2, 3, 4, 5, 6\}.$$

The random number X is:

- *upper bounded* if I(X) is upper bounded ($\sup I(X) < +\infty$);
- *lower bounded* if I(X) is lower bounded ($\inf I(X) > -\infty$);
- *bounded* if I(X) is both upper and lower bounded ($\sup I(X) < +\infty$, $\inf I(X) > -\infty$).

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8_1

Given two random numbers X and Y, we denote by $I(X, Y)$ the set of pairs of values that (X, Y) can attain. In general given n random numbers $X_1, \ldots, X_n$, we denote by $I(X_1, \ldots, X_n)$ the set of possible values that $(X_1, \ldots, X_n)$ can attain.

The random numbers X and Y are said to be *logically independent* if

$$I(X, Y) = I(X) \times I(Y),$$

where $I(X) \times I(Y)$ denotes the Cartesian product of $I(X)$ and $I(Y)$.

Similarly the random numbers $(X_1 \ldots, X_n)$ are said to be *logically independent* if $I(X_1, \ldots, X_n) = I(X_1) \times \cdots \times I(X_n)$.

Example 1.1.2 In a lottery two balls are consecutively drawn without substitution from an urn that contains 90 balls numerated from 1 to 90. Let X and Y represent the random numbers corresponding respectively to the first and the second drawing. The set of possible pairs is then

$$I(X, Y) = \{(i, j) | 1 \leq i \leq 90, 1 \leq j \leq 90, i \neq j\}.$$

Clearly $I(X, Y) \neq I(X) \times I(Y)$ as $I(X, Y)$ does not contain pairs of the type (i, i), with $i \in \{1, \ldots, 90\}$. The random numbers X and Y therefore are *not* logically independent.

By using random numbers we can perform usual arithmetic operations, obtaining again random numbers. We introduce the following operations that we will apply to random numbers. For real x and y

1. $x \vee y := \max(x, y)$;
2. $x \wedge y := \min(x, y)$;
3. $\tilde{x} := 1 - x$.

As it is easy to verify, these operations satisfy the following properties:

1. distributive property

$$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z), \tag{1.1}$$

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z); \tag{1.2}$$

2. associative property

$$x \vee (y \vee z) = (x \vee y) \vee z, \tag{1.3}$$

$$x \wedge (y \wedge z) = (x \wedge y) \wedge z; \tag{1.4}$$

3. commutative property

$$x \vee y = y \vee x, \tag{1.5}$$

$$x \wedge y = y \wedge x; \tag{1.6}$$

4. furthermore

$$\tilde{\tilde{x}} = x, \tag{1.7}$$

$$(x \vee y)\tilde{} = \tilde{x} \wedge \tilde{y}, \tag{1.8}$$

$$(x \wedge y)\tilde{} = \tilde{x} \vee \tilde{y}\,. \tag{1.9}$$

These properties are easily extended to operations to n real numbers $x_1, \ldots, x_n$.

1.2 Events

Events are a particular case of random numbers. An event E is a random number such that $I(E) \subseteq \{0, 1\}$. In the case of two events E and F, $E \vee F$ is called *logical sum* and $E \wedge F$ *logical product*. It is easy to verify that:

1. $E \vee F = E + F - EF$;
2. $E \wedge F = EF$.

Given an event E, one defines the *complementary* event E by

$$\tilde{E} = 1 - E.$$

From (1.7) we have $\tilde{\tilde{E}} = E$. From (1.8) we have

$$(E \vee F)\tilde{} = \tilde{E} \wedge \tilde{F} = (1 - E)(1 - F) = 1 - E - F + EF,$$

so that

$$E \vee F = E + F - EF.$$

Analogously

$$\begin{aligned}(E \vee F \vee G)\tilde{} = \tilde{E} \wedge \tilde{F} \wedge \tilde{G} &= (1 - E)(1 - F)(1 - G) \\ &= 1 - E - F - G + EF + EG + FG - EFG,\end{aligned}$$

so that

$$E \vee F \vee G = E + F + G - EF - EG - FG + EFG.$$

Other two operations on events are:

1. *Difference* of E and F: $E \setminus F = E - EF$.
2. *Symmetric difference* of E and F: $E \triangle F = (E \setminus F) \vee (F \setminus E) = E + F (\mathrm{mod}\ 2)$.

From now on we shall use the symbol $\vdash$ to indicate that what follows is certainly true. For example, $\vdash X \leq Y$ indicates that $I(X, Y) \subset \{(x, y) |\ x \leq y\}$.

We use the notation

$$E \subset F \text{ for } \vdash E \leq F,$$

and

$$\vdash E = F \text{ for } E \equiv F$$

that is equivalent to $E \subset F$ and $F \subset E$. When an event E is equal to 1 we say that E happens, when E is equal to 0 we say that it does not happen. The logical sum $E \vee F$ happens if and only if at least one of the events E and F takes place, whereas the logical product $E \wedge F = EF$ happens if and only if both E and F take place. The complementary event $\tilde{E}$ happens if and only if E does not happen. Note that $E \subset F$ means that E implies F, i.e. when E takes place also F does.

Definition 1.2.1 We define the following relations for events:

1. **incompatibility**: E, F are said to be incompatible if $\vdash EF = 0$;
2. **exhaustivity**: $E_1, \ldots, E_n$ are said to be exhaustive if $\vdash E_1 + \cdots + E_n \geq 1$;
3. **partition**: $E_1, \ldots, E_n$ are said to be a partition if $\vdash E_1 + \cdots + E_n = 1$ (i.e. they are exhaustive and two by two incompatible).

Example 1.2.2 An event E and its complementary $\tilde{E}$ are a partition.

Given n events $E_1, \ldots, E_n$, we can always build up a partition combining them and their complementary sets. This partition is called *partition of constituents*. We introduce the following notation. Given an event E, we put

$$E_i^* = \begin{cases} E_i \\ \tilde{E}_i \, . \end{cases}$$

A constituent of $E_1, \ldots, E_n$ is a product

$$Q = E_1^* \cdots E_n^* \, .$$

It easy to check that the set of all constituents are a partition.

In general, not all constituents are possible. If $I(E_i) = \{0, 1\}$ for $i = 1, \ldots, n$, all constituents are possible if and only if $E_1, \ldots, E_n$ are logically independent. The possible constituents are a partition. Indeed

$$1 = (E_1 + \tilde{E}_1) \ldots (E_n + \tilde{E}_n) = \sum_{Q \text{ constituent}} Q \, .$$

Impossible constituents can be obviously skipped in the sum.

If $E_1, \ldots, E_n$ are already a partition, then the possible constituents are:

$$\begin{gathered} E_1 \tilde{E}_2 \ldots \tilde{E}_n, \\ \tilde{E}_1 E_2 \tilde{E}_3 \ldots \tilde{E}_n, \\ \ldots, \\ \tilde{E}_1 \ldots \tilde{E}_{n-1} E_n, \end{gathered}$$

in this case the constituents can be identified with the events themselves.

Let us now introduce the concept of logical dependence and independence of an event E from n given events $E_1, \ldots, E_n$. The constituents Q of $E_1, \ldots, E_n$ can be classified in the following way with respect to a given event E:

(i) constituent of I type if $Q \subset E$;
(ii) constituent of II type if $Q \subset \tilde{E}$;
(iii) constituent of III type otherwise.

We say that the event E is:

- *logically dependent* from $E_1,\ldots,E_n$ if all constituents of $E_1,\ldots,E_n$ are of I or II type;
- *logically independent* from $E_1,\ldots,E_n$ if all constituents of $E_1,\ldots,E_n$ are of the III type;
- *logically semidependent* from $E_1,\ldots,E_n$ otherwise.

If E is logically dependent from $E_1,\ldots,E_n$, then we can write

$$E = \sum_{\substack{Q \text{ of I type} \\ Q \subset E}} Q\,.$$

Example 1.2.3 Let us consider two events E_1, E_2. The logical sum $(E_1 \vee E_2)$ can be written as

$$E_1 \vee E_2 = E_1 E_2 + \tilde{E}_1 E_2 + E_1 \tilde{E}_2\,.$$

In general an event E is logically dependent from $E_1,\ldots,E_n$ if and only if E can be written as $E = \Phi(E_1, \ldots, E_n)$ for some function Φ.

Example 1.2.4 Let us throw five times a coin. Let E_i be the event that we get head at the ith trial, i.e. $E_i = 1$. Set $Y = E_1 + E_2 + E_3 + E_4 + E_5$ (Y is the total number of heads in the five throws) and consider the event

$$E = (Y \geq 3).$$

Then E is logically semidependent from $E_1E_2E_3$. Indeed there are constituents of the

I type: $E_1E_2E_3 \subset E$;
II type: $\tilde{E}_1\tilde{E}_2\tilde{E}_3 \subset \tilde{E}$;
III type: $\tilde{E}_1E_2\tilde{E}_3$.

1.3 Expectation

Given a random number X, we look for a non-random number that expresses our evaluation of X. We call this quantity *expectation* of X. In economic terms, if we think of the expectation of X as a non-random gain that we judge equivalent to X.

Following de Finetti [1] the expectation $\mathbf{P}(X)$ assigned to the random number X can be defined in an operative way as follows.

Two equivalent operative definitions can be used to define the expectation:

1. **Bet method**: we think of X as a random gain (or loss, if it is negative). We have to choose a value $\mathbf{P}(X)$ (non-random) that we judge equivalent to X.
   After this choice is made, we must accept any bet with gain (or loss) given by
   $$\lambda(X - \bar{x}),$$
   where $\lambda \in \mathbb{R}$ is a constant. The corresponding coherence principle is that no choice is allowed for which there is a bet giving a certain loss. The chosen value $\bar{x}$ is our evaluation for the expectation of X.
2. **Penalty method**: in this case we choose a value $X - \bar{\bar{x}}$ and we accept to pay a penalty given by
   $$-\lambda(X - \bar{\bar{x}})^2,$$
   where $\lambda \in \mathbb{R}^+$ is a proportionality coefficient. In this case the coherence principle is that $\bar{\bar{x}}'$ is not allowed if there exits a different value $X - \bar{\bar{x}}'$ such that $\lambda(X - \bar{\bar{x}}')^2$ is certainly less than $\lambda(X - \bar{\bar{x}})^2$. The value $\bar{\bar{x}}$ that we can choose is our evaluation of the expectation $\mathbf{P}(X)$.
   It can be shown that these two operative definitions are equivalent (see [1]).

Proposition 1.3.1 (Properties of the expectation) *Given a random number X, the expectation $\mathbf{P}(X)$ has the following properties:*

1. *monotonicity:* $\inf I(X) \leq \mathbf{P}(X) \leq \sup I(X)$*;*
2. *linearity: if $X = \alpha_1 X_1 + \cdots + \alpha_n X_n$, then $\mathbf{P}(X) = \alpha_1\mathbf{P}(X_1) + \cdots + \alpha_n\mathbf{P}(X_n)$.*

Proof 1. *Monotonicity*: Assume that $\bar{x} < \inf I(X)$, then for $\lambda < 0$:

$$\vdash \lambda(X - \bar{x}) < 0.$$

If $\bar{x} > \sup I(X)$, then for $\lambda > 0$ we again get:

$$\vdash \lambda(X - \bar{x}) < 0\,,$$

i.e. a certain loss. It follows that these choices are not coherent according to the first criterium. If

$$\inf I(X) \le \bar{x} \le \sup I(X),$$

then $\vdash \left(X - \bar{\bar{x}}\right)^2 < (X - \inf I(x))$ or $\vdash \left(X - \bar{\bar{x}}\right)^2 < (X - \sup I(x))$ respectively. In this case these choices are not coherent according to the second criterium.

2. *Linearity*: Let $Z = X + Y$. Assume that we choose $\bar{z} = \mathbf{P}(Z)$, $\bar{x} = \mathbf{P}(X)$, $\bar{y} = \mathbf{P}(Y)$, then according to the bet method we are ready to accept any combination of bets on X, Y and Z that gives a total gain

$$\begin{aligned} G &= c_1(X - \bar{x}) + c_2(Y - \bar{y}) + c_3(Z - \bar{z}) \\ &= (c_1 + c_3)X + (c_2 + c_3)Y - c_1\bar{x} - c_2\bar{y} - c_3\bar{z} \end{aligned}$$

where c_1, c_2, c_3 are arbitrary constants. If we choose

$$c_1 = c_2 = -c_3,$$

(so that the random part of G cancels), then we have that the total gain is: $G = c_3(\bar{x} + \bar{y} - \bar{z})$. Then if $\bar{x} + \bar{y} - \bar{z} = 0$, one can choose c_3 so that $\vdash G < 0$. In this case this choice is not coherent according to the first criterium. On the other side if we follow the penalty method we will pay a penalty proportional to

$$-[(X - \bar{x})^2 + (Y - \bar{y})^2 + (Z - \bar{z})^2] = -[(X - \bar{x})^2 + (Y - \bar{y})^2 + (X + Y - \bar{z})^2].$$

The orthogonal projection P' of $P = (\bar{x}, \bar{y}, \bar{z})$ on the plane $z = x + y$ has a distance less or equal to the distance of P from every possible (X, Y, Z), that lies on the plane, with a strict inequality if P does not lie on the plane. Therefore by the second criterium we obtain $\bar{z} = \bar{x} + \bar{y}$. The proof that $Z = \alpha X$, $\alpha \in \mathbb{R}$, by the first or the second criterium is completely analogous.
In general, if $X = \alpha_1 X_1 + \cdots + \alpha_n X_n$, it follows that

$$\mathbf{P}(X) = \alpha_1 \mathbf{P}(X_1) + \cdots + \alpha_n \mathbf{P}(X_n). \qquad \square$$

The monotonicity of expectation implies that:

$\vdash X \ge c \Longrightarrow \mathbf{P}(X) \ge c$;
If $c_1 \le c_2, \vdash c_1 \le X \le c_2 \Longrightarrow c_1 \le \mathbf{P}(X) \le c_2$;
$\vdash X = c \Longrightarrow \mathbf{P}(X) = c$.

Remark 1.3.2 For unbounded random numbers X (for which $\inf I(X) = -\infty$, or $\sup I(X) = \infty$, or both) an evaluation of $\mathbf{P}(X)$ is not necessarily finite or even may not exist. We refer to [1] for a discussion on the definition of the expectation for unbounded random numbers.

1.4 Probability of Events

If E is an event, i.e. a random number such that $I(E) \subset \{1, 0\}$, then its expectation $\mathbf{P}(E)$ is also called *probability* of E. From monotonicity it follows that:

1. the probability of an event E is a number between 0 and 1, $0 \leq \mathbf{P}(E) \leq 1$;
2. $E \equiv 0 \Longrightarrow \mathbf{P}(E) = 0$;
3. $E \equiv 1 \Longrightarrow \mathbf{P}(E) = 1$.

When $E \equiv 1$, E is called *certain event*. If $E \equiv 0$, E is called *impossible event*. Furthermore for any given events E_1, E_2 we have that

$$\mathbf{P}(E_1 \vee E_2) = \mathbf{P}(E_1 + E_2 - E_1 E_2) \leq \mathbf{P}(E_1 + E_2)$$

and that

$$\mathbf{P}(E_1 + E_2) = \mathbf{P}(E_1) + \mathbf{P}(E_2).$$

In general for a partition $E_1, \ldots, E_n$, i.e. if $\vdash E_1 + \cdots + E_n = 1$, we have

$$\sum_{i=1}^{n} \mathbf{P}(E_i) = 1.$$

The function that assigns to the events of a partition their probabilities is called *probability distribution* of the partition. If E is logically dependent from the events $\{E_1, \ldots, E_n\}$ of a partition, then we can express the probability of E in terms of the probabilities of $E_1, \ldots, E_n$. Indeed we have

$$E = \sum_{E_i \subset E} E_i$$

so that

$$\mathbf{P}(E) = \sum_{E_i \subset E} \mathbf{P}(E_i).$$

Let us now compute the expectation of a random number X with a finite number of possible values $I(X) = \{x_1, \ldots, x_n\}$ in terms of the probabilities of events $E_i := (X = x_i)$. We use the convention that some proposition within brackets represents a quantity which is 1 when the proposition is true and 0 when it is false. We have:

$$\mathbf{P}(X) = \sum_{i=1}^{n} x_i \mathbf{P}(X = x_i). \tag{1.10}$$

Indeed

$$\begin{aligned}
\mathbf{P}(X) &= \mathbf{P}(X(E_1 + \cdots + E_n)) \\
&= \mathbf{P}(XE_1) + \cdots + \mathbf{P}(XE_n) \\
&= \sum_{i=1}^{n} \mathbf{P}(XE_i) = \sum_{i=1}^{n} \mathbf{P}(x_i E_i) \\
&= \sum_{i=1}^{n} x_i \mathbf{P}(E_i) = \sum_{i=1}^{n} x_i \mathbf{P}(X = x_i),
\end{aligned}$$

where we have used the fact that XE_i is a random number that is equal to x_i when $E_i = 1$ and to 0 when $E_i = 0$, i.e. $XE_i = x_i E_i$.

In general, if ϕ is any function $\phi : \mathbb{R} \to \mathbb{R}$, we have

$$\mathbf{P}(\phi(X)) = \sum_{i=1}^{n} \phi(x_i) \mathbf{P}(X = x_i). \tag{1.11}$$

The proof is completely analogous to the one of (1.10), which deals with the particular case $\phi(x) = x$.

Example 1.4.1 Let X be a random number representing the result of throwing a symmetric die with faces numbered from 1 to 6. By symmetry it is natural to assign the same probability (that must be $\frac{1}{6}$) to all possible values. In this case:

$$\mathbf{P}(X) = \frac{1}{6} \sum_{i=1}^{6} i = \frac{6 \cdot 7}{6 \cdot 2} = \frac{7}{2}.$$

Note that in this case the expectation does not coincide with one of the possible values of X.

Example 1.4.2 Let us throw a symmetric coin. Let $X = 1$ if the result is head and $X = 0$ if we obtain tail. Also in this case by symmetry it is natural to assign the same probability (that must be equal to $\frac{1}{2}$) to both values. In this case

$$\mathbf{P}(X) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = \frac{1}{2}.$$

1.5 Uniform Distribution on Partitions

In some situations, for reasons of symmetry, it is natural to assign the same probability to all events of a partition. This is the case of hazard games. If the events $E_1, \ldots, E_n$ are assigned the same probability, we say that the partition has the *uniform distribution*. Since the probabilities of a partition add up to 1, we have then

$$\mathbf{P}(E_i) = \frac{1}{n}.$$

Let E an event which depends logically from the partition $E_1, \ldots, E_n$, then the probability of E is given by:

$$\mathbf{P}(E) = \mathbf{P}\left(\sum_{E_i \subset E} E_i\right) = \frac{\sharp\{i|E_i \subset E\}}{n}.$$

In the case of uniform distribution on the partition, we have

$$\mathbf{P}(E) = \frac{\{i|\ E_i \subset E\}}{n}.$$

This formula is commonly expressed by saying that the probability is given by the number of favorable cases (i.e. the elements E_i contained in E), divided by the number of possible cases (i.e. the total number of E_i), as shown below:

$$\mathbf{P}(E) = \frac{\sharp\, favorable\ cases}{\sharp\, possible\ cases}. \tag{1.12}$$

This identity is valid only if the events of the partition are judged equiprobable.

Example 1.5.1 A symmetric coin is thrown n times. Let X be the random number that counts the number of heads in the n throws and let E_i be the event that the ith throw gives head. We consider the event

$$E := (X = k) = \sum_{Q \subset E} Q,$$

where Q ranges over all constituents $E_1^* \ldots E_n^*$ of $E_1, \ldots, E_n$. The symmetry of the coin leads to assign the same probability to all constituents. The probability of E is then obtained by formula (1.12). The possible cases are 2^n since a constituent is determined by n two-valued choices.

The favorable cases are $\binom{n}{k}$, since they are determined by choosing k elements out of n where the result head is obtained. Therefore

$$\mathbf{P}(E) = \binom{n}{k} \frac{1}{2^n}.$$

It follows from the properties of binomial coefficients that when n is even, the largest value for $\mathbf{P}(E)$ is obtained for $k = \frac{n}{2}$. If n is odd, the largest value for $\mathbf{P}(E)$ is obtained for $k = \frac{n-1}{2}$ and $k = \frac{n+1}{2}$.

Example 1.5.2 We perform n drawings with replacement from an urn containing N identical balls. In the urn there are H white balls and $(N - H)$ black balls. Let X be the random number of white balls which is obtained after n drawings. The set $I(x)$ of possible values of X is clearly $\{0, \dots, n\}$. In order to compute $\mathbf{P}(X = k)$ for $0 \leqslant k \leqslant n$ we can use formula (1.12), provided that we assign by symmetry reasons the same probability to the N^n sequences of length n that have exactly k white balls; their number is

$$\binom{n}{k} H^k (N - H)^{n-k},$$

since the position of the k white balls can be chosen in $\binom{n}{k}$ ways and after that we must choose a sequence of length k from the set of H white balls and one of lenght $n - k$ from the set of $N - H$ black balls. We have therefore

$$\mathbf{P}(X = k) = \binom{n}{k} \frac{H^k (N - H)^{n-k}}{N^n}.$$

Let us now consider the same problem, in the case when the drawings are made without replacement. In this case n must be less than or equal to N, as we cannot perform more than N drawings without replacement. Also X has some extra constraints, as the number X of the extracted white balls must be less than or equal to H and the number $n - X$ of extracted black balls must be less than or equal to $N - H$. Therefore

$$I(X) = \{0 \vee (n - (N - H)), \dots, n \wedge H\}.$$

In this case the possible cases are represented by all possible sets of extracted balls. An event corresponds to a set of extracted balls. The number of possible cases is then

$$\binom{N}{n}.$$

Also here by symmetry it is natural to assign the same probability to all events. If we do so, we can apply formula (1.12) and get

$$\mathbf{P}(X = k) = \frac{\binom{H}{k}\binom{N-H}{n-k}}{\binom{N}{n}},$$

for $k \in I(X)$, as the favorable cases are determined by a choice of k elements from the H white balls and $n - k$ from the $N - H$ black balls.

We could instead consider as possible cases the set of sequences of length n with distinct elements, i.e. we can take into account the order of the drawings. Of course in this case we have to take into account the order also when we count the favorable cases. The final result is the same.

1.6 Conditional Probability and Expectation

Conditional expectation and probability are very important concepts of probability. We now introduce the definition of expectation and probability under the condition that an event takes place. Let X be a random number and H an event. Conditional expectation can be defined in an operative way as ordinary expectation using bets or penalties.

1. **Bet method**: we have to choose a quantity with the agreement that we must be ready to accept any bet with gain

$$G = cH(X - \bar{x}),$$

where c is a constant (positive or negative). The chosen value is then our evaluation of the *conditional expectation* of X given by H and denoted by $\mathbf{P}(X|H)$.
2. **Penalty method**: Here we have to choose a value $\bar{\bar{x}}$ with the condition that we accept to pay a penalty.

$$P = \lambda H(X - \bar{\bar{x}})^2,$$

where λ is a *positive* constant. Note that the penalty is null when the event H does not take place, similarly as in the definition based on bets. According to this definition $\bar{\bar{x}}$ is our evaluation of the conditional expectation $\mathbf{P}(X|H)$ of X.

It can be shown, as in the case of ordinary expectation, that the two definitions are equivalent.

In the particular case when we consider an event E we speak about the *conditional probability* $\mathbf{P}(E|H)$ of E given H.

Let $I(X|H) \subseteq I(X)$ denote the set of possible values of X when H takes place. Conditional expectation enjoys the same properties as ordinary expectation, i.e. for X, Y random numbers, λ a constant and H an event, we have:

- $\inf I(X|H) \leq \mathbf{P}(X|H) \leq \sup I(X|H)$;
- $\mathbf{P}(X+Y|H) = \mathbf{P}(X|H) + \mathbf{P}(Y|H)$;
- $\mathbf{P}(\lambda X|H) = \lambda \mathbf{P}(X|H)$,

as it is easily obtained from the coherence principles.

1.7 Formula of Composite Expectation and Probability

Let X be a random number and H an event, then

$$\mathbf{P}(XH) = \mathbf{P}(H)\mathbf{P}(X|H). \tag{1.13}$$

We call (1.13) the *formula of composite expectation*. If X is also an event, (1.13) is said to be the *formula of composite probability*. In order to show that it follows from the coherence principle, let us put $z = P(XH)$, $x = P(H)$ and $y = P(X|H)$. Following the definition based on bets, this means that we are willing to accept any combination of bets with total gain:

$$\begin{aligned} G &= c_1(H-x) + c_2 H(X-y) + c_3(XH-z) \\ &= H(c_1 + (c_2+c_3)X - c_2 y) - c_1 x - c_3 z\,, \end{aligned}$$

where c_1, c_2 and c_3 are arbitrary constants. As in previous cases, let us fix c_1, c_2 and c_3 in such a way that the random part of G cancels: $c_2 = -c_3$ and $c_1 = c_2 y$. Then

$$G \;=\; -c_1 x - c_3 z \;=\; c_2(z - xy).$$

If $z \neq xy$, then it is possible to choose c_2 so that $\vdash G < 0$. Therefore by coherence principle

$$z = xy.$$

Analogously this equality follows by using the definition based on penalty. If $\mathbf{P}(H) > 0$, then

$$\mathbf{P}(X|H) = \frac{\mathbf{P}(XH)}{\mathbf{P}(H)}\,.$$

In the case of an event E the formula

$$\mathbf{P}(E|H) = \frac{\mathbf{P}(EH)}{\mathbf{P}(H)}$$

has a logical meaning, as EH is the logical product of E and H, i.e. the event that both E and H take place. In particular:

1. $E \subset H \quad \Rightarrow \quad \mathbf{P}(E|H) = \dfrac{\mathbf{P}(E)}{\mathbf{P}(H)}$;
2. $H \subset E$, that means $I(E|H) = \{1\} \quad \Rightarrow \quad \mathbf{P}(E|H) = 1$;
3. $H \subset \tilde{E}$, that means $I(E|H) = \{0\} \quad \Rightarrow \quad \mathbf{P}(E|H) = 0$.

1.8 Formula of Total Expectation and Total Probability

Given X a random number and $H_1, \ldots, H_n$ a partition, then

$$\mathbf{P}(X) = \sum_{i=1}^{n} \mathbf{P}(X|H_i)\mathbf{P}(H_i)\,. \tag{1.14}$$

We call (1.14) the *fomula of total expectation*. If X is also an event, (1.14) is said to be the *formula of total probability*. Indeed,

$$\begin{aligned}\mathbf{P}(X) &= \mathbf{P}(X \cdot 1) = \mathbf{P}(X(H_1 + \ldots + H_n)) \\ &= \mathbf{P}(XH_1 + XH_2 + \cdots + XH_n) \\ &= \sum_{i=1}^{n} \mathbf{P}(XH_i) = \sum_{i=1}^{n} \mathbf{P}(X|H_i)\mathbf{P}(H_i).\end{aligned}$$

1.9 Bayes Formula

Let E, H be events with $\mathbf{P}(H) > 0$. By applying twice the formula of total probability we obtain *Bayes' formula*:

$$\mathbf{P}(E|H) = \frac{\mathbf{P}(EH)}{\mathbf{P}(H)} = \frac{\mathbf{P}(H|E)\mathbf{P}(E)}{\mathbf{P}(H)}\,.$$

This formula is a fundamental tool in statistical inference.

Example 1.9.1 Consider an urn contain N identical balls of which some are white and some are black. Let Y be the random number of the white balls present in the urn (the composition of the urn is unknown).

The events $H_i = (Y = i)$, for $i = 0, \ldots, N$ form a partition. Let E be the event that we obtain a white ball in a drawing from the urn. Using the formula of total probability (1.14) we obtain:

$$\mathbf{P}(E) = \sum_{i=0}^{N} \mathbf{P}(E|H_i)\mathbf{P}(H_i) = \sum_{i=0}^{N} \frac{i}{N}\mathbf{P}(H_i)\,.$$

Indeed if the composition of the urn is known, i.e. if we condition with respect to H_i for some i, we can apply usual symmetry considerations and get $\mathbf{P}(E|H_i) = \frac{i}{N}$.

In the case we assign to the partition $H_0, \ldots, H_N$ the uniform distribution

$$\mathbf{P}(H_i) = \frac{1}{N+1},\ i = 0, \ldots, N$$

we get

$$\mathbf{P}(E) = \sum_{i=0}^{N} \frac{i}{N(N+1)} = \frac{1}{2}\,.$$

We now evaluate the probability that the urn contains i white balls if we have extracted a white ball. This question is answered by Bayes' formula:

$$\mathbf{P}(H_i|E) = \frac{\mathbf{P}(E|H_i)\mathbf{P}(H_i)}{\mathbf{P}(E)} = \frac{\frac{i}{N}\frac{1}{N+1}}{\frac{1}{2}} = \frac{2i}{N(N+1)}\,.$$

We see that distribution on the partition conditional to the event that a white ball is drawn is no longer uniform, but it gives higher probabilities to compositions with a large number of white balls.

1.10 Correlation Between Events

An event E is said to be *positively correlated* with the the event H if

$$\mathbf{P}(E|H) > \mathbf{P}(E).$$

Analogously E is said to be *negatively correlated* with H if

$$\mathbf{P}(E|H) < \mathbf{P}(E).$$

If $\mathbf{P}(E|H) = \mathbf{P}(E)$, we say that E is *non-correlated* with H.

If E is positively (resp. negatively) correlated with H, the information that H takes place increases (resp. decreases) our evaluation of the probability of E. When E is not correlated with H, our evaluation does not change.

When $\mathbf{P}(H) > 0$ and $\mathbf{P}(E) > 0$, one can give a symmetric formulation of correlation as it follows from the formula of composite probability. E and H are said to be:

- *positively correlated* if $\mathbf{P}(EH) > \mathbf{P}(E)\mathbf{P}(H)$;
- *negatively correlated* if $\mathbf{P}(EH) < \mathbf{P}(E)\mathbf{P}(H)$;
- *non-correlated* if $\mathbf{P}(EH) = \mathbf{P}(E)\mathbf{P}(H)$.

If E is positively correlated with H, so is $\tilde{E}$. Indeed in this case

$$\mathbf{P}(\tilde{E}|H) = 1 - \mathbf{P}(E|H) < 1 - \mathbf{P}(E) = \mathbf{P}(\tilde{E}).$$

In the same way, if E is non-correlated with H, so is $\tilde{E}$.

Example 1.10.1 We consider an urn with H white balls and $N - H$ black balls. We perform two drawings. Let E_i be the event that a white ball is extracted at the ith extraction, $i = 1, 2$. For drawings with replacement we have

$$\mathbf{P}(E_1) = \mathbf{P}(E_2) = \frac{H}{N}.$$

Indeed the urn composition in the two drawings is the same. In this case E_1 and E_2 are non-correlated, as by (1.12)

$$\mathbf{P}(E_1 E_2) = \frac{H^2}{N^2} = \mathbf{P}(E_1)\mathbf{P}(E_2).$$

Let us now consider the case of drawings without replacement. We use again formula (1.12) to compute probabilities and conditional probabilities. We have $\mathbf{P}(E_1) = \dfrac{H}{N}$ and by the formula of total probability (1.14) applied to the event E_2 and the partition E_1, $\tilde{E}_1$ we get

$$\begin{aligned}\mathbf{P}(E_2) &= \mathbf{P}(E_2|E_1)\mathbf{P}(E_1) + \mathbf{P}(E_2|\tilde{E}_1)\mathbf{P}(\tilde{E}_1) \\ &= \frac{H-1}{N-1}\frac{H}{N} + \frac{H}{N-1}(1 - \frac{H}{N}) = \frac{H}{N}.\end{aligned}$$

Here $\mathbf{P}(E_1)$ and $\mathbf{P}(E_2)$ are both equal to $\frac{H}{N}$ and $\mathbf{P}(E_1)$, $\mathbf{P}(E_2)$ are negatively correlated, as

$$\mathbf{P}(E_2|E_1) = \frac{H-1}{N-1} < \frac{H}{N} = \mathbf{P}(E_2)$$

if $0 < H < N$.

We say that two events are stochastically independent if

$$\mathbf{P}(E_1 E_2) = \mathbf{P}(E_1)\mathbf{P}(E_2).$$

When $\mathbf{P}(E_1) > 0$ and $\mathbf{P}(E_2) > 0$ this definition coincides with non-correlation. When one or both of E_1 and E_2 have 0 probability, then E_1 and E_2 are stochastically independent, as in this case $\mathbf{P}(E_1)\mathbf{P}(E_2) = 0$ and

$$\mathbf{P}(E_1 E_2) \leq \mathbf{P}(E_1) \wedge \mathbf{P}(E_2) = 0.$$

The definition of stochastic independence extends to the case of an arbitrary number of events.

Definition 1.10.2 The events $E_1, \ldots, E_n$ are said to be *stochastically independent* if for every subset $\{i_1, \ldots, i_k\}$ in $\{1, \ldots, n\}$ we have

$$\mathbf{P}(E_{i_1} \cdots E_{i_k}) = \mathbf{P}(E_{i_1}) \cdots \mathbf{P}(E_{i_k}). \tag{1.15}$$

We remark that in general n events are not stochastically independent if the events are only pairwise stochastically independent.

We shall see that if the events $E_1, \ldots, E_n$ are stochastically independent, then the events $E_1^*, \ldots, E_n^*$ are stochastically independent for every possible choice of E_i^* between E_i and $\tilde{E}_i$, for $i = 1, \ldots, n$.

Definition 1.10.3 Let $\mathcal{H} = \{H_1, \ldots, H_n\}$ be a partition. The events E_1, E_2 are said to be *stochastically independent conditionally to the partition* $\mathcal{H}$ if

$$\mathbf{P}(E_1 E_2 | H_i) = \mathbf{P}(E_1 | H_i)\mathbf{P}(E_2 | H_i) \quad \text{for all } i = 1, \ldots n\,.$$

Example 1.10.4 Let us consider an urn with unknown composition containing N identical balls, of which some are white and some are black. Let Y be the random number of white balls in the urn. We perform two drawings with replacement. Let E_i, $i = 1, 2$, be the event that in the ith drawing we extract a white ball.

Consider the partition

$$H_i = (Y = i) \quad i = 0, \ldots N.$$

It is easy to see that the events E_1 and E_2 are stochastically independent conditionally to the partition $\mathcal{H}$. We want to see whether E_1 and E_2 are stochastically independent, assuming that we assign the uniform distribution to H, i.e. $\mathbf{P}(H_i) = \frac{1}{N+1}$ for $i = 0, 1, \ldots, N$. We compute:

1. the probability of the first drawing:

$$\begin{aligned}
\mathbf{P}(E_1) &= \sum_{i=0}^{N} \mathbf{P}(E_1 | H_i)\mathbf{P}(H_i) \\
&= \frac{1}{N+1} \sum_{i=0}^{N} \frac{i}{N} \\
&= \frac{1}{N+1} \, \frac{N(N+1)}{2N} \\
&= \frac{1}{2}\,;
\end{aligned}$$

2. the probability of the second drawing:

$$\mathbf{P}(E_2) = \mathbf{P}(E_1) = \frac{1}{2};$$

3. the probability that we draw a white ball in both drawings:

$$\begin{aligned}\mathbf{P}(E_1 E_2) &= \sum_{i=0}^{N} \mathbf{P}(E_1 E_2 | H_i)\mathbf{P}(H_i) \\ &= \frac{1}{N+1} \sum_{i=0}^{N} \mathbf{P}(E_1|H_i)\mathbf{P}(E_2|H_i) \\ &= \frac{1}{N+1} \sum_{i=0}^{N} \frac{i^2}{N^2}.\end{aligned}$$

Using the fact that

$$(i+1)^3 - i^3 = 3i^2 + 3i + 1$$

we have

$$\sum_{i=0}^{N} i^2 = \sum_{i=0}^{N} \frac{(i+1)^3 - i^3}{3} - \sum_{i=0}^{N} i - \sum_{i=0}^{N} \frac{1}{3} = \frac{(N+1)^3}{3} - \frac{N(N+1)}{2} - \frac{N}{3},$$

and

$$\mathbf{P}(E_1 E_2) = \frac{(N+1)^2}{3N^2} - \frac{1}{2N} - \frac{1}{3N(N+1)}.$$

For $N \to +\infty$, $\mathbf{P}(E_1 E_2)$ tends to $\frac{1}{3}$. Therefore at least for large N, E_1 and E_2 are positively correlated. This shows that stochastic independence conditionally to a partition does not imply stochastic independence.

1.11 Stochastic Independence and Constituents

Proposition 1.11.1 *The events $E_1, \ldots, E_n$ are stochastically independent if and only if*

$$\mathbf{P}(Q) = \mathbf{P}(E_1^*) \cdots \mathbf{P}(E_n^*) \tag{1.16}$$

for every constituent $Q = E_1^* \cdots E_n^*$ of $E_1, \ldots, E_n$.

Proof $\Rightarrow$) Let $Q = E_1^* \cdots E_n^*$ be a constituent of $E_1, \dots, E_n$. Developing the products, we can express Q as a polynomial ϕ of $E_1, \dots, E_n$ where the degree in every variable is 1:

$$E_1^* \cdots E_n^* = \phi(E_1, \dots, E_n).$$

For example, consider the constituent Q of the events E_1, E_2, E_3 given by

$$Q = \tilde{E}_1 E_2 E_3 = (1 - E_1) E_2 E_3 = E_2 E_3 - E_1 E_2 E_3 .$$

Here $\phi(x_1, x_2, x_3) = x_2 x_3 - x_1 x_2 x_3 = (1 - x_1) x_2 x_3$.
If the events $E_1, \dots, E_n$ are stochastically independent, the probabilities of products factorize into products of probabilities so that

$$\begin{aligned} \mathbf{P}(Q) &= \mathbf{P}\left(\phi(E_1, \dots, E_n)\right) \\ &= \phi\left(\mathbf{P}(E_1), \dots, \mathbf{P}(E_n)\right) \\ &= \mathbf{P}(E_1^*) \cdots \mathbf{P}(E_n^*), \end{aligned}$$

where the last equality is obtained by collecting terms in ϕ and using that $\mathbf{P}(\tilde{E}_i) = 1 - \mathbf{P}(E_i)$. In the example $Q = \tilde{E}_1 E_2 E_3$. We have

$$\begin{aligned} \mathbf{P}(Q) &= \mathbf{P}\left(\tilde{E}_1 E_2 E_3\right) \\ &= \mathbf{P}\left(E_2 E_3 - E_1 E_2 E_3\right) \\ &= \mathbf{P}(E_2)\mathbf{P}(E_3) - \mathbf{P}(E_1)\mathbf{P}(E_2)\mathbf{P}(E_3) = \phi(\mathbf{P}(E_1), \mathbf{P}(E_2), \mathbf{P}(E_3)) \\ &= (1 - \mathbf{P}(E_1))\mathbf{P}(E_2)\mathbf{P}(E_3) = \mathbf{P}(\tilde{E}_1)\mathbf{P}(E_2)\mathbf{P}(E_3). \end{aligned}$$

$\Leftarrow$) We assume that (1.16) holds for all constituents of the events $E_1, \dots, E_n$. Let $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$ and $\{j_1, \dots, j_{n-k}\} = \{1, \dots, n\} \setminus \{i_1, \dots, i_k\}$. Then

$$\begin{aligned} \mathbf{P}(E_{i_1} \cdots E_{i_k}) = \mathbf{P}\left(\sum_{Q \subset E_{i_1} \cdots E_{i_k}} Q\right) \\ = \mathbf{P}(E_{i_1}) \cdots \mathbf{P}(E_{i_k}) \sum \mathbf{P}(E_{j_1}^\star \dots E_{j_{n-k}}^\star) \end{aligned}$$

where the sum ranges over all possible choices of $E_{j_l}^\star$ for $l = 1, \dots, n-k$. By collecting terms we get:

$$\begin{aligned} \mathbf{P}(E_{i_1} \dots E_{i_k}) &= \mathbf{P}(E_{i_1}) \dots \mathbf{P}(E_{i_k})[(\mathbf{P}(E_{j_1}) + \mathbf{P}(\tilde{E}_{j_1})] \dots [\mathbf{P}(E_{j_{n-k}}) + \mathbf{P}(\tilde{E}_{j_{n-k}})] \\ &= \mathbf{P}(E_{i_1}) \dots \mathbf{P}(E_{i_k}), \end{aligned}$$

since the last $n - k$ factors are all equal to 1. $\square$

1.12 Covariance and Variance

Given two random numbers X and Y, the *covariance* between X and Y is defined by

$$\mathbf{cov}(X, Y) = \mathbf{P}\left((X - \mathbf{P}(X))(Y - \mathbf{P}(Y))\right).$$

X and Y are said to be:

- *positively correlated* if $\mathbf{cov}(X, Y) > 0$;
- *negatively correlated* if $\mathbf{cov}(X, Y) < 0$;
- *non-correlated* if $\mathbf{cov}(X, Y) = 0$.

By developing the product in the definition of the covariance, we obtain:

$$\mathbf{cov}(X, Y) = \mathbf{P}(XY - \mathbf{P}(X)Y - X\mathbf{P}(Y) + \mathbf{P}(X)\mathbf{P}(Y)) = \mathbf{P}(XY) - \mathbf{P}(X)\mathbf{P}(Y).$$

The *variance* of a random number X is defined by

$$\boldsymbol{\sigma}^2(X) = \mathbf{cov}(X, X).$$

Other notations for the variance of X are $\mathbf{var}(X)$ and $\mathbf{D}(X)$. From the two expressions for the covariance we get two expressions for the variance: $\boldsymbol{\sigma}^2(X) = \mathbf{P}(X^2) - \mathbf{P}(X)^2$ and $\boldsymbol{\sigma}^2(X) = \mathbf{P}\left((X - \mathbf{P}(X))^2\right)$. From the second expression we see that

$$\boldsymbol{\sigma}^2(X) \geq 0,$$

as it is the expectation of a non-negative random number. We also define:

- *quadratic expectation*:
$$P_Q(X) = \sqrt{\mathbf{P}(X^2)};$$
- *standart deviation*:
$$\boldsymbol{\sigma}(X) = \sqrt{\boldsymbol{\sigma}^2(X)} = \mathbf{P}_Q(X - \mathbf{P}(X)).$$

Proposition 1.12.1 (Properties of covariance and variance) *Covariance and variance satisfy the following properties:*

1. *bilinearity:*
$$\mathbf{cov}(X + Y, Z) = \mathbf{cov}(X, Z) + \mathbf{cov}(Y, Z); \tag{1.17}$$

2. *behavior with respect to linear transformations:*
$$\mathbf{cov}(aX + b, cY + d) = ac\, \mathbf{cov}(X, Y), \tag{1.18}$$
$$\boldsymbol{\sigma}^2(aX + b) = a^2\boldsymbol{\sigma}^2(X). \tag{1.19}$$

Proof 1. From the definition of covariance we have

$$\begin{aligned}\mathbf{cov}(X+Y,Z) &= \mathbf{P}((X+Y)-\mathbf{P}(X+Y), Z-\mathbf{P}(Z)\\ &= \mathbf{P}((X+Y)-\mathbf{P}(X)-\mathbf{P}(Y))(Z-\mathbf{P}(Z))\\ &= \mathbf{P}(((X-\mathbf{P}(X))(Z-\mathbf{P}(Z)))+\mathbf{P}(((Y-\mathbf{P}(Y))(Z-\mathbf{P}(Z)))\\ &= \mathbf{cov}(X,Z)+\mathbf{cov}(Y,Z)\,.\end{aligned}$$

2. Again from the definition of covariance and the linearity of the expectation we have:

$$\begin{aligned}\mathbf{cov}(aX+b,cY+d) &= \mathbf{P}\left((aX+b-\mathbf{P}(aX+b))\left(cY+d-\mathbf{P}(cY+d)\right)\right)\\ &= \mathbf{P}\left((aX+b-a\mathbf{P}(X)-b)\left(cY+d-c\mathbf{P}(Y)-d\right)\right)\\ &= \mathbf{P}\left(a\left(X-\mathbf{P}(X)\right)c\left(Y-\mathbf{P}(Y)\right)\right)\\ &= ac\,\mathbf{cov}(X,Y).\end{aligned}$$

□

Proposition 1.12.2 (Variance of the sum of random numbers) *Let $X_1, \ldots, X_n$ n be random numbers, then:*

$$\begin{aligned}\sigma^2(X_1+\cdots+X_n) &= \sum_{i=1}^{n}\sigma^2(X_i)+\sum_{\substack{i,j\\ i\neq j}}\mathbf{cov}(X_i,X_j)\\ &= \sum_{i=1}^{n}\sigma^2(X_i)+2\sum_{i<j}\mathbf{cov}(X_i,X_j).\end{aligned}$$

Proof By the bilinearity property (1.17) we have:

$$\begin{aligned}\sigma^2(X_1+\cdots+X_n) &= \mathbf{cov}(X_1+\cdots+X_n, X_1+\cdots+X_n)\\ &= \sum_{i=1}^{n}\mathbf{cov}(X_i,X_i)+\sum_{i\neq j}\mathbf{cov}(X_i,X_j)\\ &= \sum_{i=1}^{n}\sigma^2(X_i)+\sum_{\substack{i,j\\ i\neq j}}\mathbf{cov}(X_i,X_j).\end{aligned}$$

□

1.13 Correlation Coefficient

It is useful to introduce an index of the correlation of two random numbers X, Y, called *correlation coefficient*. As we shall see, it has the property that if X and Y correspond to observed quantities, it does not depend on the units of measure of X and Y.

Definition 1.13.1 For X, Y random numbers with $\boldsymbol{\sigma}(X) > 0$, $\boldsymbol{\sigma}(Y) > 0$ the *correlation coefficient* of X and Y is defined by

$$\rho(X, Y) = \frac{\mathbf{cov}(X, Y)}{\boldsymbol{\sigma}(X)\,\boldsymbol{\sigma}(Y)}.$$

Let us state two important properties of the *correlation coefficient*:

1. If X, Y are random numbers with $\boldsymbol{\sigma}(X) > 0$, $\boldsymbol{\sigma}(Y) > 0$ and a, b, c, d are constants with $a \neq 0$ and $c \neq 0$, we have

$$\rho(aX + b,\ cY + d) = \operatorname{sgn}(ac)\,\rho(X, Y),$$

where $\operatorname{sgn}(x) = 1$ for $x > 0$ and $\operatorname{sgn}(x) = -1$ for $x < 0$.

Proof By using the properties (1.18) and (1.19) we get

$$\begin{aligned}\rho(aX + b,\ cY + d) &= \frac{\mathbf{cov}(aX + b,\ cY + d)}{\sqrt{\boldsymbol{\sigma}^2(aX + b)\,\boldsymbol{\sigma}^2(cY + d)}} \\ &= \frac{ac\,\mathbf{cov}(X, Y)}{|ac|\,\sqrt{\boldsymbol{\sigma}^2(X)\,\boldsymbol{\sigma}^2(Y)}} \\ &= \operatorname{sgn}(ac)\,\rho(X, Y).\end{aligned}$$

□

2. $-1 \leq \rho(X, Y) \leq 1$.
Let

$$X^* = \frac{X - \mathbf{P}(X)}{\boldsymbol{\sigma}(X)}, \qquad Y^* = \frac{Y - \mathbf{P}(Y)}{\boldsymbol{\sigma}(Y)}.$$

These are the so-called standardized random numbers: they are obtained from X, Y by means of suitable linear transformation such that $\mathbf{P}(X^*) = 0$, $\mathbf{P}(Y^*) = 0$ and $\boldsymbol{\sigma}^2(X^*) = 1$, $\boldsymbol{\sigma}^2(Y^*) = 1$ by using linearity of the expectation and (1.19). By (1.18) we get

$$\mathbf{cov}(X^*, Y^*) = \frac{\mathbf{P}\,(X^*\,Y^*)}{\boldsymbol{\sigma}(X)\,\boldsymbol{\sigma}(Y)} = \rho(X, Y).$$

Computing the variance of $X^* + Y^*$ using Proposition 1.12.1 we get:

$$\begin{aligned}0 \leq \boldsymbol{\sigma}^2(X^* + Y^*) &= \boldsymbol{\sigma}^2(X^*) + \boldsymbol{\sigma}^2(Y^*) + 2\,\mathbf{cov}(X^*, Y^*) \\ &= 2 + 2\rho(X, Y),\end{aligned}$$

so that $\rho(X, Y) \geq -1$. Similarly computing the variance of $X^* - Y^*$, we obtain

$$0 \leq \sigma^2(X^* - Y^*) = \sigma^2(X^*) + \sigma^2(-Y^*) + 2\,\mathbf{cov}(X^*, -Y^*)$$
$$= 2 - 2\rho(X, Y).$$

so that $\rho(X, Y) \leq 1$.

1.14 Chebychev's Inequality

The Chebychev's inequality allows to estimate the probability that a random number takes value far from its expectation. It can be formulated in two ways:

1. Let X be a random number with $P_Q(X) > 0$. For every $t > 0$

$$\mathbf{P}\left(|X| \geq t P_Q(X)\right) \leq \frac{1}{t^2}.$$

2. Let X be a random number with $\sigma^2(X) > 0$. Let $m = \mathbf{P}(X)$, $\forall t > 0$:

$$\mathbf{P}\left(|X - m| \geq \sigma(X)t\right) \leq \frac{1}{t^2}.$$

Proof 1. Let E be the event $E = \left(|X| \geq t P_Q(X)\right)$. We compute $\mathbf{P}\left(X^2\right)$ using the formula of total expectation with respect to the partition $E, \tilde{E}$:

$$\mathbf{P}\left(X^2\right) = \mathbf{P}\left(X^2|E\right)\mathbf{P}(E) + \mathbf{P}\left(X^2|\tilde{E}\right)\mathbf{P}\left(\tilde{E}\right).$$

Since X^2 is non-negative, the last term on the right-hand side is non-negative. Moreover $\inf I(X^2|E) \geq t^2$. Then $P_Q(X)^2 = t^2 P(X)^2$ in force of the definition of E. Therefore we have $\mathbf{P}(X)^2 \geq t^2\mathbf{P}(X^2)P(E)$. This implies the first inequality.

2. The second inequality follows from the first by applying it to the random number $Y = X - m$ and using that $\mathbf{P}_Q(Y) = \sigma(X)$. □

1.15 Weak Law of Large Numbers

Theorem 1.15.1 *(Weak law of large numbers). Let $(X_n)_{n=1,2,\ldots}$ be a sequence of random numbers such that all have the same expectation, $\mathbf{P}(X_i) = m$, the same variance $\sigma^2(X_i) = \sigma^2$ and $\mathbf{cov}(X_i, X_j) = 0$, $\forall i, j$ with $i \neq j$. If we put $S_n = X_1 + \cdots + X_n$, we have that for all $\lambda > 0$*

$$\lim_{n\to+\infty} \mathbf{P}\left(|\frac{S_n}{n} - m| \geq \lambda\right) = 0\,.$$

Proof The proof is based on the second form of Chebychev's inequality. First we compute the expectation of $\frac{S_n}{n}$:

$$\mathbf{P}\left(\frac{S_n}{n}\right) = \frac{1}{n}\left(\mathbf{P}(X_1) + \cdots + \mathbf{P}(X_n)\right) = m$$

and its variance

$$\boldsymbol{\sigma}^2\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\boldsymbol{\sigma}^2(S_n) = \frac{1}{n^2}(\boldsymbol{\sigma}^2(X_1) + \cdots + \boldsymbol{\sigma}^2(X_n)) = \frac{\boldsymbol{\sigma}^2}{n},$$

where we have used Proposition 1.12.2 and the fact that random numbers of the sequence are pairwise uncorrelated. From the second form of Chebychev's inequality we get

$$\mathbf{P}\left(|\frac{S_n}{n} - m| \geq \frac{\boldsymbol{\sigma}}{\sqrt{n}}\,t\right) \leq \frac{1}{t^2}\,.$$

Putting $\lambda = \frac{\boldsymbol{\sigma}}{\sqrt{n}}\,t$, we obtain $\frac{1}{t^2} = \frac{\boldsymbol{\sigma}^2}{n\lambda^2}$. Therefore

$$\mathbf{P}\left(|\frac{S_n}{n} - m| \geq \lambda\right) \leq \frac{\boldsymbol{\sigma}^2}{n\lambda^2}\,,$$

that tends to 0 as $n \to +\infty$. □

The quantity $\frac{S_n}{n} = \frac{E_1 + \cdots + E_n}{n}$ is called *frequence*. In this case the weak law of large numbers shows that for a large sequence of trials (events) the frequence of success is close to the probability of a single event with large probability.

Example 1.15.2 In particular one can apply the weak law of large numbers to the case of a sequence of uncorrelated events $(E_i)_{i=1,2,\ldots}$ with the same probability $\mathbf{P}(E_i) = p$. Note that for an event E_i,

$$\boldsymbol{\sigma}^2(E_i) = \mathbf{P}(E_i^2) - \mathbf{P}(E_i)^2 = \mathbf{P}(E_i) - \mathbf{P}(E_i)^2 = p(1-p)$$

so the E_i's have automatically the same variance. Hence for all $\lambda > 0$ we have

$$\mathbf{P}\left(|\frac{S_n}{n} - p| \geq \lambda\right) \to 0$$

for $n \to \infty$.

Chapter 2
Discrete Distributions

2.1 Random Numbers with Discrete Distribution

The distribution of a random number X is said to be *discrete* if there is a finite or enumerable set $A \subset I(X)$ such that $\mathbf{P}(X \in A) = 1$. This is obviously the case when $I(X)$ is itself finite or enumerable, since in this case we may take $A = I(X)$. Let $A = \{x_1, x_2, \ldots\}$ and define $p(x_i) = \mathbf{P}(X = x_i)$. In the examples of discrete distributions that we shall consider, we always have

$$\sum_{i=1}^{\infty} p(x_i) = 1 .$$

This property is not a consequence of the basic properties of expectation that we have derived from the coherence principles (from linearity and monotonicity we only get that $\sum_{i=1}^{\infty} p(x_i) \leq 1$). It can be considered as a regularity property of the expectation. See [dF] for a thorough discussion of this problem. In the following we introduce some of the most common discrete distributions.

2.2 Bernoulli Scheme

A simple and useful model from which some discrete distributions can be derived is the *Bernoulli scheme*. It can be thought of as a potentially infinite sequence of trials, each of them with two possible outcomes called success and failure. Each trial is performed in the same known conditions and we assume that there is no influence between different trials. Formally a *Bernoulli scheme* with parameter p, $0 < p < 1$, is a sequence $E_1, E_2, \ldots$ of stochastically independent equiprobable events with $\mathbf{P}(E_1) = p$.

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8_2

Example 2.2.1 A concrete example for which one can use as a model a Bernoulli scheme with $p = \frac{1}{2}$ is a sequence of throws of a symmetric coin, where E_i is the event that one gets head at the ith throw.

2.3 Binomial Distribution

Given a Bernoulli scheme $(E_i)_{i\in\mathbb{N}}$ with $\mathbf{P}(E_i) = p$, let S_n the random number of successes in the first n trials. S_n can be written as

$$S_n = E_1 + \cdots + E_n .$$

The set of possible values of S_n is $I(S_n) = \{0, \ldots, n\}$.

Let us compute, using the constituents of the events $E_1, \ldots, E_n$, the probability distribution of S_n:

$$\mathbf{P}(S_n = k) = \sum_{Q\subset(S_n=k)} \mathbf{P}(Q)\,.$$

We must determine the probability of a constituent of I type with respect to the event $(S_n = k)$. An example of such a constituent is

$$Q = E_1 \ldots E_k \tilde{E}_{k+1} \ldots \tilde{E}_n, \tag{2.1}$$

that is the event that k successes are obtained in the first k trials, whereas the remaining $n - k$ trials yield failures.

Analogously, any other constituent of I type will be a product of the same kind as in (2.1). Since the events are stochastically independent, in force of Proposition 1.11.1, every constituent Q of I type has the same probability, given by

$$\mathbf{P}(Q) = \underbrace{p \cdots p}_{k \text{ times}} \underbrace{(1 - p) \cdots (1 - p)}_{(n-k) \text{ times}} = p^k (1 - p)^{n-k}\,.$$

In order to compute $\mathbf{P}(S_n = k)$ we must therefore multiply this value times the number of constituents of I type. This is equal to $\binom{n}{k}$, that is the number of ways of choosing a subset of k elements out of n trials. Therefore we have

$$\mathbf{P}(S_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}\,.$$

S_n is said to have *binomial distribution* $\text{Bn}(n, p)$ with parameters n, p.

It is easy to check that $\sum_{k=0}^{n} \mathbf{P}(S_n = k) = 1$, as it must be since the events $(S_n = k)$, $k = 0, \ldots, n$, make up a partition. Indeed, using Newton's formula, we have:

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

The simplest way to compute the expectation of S_n is through the linearity of expectation:

$$\mathbf{P}(S_n) = \mathbf{P}(E_1 + \cdots + E_n) = \sum_{i=1}^{n} \mathbf{P}(E_i) = np .$$

Example 2.3.1 Consider an urn containing N identical balls, of which H are white and $N - H$ are black. We perform a sequence of n drawings with replacement. It is easy to check that by symmetry the sequence of events $(E_i)_{i=1,2,\ldots}$ where $E_i =$ (a white ball is drawn at the ith drawing) makes up a Bernoulli scheme, with parameters $p = \frac{H}{N}$. Indeed for $1 \leq i_1 < i_2 < \ldots < i_k$

$$\mathbf{P}(E_{i_1} \ldots E_{i_n}) = \frac{H^k}{N^k} = \left(\frac{H}{N}\right)^k ,$$

where the possible cases correspond to the N^k sequences of balls that may be drawn in the drawings $i_1, \ldots, i_k$, whereas the favorable cases correspond to the H^k sequences where white balls are drawn.

2.4 Geometric Distribution

Let $(E_i)_{i=1,2,\ldots}$ be a Bernoulli scheme; let T be the random number representing the number of the trial when the first success is obtained, i.e. $T = \min\{n \mid E_n = 1\}$. The set of possible values of T is given by:

$$I(T) = \mathbb{N} \setminus \{0\} \cup \{\infty\} .$$

It is easy to see that $\mathbf{P}(T = \infty) = 0$ since for all $n > 0$, $(T = \infty) \subseteq \tilde{E}_1 \ldots \tilde{E}_n$ so that $\mathbf{P}(T = \infty) \leq \mathbf{P}(\tilde{E}_1 \ldots \tilde{E}_n) = (1-p)^n$ for every n. Let us compute the probability distribution of T for finite values:

$$\mathbf{P}\,(T = i) = \mathbf{P}\left(\tilde{E}_1 \ldots \tilde{E}_{i-1} E_i\right) = \mathbf{P}\left(\tilde{E}_1\right) \ldots \mathbf{P}\left(\tilde{E}_{i-1}\right) \mathbf{P}\,(E_i) = (1-p)^{i-1} p .$$

T is said to have *geometric distribution* with parameter p. Using the formula for the sum of geometric series (see Appendix G.1), one verifies that

$$\sum_{i=1}^{+\infty} \mathbf{P}(T = i) = \sum_{i=1}^{+\infty} (1-p)^{i-1} p = p \sum_{k=0}^{+\infty} (1-p)^k = p \cdot \frac{1}{1 - (1-p)} = 1.$$

The expectation of T can be computed by an extension of formula (1.10) to the case of enumerable set of values. This can be justified (providing that the series converges) as a regularity property, thinking that T can be approximated with random numbers with a finite but arbitrarily large number of values. We then get

$$\mathbf{P}(T) = \sum_{i=1}^{+\infty} i\mathbf{P}(T = i) = \sum_{i=1}^{+\infty} i(1-p)^{i-1}p = p\sum_{i=1}^{+\infty} i(1-p)^{i-1} = \frac{p}{p^2} = \frac{1}{p}\,,$$

where we used that for $|x| < 1$

$$\sum_{i=1}^{+\infty} ix^{i-1} = \sum_{i=1}^{+\infty} \frac{d}{dx}[x^i] = \frac{d}{dx}\left(\sum_{i=0}^{+\infty} x^i\right) = \frac{d}{dx}\left(\frac{1}{1-x}\right) = \frac{1}{(1-x)^2}\,.$$

The geometric distribution is said to be "*memoryless*". Indeed for $m > 0, n > 0$

$$\mathbf{P}(T > m + n \mid T > n) = \mathbf{P}(T > m),$$

i.e. the conditional probability of no success up and including the $(m+n)$th trial given that there was no success up and including the nth trial is equal to the probability of no success up and including the m trial: everything starts from scratch. We have namely that

$$\mathbf{P}(T > m + n \mid T > n) = \frac{\mathbf{P}(T > m + n, T > n)}{\mathbf{P}(T > n)} = \frac{\mathbf{P}(T > m + n)}{\mathbf{P}(T > n)}\,.$$

But $\mathbf{P}(T > n) = (1-p)^n$ since $(T > n) = \tilde{E}_1 - \tilde{E}_n$. Hence

$$\mathbf{P}(T > m + n \mid T > n) = \frac{(1-p)^{m+n}}{(1-p)^n} = (1-p)^m = \mathbf{P}(T > m)\,.$$

2.5 Poisson Distribution

A random number X is said to have *Poisson distribution* with parameters λ, $\lambda \in \mathbb{R}_+$, if $I(X) = \mathbb{N}$ and

$$\mathbf{P}(X = i) = \frac{\lambda^i}{i!}e^{-\lambda}\,.$$

As in the case of a geometric distribution $\sum_{i=0}^{+\infty} \mathbf{P}(X = i) = 1$. Indeed

$$\sum_{i=0}^{+\infty} \mathbf{P}(X = i) = \sum_{i=0}^{+\infty} \frac{\lambda^i}{i!}e^{-\lambda} = e^{-\lambda}\sum_{i=0}^{+\infty} \frac{\lambda^i}{i!} = e^{-\lambda}e^{\lambda} = 1\,.$$

In order to compute the expectation, we use the extension of the formula for random numbers with a finite number of possible values to the case of a enumerable set of possible values as we did for geometric distribution and as we will do in similar cases (provided that the series is convergent). We obtain

$$\begin{aligned}\mathbf{P}(X) &= \sum_{i=0}^{+\infty} i\mathbf{P}(X=i) = \sum_{i=0}^{+\infty} i\frac{\lambda^i}{i!}e^{-\lambda} = \lambda e^{-\lambda}\sum_{i=1}^{+\infty}\frac{\lambda^{i-1}}{(i-1)!}\\ &= \lambda e^{-\lambda}\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!} = \lambda e^{-\lambda}e^{\lambda} = \lambda\,.\end{aligned}$$

2.6 Hypergeometric Distribution

Consider an urn containing N balls of which H are white and $N-H$ black, where $0 < H < N$. We perform n drawings without replacement from the urn with $n \leq N$. Let X be the random number that counts the number of white balls in the sample that we draw.

Since we perform drawings without replacement, X is less than or equal to H and $n-X$, the number of black balls in the sample, is less than or equal to $N-H$. From this it follows that the set of possible values of X is given by

$$I(X) = \{0 \vee n-(N-H), \dots, n \wedge H\}\ .$$

Let $i \in I(X)$. Due to the symmetry of the situation with respect to interchange of balls, we evaluate $\mathbf{P}(X=i)$ using formula (1.12). When defining possible cases and consequently favorable cases, we can consider the set of the n drawn balls, i.e. we can avoid to consider the order of drawings, as the event does not involve the order. In this way the possible cases correspond to the subset of size n from a set of N elements:

$$\sharp\ possible\ cases = \binom{N}{n}\ .$$

A sample with i white balls contains $(n-i)$ black balls. The number of favorable cases that correspond to such samples is therefore given by:

$$\sharp\ favorable\ cases = \binom{H}{i}\binom{N-H}{n-i}\ .$$

The random number X is said to have *hypergeometric distribution* with parameters n, H, N. By the former discussion we have:

$$\mathbf{P}(X = i) = \frac{\binom{H}{i}\binom{N-H}{n-i}}{\binom{N}{n}} .$$

In order to compute the expectation of X, it is convenient to decompose it as

$$X = \sum_{i=1}^{n} E_i$$

where E_i is the event that a white ball is chosen at ith drawing. Therefore by the linearity of the expectation

$$\mathbf{P}(X) = \mathbf{P}(E_1) + \cdots + \mathbf{P}(E_n).$$

In the evaluation of $\mathbf{P}(E_i)$ we can still use symmetry by interchange of balls, but when defining possible cases we must take into account the order, since the event depends on the order of the drawings. Possible cases correspond to sequences of length n of distinct elements from a set of N elements. Their number is $D_n^N = (N)_n = N(N-1) - (N-n+1)$. Favorable cases correspond to those sequences that have a white ball at the ith place. This ball can be chosen in H ways. The remaining balls form a sequence of lenght $n-1$ of distinct elements from a set of $N-1$ elements. Therefore

$$\mathbf{P}(E_i) = \frac{\sharp\, favorable\ cases}{\sharp\, possible\ cases} = \frac{H D_{n-1}^{N-1}}{D_n^N} = \frac{H}{N}$$

and

$$\mathbf{P}(X) = n\frac{H}{N} .$$

2.7 Independence of Partitions

Two partitions $\mathcal{H} = (H_1, \ldots, H_m)$, $\mathcal{L} = (L_1, \ldots, L_n)$ are said to be *stochastically independent* if for every i, j with $1 \le i \le m$, $1 \le j \le n$

$$\mathbf{P}\left(H_i L_j\right) = \mathbf{P}\left(H_i\right)\mathbf{P}\left(L_j\right) .$$

Stochastic independence can be extended to the case of r partitions $\mathcal{H}_1, \ldots, \mathcal{H}_r$. Consider the partitions

$$\mathcal{H}_l = (H_1^{(l)}, \ldots, H_{n_l}^{(l)})$$

for $1 \le l \le r$. $\mathcal{H}_1, \ldots, \mathcal{H}_r$ are said to be *stochastically independent* if for every $i_1, \ldots, i_r$ with $1 \le i_l \le n_l, \ldots, 1 \le l \le r$

$$\mathbf{P}\left(H_{i_1}^{(1)} \dots H_{i_r}^{(r)}\right) = \mathbf{P}\left(H_{i_1}^{(1)}\right) \dots \mathbf{P}\left(H_{i_r}^{(r)}\right) .$$

Partitions can be thought as *pluri-events*, with a certain number of possible results, such as in the case of drawings from an urn containing balls of several colors. In the case of partitions with two events, one can select an event from each partition. In this case stochastic independence of partitions is equivalent to stochastic independence of the selected events.

2.8 Generalized Bernoulli Scheme

Let $\mathcal{H}_1, \mathcal{H}_2, \dots$ be a sequence of partitions, each composed by r events, $\mathcal{H}_i = \left\{E_1^{(i)}, \dots, E_r^{(i)}\right\}$ for $i \geq 1$. We assume that $\mathcal{H}_1, \dots, \mathcal{H}_n$ are stochastically independent for every n and that $\mathbf{P}(E_k^{(i)}) = p_k$, $k = 1, \dots, r$, for all $i \geq 1$, with $p_1 + \cdots + p_r = 1$. The sequence $\mathcal{H}_1, \mathcal{H}_2, \dots$ is called a *generalized Bernoulli scheme*. In the case $r = 2$ a generalized Bernoulli scheme is equivalent to the ordinary Bernoulli scheme $(F_1, F_2, \dots)$ where $F_i = E_1^{(i)}$ with parameter $p = p_1$. We can represent a generalized Bernoulli scheme in an array:

$$\begin{array}{c}
E_1^{(1)}, \dots, E_r^{(1)} \\
E_1^{(2)}, \dots, E_r^{(2)} \\
\vdots \;, \dots, \; \vdots \\
\vdots \;, \dots, \; \vdots \\
E_1^{(n)}, \dots, E_r^{(n)},
\end{array}$$

where the events belonging to the same column are equiprobable, whereas the events of each row constitute stochastically independent partitions.

2.9 Multinomial Distribution

Starting from a generalized Bernoulli scheme, as defined in Sect. 2.2, we can now define the *multinomial distribution* in the same way as the binomial distribution can be defined starting from an ordinary Bernoulli scheme. Given $n > 0$, let us consider the random numbers $Y_1, \dots, Y_r$ defined by

$$Y_l = \sum_{i=1}^{n} E_l^{(i)} \, , \; l = 1, \dots, r \, .$$

In the array of the previous section, the Y_l's are obtained by adding up the events along the columns. We have

$$\sum_{l=1}^{r} Y_l = \sum_{l=1}^{r}\sum_{i=1}^{n} E_l^{(i)} = \sum_{i=1}^{n}\underbrace{\sum_{l=1}^{r} E_l^{(i)}}_{1} = n\,.$$

The idea of constituents can be extended in a natural fashion from events to partitions. A *constituent* of the partition $\mathcal{H}_1, \ldots, \mathcal{H}_n$ is an event of the form

$$Q = \Pi_{i=1}^{n} H_*^i\,,$$

where H_*^i is an event of the partition $\mathcal{H}_i$. If $\mathcal{H}_1, \ldots, \mathcal{H}_n$ are stochastically independent (as in the case of generalized Bernoulli scheme) we have:

$$\mathbf{P}(Q) = \mathbf{P}(H_*^1)\ldots\mathbf{P}(H_*^n)\,.$$

We want to compute

$$\mathbf{P}\,(Y_1 = k_1, \ldots, Y_r = k_r)$$

for $k_1 \geq 0, k_r \geq 0$ such that $k_1 + \cdots + k_r = n$. We can decompose this probability in terms of constituents of I type:

$$\mathbf{P}(Y_1 = k_1, \ldots, Y_r = k_r) = \sum_{Q} \mathbf{P}(Q),$$

where Q varies among the constituents of I type contained in the event $(Y_1 = k_1, \ldots, Y_r = k_r)$. In the product defining a constituent of I type there will be k_l events of index l with $1 \leq l \leq r$. Therefore since the partitions are stochastically independent, the probability of a constituent of I type is given by:

$$\mathbf{P}(Q) = p_1^{k_1}, \ldots, p_r^{k_r}\,.$$

The number of constituents of I type is equal to the way of partitioning a set of n elements into r subsets with $k_1, \ldots, k_r$ elements, i.e. $\dfrac{n!}{k_1!\ldots k_r!}$. We have therefore:

$$\mathbf{P}\,(Y_1 = k_1, \ldots, Y_r = k_r) = \sum_{Q \text{ I type}} \mathbf{P}(Q) = \underbrace{\frac{n!}{k_1!\ldots k_r!}}_{\text{number of constituents}} \quad \underbrace{p_1^{k_1}\ldots p_r^{k_r}}_{\mathbf{P}(Q)}\,.$$

The multinomial distribution depends on the parameters $r, p_1, \ldots, p_{r-1}$, since $p_r = 1 - \sum_{i=1}^{r-1} p_i$. For $r = 2$ the multinomial distribution reduces to the binomial one.

2.10 Stochastic Independence for Random Numbers with Discrete distribution

Let X and Y be two random numbers with $I(X) = \{x_1, \dots, x_m\}$ and $I(Y) = \{y_1, \dots, y_n\}$. We consider the partitions $\mathcal{H}$ and $\mathcal{K}$ generated by the events $H_i = (X = x_i)$, for $i = 1, \dots, m$, and $K_j = (Y = y_j)$, for $j = 1, \dots, n$.

The random numbers X and Y are said to be *stochastically independent* if the partitions $\mathcal{H}$ and $\mathcal{K}$ are stochastically independent.

2.11 Joint Distribution

Let us consider two random numbers X and Y, that we can look at as a *random vector* (X, Y), assuming a finite number of possible values $I(X, Y)$. If $I(X) = \{x_1, \dots, x_m\}$ and $I(Y) = \{y_1, \dots, y_n\}$ we define the *joint distribution* of X and Y. This is the function

$$p(x_i, y_j) = \mathbf{P}(X = x_i, Y = y_j)$$

defined on $I(X) \times I(Y)$. We can associate to it the matrix

$$\begin{pmatrix} p(x_1, y_1) & \dots & p(x_1, y_n) \\ \vdots & \ddots & \vdots \\ p(x_m, y_1) & \dots & p(x_m, y_n) \end{pmatrix}.$$

The *marginal distribution* of X is the function

$$p_1(x_i) = \mathbf{P}(X = x_i)$$

for $i = 1, \dots, m$. The marginal distribution can be obtained from the joint distribution:

$$p_1(x_i) = \mathbf{P}(X = x_i) = \sum_{j=1}^{n} \mathbf{P}(X_i, Y_j) = \sum_{j=1}^{n} p(x_i, y_j),$$

i.e. adding up the elements on the rows of the matrix. It is called marginal because it is customarily written at the margin of the matrix. Similarly the *marginal distribution* of Y is defined by:

$$p_2(y_j) = \mathbf{P}(Y = y_j) = \sum_{i=1}^{m} p(x_i, y_j).$$

It follows that two random numbers X and Y are stochastically independent if and only if

$$p(x_i, y_j) = p_1(x_i)p_2(y_j) \tag{2.2}$$

for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. Given $\psi : \mathbb{R}^2 \longrightarrow \mathbb{R}$, the expectation of the random number $Z = \psi(X, Y)$ can be obtained from the joint distribution of X, Y:

$$\mathbf{P}(Z) = \mathbf{P}(\psi(X, Y)) = \sum_{i=1}^{m}\sum_{j=1}^{n} \psi(x_i, y_j)p(x_i, y_j)\,. \tag{2.3}$$

The proof is completely analogous to that one in the case of a single random number. For example, we can compute $\mathbf{P}(XY)$:

$$\mathbf{P}(XY) = \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j p(X = x_i, Y = y_j).$$

If X and Y are stochastically independent and ϕ_1, ϕ_2 are two real functions $\phi_i : \mathbb{R} \longrightarrow \mathbb{R}$ with $i = 1, 2$, we have that

$$\mathbf{P}(\phi_1(X)\phi_2(Y)) = \mathbf{P}(\phi_1(X))\mathbf{P}(\phi_2(Y)). \tag{2.4}$$

Indeed

$$\begin{aligned}
\mathbf{P}(\phi_1(X)\phi_2(Y)) &= \sum_{i=1}^{m}\sum_{j=1}^{n} \phi_1(x_i)\phi_2(y_j)\mathbf{P}(X = x_i, Y = y_j) \\
&= \sum_{(x_i, y_j)\in I(X)\times I(Y)} \phi_1(x_i)\phi_2(y_j)p_1(x_i)p_2(y_j) \\
&= \sum_{x_i \in I(X)} \phi_1(x_i)p_1(x_i) \sum_{y_j \in I(Y)} \phi_2(y_j)p_2(y_j) \\
&= \mathbf{P}(\phi_1(X))\mathbf{P}(\phi_2(Y))\,.
\end{aligned}$$

2.12 Variance of Discrete Distributions

We compute the variances of the distributions that we have previously introduced.

1. *Variance of an event*:

$$\boldsymbol{\sigma}^2(E) = \mathbf{P}\left(E^2\right) - \mathbf{P}\left(E\right)^2 = \mathbf{P}(E) - \mathbf{P}(E)^2 = \mathbf{P}(E)(1 - \mathbf{P}(E))\,,$$

where we use that for an event $E^2 = E$ since E can take only values 0 and 1.

2. *Binomial distribution*: For X with binomial distribution with parameters n_x and p, we use the representation $X = E_1 + \ldots + E_n$, where the E_i's are stochastically independent and hence pairwise uncorrelated. We get:

$$\sigma^2(E_1 + \ldots + E_n) = \sum_{i=1}^{n} \sigma^2(E_i) = np(1-p)\,.$$

3. *Geometric distribution*: we need to compute $\mathbf{P}(X^2)$ as we have already computed:

$$\mathbf{P}(X) = \sum_{i=1}^{+\infty} ip(1-p)^{i-1} = \frac{1}{p}.$$

Hence

$$\begin{aligned}
\mathbf{P}(X^2) = p\sum_{i=1}^{+\infty} i^2(1-p)^{i-1} &= \left(p\sum_{i=1}^{+\infty} i(i-1)(1-p)^{i-1}\right) + p\sum_{i=1}^{+\infty} i(1-p)^{i-1} \\
&= p(1-p)\sum_{i=2}^{+\infty} i(i-1)(1-p)^{i-2} + \frac{1}{p} \\
&= p(1-p)\frac{d^2}{d^2p}\left(\sum_{i=2}^{+\infty}(1-p)^i\right) + \frac{1}{p} \\
&= p(1-p)\left(\frac{d^2}{d^2p}\right)\left(\frac{1}{1-(1-p)} - 1 - (1-p)\right) + \frac{1}{p} \\
&= \frac{2(1-p)}{p^2} + \frac{1}{p} \\
&= \frac{2}{p^2} - \frac{1}{p}\,.
\end{aligned}$$

Therefore the variance of the geometric distribution is given by

$$\sigma^2(X) = \mathbf{P}[X^2] - \mathbf{P}(X)^2 = \frac{(1-p)}{p^2}\,.$$

4. *Poisson distribution*: if X has Poisson distribution with parameter λ, we have:

$$\begin{aligned}
P(X^2) &= \sum_{i=0}^{+\infty} i^2\mathbf{P}(X=i) = \sum_{i=0}^{+\infty} i^2\frac{\lambda^i}{i!}e^{-\lambda} = e^{-\lambda}\sum_{i=0}^{+\infty} i(i-1)\frac{\lambda^i}{i!} + \lambda e^{-\lambda}\sum_{i=0}^{+\infty}\frac{\lambda^i}{i!} \\
&= \lambda^2 e^{-\lambda}\sum_{i=2}^{+\infty}\frac{\lambda^{i-2}}{(i-2)!} + \lambda = \lambda^2 e^{-\lambda}\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!} + \lambda = \lambda^2 + \lambda
\end{aligned}$$

where we have used the computation of the expectation of the Poisson distribution.

We have then

$$\boldsymbol{\sigma}^2(X) = \mathbf{P}(X^2) - \mathbf{P}(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda\,.$$

5. *Hypergeometric distribution*: with the notation of Sect. 2.6, we use the representation $X = E_1 + \cdots + E_n$. The events E_i's in this case are not stochastically independent and are actually pairwise negatively correlated. Indeed, for $0 < H < N$ for every pair i, j with $i \neq j$, we have:

$$\mathbf{cov}(E_i, E_j) = \mathbf{P}(E_i E_j) - \mathbf{P}(E_i)\mathbf{P}(E_j) = \frac{H}{N^2}\frac{H-N}{N-1} < 0$$

as

$$\mathbf{P}(E_i E_j) = \frac{H(H-1)D_{n-2}^{N-2}}{D_n^N} = \frac{H(H-1)}{N(N-1)}\frac{D_{n-2}^{N-2}}{D_{n-2}^{N-2}} = \frac{H(H-1)}{N(N-1)}.$$

Here we have used formula (1.12); possible cases are sequences with no repetition of length n from a set of N elements, whereas in counting favorable cases we first select two different white balls for the ith and the jth drawings and then the remaining $n-2$ balls from a set of $N-2$ elements.
The variance of X is then obtained by means of the formula for the variance of the sum of n random numbers:

$$\begin{aligned}\boldsymbol{\sigma}^2(X) &= \sum_{i=1}^{n} \boldsymbol{\sigma}^2(E_i) + \sum_{\substack{i,j \\ i\neq j}} \mathbf{cov}(E_i, E_j) \\ &= n\frac{H}{N}\left(1 - \frac{H}{N}\right) + n(n-1)\frac{H}{N^2}\frac{H-N}{N-1} = n\frac{N-n}{N-1}\frac{H}{N}\left(1-\frac{H}{N}\right),\end{aligned}$$

where $n(n-1)$ is the number of ordered pairs i, j, with $i \neq j$, which can be chosen out of $\{1, \ldots, n\}$.

2.13 Non-correlation and Stochastic Independence

Let us consider two random numbers X and Y with discrete joint distribution given by:

$$p(X_i, Y_j) = \mathbf{P}(X = x_i, Y = y_j) = p_{i,j}$$

and marginal distributions given by:

$$p_1(x_i) = \mathbf{P}(X = x_i) = p_i \qquad i = 1, \ldots, m\,,$$

$$p_2(y_j) = \mathbf{P}(Y = y_j) = q_j \qquad j = 1, \ldots, n\,.$$

X and Y are non-correlated if

$$\mathbf{P}(XY) = \mathbf{P}(X)\mathbf{P}(Y)$$

i.e. if

$$\sum_i \sum_j x_i y_j p_{i,j} = \sum_i x_i p_i \sum_j y_j q_j \, .$$

Moreover, the following relations are satisfied:

$$\begin{array}{lll} \sum_i p_i = 1 & \text{and } \sum_j p_{i,j} = p_i & \text{for } i = 1, \ldots, m \, , \\ \sum_j q_j = 1 & \text{and } \sum_i p_{i,j} = q_j & \text{for } j = 1, \ldots, n \, , \end{array}$$

$$\sum_i \sum_j p_{i,j} = 1 \, .$$

Assume that we want to find values $p_{i,j}$ of the joint distribution, such that X and Y are non-correlated and have two fixed marginal distributions $\{p_i\}_{i=1,\ldots,m}$ and $\left\{q_j\right\}_{j=1,\ldots,n}$. We observe first of all that $p_{i,j}$ must satisfy the relation $\sum_{i,j} p_{i,j} = 1$. In order to determine the marginal distributions we must verify other additional $(m-1)+(n-1)$ linear relations. We have $(m-1)+(n-1)$ and not $m+n$, since once $(m-1)+(n-1)$ relations are satisfied, the last two follow from the fact that $\sum_{i,j} p_{i,j} = 1$, $\sum_i p_i = 1$, $\sum_j q_j = 1$. Finally in order to impose non-correlation, an extra linear relation must be verified on the $p_{i,j}$'s:

$$\sum_i \sum_j p_{i,j} x_i y_j = m_1 m_2,$$

where $m_1 = \sum_i^m x_i p_i$ and $m_2 = \sum_j^n y_j q_j$. We have therefore a system of $1 + (m-1) + (n-1) + 1 = m + n$ linear equations for mn unknowns. This system has the solution $p_{i,j} = p_i q_j$, for which X and Y are stochastically independent. This will be the only solution if the number of linearly independent equations is equal to the number of the unknowns, i.e. if $m + n = mn$, or $mn - m - n = (m-1)(n-1) - 1 = 0$. This happens only if $m = n = 2$. It follows that non-correlation does not imply in general stochastic independence. If $m = n = 2$, then there is just one solution so that non-correlation and stochastic independence coincide. This is the case of events: *two events are non-correlated if and only if they are stochastically independent*.

In Sect. 2.11 we have shown that stochastic independence implies non-correlation and that in fact it implies non-correlation of any two functions of the random numbers.

2.14 Generating Function

Let X be a random number with discrete distribution on a subset of $\mathbb{N}$. The *generating function* of X is defined for $u \in \mathbb{C}$, $|u| \leq 1$, by

$$\phi_X(u) := \mathbf{P}(u^X) = \sum_{k \in I(X)} u^k \mathbf{P}(X = k). \tag{2.5}$$

The expectation of a complex random variable is defined as the expectation of the real part plus i times the expectation of the imaginary part. The condition $|u| \leq 1$ guarantees that the series (2.5) is convergent in the case of infinitely many possible values. We will use characteristic functions just for real values of u. We have that

$$\phi_X(0) = \mathbf{P}(X = 0).$$

In general, computing the nth derivative of (2.5) in $u = 0$, we obtain

$$\mathbf{P}(X = n) = \frac{1}{n!} \frac{d^n \phi_X(u)}{dx^n}\Big|_{u=0},$$

for every $n \in \mathbb{N}$. This shows that the probability distribution of X can be obtained from its generating function.

Proposition 2.14.1 *If* $\mathbf{P}(X) = \sum_{k \in I(X)} k\mathbf{P}(X = k) < \infty$, *then* $\mathbf{P}(X) = \lim_{u \to 1^-} \phi'_X(u)$. *Moreover* $\mathbf{P}(X) = \sum_{k \in I(X)} k\mathbf{P}(X = k) = +\infty$ *if and only if* $\lim_{u \to 1} \phi'_X(u) = \infty$.

This is a particular case of the following result.

Proposition 2.14.2 *If* $\mathbf{P}(X(X-1)\dots(X-k+1)) = \sum_{k \in I(X)} (k(k-1)\dots(k-n+1))\mathbf{P}(X = k) < \infty$, *then*

$$\mathbf{P}(X(X-1)\dots(X-k+1)) = \lim_{u \to 1^-} \phi_X^{(n)}(u).$$

Furthermore $\sum_{k \in I(X)} (k(k-1)\dots(k-n+1))\mathbf{P}(X = k) = \infty$ *if and only if* $\lim_{u \to 1^-} \phi_X^{(n)}(u) = \infty$.

Previous results are easily obtained by taking the derivatives of the generating function. In particular the variance of X can be obtained from the generating function:

$$\sigma^2(X) = \mathbf{P}(X^2) - \mathbf{P}(X)^2 = \lim_{u \to 1^-} \left(\phi''_X(u) + \phi'_X(u) - \left(\phi'_X(u)\right)^2 \right),$$

where ϕ'_X and ϕ''_X denote respectively the first and the second derivatives of ϕ_X. Generating functions of some common discrete distributions are easily obtained:

1. Event E with probability p

$$\phi_E(u) = up + (1 - p).$$

2. Binomial distribution $\mathrm{Bn}(n, p)$ with parameters n, p:

$$\begin{aligned}\phi_X(u) &= \sum_{k=0}^{n} u^k \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} (up)^k (1-p)^{n-k} = (up + (1-p))^n ,\end{aligned}$$

where Newton's binomial formula has been used.
3. Geometric distribution with parameter p:

$$\begin{aligned}\phi_X(u) &= \sum_{k=1}^{\infty} u^k p (1-p)^{k-1} \\ &= up \sum_{k=1}^{\infty} [u(1-p)]^{k-1} = \frac{up}{(1-u(1-p))},\end{aligned}$$

where the formula for the sum of geometric series has been used.
4. Poisson distribution with parameter λ:

$$\begin{aligned}\phi_X(u) &= \sum_{k=0}^{\infty} u^k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(u\lambda)^k}{k!} = e^{-\lambda(1-u)}.\end{aligned}$$

If X and Y are two stochastically independent random numbers with values in $\mathbb{N}$, i.e. $\mathbf{P}(X = i, Y = j) = \mathbf{P}(X = i)\mathbf{P}(Y = j)$ for all $i, j \in I(X) \times I(Y)$, then it is easy to show that

$$\phi_{X+Y}(u) = \phi_X(u)\phi_Y(u).$$

Indeed:

$$\begin{aligned}\phi_{X+Y}(u) &= \mathbf{P}(u^{X+Y}) = \mathbf{P}(u^X u^Y) \\ &= \sum_i \sum_j u^i u^j \mathbf{P}(X = i, Y = j) \\ &= \sum_i \sum_j u^i u^j \mathbf{P}(X = i)\mathbf{P}(Y = j) \\ &= \left(\sum_i u^i \mathbf{P}(X = i)\right)\left(\sum_j u^j \mathbf{P}(Y = j)\right) \\ &= \phi_X(u)\phi_Y(u) .\end{aligned}$$

Of course if we have n stochastically independent random numbers we obtain similarly: $\phi_{X_1+\cdots+X_n}(u) = \phi_{X_1}(u)\ldots\phi_{X_n}(u)$. One can also consider the case of the sum of a random number N of stochastically independent random numbers. Let $X_1, X_2, \ldots$ be an infinite sequence of stochastically independent random numbers with values in $\mathbb{N}$. This means that if we take any finite number of them, they are stochastically independent. We assume that $X_1, X_2, \ldots$ are identically distributed. Let N be a random number with values in $\mathbb{N}$, such that

$$N, X_1, X_2, \ldots$$

are stochastically independent. Let S_N be defined by

$$S_N = X_1 + \cdots + X_N.$$

We now compute the generating function of S_N:

$$\begin{aligned}
\phi_{S_N}(u) = \mathbf{P}(u^{S_N}) &= \sum_{k\in I(N)} \mathbf{P}(u^{S_N}|N=k)\mathbf{P}(N=k) \\
&= \sum_{k\in I(N)} \mathbf{P}(u^{S_k})\mathbf{P}(N=k) \\
&= \sum_{k\in I(N)} \mathbf{P}(N=k)\mathbf{P}(u^{X_1+\cdots+X_k}) \\
&= \sum_{k\in I(N)} \mathbf{P}(N=k)\phi_{X_1}(u)\ldots\phi_{X_k}(u) \\
&= \sum_{k\in I(N)} \mathbf{P}(N=k)\phi_{X_1}(u)^k \\
&= \phi_N\left(\phi_{X_1}(u)\right),
\end{aligned}$$

where ϕ_N is the generating function of X and we have used the fact that the random numbers X_i have the same distribution and hence the same generating function. See e.g. [3] or [6] for a more complete treatment of generating functions.

Chapter 3
One-Dimensional Absolutely Continuous Distributions

3.1 Introduction

For random numbers with discrete distribution, the distribution is completely specified by the probabilities of taking single values. If we want to introduce random numbers that take values on intervals or on the whole line, then the specification of the probabilities of taking single values is no longer sufficient to determine their distributions. For example for a random number corresponding to a random choice in an interval $[a, b]$, the probabilities of taking single values must be clearly equal to 0, but that in no way specifies the probability of taking value in a subinterval of $[a, b]$. In the following we will see how it is possible to describe the distribution of a random number in general.

3.2 Cumulative Distribution Function

Given a random number X, its *cumulative distribution function* (c.d.f) is defined by:

$$F(x) = \mathbf{P}(X \leq x), \text{ for } x \in \mathbb{R}.$$

The cumulative distribution function $F(x)$ verifies the following properties:

1. $0 \leq F(x) \leq 1$ since it is the probability of an event.
2. It is *non-decreasing*: for $a < b$ we have $F(b) - F(a) = \mathbf{P}(a < X \leq b) \geq 0$, so that $F(a) \leq F(b)$.

We introduce now some further properties that are usually assumed to be verified by cumulative distribution function. They can be thought of as regularity properties, as they state that the probability of an event E is equal to the limit of the sequence $P(E_n)$, where E_n is a monotonic sequence converging to E. In particular:

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8_3

1. *continuity from the right*: $F(x) = \lim_{y \to x^+} F(y)$;
2. *limit from the left*: $\lim_{y \to x^-} F(y) = \mathbf{P}(X < x)$;
3. $\lim_{x \to +\infty} F(x) = 1$;
4. $\lim_{x \to -\infty} F(x) = 0$.

In all examples of p.d.f.'s these extra properties will be satisfied, even if it is possible to consider cases where they do not hold true. It follows from 1 and 2 that

$$\mathbf{P}(X = x_0) = \mathbf{P}((X \leq x_0) - (X < x_0)) = F(x_0) - F(x_0^-)$$

where $F(x_0^-)$ denotes $\lim_{x \to x_0^-} F(x)$. This limit always exists as $F(x)$ is bounded non-decreasing.

Example 3.2.1 (Discrete case) In the case of a random number X with discrete distribution $I(X) = \{x_1, x_2, \ldots\}$ one has:

$$F(x) = \mathbf{P}(X \leq x) = \sum_{x_i \leq x} \mathbf{P}(X = x_i).$$

The probability that a random number X takes value in an interval $(a, b]$ can be obtained from its c.d.f. F by:

$$\begin{aligned}\mathbf{P}(a < X \leq b) &= \mathbf{P}((X \leq b) - (X \leq a)) \\ &= \mathbf{P}(X \leq b) - \mathbf{P}(X \leq a) \\ &= F(b) - F(a).\end{aligned}$$

3.3 Absolutely Continuous Distributions

Let X be a random number. We say that X has *absolutely continuous distribution* if there exists a function $f : \mathbb{R} \to \mathbb{R}^+$ with such that the c.d.f. $F(x)$ of X can be written as:

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

The function f is the called a *probability density function* (p.d.f.) of X. Note that f is not unique. Indeed if the values of f are changed on a finite set of points, the new function is still a density of X, as its integrals are the same. It follows from fundamental theorem of calculus that if x is a continuity point of f, then

$$f(x) = F'(x).$$

Since $\lim_{x\to\infty} F(x) = 1$, then if f is a p.d.f. of X we have

$$\int_{-\infty}^{\infty} f(x)dx = \lim_{x\to\infty} \int_{-\infty}^{x} f(y)dy = \lim_{x\to\infty} F(x) = 1.$$

If x is a continuity point of f, then $f(x) \geq 0$. Indeed assume that $f(x) < 0$, then by continuity there would be a neighborhood (a, b) of x where f is still strictly negative but then

$$F(b) = F(a) + \int_{a}^{b} f(x)dx < F(a),$$

so that F would not be non-decreasing. We have that for $a < b$

$$\mathbf{P}(a < X \leq b) = F(b) - F(a) = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = \int_{a}^{b} f(x)dx.$$

Let us now see how to compute the expectation of X from the p.d.f. f. We consider the particular case when $I(X)$ is contained in some interval $[a, b]$ and the p.d.f. f is continuous (and zero outside $[a, b]$). We subdivide $[a, b]$ into n intervals I_i, $i = 1, \ldots, n$ of length $\frac{b-a}{n}$. It is not important that the extremes are included: we assume that the intervals are closed on the r.h.s. and open on the l.h.s. except for I_1 that is closed on both sides. We define two random numbers with discrete distribution $X_-^{(n)}$ and $X_+^{(n)}$: if X takes value in I_i, then $X_-^{(n)}$ is equal to the left endpoint of I_i, $X_+^{(n)}$ is equal to the right endpoint. Since $X_-^{(n)}$ and $X_+^{(n)}$ have discrete distribution with a finite number of possible values, we can compute their expectations using the formula (1.10). They are given by:

$$\mathbf{P}(X_-^{(n)}) = \sum_{j=0}^{n-1} \left(a + j\frac{b-a}{n}\right) \int_{a+j\frac{b-a}{n}}^{a+(j+1)\frac{b-a}{n}} f(x)dx;$$

$$\mathbf{P}(X_+^{(n)}) = \sum_{j=0}^{n-1} \left(a + (j+1)\frac{b-a}{n}\right) \int_{a+j\frac{b-a}{n}}^{a+(j+1)\frac{b-a}{n}} f(x)dx.$$

Since $X_-^{(n)} \leq X \leq X_+^{(n)}$, then

$$\mathbf{P}(X_-^{(n)}) \leq \mathbf{P}(X) \leq \mathbf{P}(X_+^{(n)}).$$

It is easy to see, using the continuity of $f(x)$, that as $n \to \infty$ both $P(X_-^{(n)})$ and $P(X_+^{(n)})$ converge to

$$\int_{a}^{b} xf(x)\mathrm{d}x = \int_{\mathbb{R}} xf(x)\mathrm{d}x,$$

that is hence the value of $\mathbf{P}(X)$. Approximation arguments lead to extend this formula to the case of a general X with absolutely continuous distribution with probability density $f(x)$ provided that

$$\int_{\mathbb{R}} |x| f(x)\mathrm{d}x < \infty, \tag{3.1}$$

i.e. one assume that, when (3.1) holds true, the expectation of X in the absolutely continuous case is given by:

$$\mathbf{P}(X) = \int_{-\infty}^{+\infty} x f(x)\mathrm{d}x.$$

Analogously if $\psi : \mathbb{R} \to \mathbb{R}$ is a real function such that $\psi(x) f(x)$ is integrable, we are lead to assign to $\mathbf{P}(\psi(X))$ the value

$$\mathbf{P}(\psi(X)) = \int_{-\infty}^{+\infty} \psi(x) f(x)\mathrm{d}x. \tag{3.2}$$

It follows that the variance can be obtained by:

$$\begin{aligned}\boldsymbol{\sigma}^2(X) &= \mathbf{P}(X^2) - \mathbf{P}(X)^2 \\ &= \int_{-\infty}^{+\infty} x^2 f(x)\,\mathrm{d}x - \left(\int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x\right)^2,\end{aligned}$$

provided that the integrals exist. In the following sections we shall introduce some of the most common one-dimensional absolutely continuous distributions.

3.4 Uniform Distribution in [0, 1]

A random number X has *uniform distribution* in [0, 1] if its c.d.f. is given by:

$$F(x) = \begin{cases} 0 & x \le 0, \\ x & 0 < x < 1, \\ 1 & x \ge 1. \end{cases}$$

It is a continuous distribution since

$$\mathbf{P}(X = x) = F(x) - F(x^-) = 0$$

for every $x \in \mathbb{R}$. Indeed it is easy to check that it is an absolutely continuous distribution with p.d.f. $f(x)$ given by:

$$f(x) = \begin{cases} 0 & x \leq 0, \\ 1 & 0 < x < 1, \\ 0 & x \geq 1. \end{cases}$$

As in the following examples the values of the p.d.f. in discontinuity points can be chosen in an arbitrary way. The expectation is given by

$$\mathbf{P}(X) = \int_{\mathbb{R}} x f(x)\,\mathrm{d}x = \int_0^1 x\,\mathrm{d}x = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2},$$

and the variance by

$$\boldsymbol{\sigma}^2(X) = \int_0^1 x^2\,\mathrm{d}x - \frac{1}{4} = \left[\frac{x^3}{3}\right]_0^1 - \frac{1}{4} = \frac{1}{12}.$$

3.5 Uniform Distribution on an Arbitrary Interval $[a, b]$

A random number X has *uniform distribution* in $[a, b]$ if its c.d.f. is given by:

$$F(x) = \begin{cases} 0 & x \leq a, \\ c(x-a) & a < x < b, \\ 1 & x \geq 1. \end{cases}$$

In order to compute the constant c, we impose the continuity in the point $x = b$ and get $c(b-a) = 1$, that is:

$$c = \frac{1}{b-a}.$$

The expectation is given by:

$$\mathbf{P}(X) = \int_a^b \frac{x}{b-a}\,\mathrm{d}x = \left[\frac{x^2}{2(b-a)}\right]_a^b = \frac{a+b}{2},$$

and the variance by:

$$\begin{aligned} \boldsymbol{\sigma}^2(X) &= \mathbf{P}((X - \mathbf{P}(X))^2 = \int_a^b \frac{1}{b-a}\left(x - \frac{a+b}{2}\right)^2 \mathrm{d}x \\ &= \frac{1}{b-a}\,\frac{1}{3}\left[\left(x - \frac{a+b}{2}\right)^3\right]_a^b \\ &= \frac{(b-a)^2}{12}. \end{aligned}$$

3.6 Exponential Distribution

A random number X has *exponential distribution* with parameter λ if its c.d.f. is given by:

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

If X is the time when a certain fact happens (for example when the atom of some isotope decays), the exponential distribution has the property of *absence of memory*. Given $x, y \geq 0$ we have:

$$\mathbf{P}(X > x + y \mid X > y) = \mathbf{P}(X > x). \tag{3.3}$$

i.e. the probability that the fact does not occur for an extra amount of time x, given that has not occurred up to time y, is the same as the probability starting from the initial time. We obtain (3.3) by using the formula of composite probability:

$$\begin{aligned} \mathbf{P}(X > x + y \mid X > y) &= \frac{\mathbf{P}(X > x + y, X > y)}{\mathbf{P}(X > y)} \\ &= \frac{\mathbf{P}(X > x + y)}{\mathbf{P}(X > y)} \\ &= \frac{e^{-\lambda(x+y)}}{e^{-\lambda y}} \\ &= e^{-\lambda x} \\ &= \mathbf{P}(X > x). \end{aligned}$$

In the following we shall see that the exponential distribution can be obtained as limit of suitably rescaled geometric distributions. Geometric distribution has also the property of absence of memory for discrete times, as we have remarked in Sect. 2.4. The expectation of exponential distribution with parameter λ is equal to

$$\mathbf{P}(X) = \int_0^{+\infty} \lambda x e^{-\lambda x}\, \mathrm{d}x = \left[-x e^{-\lambda x}\right]_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x}\, \mathrm{d}x = \frac{1}{\lambda}.$$

The variance is equal to

$$\begin{aligned} \sigma^2(X) &= \mathbf{P}(X^2) - \mathbf{P}(X)^2 \\ &= \int_0^{+\infty} \lambda x^2 e^{-\lambda x}\, \mathrm{d}x - \frac{1}{\lambda^2} \\ &= \left[-x^2 e^{-\lambda x}\right]_0^{+\infty} + 2\int_0^{+\infty} x e^{-\lambda x}\, \mathrm{d}x - \frac{1}{\lambda^2} \\ &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \\ &= \frac{1}{\lambda^2}. \end{aligned}$$

3.7 A Characterization of Exponential Distribution

The exponential distribution can be characterized in terms of its *hazard rate*.

Given a non-negative random variable with absolutely continuous distribution that describes the time of occurence of some fact, its *hazard rate* $h(x)$ at time x is defined by:

$$h(x) = \lim_{h\to 0} \frac{\mathbf{P}(x < X < x + h | X > x)}{h}.$$

We can express $h(x)$ in terms of the probability density. Let

$$F(x) = \mathbf{P}(X \leq x) = \int_{-\infty}^{x} f(y)\mathrm{d}y.$$

Then

$$\lim_{h\to 0} \frac{\mathbf{P}(x < X < x + h)}{h\mathbf{P}(X > x)} = \frac{f(x)}{1 - F(x)} = -\frac{d}{dx}\log(1 - F(x)).$$

For exponential distribution with parameter λ, it is easy to see that the hazard rate is equal to λ for all x. Indeed:

$$h(x) = \frac{f(x)}{1 - F(x)} = \frac{\lambda e^{-\lambda x}}{e^{-\lambda x}} = \lambda.$$

Exponential distribution can be characterized as the unique distribution with *constant hazard rate*. To see that, we first show that c.d.f. can be obtained from the hazard rate.

Since X is assumed to be non-negative and with absolutely continuous distribution, we have: $F(0) = \mathbf{P}(X \leq 0) = 0$. Using that

$$h(x) = -\frac{d}{dx}\log(1 - F(x)),$$

we have that for $x \geq 0$

$$\log(1 - F(x)) = -\int_0^x h(y)\mathrm{d}y \tag{3.4}$$

$$= 1 - F(x) = \exp\left(-\int_0^x h(y)\mathrm{d}y\right) \tag{3.5}$$

$$= F(x) = 1 - \exp\left(-\int_0^x h(y)\mathrm{d}y\right). \tag{3.6}$$

If the hazard rate is constant equal to $\lambda > 0$, then

$$F(x) = 1 - e^{-\lambda x}, \qquad x > 0,$$

since X is non-negative, $F(x) = 0$ for $x < 0$. Therefore X has exponential distribution with parameter λ.

3.8 Normal Distribution

A random number X has standard *normal distribution* $N(0, 1)$ if its probability density function is:

$$n(x) = Ke^{-\frac{x^2}{2}}, \; x \in \mathbb{R}.$$

Although the indefinite integral of $e^{-\frac{x^2}{2}}$ cannot be expressed in terms of elementary functions, it can still be computed over the whole line and so the constant K. We have:

$$\begin{aligned}
\left(\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}\,\mathrm{d}x\right)^2 &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} e^{-\frac{y^2}{2}}\,\mathrm{d}x\mathrm{d}y \\
&= \int\!\!\int e^{-\frac{x^2+y^2}{2}}\,\mathrm{d}x\mathrm{d}y \\
&= \int_0^{2\pi}\int_0^{+\infty} e^{-\frac{r^2}{2}}\, r\,\mathrm{d}r\mathrm{d}\theta \\
&= 2\pi \int_0^{+\infty} e^{-\frac{r^2}{2}}\, r\,\mathrm{d}r \\
&= 2\pi \left[-e^{-\frac{r^2}{2}}\right]_0^{+\infty} \\
&= 2\pi,
\end{aligned}$$

where a change to polar coordinates $x = r\cos\theta,\; y = r\sin\theta$ has been used. The Jacobian determinant of this change of variable is r (see Appendix H).

It follows that $\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} = \sqrt{2\pi}$ and so

$$K = \frac{1}{\sqrt{2\pi}}.$$

The cumulative distribution function will be denoted by $\mathcal{N}(x)$:

$$\mathcal{N}(x) := \int_{-\infty}^{x} n(t)\,\mathrm{d}t.$$

Since n is an even function and its integral over the whole line is equal to 1, we have:

$$\mathcal{N}(-x) = 1 - \mathcal{N}(x).$$

Therefore in tables of $\mathcal{N}(x)$, only values for positive values of x are usually tabulated. The expectation of standard normal distribution is

$$\mathbf{P}(X) = \int_{-\infty}^{+\infty} x\, n(x)\, \mathrm{d}x = 0,$$

as it follows immediately since $f(x) = -f(x)$, where $f(x) = xn(x)$, $x \in \mathbb{R}$. The variance of standard normal distribution is obtained by integration by parts, using the fact that $n'(x) = -xn(x)$:

$$\boldsymbol{\sigma}^2(X) = \mathbf{P}(X^2) = \int_{-\infty}^{+\infty} x^2 n(x)\, \mathrm{d}x = [-xn(x)]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} n(x)\, \mathrm{d}x = 1.$$

We introduce now the general *normal distribution* which has two parameters m, σ^2 and will be denoted by $N(m, \sigma^2)$. We start with $X \sim N(0, 1)$ and consider $Y = m + \sigma X$, where $\sigma > 0$. Then Y has normal distribution $N(m, \sigma^2)$. The c.d.f. of Y is given by:

$$\begin{aligned} F_Y(y) &= \mathbf{P}(Y \le y) \\ &= \mathbf{P}(m + \sigma X \le y) \\ &= \mathbf{P}\left(X \le \frac{y-m}{\sigma}\right) \\ &= \mathcal{N}\left(\frac{y-m}{\sigma}\right). \end{aligned}$$

The probability density function of Y is obtained by chain rule for the derivative of a composite function:

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}\mathcal{N}\left(\frac{y-m}{\sigma}\right) = \frac{1}{\sigma}n\left(\frac{y-m}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(y-m)^2}{2\sigma^2}}.$$

The expectation and the variance of Y are obtained as follows:

$$\mathbf{P}(Y) = \mathbf{P}(\sigma X + m) = \sigma\mathbf{P}(X) + m = m$$

$$\boldsymbol{\sigma}^2(Y) = \boldsymbol{\sigma}^2(\sigma X + m) = \sigma^2\boldsymbol{\sigma}^2(X) = \sigma^2.$$

3.9 Normal Tail Estimate

As we have said, there is no formula in terms of elementary functions for $\mathcal{N}(x)$ and therefore for the probability that a random number $X \sim N(0, 1)$ is greater than some $x > 0$. It is however possible to give asymptotic estimates for this probability as x tends to infinity.

Proposition 3.9.1 *Let X be a random number with standard normal distribution. For every $x > 0$, we have:*

$$\frac{n(x)}{x} - \frac{n(x)}{x^3} < \mathbf{P}(X \geq x) < \frac{n(x)}{x},$$

where $n(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

The upper bound is obtained by integration by parts:

$$\begin{aligned}\mathbf{P}(X \geq x) &= \int_x^{+\infty} n(t)\,\mathrm{d}t = \int_x^{+\infty} t\frac{n(t)}{t}\,\mathrm{d}t \\ &= \underbrace{\left[-\frac{n(t)}{t}\right]_x^{+\infty}}_{\frac{n(x)}{x}} - \int_x^{+\infty} \underbrace{\frac{n(t)}{t^2}}_{>0}\,\mathrm{d}t < \frac{n(x)}{x}.\end{aligned}$$

A second integration by parts gives the lower bound:

$$\begin{aligned}\mathbf{P}(X \geq x) &= \frac{n(x)}{x} - \int_x^{+\infty} t\frac{n(t)}{t^3}\,\mathrm{d}t \\ &= \frac{n(x)}{x} - \underbrace{\left[-\frac{n(t)}{t^3}\right]_x^{+\infty}}_{\frac{n(x)}{x^3}} + \int_x^{+\infty} \underbrace{\frac{3n(t)}{t^4}}_{>0}\,\mathrm{d}t > \frac{n(x)}{x} - \frac{n(x)}{x^3}.\end{aligned}$$

3.10 Gamma Distribution

Let α and λ be strictly positive real numbers. The random number X is said to have *gamma distribution* $\Gamma(\alpha, \lambda)$ if its probability density function is given by

$$g_{\alpha,\lambda}(x) = \begin{cases} K\,x^{\alpha-1}\,e^{-\lambda x} & x > 0, \\ 0 & x \leq 0. \end{cases}$$

Note that exponential distribution is a particular case of gamma distribution corresponding to the choice $\alpha = 1$.

The normalizing constant K can be expressed in terms of *Euler's gamma function* $\Gamma(\alpha)$:

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

for $\alpha > 0$. The function Γ satisfies the recursive property:

1. $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$, since

$$\begin{aligned}\Gamma(\alpha+1) &= \int_0^{+\infty} x^{\alpha}\, e^{-x}\,\mathrm{d}x \\ &= \left[-x^{\alpha}\, e^{-x}\right]_0^{+\infty} + \int_0^{+\infty} \alpha x^{\alpha-1}\, e^{-x}\,\mathrm{d}x \\ &= \alpha\,\Gamma(\alpha).\end{aligned}$$

2. It follows by iteration that for integer $\alpha > 0$

$$\Gamma(\alpha) = (\alpha - 1)!$$

since $\Gamma(1) = \int_0^{+\infty} e^{-x}\mathrm{d}x = 1$.

Now for the p.d.f. $g_{\alpha,\lambda}$ we have

$$1 = \int_{-\infty}^{+\infty} g_{\alpha,\lambda}(x)\,\mathrm{d}x = K\int_0^{+\infty} x^{\alpha-1}\, e^{-\lambda x}\mathrm{d}x = \frac{K}{\lambda^{\alpha}}\int_0^{+\infty} y^{\alpha-1}e^{-y}\,\mathrm{d}y = \frac{K}{\lambda^{\alpha}}\Gamma(\alpha).$$

Hence

$$K = \frac{\lambda^{\alpha}}{\Gamma(\alpha)}.$$

The expectation and the variance of the gamma distribution can be computed using the recurrence property of gamma function:

$$\begin{aligned}\mathbf{P}(X) &= \int_{-\infty}^{+\infty} x g_{\alpha,\lambda}(x)\,\mathrm{d}x \\ &= \frac{\lambda^{\alpha}}{\Gamma(\alpha)}\int_0^{+\infty} x^{\alpha}\, e^{-\lambda x}\,\mathrm{d}x \\ &= \frac{\lambda^{\alpha}}{\Gamma(\alpha)}\,\frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} \\ &= \frac{\alpha}{\lambda}.\end{aligned}$$

It follows that:

$$\boldsymbol{\sigma}^2(X) = \mathbf{P}(X^2) - \mathbf{P}(X)^2 = \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha^2}{\lambda^2}.$$

3.11 χ^2-Distribution

From the normal distribution we can derive another distribution of wide use in statistics, the χ^2-*distribution*. In this section we introduce the χ^2-distribution with parameter $\nu = 1$. In Chap. 4 we shall consider general χ^2-distributions with parameter $\nu \in \mathbb{N} \setminus \{0\}$.

Let X be a random number with standard normal distribution $N(0, 1)$ and let $Y = X^2$. We first consider the c.d.f. of Y. If $y < 0$

$$F_Y(y) = \mathbf{P}(Y \leq y) = 0$$

since Y is non-negative. If $y \geq 0$, then

$$\begin{aligned} F_Y(y) = \mathbf{P}(Y \leq y) &= \mathbf{P}(X^2 \leq y) \\ &= \mathbf{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \mathcal{N}(\sqrt{y}) - \mathcal{N}(-\sqrt{y}) \\ &= \mathcal{N}(\sqrt{y}) - (1 - \mathcal{N}(\sqrt{y})) \\ &= 2\mathcal{N}(\sqrt{y}) - 1. \end{aligned}$$

The c.d.f. of Y is therefore

$$F_Y(y) = \begin{cases} 0 & \textit{for } y < 0, \\ 2\mathcal{N}(\sqrt{y}) - 1 & \textit{for } y \geq 0. \end{cases}$$

Let us compute the p.d.f. f_Y of Y (for $y > 0$):

$$\begin{aligned} f_Y(y) \;=\; F_Y'(y) \;&=\; 2n(y)\frac{1}{\sqrt{y}} \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{y}}e^{-\frac{y}{2}} \;=\; \frac{1}{\sqrt{2\pi}}y^{\frac{1}{2}-1}e^{-\frac{1}{2}y}, \end{aligned}$$

where the derivative has been computed by using chain rule for the derivative of composite functions. The density $f_Y(y)$ is of course zero for negative y. It follows that Y has distribution $\Gamma(\frac{1}{2}, \frac{1}{2})$. Moreover by comparing the normalizing constants, we get

$$\frac{1}{\sqrt{2}}\frac{1}{\sqrt{\pi}} = \left(\frac{1}{2}\right)^{\frac{1}{2}}\frac{1}{\Gamma\left(\frac{1}{2}\right)},$$

so that

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

By using the recurrence formula $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, we have:

$$\Gamma\left(\frac{2k+1}{2}\right) = \frac{(2k-1)(2k-3)\cdots 1}{2^k}\frac{\sqrt{\pi}}{2}$$

for $k = 1, 2, \ldots$.

3.12 Cauchy Distribution

We now consider a distribution for which the expectation defined in Sect. 3.3 does not exist. This is the *Cauchy distribution*. This is the distribution of a random number $Y = \tan\Theta$, where the random number Θ has uniform distribution in the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$. We have for $y \in \mathbb{R}$ that:

$$\begin{aligned} F_Y(y) = \mathbf{P}(Y \leq y) &= \mathbf{P}(\tan\Theta \leq y) \\ &= \mathbf{P}(\Theta \leq \arctan y). \end{aligned}$$

The p.d.f. of Y f_Y is obtained by deriving F_Y:

$$f_Y(y) = \frac{1}{\pi\left(1+y^2\right)}.$$

The formula for the expectation of Y gives an integral

$$\int \frac{y}{\pi\left(1+y^2\right)}dy$$

which is undefined, as the integrand behaves like $\frac{1}{y}$ for $y \to \infty$.

3.13 Mixed Cumulative Distribution Functions

In addition to discrete and absolutely continuous c.d.f.'s, there are continuous but not absolutely continuous c.d.f.'s. These will be not considered in this elementary book. Here we briefly speak about mixed c.d.f.'s that are convex linear combinations of discrete and absolutely continuous c.d.f.'s.

For $0 < p < 1$, let F_1 be a discrete c.d.f. and F_2 be an absolutely continuous c.d.f. Then we can consider a c.d.f. $F(x)$:

$$F(x) = pF_1(x) + (1-p)F_2(x),$$

which is neither of discrete nor of absolutely continuous type. $F(x)$ is said to be a *mixed c.d.f.* If X is a random number with c.d.f. $F(x)$, it is easy to see that the expectation of a function $\phi(X)$ is given by

$$\mathbf{P}(\phi(X)) = p\mathbf{P}(\phi(X_1)) + (1-p)\mathbf{P}(\phi(X_2)),$$

where X_1 and X_2 are random numbers with c.d.f. F_1 and F_2 respectively, provided that the terms on the right-hand side both make sense. The first term is expressed by a sum or a series, while the second by an integral.

An example of random number with mixed c.d.f is the time T of function of some device, for example a lamp, when there is a positive probability p that the device does not work already at the initial time and otherwise the distribution is absolutely continuous, for example exponential with parameter λ. The c.d.f of T is then given by:

$$F_T(t) = \begin{cases} 0 & \text{for } t < 0, \\ p + (1-p)(1-e^{-\lambda t}) & \text{for } t \geq 0. \end{cases}$$

It is easy to check that $\mathbf{P}(T) = \dfrac{1-p}{\lambda}$.

Chapter 4
Multi-dimensional Absolutely Continuous Distributions

4.1 Bidimensional Distributions

Let X, Y be two random numbers that we can consider as a random vector (X, Y). The *joint cumulative distribution function* (j.c.d.f.) is defined as:

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

Then F is a map from $\mathbb{R}^2$ to [0, 1]:

$$F : \mathbb{R}^2 \longrightarrow [0, 1].$$

The probability that (X, Y) belong to the rectangle $(a_1, b_1] \times (a_2, b_2]$ is given by:

$$\begin{aligned} \mathbf{P}(a_1 < X \leq b_1, a_2 < Y \leq b_2) &= \mathbf{P}\left[((X \leq b_1) - (X \leq a_1))\,((Y \leq b_2) - (Y \leq a_2))\right] \\ &= \mathbf{P}(X \leq b_1, Y \leq b_2) - \mathbf{P}(X \leq a_1, Y \leq b_2) \\ &\quad - \mathbf{P}(X \leq b_1, Y \leq a_2) + \mathbf{P}(X \leq a_1, Y \leq a_2) \\ &= F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2). \end{aligned} \tag{4.1}$$

We shall always assume that the following *continuity properties* are verified:

1. $\lim\limits_{\substack{x \to +\infty \\ y \to +\infty}} F(x, y) = 1;$
2. $\lim\limits_{x \to -\infty} F(x, y) = \lim\limits_{y \to -\infty} F(x, y) = 0;$
3. $\lim\limits_{\substack{x \to x_0^+ \\ y \to y_0^+}} F(x, y) = F(x_0, y_0);$
4. $\mathbf{P}(X = x_0, Y = y_0) = F(x_0, y_0) - F(x_0^-, y_0) - F(x_0, y_0^-) + F(x_0^-, y_0^-),$

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8_4

where $F(x_0^-, y_0) := \lim_{x \to x_0^-} F(x, y_0)$, $F(x_0, y_0^-) := \lim_{y \to y_0^-} F(x_0, y)$ and $F(x_0^-, y_0^-) := \lim_{\substack{x \to x_0^- \\ y \to y_0^-}} F(x, y)$.

Other analogous properties will also be assumed. We shall quote them when they will be needed.

4.2 Marginal Cumulative Distribution Functions

Given two random numbers X, Y with j.c.d.f. $F(x, y)$, the c.d.f.'s F_1, F_2 of X and Y are called *marginal cumulative distribution functions* (m.c.d.f.'s).

The m.c.d.f. of X is obtained from the j.c.d.f. by taking the limit:

$$F_1(x) = \mathbf{P}(X_1 \leq x) = \lim_{y \to +\infty} F(x, y),$$

as follows by usual continuity hypothesis. Similarly the m.c.d.f. of Y is obtained by:

$$F_2(y) = \mathbf{P}(Y \leq y) = \lim_{x \to +\infty} F(x, y).$$

Two numbers are said to be *stochastically independent* if:

$$F(x, y) = F_1(x) F_2(y)$$

for every $(x, y) \in \mathbb{R}^2$.

4.3 Absolutely Continuous Joint Distributions

Two random numbers X, Y or equivalently the random vector (X, Y) has an absolutely continuous distribution if there exists a function f

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R}$$

such that the j.c.d.f. F of X, Y can be expressed as:

$$f(X, Y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s, t) \, \mathrm{d}s \mathrm{d}t.$$

Such function f is called *joint probability density (j.p.d.)*. Applying formula (4.1) for the probability that (X, Y) belong to a rectangle $(a, b] \times (c, d]$, we get:

$$
\begin{aligned}
\mathbf{P}(a < X \le b, c < Y \le d) &= F(b,d) - F(a,d) - F(c,b) + F(a,c) \\
&= \int_{\infty}^{b}\int_{\infty}^{d} f(s,t)\,\mathrm{d}s\mathrm{d}t - \int_{\infty}^{a}\int_{\infty}^{d} f(s,t)\,\mathrm{d}s\mathrm{d}t \\
&\quad - \int_{\infty}^{b}\int_{\infty}^{c} f(s,t)\,\mathrm{d}s\mathrm{d}t + \int_{\infty}^{a}\int_{\infty}^{c} f(s,t)\,\mathrm{d}s\mathrm{d}t \\
&= \int_{a}^{b}\int_{c}^{d} f(s,t)\,\mathrm{d}s\mathrm{d}t.
\end{aligned}
$$

By usual limiting procedure one gets that the probability that a random vector (X, Y) belongs to a sufficiently regular region A of $\mathbb{R}^2$ is given by the integral of the j.p.d.f. over A, i.e.

$$\mathbf{P}((X,Y) \in A) = \int\int_A f(s,t)\,\mathrm{d}s\mathrm{d}t.$$

Moreover, if $\psi : \mathbb{R}^2 \to \mathbb{R}$ is a sufficiently regular function such that the function ψf is integrable, then, as in the one-dimensional case, we have that for $Z = \psi(X, Y)$

$$\mathbf{P}(Z) = \int\int_{\mathbb{R}^2} \psi(s,t) f(s,t)\,\mathrm{d}s\mathrm{d}t. \tag{4.2}$$

For example, if $Z = XY$ we get

$$\mathbf{P}(XY) = \int\int_{\mathbb{R}^2} st f(s,t)\,\mathrm{d}s\mathrm{d}t,$$

if the integrand function $stf(s,t)$ is integrable. In order to derive probability densities of X and Y, that are called *marginal probability densities*, we start by deriving their c.d.f.'s:

$$F_X(x) = \mathbf{P}(X \le x) = \int_{-\infty}^{+\infty}\int_{-\infty}^{x} f(s,t)\,\mathrm{d}s\mathrm{d}t.$$

It follows that the marginal probability density of X is given by:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,t)\,\mathrm{d}t.$$

Analogously f_Y, the marginal probability density of Y, is given by

$$f_Y(y) = \int_{-\infty}^{+\infty} f(s,y)\,\mathrm{d}s.$$

It is easy to check out that, if $f(X, Y)$ can be expressed as a product of two functions,

$$f(x,y) = u(x)v(y),$$

then X and Y are stochastically independent and their marginal probability densities are proportional to $u(x)$ and $v(y)$. Conversely if X and Y are stochastically independent and their joint distribution is absolutely continuous, then their joint probability density can be expressed as the product of their marginal probability densities:

$$f(x, y) = f_X(x) f_Y(y). \tag{4.3}$$

As in the case of discrete distributions, it follows from (4.3) that if X, Y are stochastically independent and ϕ_1, ϕ_2 are real functions such that $\phi_1 f_X$ and $\phi_2 f_Y$ are integrable, then by Fubini's theorem we obtain:

$$\mathbf{P}(\phi_1(X)\phi_2(Y)) = \mathbf{P}(\phi_1(X))\mathbf{P}(\phi_2(Y)).$$

4.4 The Density of $Z = X + Y$

Let X and Y be two random numbers with joint probability density $f(x, y)$. We want to determine the density of

$$Z = X + Y.$$

First we compute the c.d.f. of Z:

$$\begin{aligned} F_Z(z) = \mathbf{P}(Z \le z) = \mathbf{P}(X + Y \le z) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} f(x, y)\,\mathrm{d}y\mathrm{d}x \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{z} f(x, t - x)\,\mathrm{d}t\mathrm{d}x = \int_{-\infty}^{z} \int_{-\infty}^{+\infty} f(x, t - x)\,\mathrm{d}x\mathrm{d}t, \end{aligned}$$

where we have made the change of variable $t = x + y$ for fixed x, that allows then to exchange the order of integration in the final equality. It follows from the last expression that

$$f_Z(z) = \int_{-\infty}^{z} f_z(t)\,\mathrm{d}t$$

with

$$f_Z(z) = \int_{-\infty}^{+\infty} f(x, z - x)\,\mathrm{d}x,$$

i.e. f_Z is the density of Z. In particular when X and Y are stochastically independent and $f(x, y) = f_X(x) f_Y(y)$, then

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z - x)\,\mathrm{d}x.$$

Hence f_Z is obtained by the *convolution of* f_X *and* f_Y and is denoted by $f_X * f_Y$. An example of application of this formula is the sum of two stochastically independent gamma distributed random numbers with parameters respectively α, λ and β, λ. Using the previous formula we obtain the probability density of $Z = X + Y$:

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)\,\mathrm{d}x \\
&= \int_{-\infty}^{+\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} I_{\{x>0\}} \frac{\lambda^\beta}{\Gamma(\beta)} (z-x)^{\beta-1} e^{-\lambda(z-x)} I_{\{(z-x)>0\}}\,\mathrm{d}x,
\end{aligned}
$$

where I_A denotes the indicator function of the set A. The integral can be written as

$$
\frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda z} \int_0^z x^{\alpha-1}(z-x)^{\beta-1}\mathrm{d}x
$$

if $z > 0$ and it is equal to 0 if $z \leq 0$. For $z > 0$ we make the change of variable $\mathrm{d}x = z\mathrm{d}t$ and obtain

$$
\begin{aligned}
f_Z(z) &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda z} \int_0^z x^{\alpha-1}(z-x)^{\beta-1}\,\mathrm{d}x \\
&= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda z} \int_0^1 (zt)^{\alpha-1}(z-zt)^{\beta-1} z\,\mathrm{d}t \\
&= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} z^{\alpha+\beta-1} e^{-\lambda z} \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t \\
&= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt\, z^{\alpha+\beta-1} e^{-\lambda z} \\
&= K\, z^{\alpha+\beta-1} e^{-\lambda z},
\end{aligned}
$$

with

$$
K = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\,\Gamma(\beta)} \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t. \tag{4.4}
$$

It follows that Z has distribution $\Gamma(\alpha+\beta, \lambda)$.

Remark 4.4.1 Since the constant K must be equal to the normalizing constant of the distribution $\Gamma(a+b, \lambda)$, by (4.4) we obtain

$$
K = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)},
$$

so that

$$
\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}.
$$

4.5 Beta Distribution $\mathcal{B}(\alpha, \beta)$

Let $\alpha > 0$ and $\beta > 0$. A random number X is said to have *beta distribution* $\mathcal{B}(\alpha, \beta)$ if its density $f(x)$ is given by

$$f(x) = \begin{cases} K\, x^{\alpha-1}\,(1-x)^{\beta-1} & x \in [0,1], \\ 0 & \text{otherwise.} \end{cases}$$

It follows from the computation at the end of the previous section that

$$K = \frac{1}{\int_0^1 x^{\alpha-1}\,(1-x)^{\beta-1}\,\mathrm{d}x} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}. \tag{4.5}$$

The expectation can be obtained from the recursion property of Euler's gamma function. If X has $\mathcal{B}(\alpha, \beta)$ distribution, then

$$\mathbf{P}(X) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \int_0^1 x\, f(x)\,\mathrm{d}x.$$

The value of the integral is obtained by (4.5) by replacing α with $\alpha + 1$ so that

$$\mathbf{P}(X) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,\frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} = \frac{\alpha\Gamma(\alpha)}{\Gamma(\alpha)}\,\frac{\Gamma(\alpha+\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)} = \frac{\alpha}{\alpha+\beta}.$$

Similarly we can compute $\mathbf{P}(X^2)$:

$$\begin{aligned} \mathbf{P}(X^2) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \int_0^1 x^{\alpha+1}\,(1-x)^{\beta-1}\,\mathrm{d}x \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,\frac{\Gamma(\alpha+2)\,\Gamma(\beta)}{\Gamma(\alpha+\beta+2)}, \end{aligned}$$

where the integral is obtained by replacing α with $\alpha + 2$ in formula (4.4). By using the recursion property of the Gamma function we get

$$\begin{aligned} \Gamma(\alpha+2) &= (\alpha+1)\alpha\,\Gamma(\alpha) \\ \Gamma(\alpha+\beta+2) &= (\alpha+\beta+1)\,(\alpha+\beta)\,\Gamma(\alpha+\beta) \end{aligned}$$

so that

$$\mathbf{P}(X^2) = \frac{\alpha(\alpha+1)}{(\alpha+\beta)\,(\alpha+\beta+1)}$$

and

$$\begin{aligned}\boldsymbol{\sigma}^2(X) &= \mathbf{P}(X^2) - \mathbf{P}(X)^2 \\ &= \frac{(\alpha+1)\,\alpha}{(\alpha+\beta+1)\,(\alpha+\beta)} - \frac{\alpha^2}{(\alpha+\beta)^2} = \frac{\alpha\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)}.\end{aligned}$$

4.6 Student Distribution

We now introduce the *Student distribution* of parameter ν. Let Z and U be stochastically independent random numbers. We assume that Z has standard normal distribution and U has gamma distribution $\Gamma\left(\frac{\nu}{2}, \frac{1}{2}\right)$ where $\nu \in \mathbb{N}$. The latter distribution is called χ^2-distribution with ν degrees of freedom and plays an important role in statistics. Let $T = Z\left(\frac{U}{\nu}\right)^{-\frac{1}{2}}$. In order to obtain the probability density of T, we first derive its c.d.f.

$$F_T(t) = \mathbf{P}(T \leq t) = \mathbf{P}(Z \leq t\sqrt{\frac{U}{\nu}}) = \int_0^\infty \int_{-\infty}^{\sqrt{\frac{u}{\nu}}} f(z, u)\mathrm{d}z\mathrm{d}u,$$

where

$$f(z, u) = \frac{1}{2^{\frac{\nu}{2}}\sqrt{2\pi}\Gamma\left(\frac{\nu}{2}\right)} e^{-\frac{z^2}{2}} u^{\frac{\nu}{2}-1} e^{-\frac{u}{2}}.$$

By taking the derivative of $F_T(t)$ with respect to t, it follows from the fundamental calculus theorem that for density of the Student distribution is given for $t > 0$ by

$$\begin{aligned}f_T(t) &= F_T'(t) = \int_0^\infty f(t\sqrt{\frac{u}{\nu}}, u)\sqrt{\frac{u}{\nu}}\mathrm{d}u \\ &= \frac{1}{2^{\frac{\nu}{2}}\sqrt{2\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \int_0^\infty u^{\frac{\nu+1}{2}-1} e^{-\frac{u}{2}(1+\frac{t^2}{\nu})}\mathrm{d}u \\ &= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)}(1+\frac{t^2}{\nu})^{-\frac{\nu+1}{2}},\end{aligned}$$

where the integral has been computed by using the formula for the normalizing constant of the gamma distribution. Note that for $\nu = 1$ the Student distribution coincides with the Cauchy distribution. Since

$$\int_{-\infty}^{+\infty} \frac{|t|}{(1+\frac{t^2}{\nu})^{\frac{\nu+1}{2}}}\mathrm{d}t$$

must be finite for the existence of $\mathbf{P}(T)$, we have that $\mathbf{P}(T)$ exists and is finite if and only if $\nu > 1$. We have that

$$\mathbf{P}(T) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \int_{-\infty}^{+\infty} t(1+\frac{t^2}{\nu})^{-\frac{\nu+1}{2}} \mathrm{d}t = 0,$$

since the integrand is an odd function.To compute the variance, we calculate

$$\begin{aligned}
\boldsymbol{\sigma}(T) = \mathbf{P}(T^2) &= \mathbf{P}\left(\frac{\nu Z^2}{U}\right) = \nu\mathbf{P}(Z^2)\mathbf{P}\left(\frac{1}{U}\right) = \nu\mathbf{P}\left(\frac{1}{U}\right) \\
&= \frac{\nu}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} \int_0^\infty \frac{1}{u} u^{\frac{\nu}{2}-1} e^{-\frac{u}{2}} \mathrm{d}u = \frac{\nu}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} \int_0^\infty u^{\frac{\nu-2}{2}-1} e^{-\frac{u}{2}} \mathrm{d}u \\
&= \frac{\nu}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} 2^{\frac{\nu-2}{2}} \Gamma\left(\frac{\nu-2}{2}\right) = \frac{\nu}{\nu-2}.
\end{aligned}$$

Hence the variance exists finitely if $\nu > 2$.

4.7 Multi-dimensional Distributions

Let $(X_1, X_2, \dots, X_n)$ be an n-dimensional random vector. The function

$$F : \mathbb{R}^n \longrightarrow [0, 1]$$

defined by:

$$F(x_1, x_2, \dots, x_n) = \mathbf{P}(X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n)$$

is called *joint cumulative distribution function* (j.c.d.f.) of $(X_1, X_2, \dots, X_n)$. In the following we shall always assume that the following continuity properties are satisfied by j.c.d.f.'s:

1. $\lim_{x_1,\dots,x_n \to +\infty} F(x_1, x_2, \dots, x_n) = 1$;
2. $\lim_{x_i \to -\infty} F(x_1, x_2, \dots, x_n) = 0$
3. If $\{i_1, \dots, i_k\} \subset \{1, 2, \dots, n\}$ and $\{j_1, \dots, j_{n-k}\} = \{1, 2, \dots, n\} \setminus \{i_1, \dots, i_k\}$ then $\lim_{x_{j_1},\dots,x_{j_{n-k}} \to +\infty} F(x_1, \dots, x_n) = \mathbf{P}(X_{i_1} \le x_{i_1,\dots,X_{i_k}} \le x_{i_k})$.

Here $F_{i_1,\dots,i_k}(x_{i_1}, \dots, x_{i_k}) := \mathbf{P}(X_{i_1} \le x_{i_1}, \dots, X_{i_k} \le x_{i_k})$, $x_{i_1}, \dots, x_{i_k} \in \mathbb{R}$, is called the *marginal cumulative distribution* function (m.c.d.f.) of $X_{i_1}, \dots, X_{i_k}$. As in the two-dimensional case the probability that $X_1, \dots, X_n$ belongs to some intervals $(a_1, b_1], \dots, (a_n, b_n]$ can be computed using the j.c.d.f. Precisely:

$$\mathbf{P}(a_1 < X_1 \le b_1, \dots, a_n < X_n \le b_n) = \sum_c (-1)^{\epsilon(c)} F(c_1, \dots, c_n)$$

with $c = (c_1, \dots, c_n)$, where c_i can be a_i or b_i, and $\epsilon(c)$ is equal to the number of i's such that $c_i = a_i$. The proof of this formula is completely analogous to the one for (4.1) in the two-dimensional case.

The random numbers $X_1, \dots, X_n$ are said to be stochastically independent if

$$F(x_1, \dots, x_n) = F_1(x_1) \dots F_n(x_n),$$

where F_i is the m.c.d.f. of X_i for $i = 1, \dots, n$. If $X_1, \dots, X_n$ are stochastically independent, then

$$\begin{aligned}\mathbf{P}(a_1 < X_1 \leq b_1, \dots, a_n < X_1 \leq b_n) &= \sum_c (-1)^{\epsilon(c)} F_1(c_1) \dots F_n(c_n) \\ &= \Pi_{i=1}^n (F_i(b_i) - F_i(a_i)) \\ &= \Pi_{i=1}^n P(a_i < X_i \leq b_i).\end{aligned}$$

4.8 Absolutely Continuous Multi-dimensional Distributions

The random vector $(X_1, \dots, X_n)$ has an absolutely continuous distribution if there exists a function

$$f : \mathbb{R}^n \longrightarrow \mathbb{R}$$

such that the j.c.d.f. F of $(X_1, X_2, \dots, X_n)$ is given by:

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(t_1, t_2, \dots, t_n)\, \mathrm{d}t_1 \mathrm{d}t_2 \dots \mathrm{d}t_n.$$

It follows from Property 1 of Sect. 4.7 that

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(t_1, \dots, t_n)\, \mathrm{d}t_1 \dots \mathrm{d}t_n = 1.$$

Moreover it can be shown that one can always choose a non-negative f. The function f is called *joint probability density* (j.p.d.) of $(X_1, \dots, X_n)$. What we have said about two-dimensional joint probability density generalizes in a natural way to the n-dimensional case.

If A is a sufficiently regular region $A \subset \mathbb{R}^n$ then

$$\mathbf{P}((X_1, \dots, X_n) \in A) = \int \cdots \int_A f(t_1, \dots, t_n) \mathrm{d}t_1 \cdots \mathrm{d}t_n.$$

If ψ is a function $\psi : \mathbb{R}^n \longrightarrow \mathbb{R}$ such that ψf is integrable, then

$$\mathbf{P}(\psi(X_1, \dots, X_n)) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \psi(t_1, \dots, t_n) \mathrm{d}t_1 \dots \mathrm{d}t_n.$$

If $f(t_1, \dots, t_n) = g_1(t_1) \cdots g_n(t_n)$, then $X_1, \dots, X_n$ are stochastically independent and the marginal density of X_i can be taken proportional to g_i for $i = 1, \dots, n$. Conversely if $X_1, \dots, X_n$ are stochastically independent with absolutely continuous distribution, the j.p.d. of $X_1, \dots, X_n$ can be taken as $f(t_1, \dots, t_n) = f_1(t_1) \dots f_n(t_n)$, where $f_1, \dots, f_n$ are marginal probability density functions of $X_1, X_2, \dots, X_n$.

4.9 Multi-dimensional Gaussian Distribution

A random vector $(X_1, X_2, \dots, X_n)$ has n-dimensional *Gaussian distribution* if its density has the form:

$$f(x_1, x_2, \dots, x_n) = K e^{-\frac{1}{2} Ax \cdot x + b \cdot x}$$

where $x = (x_1, x_2, \dots, x_n)^t \in \mathbb{R}^n$, $b = (b_1, b_2, \dots, b_n)^t \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix.[1] The symbol A^t denotes the *transpose matrix* of A, with elements

$$[A^t]_{i,j} = [A]_{j,i}.$$

We remind that $b \cdot x$ is the *scalar product* of b and x, given by

$$b \cdot x = \sum_{i=1}^{n} b_i x_i$$

and that Ax is the vector with elements

$$[Ax]_i = \sum_j a_{ij} x_j.$$

Let $a_{i,j}$ denote $[A]_{i,j}$. The expression $Ax \cdot x$ is a quadratic form

$$\sum_{i,j} a_{ij} x_i x_j.$$

[1]Recall that a matrix $A \in \mathbb{R}^{n \times n}$ is

- symmetric if $A^t = A$, i.e. $a_{ij} = a_{ji}$,
- positive definite if $Ax \cdot x > 0$ for all $x \neq 0$, $x \in \mathbb{R}$.

If we have a quadratic form

$$Bx \cdot x = \sum_{i,j} b_{ij} x_i x_j$$

we can always replace the matrix B with a symmetric matrix A such that

$$Ax \cdot x = \sum_{i,j} a_{ij} x_i x_j = Bx \cdot x,$$

where a_{ij} is defined by

$$a_{ij} = \begin{cases} b_{ii} & \text{for } i = j, \\ (b_{ij} + b_{ji})/2 & \text{for } i \neq j. \end{cases}$$

We consider first the simplest case:

Case 1: *A* diagonal and *b* = 0
Let

$$A = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}$$

and $b = 0$. We obtain[2]

$$f(x_1, x_2, \ldots, x_n) = K \exp\left(-\left(\lambda_1 \frac{x_1^2}{2} + \lambda_2 \frac{x_2^2}{2} + \cdots + \lambda_n \frac{x_n^2}{2}\right)\right).$$

By computing the marginal densities, it is easy to get

$$f(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1)\, f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

where

$$f_{X_i}(x_i) = \sqrt{\frac{\lambda_i}{2\pi}} \exp\left(-\frac{\lambda_i\, x_i^2}{2}\right)$$

[2]Here the notation $\exp(x)$ is introduced to denote the exponential function e^x.

is the marginal density of X_i. It follows that

1. $X_1, \ldots, X_n$ are stochastically independent;
2. X_i has gaussian density $N\left(0, \dfrac{1}{\lambda_i}\right)$;
3. the normalizing constant is given by:

$$K = \sqrt{\frac{\lambda_1}{2\pi}}\sqrt{\frac{\lambda_2}{2\pi}}\cdots\sqrt{\frac{\lambda_n}{2\pi}} = \sqrt{\frac{\det A}{(2\pi)^n}}.$$

The *expectation vector* is given by

$$(\mathbf{P}(X_1), \ldots, \mathbf{P}(X_n)) = (0, \ldots, 0)$$

and the *covariance matrix* is:

$$C = \begin{pmatrix} \boldsymbol{\sigma}^2(X_1) & \mathbf{cov}(X_1, X_2) & \cdots & \mathbf{cov}(X_1, X_n) \\ \mathbf{cov}(X_2, X_1) & \boldsymbol{\sigma}^2(X_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{cov}(X_{n-1}, X_n) \\ \mathbf{cov}(X_n, X_1) & \cdots & \mathbf{cov}(X_n, X_{n-1}) & \boldsymbol{\sigma}^2(X_n) \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\lambda_n} \end{pmatrix}$$

$$= A^{-1}.$$

Case 2: Computation of the expectation vector in the general case

Let now A be symmetric and positive definite and $b \neq 0$. By making a *translation* we can reduce the density to the case $b = 0$. Let $U = X - c$ with $c \in \mathbb{R}$. The j.c.d.f. of the random vector U can be expressed in terms of that of X:

$$F_U(u) = \mathbf{P}(U \leq u) = \mathbf{P}(X - c \leq u) = \mathbf{P}(X \leq u + c) = F_X(u + c).$$

It follows that the joint probability density can be similarly obtained from that of X:

$$\begin{aligned}
f_U(u_1, u_2, \ldots, u_n) &= f_X(u_1 + c_1, u_2 + c_2, \ldots, u_n + c_n) \\
&= K' \exp\left[-\frac{1}{2}A(u+c)\cdot(u+c) + b\cdot(u+c)\right] \\
&= K' \exp\left(-\frac{1}{2}Au\cdot u - \frac{1}{2}Au\cdot c - \frac{1}{2}Ac\cdot u - \frac{1}{2}Ac\cdot c + b\cdot u + b\cdot c\right) \\
&= \underbrace{K' \exp\left(-\frac{1}{2}Ac\cdot c + b\cdot c\right)}_{\text{constant}} \exp\left(-\frac{1}{2}Au\cdot u + (b - Ac)\cdot u\right),
\end{aligned}$$

where we have used the fact that

$$Ac\cdot u = Au\cdot c,$$

since A is symmetric. In order to reduce the density to the case $b = 0$, we must choose c so that the first degree part cancels, i.e.:

$$b - Ac = 0.$$

We choose therefore

$$c = A^{-1}b.$$

Note that A is invertible since it is positive definite. For this choice of c the density $f_U(u_1, u_2, \ldots, u_n)$ is given by:

$$\begin{aligned}
f_U(u_1, u_2, \ldots, u_n) &= f_X(u_1 + c_1, u_2 + c_2, \ldots, u_n + c_n) \\
&= K' \exp\left(A^{-1}b\cdot b - \frac{A(A^{-1}b)\cdot A^{-1}b}{2}\right) \exp\left(-\frac{1}{2}Au\cdot u\right) \\
&= \underbrace{K \exp\left(\frac{1}{2}A^{-1}b\cdot b\right)}_{K'} \exp\left(-\frac{1}{2}Au\cdot u\right) \\
&= K' \exp\left(-\frac{1}{2}Au\cdot u\right).
\end{aligned}$$

It is easy to see that $\mathbf{P}(U_i) = 0$ for $i = 1, 2, \ldots, n$, since the density of $-U$ and U are the same. Using previous results, we obtain that

$$\mathbf{P}(X_i) = \mathbf{P}(U_i + c_i) = \mathbf{P}(U_i) + c_i = c_i = (A^{-1}b)_i,$$

i.e. in vectorial notation:

$$\mathbf{P}(X) = A^{-1}b;$$

where the expectation of a random vector is defined as the vector of the expectations of its components. The normalizing constant is

$$K = K' \exp\left(\frac{1}{2} A^{-1} b \cdot b\right),$$

where K' is the normalizing constant for the case with $b = 0$. The covariance matrix of X is equal to one of U, as a translation leaves variances and covariances unchanged:

$$C = \begin{pmatrix} \boldsymbol{\sigma}^2(X_1) & \mathbf{cov}(X_1, X_2) & \cdots & \mathbf{cov}(X_1, X_n) \\ \mathbf{cov}(X_2, X_1) & \boldsymbol{\sigma}^2(X_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{cov}(X_{n-1}, X_n) \\ \mathbf{cov}(X_n, X_1) & \cdots & \mathbf{cov}(X_n, X_{n-1}) & \boldsymbol{\sigma}^2(X_n) \end{pmatrix}$$

$$= \begin{pmatrix} \boldsymbol{\sigma}^2(U_1) & \mathbf{cov}(U_1, U_2) & \cdots & \mathbf{cov}(U_1, U_n) \\ \mathbf{cov}(U_2, U_1) & \boldsymbol{\sigma}^2(U_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{cov}(U_{n-1}, U_n) \\ \mathbf{cov}(U_n, U_1) & \cdots & \mathbf{cov}(U_n, U_{n-1}) & \boldsymbol{\sigma}^2(U_n) \end{pmatrix}.$$

Case 3: computation of covariance matrix and normalization constant in the general case

As it is shown we can reduce to the case $b = 0$ by making a translation. Since A is symmetric, there exists an orthogonal matrix O, i.e. such that $O^t A O = D$, where D is diagonal.

If U is the random vector $U = O^{-1} X$, its density is given by

$$\begin{aligned} f(u_1, \ldots, u_n) &= K \exp\left(-\frac{1}{2} A\, Ou \cdot Ou\right) \\ &= K \exp\left(-\frac{1}{2} O^t A O u \cdot u\right) \\ &= K \exp\left(-\frac{1}{2} Du \cdot u\right). \end{aligned}$$

Now for U we are in the situation of a diagonal matrix that we have already considered. The covariance matrix of X, is given by:

$$C = \mathbf{P}(XX^t) = \mathbf{P}(OU\,(OU)^t)$$
$$= O\,\mathbf{P}(U\,U^t)\,O^t = O\,D^{-1}\,O^t = A^{-1}.$$

Here the expectation of a random matrix denotes a matrix whose entries are the expectations of the corresponding entries. We have used the easily verifiable fact that if Z is a random matrix and A, B are constant matrices, such that the product AZB is defined, then $\mathbf{P}(AZB) = A\mathbf{P}(Z)B$.

We have found that in the general case

1. the *normalization constant* is

$$K = \sqrt{\frac{\det A}{(2\pi)^n}}\, e^{-\frac{1}{2}A^{-1}b\cdot b}\,;$$

2. the *expectation* is

$$\mathbf{P}(X) = A^{-1}b\,;$$

3. the *covariance matrix* is

$$C = A^{-1}.$$

Remark 4.9.1 It is easy to check that the marginal distribution of the X_i's and of subsets of the X_i's are gaussian. In particular, if $\mathbf{cov}(X_i, X_j) = 0$ for some $i, j\; i \neq j$, then the covariance matrix of (X_i, X_j) is diagonal, so that X_i and X_j are stochastically independent as it is shown in the next remark.

Remark 4.9.2 When $n = 2$, the covariance matrix is given by:

$$C = \begin{pmatrix} \sigma_1^2 & \rho\,\sigma_1\sigma_2 \\ \rho\,\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

where $\sigma_1^2 = \boldsymbol{\sigma}^2(X_1), \sigma_2^2 = \boldsymbol{\sigma}^2(X_2)$ and $\rho = \boldsymbol{\rho}(X_1, X_2)$. The matrix A can be obtained as:

$$A = C^{-1} = \frac{1}{\det C}\begin{pmatrix} \sigma_2^2 & -\rho\,\sigma_1\sigma_2 \\ -\rho\,\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$$
$$= \frac{1}{\sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2}\begin{pmatrix} \sigma_2^2 & -\rho\,\sigma_1\sigma_2 \\ -\rho\,\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$$

$$= \frac{1}{1-\rho^2} \begin{pmatrix} \dfrac{1}{\sigma_1^2} & -\dfrac{\rho}{\sigma_1\sigma_2} \\ -\dfrac{\rho}{\sigma_1\sigma_2} & \dfrac{1}{\sigma_2^2} \end{pmatrix}.$$

The density of two-dimensional gaussian distribution with parameters $m_1 = \mathbf{P}(X_1)$, $m_2 = \mathbf{P}(X_2)$, $\sigma_1^2 = \boldsymbol{\sigma}^2(X_1)$, $\sigma_2^2 = \boldsymbol{\sigma}^2(X_2)$, $\rho = \boldsymbol{\rho}(X_1, X_2)$ is therefore given by:

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-m_1)^2}{\sigma_1^2} - 2\frac{\rho(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\right)\right).$$

Chapter 5
Convergence of Distributions

5.1 Convergence of Cumulative Distribution Functions

It is natural to introduce a notion of *convergence for sequences of cumulative distribution functions*, i.e. to give a meaning to the expression $F_n \to F$. One possible meaning could be pointwise convergence, i.e.: $F_n(x) \to F(x)$ for every $x \in \mathbb{R}$. However this notion of convergence turns out to be too restrictive. For example consider the sequence $F_n(x)$ defined as:

$$F_n(x) = \begin{cases} 1 & \text{for } x \geq \frac{1}{n}, \\ 0 & \text{for } x < 0. \end{cases}$$

If the random number X_n has c.d.f. F_n, then $\mathbf{P}(X_n = \frac{1}{n}) = 1$. For a reasonable convergence notion we should have $F_n \to F$, where

$$F(x) = \begin{cases} 1 & \text{for } x \geq 0, \\ 0 & \text{for } x < 0. \end{cases}$$

However it is not true in this case that $F_n(x) \to F(x)$ for every $x \in \mathbb{R}$. Indeed $F_n(0) = 0$ for every n, whereas $F(0) = 1$. Therefore it is natural to introduce a weaker definition of convergence.

Definition 5.1.1 We say that $F_n \to F$ for every x if for every $\epsilon > 0$ there exists N such that for $n \geq N$

$$F(x - \epsilon) - \epsilon < F_n(x) < F(x + \epsilon) + \epsilon.$$

If x is a continuity point of F, this definition implies that

$$\lim_{n \to \infty} F_n(x) = F(x).$$

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8_5

Conversely if for every continuity point x of F $\lim_{n\to\infty} F_n(x) = F(x)$, then $F_n \to F$. For a cumulative distribution, continuity points make up an everywhere dense set since discontinuity points are denumerable. Indeed there cannot be more than n discontinuity points with jump larger than or equal to $\frac{1}{n}$ because F is bounded by 1 from above. Let then $x \in \mathbb{R}$ and $\epsilon > 0$. There exist two continuity points x_0, x of F such that $x - \epsilon < x_0 < x < x_1 < x + \epsilon$. We have then

$$F(x-\epsilon) \le \lim_{n\to\infty} F_n(x_0) = F(x_0)$$

and also

$$\lim_{n\to\infty} F_n(x_1) = F(x_1) \le F(x+\epsilon).$$

On the other side for every n

$$F_n(x_0) \le F_n(x) \le F_n(x_1).$$

Therefore for n sufficiently large we have

$$F(x-\epsilon) - \epsilon < F_n(x) < F(x+\epsilon) + \epsilon.$$

It is easy to build up examples of sequences of absolutely continuous c.d.f.'s converging to a discrete (pure jump) c.d.f. For example if $F_n(x) = \mathcal{N}(\sqrt{n}x)$, the c.d.f. of normally distributed X_n with $\mathbf{P}(X) = 0$ and $\sigma^2(X_n) = \frac{1}{n}$, then $F_n \to F$ with

$$F(x) = \begin{cases} 1 & \text{for } x \ge 0, \\ 0 & \text{for } x < 0. \end{cases}$$

Conversely we can build up examples of discrete c.d.f's converging to an absolutely continuous c.d.f. For example if

$$F_n(x) = \begin{cases} 0 & \text{for } x \le 0, \\ \frac{[nx]}{n} & \text{for } 0 < x \le 1, \\ 1 & \text{for } x > 1, \end{cases}$$

where $[x]$ denotes the integer part of x, then $F_n \to F$, where F is the c.d.f. of the uniform distribution in $[0, 1]$:

$$F(x) = \begin{cases} 0 & \text{for } x \le 0, \\ x & \text{for } 0 < x \le 1, \\ 1 & \text{for } x > 1. \end{cases}$$

5.2 Convergence of Geometric Distribution to Exponential Distribution

We have seen that geometric and exponential distributions share the property of absence of memory, the former among discrete distributions, the latter among absolutely continuous distributions. Let us now consider a sequence $(X_n)_{n\in\mathbb{N}}$ of random numbers with geometric distributions with parameters p_n:

$$\mathbf{P}(X_n = k) = p_n(1 - p_n)^{k-1}, \quad \forall k \geq 1.$$

We assume that np_n converges to $\lambda > 0$, as $n \to \infty$. We put $Y_n = \frac{X_n}{n}$ and denote by F_{Y_n} the c.d.f. of Y_n. We have that

$$F_{Y_n} \to F,$$

where F is the c.d.f. of exponential distribution with parameter $\lambda > 0$, i.e.:

$$F(x) = \begin{cases} 0 & \text{for } x < 0, \\ 1 - e^{-\lambda x} & \text{for } x \geq 0. \end{cases}$$

Indeed for $x < 0$, $F_{Y_n} \equiv 0$, as $\vdash Y_n \geq 0$. For $x \geq 0$

$$\begin{aligned} F_{Y_n}(x) &= \mathbf{P}(Y_n \leq x) = \mathbf{P}(X_n \leq nx) \\ &= 1 - \sum_{k=[nx]+1}^{\infty} p_n(1 - p_n)^{k-1} \\ &= 1 - p_n(1 - p_n)^{[nx]} \sum_{i=0}^{\infty} (1 - p_n)^i \\ &= 1 - p_n(1 - p_n)^{[nx]} \frac{1}{1 - (1 - p_n)} \\ &= 1 - (1 - p_n)^{[nx]}, \end{aligned}$$

where we have used the formula for the sum of geometric series. We write

$$nx = [nx] + \delta_n, \quad \text{with } 0 \leq \delta_n < 1.$$

We obtain therefore

$$F_{Y_n}(x) = 1 - (1 - p_n)^{nx - \delta_n},$$

which tends to $1 - e^{-\lambda x}$ for $n \to \infty$, since

$$\log(1 - p_n)^{nx} = nx \log(1 - p_n) = -xnp_n + o(np_n)$$

which tends to $-\lambda x$ for $n \to \infty$, whereas

$$(1 - p_n)^{\delta_n} \xrightarrow[n\to\infty]{} 1,$$

as $0 \le \delta_n < 1$ and $p_n \xrightarrow[n\to\infty]{} 0$.

5.3 Convergence of Binomial Distribution to Poisson Distribution

We now provide an approximation of the binomial distribution when we consider the number of successes in a large number of trials.

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of binomially distributed random numbers with parameters n, p_n such that $np_n \to \lambda$ with $\lambda > 0$ as $n \to \infty$. For example X_n represent the number of successes in n Bernoulli trials with parameter p_n. As the number of trials grows to infinity we send to 0 the probability of success in a single trial. For $0 \le k \le n$:

$$\begin{aligned}
\mathbf{P}(X_n = k) &= \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\
&= \underbrace{\frac{n!}{k!(n-k)!} p_n^k (1 - p_n)^{n-k} \frac{n^k}{n^k}}_{\text{multiplication and division by } n^k} \\
&= \frac{1}{k!}\left(1 - \frac{1}{n}\right)\cdots\left(1 - \frac{k-1}{n}\right)(np_n)^k (1 - p_n)^{n-k}.
\end{aligned}$$

We observe that:

- $\left(1 - \frac{1}{n}\right)\cdots\left(1 - \frac{k-1}{n}\right)$ tends to 1 as $n \to \infty$;
- $(np_n)^k$ tends to λ^k for $n \to \infty$;
- $(1 - p_n)^{-k}$ tends to 1 for $n \to \infty$;
- $(1 - p_n)^n$ tends to $e^{-\lambda}$ for $n \to \infty$ as

  $$\log(1 - p_n)^n = n\log(1 - p_n) = -np_n + o(np_n)$$

  tends to $-\lambda$.

It follows that for $k \in \mathbb{N}$

$$\mathbf{P}(X_n = k) \xrightarrow[n\to\infty]{} \frac{\lambda^k}{k!}e^{-\lambda},$$

and therefore the sequence of binomial c.d.f.'s with parameters n, p_n tends to Poisson c.d.f. with parameter λ.

5.4 De Moivre-Laplace Theorem

We consider now another type of convergence for sequences of binomial c.d.f's. We send the number of trials to infinity but this time we keep fixed the probability of success in a single trial. In order to obtain convergence we need to perform a linear rescaling.

Theorem 5.4.1 *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random numbers with binomial distribution $Bn(n, p)$ with $0 < p < 1$ and let X_n^* be the corresponding* standardized *random numbers given by*

$$X_n^* = \frac{X_n - \mathbf{P}(X_n)}{\sigma(X_n)} = \frac{X_n - np}{\sqrt{np\tilde{p}}}$$

for $n \in \mathbb{N} \setminus \{0\}$. Where $\tilde{p} = 1 - p$. Then we have for all $n \in \mathbb{N} \setminus \{0\}$

$$\mathbf{P}(X_n^* = x) = \frac{h_n}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}e^{E_n(x)},$$

where $h_n = \dfrac{1}{\sqrt{np\tilde{p}}}$ and the error $E_n(x)$ tends uniformly to 0 when x ranges on $I(X_n^*)\bigcap[-K, K]$ for any fixed constant K.

Proof The set $I(X_n)$ of the possible values of X_n is $I(X_n) = \{0, 1, \dots, n\}$. Therefore

$$I(X_n^*) = \{h_n(-np), h_n(1 - np), \dots, h_n(n - np)\}$$

where $h_n = \dfrac{1}{\sqrt{np\tilde{p}}}$ is the spacing between possible values of X_n^*.

We define $\phi_n(x) = \log \mathbf{P}(X_n^* = x)$ for $x \in I(X_n^*)$ and consider its incremental ratio:

$$\frac{\phi_n(x + h_n) - \phi_n(x)}{h_n} = \frac{1}{h_n}\log\frac{\mathbf{P}(X_n^* = x + h_n)}{\mathbf{P}(X_n^* = x)}.$$

Putting $k = np + x\sqrt{np\tilde{p}}$, we obtain

$$\begin{aligned}
\frac{1}{h_n}\log\frac{\mathbf{P}(X_n^* = x + h_n)}{\mathbf{P}(X_n^* = x)} &= \frac{1}{h_n}\log\frac{\mathbf{P}(X_n = k+1)}{\mathbf{P}(X_n = k)}\\
&= \frac{1}{h_n}\log\frac{(n-k)p}{(k+1)\tilde{p}}\\
&= \sqrt{np\tilde{p}}\log\frac{n\tilde{p} - x\sqrt{np\tilde{p}}}{n+1+x\sqrt{np\tilde{p}}}\frac{p}{\tilde{p}}\\
&= \sqrt{np\tilde{p}}\log\frac{1 - x\sqrt{\dfrac{p}{n\tilde{p}}}}{1 + \dfrac{1}{np} + x\sqrt{\dfrac{\tilde{p}}{np}}}\,.
\end{aligned}$$

Using the 1-order expansion of the logarithm $\log(1+x) = x + O(x^2)$, we obtain

$$\begin{aligned}
&\sqrt{np\tilde{p}}\log\frac{1 - x\sqrt{\dfrac{p}{n\tilde{p}}}}{1 + \dfrac{1}{np} + x\sqrt{\dfrac{\tilde{p}}{np}}}\\
&= \sqrt{np\tilde{p}}\left[-x\sqrt{\frac{p}{n\tilde{p}}} + O(\frac{x^2}{n}) - x\sqrt{\frac{p}{n\tilde{p}}} + O(\frac{x^2+1}{n})\right]\\
&= -xp - x\tilde{p} + O(\frac{x^2+1}{\sqrt{n}})\\
&\quad - x + O(\frac{x^2+1}{\sqrt{n}})\,.
\end{aligned}$$

The function $\phi_n(x)$ is not defined everywhere, but only for x in $I(X_n^*)$. We can extend it to values between two elements of $I(X_n^*)$ by linear interpolation. In this way we can write

$$\phi_n(x) = \phi_n(0) + \int_0^x \phi_n'(y)dy\,.$$

If $x \le y \le x + h_n$, then $\phi_n'(y) = \Delta_{h_n}\phi_n(x) = -x + O(\frac{x^2+1}{\sqrt{n}}) = -y + O(\frac{x^2+1}{\sqrt{n}})$ so that:

$$\begin{aligned}
\phi_n(x) &= \phi_n(0) + \int_0^x \phi_n'(y)dy\\
&= \phi_n(0) + \int_0^x (-y)dy + O(\frac{|x|^3 + |x|}{\sqrt{n}})\\
&= \phi_n(0) - \frac{x^2}{2} + O(\frac{|x|^3 + |x|}{\sqrt{n}})\,.
\end{aligned}$$

Since $\phi_n(x) = \log \mathbf{P}(X_n^* = x)$, we obtain

$$\log \mathbf{P}(X_n^* = x) = e^{\phi_n(0)} e^{-\frac{x^2}{2}} e^{E_n(x)}$$

where $E_n(x) = O\left(\frac{|x|^3 + |x|}{\sqrt{n}}\right)$.

We can estimate $e^{\phi_n(0)}$ in the following way: X_n^* is a standardized random number, i.e. $\mathbf{P}(X_n^*) = 0$ and $\sigma^2(X_n^*) = 1$. By the Chebychev inequality, we have that:

$$\mathbf{P}(|X_n^*| \geq K) \leq \frac{1}{K^2}\,.$$

K can be chosen so that this probability is arbitrary small, that is for every $\epsilon > 0$ there is K such that:

$$1 - \epsilon = 1 - \frac{1}{K^2} \leq \mathbf{P}(|X_n^*| < K) \leq 1.$$

Since $\mathbf{P}(|X_n^*| < K) = \sum_{x,|x|<K} \mathbf{P}(|X_n^*| = x)$, it follows that:

$$1 - \epsilon \leq \sum_{x,|x|<K} \mathbf{P}(X_n^* = x) \leq 1\,.$$

Moreover

$$\mathbf{P}(|X_n^*| < K) = \sum_{x,|x|<K} \mathbf{P}(|X_n^*| = x) = \sum_{x,|x|<K} h_n e^{-\frac{x^2}{2}}\,.$$

Since $E_n(X)$ tends uniformly to 0 on bounded interval and $\sum_{x,|x|<K} h_n e^{-\frac{x^2}{2}}$ is the Riemann sum for the function $e^{-\frac{x^2}{2}}$ and tends to $\int_{-K}^{K} e^{-\frac{x^2}{2}} dx$, we have for n sufficiently large that:

$$1 - 2\epsilon \leq \frac{e^{\phi_n(0)}}{h_n} \int_{-K}^{K} e^{-\frac{x^2}{2}} dx \leq 1\,.$$

Let K tending to infinity, we obtain

$$1 - 3\epsilon \leq \frac{e^{\phi_n(0)}}{h_n} \sqrt{2\pi} \leq 1$$

so that

$$\frac{e^{\phi_n(0)}}{h_n} \sqrt{2\pi} \xrightarrow[n\to\infty]{} 1\,.$$

It follows that

$$\mathbf{P}(|X_n^*| = x) = \frac{h_n}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{E_n(x)} ,$$

where $E_n(x)$ is an error that tends uniformly to 0 for x ranging on the possible values of X_n^* in a bounded interval.

As application of the theorem, one obtains an approximation of the c.d.f. of the binomial distribution. Given $a, b, a < b$:

$$\mathbf{P}(a \leq X_n^* \leq b) = \sum_{a \leq x \leq b} \mathbf{P}(X_n^* = x) = \sum_{a \leq x \leq b} \frac{h_n}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{E_n(x)} .$$

This is the Riemann sum of $n(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, therefore it converges to

$$\frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx = \mathcal{N}(b) - \mathcal{N}(a) ,$$

where $\mathcal{N}(x)$ is the c.d.f. of standard Gaussian distribution. The c.d.f. $F_n(x)$ of X_n^* converges to $\mathcal{N}(x)$ since

$$\begin{aligned} F_n(x) &= \mathbf{P}(X_n^* \leq x) = \mathbf{P}(-k < X_n^* \leq x) + \mathbf{P}(X_n^* \leq -k) \\ &= \mathcal{N}(x) - \mathcal{N}(-k) + \mathbf{P}(X_n^* \leq -k) + E_n'(x) \end{aligned}$$

with $\lim_{n \to \infty} E_n'(x) = 0$. The Chebychev inequality states that $\mathbf{P}(X_n^* \leq -k)$ can be made arbitrarily small. Also $\mathcal{N}(-k)$ tends to 0 for $k \longrightarrow \infty$. Therefore the c.d.f. of standardized binomial distributions tend to $\mathcal{N}$.

Chapter 6
Discrete Time Markov Chains

6.1 Homogeneous Discrete Time Markov Chains with Finite State Space

We define a *homogeneous Markov chain* with *finite state space* $S \subset \mathbb{R}$ as a sequence of random numbers $(X_i)_{i \in \mathbb{N}} \subset S$ for $i \in \mathbb{N}$ such that:

$$\mathbf{P}(X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n) = \rho_{s_0} p_{s_0,s_1} p_{s_1,s_2} \cdots p_{s_{n-1},s_n},$$

where

1. ρ_{s_i}, $s_i \in S$, is called *initial distribution*:

$$\rho_{s_i} = \mathbf{P}(X_0 = s_i) \quad \text{and} \quad \sum_{s \in S} \rho_s = 1\,.$$

2. $p_{s,s'} = [P]_{s,s'}$, are called *transition probabilities* and satisfy:
   - $0 \le p_{ij} \le 1$;
   - $\sum_{s' \in S} p_{s,s'} = 1 \quad$ for every $s \in S$.

They can be arranged in a matrix P called *transition probability matrix* of entries

$$[P]_{s,s'} =: p_{s,s'}.$$

The Markov chain $(X_i)_{i \in \mathbb{N}}$ can be seen as representing the evolution of a system that moves from one state to another in a random fashion. We have assumed that $S \subset \mathbb{R}$, but it may be convenient to consider in some situations a general finite set S. In this case X_i are not random numbers, but random entities. However what follows goes through without any change.

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8_6

We show now that $p_{s,s'}$ is the probability to go from state s to state s'. Moreover we show that the probability that $X_{r+1} = s'$ conditional to all previous history $X_0 = s_0, \dots, X_{r-1} = s_{r-1}, X_r = s$ depends just on s and is equal to $p_{s,s'}$ (*Markov property*). Indeed:

$$\begin{aligned}\mathbf{P}(X_{r+1} = s'|X_r = s, X_{r-1} = s_{r-1}, \dots, X_0 = s_0) \\ &= \frac{\mathbf{P}(X_{r+1} = s', X_r = s, X_{r-1} = s_{r-1}, \dots, X_0 = s_0)}{\mathbf{P}(X_r = s, X_{r-1} = s_{r-1}, \dots, X_0 = s_0)} \\ &= \frac{\rho_{s_0} p_{s_0,s_1} \cdots p_{s_{r-1},P_{s,s'}}}{\rho_{s_0} p_{s_0,s_1} \cdots p_{s_{r-1},s}, p_{s,s'}} \\ &= p_{s,s'},\end{aligned}$$

provided that the probability of the conditioning event, which is at denominator, is positive (this is required to compute the conditional probability).

Example 6.1.1 (Random walk). A *random walk* in an integer interval $[a, b] \subset \mathbb{Z}$, with *absorbing boundary conditions* is a Markov chain with state space $S = [a, b]$ and transition probability matrix:

$$P = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ 1-p & 0 & p & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & 1-p & 0 & p \\ 0 & \cdots & \cdots & 0 & 0 & 1 \end{pmatrix},$$

where $0 < p < 1$. Boundary conditions are determined by the transition probabilities from state a and b. Other boundary conditions can be considered: reflecting, mixed, …In the case $p = \frac{1}{2}$ we speak of *symmetric random walk.*

Example 6.1.2 (Bernoulli-Laplace chain). Let us consider two urns A and B, each containing N balls. The balls are assumed to be identical apart from their colors. Among the balls there are N white balls and N black balls. At each integer time we choose one ball from each urn and exchanges them.

Let X_i be the random number of white balls in A at time i. The state space is

$$S = \{0, 1, \dots, N\}.$$

The transition probability from state k to state l is given by:

$$\begin{aligned}p_{k,k} &= \mathbf{P}(\text{two white balls or two black balls are drawn}) \\ &= \frac{k}{N}\frac{N-k}{N} + \frac{N-k}{N}\frac{k}{N} = 2\frac{k}{N}\frac{N-k}{N};\end{aligned} \tag{6.1}$$

$$p_{k,k+1} = \mathbf{P}(1 \text{ black ball from urn } A \text{ and } 1 \text{ white ball from } B\ )$$
$$= \frac{N-k}{N}\,\frac{N-k}{N} = \frac{(N-k)^2}{N^2}; \tag{6.2}$$
$$p_{k,k-1} = \mathbf{P}(1 \text{ white ball from } A \text{ and } 1 \text{ black ball from } B\ )$$
$$= \frac{k}{N}\,\frac{k}{N} = \frac{k^2}{N^2}\,. \tag{6.3}$$

The transition probabilities to other states are zero. This applies also to the case $k = 0$ and $k = N$. The transition matrix is therefore:

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ \frac{1}{N^2} & \frac{2(N-1)}{N^2} & \left(\frac{N-1}{N}\right)^2 & 0 & \cdots & 0 \\ 0 & \frac{4}{N^2} & \frac{4(N-2)}{N^2} & \left(\frac{N-2}{N}\right)^2 & \cdots & 0 \\ \vdots & & & & \vdots & \\ 0 & \cdots & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

6.2 Transition Probability in n Steps

By using composite probability formula we can compute the probability for a Markov chain to go from state s to state s' in n steps. Let $s_0, s_1, \ldots, s_{m-1}, s$ be a sequence of states such that $\rho_{s_0}\, p_{s_0,s_1}\, p_{s_1,s_2} \cdots\, p_{s_{m-1},s}$ are strictly positive. We have:

$$\mathbf{P}(X_{m+n} = s' | X_m = s, X_{m-1} = s_{m-1}, \ldots, X_0 = s_0)$$

$$= \frac{\mathbf{P}(X_{m+n} = s', X_m = s, X_{m-1} = s_{m-1}, \ldots, X_0 = s_0)}{\mathbf{P}(X_m = s, X_{m-1} = s_{m-1}, \ldots, X_0 = s_0)}$$

$$= \frac{\sum_{s_{m+1},\ldots,s_{m+n-1}} \mathbf{P}(X_{m+n} = s', X_{m+n-1} = s_{m+n-1}, \ldots, X_0 = s_0)}{\mathbf{P}(X_m = s, X_{m-1} = s_{m-1}, \ldots, X_0 = s_0)}$$

$$= \frac{\sum_{s_{m+1},\ldots,s_{m+n-1}} \rho_{s_0}\, p_{s_0,s_1} \cdots\, p_{s_{m-1},s}\, p_{s,s_{m+1}} \cdots\, p_{s_{m+n-1},s'}}{\rho_{s_0}\, p_{s_0,s_1}\, p_{s_1,s_2} \cdots\, p_{s_{m-1},s}}$$

$$= \sum_{s_{m+1},\ldots,s_{m+n-1}} p_{s,s_{m+1}} \cdots\, p_{s_{m+n-1},s'}$$

$$= \left[P^n\right]_{s,s'}\,.$$

This probability *does not* depend on m, but just on n, that is the number of intermediate steps. It is obtained as the element with coordinates s, s' of the n-th power of the transition matrix P. In the following we will use the common notation $p^{(n)}_{s,s'}$ for this probability:

$$p^{(n)}_{s,s'} := \mathbf{P}(X_{m+n} = s' | X_m = s) = \left[P^n\right]_{s,s'} .$$

By convention one defines:

$$p^{(0)}_{s,s'} := p_{s,s'} = \begin{cases} 1 & \text{if } s = s', \\ 0 & \text{otherwise.} \end{cases}$$

6.3 Equivalence Classes

Let $(X_i)_{i\in\mathbb{N}}$ be a homogeneous Markov chain. We say that the state s *communicates* with the state s' if there exists $n > 0$ such that

$$p^{(n)}_{s,s'} > 0,$$

that is if there exists a path $s, s_1, \ldots, s_{n-1}, s'$ such that all transition probabilities $p_{s,s_1}, p_{s_1,s_2}, \ldots, p_{s_{n-1},s_n}$ are strictly positive. We will use the notation $s \prec s'$ to indicate that s communicates with s'.

Two states s, s', are said to be *equivalent* if $s \prec s'$ and $s' \prec s$. This is an equivalence relation, i.e. it is reflexive, symmetric and transitive. The first two properties are evident. Transitivity follows from transitivity of communication. Assume that $s \prec s'$ and $s' \prec s''$. Then there are n_1, n_2 such that $p^{n_1}_{s,s'} > 0$ and $p^{n_2}_{s',s''} > 0$. It follows that $s \prec s''$. Indeed:

$$p^{(n_1+n_2)}_{s,s''} = \left[P^{n_1+n_2}\right]_{s,s''} = \sum_{s_1} p^{(n_1)}_{s,s_1}\, p^{(n_2)}_{s_1,s''} \geq \underbrace{p^{(n_1)}_{s,s'}}_{>0}\, \underbrace{p^{(n_2)}_{s',s''}}_{>0} > 0 .$$

The communication relation $\prec$ between states can be extended without ambiguity to equivalence classes. We indicate with $[s]$ the equivalence class of the state s, i.e. the set of all states s' equivalent to s according to the previously introduced relation. When $s \prec s'$ we say that s' *follows* s. We say that $[s]$ communicates with $[s']$ and write $[s] \prec [s']$ if $s \prec s'$. Using the transitivity property it is easy to check that this is a well-posed definition, i.e. it does not depend on the choices of the representatives in the equivalence classes.

An equivalence class is said to be *maximal* if it is not followed by any other class with respect to the communication relation. If a Markov time is in a state of a maximal equivalence class, then at all subsequent times, it will be in states of the same class with probability 1.

Another characteristic of a state of a Markov chain is its *period*. Let $s \in S$ be a state of a Markov chain and let:

$$A_s^+ = \{n > 0 \,|\, p_{s,s}^{(n)} > 0\}\,.$$

If $A_s^+ \neq \emptyset$, we define the *period* of s as the greatest common divisor (GCD) of the elements of A_s^+. If the period of s is 1, we say that s is an *aperiodic* state. For example, in the random walk on the interval $[a, b]$ with absorbing boundary conditions all states s with $a < s < b$ have period 2.

All states of an equivalence class have the same period. Therefore one can speak of the period of an equivalence class.

Proof Let us consider two equivalent states $s \sim s'$, and q, q' their periods. It is enough to show that q' divides every $n \in A_s^+$. In force of the equivalence, there is n_1 such that $p_{s,s'}^{(n_1)} > 0$ and there is n_2 such that $p_{s',s}^{(n_2)} > 0$. Then $(n_1 + n_2) \in A_s^+$ since

$$p_{s,s}^{(n_1+n_2)} = \sum_{s_1} p_{s,s_1}^{(n_1)} \, p_{s_1,s}^{(n_2)} \geq p_{s,s'}^{(n_1)} \, p_{s',s}^{(n_2)} > 0.$$

Similarly $(n_1 + n_2) \in A_{s'}^+$; hence q and q' both divide $(n_1 + n_2)$. Moreover for all $n \in A_s^+$, $(n + n_1 + n_2) \in A_{s'}^+$, since

$$p_{s',s'}^{(n+n_1+n_2)} \geq p_{s',s}^{(n_2)} \, p_{s,s}^{(n)} \, p_{s,s'}^{(n_1)} > 0\,.$$

Hence, q and q' divide $(n + n_1 + n_2)$ for all $n \in A_s^+$ and for all $n \in A_{s'}^+$. Since n_1, n_2 are divisible by q and by q', q and q' are both common divisors of A_s^+ and of $A_{s'}^+$, so that

$$q = q'\,.$$

An equivalence class C of period $q < \infty$ can be decomposed in q subsets:

$$C = C_0 \cup C_1 \cup \cdots \cup C_{q-1}$$

with the property that if $s \in C_i$, $s' \in C_j$ and $p_{s,s'}^{(n)} > 0$ then

$$n \equiv (j - i) \pmod{q}\,.$$

If a maximal equivalence class C has period q, $C_0, C_1, \ldots, C_{q-1}$ are cyclically visited by the Markov chain: i.e. if $X_0 \in C_i$, then $X_1 \in C_{[i+1]_{mod_q}}$, $X_2 \in C_{[i+2]_{mod_q}}$ with probability 1, where we use the notation $[k]_q$ for the element of the set $\{0, \ldots, q-1\}$ that is equivalent to k modulo q.

6.4 Ergodic Theorem

We want to study the behavior of a Markov chain as time proceeds.

An important result states that a Markov chain with finite state space and a single aperiodic equivalence class has the property that the distribution on state space converges to a limit that does not depend on the initial state. This is the result of the following theorem called *ergodic* theorem (see e.g. Gnedenko (1997) for a proof).

Theorem 6.4.1 (Ergodic theorem) *Let $(X_i)_{i\in\mathbb{N}}$ be a homogeneous Markov chain with a finite state space. If the chain is irreducible (i.e. there is a unique equivalence class) and aperiodic (i.e. the period is 1), then there is a probability distribution $\Pi = (\pi_s)_{s\in S}$ on the state space and constants $C > 0$ and $0 \le \delta < 1$ such that for all $s', s \in S$:*

$$|p^{(n)}_{s',s} - \pi_s| \le C\delta^n .$$

In other words there are π_s, $s \in S$, such that:

1. $0 \le \pi_s \le 1$;
2. $\sum_{s\in S} \pi_s = 1$,

and $\forall\, s' \in S$

$$\lim_{n\to+\infty} p^{(n)}_{s',s} = \pi_s$$

with exponential speed.

This theorem can be used also in the case when the period q is strictly larger than 1, by considering the Markov chain with transition matrix P^q. Indeed, the restriction of this chain to each of the subsets $C_0, C_1, \dots, C_{q-1}$ satisfies the hypothesis of ergodic theorem.

The probability distribution Π that appears in the statement of the ergodic theorem is an *invariant* (or stationary) distribution for the Markov chain: this means that if we take it as initial distribution, so that $\mathbf{P}(X_0 = s) = \pi_s$ for every $s \in S$, then for every $s \in S$ and for every $n \ge 0$

$$\mathbf{P}(X_n = s) = \pi_s .$$

This property allows us to compute π_s as the solution of a system of linear equations. Indeed:

$$\begin{aligned}\pi_s &= \mathbf{P}(X_1 = s)\\ &= \sum_{s'\in S} \mathbf{P}(X_0 = s')\, p_{s',s}\\ &= \sum_{s'\in S} \pi_{s'}\, p_{s',s} .\end{aligned}$$

Moreover, since π_s is a probability distribution, we have

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

Under the hypothesis of the ergodic theorem one can show that there is one and only one solution for this system of $|S| + 1$ equations in $|S|$ unknowns; one of the equations, in this case one of the first $|S|$ equations, is a linear combination of the others and therefore it can be skipped in the solution of the system:

$$\begin{cases} \Pi^t = \Pi^t P \\ \sum_{\pi_s \in S} \pi_s = 1 , \end{cases} \tag{6.4}$$

where we have represented Π as $|S|$-dimensional vector. The ergodic theorem tells us that as time advances the Markov chain forgets the initial state and reaches an equilibrium. We show now the *uniqueness* of the invariant measure.

Proof Let us assume that $(\mu_s)_{s \in S}$ is another probability distribution on the state space satisfying system (6.4). We have

$$\begin{cases} \mu^t = \mu^t P \\ \sum_{s \in S} \mu_s = 1 , \end{cases}$$

where we have represented the distribution $(\mu_s)_{s \in S}$ as the $|S|$-dimensional column vector μ. We have

$$\mu^t = \mu^t P \quad \Rightarrow \quad \mu^t = \mu^t P = \mu^t P^2 = \cdots = \mu^t P^n .$$

If n tends to infinity, by ergodic theorem P^n converges to the matrix:

$$\begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_n \\ \pi_1 & \pi_2 & \cdots & \pi_n \\ \vdots & \vdots & & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_n \end{pmatrix}$$

therefore for $s \in S$

$$\mu_s = \sum_{s'} \mu_{s'} p_{s',s} = \sum_{s'} \mu_{s'} p^{(n)}_{s',s} .$$

By taking the limit $\lim_{n \to +\infty} p^{(n)}_{s',s} = \pi_s$, we have

$$\mu_s = \sum_{s'} \mu_{s'} \pi_s = \pi_s \underbrace{\sum_{s'} \mu_{s'}}_{1} = \pi_s .$$

Chapter 7
Continuous Time Markov Chains

7.1 Introduction

In this chapter we shall introduce some simple queueing systems. For further reading, we refer to [7, 8].

A *queueing system* can be described in terms of servers and a flow of clients who access servers and are served according to some pre-established rules. The clients after service can either stay in the system or leave it, also according to some established rules.

The simplest case is when there is a single set of servers and a flow of clients accessing to it. If there is at least one free server, then an incoming client is served right away. Otherwise, i.e. if all servers are engaged, he is put in a queue and waits for his turn. Once a client is served, he leaves the system.

Usual hypotheses are that service times are stochastically independent, identically distributed, and moreover that they are stochastically independent from the flow of clients' arrivals. One would like to obtain the probabilities that, at given times, there are some numbers of clients in the system. For this one needs to introduce a random number for each time t; this leads us to introduce the notion of *stochastic process*.

Definition 7.1.1 *A stochastic process* $(X_t)_{t\in I}$ *with* I *interval of* $\mathbb{R}$*, is a family of random numbers with index varying in some interval* I *of* $\mathbb{R}$*.*

Speaking of stochastic processes therefore one refers to continuous index space, where the index is usually interpreted as time. Markov chains, introduced in previous chapter, can be considered as *discrete time stochastic processes*.

We model the flow of incoming clients by a stochastic process N_t, representing the number of clients arrived before time t, which is assumed to be stochastically independent from service times. For fixed t, X_t represents the random number of the clients who are present in the system at time t.

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8_7

In order to characterize a system such as that we have described, one needs to specify:

1. the stochastic process ruling the flow of incoming clients;
2. the distribution of service times;
3. the number of servers.

It is customary to adopt the following notation to indicate the specifications of a given queueing system:

1. M denotes the Poisson process for the flow of incoming clients or exponential distribution for service times;
2. E_r denotes the Erlang distribution with parameter r for the inter-arrival times of clients (that are supposed to be stochastically independent and identically distributed) or for service times. The Erlang distribution with parameter r is the distribution of a sum of r stochastically independent exponential random numbers with the same parameter;
3. D denotes deterministic (non-random) inter-arrival times or service times;
4. G indicates that one does not make any particular hypothesis on the inter-arrival times or service times (that however are always assumed to be stochastically independent).

A process of the type we have described will be indicated by three symbols separated by two slashes. The first symbol refers to the distribution of inter-arrival times, always assumed to be stochastically independent and identically distributed (i.i.d.). The second symbol refers to the distribution of service times. The third symbol indicates the number of servers; it can possibly take the value ∞.

We shall consider three examples of queueing systems and precisely the systems $M/M/1$, $M/M/n$ with $n > 1$ and $M/M/\infty$. Before that we shall speak about continuous time Markov chains with countable state space, and in particular introduce the Poisson process $N_t, t \geq 0$, that for these queueing systems represents the number of clients who entered the system before time t.

7.2 Homogeneous Continuous Time Markov Chains with Countable State Space

An homogeneous continuous time Markov chain is a stochastic process $(X_t)_{t\geq 0}$ with $I(X_t) = \mathbb{N}$ characterized by the initial distribution $(\rho_s)_{s\in\mathbb{N}}$, and for every $t > 0$ a transition matrix $p(t)_{s,s'} = [\Pi]_{ss'}(t)$. As in the case of discrete time case they must respectively satisfy

$$0 \leq \rho_s \leq 1, \quad \sum_{s\in\mathbb{N}} \rho_s = 1,$$
$$0 \leq p(t)_{s,s'} \leq 1, \quad \sum_{s'\in\mathbb{N}} p(t)_{s,s'} = 1,$$

for every $t > 0$. If $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n$, then

$$\mathbf{P}(X_0 = s, X_{t_1} = s_1, \ldots, X_{t_n} = s_n)$$
$$= \rho_{s_0} p_{s_0,s_1}(t_1) p_{s_1,s_2}(t_2 - t_1) \ldots p_{s_{n-1},s_n}(t_n - t_{n-1}).$$

It follows from conditions of compatibility that transition matrices are related by *Chapman-Kolmogorov equations* that can be expressed in synthetic form by:

$$\Pi(t + t') = \Pi(t)\,\Pi(t') \qquad \forall\, t, t' \geq 0$$

or explicitly:

$$p_{s,s'}(t + t') = \sum_{s''} p_{s,s''}(t)\, p_{s'',s'}(t').$$

In order to treat interesting examples, such as those arising from queueing theory, we need to consider the case of strictly denumerable state spaces. In this case $\Pi(t)$ is a matrix with infinitely many rows and columns with non-negative entries, such that the sum of the series of the elements of each row is equal to 1.

The product of two matrices of this kind can be defined according to the usual row times column rule, where the finite sum is replaced by a series. It is easy to check that the result is still a matrix of this kind.

In the case of discrete time the transition probabilities in more steps can be obtained from those in one step. In the case of continuous time analogously transition probabilities in a finite time t can be obtained starting from their behavior as t becomes infinitely small. The simplest case is the Poisson process.

7.3 Poisson Process

A *Poisson process* is a continuous time Markov chain with state space $S = \mathbb{N}$. In the following we shall use a Poisson process as a model for the flow of clients entering a queueing system. For the quantities that we shall consider the order in which clients are served does not matter. A Poisson process $N = (N_t)_{t \geq 0}$ with parameter λ, where $\lambda > 0$, is characterized by the following properties:

1. $p_{s,s}(h) = 1 - \lambda h + o(h)$;
2. $p_{s,s+1}(h) = \lambda h + o(h)$;
3. $p_{s,s'}(h) = o(h) \quad$ for $s' \notin \{s, s+1\}$,

where $o(h)$ is infinitesimal of order larger than h, uniformly in s and s'.

Starting from this hypothesis we can obtain the *Kolmogorov forward equations*, a system of infinitely many differential equations for transition probabilities. Let us fix $\bar{s} = 0$ and the initial distribution $\rho_0 = 1$, $\rho_s = 0$ for $s \neq 0$, i.e. $\mathbf{P}(N_0 = 0) = 1$.

We put:

$$\mu_s(t) \;=\; p_{0,s}(t) \quad \text{for } s \in \mathbb{N}$$

and denote by μ_s' the first derivative of μ_s. The functions μ_s verify the system of equations:

$$\begin{cases} \mu_0'(t) \;=\; -\lambda\mu_0(t) \\ \mu_s'(t) \;=\; -\lambda\mu_s(t) + \lambda\mu_{s-1}(t) \;\; \text{for } s \geq 1, \end{cases} \tag{7.1}$$

as we now show. Consider for $s > 0$ the incremental ratio $\dfrac{\mu_s(t+h) - \mu_s(t)}{h}$ for $h > 0$. We have:

$$\begin{aligned}
\frac{\mu_s(t+h) - \mu_s(t)}{h} &= \frac{p_{0,s}(t+h) - p_{0,s}(t)}{h} \\
&= \frac{\sum_j p_{0,j}(t)p_{j,s}(h) - p_{0,s}(t)}{h} \\
&= \frac{1}{h}((1 - \lambda h + o(h))p_{0,s}(t) + (\lambda h + o(h))p_{0,s-1}(t)) \\
&\quad + \frac{1}{h}\left(\sum_{\substack{j \\ j\neq s, j\neq s-1}} p_{0,j}(t)p_{j,s}(h) - p_{0,s}(t)\right) \\
&= -\lambda p_{0,s}(t) + \lambda p_{0,s-1}(t) + \frac{o(h)}{h} = -\lambda\mu_s(t) + \lambda\mu_{s-1}(t) + \frac{o(h)}{h}\,.
\end{aligned}$$

By taking the limit $h \downarrow 0$, we obtain an equation for the right derivative:

$$\mu_s'(t) \;=\; -\lambda\mu_s(t) + \lambda\mu_{s-1}(t) \quad \text{for } s \geq 1\,,$$

where we have used the notation for the derivative since it is easy to show that it exists. For $s = 0$, we obtain for $h > 0$:

$$\begin{aligned}
\frac{\mu_0(t+h) - \mu_0(t)}{h} &= \frac{p_{0,0}(t+h) - p_{0,0}(t)}{h} \\
&= \frac{\sum_j p_{0,j}(t)p_{j,0}(h) - p_{0,0}(t)}{h} \\
&= \frac{(1 - \lambda h + o(h))p_{0,0}(t) + \sum_{j\neq 0} p_{0,j}(t)p_{j,0}(h) - p_{0,0}(t)}{h} \\
&= -\lambda p_{0,0}(t) + \frac{o(h)}{h} = -\lambda\mu_0(t) + \frac{o(h)}{h}
\end{aligned}$$

that in the limit $h \downarrow 0$ converges to the equation

$$\mu_0'(t) = -\lambda \mu_0(t).$$

As we show below the solution of the system is given by

$$\mu_s(t) = p_{0,s}(t) = \frac{(\lambda t)^s}{s!} e^{-\lambda t},$$

i.e. for each t we have that N_t has Poisson distribution with parameter λt.

If we take $\rho_{\bar{s}} = 1$ and $\rho_s = 0$ for $s \neq \bar{s}$ i.e. assume that $\mathbf{P}(N_0 = \bar{s})$ for some arbitrary state $\bar{s}$, then we obtain the transition probabilities starting from $\bar{s}$:

$$\begin{cases} p_{\bar{s},s}(t) = 0 & \text{for } s < \bar{\bar{s}}, \\ p_{\bar{s},s}(t) = \dfrac{(\lambda t)^{s-\bar{s}}}{(s-\bar{s})!} e^{-\lambda t} & \text{for } s \geq \bar{\bar{s}}. \end{cases} \tag{7.2}$$

Let us prove that (7.2) provides a solution for the system with initial state $\bar{s}$. Let us consider the generating function:

$$\Phi(z, t) = \sum_s p_{\bar{s},s}(t) z^s.$$

We derive $\Phi(z, t)$ with respect to t. It is easy to see that the derivative can be exchanged with the series. By applying the system of equation for $\mu_s(t) = p_{\bar{s},s}(t)$, we obtain

$$\frac{\partial}{\partial t}\Phi(z, t) = \sum_s \mu_s'(t) z^s = -\lambda \sum_{s=0}^{\infty} \mu_s(t) z^s + \lambda \sum_{s=1}^{\infty} \mu_{s-1}(t) z^s = \lambda(z-1)\Phi(z, t).$$

Therefore

$$\frac{1}{\Phi(z, t)} \frac{\partial}{\partial t}\Phi(z, t) = \frac{\partial}{\partial t} \log \Phi(z, t) = \lambda(z-1),$$

so that

$$\log \Phi(z, t) = \lambda(z-1)t + K,$$

that is $\Phi(z, t) = e^K e^{\lambda(z-1)t}$. Since $\mu_{\bar{s}}(0) = 1$ and $\mu_s(0) = 0$ for $s \neq \bar{s}$, we have $\Phi(z, 0) = z^{\bar{s}}$. We have therefore:

$$\Phi(z, t) = z^{\bar{s}} e^{\lambda(z-1)t} = e^{-\lambda t} z^{\bar{s}} e^{\lambda z t} = e^{-\lambda t} \sum_k z^{\bar{s}+k} \frac{(\lambda t)^k}{k!}.$$

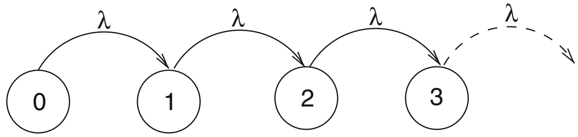


Fig. 7.1 Scheme of Poisson process with parameter λ

It follows that $p_{\bar{s},s}(t) = 0$ for $s < \bar{s}$, $p_{\bar{s},s}(t) = \dfrac{(\lambda t)^{s-\bar{s}}}{(s-\bar{s})!}e^{-\lambda t}$ for $s \geq \bar{s}$. The Poisson process is non-decreasing with probability 1. It can be represented as in Fig. 7.1, when an arrow connecting two states with superscript λ indicates that the transition intensity from one state to the other one is equal to λ. We observe that an arrow enters every state s with $s \geq 1$. These two arrows, one in-coming and one exiting, correspond to two terms, one with plus sign and one with minus sign, on the right-hand side of the differential equation. For $s = 0$ there is just an out-coming arrow, corresponding to the single term, with minus sign, on the right-hand side of the differential equation.

If we indicate with $P_s(t) = P(N_t = s)$ the probability that Poisson process at time t is in the state s, then we have

$$P_s(t) = \sum_{s \in \mathbb{N}} \rho_{\bar{s}}\, p_{\bar{s},s}(t)\,,$$

where ρ_s is the initial distribution. It follows that for every initial distribution the functions $(P_s(t))_{s\in\mathbb{N}}$ satisfy the same system of differential equations.

$$\begin{cases} P_0'(t) = -\lambda P_0(t) \\ P_s'(t) = -\lambda P_s(t) + \lambda P_{s-1}(t) \quad \text{for } s \geq 1. \end{cases}$$

The functions $(p_{\bar{s},s}(t))_{s\in\mathbb{N}}$ can be considered as particular cases in which $\rho_{\bar{s}} = 1$ and $\rho_s = 0$ for $s \neq \bar{s}$.

7.4 Queueing Processes

We now consider some examples of continuous time Markov chains that serve as models of queueing processes. As we have said in Sect. 7.1, in queueing theory there is a symbolic notation to indicate the type of a queueing system. In the examples we consider the flow of incoming clients follows a Poisson process with parameter λ. Clients who find a free server start a service time and after service leave the system. When an arriving client finds all servers engaged, he is put in a queue. When a server becomes free, if there are clients waiting in queue, one of them starts its service time.

For what we are interested in, the order in which clients access the service does not matter; we can assume, for example, that the order is randomly chosen, but other possible choices would not change the results. We assume that service times are stochastically independent, identically distributed and stochastically independent from the Poisson process ruling the flow of arrivals. We also assume that service times are exponentially distributed with some parameter μ.

A process of this type will be indicated with the symbol $M/M/n$. The first M means that the flow of arrivals is Poisson, the second M means that service times are exponentially distributed, while n denotes the number of servers and can vary from 1 to ∞ (∞ is an admissible value).

7.5 $M/M/\infty$ Queueing Systems

We consider an idealized situation in which there are infinitely many servers. The flow of arrivals is ruled by a Poisson process with parameter λ and service times are exponentially distributed with parameter μ.

Let $X = (X_t)_{t\geq 0}$ be the process indicating the number of clients who are in the system at time t. As initial distribution we assume that:

$$\begin{cases} \mathbf{P}(X_0 = 0) = 1\,, \\ \mathbf{P}(X_0 = i) = 0 \quad \text{for } i > 0, \end{cases}$$

i.e. no client is present in the system at time 0. As stated in previous section, service times are stochastically independent between themselves and from the arrivals' process. In order to compute the intensity of service process, we obtain the probability that a client is served in time interval $(t, t+h)$, given that he has not been served up to time t. If T is service time for a client, we have:

$$\begin{aligned} \mathbf{P}(T \leq t+h | T > t) &= \frac{\mathbf{P}(t < T \leq t+h)}{\mathbf{P}(T > t)} \\ &= \frac{e^{-\mu t} - e^{-\mu(t+h)}}{e^{-\mu t}} \\ &= 1 - e^{-\mu h} \\ &= 1 - (1 - \mu h + o(h)) \\ &= \mu h + o(h), \end{aligned}$$

where we have used first order expansion of the exponential $e^{-\mu h} = 1 - \mu h + o(h)$ for small h. Assume that there are n clients in the system. If no one of them has been served up to time t, the probability that at least one of them is served in time interval $(t, t+h)$ is then:

$$\begin{aligned} &1 - \mathbf{P}(T_1 > t+h, \ldots, T_n > t+h | T_1 > t, \ldots, T_n > t) \\ &\quad = 1 - \mathbf{P}(T > t+h | T > t)^n = 1 - e^{-n\mu h} = n\mu h + o(h)\,, \end{aligned}$$

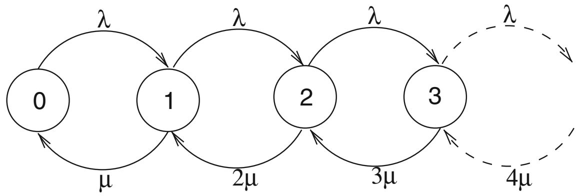


Fig. 7.2 Graphical representation of a $M/M/\infty$ queueing system

where $T_1, \ldots, T_n$ denote the service times of the clients and we have used the fact that they are stochastically independent and identically distributed. Therefore a client exits the system with an intensity which is proportional to the number of clients present in the system. The process can be represented as in Fig. 7.2.

Putting $p_{0,s}(t) = \mu_s(t)$, we can write forward Kolmogorov equations by using the rule described in Sect. 7.3:

$$\begin{cases} \mu_0'(t) = \mu\mu_1(t) - \lambda\mu_0(t) \\ \mu_i'(t) = -(\lambda + i\mu)\,\mu_i(t) + \lambda\,\mu_{i-1}(t) + (i+1)\mu\,\mu_{i+1}(t) \end{cases}$$

for $i \geq 1$, where $\mu_i'(t)$ denotes the derivative of μ_i.

We have seen that n-steps transition probabilities of discrete time Markov chains satisfying the hypothesis of ergodic theorem converge as $n \to \infty$ to the stationary distribution. Analogous results hold for continuous time Markov chains. Therefore we look for a stationary solution $(p_i)_{i\geq 0}$ of the system of equations, that is a solution which does *not* depend on the time. We impose $\mu_i'(t) = 0$ so that $\mu_i(t) = p_i$ and obtain:

$$\begin{cases} 0 = \mu p_1 - \lambda p_0 \\ 0 = -(\lambda + i\mu)p_i + \lambda p_{i-1} + (i+1)\mu p_{i+1}, \quad \text{for } i \geq 1 \\ \displaystyle\sum_{i=0}^{+\infty} p_i = 1\,. \end{cases}$$

By adding up the equations up to the i-th one, we obtain the recursive formula:

$$p_i = \frac{\lambda}{i\mu} p_{i-1} = \frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i p_0\,.$$

By imposing the condition $\displaystyle\sum_{i=0}^{+\infty} \frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i p_0 = 1$, we obtain:

$$p_0 \sum_{i=0}^{+\infty} \frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i = 1.$$

Since $\sum_{i=0}^{+\infty} \frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i = e^{\frac{\lambda}{\mu}}$, therefore

$$p_i = e^{-\frac{\lambda}{\mu}} \quad \text{and} \quad p_i = \frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i e^{-\frac{\lambda}{\mu}},$$

which is Poisson distribution with parameter $\frac{\lambda}{\mu}$. We come to the conclusion that for $M/M/\infty$ the queueing system stationary distribution exists for all values of λ and μ.

7.6 $M/M/1$ Queueing Systems

Also for $M/M/1$ service times are assumed to be stochastically independent and identically distributed with exponential distribution with parameter μ. The arrival flow of clients is ruled by a Poisson process with parameter λ which is stochastically independent from service times.

For this system there is just one server. Therefore the intensity for a client to exit the system is equal to μ independently from the number of clients present in the system. $M/M/1$ queueing system can be graphically represented as shown in Fig. 7.3.

The system of differential equations for the function $\mu_s(t) = p_{\bar{s},s}(t)$, where $\bar{s}$ is some fixed state, is then:

$$\begin{cases} \mu_0'(t) = \mu\mu_1(t) - \lambda\mu_0(t) \\ \mu_1'(t) = -(\lambda+\mu)\,\mu_1(t)\,. \end{cases}$$

Also in this case we look for a *stationary solution*, i.e. such that $\mu_i'(t) = 0$ for $i \in \mathbb{N}$ with $\mu_i(t) = p_i$, where p_i is a probability distribution. We obtain then the system of linear equations:

$$\begin{cases} 0 = \mu p_1 - \lambda p_0 \\ 0 = -(\lambda+\mu)p_i + \lambda p_{i-1} + \mu p_{i+1} \\ \sum_{i=0}^{+\infty} p_i = 1\,. \end{cases}$$

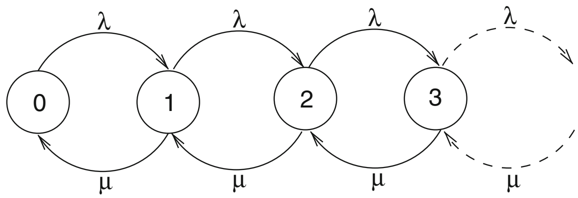


Fig. 7.3 Scheme of $M/M/1$ queueing system

From this system we obtain, by adding up the first n equations, the recursive relation

$$p_n = \frac{\lambda}{\mu} p_{n-1} = \left(\frac{\lambda}{\mu}\right)^n p_0 .$$

By imposing the condition $\sum_{i=0}^{+\infty} p_i = 1$, we obtain

$$\left(\sum_{i=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^i \right) p_0 = 1.$$

This series is convergent if $\dfrac{\lambda}{\mu} < 1$. In this case we get

$$p_0 = 1 - \frac{\lambda}{\mu}.$$

The stationary probability distribution is then

$$p_i = \left(\frac{\lambda}{\mu}\right)^i \left(1 - \frac{\lambda}{\mu}\right), \qquad \text{for } i = \frac{\lambda}{\mu}.$$

The stationary probability distribution is a shifted geometric distribution with parameter $\frac{\lambda}{\mu}$ (the set of possible values is $\mathbb{N}$ instead of $\mathbb{N} \setminus \{0\}$). It exists if and only if $\frac{\lambda}{\mu} < 1$ or $\lambda < \mu$, i.e. if the intensity of arrivals of clients is strictly less than the parameter of the exponential distribution of service times.

7.7 $M/M/n$ Queueing Systems

We finally consider $M/M/n$ queueing systems with $n \geq 2$, i.e. with a finite number of servers larger than 1. From considerations similar to those developed for the other cases we obtain the following system of equations for transition probabilities $\mu_s'(t) = p_{\bar{s},s}(t)$, where $\bar{s}$ is some fixed state:

$$\begin{cases} \mu_0'(t) = \mu\mu_1(t) - \lambda\mu_0(t) \\ \mu_1'(t) = -(\lambda + \mu)\,\mu_1(t) + \lambda\,\mu_0(t) + 2\mu\,\mu_2(t) \\ \cdots \\ \mu_{n-1}'(t) = -(\lambda + (n-1)\mu)\,\mu_{n-1}(t) + \lambda\,\mu_{n-1}(t) + n\mu\,\mu_n(t) \\ \mu_n'(t) = -(\lambda + n\mu)\,\mu_n(t) + \lambda\,\mu_{n-1}(t) + n\mu\,\mu_{n+1}(t) \\ \mu_{n+1}'(t) = -(\lambda + n\mu)\,\mu_{n+1}(t) + \lambda\,\mu_n(t) + n\mu\,\mu_{n+2}(t) \\ \cdots, \end{cases}$$

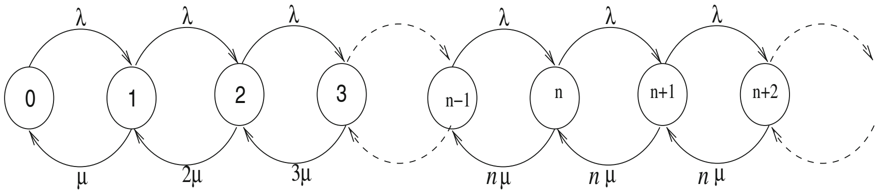


Fig. 7.4 Scheme of a $M/M/n$ queueing system with initial state in 0

where λ and μ are, as in previous cases, respectively the parameter of the Poisson process ruling the arrival of clients and of the exponential distribution of service times. The system is graphically represented in Fig. 7.4.

Let us now look for the stationary distribution by imposing $\mu_i'(t) = 0$ for all $i \in \mathbb{N}$. If we denote $p_i \equiv \mu_i(t)$, we obtain the following system of linear equations:

$$\begin{cases} 0 = \mu p_1 - \lambda p_0 \\ 0 = 2\mu p_2 - \lambda p_1 \\ \cdots \\ 0 = (n-1)\mu p_{n-1} - \lambda p_{n-2} \\ 0 = n\mu p_n - \lambda p_{n-1} \\ 0 = n\mu p_{n+1} - \lambda p_n \\ \cdots \\ \sum_{i=0}^{+\infty} p_i = 1\,. \end{cases}$$

We obtain the following recursive equations:

$$p_i = \frac{\lambda}{i\mu} p_{i-1} \quad \text{for } i = 1, \ldots, n;$$
$$p_i = \frac{\lambda}{n\mu} p_{i-1} \quad \text{for } i \geq n+1.$$

Therefore we have:

$$p_i = \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!}\, p_0 \quad \text{for } i = 0, \ldots, n,$$
$$p_i = \left(\frac{\lambda}{\mu}\right)^i \frac{1}{n!n^{i-n}}\, p_0 \quad \text{for } i \geq n+1.$$

A solution of the system exists if

$$\sum_{i=0}^{n-1} \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!} + \sum_{i=n}^{\infty} \left(\frac{\lambda}{\mu}\right)^i \frac{1}{n!n^{i-n}} < +\infty.$$

The first term on the left-hand side is a finite sum. The series of the second term can be rewritten by putting $j = i - n$ as

$$\frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n \sum_{j=0}^{\infty}\left(\frac{\lambda}{n\mu}\right)^j .$$

The condition of convergence is therefore $\frac{\lambda}{n\mu} < 1$, i.e. $\lambda < n\mu$. This result answers the problem of how many servers are needed for a queueing system with some fixed Poisson flow of incoming clients so that the queue stabilizes (so that a stationary distribution exists). For $\lambda < n\mu$ we have:

$$p_0 = \left(\sum_{i=0}^{n-1}\left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!} + \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n \frac{1}{1-\frac{\lambda}{n\mu}}\right)^{-1} \tag{7.3}$$

$$p_i = \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!}\, p_0 \qquad \text{for } i = 1, \ldots, n, \tag{7.4}$$

$$p_i = \left(\frac{\lambda}{\mu}\right)^i \frac{1}{n!n^{i-n}}\, p_0 \qquad \text{for } i \geq n+1. \tag{7.5}$$

7.8 Queueing Systems in Stationary Regime and Little's Formulas

For Markov queueing systems introduced in the previous sections the existence of an invariant distribution allows us to consider a stationary regime for the process X representing the number of clients present in the system. In the stationary regime probabilistic characteristics of the process don't vary in time. The stationary regime is obtained by taking as initial distribution the stationary distribution.

It can be shown that, when a stationary distribution exists, these queueing systems evolve towards stationary regime and moreover temporal averages of observables tend, as the length of the temporal interval tends to infinity, to the expectations of the observables computed in stationary regime. All this should be precisely stated and supported with proofs. We limit ourselves to accept it and to reason at intuitive level. We now consider some quantities or observables, which are relevant for the study of queueing systems and their efficiency, and establish some useful relations. From now on we shall always refer to queueing systems in stationary regime.

In order to evaluate the efficiency of a queueing system, we introduce the *utilization factor* ρ. This quantity is defined as the client's average arrival rate λ times the average service time $\bar{T}$ divided by the number m of servers. It can be shown that the utilization factor is equal to the average percentage rate of utilization of servers. For a non-deterministic system in stationary regime it is known that $\rho < 1$, see also [12], i.e. that with probability one servers do not work full time. A server will be free for a positive percentage of time. Other interesting quantities are:

1. the average number L of clients present in the system;
2. the average number L_q of clients waiting in queues;
3. the average time W that a client spends in the system;
4. the average time W_q that a client spends waiting in queues.

The last two quantities are related by the equation

$$W = W_q + \bar{T},$$

where $\bar{T}$ is the equation of service time.

Let us assume that every client pays an amount equal to the time he spends in the system. In a time interval of length t the expectation of the amount paid by clients is given, apart from quantities of order smaller than t, by λt (expectation of the number of clients entering the system in a time interval of length t) times W (expectation of the time a client spends in the system). Alternatively the same quantity is given by Lt. By equating the expressions and letting t tend to infinity, we get first *Little's formula* $L = \lambda W$.

Analogously if we assume that a client pays an amount equal to the time he spends in queue, we get the second *Little's formula* $L_q = \lambda W_q$.

Little's formulas apply to a large class of queueing systems in stationary regime.

Let us consider the case of $M/M/1$ queueing system. As we have seen, this system has an invariant distribution if and only if $\lambda < \mu$, where λ is the parameter of Poisson process of incoming clients and μ is the parameter of the exponential distribution of service time. In this case the stationary distribution for the number of clients present in the system is given by:

$$\rho_k = \left(\frac{\lambda}{\mu}\right)^k \left(1 - \frac{\lambda}{\mu}\right).$$

We have therefore

$$L = \sum_{k=1}^{\infty} k\rho_k = \sum_{k=1}^{\infty} k \left(\frac{\lambda}{\mu}\right)^k \left(1 - \frac{\lambda}{\mu}\right) = \frac{\lambda}{\mu - \lambda}$$

and

$$L_q = \sum_{k=2}^{\infty} (k-1)\rho_k = \sum_{k=1}^{\infty} (k-1) \left(\frac{\lambda}{\mu}\right)^k \left(1 - \frac{\lambda}{\mu}\right) = \frac{\lambda^2}{\mu(\mu - \lambda)}.$$

Therefore by using Little's formulas we have:

$$W = \frac{1}{\mu - \lambda}, \quad W_q = \frac{\lambda}{\mu(\mu - \lambda)},$$

that satisfy the equation $W = W_c + \bar{T}$, where $\bar{T} = \dfrac{1}{\mu}$ (expectation of exponential distribution with parameter μ).

In this case the utilization factor is $\dfrac{\lambda}{\mu}$. We observe that, as ρ tends to 1, the average number of clients present in the system and waiting in queue, as well as the average time spent by a client in the system, all tend to infinity. This is a general characteristics of random queueing systems. If one tries to increase utilization factor, one has to pay the price of an increase of the number of clients in queue and of their typical waiting times. Value 1 for the utilization factor is not reachable by a random queueing system in stationary regime, but it can be obtained by a deterministic system with one server where clients arrive at regular time intervals equal to the service time.

Chapter 8
Statistics

We now introduce some basic notions in Bayesian statistics. For further reading, we refer to [5, 9, 10].

8.1 Bayesian Statistics

Assume that we know the value x_i of some characteristics, for example the height for every individual i of a population $i = 1, \ldots, N$. We can then build up a cumulative distribution function $F(x)$ defined by

$$F(x) = \frac{\sharp\{i|\ x_i \leq x\}}{N}.$$

$F(x)$ can be interpreted as the c.d.f. of a random number X, where X is the height of an individual randomly chosen from the population (every individual is chosen with equal probability $\frac{1}{N}$). Some relevant quantities can be extracted from $F(x)$, such as the expectation, the variance, the median and others.

$F(x)$ (called *empirical c.d.f.*) will always be of discrete type, but for large N it is possible that it is well approximated by an absolutely continuous c.d.f. Similarly for two quantities x_i, y_i for example height and weight relative to each individual, we can obtain the joint c.d.f. $F(x, y)$ defined by

$$F(x, y) = \frac{\sharp\{i|\ x_i \leq x, y_i \leq y\}}{N}.$$

$F(x, y)$ is the joint c.d.f. of the random vector (X, Y), where X and Y are respectively the height and the weight of a randomly chosen individual in the population. Also in this case relevant indices such as covariance, correlation coefficient, etc. can be

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8_8

extracted from $F(x, y)$. The study of empirical c.d.f.'s is part of *descriptive statistics* and is obviously related to the study of probability distributions.

Often the data about the entire population we are interested in are not available. In this case one tries to form an evaluation of the distributions of quantities in the whole population starting from results obtained by sampling (that is by randomly extracting a subset of individuals of the population). These methods are part of what is called *statistical inference* or *statistical induction*, in the Bayesian approach, that we shall follow in this chapter. They are an application of Bayes' Formula and therefore are part of Probability Theory. We deal here just with a few relevant examples in which a model based on some distribution is assumed to be fixed and one makes inference on one or a certain number of unknown parameters, that in Bayesian approach are treated as *random numbers*.

8.2 Conditional Density for Two Random Numbers

We now introduce the *conditional density* of a random number Y given another random number X. Let $f(x, y)$ be the joint probability density function of (X, Y) and f_X, f_Y the probability density functions of X, Y, respectively. The conditional probability of the event $(a \le Y \le b)$ given $(x - h \le X \le x + h)$ is then given by

$$P(a \le Y \le b|\, x - h < X < x + h) = \frac{\int_{x-h}^{x+h} \int_a^b f(s, t) ds dt}{P(x - h < X < x + h)}.$$

In order to give a meaning to the conditional probability given $(X = x)$, we let h tend to 0. Assume that $f(x, y)$ satisfies the following conditions:

1. $f(x, y)$ is continuous;
2. $f_X(x)$ is continuous.

Then it is easy to see that if $f_X(x) > 0$

$$\lim_{h \to 0} P(a \le Y \le b|\, x - h < X < x + h) = \int_a^b \frac{f(x, t)}{f_X(x)} dt.$$

Previous argument justifies the definition of conditional density $f_X(y|x)$ of Y given $X = x$ under the condition $f_X(x) > 0$ as given by

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_{X(x)}}$$

We obtain then *Bayes' formula for densities*. From

$$f(x, y) = f_{Y|X}(y|x)\, f_X(x)$$

and

$$f(x, y) = f_{X|Y}(x|y)\, f_Y(y),$$

we get

$$f_{Y|X}(y|x) \;=\; f_Y(y)\, \frac{f_{X|Y}(x|y)}{f_X(x)}.$$

These formulas generalize to the n-dimensional case. Let $X_1, \dots, X_n$ be random numbers with joint probability density $f(x_1, \dots, x_n)$. Let $\{i_1, \dots, i_k\}$ be a proper subset of $\{1, \dots, n\}$ and assume that the marginal density function $f_{i_1,\dots,i_k}(x_{i_1}, \dots, x_{i_k})$ of $X_{i_1}, \dots, X_{i_k}$ is strictly positive at the point $(x_{i_1}, \dots, x_{i_k})$. Let $\{j_1, \dots, j_{n-k}\} = \{1, \dots, n\} \setminus \{i_1, \dots, i_k\}$. Then the conditional density of $X_{j_1}, \dots, X_{n-k}$ given $(X_{i_1} = x_{i_1}, \dots, X_{i_k} = x_{i_k})$, provided that $f_{i_1,\dots,x_{i_k}}(x_{i_1}, \dots, x_{i_k}) > 0$, is defined by

$$\begin{aligned} &f_{j_1,\dots,j_{n-k}|i_1,\dots,i_k}(x_{j_1}, \dots, x_{j_{n-k}}|x_{i_1}, \dots, x_{i_k}) \\ &\quad= \frac{f(x_1, \dots, x_n)}{f_{i_1,\dots,i_k}(x_{i_1}, \dots, x_{i_k})}. \end{aligned}$$

As in the two-dimensional case we get Bayes' formula for densities

$$\begin{aligned} &f_{j_1,\dots,j_{n-k}|i_1,\dots,i_k}(x_{j_1}, \dots, x_{j_{n-k}}|x_{i_1}, \dots, x_{i_k}) \\ &\quad= \frac{f_{i_1,\dots,i_k|j_1,\dots,j_{n-t}}(x_{i_1}, \dots, x_{i_k}|x_{j_1}, \dots, x_{j_{n-k}})\, f_{j_1,\dots,j_{n-k}}(x_{j_1}, \dots, x_{j_{n-k}})}{f_{i_1,\dots,i_k}(x_{i_1}, \dots, x_{i_k})} \end{aligned}$$

This formula is applied to statistical inference in the Bayesian approach that will be treated in following sections.

8.3 Statistical Induction on Bernoulli Distribution

Let us consider a sequence of events $(E_i)_{i=1,2,\dots}$ stochastically independent conditionally on the knowledge of a parameter Θ such that

$$\mathbf{P}(E_i = 1|\Theta = \theta) = \theta$$

where $0 < \theta < 1$.

The events E_i can be thought of as the result of experiments; their stochastic independence conditionally on the knowledge of the value of Θ means that

$$\mathbf{P}(E_1 = \epsilon_1, \dots, E_n = \epsilon_n|\Theta = \theta) \;=\; \prod_{i=1}^{n} \mathbf{P}(E_i = \epsilon_i|\Theta = \theta)$$

for any $\epsilon_i \in \{0, 1\}$ for $i = 1, \dots, n$.

Let Θ have an a priori *probability density*. We want to find out how the distribution of Θ changes after n experiments are performed. Assume that the results are $E_1 = \epsilon_1, \dots, E_n = \epsilon_n$. The conditional density of Θ given $E_1 = \epsilon_1, \dots, E_n = \epsilon_n$ is denoted by

$$\pi_n(\theta | E_1 = \epsilon_1, \dots, E_n = \epsilon_n)$$

and it is called a posteriori *density*. By the composite probability law we have, given $0 \le a < b \le 1$ that

$$\mathbf{P}(\theta \in [a, b] | E_1 = \epsilon_1, \dots, E_n = \epsilon_n) = \frac{\mathbf{P}(\theta \in [a, b], E_1 = \epsilon_1, \dots, E_n = \epsilon_n)}{\mathbf{P}(E_1 = \epsilon_1, \dots, E_n = \epsilon_n)}. \tag{8.1}$$

By using the formula of total probabilities, that can be easily extended to this continuous case, and the conditional independence of $E_1, \dots, E_n$ given $\Theta = \theta$ we can rewrite the right-hand side of (8.1) as

$$\frac{\int_a^b \theta^{\epsilon_1+\dots+\epsilon_n} (1-\theta)^{n-(\epsilon_1+\dots+\epsilon_n)} \pi_0(\theta) d\theta}{\int_0^1 \theta^{\epsilon_1+\dots+\epsilon_n} (1-\theta)^{n-(\epsilon_1+\dots+\epsilon_n)} \pi_0(\theta) d\theta}.$$

Therefore we have

$$\begin{aligned} &\pi_n(\theta | E_1 = \epsilon_1, \dots, E_n = \epsilon_n) \\ &\quad = \frac{1}{c} \pi_0(\theta) \theta^{\epsilon_1+\dots+\epsilon_n} (1-\theta)^{n-\epsilon_1-\dots-\epsilon_n} \end{aligned}$$

for $0 \le \theta \le 1$ where

$$c = \mathbf{P}(E_1 = \epsilon_1, \dots, E_n = \epsilon_n) = \int_0^1 \theta^{\epsilon_1+\dots+\epsilon_n} (1-\theta)^{n-\epsilon_1-\dots-\epsilon_n} \pi_0(\theta) \, d(\theta).$$

In particular, if a priori distribution of Θ is beta $\mathcal{B}(\alpha, \beta)$ with parameters α and β, the a posteriori distribution will also be beta $\mathcal{B}(\alpha', \beta')$ with parameters

$$\alpha' = \alpha + \sum_{i=1}^n \epsilon_i \quad \text{and} \quad \beta' = \beta + n - \sum_{i=1}^n \epsilon_i$$

where $\sum_{i=1}^n \epsilon_i$ and $n - \sum_{i=1}^n \epsilon_i$ are respectively the number of events that have and have not taken place. Therefore

$$\pi_n(\theta | E_1 = \epsilon_1, \dots, E_n = \epsilon_n) = \begin{cases} \frac{\Gamma(\alpha'+\beta')}{\Gamma(\alpha')\Gamma(\beta')} \theta^{\alpha'-1} (1-\theta)^{\beta'-1} & \theta \in [0, 1], \\ 0 & \text{otherwise.} \end{cases}$$

8.4 Statistical Induction on Expectation of Normal Distribution

Let $(X_i)_{i=1,2,\ldots}$ be a sequence of random numbers that are stochastically independent given the knowledge of a parameter Θ with conditional probability density

$$f(x|\theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right)$$

for some $\sigma > 0$.

By using Bayes' formula for densities on $X_1, \ldots, X_n, \Theta$ we get an expression for the a posteriori density of Θ, i.e. the conditional density given $X_1 = x_1, \ldots, X_n = x_n$:

$$\begin{aligned}\pi_n(\theta|x_1, \ldots, x_n) &= \frac{\pi_0(\theta) \prod_{i=1}^n f(x_i|\theta)}{p_n(x_1, \ldots, x_n)} \\ &= K\, \pi_0(\theta) \prod_{i=1}^n f(x_i|\theta),\end{aligned}$$

where $p_n(x_1, \ldots, x_n)$ is the marginal density of $X_1, \ldots, X_n$ and we have denoted by a constant K the quantity $p_n(x_1, \ldots, x_n)^{-1}$, since it does not depend on θ and can therefore thought of as a normalizing constant for the probability density $\pi_n(\theta|x_1, \ldots, x_n)$. In the future we shall denote any normalization constant by K, even if its value changes from one formula to the other, in order not to introduce too many constants.

If the a priori distribution of Θ is Gaussian $N(\mu_0, \sigma_0^2)$, we obtain

$$\begin{aligned}\pi_n(\theta|x_1, \ldots, x_n) &:= \\ &= K\pi_0(\theta) \prod_{i=1}^n f(x_i|\theta) \\ &= K e^{-\dfrac{(\theta-\mu_0)^2}{2\sigma_0^2}} \exp\left(-\sum_{i=1}^n \frac{(x_i-\theta)^2}{2\sigma^2}\right) \\ &= K \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)\theta^2 - 2\theta\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^n x_i}{\sigma^2}\right)\right]\right\} \\ &= K \exp\left\{-\frac{1}{2}\frac{(\theta-m_n)^2}{\sigma_n^2}\right\},\end{aligned}$$

where

$$m_n = \frac{\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^n x_i}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}, \quad \sigma_n^2 = \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}$$

and K is the normalizing constant. If $\bar{x}$ denotes the *sample average* $\bar{x} = \frac{x_1+\cdots+x_n}{n}$, the a posteriori distribution of Θ is Gaussian

$$N\left(\frac{\mu_0\sigma_0^{-2} + \bar{x}n\sigma^{-2}}{\sigma_0^{-2} + n\sigma^{-2}}, \frac{1}{\sigma_0^{-2} + n\sigma^{-2}}\right).$$

The expectation can be thought of as a weighted average of μ_0 and $\bar{x}$ with weights σ_0^{-2} and $n\sigma^{-2}$.

8.5 Statistical Induction on Variance of Normal Distribution

We consider now statistical induction on the variance of normal distribution. It is convenient to use as parameter the inverse of the variance, called *precision*; it is clear that precision carries the same amount of information as the variance. The term precision is related to the interpretation of random numbers as measurements of some quantity. Let $(X_n)_{n=1,2,\ldots}$ be a sequence of random numbers stochastically independent conditionally on the knowledge of the value of the parameter Φ.

Assume that the conditional probability density of each of the X_i, given that $(\Phi = \phi)$, is equal to

$$f(x|\phi) = f(x_i|\phi) = \frac{\phi^{\frac{1}{2}}}{\sqrt{2\pi}} \exp\left(-\frac{\phi}{2}(x-\mu)^2\right),$$

where μ is some constant. The conditional density of $X_1, \ldots, X_n$ given $(\Phi = \phi)$ called the *likelihood factor* is given by

$$\prod_{i=1}^n f(x_i|\phi) = K\phi^{\frac{n}{2}} \exp\left(-\frac{\phi}{2}\sum_{i=1}^n (x_i-\mu)^2\right)$$
$$= K\phi^{\frac{n}{2}} \exp\left(-\frac{nS^2\phi}{2}\right),$$

where

$$S^2 := \frac{\sum_{i=1}^n (x_i-\mu)^2}{n}$$

is the *average of the squares of the deviations* of the x_i's from μ. If we assume that the a priori distribution of Φ is $\Gamma(\alpha_0, \lambda_0)$, then the a posteriori density of Φ, given that $X_1 = x_1, \ldots, X_n = x_n$, is given by:

$$\pi_n(\phi|x_1, \ldots, x_n) = K\phi^{\frac{n}{2}+\alpha_0-1} \exp\left(-\phi(\lambda_0 + \frac{nS^2}{2})\right)$$

for $\phi > 0$ and 0 otherwise. That is the a posteriori distribution of Φ is gamma $\Gamma(\alpha_0 + \frac{n}{2}, \lambda_0 + \frac{nS^2}{2})$.

8.6 Improper Distributions

Let us go back to the induction on the expectation of normal distribution. We want to describe a vague initial state of information. This can be achieved by choosing an a priori distribution with large variance. We can let the variance tend to infinity. In the limit we do not get a probability distribution.

Nonetheless we observe that the corresponding a posteriori distributions converge. The limiting a posteriori distribution can be alternatively obtained by introducing as a priori distribution the so called *uniform improper distribution* density $\pi_0(\theta) = K$. This π_0 does not correspond to a probability distribution, but it must be interpreted in terms of the limiting procedure we have just described.

8.7 Statistical Induction on Expectation and Variance of Normal Distribution

Let us now consider the case of statistical induction on both expectation and variance of a normal distribution. Assume that we are in a state of vague information, that, as we have said, can be described by means of an improper distribution. We have now two unknown parameters Θ and Φ, respectively the expectation and the precision, that is the inverse of the variance. Since Φ can take only positive values, we consider as a priori distribution an improper uniform distribution for Θ and $\log \Phi$. This corresponds to the improper density:

$$\pi_0(\theta, \phi) = K\phi^{-1}, \quad (\phi > 0).$$

Assume that we have a sequence of random numbers that are stochastically independent conditionally on the event that Θ and Φ take some definite values θ and ϕ and that their conditional density is:

$$f(x|\theta, \phi) = \frac{1}{\sqrt{2\pi}}\phi^{\frac{1}{2}} \exp\left(-\frac{\phi}{2}(x-\theta)^2\right).$$

The conditional joint density of $X_1, \dots, X_n$, given $\Theta = \theta, \Phi = \phi$, which is called likelihood factor, is then

$$
\begin{aligned}
f(x_1, \dots, x_n|\theta, \phi) &= K\phi^{\frac{n}{2}} \exp\left(-\frac{\phi}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right) \\
&= K\phi^{\frac{n}{2}} \exp\left(-\frac{\phi}{2}\left((\bar{x} - \theta)^2 + \nu s^2\right)\right),
\end{aligned}
$$

where $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, $\nu = n - 1$, $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\nu}$. The joint a posteriori density of Θ, Φ is obtained by Bayes' formula for densities and is given by:

$$
\pi_n(\theta, \phi|x_1, \dots, x_n) = K\phi^{\frac{n}{2}-1} \exp\left(-\frac{\phi}{2}\left((\bar{x} - \theta)^2 + \nu s^2\right)\right).
$$

From joint a posteriori probability density of Θ and Φ we can get their marginal densities by integrating with respect to the other variable. The integral with respect to ϕ reduces to the integral of the gamma function. After collecting in the constant K all factors that do not depend on θ, we obtain

$$
\pi_n(\theta|x_1, \dots, x_n) = \int_0^{+\infty} \pi_n(\theta, \phi|x_1, \dots, x_n)d\phi = \frac{K}{\left((\bar{x} - \theta)^2 + \nu s^2\right)}.
$$

From this it follows that the random number $T = \frac{\bar{x} - \theta}{s\sqrt{\nu}}$ has Student t density with ν degrees of freedom

$$
f_T(t) = K\left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}.
$$

Analogously we obtain the a posteriori density of Φ by integrating the conditional density $\pi_n(\theta, \phi|x_1, \dots, x_n)$ with respect to θ. It is a Gaussian integral that, apart from constant factors, gives a factor $\phi^{-\frac{1}{2}}$. The a posteriori marginal probability density of Φ is

$$
\pi_n(\phi|x_1, \dots, x_n) = \int_0^{+\infty} \pi_n(\theta, \phi|x_1, \dots, x_n)d\theta = K\phi^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu s^2 \phi}{2}\right),
$$

with $\phi > 0$. By making a linear change of variable we see that the random number $\nu s^2 \Phi$ has a posteriori distribution with density

$$
Ku^{\frac{\nu}{2}-1} \exp\left(-\frac{u}{2}\right),
$$

with $u > 0$, i.e. is χ^2-distribution with ν degrees of freedom. The normalizing constant is therefore given by $K = \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}$.

8.8 Bayesian Confidence Intervals and Hypotheses' Testing

A synthetic description of a posteriori distribution can be achieved by means of confidence intervals or, in the multidimensional case, confidence regions. Given $0 < \alpha < 1$, an α-*level confidence interval* or *confidence region* is an interval or respectively a region whose a posteriori probability is $1 - \alpha$. The choice of an interval or a region with this property is clearly arbitrary. In concrete situations one can base the choice on symmetry criteria if the a posteriori density is symmetric or alternatively one can choose the region with minimal volume in parameters' space.

In the Bayesian approach to statistics, hypotheses' testing can be related to the definitions of confidence intervals or regions. The hypotheses that parameters have a given value is rejected if the value does not belong to the confidence interval or region. This procedure, it must be stressed, is arbitrary, since, as we have said, the interval or region can be arbitrarily chosen. Nevertheless, since in many situations there is a preferential choice, the use of hypotheses' testing in Bayesian approach can be accepted as a shortened and less precise form of induction with respect to the complete analysis based on a posteriori distribution.

8.9 Comparison of Expectations for Normal Distribution

Assume that we have two samples of size respectively n_1 and n_2 that, conditionally on the knowledge that the parameters Θ_1 and Θ_2 are equal respectively to θ_1 and θ_2, are stochastically independent samples with Gaussian distribution $N(\theta_1, \sigma_1^2)$ and $N(\theta_2, \sigma_2^2)$ respectively. If the a priori density of Θ_1 and Θ_2 is uniform improper, Θ_1 and Θ_2 are stochastically independent a posteriori with Gaussian distribution $N(\bar{x}_1, \frac{\sigma_1^2}{n})$ and $N(\bar{x}_2, \frac{\sigma_2^2}{n})$ respectively, where $\bar{x}_1, \bar{x}_2$ are the sample averages of the samples.

Indeed, since the samples are stochastically independent and Θ_1 and Θ_2 are stochastically independent in the a priori distribution, we can separately apply to the samples the results on the induction on the expectation of normal distribution in the case of uniform improper a priori distribution. If we define $\Theta = \Theta_2 - \Theta_1$, then the a posteriori distribution of Θ is in $N(\bar{x}_2 - \bar{x}_1, \dfrac{\sigma_2^2}{n_2} + \dfrac{\sigma_1^2}{n_1})$.

Let us now consider the case when there is an extra parameter Φ such that conditionally on the knowledge that $\Phi = \phi$ and $\Theta_1 = \theta_1$, $\Theta_2 = \theta_2$, the two samples are stochastically independent with distributions respectively $N(\theta_1, \phi^{-1})$ and $N(\theta_2, \phi^{-1})$. The conditional probability densities of the random numbers of the first and the second sample are then respectively

$$f_1(x|\theta_1,\theta_2,\phi) = \frac{1}{\sqrt{2\pi}}\phi^{\frac{1}{2}}\exp\left(-\frac{\phi}{2}(x-\theta_1)^2\right),$$

$$f_2(x|\theta_1,\theta_2,\phi) = \frac{1}{\sqrt{2\pi}}\phi^{\frac{1}{2}}\exp\left(-\frac{\phi}{2}(x-\theta_2)^2\right).$$

Also here we consider the case of improper a priori distribution and precisely we assume that Θ_1, Θ_2, $\log\Phi$ are stochastically independent with uniform improper distribution on $\mathbb{R}$. This corresponds for Θ_1, Θ_2, Φ to an a priori improper density

$$\pi_0(\theta_1,\theta_2,\phi) = K\phi^{-1} \quad (\phi > 0).$$

Consider first statistical induction for Φ. Here we can apply without any essential change what we have seen about the induction for normal distributions with two unknown parameters and obtain that the a posteriori density of Φ is given by:

$$K\phi^{\frac{\nu_1+\nu_2}{2}-1}\exp\left(-\frac{s^2\phi}{2}\right),$$

where $s^2 = \nu_1 s_1^2 + \nu_2 s_1^2$ with $\nu_i = n_i - 1$, $i = 1, 2$,

$$s_i^2 = \frac{\sum_{j=1}^{n_i}(x_{i,j}-\bar{x}_i)^2}{v_i},$$

and $x_{i,j}$ is the j-th value of the i-th sample. By combining these results we can obtain the a posteriori probability density of $\Theta = \Theta_2 - \Theta_1$ in the case when Φ is unknown. Indeed we have:

$$\begin{aligned}
\pi(\theta|x_1,x_2) &= K\int_{\mathbb{R}^+}\phi^{\frac{1}{2}}\exp\left(-\frac{\phi}{2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}(\theta-(\bar{x}_2-\bar{x}_1))^2\right)\phi^{\frac{\nu_1+\nu_2}{2}-1}\exp\left(-\frac{s^2\phi}{2}\right)d\phi \\
&= K2^{\frac{\nu_1+\nu_2+1}{2}}\left[\frac{(\theta-(\bar{x}_2-\bar{x}_1))^2}{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}+s^2\right]^{-\frac{\nu_1+\nu_2+1}{2}}\int_{\mathbb{R}^+}y^{\frac{\nu_1+\nu_2+1}{2}-1}e^{-y}dy \\
&= K2^{\frac{\nu_1+\nu_2+1}{2}}\Gamma(\frac{\nu_1+\nu_2+1}{2})\left[\frac{(\theta-(\bar{x}_2-\bar{x}_1))^2}{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}+s^2\right]^{-\frac{\nu_1+\nu_2+1}{2}}, \\
&= K(2/s^2)^{\frac{\nu_1+\nu_2+1}{2}}\Gamma(\frac{\nu_1+\nu_2+1}{2})\left[\frac{(\theta-(\bar{x}_2-\bar{x}_1))^2}{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}+1\right]^{-\frac{\nu_1+\nu_2+1}{2}}
\end{aligned}$$

where we have used the change of variable $y = \frac{1}{2}\left[\frac{(\theta-(\bar{x}_2-\bar{x}_1))^2}{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} + s^2\right]\phi$ to express the integral in terms of a Gamma function. We obtain

$$\pi(\theta|\bar{x}_1, \bar{x}_2) = K\left[\frac{(\theta-(\bar{x}_2-\bar{x}_1))^2}{\nu s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} + 1\right]^{-\frac{\nu+1}{2}},$$

where $\nu = \nu_1 + \nu_2$ and K is now a suitable normalization constant. If we define

$$T = \frac{\theta-(\bar{x}_2-\bar{x}_1)}{s\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{\frac{1}{2}}},$$

we see that the a posteriori distribution of T is Student with $\nu = \nu_1 + \nu_2$ degrees of freedom. This allows us to use Student distribution's table to obtain confidence intervals for Θ.

Part II
Exercises

Chapter 9
Combinatorics

Exercise 9.1 The game of bridge is played with 52 cards. Compute:

1. The number of different ways a player can receive an handful of 13 cards.
2. The number of different ways the cards can be distributed among 4 players.
3. The number of different ways a player can receive an handful of 13 cards all different in values. Which is the number of different ways in which all 4 players receive cards all different in values?
4. The number of different ways a player can receive flush number cards of the same sign. In how many ways can a player obtain at least 2 cards with equal value?

Solution 9.1 1. The number of different ways a player can receive an handful of 13 cards is given by the simple combinations

$$\binom{52}{13}.$$

Namely one has to choose 13 elements out of 52 without repetitions and without taking in account of the order.

2. For the first player we have already computed the number of different ways she can receive an handful of 13 cards. For the second player we can choose 13 cards out of the $52 - 13 = 39$ remaining ones. Analogously for the third player. The fourth player receives the remaining 13 cards. The number of different ways in which all 4 players receive cards all different in values is then

$$\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13} = \frac{52!}{(13!)^4}.$$

The multinomial coefficient counts the number of ways of making 4 groups of 13 elements each out of a set of 52 cards.

3. Since the cards are all different in values, we can think that they are in increasing order. For the first card, we can choose one of the four aces. For the second one, one of the 4 twos, and so on. The number of different ways a player can receive

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8_9

an handful of 13 cards all different in values is then in

$$\underbrace{4 \cdot 4 \cdot \cdots \cdot 4}_{13 \text{ times}} = 4^{13}$$

different ways. If we consider all 4 players, we have that for the second player the choices for each card will reduce to

$$\underbrace{3 \cdot 3 \cdot \cdots \cdot 3}_{13 \text{ times}} = 3^{13}$$

different ways. Then the 4 players receive cards all different in values in

$$4^{13} \cdot 3^{13} \cdot 2^{13} \cdot 1^{13} = (4!)^{13}$$

different ways.

4. A player can receive flush number cards of the same sign in 4 different ways, since there exist flush number cards of 4 different signs. If we consider 4 players, the number of ways of assigning them flush number cards of the same sign is given by the number of permutations of the 4 signs, i.e.

$$4 \cdot 3 \cdot 2 \cdot 1 = 4!.$$

The number of ways in which a player can obtain at least 2 cards with equal value is equal to

$$\binom{52}{13} - 4^{13}$$

that is the number of all possible choices minus the number of ways of obtaining an handful of all different cards.

□

Exercise 9.2 At the ticket counter of a theatre there are available tickets with numbers from 1 to 100. The tickets are randomly distributed among the buyers. Four friends A, B, C, D buy separately a ticket each.

1. Which is the probability that they have received the tickets with numbers $31, 32, 33$ and 34?
2. Which is the probability that they have received the tickets $31, 32, 33$ and 34 in this order?
3. Which is the probability that they have received tickets with 4 consecutive numbers?
4. Which is the probability that A, B, C receive tickets with a number greater than 50?

Solution 9.2 1. To compute this probability we use the formula

$$\frac{\sharp \text{ favorable cases}}{\sharp \text{ possible cases}}.$$

The possible cases are all the ways of choosing 4 numbers out of 100, i.e.

$$\binom{100}{4}.$$

There exists only 1 favorable case, i.e. to choose the numbers 31, 32, 33 and 34. Hence the probability that the three friends have received the tickets with numbers 31, 32, 33, 34 is given by

$$p = \frac{1}{\binom{100}{4}}.$$

2. Here the number of possible cases is given by

$$\mathcal{D}_4^{100} = \frac{100!}{96!}.$$

The probability that the 4 friends receive the tickets 31, 32, 33, 34 in this order is then

$$p = \frac{1}{\mathcal{D}_4^{100}} = \frac{96!}{100!}.$$

3. One can obtain tickets with consecutive numbers in

$$100 - 3 = 97$$

different ways. We need also to consider the case {97, 98, 99, 100}. The probability of receiving 4 consecutive tickets is then

$$\frac{97}{\binom{100}{4}} = \frac{97!4!}{100!}.$$

4. The probability that A, B and C receive tickets with numbers greater than 50 is

$$p = \frac{50}{100}\,\frac{49}{99}\,\frac{48}{98}.$$

For the first case the are 50 favorable cases (all tickets with number from 51 up to 100) out of 100. For the second ticket there are 49 possibilities out of the 99 tickets left. And so on.

Exercise 9.3 A credit card PIN consists of 5 numbers. We assume that every sequence of 5 digits is generated with the same probability. Compute:

1. The probability that the numbers composing the PIN are all different.
2. The probability that the PIN contains at least 2 numbers which are equal.
3. The probability that the numbers composing the PIN are all different if the first digit is different from 0.
4. The probability that the PIN contains exactly 2 numbers which are equal, if the first digit is different from 0.

Solution 9.3 1. A PIN differs from another one if the digits are in different order. The possible cases are given by

$$10^5 .$$

The favorable cases, when all digits are different, are

$$\mathcal{D}_5^{10} = \frac{10!}{5!} .$$

The probability that the numbers composing the PIN are all different is then

$$p_1 = \frac{\mathcal{D}_5^{10}}{10^5} .$$

2. The probability that the PIN contains at least 2 numbers which are equal is

$$p = 1 - p_1 = 1 - \frac{10!}{5!\,10^5} ,$$

where p_1 is the probability that the numbers composing the PIN are different.

3. In this case the number of possible cases is

$$9 \cdot 10 \cdot 10 \cdot 10 \cdot 10.$$

For the first digit we have 9 possibilities (all numbers from 1 to 9). We need to choose the remaining digits without repetitions and taking in account the order: we have $\mathcal{D}_4^9$ ways. The number of favorable cases is then

$$9 \cdot 9 \cdot 8 \cdot 7 \cdot 6 = 9 \cdot \mathcal{D}_4^9 .$$

The probability that the numbers composing the PIN are all different if the first digit is different from 0 is then

$$\frac{9 \cdot \mathcal{D}_4^9}{9 \cdot 10^4} = \frac{\mathcal{D}_4^9}{10^4} .$$

4. The possible cases are still given by

$$9 \cdot 10^4.$$

In order to compute the number of ways in which the PIN contains exactly 2 numbers which are equal, if the first digit is different from 0, we can proceed as follows:

(a) For the digit that is repeated: without loss of generality we can think that it is equal to the first digit in the string. There are 9 ways of choosing it (remember: the 0 is now excluded).

(b) We choose the place of the repeated digit in the string: there are

$$\binom{4}{1}$$

positions where it can be placed.

(c) The other digits must be different from the first one. They can be placed in

$$\mathcal{D}_3^9$$

different ways in the string.
In total we have

$$9 \cdot \binom{4}{1} \cdot \mathcal{D}_3^9$$

possibilities. The procedure illustrated in (a),(b), and (c) will be called in the sequel the *string rule*.

(d) If the repeated digit is different from the first one, we have
- 9 ways of choosing the first digit;
- 9 ways of choosing the repeated digit;
- $\binom{4}{2}$ ways of choosing the place in the string;
- $\mathcal{D}_2^8$ ways of placing the remaining digits.

Totally we have

$$9 \cdot 9 \cdot \mathcal{D}_2^8 \cdot \binom{4}{2}.$$

The total number of favorable cases is then

$$9 \cdot \binom{4}{1} \cdot \mathcal{D}_3^9 + 9 \cdot 9 \cdot \mathcal{D}_2^8 \cdot \binom{4}{2} = 9 \cdot \left[\binom{4}{1} + \binom{4}{2}\right] \cdot \mathcal{D}_3^9 = 9 \cdot \binom{5}{2} \cdot \mathcal{D}_3^9,$$

where we have used the formula

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}.$$

□

Exercise 9.4 Four fair dice are thrown at the same time. Their faces are numbered from 1 to 6. Compute:

(a) The probability of obtaining four different faces.
(b) The probability of obtaining at least 2 equal faces.
(c) The probability of obtaining exactly 2 equal faces.
(d) The probability that the sum of the faces is equal to 5.
(e) We throw only 2 dice. Compute the probability that the sum of the faces is an odd number.

Solution 9.4 (a) To compute the probability we use the formula

$$p = \frac{\sharp\text{favorable cases}}{\sharp\text{possible cases}}. \tag{9.1}$$

The possible cases are given by

$$\text{possible cases} = 6 \cdot 6 \cdot 6 \cdot 6 = 6^4.$$

This is given by the number of all possible dispositions of 4 elements out of 6. The favorable cases are given by the simple dispositions of 4 elements out of 6, since the faces are required to be different from each other:

$$\text{favorable cases} = 6 \cdot 5 \cdot 4 \cdot 3 = D_4^6.$$

The probability of obtaining four different faces is then

$$\mathbf{P}(\text{all the thrown dies have different faces}) = \frac{D_4^6}{6^4} = \frac{5}{18}.$$

(b) The probability of obtaining at least 2 equal faces can be computed by using the probability obtained above, since:

$$\begin{aligned}
&\mathbf{P}(\text{the thrown dice have at least 2 same faces}) \\
&\quad = 1 - \mathbf{P}(\text{all the thrown dice have different faces}) \\
&\quad = 1 - \frac{D_4^6}{6^4} = \frac{13}{18}.
\end{aligned}$$

c) Also in this case we use the string rule as in Exercise 9.3. The number of the ways of obtaining exactly 2 equal faces is then:

$$\binom{4}{2} \cdot 6 \cdot D_2^5,$$

where

$$\begin{aligned}\binom{4}{2} &= \sharp \text{ ways of choosing 2 dice with equal faces,}\\ 6 &= \sharp \text{ ways of choosing the face which is repeated,}\\ D_2^5 &= \sharp \text{ ways of choosing the remaining faces.}\end{aligned}$$

Recall that the remaining faces must be different among each other and with respect to the one which is repeated.

(d) In order to have the sum of the faces equal to 5, the only possibility is that 3 faces present the number 1 and one the number 2, since we are dealing with 4 dice. We compute first the favorable cases. After having chosen the places for the number 1, it remains only one possibility for the number 2, i.e. we have

$$\binom{4}{3} \cdot 1 = 4$$

favorable cases. The possible cases are given by 6^4 ways of having a configuration of 4 dice. Hence the probability that the sum of the faces is equal to 5 is given by

$$p = \frac{4}{6^4}.$$

(e) The sum of the faces is odd if one of the faces presents an odd number and the other one an even number. Hence

$$\sharp\text{favorable cases} = \binom{2}{1} \cdot 3 \cdot 3 = 18,$$

where $\binom{2}{1}$ counts the number of ways for a die to come out with an even face. Hence

$$\mathbf{P}(\text{the sum of the faces is given by an odd number}) = \frac{2 \cdot 3^2}{6^2} = \frac{1}{2}.$$

More simply, one can consider that the sum of the faces can be either odd or even. Hence:

$$\sharp \text{ possible cases} = 2$$
$$\sharp \text{ favorable cases} = 1,$$

and consequently

$$\mathbf{P}(\text{the sum of the faces is given by an odd number}) = \frac{1}{2}.$$

□

Exercise 9.5 Two factories A and B produce garments for the same trademark Y. For the factory A, 5 % of the garments present some production defect; for the factory B, 7 % of the garments present some production defect. Furthermore 75 % of the garments sold by Y derive from the the factory A, while the remaining 25 % comes from the factory B. We suppose that a garment is chosen randomly with equal probability among all the garments on sale. Compute:

1. The probability of purchasing a garment of the trademark Y which presents some production defect.
2. The probability that the garment comes from the factory A, subordinated to the fact that it presents some production defect.

Solution 9.5 We denote by:

- with A the event

$$A = \{\text{the garment comes from the factory A}\};$$

- with B the event

$$B = \{\text{the garment comes from the factory B}\};$$

- with D the event

$$D = \{\text{the garment presents some production defect}\}.$$

1. The probability of purchasing a garment of the trademark Y which presents some production defect can be computed with the formula of the total probabilities, since we do not know whether it comes from factory A or B. Hence

$$\mathbf{P}(D) = \mathbf{P}(D|A)\,\mathbf{P}(A) \;+\; \mathbf{P}(D|B)\,\mathbf{P}(B)$$

$$= \frac{5}{100}\,\frac{75}{100} + \frac{7}{100}\,\frac{25}{100} = \frac{11}{200}.$$

2. The probability that the garment comes from the factory A, if it presents some production defect, is given by:

$$\mathbf{P}(A|D) = \frac{\mathbf{P}(D|A)\,\mathbf{P}(A)}{\mathbf{P}(D)} = \frac{15}{22}.$$

This subordinated probability has been computed with Bayes' Formula.

□

Exercise 9.6 We consider 3 different elementary schools E, M, S. The percentage of pupils wearing glasses is 10 % in the school E, 25 % in the school M and 40 % in the school S. Compute:

1. The probability that by choosing randomly 3 pupils, one out of each school, at least one of them wears glasses.
2. The probability that a pupil wears glasses, if we randomly choose her or him out of the three schools (each school can be picked up with the same probability).
3. The probability that the pupil belongs to school E, if she wears glasses.

Solution 9.6 1. The quickest method to compute the probability that by choosing by chance 3 pupils, one out of each school, at least one of them wears glasses, is to evaluate the probability that none of them wears glasses. If B is the event that at least one of the 3 pupils wears glasses, then

$$\mathbf{P}(B) = 1 - \mathbf{P}\left(\tilde{B}\right).$$

In this case $\mathbf{P}\left(\tilde{B}\right) = \frac{90}{100}\,\frac{75}{100}\,\frac{60}{100} = \frac{81}{200}$, from which

$$\mathbf{P}(B) = 1 - \frac{81}{200} = \frac{119}{200}.$$

2. Let O be the event

$$O = \{\text{the pupil wears glasses}\}.$$

The probability of O can be computed by using the formula of the total probability, since we do not know which school the pupil belongs to. We set

- $E = \{$the pupil belongs to school E$\}$;
- $M = \{$the pupil belongs to school M$\}$;
- $S = \{$the pupil belongs to school S$\}$.

We then have:

$$\begin{aligned}\mathbf{P}(O) &= \mathbf{P}(O|E)\,\mathbf{P}(E) + \mathbf{P}(O|M)\,\mathbf{P}(M) + \mathbf{P}(O|S)\,\mathbf{P}(S)\\ &= \frac{1}{10}\,\frac{1}{3} + \frac{1}{4}\,\frac{1}{3} + \frac{2}{5}\,\frac{1}{3} = \frac{1}{4}.\end{aligned}$$

Note that we have assumed that each school can be picked up with the same probability.

3. The probability that the pupil belongs to school E, if she wears glasses, can be computed by using Bayes' formula:

$$\mathbf{P}(E|O) = \frac{\mathbf{P}(O|E)\,\mathbf{P}(E)}{\mathbf{P}(O)} = \frac{2}{15}\,.$$

□

Chapter 10
Discrete Distributions

Exercise 10.1 Two friends A and B are playing with a deck of cards consisting of 52 cards, 13 for each sign. They choose out 2 cards each. Player A starts. In order to win, the player has to be the first to extract the ace of spade or 2 cards of diamonds. After having chosen the 2 cards, they put the 2 cards back in the deck and mix it. Compute the probability that:

(a) Player A wins after 3 trials (i.e. after each player has done 2 extractions).
(b) Player A wins, player B wins, nobody wins.
(c) Let T be the random number representing the number of the trial, when one of the player first wins. Compute the expectation of T.
(d) Which is the probability distribution of T?

Solution 10.1 (a) The trials of the 2 players can be represented as a sequence of stochastically independent and equally distributed random trials. The probability that player A wins after 3 trials (i.e. after each player has done 2 extractions) is then equal to the probability of first success after

$$2 + 2 + 1 = 5$$

trials. The player A wins if she extracts the ace of spade or 2 cards of diamonds. The probability of this event is given by

$$p = \frac{51}{\binom{52}{2}} + \frac{\binom{13}{2}}{\binom{52}{2}} \tag{10.1}$$

where we have used the fact that the events are incompatible and that:

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8_10

1. The probability of extracting the ace of spade, is given by $\dfrac{1\cdot\binom{51}{1}}{\binom{52}{2}}$.
2. The probability of extracting 2 cards of diamonds, is given by $\dfrac{\binom{13}{2}}{\binom{52}{2}}$.

Let T be the random number representing the first time of success. The probability that A wins at the third trial is

$$\mathbf{P}(T = 5) = p(1-p)^4 ,$$

where p is given by (10.1).

(b) If A wins, the game stops with an odd trial. The probability that A wins is then

$$\begin{aligned}\mathbf{P}(A \text{ wins}) &= \sum_{k=0}^{\infty} \mathbf{P}(T = 2k+1)\\ &= \sum_{k=0}^{\infty} p(1-p)^{2k} = p\frac{1}{1-(1-p)^2} .\end{aligned}$$

If B wins, the game stops with an even trial. The probability that B wins is then

$$\begin{aligned}\mathbf{P}(B \text{ wins}) &= \sum_{k=1}^{\infty} \mathbf{P}(T = 2k)\\ &= p\sum_{k=1}^{\infty}(1-p)^{2k-1}\\ &= \frac{p}{1-p}\left(\frac{1}{1-(1-p)^2} - 1\right)\\ &= \frac{p}{1-p}\frac{(1-p)^2}{1-(1-p)^2}\\ &= \frac{p(1-p)}{1-(1-p)^2} = \frac{1-p}{2-p} .\end{aligned}$$

The probability that nobody wins is given by

$$\begin{aligned}
\mathbf{P}(\text{nobody wins}) &= 1 - \mathbf{P}(A \text{ wins}) - \mathbf{P}(B \text{ wins}) \\
&= 1 - \sum_{k=0}^{\infty} \mathbf{P}(T = 2k+1) - \sum_{k=1}^{\infty} \mathbf{P}(T = 2k) \\
&= 1 - \sum_{k=1}^{\infty} \mathbf{P}(T = k) \\
&= 0\,.
\end{aligned}$$

(c)–(d) The random number T that represents the time when the game is decided, has a geometric distribution of parameter p since it denotes the first time of success in a sequence of stochastically independent and identically distributed trials. Hence the expectation of T is given by:

$$\mathbf{P}(T) = \frac{1}{p} = \frac{\binom{52}{2}}{1 + \binom{13}{2}}\,.$$

□

Exercise 10.2 Let X, Y be two stochastically independent random numbers with Poisson distribution with parameters μ and σ, respectively.

1. Let $Z = X + Y$. Compute the expectation and the variance of Z.
2. What is the set $I(Z)$ of possible values for Z?
3. Compute $\mathbf{P}(Z = i)$, for $i \in I(Z)$.
4. Compute $\mathbf{cov}(Z, X,)$.
5. Let $u > 0$; compute the generating function $\phi_Z(u) = \mathbf{P}(u^Z)$ of Z.

Solution 10.2 1. By the linearity of the expectation we obtain

$$\mathbf{P}(Z) \;=\; \mathbf{P}(X+Y) \;=\; \mathbf{P}(X) + \mathbf{P}(Y) \;=\; \mu + \sigma\,.$$

To compute the variance, we use the formula of the variance of the sum

$$\boldsymbol{\sigma}^2(X+Y) \;=\; \boldsymbol{\sigma}^2(X) + \boldsymbol{\sigma}^2(Y) + 2\,\mathbf{cov}(X, Y)\,.$$

Since X, Y are stochastically independent, we have

$$\mathbf{cov}(X, Y) = 0\,.$$

Hence

$$\boldsymbol{\sigma}^2(X+Y) \;=\; \boldsymbol{\sigma}^2(X) + \boldsymbol{\sigma}^2(Y) \;=\; \mu + \lambda\,.$$

2. The set $I(Z)$ of possible values for Z is given by

$$I(Z) \;=\; \mathbb{N} \;=\; \{\inf(X) + \inf(Y), \ldots\}\,.$$

3. We now compute the probability distribution of Z. The event $\{Z = i\}$ can be written as

$$\begin{aligned}\{Z = i\} &= \{X = 0, Y = i\} + \{X = 1, Y = i - 1\} + \cdots + \{X = i, Y = 0\} \\ &= \sum_{k=0}^{i} \{X = k, Y = i - k\},\end{aligned}$$

since the events $\{X = k, Y = i - k\}$ are disjoint for $k = 0, \ldots, i$. By the linearity of the expectation we obtain

$$\mathbf{P}(Z = i) \;=\; \sum_{k=0}^{i} \mathbf{P}(X = k, Y = i - k)\,.$$

Furthermore X, Y are stochastically independent, hence

$$\mathbf{P}(X = k, Y = i - k) \;=\; \mathbf{P}(X = k)\,\mathbf{P}(Y = i - k),$$

so that

$$\begin{aligned}\mathbf{P}(Z = i) &= \sum_{k=0}^{i} \mathbf{P}(X = k)\,\mathbf{P}(Y = i - k) \\ &= \sum_{k=0}^{i} \frac{\mu^k}{k!}\, e^{-\mu}\, \frac{\sigma^{(i-k)}}{(i-k)!}\, e^{-\sigma} \\ &= \frac{e^{-(\mu+\sigma)}}{i!} \sum_{k=0}^{i} \frac{i!}{k!\,(i-k)!}\, \mu^k\, \sigma^{(i-k)} \\ &= \frac{(\mu+\sigma)^i}{i!}\, e^{-(\mu+\sigma)}\,,\end{aligned}$$

where we have used Newton's binomial formula. Therefore Z has Poisson distribution with parameter $\mu + \lambda$.

4. In order to compute the covariance between Z and X, we proceed as follows:

$$\begin{aligned}\mathbf{cov}(Z, X) &= \mathbf{P}(ZX) - \mathbf{P}(Z)\,\mathbf{P}(X) \\ &= \mathbf{P}\left((X + Y)X\right) - \left(\mathbf{P}(X) + \mathbf{P}(Y)\right)\,\mathbf{P}(X) \\ &= \mathbf{P}(X^2) + \mathbf{P}(XY) - \mathbf{P}(X)^2 - \mathbf{P}(Y)\mathbf{P}(X) \\ &= \boldsymbol{\sigma}^2(X) \\ &= \mu\,.\end{aligned}$$

5. For $\mu > 0$, the generating function of Z is given by

$$\phi_Z(u) \;=\; \mathbf{P}(u^Z) \;=\; \mathbf{P}(u^{X+Y})\,.$$

Since X, Y are stochastically independent, we have

$$\mathbf{P}(u^{X+Y}) \;=\; \mathbf{P}(u^X \cdot u^Y) \;=\; \mathbf{P}(u^X) \cdot \mathbf{P}(u^Y)\,.$$

We now compute $\mathbf{P}(u^X)$ by using the formula for the expectation of a function of X:

$$\begin{aligned}\mathbf{P}(u^X) &= \sum_{i=0}^{+\infty} u^i \mathbf{P}(X = i)\\ &= e^{-\mu} \sum_{i=0}^{+\infty} \frac{(u\mu)^i}{i!}\\ &= e^{(u-1)\,\mu},\end{aligned}$$

where in the last step we have used the series:

$$\sum_{i=0}^{+\infty} \frac{x^i}{i!} \;=\; e^x\,.$$

It follows that

$$\begin{aligned}\phi_Z(u) &= \mathbf{P}(u^X) \cdot \mathbf{P}(u^Y)\\ &= e^{(u-1)\,\mu}\, e^{(u-1)\,\sigma}\\ &= e^{(u-1)\,(\mu+\sigma)}.\end{aligned}$$

Since the generating function uniquely identifies the distribution, this proves that Z has Poisson distribution with the parameter $\mu + \sigma$. □

Exercise 10.3 In a small village with 200 inhabitants, 5 inhabitants are affected by a particular genetic disease. A sample of 3 individuals is chosen randomly among the population (all subsets have the same probability of being chosen). Let X be the number of individuals in the sample who are affected by the disease.

1. Determine the set $I(X)$ of possible values for X.
2. Determine the probability distribution of X.
3. Compute the expectation and the variance of X.

Solution 10.3 1. The possible values of X are $0, 1, 2$ and 3, i.e. the minimum number of people affected by the disease in the sample is 0 and the maximum number is 3.

2. Consider the event $\{X = i\},\ i \in I(X)$. To determine the probability distribution of X, we need to compute

$$\mathbf{P}(X = i), \quad i \in I(X)\,.$$

To this purpose we use the formula

$$\frac{\sharp \text{ favorable cases}}{\sharp \text{ possible cases}}\,.$$

The number of possible cases is given by the number of ways of choosing 3 people out of 200 inhabitants, i.e.

$$\binom{200}{3}.$$

The number of favorable cases is given by the number of ways of choosing i people out of the group of inhabitants affected by the disease and $(3 - i)$ people out of the group of 'healthy' people, i.e.

$$\binom{5}{i}\binom{195}{3-i}.$$

We obtain

$$\mathbf{P}(X = i) = \frac{\binom{5}{i}\binom{195}{3-i}}{\binom{200}{3}}\,.$$

The distribution of X is then hypergeometric.

3. We can compute directly the expectation of X, since $I(X)$ consists only of 4 values:

$$\begin{aligned}
\mathbf{P}(X) &= \sum_{i=0}^{3} i\,\mathbf{P}(X = i) \\
&= \frac{1}{\binom{200}{3}}\left(5\binom{195}{2} + 20\binom{195}{1} + 30\right) \\
&= \frac{3}{40}\,.
\end{aligned}$$

For the variance the computations is analogous. It is sufficient to apply the formula

$$\sigma^2(X) \;=\; \mathbf{P}(X^2) - \mathbf{P}(X)^2$$

and to compute $\mathbf{P}(X^2) = \sum_{i=0}^{3} i^2\,\mathbf{P}(X = i)$. □

Exercise 10.4 At a horse race there are 10 participants. Gamblers can win if they correctly predict the first 3 horses in order of arrival. We suppose that all the orders have the same probability of occurrence and that the gamblers choose independently of each other and with the same probability the 3 horses on which to bet.

1. Compute the probability that one of the gamblers wins.
2. If the gamblers are 100 in total, let X be the random numbers counting the number of gamblers who win. Determine $I(X)$ and $\mathbf{P}(X = i)$ for $i = 1, 2, 3$.
3. Compute expectation and variance of X.
4. Suppose that the gamblers are numbered from 1 to 100. Compute the probability that there is at least one winner and that the winner with the minimal number has a number greater or equal to 50.

Solution 10.4 1. The probability that a gambler wins can be computed with the formula

$$\frac{\sharp \text{ favorable cases}}{\sharp \text{ possible cases}}\,.$$

In this case, the possible cases are given by the simple dispositions of 3 elements out of 10. They represent the number of ways of assuming the first 3 positions for the 10 horses. Only one is the winning triplet, hence the probability of winning for a gambler is given by

$$p \;=\; \frac{1}{\mathcal{D}_3^{10}} \;=\; \frac{7!}{10!} \;=\; \frac{1}{720}\,.$$

2. If X is the random numbers counting the number of gamblers who win, we can write

$$X \;=\; E_1 + E_2 + \cdots + E_{100},$$

where the event E_i is verified if the i-th gambler wins. The events E_i, $i = 1, \ldots, 100$, are stochastically independent and identically distributed since the gamblers choose independently of each other and with the same probability the 3 horses on which to bet. Hence X has binomial distribution $\mathrm{Bn}(n, p)$ of parameters $n = 100$ and $p = \dfrac{1}{720}$. The set of possible values is then

$$I(X) = \{0, 1, \ldots, 100\}$$

and

$$\mathbf{P}(X = i) = \binom{100}{i} \left(\frac{1}{720}\right)^{i} \left(1 - \frac{1}{720}\right)^{100-i}.$$

In particular, we obtain:

i = 1

$$\mathbf{P}(X = 1) = 100 \cdot \frac{1}{720} \cdot \left(\frac{719}{720}\right)^{99}.$$

i = 2

$$\mathbf{P}(X = 2) = \binom{100}{2} \left(\frac{1}{720}\right)^{2} \cdot \left(\frac{719}{720}\right)^{98}.$$

i = 3

$$\mathbf{P}(X = 3) = \binom{100}{3} \left(\frac{1}{720}\right)^{3} \cdot \left(\frac{719}{720}\right)^{97}.$$

3. The expectation of X is given by linearity by

$$\begin{aligned}\mathbf{P}(X) &= \mathbf{P}(E_1 + \cdots + E_{100}) \\ &= \sum_{i=1}^{100} \mathbf{P}(E_i) = 100 \cdot \frac{1}{720} = \frac{5}{36}.\end{aligned}$$

Analogously by the formula of the variance of the sum of n random numbers, we have:

$$\begin{aligned}\boldsymbol{\sigma}^2(X) &= \boldsymbol{\sigma}^2(E_1 + \cdots + E_{100}) \\ &= \sum_{i=1}^{100} \boldsymbol{\sigma}^2(E_i) + \underbrace{\sum_{i,j=1}^{100} \mathbf{cov}(E_i, E_j)}_{0} \\ &= 100 \cdot \frac{1}{720} \cdot \left(1 - \frac{1}{720}\right).\end{aligned}$$

Here we have used that the events E_i are stochastically independent.

4. In order to have a winner with minimal number greater than or equal to 50, we need that the first 49 gamblers do not win and that at least one of the gamblers with number from 50 to 100 wins. Let E be the event that all the all the gamblers with number from 1 to 49 lose and F the event that at least one of the gamblers

with number from 50 to 100 wins. The probability that there is at least one winner and that the winner with the minimal number has a number greater than or equal to 50 is then

$$\mathbf{P}(EF) \;= \mathbf{P}(E)\mathbf{P}(F) = \left(\frac{719}{720}\right)^{49} \left[1 - \left(\frac{719}{720}\right)^{51}\right],$$

where $\mathbf{P}(F) = 1 - \mathbf{P}(\widetilde{F})$ and $\widetilde{F}$ is the event that no gambler with number from 50 to 100 wins. □

Exercise 10.5 In an opinion poll 100 people are asked to answer a questionnaire with 5 questions. Each question can be answered only yes or no. For each person the probability of all possible answers is the same and their choices are stochastically independent. Let N be the number of interviewed people that answer yes to the first questions or answer yes at least to 4 questions.

1. Which is the probability distribution of N?
2. Compute the expectation, the variance and the generating function of N.

Solution 10.5 1. Let E_i be the event that the i-th interviewed person has answered yes to the first questions or yes at least to 4 questions. We can rewrite N as

$$N \;=\; E_1 + E_2 + \cdots + E_{100}\,.$$

The events are stochastically independent and identically distributed since every person answers independently of the other ones. Furthermore we have assumed that the probability of all possible answers is the same. It is sufficient to compute the probability of each E_i. We put:

- F_i= {the i-th interviewed person answers yes to the first question};
- G_i= {the i-th interviewed person answers yes at least to 4 questions}.

We obtain that $E_i \;=\; F_i \vee G_i$ e

$$\mathbf{P}(E_i) \;=\; \mathbf{P}(F_i) + \mathbf{P}(G_i) - \mathbf{P}(F_i \wedge G_i)\,.$$

The probability of F_i, G_i e $F_i \wedge G_i$ are given by:

(a)

$$\mathbf{P}(F_i) \;=\; \frac{1}{2}\cdot\frac{1}{2} \;=\; \frac{1}{4}\,.$$

For all question we have 2 possible cases (yes and no), while for the first question we have only one possible choice (yes).

(b)

$$\mathbf{P}(G_i) \;=\; \binom{5}{4}\left(\frac{1}{2}\right)^5 + \binom{5}{5}\left(\frac{1}{2}\right)^5\,.$$

A person answers yes at least to 4 questions if she answers yes to exactly 4 questions or to exactly 5 questions.

(c)

$$\mathbf{P}(F_i \wedge G_i) = \binom{3}{2}\left(\frac{1}{2}\right)^5 + \binom{3}{3}\left(\frac{1}{2}\right)^5 .$$

In the case the events happen at the same time, we need to choose only the other 2, respectively 3 questions to which the candidate answer yes.

Finally

$$\mathbf{P}(E_i) = \frac{1}{4} + \binom{5}{4}\frac{1}{2^5} + \frac{1}{2^5} - \frac{3}{2^5} - \frac{1}{2^5} = \frac{1}{2^2} + \frac{1}{2^4} = \frac{5}{16} .$$

We obtain that $I(N) = \{0, \ldots, 100\}$ and

$$\mathbf{P}(N = i) = \binom{100}{i}\left(\frac{5}{16}\right)^i \left(1 - \frac{5}{16}\right)^{100-i} ,$$

i.e. N has binomial distribution $\mathrm{Bn}(100, \frac{5}{16})$.

2. The expectation of N is given by

$$\mathbf{P}(N) = \sum_{i=1}^{100} i\mathbf{P}(E_i) = 100 \cdot \frac{5}{16} = \frac{125}{4} .$$

The variance of N is given by

$$\begin{aligned}\boldsymbol{\sigma}^2(N) &= \sum_{i=1}^{100} \boldsymbol{\sigma}^2(E_i) + \underbrace{\sum_{i,j=1}^{100} \mathbf{cov}(E_i, E_j)}_{0} \\ &= 100 \cdot \frac{5}{16} \cdot \left(1 - \frac{5}{16}\right) .\end{aligned}$$

The generating function of N is given by

$$\begin{aligned}\phi_N(t) &= \mathbf{P}(t^N) \\ &= \sum_{i=0}^{100} \binom{100}{i} \cdot \left(\frac{5t}{16}\right)^i \cdot \left(1 - \frac{5}{16}\right)^{100-i} \\ &= \left(\frac{5t}{16} + 1 - \frac{5}{16}\right)^{100} ,\end{aligned}$$

where we have used Newton's binomial formula. □

Exercise 10.6 A box contains 8 balls: 4 white and 4 black. We draw 4 balls. Let E_i be the event that the i-th ball extracted is white. Let $X = E_1 + E_2$, $Y = E_3 + E_4$.

(a) Compute the joint distribution of X and Y.
(b) Compute $\mathbf{P}(X)$, $\mathbf{P}(Y)$, $\sigma^2(X)$, $\sigma^2(Y)$.
(c) Compute $\mathbf{cov}(X, Y)$, the correlation coefficient $\rho(X, Y)$. Are X and Y stochastically independent?

Solution 10.6 (a) Consider the random vector (X, Y). The set of possible values for (X, Y) is given by

$$I(X, Y) = \{(i, j)|\ i = 0, 1, 2,\ \ j = 0, 1, 2\}\,.$$

To compute the joint distribution of (X, Y), we need to calculate

$$\mathbf{P}(X = i, Y = j) = \mathbf{P}(Y = j|\ X = i)\mathbf{P}(X = i)$$

for all $(i, j) \in I(X, Y)$. The probability of extracting a white ball in the first 2 extractions is given by

$$\mathbf{P}(X = i) = \frac{\binom{4}{i}\binom{4}{2-i}}{\binom{8}{2}}\,.$$

Here the possible cases are $\binom{8}{2}$ since we consider only the first 2 extractions. Moreover

$$\mathbf{P}(Y = j|\ X = i) = \frac{\binom{4-i}{j}\binom{4-(2-i)}{2-j}}{\binom{6}{2}} = \frac{\binom{4-i}{j}\binom{2+i}{2-j}}{\binom{6}{2}}\,.$$

After the first 2 extractions, only 6 balls are left in the box. We have to draw 2 more balls, j among the remaining white ones $(4 - i)$ and $(2 - j)$ among the remaining black ones $4 - (2 - i) = 2 + i$. The joint distribution of X and Y is then

$$\mathbf{P}(X=i, Y=j) = \frac{\binom{4-i}{j}\binom{2+i}{2-j}}{\binom{6}{2}} \cdot \frac{\binom{4}{i}\binom{4}{2-i}}{\binom{8}{2}}.$$

(b) To compute $\mathbf{P}(X)$ and $\mathbf{P}(Y)$ we use the fact that the events E_i have equal probability (but they are not stochastically independent!), hence

$$\mathbf{P}(X) = \mathbf{P}(E_1) + \mathbf{P}(E_2) = 2 \cdot \frac{4}{8} = 1$$

and

$$\mathbf{P}(X) = \mathbf{P}(Y) = 1\,.$$

The events E_1 and E_2 (and consequently also E_3 e E_4) are negatively correlated with covariance:

$$\begin{aligned}\mathbf{cov}(E_1, E_2) &= \mathbf{P}(E_1E_2) - \mathbf{P}(E_1)\mathbf{P}(E_2)\\ &= \mathbf{P}(E_2|E_1)\mathbf{P}(E_1) - \mathbf{P}(E_1)\mathbf{P}(E_2)\\ &= -\frac{1}{28}\,.\end{aligned}$$

The variance of X is then

$$\begin{aligned}\boldsymbol{\sigma}^2(X) = \boldsymbol{\sigma}^2(E_1+E_2) &= \boldsymbol{\sigma}^2(E_1) + \boldsymbol{\sigma}^2(E_2) + 2\mathbf{cov}(E_1, E_2)\\ &= \frac{1}{4} + \frac{1}{4} - \frac{1}{28} = \frac{13}{28}\,.\end{aligned}$$

Also in this case $\boldsymbol{\sigma}^2(Y) = \boldsymbol{\sigma}^2(X) = \frac{13}{28}$.

(c) We have:

$$\begin{aligned}\mathbf{cov}(X, Y) &= \mathbf{cov}(E_1+E_2, E_3+E_4)\\ &= \mathbf{cov}(E_1, E_3) + \mathbf{cov}(E_1, E_4) + \mathbf{cov}(E_2, E_3) + \mathbf{cov}(E_2, E_4)\\ &= 4 \cdot \left(-\frac{1}{28}\right) = -\frac{1}{7}\,.\end{aligned}$$

Here we have used that fact that the covariance is a bilinear function. Finally, the coefficient of correlation between X and Y is equal to:

$$\rho(X, Y) = \frac{\mathbf{cov}(X, Y)}{\boldsymbol{\sigma}(X)\boldsymbol{\sigma}(Y)} = \frac{-\frac{1}{7}}{\sqrt{\frac{13}{28}} \cdot \sqrt{\frac{13}{28}}} = -\frac{4}{13}\,.$$

□

Exercise 10.7 Let E_1, E_2, F_1, F_2 be stochastically independent events with $\mathbf{P}(E_1) = \mathbf{P}(E_2) = \frac{1}{4}$, $\mathbf{P}(F_1) = \mathbf{P}(F_2) = \frac{1}{3}$. Let $X = E_1 + E_2$, $Y = F_1 + F_2$.

(a) Compute the set of possible values and the probability distributions of X and Y.
(b) Compute $\mathbf{P}(X + Y)$, $\boldsymbol{\sigma}^2(X + Y)$.
(c) Compute $\mathbf{P}(X = Y)$, $\mathbf{P}(X = -Y)$.

Solution 10.7 (a) Since E_1, E_2 are events, i.e. random numbers that can assume only the values 0 and 1, we have that the set of possible values of X is given by

$$I(X) = \{0, 1, 2\}.$$

Analogously for Y

$$I(Y) = \{0, 1, 2\}.$$

To compute the probability distribution of X means that we have to calculate with which probability X assumes each of the possible values. For example, we have that

$$\begin{aligned}\mathbf{P}(X = 0) &= \mathbf{P}(E_1 + E_2 = 0) \\ &= \mathbf{P}(E_1 = E_2 = 0) = \mathbf{P}(\tilde{E}_1)\mathbf{P}(\tilde{E}_2) = \frac{9}{16}.\end{aligned}$$

Since X is equal to the sum of 2 stochastically independent events with the same probability, we can immediately say that the distribution of X is binomial $\mathrm{Bn}(n, p)$ with parameters $n = 2$ and $p = \frac{1}{4}$. Analogously Y has binomial distribution $\mathrm{Bn}(2, \frac{1}{3})$ and we have that

$$\begin{aligned}\mathbf{P}(X = i) &= \binom{2}{i}\left(\frac{1}{4}\right)^i \left(\frac{3}{4}\right)^{2-i}, \quad i = 0, 1, 2 \\ \mathbf{P}(Y = j) &= \binom{2}{j}\left(\frac{1}{3}\right)^j \left(\frac{2}{3}\right)^{2-j}, \quad j = 0, 1, 2.\end{aligned}$$

(b) To compute the expectation, we can use the linearity

$$\begin{aligned}\mathbf{P}(X + Y) &= \mathbf{P}(E_1 + E_2 + F_1 + F_2) \\ &= \mathbf{P}(E_1) + \mathbf{P}(E_2) + \mathbf{P}(F_1) + \mathbf{P}(F_2) \\ &= 2 \cdot \frac{1}{4} + 2 \cdot \frac{1}{3} = \frac{7}{6}.\end{aligned}$$

For the variance, we use the formula of the variance of a sum:

$$\boldsymbol{\sigma}^2(X + Y) = \boldsymbol{\sigma}^2(X) + \boldsymbol{\sigma}^2(Y) + 2\,\mathbf{cov}(X, Y).$$

Since X and Y have binomial distribution, we have

$$\boldsymbol{\sigma}^2(X) = 2 \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{8},$$

$$\boldsymbol{\sigma}^2(Y) = 2 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{4}{9}.$$

To compute the covariance between X and Y we use the fact that the events E_1, E_2, F_1, F_2 are stochastically independent in the following way:

$$\begin{aligned}
\mathbf{cov}(X, Y) &= \mathbf{cov}(E_1 + E_2, F_1 + F_2) \\
&= \mathbf{cov}(E_1, F_1) + \mathbf{cov}(E_1, F_2) + \mathbf{cov}(E_2, F_1) + \mathbf{cov}(E_2, F_2) \\
&= 0.
\end{aligned}$$

Hence

$$\boldsymbol{\sigma}^2(X + Y) = \frac{3}{8} + \frac{4}{9} = \frac{59}{72}.$$

(c) To compute $\mathbf{P}(X = Y)$ we note that the event

$$(X = Y)$$

is given by

$$(X = Y) = (X = 0, X = 0) + (X = 1, Y = 1) + (X = 2, Y = 2).$$

Hence

$$\begin{aligned}
\mathbf{P}(X = Y) &= \sum_{i=0}^{2} \mathbf{P}(X = i, Y = i) \\
&= \sum_{i=0}^{2} \mathbf{P}(X = i)\mathbf{P}(Y = i) \\
&= \sum_{i=0}^{2} \binom{2}{i} \left(\frac{1}{4}\right)^i \left(\frac{3}{4}\right)^{2-i} \binom{2}{i} \left(\frac{1}{3}\right)^i \left(\frac{2}{3}\right)^{2-i} \\
&= \sum_{i=0}^{2} \binom{2}{i}^2 \left(\frac{1}{12}\right)^i \left(\frac{1}{2}\right)^{2-i} \\
&= \frac{1}{4} \sum_{i=0}^{2} \binom{2}{i}^2 \left(\frac{1}{6}\right)^i = \frac{61}{144}.
\end{aligned}$$

On the other side the event

$$(X = -Y)$$

is verified only if

$$(X = -Y) = (X = 0, Y = 0)\,.$$

Hence

$$\mathbf{P}(X = -Y) = \mathbf{P}(X = 0, Y = 0) = \mathbf{P}(X = 0)\mathbf{P}(Y = 0) = \frac{1}{4}\,. \qquad \square$$

Chapter 11
One-Dimensional Absolutely Continuous Distributions

Exercise 11.1 The random numbers X, Y and Z are stochastically independent with exponential distribution of parameter $\lambda = 2$.

(a) Compute the density of the probability of $X + Y$ and of $X + Y + Z$.
(b) Let E, F, G be the events $E = (X \leq 2)$, $F = (X + Y > 2)$, $G = (X + Y + Z \leq 3)$. Compute $\mathbf{P}(E)$, $\mathbf{P}(F)$, $\mathbf{P}(G)$ e $\mathbf{P}(EF)$.
(c) Determine if E, F and G are stochastically independent.

Solution 11.1 (a) The exponential distribution is a particular case of the gamma distribution with parameter $1, \lambda$. If X, Y and Z are stochastically independent random numbers with exponential distribution of parameter $\lambda = 2$, i.e. Gamma distribution $\Gamma(1, 2)$, we can use the following property of the the sum of stochastically independent random numbers with Gamma distribution

$$\Gamma(\alpha, \lambda) + \Gamma(\beta, \lambda) \sim \Gamma(\alpha + \beta, \lambda).$$

Hence $W_1 = X + Y$ has distribution $\Gamma(2, 2)$. We can iterate this procedure and obtain that

$$W_2 = X + Y + Z = W_1 + Z$$

has distribution $\Gamma(3, 2)$.

(b) We have:

$$\begin{aligned}\mathbf{P}(E) = \mathbf{P}(X \leq 2) &= \int_0^2 2e^{-2x}\mathrm{d}x \\ &= 1 - e^{-4};\end{aligned}$$

$$\begin{aligned}\mathbf{P}(F) = \mathbf{P}(X + Y > 2) &= \int_2^{+\infty} 4xe^{-2x}\mathrm{d}x \\ &= \left[-2xe^{-2x}\right]_2^{+\infty} + 2\int_2^{+\infty} e^{-2x}\mathrm{d}x\end{aligned}$$

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8_11

$$= 4e^{-4} + \left[-e^{-2x}\right]_2^{+\infty} = 5e^{-4}\,;$$

$$\begin{aligned}\mathbf{P}(G) = \mathbf{P}(X+Y+Z \le 3) &= \int_0^3 4x^2 e^{-2x}\mathrm{d}x \\ &= 1 - \int_3^{+\infty} 4x^2 e^{-2x}\mathrm{d}x \\ &= 1 - \left\{\left[-2x^2 e^{-2x}\right]_3^{+\infty} + 4\int_3^{+\infty} x e^{-2x}\mathrm{d}x\right\} \\ &= 1 - 25e^{-6},\end{aligned}$$

and

$$\begin{aligned}\mathbf{P}(EF) &= \mathbf{P}(X \le 2, X+Y > 2) \\ &= \mathbf{P}(X \le 2, Y > 2 - X) = \mathbf{P}(X \le 2, Y > 0) \\ &= \mathbf{P}(X \le 2)\mathbf{P}(Y > 0) = \mathbf{P}(X \le 2)\,.\end{aligned}$$

Here we have used the fact that X and Y are assumed to be stochastically independent, as well as that the product of 2 events denotes that both conditions must be simultaneously satisfied.

(c) To determine if E, F, G are stochastically independent, we need to verify all the following conditions:

$$\begin{aligned}\mathbf{P}(EF) &= \mathbf{P}(E)\mathbf{P}(F); \\ \mathbf{P}(EG) &= \mathbf{P}(E)\mathbf{P}(G); \\ \mathbf{P}(FG) &= \mathbf{P}(F)\mathbf{P}(G); \\ \mathbf{P}(EFG) &= \mathbf{P}(E)\mathbf{P}(F)\mathbf{P}(G)\,.\end{aligned}$$

If one of them is <u>not</u> verified, then the events are <u>not</u> stochastically independent. We can immediately see that

$$\mathbf{P}(EF) \neq \mathbf{P}(E)\mathbf{P}(F)$$

by using the results above. Hence the three events are not stochastically independent. □

Exercise 11.2 Let X be a random number with standard normal distribution. Let $Y = 3X + 2$ and $Z = X^2$.

1. Compute the c.d.f. and the density of Y.
2. Estimate $\mathbf{P}(Y \ge y)$, where $y > 0$.
3. Compute the expectation and the variance of Z.
4. Compute the c.d.f. and the density of Z.

Solution 11.2 1. Put

$$n(t) = \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{t^2}{2}}$$

We compute the c.d.f. F_Y of $Y = 3X + 2$. Given $y \in \mathbb{R}$

$$F_Y(y) = \mathbf{P}(Y \le y) = \mathbf{P}(3X + 2 \le y) =$$

$$\mathbf{P}\left(X \le \frac{y-2}{3}\right) = \int_{-\infty}^{\frac{y-2}{3}} n(t)\mathrm{d}t = \int_{-\infty}^{y} \frac{1}{3\sqrt{2\pi}} e^{-\frac{(z-2)^2}{18}}\,\mathrm{d}z\,,$$

where we have used the change of variable $t = \dfrac{z-2}{3}$. The density f_Y of Y is obtained by the derivation of F_Y:

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \frac{1}{3}\,\frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{(y-2)^2}{2\cdot 9}}\,.$$

It follows that Y has normal distribution $N(2, 9)$.

2. To estimate the probability $\mathbf{P}(Y \ge y)$, $y > 0$, we use that

$$\frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{x^2}{2}} \left(\frac{1}{x} - \frac{1}{x^3}\right) \;\le\; \mathbf{P}(X \ge x) \le \frac{1}{x}\,\frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{x^2}{2}}\,,$$

if X has standard normal distribution. Since $\mathbf{P}(Y \ge y) = \mathbf{P}\left(X \ge \frac{y-2}{3}\right)$ for $y > 0$, we obtain

$$\frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{(y-2)^2}{2\cdot 9}} \left(\frac{3}{y-2} - \frac{27}{(y-2)^3}\right) \le \mathbf{P}(Y > y) \le \frac{3}{y-2}\,\frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{(y-2)^2}{2\cdot 9}}\,.$$

3. The expectation of Z is given by:

$$\mathbf{P}(Z) = \mathbf{P}(X^2) = \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{x^2}{2}}\,\mathrm{d}x = \boldsymbol{\sigma}^2(X) = 1,$$

where we have used the formula $\mathbf{P}(\psi(x)) = \int \psi(x)\, f_X(x)\,\mathrm{d}x$. To compute the variance of Z, we use the formula

$$\boldsymbol{\sigma}^2(Z) = \mathbf{P}(Z^2) - \mathbf{P}(Z)^2\,.$$

It remains to compute

$$\begin{aligned}
\mathbf{P}(Z^2) &= \mathbf{P}((X^2)^2) = \mathbf{P}(X^4) \\
&= \int_{-\infty}^{+\infty} x^4 \, \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{x^2}{2}} \, \mathrm{d}x \\
&= \left[-x^3 \, \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{x^2}{2}} \right]_{-\infty}^{+\infty} + 3 \int_{-\infty}^{+\infty} x^2 \, \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{x^2}{2}} \, \mathrm{d}x \\
&= 3.
\end{aligned}$$

4. To compute the c.d.f. F_Z of Z, we proceed as above, i.e.

$$F_Z(z) \;=\; \mathbf{P}(Z \le z) \;=\; \mathbf{P}(X^2 \le z)\,.$$

Since $Z = X^2$ is a non negative random number, we can distinguish 2 cases:

(a) for $z < 0$ we have that $F_Z(z) = 0$;
(b) if $z \ge 0$

$$\begin{aligned}
F_Z(z) &= \mathbf{P}\left(X^2 \le z\right) \\
&= \mathbf{P}\left(-\sqrt{z} \le X \le \sqrt{z}\right) \\
&= \mathbf{P}\left(X \le \sqrt{z}\right) - \mathbf{P}\left(X \le -\sqrt{z}\right) \\
&= \int_{-\infty}^{\sqrt{z}} \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{t^2}{2}} \, \mathrm{d}t - \int_{-\infty}^{-\sqrt{z}} \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{t^2}{2}} \, \mathrm{d}t \\
&= \int_{-\sqrt{z}}^{\sqrt{z}} \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{t^2}{2}} \, \mathrm{d}t\,.
\end{aligned}$$

Finally we get

$$F_Z(z) \;=\; \begin{cases} 0 & z < 0\,, \\ \int_{-\sqrt{z}}^{\sqrt{z}} \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{t^2}{2}} \, \mathrm{d}t & z \ge 0. \end{cases}$$

To compute the density f_Z, we can take the derivative of the c.d.f.. For $z \ge 0$

$$\begin{aligned}
f_Z(z) &= \frac{\mathrm{d}}{\mathrm{d}z} \left(\int_{-\infty}^{\sqrt{z}} \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{t^2}{2}} \, \mathrm{d}t - \int_{-\infty}^{-\sqrt{z}} \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{t^2}{2}} \, \mathrm{d}t \right) \\
&= \frac{1}{2} z^{-\frac{1}{2}}\, n(\sqrt{z}) - \left(-\frac{1}{2} z^{-\frac{1}{2}} \right) n(-\sqrt{z}) \\
&= z^{-\frac{1}{2}} \cdot n(\sqrt{z}) \\
&= z^{-\frac{1}{2}} \, \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{z}{2}}\,.
\end{aligned}$$

We obtain

$$f_Z(z) = \begin{cases} 0 & z < 0\,, \\ \dfrac{1}{\sqrt{2\,\pi}}\, z^{-\frac{1}{2}}\, e^{-\frac{z}{2}} & z \geq 0. \end{cases}$$

Hence Z has Gamma distribution of parameters $\Gamma(\frac{1}{2}, \frac{1}{2})$, i.e. χ^2-distribution of parameter 1. □

Exercise 11.3 Let X be a random number with exponential distribution with parameter $\lambda = 2$.

1. Compute the moments of order n of X, i.e. $\mathbf{P}(X^n)$, $n \in \mathbb{N}$.
2. Consider the family of random numbers $Z_u = e^{uX}$, $u < \lambda$. Given a fixed $u < \lambda$, compute the expectation $\Psi_X(u) = \mathbf{P}(e^{uX})$ of Z_u. The function $\Psi_X(u)$ is called *moment generating function* of X.

Solution 11.3 1. The moment of order $n \in \mathbb{N}$ for X can be computed with the formula

$$\mathbf{P}(\Psi(x)) = \int \Psi(x)\, f_X(x)\,\mathrm{d}x\,,$$

for a given function $\Psi : \mathbb{R} \longrightarrow \mathbb{R}$ such that the integral above exists and is finite. In this case $\Psi(x) = x^n$. We then obtain

$$\begin{aligned} \mathbf{P}(X^n) &= \int_0^{+\infty} x^n\,\lambda\,e^{-\lambda x}\,\mathrm{d}x \\ &= \lambda \int_0^{+\infty} x^n e^{-\lambda x}\,\mathrm{d}x = \lambda \frac{\Gamma(n+1)}{\lambda^{n+1}} \\ &= \frac{n!}{\lambda^n}. \end{aligned}$$

In particular for $n = 1$ we have that $\mathbf{P}(X) = \dfrac{1}{\lambda}$.

2. We compute the expectations of $Z_u = e^{uX}$, $u \in \mathbb{R}$.

$$\begin{aligned} \mathbf{P}(Z_u) &= \mathbf{P}(e^{uX}) \\ &= \int_0^{+\infty} \lambda\,e^{ux}\,e^{-\lambda x}\,\mathrm{d}x \\ &= \int_0^{+\infty} \lambda\,e^{(u-\lambda)x}\,\mathrm{d}x\,. \end{aligned}$$

Note that here u is a given parameter. The integral is well-defined since $u < \lambda$. We obtain that

$$\mathbf{P}(Z_u) = \frac{\lambda}{u-\lambda}\left[e^{(u-\lambda)x}\right]_0^{+\infty} = \frac{\lambda}{u-\lambda}\,. \qquad \square$$

Exercise 11.4 The random number X has uniform distribution on the interval $[-1, 1]$.

(a) Write the density of X.

Let $Z = \log|X|$.

(b) Compute $I(Z)$ e $\mathbf{P}(Z)$.
(c) Compute the c.d.f. and the density of Z.
(d) Calculate $\mathbf{P}(Z < -\frac{1}{2}|X > -\frac{1}{2})$.

Solution 11.4 (a) The density of X is equal to

$$f(x) = \begin{cases} \frac{1}{2} & \text{per } x \in (-1, 1), \\ 0 & \text{otherwise.} \end{cases}$$

(b) The random number X has as set of possible values

$$I(X) = [-1, 1],$$

hence the set of possible values for $Z = \log|X|$ is given by

$$I(Z) = (-\infty, 0]\,.$$

The random number Z is not defined if X assumes the value $0 \in I(X)$. To compute the expectation we can proceed as follows:

$$\mathbf{P}(Z) = \mathbf{P}(Z|X > 0)\mathbf{P}(X > 0) + \mathbf{P}(Z|X < 0)\mathbf{P}(X < 0)$$

$$= \mathbf{P}(\log X|X > 0) \cdot \frac{1}{2} + \mathbf{P}(\log(-X)|X < 0) \cdot \frac{1}{2}\,,$$

where we have used the fact that

$$\mathbf{P}(X > 0) = \mathbf{P}(X < 0) = \frac{1}{2}\,.$$

Verify this by direct computation!

We need only to calculate

$$\mathbf{P}(\log X|X > 0) = \int_0^1 \log x \mathrm{d}x \qquad (11.1)$$
$$= [x \log x - x]_0^1 = -1 \,,$$

$$\mathbf{P}(\log(-X)|X < 0) = \int_{-1}^0 \log(-x)\mathrm{d}x \qquad (11.2)$$
$$= \int_0^1 \log y \mathrm{d}y = -1,$$

hence

$$\mathbf{P}(Z) = \mathbf{P}(\log X) = -1 \,.$$

(c) To compute the c.d.f. of Z, we need to exclude again the value 0. We have

$$F_Z(z) = \mathbf{P}(Z \le z)$$
$$= \mathbf{P}(Z \le z, X > 0) + \mathbf{P}(Z \le z, X < 0) \,.$$

If $z \ge 0$, then $F_Z(z) = 1$. Let $z < 0$. We obtain:

$$F_Z(z) = \mathbf{P}(\log X \le z, X > 0) + \mathbf{P}(\log(-X) \le z, X < 0) \,.$$

We now compute

$$\mathbf{P}(\log X \le z, X > 0) = \mathbf{P}(X \le e^z, X > 0) \qquad (11.3)$$
$$= \mathbf{P}(0 < X \le e^z)$$
$$= \int_0^{e^z} \frac{1}{2}\mathrm{d}x = \frac{1}{2}e^z$$

and

$$\mathbf{P}(\log(-X) \le z, X < 0) = \mathbf{P}(X \ge -e^z, X < 0) \qquad (11.4)$$
$$= \mathbf{P}(-e^z \le X < 0)$$
$$= \int_{-e^z}^0 \frac{1}{2}\mathrm{d}x = \frac{1}{2}e^z \,.$$

Hence

$$F_Z(z) = e^z \quad \text{if } z < 0 \,.$$

The density of Z is given by

$$f_Z(z) = \begin{cases} e^z & \text{for } z < 0, \\ 0 & \text{otherwise}. \end{cases}$$

(d) We evaluate $\mathbf{P}(Z < -\frac{1}{2}|X > -\frac{1}{2})$ by using the formula of the conditional probability:

$$\mathbf{P}\left(Z < -\frac{1}{2}\middle|X > -\frac{1}{2}\right) = \frac{\mathbf{P}\left(Z < -\frac{1}{2}, X > -\frac{1}{2}\right)}{\mathbf{P}\left(X > -\frac{1}{2}\right)},$$

where

$$\mathbf{P}\left(Z < -\frac{1}{2}, X > -\frac{1}{2}\right) = \mathbf{P}\left(\log|X| < -\frac{1}{2}, X > -\frac{1}{2}\right) =$$
$$\mathbf{P}\left(\log X < -\frac{1}{2}, X > 0\right) + \mathbf{P}\left(\log(-X) < -\frac{1}{2}, -\frac{1}{2} < X < 0\right),$$

Here we have used that

$$\left(X > -\frac{1}{2}\right) = (X > 0) + \left(-\frac{1}{2} < X < 0\right).$$

It follows that

$$\mathbf{P}\left(\log X < -\frac{1}{2}, X > 0\right) = \mathbf{P}\left(0 < X < e^{-\frac{1}{2}}\right) = \frac{e^{-\frac{1}{2}}}{2}$$

and furthermore

$$\mathbf{P}\left(\log(-X) < -\frac{1}{2}, -\frac{1}{2} < X < 0\right) = \mathbf{P}\left(X > -e^{-\frac{1}{2}}, -\frac{1}{2} < X < 0\right)$$
$$= \mathbf{P}\left(-\frac{1}{2} < X < 0\right) = \int_{-\frac{1}{2}}^{0} \frac{1}{2}\mathrm{d}x = \frac{1}{4}.$$

Finally

$$\mathbf{P}\left(Z < -\frac{1}{2}|X > -\frac{1}{2}\right) = \frac{1}{2}\left(\frac{1}{\sqrt{e}} + \frac{1}{2}\right).$$

□

Chapter 12
Absolutely Continuous and Multivariate Distributions

Exercise 12.1 Let X be the random number with density

$$f(x) = \begin{cases} Kx^2 & \text{for } -1 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Compute K.
(b) Compute the c.d.f., the expectation and the variance of X.
(c) Let Y be a random number which is stochastically independent and has exponential distribution with parameter $\lambda = 2$. Write the joint density function and the joint c.d.f. of (X, Y).

Solution 12.1 (a) The normalization constant K is such that

$$\int_{-1}^{1} Kx^2 \mathrm{d}x = 1 .$$

Hence

$$K = \frac{1}{\int_{-1}^{1} x^2 \mathrm{d}x} = \frac{3}{2} .$$

(b) The c.d.f. of X is given by

$$F(x) = \mathbf{P}(X \leq x) = \int_{-\infty}^{x} f(t)\mathrm{d}t .$$

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8_12

Hence

$$F(x) = \begin{cases} 0 & \text{for } x \leq 1, \\ \int_{-1}^{x} \frac{3}{2}t^2 \mathrm{d}t = \frac{1}{2}(x^3 + 1) & \text{for } x \in [-1, 1] \\ 1 & \text{for } x \geq 1. \end{cases}$$

Furthermore the expectation of X is equal to

$$\mathbf{P}(X) = \int_{\mathbb{R}} t f(t) \mathrm{d}t = \int_{-1}^{1} \frac{3}{2} x^3 \mathrm{d}x = 0\,.$$

The variance is given by

$$\begin{aligned} \boldsymbol{\sigma}^2(X) &= \mathbf{P}(X^2) - \mathbf{P}(X)^2 \\ &= \mathbf{P}(X^2) = \int_{-1}^{1} x^2 \cdot \frac{3}{2} x^2 \mathrm{d}x \\ &= \frac{3}{2} \int_{-1}^{1} x^4 \mathrm{d}x = \frac{3}{5}\,. \end{aligned}$$

(c) The density of Y is given by

$$g(y) = \begin{cases} 2e^{-2y} & \text{for } y \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

If X and Y are stochastically independent, then the joint density is given by the product of the marginal densities:

$$f(x, y) = f_X(x) g_Y(y) = \begin{cases} 2e^{-2y} \frac{3}{2} x^2 = 3e^{-2y} x^2 & \text{for } x \in [-1, 1] \text{ and } y \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Analogously the joint c.d.f. coincides with the product of the marginal distribution functions:

$$F(x, y) = F_X(x) F_Y(y) = \begin{cases} (1 - e^{-2y}) \dfrac{x^3 + 1}{2} & \text{for } x \in [-1, 1] \text{ and } y \geq 0, \\ 1 - e^{-2y} & \text{for } x > 1 \text{ and } y \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

□

Exercise 12.2 Let (X, Y) be a random vector with uniform distribution on the disk of radius 1 and center at the origin of the axes.

1. Compute the joint density function $f(x, y)$ of (X, Y).
2. What is the marginal density f_X of X?
3. Let $Z = X^2 + Y^2$, compute $\mathbb{P}(\frac{1}{4} \le Z \le 1)$.
4. Compute the c.d.f. and the density of Z.

Solution 12.2 1. Since (X, Y) have uniform distribution on the disk

$$D_1 = \left\{(x, y) : x^2 + y^2 \le 1\right\} .$$

the joint density $f(x, y)$ is constant on D_1 and 0 outside. We obtain that

$$f(x, y) = \begin{cases} \frac{1}{\text{area } D_1} = \dfrac{1}{\pi} & \text{for } (x, y) \in D_1 , \\ 0 & \text{otherwise.} \end{cases}$$

The density domain is shown in Fig. 12.1.
The value of the density f on D_1 can be determined by imposing that

$$1 = \int\int_{\mathbb{R}^2} f(x, y)\,\mathrm{d}x\,\mathrm{d}y = \int\int_{D_1} c\,\mathrm{d}x\,\mathrm{d}y,$$

i.e.

$$c = \frac{1}{\int\int_{D_1} \mathrm{d}x\,\mathrm{d}y} = \frac{1}{\text{area } D_1} = \frac{1}{\pi}.$$

2. To compute the marginal density of X, we distinguish 4 cases as follows.

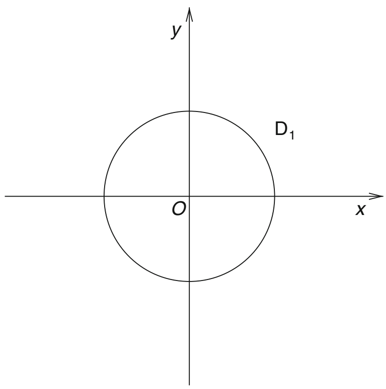


Fig. 12.1 Representation of the area D_1 on the plane

Fig. 12.2 Case $0 \le x \le 1$

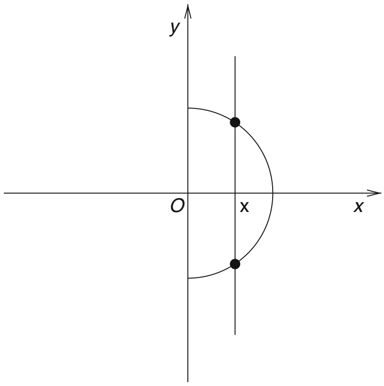
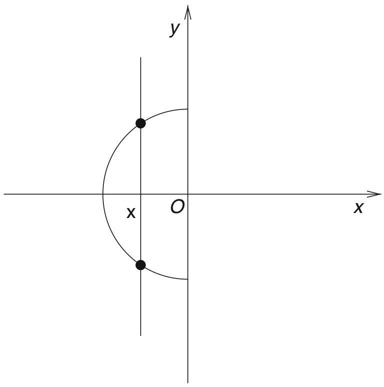


Fig. 12.3 Case $-1 \le x \le 0$



- Case $x > 1$: $f_X(x) = 0$.
- Case $1 \ge x \ge 0$: set the x coordinate; y varies along the line orthogonal to the x-axis and passing through $(x, 0)$. The extremes are the points where this line intersects the graph of D_1 as shown in Fig. 12.2. We obtain:

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} f(x,t)\,\mathrm{d}t = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\,\mathrm{d}t = \frac{2\sqrt{1-x^2}}{\pi}\,.$$

- Case $-1 \le x < 0$: by symmetry, we obtain as shown in Fig. 12.3, that

$$f_X(x) = \frac{2\sqrt{1-x^2}}{\pi}\,.$$

- Case $x < -1$: also here we have $f_X(x) = 0$.

Summing up:

$$f_X(x) = \begin{cases} \frac{2\sqrt{1-x^2}}{\pi} & \text{for } x \in [-1, 1] \\ 0 & \text{otherwise.} \end{cases}$$

3. Let $Z = X^2 + Y^2$; to compute $\mathbf{P}\left(\frac{1}{4} \leq Z \leq 1\right)$ is equivalent to calculate the probability that the random vector (X, Y) belongs to the region A of the plane between the disk with center O and radius $\frac{1}{2}$ and the disk with center O and radius 1, i.e.

$$\mathbf{P}\left(\frac{1}{4} \leq Z \leq 1\right) = \mathbf{P}\left(\frac{1}{4} \leq X^2 + Y^2 \leq 1\right).$$

Hence

$$\mathbf{P}\left(\frac{1}{4} \leq X^2 + Y^2 \leq 1\right) = \int\int_A f(x, y)\, \mathrm{d}x\, \mathrm{d}y.$$

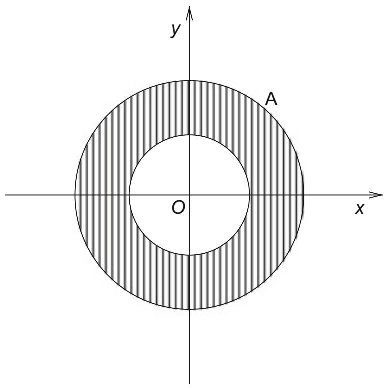
We can compute this integral by passing to the polar coordinates

$$x = \rho \cos\theta, \qquad y = \rho \sin\theta.$$

To perform the change of variables in the integral, we need to take account of the absolute value of the Jacobian determinant (Fig. 12.4). In the case of polar coordinates, this is equal to

$$|J| = \rho.$$

Fig. 12.4 Area of the region $\left\{(x, y) | \frac{1}{4} \leq x^2 + y^2 \leq 1\right\}$



It follows that

$$\int\int_{A(x,y)} f(x, y)\,\mathrm{d}x\,\mathrm{d}y = \int\int_{A(\rho,\theta)} f(\rho, \theta)\,\mathrm{d}\rho\,\mathrm{d}\theta$$
$$= \int_0^{2\pi} \mathrm{d}\theta \int_{\frac{1}{2}}^{1} \frac{1}{\pi}\,\mathrm{d}\rho = \int_{\frac{1}{2}}^{1} 2\rho\,\mathrm{d}\rho = \left[\rho^2\right]_{\frac{1}{2}}^{1} = \frac{3}{4}\,.$$

4. To compute the c.d.f. $F_Z(z)$ of Z we use again spherical symmetry.
   - $z < 0$: In this case $F_Z(z) = 0$.
   - $1 \geq z \geq 0$:

$$\begin{aligned} F_Z(z) &= \mathbf{P}(Z \leq z) \\ &= \mathbf{P}(X^2 + Y^2 \leq z) \\ &= \int\int_{D_z} f(x, y)\,\mathrm{d}x\,\mathrm{d}y, \end{aligned}$$

   where $D_z = \{(x, y) : x^2 + y^2 \leq z\}$. It follows that

$$F_Z(z) = \int_0^{2\pi} \int_0^{\sqrt{z}} \frac{1}{\pi}\,\rho\,\mathrm{d}\rho\,\mathrm{d}\theta = \int_0^{\sqrt{z}} 2\rho\,\mathrm{d}\rho = \left[\rho^2\right]_0^{\sqrt{z}} = z\,.$$

   - $z > 1$: In this case $F_Z(z) = P\left(X^2 + Y^2 \leq z\right) = 1$.

   Summing up:

$$F_Z(z) = \begin{cases} 0 & \text{for } z < 0, \\ z & \text{for } 0 \leq z < 1, \\ 1 & \text{for } z > 1\,. \end{cases}$$

   The density function of Z is given by

$$f_Z(z) = \begin{cases} 1 & \text{for } 0 \leq z \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

The random number Z has therefore a uniform density in [0, 1]. □

Exercise 12.3 Let (X, Y) be a random vector with joint density

$$f(x, y) = \begin{cases} k\,xy & (x, y) \in T, \\ 0 & \text{otherwise.} \end{cases}$$

where $T = \{(x, y) \in \mathbb{R}^2 |\ 0 \le y \le -x + 2,\ 0 < x < 2\}$.

1. Compute the normalization constant k.
2. Compute the probability $\mathbf{P}(X > 1, Y < \frac{1}{2})$ and the conditional probability $\mathbf{P}(X > 1 | Y < \frac{1}{2})$.
3. Let $Z = X + Y$. Compute the probability that $\mathbf{P}(0 < Z < 1)$.
4. Compute the p.d.f. and the density of Z.

Solution 12.3 1. To compute the normalization constant k we impose that

$$\int\int_{\mathbb{R}^2} f(x, y)\, \mathrm{d}x\, \mathrm{d}y = 1\,.$$

The integral of f can be computed by using Fubini-Tonelli Theorem:

$$\int\int_{\mathbb{R}^2} f(x, y)\, \mathrm{d}x\, \mathrm{d}y = k \int_0^2 x \int_0^{-x+2} y\, \mathrm{d}y\, \mathrm{d}x$$

$$= k \int_0^2 x \left[\frac{y^2}{2}\right]_0^{-x+2} \mathrm{d}x = k \int_0^2 x\, \frac{1}{2}\, (2 - x)^2\, \mathrm{d}x$$

$$= \frac{k}{2} \int_0^2 (4x - 4x^2 + x^3)\, \mathrm{d}x = \frac{k}{2} \left[2x^2 - \frac{4}{3}x^3 + \frac{1}{4}x^4\right]_0^2 = \frac{2}{3}\, k\,.$$

It follows that

$$k = \frac{3}{2}\,.$$

2. The probability $\mathbf{P}(X > 1, Y < \frac{1}{2})$ is given by the integral of the joint density on the region D given by the intersection

$$D = \{(x, y) \in \mathbb{R}^2 \,|\, x > 1,\, y < \frac{1}{2}\} \cap T\,,$$

see Figs. 12.5 and 12.6.
To find the extremes, it is easier this time to fix y and let x vary. The extremes are given by the intersection of the border of D with the line passing in $(0, y)$ which is parallel to the x-axis, as we can see in Fig. 12.7.

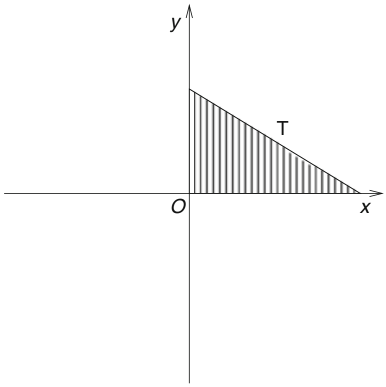


Fig. 12.5 Representation of the area T on the plane

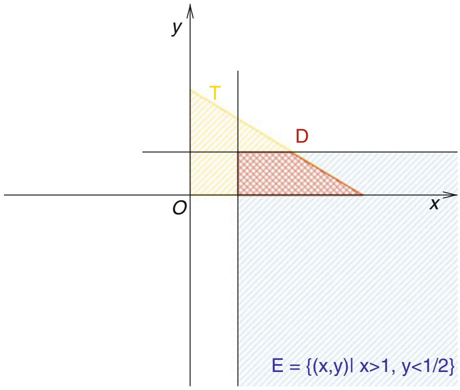


Fig. 12.6 Representation of the area D on the plane

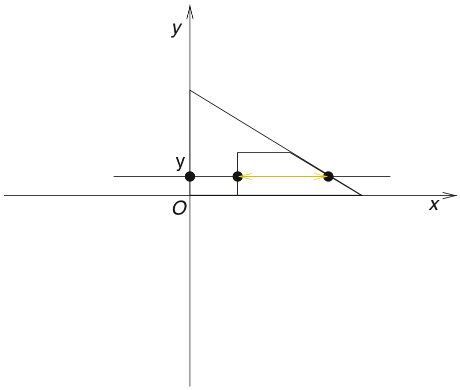


Fig. 12.7 Extremes of variation of x

$$
\begin{aligned}
\mathbf{P}(X > 1, Y < \frac{1}{2}) &= \int\int_D f(x, y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_0^{\frac{1}{2}} \left(y \int_1^{-y+2} \frac{3}{2}\,x\,\mathrm{d}x \right) \mathrm{d}y \\
&= \int_0^{\frac{1}{2}} y \left[\frac{3}{4}\,x^2 \right]_1^{-y+2} \mathrm{d}y \\
&= \frac{3}{4} \int_0^{\frac{1}{2}} y\,(3 - 4y + y^2)\,\mathrm{d}y \\
&= \frac{3}{4} \left[\frac{3}{2}\,y^2 - \frac{4}{3}\,y^3 + \frac{1}{4}\,y^4 \right]_0^{\frac{1}{2}} \\
&= \frac{43}{256}\,.
\end{aligned}
$$

The conditional probability $\mathbf{P}(X > 1|Y < \frac{1}{2})$ can be obtained as follows:

$$
\mathbf{P}(X > 1|Y < \frac{1}{2}) \;=\; \frac{\mathbf{P}(X > 1, Y < \frac{1}{2})}{\mathbf{P}(Y < \frac{1}{2})}\,.
$$

We simply need to compute $\mathbf{P}(Y < \frac{1}{2})$. To this purpose we do not necessarily need to know the marginal density of Y. This probability is given by the integral of the *joint* probability $f(x, y)$ on the domain D_1 given by the intersection of $E_1 = \{(x, y) \in \mathbb{R}^2 |\ y < \frac{1}{2}\}$ and of T, i.e.

$$
D_1 \;=\; E_1 \cap T\,,
$$

see Fig. 12.8.
We can obtain the probability that Y is less than $\frac{1}{2}$ by computing the joint probability that there are no restrictions on X and that Y is less than $\frac{1}{2}$. We obtain:

$$
\begin{aligned}
\mathbf{P}(Y < \frac{1}{2}) &= \int\int_{D_1} f(x, y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_0^{\frac{1}{2}} \frac{3}{2}\,y \int_0^{-y+2} x\,\mathrm{d}x\,\mathrm{d}y = \frac{3}{4} \int_0^{\frac{1}{2}} y\,(4 - 4y + y^2)\,\mathrm{d}y \\
&= \frac{3}{4} \left[2y^2 - \frac{4}{3} y^3 + \frac{1}{4} y^4 \right]_0^{\frac{1}{2}} = \frac{67}{256}\,.
\end{aligned}
$$

The conditional probability is then given by

$$
\mathbf{P}(X > 1|Y < \frac{1}{2}) \;=\; \frac{\mathbf{P}(X > 1, Y < \frac{1}{2})}{\mathbf{P}(Y < \frac{1}{2})} \;=\; \frac{43}{67}\,.
$$

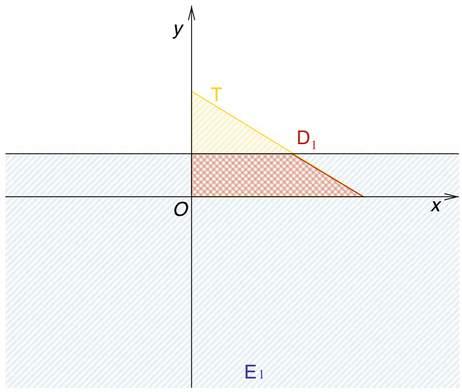


Fig. 12.8 Representation of the area D_1 on the plane

3. We now consider the random number $Z = X + Y$. To compute the probability $\mathbf{P}(0 < Z < 1)$ we can use the joint density of (X, Y). We obtain

$$\begin{aligned}\mathbf{P}(0 < Z < 1) &= \mathbf{P}(0 < X + Y < 1)\\ &= \mathbf{P}(-Y < X < 1 - Y)\\ &= \mathbf{P}(0 < X < 1 - Y).\end{aligned}$$

   Note that in this case X and Y are both positive, hence the condition $X > -Y$ reduces to $X > 0$. In Fig. 12.9 we represent the region where the integral of the joint density of X, Y must be calculated to obtain $\mathbf{P}(0 < Z < 1)$.

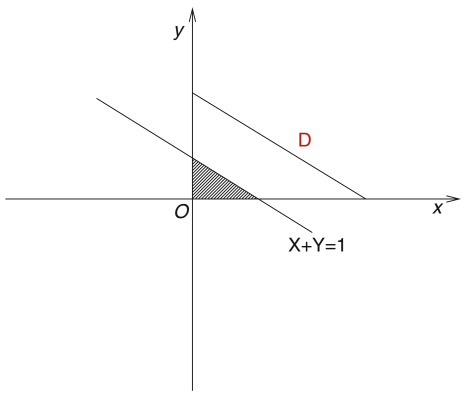


Fig. 12.9 Region where $0 < Z < 1$

$$\begin{aligned}\mathbf{P}(0 < X < 1 - Y) &= \int_0^1 \frac{3}{2}\, y \int_0^{1-y} x\,\mathrm{d}x\,\mathrm{d}y \\ &= \frac{3}{4}\int_0^1 y\,(1-y)^2\,\mathrm{d}y \\ &= \frac{3}{4}\int_0^1 (y - 2y^2 + y^3)\,\mathrm{d}y \\ &= \frac{3}{4}\left[\frac{1}{2}\,y^2 - \frac{2}{3}\,y^3 + \frac{1}{4}\,y^4\right]_0^1 \\ &= \frac{1}{16}\,.\end{aligned}$$

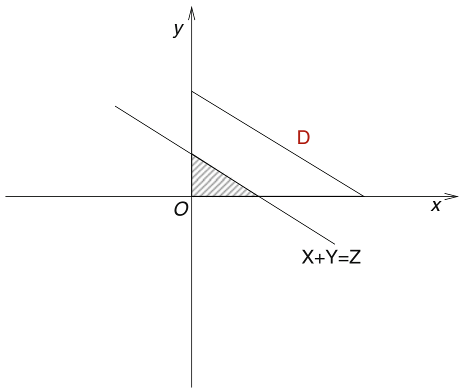
4. The m.d.f. of Z is given by

$$F_Z(z) \;=\; \mathbf{P}(Z \le z) \;=\; \mathbf{P}(X + Y \le z) \;=\; \mathbf{P}(X \le z - Y)\,.$$

If we consider the line $x + y - z = 0$, the distribution function of Z is given by the integral of the joint density of X, Y on the region R delimited by this line on T, as shown by Fig. 12.10.
We obtain:

- for $z < 0$: $\mathbf{P}(Z < z) \;=\; 0$;
- for $z > 2$: $\mathbf{P}(Z < z) \;=\; 1$;
- for $0 \le z \le 2$:

Fig. 12.10 Region R



$$
\begin{aligned}
\mathbf{P}(Z < z) &= \int_0^z \frac{3}{2}\, y \int_0^{z-y} x \,\mathrm{d}x \,\mathrm{d}y \\
&= \frac{3}{4} \int_0^z y\,(z-y)^2 \,\mathrm{d}y \\
&= \frac{3}{4} \int_0^z (z^2 y - 2zy^2 + y^3) \,\mathrm{d}y \\
&= \frac{3}{4} \left[\frac{1}{2}\, z^2 y^2 - \frac{2}{3}\, zy^3 + \frac{1}{4}\, y^4 \right]_0^z \\
&= \frac{3}{4} \left(\frac{1}{2}\, z^4 - \frac{2}{3}\, z^4 + \frac{1}{4}\, z^4 \right) \\
&= \frac{z^4}{16}\,.
\end{aligned}
$$

Summing up:

$$
F_Z(z) \;=\; \begin{cases} 0 & \text{for } z < 0, \\[2ex] \dfrac{z^4}{16} & \text{for } 0 \le z \le 2, \\[2ex] 1 & \text{for } z > 2\,. \end{cases}
$$

The density can be obtained by deriving the distribution function

$$
f_Z(z) \;=\; \begin{cases} 0 & \text{for } z < 0, \\[2ex] \dfrac{z^3}{4} & \text{for } 0 \le z \le 2, \\[2ex] 0 & \text{for } z > 2, \end{cases}
$$

or by means of the formula

$$
f_Z(z) \;=\; \int_{\mathbb{R}} f(x, z-x) dx\,. \qquad \square
$$

Exercise 12.4 Let X, Y be two random numbers with joint distribution function

$$
f(x, y) = \begin{cases} Kx & \text{for } y \le x \le y+1, \quad 0 \le y \le 2, \\ 0 & \text{otherwise.} \end{cases}
$$

(a) Compute K.
(b) Compute the m.d.f. and the expectation of X.

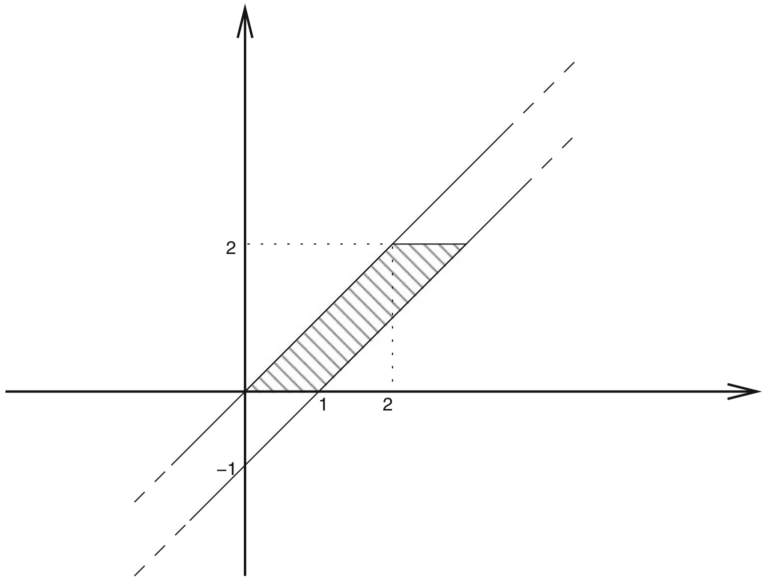


Fig. 12.11 Region R of definition of the density

(c) Compute $\mathbf{cov}(X, Y)$.
(d) Compute $\mathbf{P}(0 < X - Y < 1)$.

Solution 12.4 (a) As in previous exercises, first we draw the picture of the region R of definition of the joint density, as shown by Fig. 12.11.

Since the integral of a density must be equal 1, the constant of normalization is given by

$$K = \frac{1}{\int\int_{\mathbb{R}^2} x\mathrm{d}x\mathrm{d}y}\,,$$

where

$$\begin{aligned}\int\int_{\mathbb{R}^2} x\mathrm{d}x\mathrm{d}y &= \int_0^2 \mathrm{d}y \int_y^{y+1} x\mathrm{d}x \\ &= \int_0^2 \left[\frac{x^2}{2}\right]_y^{y+1} \mathrm{d}y \\ &= \int_0^2 \left(\frac{(y+1)^2}{2} - \frac{y^2}{2}\right) \mathrm{d}y \\ &= \frac{1}{6}\left[(y+1)^3 - y^3\right]_0^2 = 3\,.\end{aligned}$$

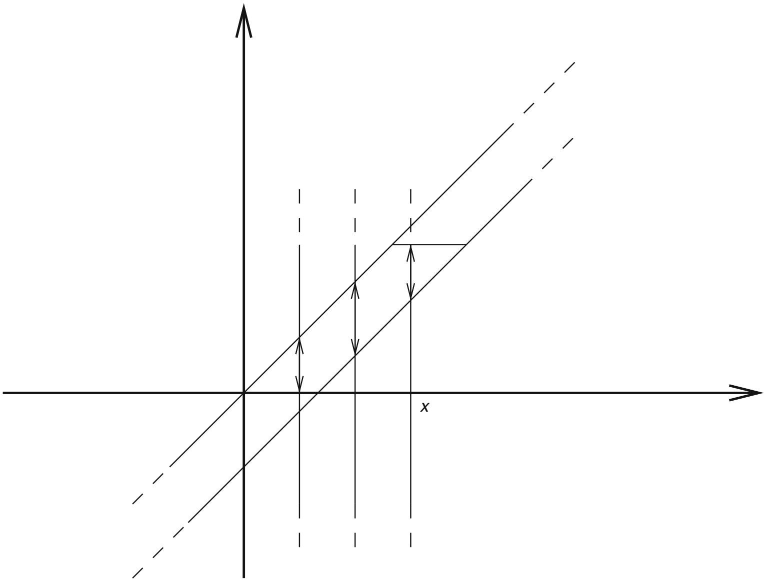


Fig. 12.12 Extremes of variation y

We conclude that $K = \frac{1}{3}$.

(b) To compute the marginal density of X we apply the formula

$$f_X(x) = \int_{\mathbb{R}} f(x, y)\mathrm{d}y\,.$$

To find the extremes of integration, we apply the general method as shown in Fig. 12.1.

We have to pay attention, since the expressions for the extremes of integration vary if $0 < x < 1,\ 1 < x < 2,\ 2 < x < 3$ (see Fig. 12.12).

We have that if $0 < x < 1$, then y varies between the lines

$$y = 0 \quad \text{e} \quad y = x\,.$$

If $1 < x < 2$, then y varies between the lines

$$y = x - 1 \quad \text{e} \quad y = x\,.$$

If $2 < x < 3$, then y varies between

$$y = x - 1 \quad \text{and} \quad y = 2.$$

- For $0 < x < 1$:

$$f_X(x) = \int_0^x \frac{1}{3} x \mathrm{d}y = \frac{1}{3} x^2 .$$

- For $1 < x < 2$:

$$f_X(x) = \int_{x-1}^x \frac{1}{3} x \mathrm{d}y = \frac{1}{3} x .$$

- For $2 < x < 3$:

$$f_X(x) = \int_{x-1}^2 \frac{1}{3} x \mathrm{d}y = \frac{1}{3} x(3 - x) .$$

Summing up:

$$f_X(x) = \begin{cases} \frac{1}{3} x^2 & \text{for } 0 < x < 1, \\ \frac{1}{3} x & \text{for } 1 < x < 2, \\ \frac{1}{3} x(3 - x) & \text{for } 2 < x < 3, \\ 0 & \text{otherwise.} \end{cases}$$

We now verify that $f_X(x)$ is a probability density. We need to have that

$$\int_{\mathbb{R}} f_X(x) \mathrm{d}x = 1,$$

Indeed

$$\begin{aligned} \int_0^1 \frac{1}{3} x^2 \mathrm{d}x + \int_1^2 \frac{1}{3} x \mathrm{d}x + \int_2^3 \frac{1}{3} x(3 - x) \mathrm{d}x &= \\ = \left[\frac{1}{9} x^3 \right]_0^1 + \left[\frac{1}{6} x^2 \right]_1^2 + \left[\frac{x^2}{2} - \frac{x^3}{9} \right]_2^3 & \\ = 1 . & \end{aligned}$$

The expectation of X is given by:

$$\begin{aligned}
\mathbf{P}(X) &= \int_{\mathbb{R}} x f(x) \mathrm{d}x \\
&= \int_0^1 x \frac{1}{3} x^2 \mathrm{d}x + \int_1^2 x \frac{1}{3} x \mathrm{d}x + \int_2^3 x \frac{1}{3} x(3-x) \mathrm{d}x \\
&= \left[\frac{1}{12} x^4 \right]_0^1 + \left[\frac{1}{9} x^3 \right]_1^2 + \left[\frac{x^3}{3} - \frac{x^4}{12} \right]_2^3 \\
&= \frac{16}{9} .
\end{aligned}$$

(c) The covariance $\mathbf{cov}(X, Y)$ is given by:

$$\mathbf{cov}(X, Y) = \mathbf{P}(XY) - \mathbf{P}(X)\mathbf{P}(Y),$$

where

$$\begin{aligned}
\mathbf{P}(XY) &= \int \int_{\mathbb{R}^2} xy f(x, y) \mathrm{d}x \mathrm{d}y = \\
&= \int_0^2 \mathrm{d}y \int_y^{y+1} xy \frac{1}{3} x \mathrm{d}x = \int_0^2 \frac{1}{9} y \left[(y+1)^3 - y^3 \right] \mathrm{d}y \\
&= \left[\frac{1}{12} y^4 + \frac{1}{9} y^3 + \frac{1}{18} y^2 \right]_0^2 = \frac{22}{9} .
\end{aligned}$$

To compute the expectation of Y, we do not need to compute the marginal distribution of Y. In fact it holds that

$$\begin{aligned}
\mathbf{P}(Y) &= \int_{\mathbb{R}} y f_Y(y) \mathrm{d}y \\
&= \int_{\mathbb{R}} \left[y \int_{\mathbb{R}} f(x, y) \mathrm{d}x \right] \mathrm{d}y \\
&= \int \int_{\mathbb{R}} y f(x, y) \mathrm{d}x \mathrm{d}y .
\end{aligned}$$

Hence

$$\begin{aligned}
\mathbf{P}(Y) &= \int \int_{\mathbb{R}} y f(x, y) \mathrm{d}x \mathrm{d}y \\
&= \int_0^2 \mathrm{d}y \int_y^{y+1} \frac{1}{3} xy \mathrm{d}x \\
&= \int_0^2 y \frac{1}{6} \left[(y+1)^2 - y^2 \right] \mathrm{d}y \\
&= \left[\frac{y^3}{6} + \frac{y^2}{12} \right]_0^2 = \frac{5}{3} .
\end{aligned}$$

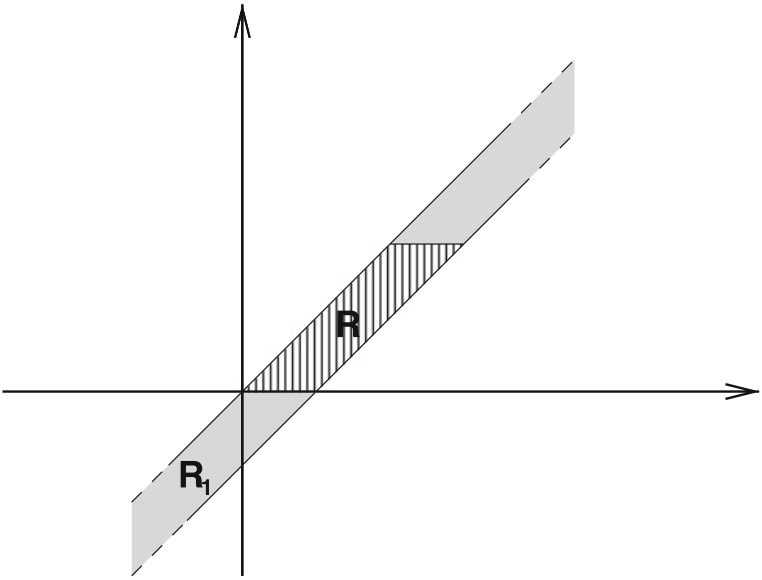


Fig. 12.13 The region R_1

We obtain

$$\mathbf{cov}(X, Y) = \frac{22}{9} - \frac{11}{12} \times \frac{5}{3} = \frac{11}{12},$$

i.e. X and Y are positively correlated.

(d) To compute $\mathbf{P}(0 < X - Y < 1)$, we note that

$$\mathbf{P}(0 < X - Y < 1) = \mathbf{P}(Y < X < Y + 1) = 1,$$

since the region

$$R_1 = \{(x, y) |\ y < x < y + 1\}$$

contains entirely the domain of definition of the density, see Fig. 12.13. □

Exercise 12.5 Let X, Y be two stochastically independent random numbers with the following marginal density:

$$f(x) = \begin{cases} K(x^3 - 1) & \text{for } 1 \leq x \leq 2, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Compute K.
(b) Compute the joint c.d.f., the expectation, the variance and the covariance of X and Y.

(c) Let $Z = X^2$. Compute the c.d.f., the expectation and the variance of Z.
(d) Compute the correlation coefficients $\rho(X, Z)$, $\rho(X + Y, Z)$.

Solution 12.5 (a) Since K is the normalization constant, we obtain

$$K = \frac{1}{\int_0^2 (x^3 - 1)\mathrm{d}x} = \frac{1}{\left[\frac{x^4}{4} - x\right]_1^2} = \frac{4}{11}\,.$$

(b) Since the random numbers X and Y are stochastically independent, their joint c.d.f. is given by the product of the marginal c.d.f.'s:

$$F(x, y) = \mathbf{P}(X \le x, Y \le y) = F_X(x)F_Y(y)\,.$$

It is sufficient to compute

$$F(x) = \mathbf{P}(X \le x) = \int_{-\infty}^{x} f(t)\mathrm{d}t\,.$$

We obtain

$$F_X(x) = \begin{cases} 0 & \mathrm{x} < 1\,, \\ \int_1^x \frac{4}{11}(t^3 - 1)\mathrm{d}t = \frac{4}{11}\left(\frac{x^4}{4} - x + \frac{3}{4}\right) & \mathrm{x} \in [1, 2], \\ 1 & \mathrm{x} \ge 2. \end{cases}$$

Hence

$$F(x, y) = \begin{cases} 0 & \text{for } x < 1 \text{ or } y < 1, \\ \left(\frac{4}{11}\right)^2 (x^4 - x + \frac{3}{4})(y^4 - y + \frac{3}{4}) & \text{for } (x, y) \in [1, 2] \times [1, 2], \\ \frac{4}{11}(x^4 - x + \frac{3}{4}) & \text{for } x \in [1, 2],\ y > 2, \\ \frac{4}{11}(y^4 - y + \frac{3}{4}) & \text{for } x > 2,\ y \in [1, 2], \\ 1 & \text{for } x > 2,\ y > 2\,. \end{cases}$$

Since X and Y are stochastically independent, we have immediately

$$\mathbf{cov}(X, Y) = 0\,.$$

Finally, we compute the expectation and the variance as follows:

1. Expectation

$$\begin{aligned}\mathbf{P}(X) = \mathbf{P}(Y) &= \int_{\mathbb{R}} t f(t)\mathrm{d}t \\ &= \frac{4}{11}\int_1^2 t(t^3-1)\mathrm{d}t = \frac{4}{11}\left[\frac{t^5}{5} - \frac{t^2}{2}\right]_1^2 = \frac{94}{55}\,.\end{aligned}$$

2. For the variance, we need first to calculate

$$\begin{aligned}\mathbf{P}(X^2) = \mathbf{P}(Y^2) &= \frac{4}{11}\int_1^2 t^2(t^3-1)\mathrm{d}t \\ &= \frac{4}{11}\left[\frac{t^6}{6} - \frac{t^3}{3}\right]_1^2 = \frac{98}{33}\,.\end{aligned}$$

   Hence

$$\sigma^2(X) = \mathbf{P}(X^2) - \mathbf{P}(X)^2 = \frac{98}{33} - \left(\frac{94}{55}\right)^2\,.$$

(c) We now compute the c.d.f. of $Z = X^2$:

$$F_Z(z) = \mathbf{P}(Z \le z) = \mathbf{P}(X^2 \le z)\,.$$

For $z < 1$, we have immediately $F_Z(z) = 0$. For $1 \le z < 4$, i.e. for $1 \le \sqrt{z} < 2$, we have

$$F_Z(z) = \mathbf{P}(Z \le z) = \mathbf{P}(X^2 \le z) = \mathbf{P}(-\sqrt{z} \le X \le \sqrt{z}) =$$
$$\frac{4}{11}\int_1^{\sqrt{z}}(t^3-1)\mathrm{d}t = \frac{4}{11}\left[\frac{t^4}{4} - t\right]_1^{\sqrt{z}} = \frac{1}{11}(z^2 - 4\sqrt{z} + 3)\,.$$

For $z \ge 4$, $F_Z(z) = 1$. Summing up:

$$F_Z(z) = \begin{cases} 0 & \text{for } z < 1, \\ \frac{1}{11}(z^2 - 4\sqrt{z} + 3) & \text{for } z \in [1,4], \\ 1 & \text{for } z \ge 4. \end{cases}$$

The expectation of Z coincides with the expectation of X^2, i.e.

$$\mathbf{P}(Z) = \mathbf{P}(X^2) = \frac{98}{33}\,.$$

To compute the variance, we note that

$$\mathbf{P}(Z^2) = \mathbf{P}(X^4) = \int_1^2 \frac{4}{11} t^4 (t^3 - 1) \mathrm{d}t =$$
$$\frac{4}{11} \left[\frac{t^8}{8} - \frac{t^5}{5} \right]_1^2 = \frac{1027}{110} .$$

Hence the variance is given by

$$\sigma^2(Z) = \mathbf{P}(Z^2) - \mathbf{P}(Z)^2 = \frac{1027}{110} - \left(\frac{98}{33}\right)^2 .$$

(d) We now compute the correlation coefficient $\rho(X, Z)$:

$$\rho(X, Z) = \frac{\mathbf{cov}(X, Z)}{\sigma(X)\sigma(Z)} .$$

Since we have already determined $\sigma^2(X)$, $\sigma^2(Z)$, we have immediately the mean square deviations $\sigma(X)$, $\sigma(Z)$. It remains to compute

$$\begin{aligned} \mathbf{cov}(X, Z) &= \mathbf{P}(XZ) - \mathbf{P}(X)\mathbf{P}(Z) \\ &= \mathbf{P}(X^3) - \mathbf{P}(X)\mathbf{P}(Z) \\ &= \frac{4}{11} \int_1^2 t^3(t^3 - 1)\mathrm{d}t - \frac{94}{55} \times \frac{98}{33} \\ &= \frac{4}{11} \left[\frac{t^7}{7} - \frac{t^4}{4} \right]_1^2 - \frac{94}{55} \times \frac{98}{33} = \frac{403}{77} - \frac{94}{55} \times \frac{98}{33} . \end{aligned}$$

Finally to obtain the correlation coefficient $\rho(X + Y, Z)$ we simply note that

$$\rho(X + Y, Z) = \rho(X, Z) + \rho(Y, Z) = \rho(X, Z) ,$$

since X and Y (hence Z and Y) are stochastically independent. □

Exercise 12.6 The random numbers X and Y are stochastically independent. The probability density $f_X(x)$ of X is given by:

$$f_X(x) = \begin{cases} 2x & \text{for } 0 \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

while the probability density of Y is given by

$$f_Y(y) = \begin{cases} e^{-y} & \text{for } y \geq 0, \\ 0 & \text{otherwise} . \end{cases}$$

(a) Compute $\mathbf{P}(X)$, $\mathbf{P}(Y)$, $\boldsymbol{\sigma}^2(X)$, $\boldsymbol{\sigma}^2(Y)$.
(b) Determine the joint c.d.f. and the joint density of (X, Y).
(c) Let $Z = X + Y$. Compute $\mathbf{P}(Z)$, $\boldsymbol{\sigma}^2(Z)$ the c.d.f. and the density of Z.

Solution 12.6 (a) We compute the first moments of X and Y:

$$\mathbf{P}(X) = \int_{\mathbb{R}} x f_X(x)\mathrm{d}x = \int_0^1 2x^2\mathrm{d}x = \frac{2}{3};$$
$$\boldsymbol{\sigma}^2(X) = \mathbf{P}(X^2) - \mathbf{P}(X)^2 = \int_0^1 2x^3\mathrm{d}x - \frac{4}{9} = \frac{1}{18}.$$

The random number Y has exponential density of parameter $\lambda = 1$, hence we can immediately write

$$\mathbf{P}(Y) = \frac{1}{\lambda} = 1, \qquad \boldsymbol{\sigma}^2(Y) = \frac{1}{\lambda^2} = 1.$$

(b) The random numbers X and Y are stochastically independent, hence their joint density is equal to

$$f(x, y) = f_X(x) f_Y(y),$$

i.e.

$$f(x, y) = \begin{cases} 2xe^{-y} & \text{for } 0 \le x \le 1 \text{ and } y \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$

We compute the joint c.d.f.

$$F(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(s, t)\mathrm{d}s\mathrm{d}t$$

after having identified the domain D of definition of the joint density as shown in Fig. 12.14.
We obtain that

$$F(x, y) = \begin{cases} \int_0^x \int_0^y 2se^{-t}\mathrm{d}s\mathrm{d}t = x^2(1 - e^{-y}) & \text{for } 0 \le x \le 1 \; e \; y \ge 0, \\ \int_0^1 \int_0^y 2se^{-t}\mathrm{d}s\mathrm{d}t = 1 - e^{-y} & \text{for } x > 1 \; e \; y \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$

(c) Consider now $Z = X + Y$. To compute $\mathbf{P}(Z)$ e $\boldsymbol{\sigma}^2(Z)$ we use:

(i) the linearity property of the expectation:

$$\mathbf{P}(Z) = \mathbf{P}(X) + \mathbf{P}(Y) = \frac{2}{3} + 1 = \frac{5}{3};$$

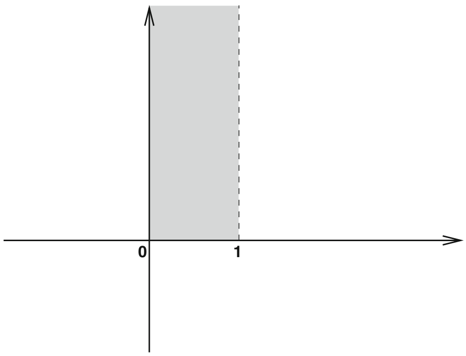


Fig. 12.14 The domain D of definition of the joint density

(ii) the formula for the variance of the sum of 2 random numbers:

$$\begin{aligned}\boldsymbol{\sigma}^2(Z) &= \boldsymbol{\sigma}^2(X+Y) = \boldsymbol{\sigma}^2(X) + \boldsymbol{\sigma}^2(Y) + 2\mathbf{cov}(X,Y)\\ &= \boldsymbol{\sigma}^2(X) + \boldsymbol{\sigma}^2(Y) = \frac{19}{18}\,.\end{aligned}$$

To compute the distribution function of $Z = X + Y$, we use the fact that

$$\begin{aligned}F_Z(z) = \mathbf{P}(Z \le z) &= \mathbf{P}(X + Y \le z)\\ &= \mathbf{P}(Y \le z - X)\\ &= \int\int_{\mathcal{D}_z} f(s,t)\mathrm{d}s\mathrm{d}t\,,\end{aligned}$$

where for all fixed z, $\mathcal{D}_z$ is the region of the plane determined by the intersection of the domain D of definition of the density and of the semi-plane

$$\mathcal{S}_z = \{(x, y)|y \le z - x\}.$$

Figures 12.15 and 12.16 show the region intersected by $\mathcal{S}_z$ on D when z varies. We obtain that:

(i) for $z < 0$, $F_z(z) = 0$;
(ii) for $0 < z < 1$,

$$\begin{aligned}F_Z(z) &= \int_0^z 2x \int_0^{z-x} e^{-y}\mathrm{d}y\mathrm{d}x\\ &= \int_0^z 2x\left(1 - e^{-(z-x)}\right)\mathrm{d}x = z^2 + 2(1-z) - 2e^{-z}\,;\end{aligned}$$

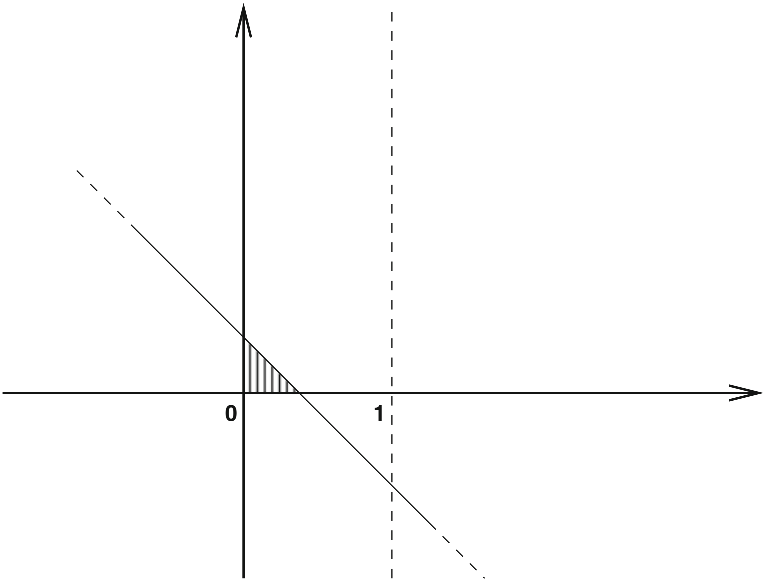
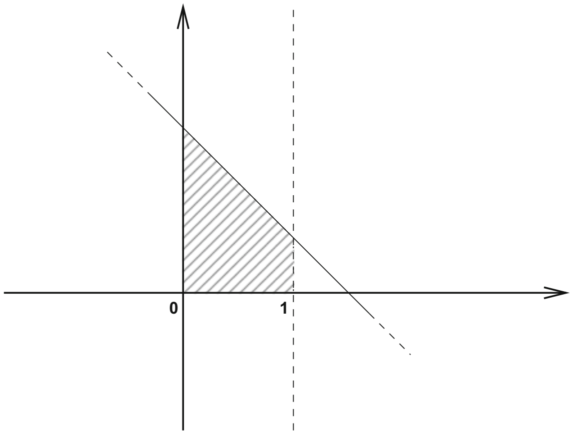


Fig. 12.15 Case $0 < z < 1$

Fig. 12.16 Case $z > 1$



(iii) for $z > 1$

$$F_Z(z) = \int_0^1 \int_0^{z-x} 2xe^{-y}\mathrm{d}y\mathrm{d}x = \int_0^1 2x\left(1 - e^{-(z-x)}\right)\mathrm{d}x = 1 - 2e^{-z}\,.$$

We obtain the density of Z by deriving the c.d.f. of Z, i.e.:

$$f_Z(z) = \begin{cases} 2z - 2 + 2e^{-z} & \text{for } 0 \le z < 1, \\ 2e^{-z} & \text{for } z > 1, \\ 0 & \text{otherwise.} \end{cases}$$

□

Exercise 12.7 The random numbers X and Y have bidimensional Gaussian density

$$p(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}\,.$$

Let $U = 2X + 3Y$ and $V = X - Y$. Compute:

1. The covariance matrix of U and V.
2. The joint density of U and V.

Solution 12.7 1. We compute the covariance matrix of U and V:

$$C = \begin{pmatrix} \boldsymbol{\sigma}^2(U) & \mathbf{cov}(U, V) \\ \mathbf{cov}(U, V) & \boldsymbol{\sigma}^2(V) \end{pmatrix}.$$

In order to compute C we use the formula of the variance of the sum of 2 random numbers and the bilinearity of the covariance:

- $\boldsymbol{\sigma}^2(U)$

$$\begin{aligned} \boldsymbol{\sigma}^2(U) &= \boldsymbol{\sigma}^2(2X + 3Y) \\ &= 4\,\boldsymbol{\sigma}^2(X) + 9\,\boldsymbol{\sigma}^2(Y) + 2 \times 6\,\mathbf{cov}(X, Y) \\ &= 13\,; \end{aligned}$$

- $\boldsymbol{\sigma}^2(V)$

$$\begin{aligned} \boldsymbol{\sigma}^2(V) &= \boldsymbol{\sigma}^2(X - Y) \\ &= \boldsymbol{\sigma}^2(X) + \boldsymbol{\sigma}^2(Y) - 2\,\mathbf{cov}(X, Y) \\ &= 2\,; \end{aligned}$$

- $\mathbf{cov}(U, V)$

$$\begin{aligned}\mathbf{cov}(U, V) &= \mathbf{cov}(2X + 3Y, X - Y) \\ &= 2\,\sigma^2(X) - 2\,\mathbf{cov}(X, Y) + 3\,\mathbf{cov}(X, Y) - 3\,\sigma^2(Y) \\ &= -1\,.\end{aligned}$$

The covariance matrix is

$$C = \begin{pmatrix} 13 & -1 \\ -1 & 2 \end{pmatrix}.$$

2. To compute the joint density of (U, V), we first compute the joint c.d.f. of (U, V) given by

$$F(u, v) = \mathbf{P}(U \leq u, V \leq v) = \mathbf{P}(2X + 3Y \leq u, X - Y \leq v)\,.$$

This probability is given by the integral of the joint density on the domain $D_{u,v}$ of $\mathbb{R}^2$ where

$$D_{u,v} = \{(x, y) \in \mathbb{R}^2 \,|\, 2x + 3y \leq u,\ x - y \leq v\}\,.$$

We obtain

$$F(u, v) = \int\int_{D_{u,v}} f(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

To solve the integral, we perform the change of variables

$$z = 2x + 3y, \qquad t = x - y\,,$$

to transform the domain $D_{u,v}$ into the region

$$\hat{D}_{u,v} = \{(x, y) \in \mathbb{R}^2 \,|\, z \leq u,\ t \leq v\}\,.$$

with sides which are parallel to the axes. If we now compute x, y as function of z and t, we obtain

$$x = \frac{1}{5}\,(z + 3t), \qquad y = \frac{1}{5}\,(z - 2t)\,.$$

It follows that the Jacobian matrix is equal to

$$J_\Psi = \begin{pmatrix} \frac{\partial \Psi_1}{\partial z} & \frac{\partial \Psi_1}{\partial t} \\ \frac{\partial \Psi_2}{\partial z} & \frac{\partial \Psi_2}{\partial t} \end{pmatrix} = \begin{pmatrix} \frac{1}{5} & \frac{3}{5} \\ \frac{1}{5} & -\frac{2}{5} \end{pmatrix},$$

where $(x, y) = \Psi(z, t) = (\Psi_1(z, t), \Psi_2(z, t)) = \left(\frac{z+3t}{5}, \frac{z-2t}{5}\right)$, with determinant

$$|\det J_\Psi| = \frac{1}{5}.$$

We obtain:

$$\begin{aligned}
F(u, v) &= \int\int_{D_{u,v}} f(x, y)\, \mathrm{d}x\, \mathrm{d}y \\
&= \int\int_{\hat{D}_{u,v}} f(\psi(z, t))\, |\det J_\Psi|\, \mathrm{d}z\, \mathrm{d}t \\
&= \int_{-\infty}^{u}\int_{-\infty}^{v} \frac{1}{2\pi} e^{-\frac{1}{2}\left(\left(\frac{z+3t}{5}\right)^2+\left(\frac{z-2t}{5}\right)^2\right)} \frac{1}{5}\, \mathrm{d}z\, \mathrm{d}t \\
&= \int_{-\infty}^{u}\int_{-\infty}^{v} \frac{1}{10\pi} e^{-\frac{1}{2}\cdot\frac{1}{25}\left(2z^2+13t^2+2zt\right)}\, \mathrm{d}z\, \mathrm{d}t\,.
\end{aligned}$$

The joint density of (U, V) is then

$$\frac{1}{10\pi} e^{-\frac{1}{50}\left(2z^2+13t^2+2zt\right)},\ z, t \in \mathbb{R}^2\,.$$

Note that (U, V) have again joint Gaussian distribution with covariance matrix equal to C.
To verify these results, compute the inverse matrix of A, where

$$A = \begin{pmatrix} \frac{2}{25} & \frac{1}{25} \\ \frac{1}{25} & \frac{13}{25} \end{pmatrix}.$$

□

Exercise 12.8 A random vector (X, Y, Z) has joint density given by

$$f(x, y, z) = k\, e^{-\frac{1}{2}\left(2x^2-2xy+y^2+z^2+2x-6y\right)}\,.$$

1. Compute k.
2. Compute the expectations $\mathbf{P}(X)$, $\mathbf{P}(Y)$ and $\mathbf{P}(Z)$.
3. Compute the density of the random vector (X, Z).
4. Compute the correlation coefficient between X and Z and between X and Y.
5. Let $W = X + Z$; compute the probability density of W.

Solution 12.8 1. If we write the density in the standard form

$$f(x, y, z) = k\, e^{-\frac{1}{2} Av\cdot v + b\cdot v}$$

where A is the symmetric matrix

$$A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

b is the vector in $\mathbb{R}^3$

$$b = \begin{pmatrix} -1 \\ 3 \\ 0 \end{pmatrix},$$

and v is given by

$$v = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

We can compute the normalization constant k as follows:

$$k = \sqrt{\frac{\det A}{(2\pi)^3}}\, e^{-\frac{1}{2} A^{-1} b\cdot b}.$$

It is now sufficient to calculate the determinant and the inverse matrix of A. We obtain:

$$\det A = 1,$$

$$A^{-1} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

from which

$$A^{-1}b = \begin{pmatrix} 2 \\ 5 \\ 0 \end{pmatrix}$$

and

$$k = e^{-\frac{1}{2}A^{-1}b\cdot b}\sqrt{\frac{\det A}{(2\pi)^3}} = e^{-\frac{13}{2}}\sqrt{\frac{1}{(2\pi)^3}}\,.$$

2. The expectations of X, Y, Z are given respectively by

$$\mathbf{P}(X) = \left[A^{-1}b\right]_1 = 2\,,$$

$$\mathbf{P}(Y) = \left[A^{-1}b\right]_2 = 5\,,$$

$$\mathbf{P}(Z) = \left[A^{-1}b\right]_3 = 0\,.$$

3. The random vector (X, Z) has bidimensional Gaussian density of covariance matrix D given by

$$D = \begin{pmatrix} \left[A^{-1}\right]_{11} & \left[A^{-1}\right]_{13} \\ \left[A^{-1}\right]_{31} & \left[A^{-1}\right]_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and vector d of expectations

$$d = \begin{pmatrix} 2 \\ 0 \end{pmatrix}.$$

To prove this, we derive the joint density $f_{X,Z}(x, z)$ from $f(x, y, z)$ as follows:

$$\begin{aligned} f_{X,Z}(x, z) &= \int_{\mathbb{R}} f(x, y, z)\,\mathrm{d}y \\ &= \int_{\mathbb{R}} k\,e^{-\frac{1}{2}\,(2x^2-2xy+y^2+z^2+2x-6y)}\,\mathrm{d}y \\ &= e^{-\frac{1}{2}\,(2x^2+z^2+2x)}\int_{\mathbb{R}} k\,e^{-\frac{1}{2}\,(y^2-2xy)+3y}\,\mathrm{d}y \\ &= e^{-\frac{1}{2}\,(2x^2+z^2)-x}\int_{\mathbb{R}} k\,e^{-\frac{1}{2}y^2+(3+x)y}\,\mathrm{d}y\,. \end{aligned}$$

Here we can consider

$$I = \int_{\mathbb{R}} k\,e^{-\frac{1}{2}y^2+(3+x)y}\,\mathrm{d}y$$

as the integral of a one-dimensional Gaussian distribution with coefficients depending on the parameter x. In the same notation as above, we obtain

$$A = 1 \qquad \text{and} \qquad b = 3 + x\,,$$

from which we obtain

$$\begin{aligned}\int_{\mathbb{R}} k\, e^{-\frac{1}{2}y^2+(3+x)y}\,\mathrm{d}y &= \sqrt{\frac{2\pi}{\det A}} e^{\frac{1}{2}A^{-1}b\cdot b} \\ &= \sqrt{2\pi} e^{\frac{1}{2}(3+x)^2}\,.\end{aligned}$$

We can obtain the same result also by completing the square

$$-\frac{1}{2}y^2 + (3+x)y$$

in the integral I. It follows that

$$\begin{aligned}f_{X,Z}(x,z) &= k\, e^{-\frac{1}{2}(2x^2+z^2)-x} \cdot \sqrt{2\pi}\, e^{\frac{1}{2}\,(3+x)^2} \\ &= \frac{e^{-\frac{13}{2}+\frac{9}{2}}}{2\pi}\, e^{-\frac{1}{2}(x^2+z^2)+2x} \\ &= \frac{e^{-2}}{2\pi}\, e^{-\frac{1}{2}(x^2+z^2)+2x}\,.\end{aligned}$$

4. The correlation coefficient between X and Z can be obtained by the formula

$$\rho(X,Z) \;=\; \frac{\mathbf{cov}(X,Z)}{\sigma(X)\,\sigma(Z)}\,.$$

From the covariance matrix we have that

$$\mathbf{cov}(X,Z) \;=\; 0\,,$$

hence

$$\rho(X,Y) \;=\; \frac{1}{\sqrt{2}\sqrt{1}} \;=\; \frac{\sqrt{2}}{2}\,.$$

5. The probability density of W can be computed via the formula

$$f_W(w) \;=\; \int_{\mathbb{R}} f_{X,Z}(x, w-x)\,\mathrm{d}x\,.$$

Hence with the same method used above:

$$\begin{aligned}
f_W(w) &= \int_{\mathbb{R}} \frac{e^{-2}}{2\pi}\, e^{-\frac{1}{2}(x^2+(w-x)^2)+2x}\, \mathrm{d}x \\
&= \frac{e^{-2}}{2\pi}\, e^{-\frac{1}{2}w^2} \int_{\mathbb{R}} e^{-\frac{1}{2}(2x^2)+(2+w)x}\, \mathrm{d}x \\
&= \frac{e^{-2}}{2\pi}\, e^{-\frac{1}{2}w^2}\, \sqrt{\pi} e^{\frac{1}{2}\times\frac{1}{2}\,(2+w)^2} \\
&= \frac{e^{-2+1}}{2\sqrt{\pi}}\, e^{-\frac{1}{4}w^2+w} \\
&= \frac{1}{2\sqrt{\pi}}\, e^{-\frac{1}{4}\,(w-2)^2}\,.
\end{aligned}$$

The random number W has normal density with expectation

$$\mathbf{P}(W) \,=\, \mathbf{P}(X) + \mathbf{P}(Z) \,=\, 2$$

and variance

$$\boldsymbol{\sigma}^2(W) \,=\, \boldsymbol{\sigma}^2(X) + \boldsymbol{\sigma}^2(Z) + 2\,\mathbf{cov}(X, Z) \,=\, 2\,. \qquad \square$$

Chapter 13
Markov Chains

Exercise 13.1 A Markov chain $(X_n)_{n\in\mathbb{N}}$ with states $S = \{1, 2, 3, 4\}$ has the following transition matrix

$$\begin{pmatrix} \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} \end{pmatrix}$$

and initial distribution

$$\mu(1) = \mu(2) = \mu(3) = \mu(4) = \frac{1}{4}.$$

(a) Determine the equivalence classes of the states and their periods.
(b) Compute $p_{2,1}^{(2)}$, $p_{1,4}^{(2)}$, $p_{1,1}^{(2)}$.
(c) Check the existence of the following limits and compute them, if they exist:

$$\lim_{n\to\infty} p_{1,3}^{(n)} \quad \text{and} \quad \lim_{n\to\infty} \mathbf{P}(X_n = 2).$$

Solution 13.1 (a) To determine equivalence classes of the states, we can draw a graph of the transition probabilities by using the matrix P. We first represent the states (see Fig. 13.1) and then connect with an arrow two states such that the transition probability from one to the other is strictly positive. For example, since

$$[P]_{1,2} = \frac{3}{4},$$

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8_13

Fig. 13.1 The states

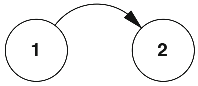


Fig. 13.2 The chain has positive probability of going from 1 to 2

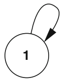


Fig. 13.3 The chain have positive probability of remaining in the state 1

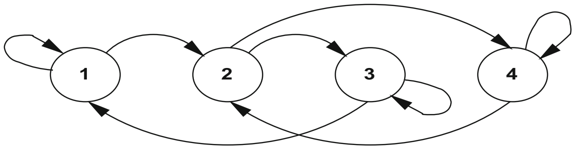


Fig. 13.4 Graph of the relations among the states

the chain has positive probability to go from the state 1 to the state 2 in one step. We represent this on the graph by connecting 1 and 2 with an arrow going from 1 to 2, see Fig. 13.2.

Analogously, $[P]_{1,1} = \dfrac{1}{4}$ means that the chain has positive probability of remaining in the state 1. This can be represented as illustrated in Fig. 13.3.
By using this procedure we can construct the graph of Fig. 13.4.
From the graph we deduce that all elements can communicate with each other, i.e. there exists paths that connect each state to all the other ones with positive probability. We conclude that there exists only one equivalence class [1].
Furthermore we can deduce from the graph that the period of the chain is 1, since there exists a path of length 1 from state 1 to itself, i.e.

$$1 \in \{n \mid p_{1,1}^{(n)} > 0\}.$$

(b) To compute $p_{2,1}^{(2)}$, i.e. the probability of going in 2 steps from the state 2 to the state 1, we write

$$p_{2,1}^{(2)} = \sum_{i \in S} p_{2,i}^{(1)} p_{i,1}^{(1)}.$$

This formula shows how the probability of going in 2 steps from the state 2 to the state 1 can be computed as the sum of the probabilities of all possible paths from 2 to 1.
From the graph relative to the matrix P we obtain

$$p_{2,1}^{(2)} = p_{2,3}^{(1)} p_{3,1}^{(1)} = \frac{2}{3} \cdot \frac{1}{4} = \frac{1}{6}.$$

Note that we can compute $p_{2,1}^{(2)}$ by taking the product of the matrix column with the matrix row

$$p_{2,1}^{(2)} = P_2 \cdot P^1$$

where P_2 denotes the second row and P^1 the first column of the matrix P. Analogously we compute $p_{1,4}^{(2)}$ and $p_{1,1}^{(2)}$.

(c) Since the chain is irreducible and aperiodic, the ergodic theorem guarantees the existence of the limit

$$\lim_{n\to\infty} p_{1,3}^{(n)} = \pi_3,$$

where π_3 can be obtained by the solution of the linear system

$$\pi = {}^t\pi P$$

$$\begin{cases} \pi_1 = \sum \pi_i p_{i,1} \\ \pi_2 = \sum \pi_i p_{i,2} \\ \pi_3 = \sum \pi_i p_{i,3} \\ \pi_1 + \pi_2 + \pi_3 + \pi_4 = 1. \end{cases}$$

In this case

$$\begin{cases} \pi_1 = \frac{1}{4}\pi_1 + \frac{1}{4}\pi_3 \\ \pi_2 = \frac{3}{4}\pi_1 + \frac{1}{3}\pi_4 \\ \pi_3 = \frac{2}{3}\pi_2 + \frac{3}{4}\pi_3 \\ \sum_{i=1}^{4} \pi_i = 1. \end{cases}$$

By using standard methods for the solution of linear systems of equations, we obtain:

$$\pi_3 = \frac{12}{25}.$$

Hence

$$\lim_{n\to\infty} p_{1,3}^{(n)} = \frac{12}{25}.$$

To compute $\lim_{n\to\infty} \mathbf{P}(X_n = 2)$, we note that

$$\mathbf{P}(X_n = 2) = \sum_{i=1}^{4} \mathbf{P}(X_n = 2 | X_0 = i)\mathbf{P}(X_0 = i) = \frac{1}{4}\sum_{i=1}^{4} p_{i,2}^{(n)}.$$

Since for all i

$$\lim_{n\to\infty} p_{i,2}^{(n)} = \pi_2,$$

we have

$$\lim_{n\to\infty} \mathbf{P}(X_n = 2) = \lim_{n\to\infty} \frac{1}{4}\sum_{i=1}^{4} p_{i,2}^{(n)} = \frac{1}{4} \cdot 4\pi_2 = \pi_2,$$

where $\pi_2 = \dfrac{9}{50}$. □

Exercise 13.2 A Markov chain $X_n, n = 0, 1, 2\ldots$ with states

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

has the following transition matrix

$$\begin{pmatrix} 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 \end{pmatrix}$$

and initial distribution

$$\mu(1) = \frac{1}{3},\ \mu(2) = \frac{2}{3},\ \mu(3) = \mu(4) = \mu(5) = \mu(6) = 0.$$

1. Determine the equivalence classes of the states and their periods.
2. Check the existence of the following limits and compute them, if they exist:

$$\lim_{n\to\infty} p_{1,5}^{(2n)},\ \lim_{n\to\infty} p_{3,5}^{(n)},\ \lim_{n\to\infty} p_{2,5}^{(2n)} \text{ and } \lim_{n\to\infty} \mathbf{P}(X_n = 5).$$

3. Compute $\mathbf{P}(X_2 < 3)$.

Solution 13.2 1. As in the previous exercise, we draw the graph of the states as in Fig. 13.5, in order to determine the equivalence classes of the states and their periods. We connect the states with an arrow in the case there exists a positive

Fig. 13.5 Graph of the states

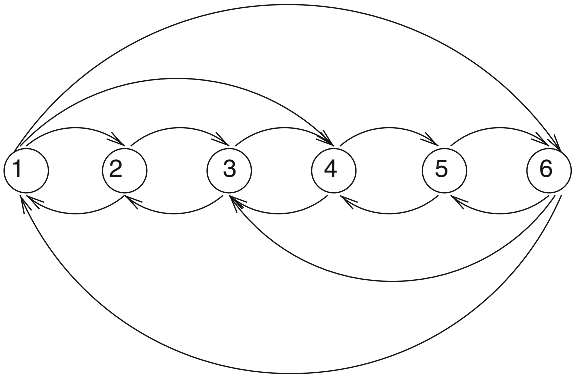


Fig. 13.6 Graph of the probabilities of transition

probability to pass from the state, where the arrow starts, to the state where the arrow ends.

By using the transition matrix P we obtain the graph shown in Fig. 13.6, where we can see that there exists only one equivalence class. We note that the number of steps needed to come back to the state from where we started is always even. Furthermore $p_{1,1}^{(2)} > 0$, hence it follows that the period of the equivalence class is 2. Namely

$$2 = \text{MCD}\, A_s^+$$

where $A_s^+ = \{n \mid p_{s,s}^{(n)} > 0\}$.

2. To study the limits, we consider the equivalence classes of the matrix P^2; we obtain two equivalence classes, each of period 1. To derive the equivalence classes, it is not necessary to compute the whole matrix P^2; for example, the equivalence class of 1 will be given by all states that communicate with 1 with an even number of steps. We obtain

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

Since 2 and 5 do not communicate with an even number of steps, we immediately have that

$$p_{2,5}^{(2n)} = 0$$

for all n, hence

$$\lim_{n\to\infty} p_{2,5}^{(2n)} = 0.$$

The state 5 belongs to the class [1] calculated with respect to P^2. Hence we can apply the ergodic theorem to that class, since it has period 1 with respect to P^2. The submatrix of P^2 relative to [1] is given by:

$$\begin{pmatrix} \frac{5}{18} & \frac{9}{18} & \frac{2}{9} \\ \frac{1}{6} & \frac{11}{18} & \frac{2}{9} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \end{pmatrix}.$$

By the ergodic theorem we have that

$$\lim_{n\to\infty} p_{1,5}^{(2n)} = \pi_5$$

where π_5 is the solution of the system

$$\begin{cases} \pi_1 = \dfrac{5}{18}\pi_1 + \dfrac{1}{6}\pi_3 + \dfrac{1}{6}\pi_5 \\ \pi_3 = \frac{9}{18}\pi_1 + \frac{11}{18}\pi_3 + \frac{1}{2}\pi_5 \\ \pi_1 + \pi_3 + \pi_5 = 1. \end{cases}$$

We obtain

$$\pi_1 = \frac{3}{16} \quad \pi_3 = \frac{9}{16} \quad \pi_5 = \frac{1}{4}.$$

Hence

$$\lim_{n\to\infty} P_{1,5}^{(2n)} = \frac{1}{4}.$$

To obtain the asymptotic behavior of $p_{3,5}^{(n)}$ for n going to infinity, we note that

(a) on the even steps, i.e. for $n = 2k$, we have

$$p_{3,5}^{(2k)} \xrightarrow[k\to\infty]{} \pi_5;$$

(b) on the odd steps, i.e. for $n = 2k+1$, we have

$$p_{3,5}^{(2k+1)} = 0$$

since the probability of going from the state 3 to the state 5 in an odd number of steps is zero. Since the limit on two subsequences is different, we can conclude that

$$\lim_{n\to\infty} p_{3,5}^{(n)}$$

does not exist. To compute

$$\lim_{n\to\infty} \mathbf{P}(X_n = 5),$$

we use the formula of total probability:

$$\mathbf{P}(X_n = 5) = \sum_{i=1}^{6} \mathbf{P}(X_n = 5|X_0 = i)\,\mathbf{P}(X_0 = i)$$
$$\sum_{i=1}^{6} p_{i,5}^{(n)}\,\mu_i = \frac{1}{3}p_{1,5}^{(n)} + \frac{2}{3}p_{2,5}^{(n)}.$$

We need to distinguish the following 2 cases:

(a) what happens on the even steps, i.e. for $n = 2k$. We have:

$$\frac{1}{3}\,p_{1,5}^{(2k)} + \frac{2}{3}\,p_{2,5}^{(2k)} = \frac{1}{3}\,p_{1,5}^{(2k)} \xrightarrow[k\to\infty]{} \frac{\pi_5}{3};$$

(b) what happens on the odd steps, i.e. for $n = 2k+1$. We have:

$$\frac{1}{3}\,p_{1,5}^{(2k+1)} + \frac{2}{3}p_{2,5}^{(2k+1)} = \frac{2}{3}\,p_{2,5}^{(2k+1)} = \frac{2}{3}\sum_{i=1}^{6} p_{2,i}^{(1)}\,p_{i,5}^{(2k)},$$

that tends to $\frac{2}{3}\,\pi_5 \sum_{i=1}^{6} p_{2,i}^{(1)} = \frac{2}{3}\pi_5$ for $k \to \infty$, since we have that $p_{2,i}^{(1)} \neq 0$ for the i such that we have $\lim_{k\to\infty} p_{i,5}^{(2k)} = \pi_5$.

Since we obtain different limits, we can conclude that the limit

$$\lim_{n\to\infty} \mathbf{P}(X_n = 5)$$

does not exist.

3. To compute $\mathbf{P}(X_2 < 3)$ we note that

$$\mathbf{P}(X_2 < 3) = \sum_{i=1}^{2} \mathbf{P}(X_2 = i),$$

since the event $(X_2 < 3) = (X_2 = 1) + (X_2 = 2)$. It is then sufficient to compute

$$
\begin{aligned}
\mathbf{P}(X_2 = 1) &= \sum_{i=1}^{6} \mathbf{P}(X_2 = 1 | X_0 = i)\, \mathbf{P}(X_0 = i) \\
&= \sum_{i=1}^{6} p_{i,1}^{(2)}\, \mu_i \\
&= \frac{1}{3} p_{1,1}^{(2)} + \frac{2}{3} p_{2,1}^{(2)} = \frac{5}{54}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{P}(X_2 = 2) &= \sum_{i=1}^{6} \mathbf{P}(X_2 = 2 | X_0 = i)\, \mathbf{P}(X_0 = i) \\
&= \sum_{i=1}^{6} p_{i,2}^{(2)}\, \mu_i \\
&= \frac{1}{3} p_{1,2}^{(2)} + \frac{2}{3} p_{2,2}^{(2)} = \frac{2}{9}.
\end{aligned}
$$

Finally, $\mathbf{P}(X_2 < 3) = \dfrac{17}{54}$. □

Exercise 13.3 A Markov chain $X_n, n = 0, 1, 2 \ldots$ with states

$$S = \{1, 2, 3, 4\}$$

has the following transition matrix

$$
P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ \frac{1}{6} & 0 & 0 & \frac{5}{6} \\ 0 & \frac{3}{4} & \frac{1}{4} & 0 \end{pmatrix}
$$

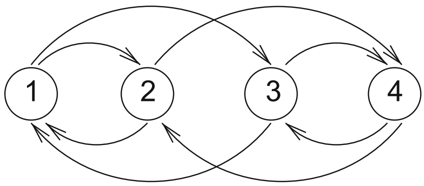
and initial distribution

$$\mu(1) = \frac{1}{3},\ \mu(2) = \frac{1}{3},\ \mu(3) = \frac{1}{3},\ \mu(4) = 0.$$

1. Determine the equivalence classes of the states and their periods.
2. Compute $\mathbf{P}(X_5 = 2 | X_2 = 3)$, $p_{1,4}^{(2)}$ and $\mathbf{P}(X_2)$.
3. Check the existence of the following limits and compute them, if they exist:

$$\lim_{n\to\infty} p_{1,3}^{(2n)},\ \lim_{n\to\infty} p_{1,4}^{(2n)},\ \lim_{n\to\infty} p_{2,3}^{(n)} \text{ and } \lim_{n\to\infty} \mathbf{P}(X_n = 2).$$

Fig. 13.7 Graph of the states



Solution 13.3 1. To find the equivalence classes we draw the graph of the states as shown in Fig. 13.7.
Since we can reach all other states by starting from the state 1 and the state 1 can be reached from all other states, there exists only one equivalence class. Furthermore by starting from 1, we return to it always with only an even number of steps and $p_{1,1}^{(2)} > 0$. We conclude that the period of the class is equal to 2.

2. To compute the conditional probability $\mathbf{P}(X_5 = 2|X_2 = 3)$ we use the fact that the chain is homogeneous. It holds that

$$\mathbf{P}(X_5 = 2|X_2 = 3) \;=\; p_{3,2}^{(3)} \;=\; [P^3]_{3,2} \;=\; 0.$$

To compute this probability, we need to calculate the element on the row 3 and column 2 of the matrix P^3, We can obtain this element by multiplicating the third row of P^2 with the second column of P.

Anyways without any computation we can immediately state that the probability of going from the state 3 to the state 2 in a odd number of steps is 0!

Furthermore

$$p_{1,4}^{(2)} \;=\; [P^2]_{1,4} \;=\; \frac{7}{12}.$$

We now compute the expectation of the state of the chain at time $t = 2$ by using the formulas of the expectation of a random number with discrete distribution and of the total probabilities:

$$\begin{aligned}
\mathbf{P}(X_2) &= \sum_{i=1}^{4} i\,\mathbf{P}(X_2 = i) \\
&= \sum_{i=1}^{4} i \sum_{j=1}^{4} \mathbf{P}(X_2 = i|X_0 = j)\,\mathbf{P}(X_0 = j) \\
&= \sum_{i=1}^{4} \frac{i}{3}\,\left(\mathbf{P}(X_2 = i|X_0 = 1) + \mathbf{P}(X_2 = i|X_0 = 2) + \mathbf{P}(X_2 = i|X_0 = 3)\right) \\
&= \sum_{i=1}^{4} \frac{i}{3}\,\left([P^2]_{1,i} + [P^2]_{2,i} + [P^2]_{3,i}\right).
\end{aligned}$$

The matrix of P^2 is given by:

$$P^2 = \begin{pmatrix} \frac{5}{12} & 0 & 0 & \frac{7}{12} \\ 0 & \frac{7}{12} & \frac{5}{12} & 0 \\ 0 & \frac{17}{24} & \frac{7}{24} & 0 \\ \frac{13}{24} & 0 & 0 & \frac{11}{24} \end{pmatrix}.$$

Hence the expectation of X_2 is then:

$$\mathbf{P}(X_2) = \frac{1}{3}\left(\frac{5}{12} + 2\left(\frac{7}{12} + \frac{17}{24}\right) + 3\left(\frac{5}{12} + \frac{7}{24}\right)\right) = \frac{41}{24}.$$

3. We now compute the limits. The Markov chain observed on the even steps can be considered as a Markov chain with transition matrix equal to P^2. We can immediately see that the state 3 cannot be reached from the state 1 with an even number of steps. In fact, the equivalence classes relative to P^2 are given by

$$[1] = \{1,\ 4\},$$

$$[2] = \{2,\ 3\}.$$

It follows that

$$\lim_{n\to\infty} p_{1,3}^{(2n)} = 0.$$

The state 4 belongs to the equivalence class [1] relative to P^2. This class has period 1. Hence we can apply the ergodic theorem to this irreducible aperiodic subchain to compute

$$\lim_{n\to\infty} p_{1,4}^{(2n)}.$$

If we put $\pi_4 = \lim_{n\to\infty} p_{1,4}^{(2n)}$ and $\pi_1 = \lim_{n\to\infty} p_{1,1}^{(2n)}$, by the ergodic theorem we obtain

$$\begin{cases} \pi_1 + \pi_4 = 1 \\ \frac{5}{12}\pi_1 + \frac{13}{24}\pi_4 = \pi_1. \end{cases}$$

The solution of the system is

$$\pi_1 = \frac{13}{27}, \qquad \pi_4 = \frac{14}{27}.$$

It follows that

$$\lim_{n\to\infty} p_{1,4}^{(2n)} = \frac{14}{27}.$$

To compute $\lim_{n\to\infty} p_{2,3}^{(n)}$ we observe the behavior of the chain on the even steps and on the odd ones.

(a) First we note that $2 \in [3]$ relative to P^2.

By the ergodic theorem we have that on the even steps (i.e. if $n = 2k$)

$$p_{2,3}^{(2k)} \xrightarrow[k\to\infty]{} \pi_3$$

where π_3 is the solution of the system

$$\begin{cases} \pi_2 + \pi_3 = 1 \\ \frac{5}{12}\pi_2 + \frac{7}{24}\pi_3 = \pi_3. \end{cases}$$

(b) There exists no path with an odd number of steps from the state 2 to the state 3, hence

$$p_{2,3}(2k+1) = 0 \qquad \forall\, k.$$

We can obtain the same result by computing

$$\begin{aligned} p_{2,3}^{(2k+1)} &= \sum_j p_{2,j}^{(2k)}\, p_{j,3}(1) \\ &= p_{2,1}^{(2k)}\, p_{1,3}(1) + p_{2,4}^{(2k)}\, p_{4,3}(1) = 0. \end{aligned}$$

Summing up

$$\begin{aligned} p_{2,3}^{(2k)} &\xrightarrow[k\to\infty]{} \pi_3 > 0 \\ p_{2,3}^{(2k+1)} &\xrightarrow[k\to\infty]{} 0. \end{aligned}$$

Hence the limit $\lim_{n\to\infty} p_{2,3}^{(n)}$ *does not* exist.

Finally, to compute $\lim_{n\to\infty} \mathbf{P}(X_n = 2)$ we proceed as in the previous case. First we use the formula of the total probability to compute $\mathbf{P}(X_n = 2)$:

$$\begin{aligned}\mathbf{P}(X_n = 2) &= \sum_{i=1}^{4} \mathbf{P}(X_n = 2|X_0 = i)\,\mathbf{P}(X_0 = i)\\ &= \frac{1}{3}\,(\mathbf{P}(X_n = 2|X_0 = 1) + \mathbf{P}(X_n = 2|X_0 = 2) + \mathbf{P}(X_n = 2|X_0 = 3))\\ &= \frac{1}{3}\,\left(p_{1,2}^{(n)} + p_{2,2}^{(n)} + p_{3,2}^{(n)}\right).\end{aligned}$$

By using the results above, we obtain

(a) if $n = 2k$

$$\frac{1}{3}\,\left(p_{1,2}^{(2k)} + p_{2,2}^{(2k)} + p_{3,2}^{(2k)}\right) = \frac{1}{3}\,\left(p_{2,2}^{(2k)} + p_{3,2}^{(2k)}\right)$$

which tends to $\frac{2}{3}\pi_2$ for $k \to \infty$;

(b) if $n = 2k + 1$

$$\begin{aligned}\frac{1}{3}\,\left(p_{1,2}^{(2k+1)} + p_{2,2}^{(2k+1)} + p_{3,2}^{(2k+1)}\right) &= \frac{1}{3}\,p_{1,2}^{(2k+1)}\\ &= \frac{1}{3}\,\sum_{i=1}^{4} p_{1,i}^{(1)}\,p_{i,2}^{(2k)}\\ &= \frac{1}{3}\,\left(p_{1,2}^{(1)}\,p_{2,2}^{(2k)} + p_{1,3}^{(1)}\,p_{3,2}^{(2k)}\right)\end{aligned}$$

which tends to $\frac{1}{3}\,\pi_2\,\left(p_{1,2}^{(1)} + p_{1,3}^{(1)}\right) = \frac{1}{3}\,\pi_2$ for $k \to \infty$.

Summing up, if we put $p_n = \mathbf{P}(X_n = 2)$, we have that

$$\begin{aligned}p_{2n} &\xrightarrow[n\to\infty]{} \frac{2}{3}\pi_2,\\ p_{2n+1} &\xrightarrow[n\to\infty]{} \frac{1}{3}\,\pi_2,\end{aligned}$$

i.e. the limit $\lim_{n\to\infty} \mathbf{P}(X_n = 2)$ does not exist. □

Exercise 13.4 A Markov chain $(X_n)_{n\in\mathbb{N}}$ with states $S = \{1, 2, 3, 4, 5\}$ has the following transition matrix

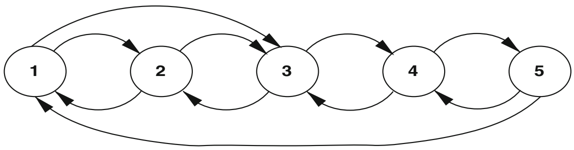


Fig. 13.8 Graph of states

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} & 0 \end{pmatrix}$$

and initial distribution

$$\mu(1) = 0,\ \mu(2) = \frac{2}{3},\ \mu(3) = \frac{1}{3},\ \mu(4) = \mu(5) = 0.$$

(a) Determine the equivalence classes of the states and their periods.
(b) Check the existence of the following limits and, if they exist, compute them:

$$\lim_{n\to\infty} p_{1,5}^{(n)},\ \lim_{n\to\infty} p_{3,5}^{(n)},\ \lim_{n\to\infty} (p_{2,3}^{(n)} + p_{3,5}^{(n)}),\ \lim_{n\to\infty} \mathbf{P}(X_n = 5).$$

(c) Compute $\mathbf{P}(X_1 \le 2)$ and $\mathbf{P}(X_2 = 5)$.

Solution 13.4 (a) To determine the equivalence classes of the states and their periods we draw the graph of the states, see Fig. 13.8.

We first note that all the states comunicate among each other. Consider the set

$$A_1^+ = \{n|\, p_{11}^{(n)} > 0\}$$

i.e. the set of the lengths of the paths that starts and ends in 1. We note that there exists a path of length 2 (for example, from 1 to 2 and from 2 to 1) and of length 3 (from 1 to 3, from 3 to 2, from 2 to 1). We have

$$2, 3 \in A_1^+.$$

The period d of the equivalence class [1] is given by

$$d = MCD(A_1^+),$$

hence d must be equal to 1 since it must divide 2 and 3. We conclude that there exists only one equivalence class of period 1.

(b) By the ergodic theorem it follows that all the limits exist since the chain has a unique equivalence class of period 1. First we note that

$$\lim_{n\to\infty} p_{1,5}^{(n)} = \lim_{n\to\infty} p_{3,5}^{(n)} = \pi_5$$

since the starting state (1 to 3) does not count. Furthermore

$$\begin{aligned}\lim_{n\to\infty}(p_{2,3}^{(n)} + p_{3,5}^{(n)}) &= \lim_{n\to\infty} p_{2,3}^{(n)} + \lim_{n\to\infty} p_{3,5}^{(n)}\\ &= \pi_3 + \pi_5\end{aligned}$$

and finally

$$\begin{aligned}\lim_{n\to\infty} \mathbf{P}(X_n = 5) &= \lim_{n\to\infty}\sum_{i=1}^{5}\mathbf{P}(X_n = 5|X_0 = i)\mathbf{P}(X_0 = i)\\ &= \lim_{n\to\infty}\sum_{i=1}^{5}\mu(i)p_{i,5}^{(n)}\\ &= \pi_5\sum_{i=1}^{5}\mu(i) = \pi_5\cdot 1 = \pi_5,\end{aligned}$$

since $\lim_{n\to\infty} p_{i,5}^{(n)} = \pi_5,\ \forall i = 1,\dots,5$ and $\sum_{i=1}^{5}\mu(i) = 1$. To obtain π_i, it is sufficient to solve the system

$$\begin{cases}\pi = \pi^t P\\ \sum_{i=1}^{5}\pi_i = 1\end{cases}$$

i.e.

$$\begin{cases}\pi_1 = \frac{1}{2}\pi_2 + \frac{2}{3}\pi_5\\ \pi_2 = \frac{1}{2}\pi_1 + \frac{2}{3}\pi_3\\ \pi_4 = \frac{1}{3}\pi_3 + \frac{1}{3}\pi_5\\ \pi_5 = \frac{1}{3}\pi_4\\ \sum_{i=1}^{5}\pi_i = 1.\end{cases}$$

In this system we have already taken out a redundant equation. We obtain

$$\pi_3 = \frac{7}{540} \quad \text{e} \quad \pi_5 = \frac{14}{135}.$$

Hence, summing up

$$\lim_{n\to\infty} p_{1,5}^{(n)} = \lim_{n\to\infty} p_{3,5}^{(n)} = \lim_{n\to\infty} \mathbf{P}(X_n = 5) = \pi_5 = \frac{14}{135},$$
$$\lim_{n\to\infty} \left(p_{2,3}^{(n)} + p_{3,5}^{(n)}\right) = \pi_3 + \pi_5 = \frac{7}{540} + \frac{14}{135} = \frac{7}{60}.$$

(c) To compute the probabilities, we note that

$$\begin{aligned}
\mathbf{P}(X_1 \le 2) &= \mathbf{P}(X_1 = 1) + \mathbf{P}(X_1 = 2) \\
&= \sum_{i=1}^{5} \mathbf{P}(X_1 = 1|X_0 = i)\mu(i) + \sum_{i=1}^{5} \mathbf{P}(X_1 = 2|X_0 = i)\mu(i) \\
&= \frac{2}{3}p_{2,1} + \frac{1}{3}p_{3,1} + \frac{2}{3}p_{2,2} + \frac{1}{3}p_{3,2} \\
&= \frac{1}{3} + \frac{2}{9} = \frac{5}{9}.
\end{aligned}$$

The second probability can be computed by using the formula of the total probability:

$$\begin{aligned}
\mathbf{P}(X_2 = 5) &= \sum_{i=1}^{5} \mathbf{P}(X_2 = 5|X_0 = i)\mu(i) \\
&= \frac{2}{3}p_{2,5}^{(2)} + \frac{1}{3}p_{3,5}^{(2)} \\
&= \frac{2}{3}[P^2]_{2,5} + \frac{1}{3}[P^2]_{3,5} \\
&= \frac{1}{3} \cdot \frac{1}{9} = \frac{1}{27}.
\end{aligned}$$

$\square$

Chapter 14
Statistics

Exercise 14.1 The events $E_1, E_2, \ldots$ are stochastically independents subordinately to the random parameter Θ with $\mathbf{P}(E_i|\Theta = \theta) = \theta$. The a priori density of Θ is given by

$$\pi_0 = \begin{cases} 3\,\theta^2 & 0 \le \theta \le 1, \\ \\ 0 & \text{otherwise.} \end{cases}$$

We observe the values of the first 4 events:

$$E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1.$$

1. Compute the a posteriori density $\pi_4(\Theta|E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1)$ of Θ.
2. Compute the a priori probability that Θ belongs to the interval $\left[\frac{1}{2}, 1\right]$.
3. Compute the a posteriori probability that Θ belongs to the interval $\left[\frac{1}{2}, 1\right]$.
4. Compute $\arg\max \pi_4(\Theta|E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1)$.[1]
5. Compute the a posteriori expectation of $E = E_5 \wedge E_6$.

Solution 14.1 1. The a posteriori density can be computed by using the formula

$$\begin{aligned}\pi_4(\Theta|E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1) \\ = k\mathbf{P}(E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1|\Theta = \theta)\pi_0(\theta).\end{aligned}$$

[1]Here and in the sequel, for a given function f we denote by $\arg\max(f)$ the points where f achieves its maximum.

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8_14

Since the events E_i are stochastically independent subordinately to the random parameter Θ, the probability

$$\mathbf{P}(E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1|\Theta = \theta)$$

can be factorized in

$$\begin{aligned}
&\mathbf{P}(E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1|\Theta = \theta) \\
&\quad = \mathbf{P}(E_1 = 0|\Theta = \theta) \cdot \mathbf{P}(E_2 = 1|\Theta = \theta)\mathbf{P}(E_3 = 1|\Theta = \theta) \cdot \mathbf{P}(E_4 = 1|\Theta = \theta) \\
&\quad = (1-\theta) \cdot \theta \cdot \theta \cdot \theta.
\end{aligned}$$

Hence the a posteriori density is given by

$$\pi_4(\Theta|E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1) = k\theta^5(1-\theta),$$

where k is a normalization constant. Since the a posteriori density π_4 corresponds to a beta distribution $\mathcal{B}(6, 2)$, we have

$$k = \frac{\Gamma(6+2)}{\Gamma(6)\Gamma(2)} = \frac{7!}{5!} = 42.$$

2. The a priori probability that Θ belongs to the interval $\left[\frac{1}{2}, 1\right]$ is given by:

$$\begin{aligned}
\mathbf{P}(\frac{1}{2} \le \Theta \le 1) &= \int_{\frac{1}{2}}^{1} \pi_0(\theta)\,\mathrm{d}\theta \\
&= \int_{\frac{1}{2}}^{1} 3\,\theta^2\,\mathrm{d}\theta = \left[\theta^3\right]_{\frac{1}{2}}^{1} = \frac{7}{8}.
\end{aligned}$$

3. The a posteriori probability that Θ belongs to the interval $\left[\frac{1}{2}, 1\right]$ is given by:

$$\begin{aligned}
&\mathbf{P}(\frac{1}{2} \le \Theta \le 1|E_1 = 0, E_2 = E_3 = E_4 = 1) \\
&= \int_{\frac{1}{2}}^{1} \pi_4(\theta|E_1 = 0, E_2 = E_3 = E_4 = 1)\,\mathrm{d}\theta \\
&= 42 \int_{\frac{1}{2}}^{1} (\theta^5 - \theta^6)\,\mathrm{d}\theta = 42\left[\frac{\theta^6}{6} - \frac{\theta^7}{7}\right]_{\frac{1}{2}}^{1} = \frac{15}{16}.
\end{aligned}$$

4. By calculating the derivative

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\pi_4(\theta|E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1) = 42\,\theta^4(5-6\theta),$$

we have that it is equal 0 in $\bar{\theta} = \dfrac{5}{6}$. Since

$$\frac{\mathrm{d}^2}{\mathrm{d}^2\theta}\pi_4\Big|_{\theta=\frac{5}{6}} \;=\; 42\,\theta^3(20-30\theta)\Big|_{\theta=\frac{5}{6}} < 0,$$

we can conclude that $\arg\max \pi_4(\Theta|E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1) = \frac{5}{6}$.

5. The a posteriori expectation of the event

$$E \;=\; E_5 \wedge E_6 = \min(E_5, E_6) = E_5E_6$$

coincides with its a posteriori probability

$$\begin{aligned}
&\mathbf{P}(E_5E_6|E_1 = 0, E_2 = E_3 = E_4 = 1)\\
&\quad= \int_0^1 \mathbf{P}(E_5E_6|\theta)\mathbf{P}(\theta|E_1 = 0, E_2 = E_3 = E_4 = 1)\,\mathrm{d}\theta\\
&\quad= \int_0^1 \mathbf{P}(E_5E_6|\theta)\pi_4(\theta|E_1 = 0, E_2 = E_3 = E_4 = 1)\,\mathrm{d}\theta\\
&\quad= \int_0^1 \mathbf{P}(E_5|\theta)\mathbf{P}(E_6|\theta)\cdot 42\,\theta^5\,(1-\theta)\,\mathrm{d}\theta\\
&\quad= \int_0^1 \theta^2\,42\,\theta^5\,(1-\theta)\,\mathrm{d}\theta\\
&\quad= 42\int_0^1 \theta^7\,(1-\theta)\,\mathrm{d}\theta\\
&\quad= \frac{\Gamma(8)}{\Gamma(6)\,\Gamma(2)}\cdot\frac{\Gamma(8)\,\Gamma(2)}{\Gamma(10)}\\
&\quad= \frac{\Gamma(8)}{\Gamma(6)}\cdot\frac{7\cdot 6\cdot\Gamma(6)}{9\cdot 8\cdot\Gamma(8)}\\
&\quad= \frac{7}{12}.
\end{aligned}$$

Here we have used the following formula[2]

$$\int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta-1}\,\mathrm{d}\theta = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \qquad \square$$

Exercise 14.2 The events $E_1, E_2, \ldots$ are stochastically independents subordinately to Θ with $\mathbf{P}(E_i|\Theta = \theta) = \theta$. The a priori density Θ is given by

[2]For this, see the proof for the density of the sum of $\Gamma(\alpha,\lambda) + \Gamma(\beta,\lambda)$, where $\Gamma(\alpha,\lambda)$, $\Gamma(\beta,\lambda)$ are stochastically independent.

$$\pi_0 = \begin{cases} K\,\theta^2\,(1-\theta) & 0 \le \theta \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

We observe the values of the first 5 events:

$$E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 0, E_5 = 1.$$

1. Compute the normalization constant K.
2. Compute the a posteriori density of the event Θ and the a posteriori probability of $(\Theta < \frac{1}{2})$.
3. Compute the a posteriori expectation of $X = E_6 + E_7$ and the a posteriori probabilities of $E = E_6 E_7$ and $F = E_6 \vee E_7$.

Solution 14.2 1. To compute the constant k, we impose that the integral of the density is equal to 1, i.e.

$$\int_0^1 \pi_0(\theta)\,\mathrm{d}\theta = 1.$$

It follows that

$$k = \frac{1}{\int_0^1 \theta^2\,(1-\theta)\,\mathrm{d}\theta}.$$

The value of this integral is well-known and equal to:

$$\int_0^1 \theta^2\,(1-\theta)\,\mathrm{d}\theta = \frac{\Gamma(3)\,\Gamma(2)}{\Gamma(3+2)},$$

hence

$$k = \frac{\Gamma(3+2)}{\Gamma(3)\,\Gamma(2)} = \frac{4!}{2!\cdot 1!} = 12.$$

2. The a posteriori density is given by

$$\begin{aligned}
&\pi_5(\Theta | E_1 = 0, E_2 = E_3 = 1, E_4 = 0, E_5 = 1) \\
&\quad = \pi_0(\theta)\mathbf{P}(E_1 = 0, E_2 = E_3 = 1, E_4 = 0, E_5 = 1|\theta) \\
&\quad = \pi_0(\theta)\mathbf{P}(E_1 = 0|\theta)\mathbf{P}(E_2 = 1|\theta)\mathbf{P}(E_3 = 1|\theta)\mathbf{P}(E_4 = 0|\theta)\mathbf{P}(E_5 = 1|\theta) \\
&\quad = c\cdot\theta^2\,(1-\theta)\cdot(1-\theta)^2\,\Theta^3 \\
&\quad = c\cdot\theta^5\,(1-\theta)^3.
\end{aligned}$$

Here we denote with c the normalization constant of the a posteriori density. In this case the a posteriori probability distribution is a beta $\mathcal{B}(6,4)$. It follows that

$$c = \frac{\Gamma(6+4)}{\Gamma(6)\,\Gamma(4)} = \frac{9!}{5!\,3!} = 7 \cdot 8 \cdot 9 = 504.$$

If we put $W = \tilde{E}_1\, E_2\, E_3\, \tilde{E}_4\, E_5$, the a posteriori density is given by

$$\pi_5(\theta) = \begin{cases} 504\,\theta^5\,(1-\theta)^3 & 0 \le \theta \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

To find the a posteriori probability of the event $(\Theta < \frac{1}{2})$ it is sufficient to integrate the a posteriori density between 0 and $\frac{1}{2}$:

$$\mathbf{P}(\Theta < \frac{1}{2}) = 504 \int_0^{\frac{1}{2}} \theta^5\,(1-\theta)^3 d\theta = 504 \int_0^{\frac{1}{2}} (\theta^5 + 3\theta^7 - 3\theta^6 - \theta^8) d\theta = \frac{65}{256}.$$

3. The a posteriori expectation of $X = E_6 + E_7$ is given by

$$\begin{aligned} \mathbf{P}(X|W) &= \sum_{i=0}^{2} i\,\mathbf{P}(X = i|W) \\ &= \mathbf{P}(X = 1|W) + 2\,\mathbf{P}(X = 2|W). \end{aligned}$$

We obtain:

$$\begin{aligned} \mathbf{P}(X = 1|W) &= \mathbf{P}(E_6 = 1,\ E_7 = 0|W) + \mathbf{P}(E_6 = 0,\ E_7 = 1|W) \\ &= 2 \int_0^1 \mathbf{P}(E_6 = 0,\ E_7 = 1|\Theta = \theta)\pi_5(\Theta = \theta|W)\,\mathrm{d}\theta \\ &= 2 \int_0^1 \frac{\Gamma(10)}{\Gamma(6)\,\Gamma(4)} \theta \cdot (1-\theta) \cdot \theta^5 \cdot (1-\theta)^3\,\mathrm{d}\theta \\ &= 2\,\frac{\Gamma(10)}{\Gamma(6)\,\Gamma(4)} \int_0^1 \theta^6 (1-\theta)^4\,\mathrm{d}\theta \\ &= 2\,\frac{\Gamma(10)}{\Gamma(6)\,\Gamma(4)}\,\frac{\Gamma(7)\,\Gamma(5)}{\Gamma(12)} \\ &= \frac{24}{55} \end{aligned}$$

and

$$\begin{aligned} \mathbf{P}(X = 2|W) &= \mathbf{P}(E_6 = 1,\ E_7 = 1|W) \\ &= \int_0^1 \mathbf{P}(E_6 = 1,\ E_7 = 1|\theta = \Theta)\pi_5(\Theta = \theta|W)\,\mathrm{d}\theta \end{aligned}$$

$$
\begin{aligned}
&= \frac{\Gamma(10)}{\Gamma(6)\,\Gamma(4)} \int_0^1 \theta^2 \cdot \theta^5 \cdot (1-\theta)^3 \,\mathrm{d}\theta \\
&= \frac{\Gamma(10)}{\Gamma(6)\,\Gamma(4)} \int_0^1 \theta^7 (1-\theta)^3 \,\mathrm{d}\theta \\
&= \frac{\Gamma(10)}{\Gamma(6)\,\Gamma(4)} \frac{\Gamma(8)\,\Gamma(4)}{\Gamma(12)} \\
&= \frac{21}{55}.
\end{aligned}
$$

The a posteriori expectation of X is equal to

$$\mathbf{P}(X|W) \;=\; 2 \cdot \frac{21}{55} + \frac{24}{55} \;=\; \frac{6}{5}.$$

Note that $X = E_6 + E_7$ is a random number, but not an event since it can assume 3 possible values: 0, 1 or 2.
The a posteriori probability of the events $E = E_6\,E_7$ e $F = E_6 \vee E_7$ can be calculated in the same way:

$$\mathbf{P}(E|W) = \mathbf{P}(E_6E_7 = 1|W) = \mathbf{P}(E_6 = E_7 = 1|W) = \frac{21}{55}$$

and

$$
\begin{aligned}
\mathbf{P}(F|W) &= \mathbf{P}(E_6 \vee E_7 = 1|W) \\
&= \mathbf{P}(E_6 = 1,\; E_7 = 0|W) \\
&\quad + \mathbf{P}(E_6 = 0,\; E_7 = 1|W) \\
&\quad + \mathbf{P}(E_6 = 1 = E_7 = 1|W) \\
&= \frac{9}{11}.
\end{aligned}
$$

□

Exercise 14.3 The events $E_1, E_2, \ldots$ are stochastically independents subordinately to Θ with $\mathbf{P}(E_i|\Theta = \theta) = \theta$. The a prior density Θ is given by

$$\pi_0(\theta) = \begin{cases} K\theta^2(1-\theta)^2 & \text{for } 0 \le \theta \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

We observe the values of the first 4 events: $E_1 = 0$, $E_2 = 1$, $E_3 = E_4 = 1$.

(a) Compute the normalization constant K.
(b) Compute the a posteriori density and the a posteriori expectation of Θ.
(c) Compute the a posteriori probability of the event $F = E_5^2$ and the a posteriori variance of expectation of $\tilde{E}_6$.

Solution 14.3 (a) To compute K we impose that

$$\int_0^1 \pi_0(\theta)\mathrm{d}\theta = 1$$

i.e.

$$K = \frac{1}{\int_0^1 \theta^2(1-\theta)^2\mathrm{d}\theta},$$

since the integral of a probability density must be equal to 1. The integral appearing at the denominator is well-known and equal to

$$\int_0^1 \theta^2(1-\theta)^2\mathrm{d}\theta = \frac{\Gamma(3)^2}{\Gamma(6)}$$

hence

$$K = \frac{\Gamma(6)}{\Gamma(3)^2} = \frac{5!}{(2!)^2} = 30.$$

(b) The a posteriori density of Θ given the events $E_1 = 0$, $E_2 = E_3 = E_4 = 1$ is given by the formula

$$\begin{aligned}
&\pi_4(\theta|E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1)\\
&\quad = K\pi_0(\theta)\mathbf{P}(E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1|\theta)\\
&\quad = \mathbf{P}(E_1 = 0|\theta)\cdot\mathbf{P}(E_2 = 1|\theta)\cdot\mathbf{P}(E_3 = 1|\theta)\cdot\mathbf{P}(E_4 = 1|\theta)\\
&\quad = K\theta^5(1-\theta)^3,
\end{aligned}$$

where $K = \dfrac{\Gamma(10)}{\Gamma(6)\Gamma(4)} = 504$ and $\theta \in [0, 1]$. For $\theta \notin [0, 1]$, the a posteriori density is equal to 0. To compute the a posteriori expectation of Θ, we apply the formula of the expectation for absolutely continuous distributions, i.e.

$$\begin{aligned}
&\mathbf{P}(\Theta|E_1 = 0, E_2 = E_3 = E4 = 1)\\
&\quad = \int_0^1 \theta\pi_4(\theta|E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1)\mathrm{d}\theta\\
&\quad = \frac{\Gamma(10)}{\Gamma(6)\Gamma(4)}\int_0^1 \theta^6(1-\theta)^3\mathrm{d}\theta\\
&\quad = \frac{\Gamma(10)}{\Gamma(6)\Gamma(4)}\cdot\frac{\Gamma(7)\Gamma(4)}{\Gamma(11)} = \frac{3}{5}.
\end{aligned}$$

(c) The event $F = E_5^2$ coincides with E_5 since it assumes only the value 0 or 1. The a posteriori probability of F is given by

$$\begin{aligned}
&\mathbf{P}(F|E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1) \\
&\quad = \mathbf{P}(E_5^2|E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1) \\
&\quad = \mathbf{P}(E_5|E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1) \\
&\quad = \int_0^1 \theta\pi_4(\theta|E_1 = 0, E_2 = 1, E_3 = 1, E_4 = 1)\mathrm{d}\theta = \frac{3}{5}.
\end{aligned}$$

To compute the a posteriori variance of $\tilde{E}_6$, we consider the usual formula for the variance. To simplify the notations, we put $A = \widetilde{E}_1 E_2 E_3 E_4$. We obtain

$$\begin{aligned}
&\boldsymbol{\sigma}^2(\tilde{E}_6|\widetilde{E}_1 E_2 E_3 E_4) \\
&\quad = \mathbf{P}(\tilde{E}_6^2|\widetilde{E}_1 E_2 E_3 E_4) - \mathbf{P}(\tilde{E}_6|\widetilde{E}_1 E_2 E_3 E_4)^2 \\
&\quad = \mathbf{P}(\tilde{E}_6|\widetilde{E}_1 E_2 E_3 E_4) - \mathbf{P}(\tilde{E}_6|\widetilde{E}_1 E_2 E_3 E_4)^2 \\
&\quad = \mathbf{P}(\tilde{E}_6|\widetilde{E}_1 E_2 E_3 E_4)(1 - \mathbf{P}(\tilde{E}_6|\widetilde{E}_1 E_2 E_3 E_4)),
\end{aligned}$$

where we have used that

$$\tilde{E}_6^2 = \tilde{E}_6.$$

We need only to compute the a posteriori expectation of $\tilde{E}_6$. For this purpose we apply the formula of the total probabilities. Hence

$$\begin{aligned}
\mathbf{P}(\tilde{E}_6^2|\widetilde{E}_1 E_2 E_3 E_4) &= \mathbf{P}(\tilde{E}_6|\widetilde{E}_1 E_2 E_3 E_4) \\
&= \mathbf{P}(\tilde{E}_6 E_5|\widetilde{E}_1 E_2 E_3 E_4) \\
&\quad + \mathbf{P}(\tilde{E}_6 \tilde{E}_5|\widetilde{E}_1 E_2 E_3 E_4) \\
&= \int_0^1 \mathbf{P}(\tilde{E}_6 E_5|\theta)\pi_4(\theta|\widetilde{E}_1 E_2 E_3 E_4)\mathrm{d}\theta \\
&\quad + \int_0^1 \mathbf{P}(\tilde{E}_6 \tilde{E}_5|\theta)\pi_4(\theta|\widetilde{E}_1 E_2 E_3 E_4)\mathrm{d}\theta \\
&= \int_0^1 \theta(1-\theta)\pi_4(\theta|\widetilde{E}_1 E_2 E_3 E_4)\mathrm{d}\theta \\
&\quad + \int_0^1 (1-\theta)(1-\theta)\pi_4(\theta|\widetilde{E}_1 E_2 E_3 E_4)\mathrm{d}\theta \\
&= \int_0^1 (1-\theta)[\theta + 1 - \theta]\pi_4(\theta|\widetilde{E}_1 E_2 E_3 E_4)\mathrm{d}\theta \\
&= \int_0^1 (1-\theta)\pi_4(\theta|\widetilde{E}_1 E_2 E_3 E_4)\mathrm{d}\theta
\end{aligned}$$

$$= 1 - \int_0^1 \theta \pi_4(\theta | \widetilde{E}_1 E_2 E_3 E_4) \mathrm{d}\theta$$
$$= \frac{2}{5}.$$

Note that the a posteriori probabilities of E_5 and E_6 coincide. □

Exercise 14.4 The events $E_1, E_2, \ldots$ are stochastically independent subordinately to Θ with $\mathbf{P}(E_i|\Theta = \theta) = \theta$. The a priori density of Θ is given by

$$\pi_0(\theta) = \begin{cases} K\theta^2\sqrt{1-\theta} & \text{for } 0 \le \theta \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

We observe the values of the first 4 events: $E_1 = 1$, $E_2 = 0$, $E_3 = 0$, $E_4 = 1$.

(a) Compute the normalization constant K.
(b) Compute the a posteriori density $\pi_4(\theta | E_1 = 1, E_2 = E_3 = 0, E_4 = 1)$ of Θ and $\arg\max \pi_4(\theta | E_1 = 1, E_2 = E_3 = 0, E_4 = 1)$.
(c) Compute the a posteriori covariance of the events E_6 and E_7.

Solution 14.4 (a) The normalization constant K makes the integral of the density equal to 1, hence

$$K = \frac{1}{\int_0^1 \theta^2 (1-\theta)^{\frac{1}{2}} \mathrm{d}\theta}.$$

We know that

$$\int_0^1 \theta^2 (1-\theta)^{\frac{1}{2}} \mathrm{d}\theta = \frac{\Gamma(3)\Gamma\left(\frac{3}{2}\right)}{\Gamma\left(3 + \frac{3}{2}\right)},$$

hence

$$K = \frac{\Gamma\left(\frac{9}{2}\right)}{\Gamma(3)\Gamma\left(\frac{3}{2}\right)} = \frac{\frac{7}{2} \cdot \frac{5}{2} \cdot \frac{3}{2} \cdot \Gamma\left(\frac{3}{2}\right)}{2!\Gamma\left(\frac{3}{2}\right)} = \frac{105}{16}.$$

(b) We compute the a posteriori density by using the fact that the events are stochastically independent subordinately to Θ. We have

$$\begin{aligned} &\pi_4(\theta | E_1 = 1, E_2 = E_3 = 0, E_4 = 1) \\ &\quad = K\mathbf{P}(E_1 = 1, E_2 = E_3 = 0, E_4 = 1|\theta)\pi_0(\theta) \\ &\quad = K\mathbf{P}(E_1 = 1|\theta) \cdot \mathbf{P}(E_2 = 0|\theta) \cdot \mathbf{P}(E_3 = 0|\theta) \cdot \mathbf{P}(E_4 = 1|\theta)\pi_0(\theta) \\ &\quad = K\theta^4(1-\theta)^{\frac{5}{2}}, \end{aligned}$$

where

$$K = \frac{\Gamma\left(5+\frac{7}{2}\right)}{\Gamma(5)\Gamma\left(\frac{7}{2}\right)} = \frac{\frac{15}{2}\cdot\frac{13}{2}\cdot\frac{11}{2}\cdot\frac{9}{2}\Gamma\left(\frac{7}{2}\right)}{\Gamma(5)\Gamma\left(\frac{7}{2}\right)} = \frac{6435}{128}.$$

Hence

$$\pi_4(\theta|E_1=1, E_2=E_3=0, E_4=1) = \begin{cases} \frac{6435}{128}\theta^4(1-\theta)^{\frac{5}{2}} & \theta\in[0,1], \\ 0 & \text{otherwise.} \end{cases}$$

The arg max can be computed by finding the zeros of the first derivative. We have

$$\begin{aligned} \pi_4'(\theta|E_1=1, E_2=E_3=0, E_4=1) &= K\left[4\theta^3(1-\theta)^{\frac{5}{2}} - \frac{5}{2}\theta^4(1-\theta)^{\frac{3}{2}}\right] \\ &= K\theta^3(1-\theta)^{\frac{3}{2}}\left[4(1-\theta)-\frac{5}{2}\theta\right] \\ &= K\frac{\theta^3}{2}(1-\theta)^{\frac{3}{2}}[8-13\theta]. \end{aligned}$$

The derivative is equal to 0 in the extremes of the interval as well in

$$\bar{\theta} = \frac{8}{13}.$$

Since $\pi_4' > 0$ for $\theta \in \left[0, \frac{8}{13}\right)$ and $\pi_4' < 0$ for $\theta \in \left(\frac{8}{13}, 1\right]$, we have that $\arg\max \pi_4(\theta|E_1=1, E_2=E_3=0, E_4=1) = \frac{8}{13}$.

(c) The a posteriori covariance of the events E_6 and E_7 is given by

$$\begin{aligned} &\mathbf{cov}(E_6, E_7|E_1=1, E_2=E_3=0, E_4=1) \\ &\quad = \mathbf{P}(E_6E_7|E_1=1, E_2=E_3=0, E_4=1) \\ &\qquad - \mathbf{P}(E_6|E_1=1, E_2=E_3=0, E_4=1)\mathbf{P}(E_7|E_1=1, E_2=E_3=0, E_4=1). \end{aligned}$$

In Exercise 14.3 we have proved that

$$\begin{aligned} &\mathbf{P}(E_6|E_1=1, E_2=E_3=0, E_4=1) \\ &\quad = \mathbf{P}(E_7|E_1=1, E_2=E_3=0, E_4=1) \\ &\quad = \int_0^1 \theta\pi_4(\theta|E_1=1, E_2=E_3=0, E_4=1)\mathrm{d}\theta = \frac{10}{17}. \end{aligned}$$

Analogously

$$\begin{aligned}&\mathbf{P}(E_6E_7|E_1=1,E_2=E_3=0,E_4=1)\\&=\int_0^1\theta^2\pi_4(\theta|E_1=1,E_2=E_3=0,E_4=1)\mathrm{d}\theta=\frac{120}{323}.\end{aligned}$$

We conclude that

$$\mathbf{cov}(E_6,E_7|E_1=1,E_2=E_3=0,E_4=1)=\frac{120}{323}-\left(\frac{10}{17}\right)^2.\qquad\square$$

Exercise 14.5 The random numbers $X_1, X_2, \ldots$ are stochastically independents subordinately to Θ with the same conditional marginal density given by

$$f(x|\theta)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-\theta)^2}{2}\right),\qquad x\in\mathbb{R}.$$

We assume that Θ has standard normal distribution. We observe the values of the first 4 experiments:

$$x_1=0.1,\qquad x_2=2,\qquad x_3=-1,\qquad x_4=0.5.$$

(a) Write the a priori density of Θ.
(b) Compute the a posteriori density $\pi_4(\theta|\,x_1=0.1,x_2=2,x_3=-1,x_4=0.5)$ of Θ and $\arg\max\pi_4(\theta|\,x_1=0.1,x_2=2,x_3=-1,x_4=0.5)$.
(c) Compute the a posteriori expectation and variance of Θ.

Solution 14.5 (a) Since Θ has a standard normal distribution as a priori distribution, we can write immediately the a priori density

$$\pi_0(\theta)=\frac{1}{\sqrt{2\pi}}e^{-\frac{\theta^2}{2}},\qquad\theta\in\mathbb{R}.$$

(b) We compute the a posteriori density by using the fact that the random numbers are stochastically independent subordinately to Θ:

$$\begin{aligned}&\pi_4(\theta|\,x_1=0.1,x_2=2,x_3=-1,x_4=0.5)\\&=kf(x_1,x_2,x_3,x_4|\theta)\pi_0(\theta)\\&=k\Pi_{i=1}^4f(x_i|\theta)\pi_0(\theta)=k\exp\left(-\frac{\sum_{i=1}^4(x_i-\theta)^2+\theta^2}{2}\right)\\&=k\exp\left(-\frac{5}{2}(\theta-\frac{8}{25})^2\right),\end{aligned}$$

Note that all factors which are independent of θ are now included in the constant k. We obtain that the a posteriori distribution is Gaussian $N(\frac{8}{25}, \frac{1}{5})$, hence

$$k = \frac{\sqrt{5}}{\sqrt{2\pi}}.$$

The graph of the a posteriori density is *bell shaped* with symmetry axis $x = \frac{8}{25}$. Then $\arg\max \pi_4(\theta |\, x_1 = 0.1, x_2 = 2, x_3 = -1, x_4 = 0.5) = \frac{8}{25}$. Verify this by computing the derivatives of the density function.

(c) The parameters of the a posteriori density provide us with:

1. the a posteriori expectation

$$\mathbf{P}(\Theta |\, x_1 = 0.1, x_2 = 2, x_3 = -1, x_4 = 0.5) = \frac{8}{25};$$

2. the a posteriori variance

$$\boldsymbol{\sigma}^2(\Theta |\, x_1 = 0.1, x_2 = 2, x_3 = -1, x_4 = 0.5) = \frac{1}{5}.$$ □

Exercise 14.6 The random numbers $X_1, X_2, \ldots$ are stochastically independent subordinately to Θ with the same conditional marginal density given by

$$f(x|\theta) = \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{(x-\theta)^2}{8}\right), \qquad x \in \mathbb{R}.$$

The a priori distribution of Θ is given by

$$\pi_0(\theta) = \frac{1}{\sqrt{4\pi}} \exp\left(-\frac{(\theta-1)^2}{4}\right), \qquad \theta \in \mathbb{R}.$$

We observe the values of the first 3 experiments:

$$x_1 = 1, \qquad x_2 = 0.5, \qquad x_3 = -1.$$

(a) Compute the likelihood factor.
(b) Compute the a posteriori density of Θ.
(c) Estimate the a posteriori probability of the event $(\Theta > 1000)$.

Solution 14.6 (a) By definition, the likelihood factor is given by

$$\begin{aligned} f(x_1, x_2, x_3|\theta) &= \Pi_{i=1}^3 f(x_i|\theta) \\ &= \frac{1}{8\sqrt{(2\pi)^3}} \exp\left(-\frac{\sum_{i=1}^3 (x_i-\theta)^2}{8}\right) \\ &= \frac{1}{8\sqrt{(2\pi)^3}} \exp\left(-\frac{1}{8}\left(3\theta^2-\theta+\frac{9}{4}\right)\right). \end{aligned}$$

(b) By the computations for the likelihood factor, we immediately obtain the a posteriori density as follows

$$\begin{aligned} &\pi_3(\theta|\, x_1 = 1, x_2 = 0.5, x_3 = -1) \\ &\quad = kf(x_1, x_2, x_3, x_4|\theta)\pi_0(\theta) \\ &\quad = k \exp\left(-\frac{1}{8}\left(3\theta^2-\theta+\frac{9}{4}\right) - \frac{(\theta-1)^2}{4}\right) \\ &\quad = k \exp\left(-\frac{5}{8}(\theta-\frac{1}{2})^2\right), \end{aligned}$$

where we have put in the constant k all terms which are independent of θ. We obtain that the a posteriori distribution is a normal distribution $N(\frac{1}{2}, \frac{4}{5})$ with normalization constant $k = \frac{\sqrt{5}}{2\sqrt{2\pi}}$.

(c) To estimate the a posteriori probability of the event $(\Theta > 1000)$ we use the tail estimation for the Gaussian distribution. To this purpose we need first to express Θ as function of a random variable Y with distribution $N(0, 1)$. Since the a posteriori distribution of Θ is Gaussian $N(\frac{1}{2}, \frac{4}{5})$, we have

$$\Theta = \frac{2\sqrt{5}}{5}Y + \frac{1}{2},$$

where $Y \sim N(0, 1)$. Hence

$$\mathbf{P}(\Theta > 1000) = \mathbf{P}(\frac{2\sqrt{5}}{5}Y + \frac{1}{2} > 1000) = \mathbf{P}(Y > \frac{\sqrt{5}}{2} \cdot 999, 5).$$

By the tail estimation of the standard normal distribution, we obtain that

$$\frac{n(x)}{x} - \frac{n(x)}{x^3} < \mathbf{P}(Y > x) < \frac{n(x)}{x},$$

where $x > 0$, $n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. To obtain an upper bound for $\mathbf{P}(\Theta > 1000)$ we can compute $\frac{n(x)}{x}$ in the point

$$x = \frac{\sqrt{5}}{2} \cdot 999, 5. \qquad \square$$

Exercise 14.7 The random numbers $X_1, X_2, \ldots$ are stochastically independent subordinately to Φ with the same conditional marginal density given by

$$f(x|\phi) = \frac{1}{\sqrt{2\pi}} \phi^{\frac{1}{2}} \exp\left(-\frac{\phi(x-1)^2}{2}\right), \qquad x \in \mathbb{R}.$$

The a priori distribution of Φ is given by a Gamma distribution $\Gamma(2, 1)$. We observe the values of the first 3 experiments:

$$x_1 = 1.5, \qquad x_2 = 0.5, \qquad x_3 = 2.$$

(a) Write the a priori density of Φ.
(b) Compute the a posteriori density $\pi_3(\phi|\, x_1 = 1.5, x_2 = 0.5, x_3 = 2)$ of Φ and $\arg\max \pi_3(\phi|\, x_1 = 1.5, x_2 = 0.5, x_3 = 2)$.
(c) Compute the a posteriori expectation and variance of Φ.

Solution 14.7 (a) Since the a priori distribution of Φ is given by a Gamma distribution $\Gamma(2, 1)$, we can write immediately the a priori density:

$$\pi_0(\phi) = \begin{cases} \phi e^{-\phi} & \phi \geq 0, \\ 0 & \phi < 0. \end{cases}$$

(b) The a posteriori density is given by

$$\begin{aligned} &\pi_3(\phi|\, x_1 = 1.5, x_2 = 0.5, x_3 = 2) \\ &\quad = k f(x_1, x_2, x_3|\phi)\pi_0(\phi) \\ &\quad = k\phi^{\frac{5}{2}} \exp\left(-\left(\frac{\sum_{i=1}^3 (x_i - 1)^2}{2} + 1\right)\phi\right) \\ &\quad = k\phi^{\frac{5}{2}} e^{-\frac{7}{4}\phi} \end{aligned}$$

for $\phi \geq 0$, 0 otherwise. Note that we have put in the constant k all factors which are independent of ϕ. The a posteriori distribution is then a Gamma distribution $\Gamma(\frac{7}{2}, \frac{7}{4})$ with normalization constant

$$k = \left(\frac{7}{4}\right)^{\frac{7}{2}} \frac{1}{\Gamma(\frac{7}{2})} = \frac{\sqrt{7^7}}{240\sqrt{\pi}}.$$

Furthermore we have that

$$\frac{d}{d\phi}\pi_3(\phi|\, x_1 = 1.5, x_2 = 0.5, x_3 = 2) = k\phi^{\frac{3}{2}} e^{-\frac{7}{4}\phi}(\frac{5}{2} - \frac{7}{4}\phi) = 0$$

if $\phi = \dfrac{10}{7}$. We immediately obtain that $\arg\max \pi_3(\phi|\, x_1 = 1.5, x_2 = 0.5, x_3 = 2) = \frac{10}{7}$ by analyzing the sign of the first derivative.

(c) The parameters of the a posteriori density provide us with

1. the a posteriori expectation

$$\mathbf{P}(\Phi|\, x_1 = 1.5, x_2 = 0.5, x_3 = 2) = 2\,;$$

2. the a posteriori variance

$$\boldsymbol{\sigma}^2(\Theta|\, x_1 = 0.1, x_2 = 2, x_3 = -1, x_4 = 0.5) = \frac{8}{7}. \qquad \square$$

Exercise 14.8 The random numbers $X_1, X_2, \ldots$ are stochastically independent subordinately to Φ with the same conditional marginal density given by

$$f(x|\phi) = \frac{1}{\sqrt{2\pi}}\phi^{\frac{1}{2}} \exp\left(-\frac{\phi x^2}{2}\right), \qquad x \in \mathbb{R}.$$

The a priori distribution of Φ is given by an exponential distribution with parameter $\lambda = 2$. We observe the values of the first 4 experiments:

$$x_1 = 1, \qquad x_2 = 2, \qquad x_3 = 0.5, \qquad x_4 = \sqrt{2}.$$

(a) Write the a priori density of Φ and the a priori probability of the event $(\Phi > 2)$.
(b) Compute the a posteriori density of Φ and the a posteriori probability of the event $(\Phi > 2)$.
(c) Compute the a posteriori expectation of $Z = \Phi^2$.

Solution 14.8 (a) Since the a priori distribution of Φ is given by an exponential distribution with parameter $\lambda = 2$, i.e. a Gamma distribution $\Gamma(1, 2)$, we can write immediately the a priori density:

$$\pi_0(\phi) = \begin{cases} 2e^{-2\phi} & \phi \geq 0, \\ 0 & \phi < 0. \end{cases}$$

The a priori probability of the event $(\Phi > 2)$ is given by

$$\mathbf{P}(\Phi > 2) = \int_2^{+\infty} 2e^{-2\phi} \mathrm{d}\phi = e^{-4}.$$

(b) The a posteriori density is given by

$$\begin{aligned}
&\pi_4(\phi |\, x_1 = 1, x_2 = 2, x_3 = 0.5, x_4 = \sqrt{2}) \\
&\quad = kf(x_1, x_2, x_3, x_4|\phi)\pi_0(\phi) \\
&\quad = k\phi^2 \exp\left(-\left(\frac{\sum_{i=1}^4 x_i^2}{2} + 2\right)\phi\right) \\
&\quad = k\phi^2 e^{-\frac{45}{8}\phi}
\end{aligned}$$

for $\phi \geq 0$, 0 otherwise. Note that we have put in the constant k all factors which are independent of ϕ. The a posteriori distribution is then a Gamma distribution $\Gamma(\alpha_4, \lambda_4) = \Gamma(3, \dfrac{45}{8})$ with normalization constant

$$k = \left(\frac{45}{8}\right)^3 \frac{1}{\Gamma(3)} = \frac{45^3}{2^{10}}.$$

The a posteriori probability of the event $(\Phi > 2)$ is given by

$$\begin{aligned}
&\mathbf{P}(\Phi > 2|\, x_1 = 1, x_2 = 2, x_3 = 0.5, x_4 = \sqrt{2}) \\
&\quad = \int_2^{+\infty} \pi_4(\phi|\, x_1 = 1, x_2 = 2, x_3 = 0.5, x_4 = \sqrt{2})\mathrm{d}\phi \\
&\quad = k \int_2^{+\infty} \phi^2 e^{-\frac{45}{8}\phi} \mathrm{d}\phi \\
&\quad = k\frac{8}{45}\left[-\left(\phi^2 + 2\phi + 2\right) e^{-\frac{45}{8}\phi}\right]_2^{+\infty} = \frac{3^4 5^3}{2^6} e^{-\frac{45}{4}}.
\end{aligned}$$

(c) To compute the a posteriori expectation of $Z = \Phi^2$ it is sufficient to note that

$$\begin{aligned}
&\mathbf{P}(\Phi^2|\, x_1 = 1, x_2 = 2, x_3 = 0.5, x_4 = \sqrt{2}) \\
&\quad = \boldsymbol{\sigma}^2(\Phi|\, x_1 = 1, x_2 = 2, x_3 = 0.5, x_4 = \sqrt{2}) \\
&\quad + \mathbf{P}(\Phi|\, x_1 = 1, x_2 = 2, x_3 = 0.5, x_4 = \sqrt{2}) \\
&\quad = \frac{\alpha_4}{\lambda_4^2} + \frac{\alpha_4^2}{\lambda_4^2} = \frac{256}{675}.
\end{aligned}$$

□

Appendix A
Elements of Combinatorics

Consider a set $\Omega = \{a_1, \ldots, a_n\}$ of n elements. We recall that the symbol $\binom{n}{r}$ is called *binomial coefficient* and that $\binom{n}{r} = \dfrac{n!}{r!n-r!}$.

A.1 Dispositions

We count the number of ways of choosing r elements out of a set of n elements with repetitions and taking in account of their order, i.e. the number of *dispositions* of r elements out of n. We have:

$$
\begin{aligned}
\text{1st element} &\longrightarrow n \text{ choices},\\[1em]
\text{2nd elements} &\longrightarrow n \text{ choices},\\[1em]
&\;\;\vdots \qquad\quad \vdots\\[1em]
r\text{th elements} &\longrightarrow n \text{ choices}\,.
\end{aligned}
$$

Totally, the dispositions are $n \cdot n \ldots n = n^r$. They count the number of functions from a set of r elements to a set of n elements.

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8

A.2 Simple Dispositions

We count the number of ways of choosing r elements out of a set of n elements without repetitions and taking in account of their order, i.e. the number of *simple dispositions* of r elements out of n. We have:

$$1^o \text{ element} \longrightarrow n \text{ choices},$$

$$2^o \text{ elements} \longrightarrow (n-1) \text{ choices},$$

$$3^o \text{ elements} \longrightarrow (n-2) \text{ choices},$$

$$\vdots \qquad \vdots$$

$$r^o \text{ elements} \longrightarrow (n-r+1) \text{ choices}.$$

Totally, the simple dispositions are $n \cdot (n-1) \dots (n-r+1) = \dfrac{n!}{(n-r)!}$ and are denoted by the symbol D_r^n or $(n)_r$. They count the number of injective functions from a set of r elements to a set of n elements. If $r = n$, they are called *permutations*.

A.3 Simple Combinations

We count the number of ways of choosing r elements out of a set of n elements without repetitions and without taking in account of their order, i.e. the number of *simple combinations* of r elements out of n. Given a simple combination of r elements out of n, we obtain $r!$ dispositions by permutating the r elements. The number of simple combinations is then

$$\frac{1}{r!} D_r^n = \frac{n!}{r!(n-r)!} = \binom{n}{r}.$$

They count the number of injective functions from a set of r elements to a set of n elements which have a different image.

A.4 Combinations

We count the number of ways of choosing r elements out of a set of n elements without taking in account of their order, i.e. the number of *combinations* of r elements out of n. Given a combination $\{a_1, \ldots, a_r\}$, without loss of generality, we can suppose that $a_1 \leq \cdots \leq a_r$. Starting from this combination, we now construct a simple combination of r elements out of $n + r - 1$ elements in the following way:

$$\begin{aligned} b_1 &= a_1 \,, \\ b_2 &= a_2 + 1 \,, \\ &\vdots \\ b_r &= a_r + r - 1 \,. \end{aligned}$$

On the other way round, we can always associate a combination to a simple combination. Hence the r-combinations are as many as the r-simple combinations in $n + r - 1$, elements, i.e. $\binom{n+r-1}{r}$.

A.5 Multinomial Coefficient

The number of ways of forming k groups of $r_1, \ldots, r_k$ elements respectively, where $r_1 + \cdots + r_k = n$ is given by the *multinomial coefficient*

$$\frac{n!}{r_1! r_2! \ldots r_k!} \,.$$

To form the first group of r_1 elements, we have $\binom{n}{r_1}$ possibilities. For the second group, we have $\binom{n-r_1}{r_2}$ ways. Analogously we proceed for the remaining groups. We obtain

$$\binom{n}{r_1}\binom{n-r_1}{r_2}\cdots\binom{n-r_1-\cdots-r_{k-1}}{r_k} = \frac{n!}{r_1! r_2! \ldots r_k!} \,.$$

Appendix B
Relations Between Discrete and Absolutely Continuous Distributions

In Table B.1 we summarize some analogies between discrete and absolutely continuous distributions.

Table B.1 Some analogies between discrete and absolutely continuous distributions

C. Discrete		C. Abs. Continuous
Probability		Density
$P(X = x)$	$\longrightarrow$	$f(x)$
Cumulative $\sum_{i \in I(X), i \leq x} P(X = i)$	distribution function $\longrightarrow$	$P(X \leq x)$ $\int_{-\infty}^{x} f(s)\,\mathrm{d}s$
$\sum_{i \in I(X)} i\, P(X = i)$	Expectation of X $\longrightarrow$	$\int_{-\infty}^{+\infty} s\, f(s)\,\mathrm{d}s$
Expectation of $\sum_{i \in I(X)} \Psi(i)\, P(X = i)$	$Y = \Psi(X)$ $\longrightarrow$	$\int_{-\infty}^{+\infty} \Psi(s)\, f(s)\,\mathrm{d}s$
	$P(X \in A)$	
$\sum_{i \in I(X), i \in A} P(X = i)$	$\longrightarrow$	$\int_A f(s)\,\mathrm{d}s$

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8

Appendix C
Some Discrete Distributions

We present in Table C.1 an overview of the discrete distributions presented in Chap. 2.

Table C.1 Some discrete distributions

Distribution	$I(X)$	$\mathbf{P}(X=k)$	$\mathbf{P}(X)$	$\sigma^2(X)$
Bernoulli p	$\{0, 1\}$	$\mathbf{P}(X=1)=p$	p	$p(1-p)$
Binomial Bn(n, p)	$\{0, \dots, n\}$	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
Geometric p	$\{1, 2, \dots\}$	$p(1-p)^{k-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Hypergeometric (n, N, b)	$\{0 \vee (n-(N-b)), \dots, n \wedge b\}$	$\dfrac{\binom{b}{k}\binom{N-b}{n-k}}{\binom{N}{n}}$	$n\frac{b}{N}$	$n\left(\frac{N-n}{N-1}\right)\frac{b}{N}\left(1-\frac{b}{N}\right)$
Poisson λ	$\mathbb{N}$	$\frac{\lambda^k}{k!}e^{-\lambda}$	λ	λ

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8

Appendix D
Some One-Dimensional Absolutely Continuous Distributions

We recall in Table D.1 the most common one-dimensional absolutely continuous distributions.

Table D.1 Some one-dimensional absolutely continuous distributions

Distribution	$I(X)$	Density	$\mathbf{P}(X)$	$\sigma^2(X)$
Uniform $[a, b]$	$[a, b]$[a]	$\frac{1}{b-a}\, I_{[a,b]}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential λ	$\mathbb{R}^+$	$\lambda\, e^{-\lambda x}\, I_{\{x\geq 0\}}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Std. normal $N(0, 1)$	$\mathbb{R}$	$\frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$	0	1
Gen. normal $N(\mu, \sigma^2)$	$\mathbb{R}$	$\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Gamma $\Gamma(\alpha, \beta)$	$\mathbb{R}^+$	$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\, I_{\{x\geq 0\}}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Beta $\mathcal{B}(\alpha, \beta)$	$[0, 1]$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1}\,(1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)}$

[a] $b > a$

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8

Appendix E
The Normal Distribution

We present in Table E.1 a summary on the normal distribution.

Table E.1 The normal distribution in a nutshell

Density	$f(x_1, \ldots, x_n) = k\, e^{(-\frac{1}{2}Ax\cdot x + b\cdot x)}$ $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, A \in \mathcal{S}(n), b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$
Normalization constant	$k = \sqrt{\frac{\det A}{(2\pi)^n}}\, e^{-\frac{1}{2}A^{-1}b\cdot b}$
Expectation	$\mathbf{P}(X) = A^{-1}b \Rightarrow \mathbf{P}(X_i) = (A^{-1}b)_i$
Variance and covariance matrix	$C = A^{-1}$
Marginal distribution of X_i	$X_i \sim N\left((A^{-1}b)_i, [A^{-1}]_{ii}\right)$

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8

Appendix F
Stirling's Formula

In this chapter we present *Stirling's formula*, which describes the asymptotic behavior of $n!$ with n increasing. It holds that:

$$\textbf{Stirling's formula:} \quad n! = \sqrt{2\pi}\, n^{n+\frac{1}{2}}\, e^{-n}(1 + O(n^{-1}))\,.$$

Different kinds of proofs can be used to prove this formula. Here we present the classical proof and a more general result, of which the Stirling's formula represents a particular case.

We start with the classical proof which can be found in [2]. Here we recall it for reader's convenience.

F.1 First Proof

Here we obtain Stirling's formula modulo a multiplicative constant. This value can be shown to be equal to $\sqrt{2\pi}$, as a consequence of Theorem 5.4.1 by approximating the probability that a random number with binomial distribution $\mathrm{Bn}(2n, \frac{1}{2})$ assumes the value n.

The Stirling's formula is equivalent to

$$\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi}\, n^{n+\frac{1}{2}}\, e^{-n}} = 1.$$

In order to compute this limit, we look for an estimation of

$$\log n! = \log(1 \cdot 2 \cdot \ldots n) = \log 1 + \log 2 + \cdots + \log n\,.$$

Since $\log x$ is an increasing function, it can be approximated as follows:

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8

$$\int_{k-1}^{k} \log x \mathrm{d}x < \log k \cdot 1 < \int_{k}^{k+1} \log x \mathrm{d}x \,.$$

Hence summing up

$$\sum_{k=1}^{n} \int_{k-1}^{k} \log x \mathrm{d}x < \sum_{k=1}^{n} \log k < \sum_{k=1}^{n} \int_{k}^{k+1} \log x \mathrm{d}x,$$

we have

$$n \log n - n < \log n! < (n+1) \log(n+1) - n \,.$$

This inequality suggests to use $\log n!$ to approximate

$$(n + \frac{1}{2}) \log n - n \,.$$

We can namely think that $(n + \frac{1}{2}) \log n$ represents a sort of average. If we put

$$d_n = \log n! - \left(n + \frac{1}{2}\right) \log n - n = \log\left(\frac{n!}{n^{n+\frac{1}{2}} e^{-n}}\right),$$

we have

$$d_n - d_{n+1} = \left(n + \frac{1}{2}\right) \log\left(\frac{n+1}{n}\right) - 1, \tag{F.1}$$

however

$$\frac{n+1}{n} = \frac{1 + \dfrac{1}{2n+1}}{1 - \dfrac{1}{2n+1}}$$

and

$$\log(x+1) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n} \,. \tag{F.2}$$

Since

$$\log\left(\frac{n+1}{n}\right) = \log\left(\frac{1 + \dfrac{1}{2n+1}}{1 - \dfrac{1}{2n+1}}\right) = \log\left(1 + \frac{1}{2n+1}\right) - \log\left(1 - \frac{1}{2n+1}\right),$$

using (F.2) with $x = \pm\dfrac{1}{2n+1}$ we obtain

$$
\begin{aligned}
d_n - d_{n+1} &= \frac{1}{2}(2n+1)\left[\log\left(1+\frac{1}{2n+1}\right) - \log\left(1-\frac{1}{2n+1}\right)\right] - 1 \\
&= \frac{1}{2}(2n+1)\left[\frac{2}{(2n+1)} + \frac{2}{3(2n+1)^3} + \frac{2}{5(2n+1)^5} + \cdots\right] - 1 \\
&= \frac{1}{3(2n+1)^2} + \frac{1}{5(2n+1)^4} + \cdots ,
\end{aligned}
$$

from which it follows

$$
d_n - d_{n+1} > 0 .
$$

Hence d_n is decreasing. It follows that the limit of d_n exists (finite or infinite). To prove that the limit is finite, we note that

$$
\begin{aligned}
0 < d_n - d_{n+1} < \frac{1}{3}\sum_{k=1}^{\infty}\left(\frac{1}{2n+1}\right)^{2k} &= \frac{1}{3}\left[\frac{1}{1-\dfrac{1}{(2n+1)^2}} - 1\right] \\
= \frac{1}{3}\frac{1}{(2n+1)^2-1} &= \frac{1}{12n} - \frac{1}{12(n+1)},
\end{aligned}
$$

i.e. the sequence

$$
a_n = d_n - \frac{1}{12n}
$$

is increasing. Since

$$
a_n \le d_n \qquad \forall n \in \mathbb{N},\ n \neq 0,
$$

and it holds that

$$
\lim_{n\to\infty} a_n = \lim_{n\to\infty}\left(d_n - \frac{1}{12n}\right) = \lim_{n\to\infty} d_n ,
$$

we obtain that the limit of d_n exists finite since the two sequences a_n e d_n are bounded by each other.

F.2 Proof by Using the Gamma Function

Consider the Gamma function given by

$$
\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha} e^{-x} dx ,
$$

where $\alpha > 0$. It represents a generalization of factorial n, since for all $\alpha > 0$ it holds that

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha).$$

This can be easily verified by integration by parts. If α is a natural number, by iteration we obtain

$$\Gamma(n+1) = n!.$$

To prove Stirling's formula we show the more general result that

$$\Gamma(\alpha+1) = \sqrt{2\pi}\,\alpha^{\alpha+\frac{1}{2}}\,e^{-\alpha}(1+O(\alpha^{-1}))\,.$$

We consider logarithm of $\phi(x) = \log(x^{\alpha}e^{-x}) = \alpha\log x - x$. We compute the Taylor expansion of $\phi(x)$ at the maximum point α:

$$\phi(x) = \alpha\log\alpha - \alpha - \frac{1}{2\alpha}(x-\alpha)^2 + \sum_{k=3}^{n}\frac{(-1)^{k-1}}{k}\frac{(x-\alpha)^k}{\alpha^{k-1}} + \alpha\frac{(-1)^n}{n+1}\frac{(x-\alpha)^{n+1}}{\xi^{n+1}},$$

where $\xi\in[\alpha,x]$. In the integral we perform the change of variable

$$u = \frac{x-\alpha}{\sqrt{\alpha}}, \qquad dx = \sqrt{\alpha}\mathrm{d}u\,.$$

We obtain

$$\Gamma(\alpha+1) = \alpha^{\alpha+\frac{1}{2}}\,e^{-\alpha}\int_{-\sqrt{\alpha}}^{+\infty} e^{-\frac{u^2}{2}+\psi(u)}\mathrm{d}u,$$

where

$$\psi(u) = \sum_{k=3}^{n}\frac{(-1)^{k-1}}{k}\frac{u^k}{\alpha^{\frac{k}{2}-1}} + \alpha^{\frac{n+3}{2}}\frac{(-1)^n}{n+1}\frac{u^{n+1}}{(\alpha+\xi\sqrt{\alpha})^{n+1}}$$

with $\xi\in[0,u]$. We divide the integral in three parts:

$$I_1 = [-\sqrt{\alpha}, -\alpha^{\delta}], \quad I_2 = [-\alpha^{\delta}, \alpha^{\delta}], \quad I_3 = [\alpha^{\delta}, +\infty],$$

where $\delta > 0$ s sufficiently small constant. For what concerns I_1, I_3 we note that $\phi(u)$ is a concave function. Hence we obtain that also the function

$$\theta(u) = -\frac{u^2}{2} + \psi(u),$$

obtained by ϕ by adding a constant and by a linear transformation of the underlying variable, is concave. For $u \le -\alpha^{\delta}$ we have

$$\theta(u) \le -\frac{u}{\alpha^{\delta}}\theta(-\alpha^{\delta})$$

and for $u \geq \alpha^\delta$

$$\theta(u) \leq \frac{u}{\alpha^\delta}\theta(\alpha^\delta).$$

By the expansion of $\psi(u)$ with $n = 2$ we note that for α sufficiently big and $\delta < \frac{1}{6}$ we have $\theta(-\alpha^\delta) < -\frac{\alpha^{2\delta}}{4}$, $\theta(\alpha^\delta) < -\frac{\alpha^{2\delta}}{4}$ Hence for $|u| \geq \alpha^\delta$ it holds that

$$\theta(u) \leq -|u| \ \frac{\alpha^\delta}{4} \,.$$

It follows that

$$\begin{aligned}\int_{I_1} e^{\theta(u)} du + \int_{I_3} e^{\theta(u)} du &\leq \int_{|u|\geq\alpha^\delta} e^{-|u|\frac{\alpha^\delta}{4}} du \\ &= \left[-\frac{8}{\alpha^\delta} e^{-|u|\frac{\alpha^\delta}{4}}\right]_{\alpha^\delta}^{+\infty} = \frac{8}{\alpha^\delta} e^{-\frac{\alpha^{2\delta}}{4}} \,.\end{aligned}$$

We now consider I_2. If we choose $n = 3$ we obtain

$$e^{\psi(u)} = \exp\left(\frac{1}{3}\frac{u^3}{\alpha^{\frac{1}{2}}} - \frac{1}{4}\frac{\alpha^3 u^4}{(\alpha + \xi\sqrt{\alpha})^4}\right) = 1 + \frac{1}{3}\frac{u^3}{\alpha^{\frac{1}{2}}} + O(\frac{u^4}{\alpha})$$

with $\xi \in [0, u] \subset I_2$ and for $|u| < \alpha^\delta$. It follows that

$$\begin{aligned}\int_{I_2} e^{-\frac{u^2}{2}+\psi(u)} du &= \int e^{-\frac{u^2}{2}} du - \int_{I_2^c} e^{-\frac{u^2}{2}} du + \frac{\alpha^{-\frac{1}{2}}}{3} + \int_{I_2} u^3 e^{-\frac{u^2}{2}} du + O(\alpha^{-1}) \\ &= \sqrt{2\pi} + O(\alpha^{-1})\end{aligned}$$

and hence

$$\Gamma(\alpha + 1) = \sqrt{2\pi}\ \alpha^{\alpha+\frac{1}{2}}\ e^{-\alpha}(1 + O(\alpha^{-1}))\,.$$

Appendix G
Elements of Analysis

In this appendix we recall some definitions and results of analysis in one variable to facilitate the theoretical comprehension and the execution of the exercises.

G.1 Limit of a Sequence

Let $(a_n)_{n \in \mathbb{N}}$ be a sequence of real numbers. This is called

1. *convergent* if
$$\lim_{n \to \infty} a_n = L < \infty,$$
i.e. if for all $\epsilon > 0$ there exists $N = N(\epsilon)$ such that for all $n > N$
$$|a_n - L| < \epsilon;$$
2. *divergent* if
$$\lim_{n \to \infty} a_n = +\infty,$$
i.e. if for all $M > 0$ there exists $N = N(M)$ such that for all $n > N$
$$a_n > M,$$
or $\lim_{n \to \infty} a_n = -\infty$ respectively, i.e. if for all $M > 0$ there exists $N = N(M)$ such that for all $n > N$
$$a_n < M.$$

A sequence may neither be convergent nor divergent. For example, the sequence $a_n = (-1)^n$ oscillates between 1 and -1.

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8

G.2 Limit of Functions

A function $f : \mathbb{R} \longrightarrow \mathbb{R}$ has:

1. *finite limit* in x if

$$\lim_{y \to x} f(y) = L < \infty,$$

   i.e. if for all $\epsilon > 0$ there exists $\delta = \delta(\epsilon)$ such that for all y with $|y - x| < \delta$

$$|f(y) - L| < \epsilon;$$

2. *infinite limit* in x if

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

   i.e. if for all $M > 0$ there exists $\delta = \delta(M)$ such that for all y with $|y - x| < \delta$

$$f(y) > M,$$

   or $\lim_{y \to x} f(y) = -\infty$ meaning that, for all $M > 0$ there exists $\delta = \delta(M)$ such that for all y with $|y - x| < \delta$

$$f(y) < M.$$

G.3 Limits of Special Interest

We recall the following limits of special interest:

1.

$$\lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n = e;$$

2. $\forall x \in \mathbb{R}$

$$\lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n = e^x;$$

3.

$$\lim_{x \to 0} \frac{\log(1 + x)}{x} = 1.$$

G.4 Series

We recall the following series:

1. the *geometric* series

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$$

   for all $|x| < 1$;
2. the series

$$\sum_{n=1}^{\infty} nx^{n-1} = \frac{1}{(1-x)^2}$$

   for all $|x| < 1$, which is obtained as derivative of the geometric series;
3. the *exponential* series

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x$$

   for all $x \in \mathbb{R}$.

G.5 Continuity

A function is said to be *continuous* in the point x_0 if

$$\lim_{x \to x_0^-} f(x) = \lim_{x \to x_0^+} f(x) = f(x_0)\,,$$

where $\lim_{x \to x_0^-} f(x)$, $\lim_{x \to x_0^+} f(x)$ are called *left limit* and *right limit* respectively. The left limit is taken over $x < x_0$, the right limit is taken over $x > x_0$.

G.6 Table of the Principal Rules of Derivation

We summarize the most common derivatives as well as the principal rules of derivation in Table G.1 and in Table G.2, respectively.

Table G.1 Derivatives

Function $f(x)$	Derivative $f'(x)$
x^n	$n\,x^{n-1}$
e^x	e^x
$\log\,x$	$\frac{1}{x}$
$\sin\,x$	$\cos\,x$
$\cos\,x$	$-\sin\,x$
$e^{-\frac{x^2}{2}}$	$-x\,e^{-\frac{x^2}{2}}$

Table G.2 Rules of derivation

$\frac{\mathrm{d}}{\mathrm{d}x}\,[f(x)+g(x)]$	$f'(x)+g'(x)$
$\frac{\mathrm{d}}{\mathrm{d}x}\,[f(x)\,g(x)]$	$f'(x)\,g(x)+f(x)\,g'(x)$
$\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{f(x)}{g(x)}\right]$	$\frac{f'(x)\,g(x)-f(x)\,g'(x)}{g^2(x)}$
$\frac{\mathrm{d}}{\mathrm{d}x}\,[f(g(x))]$	$f'(g(x))\cdot g'(x)$

G.7 Integrals

1. **Integration by parts formula**

$$\int_a^b f(x)\,g'(x)\,\mathrm{d}x \;=\; [f(x)\,g(x)]_a^b - \int_a^b f'(x)\,g(x)\,\mathrm{d}x\,.$$

2. **Change of variable**

$$x = g(y) \Rightarrow dx = g'(y)\,\mathrm{d}y,$$

$$\int_a^b f(x)\,\mathrm{d}x \;=\; \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(y))\,g'(y)\,\mathrm{d}y\,.$$

Appendix H
Bidimensional Integrals

In this appendix we recall some notions of analysis in several variables to facilitate the comprehension of the text and the execution of the exercises.

H.1 Areas of Bidimensional Regions

Let A be a region of the plane (Fig. H.1). The area of A is given by

$$\text{area } A = \int\int_A \mathrm{d}x\mathrm{d}y .$$

This is analogous to the one-dimensional case, where the length of a segment $[a, b]$ is given by

$$l([a, b]) = \int_a^b \mathrm{d}x .$$

H.2 Integrals of Functions of Two Variables

Let $f : \mathbb{R}^2 \to \mathbb{R}$ and put $z = f(x, y)$. A function in two variables describes a surface in $\mathbb{R}^3$ of coordinates $(x,\ y,\ f(x, y))$. We want to calculate the volume between the surface described by the function and the plane xy. This volume is given by the double integral

$$\int\int_{\mathbb{R}^2} f(x, y)\,\mathrm{d}x\mathrm{d}y ,$$

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8

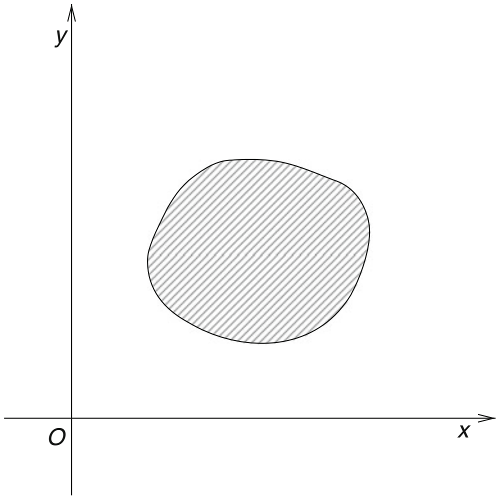


Fig. H.1 A region of the plane

if f is sufficiently regular (for example, if f is continuous). We can compute a double integral as two nested one-dimensional integrals, i.e.

$$\int \mathrm{d}x \int f(x, y)\,\mathrm{d}y = \int \mathrm{d}y \int f(x, y)\,\mathrm{d}x = \int\int f(x, y)\,\mathrm{d}x\mathrm{d}y\,.$$

This result holds for sufficiently regular functions f (for example, if f is continuous) and is known as *Fubini-Tonelli theorem*. We refer to [11] for further details.

Example H.2.1 Let $A = \{1 < x < 2,\ 3 < y < 4\}$. We compute the following integral on A.

$$\begin{aligned}\int\int_A x^2 y\,\mathrm{d}x\mathrm{d}y &= \int_1^2 \mathrm{d}x \underbrace{\int_3^4 x^2 y\,\mathrm{d}y}_{x \text{ is a parameter!}} \\ &= \int_1^2 x^2 \left(\int_3^4 y\,\mathrm{d}y\right)\mathrm{d}x \\ &= \int_1^2 x^2 \left[\frac{1}{2}y^2\right]_3^4 \mathrm{d}x \\ &= \frac{7}{2}\int_1^2 x^2\,\mathrm{d}x \\ &= \frac{49}{6}\,.\end{aligned}$$

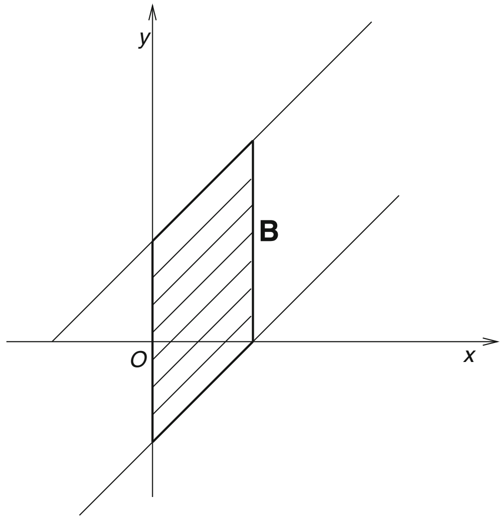


Fig. H.2 The region B

Example H.2.2 Let $B = \{0 < x < 1,\ x - 1 < y < x + 1\}$, see Fig. H.2. We calculate the following integral on B.

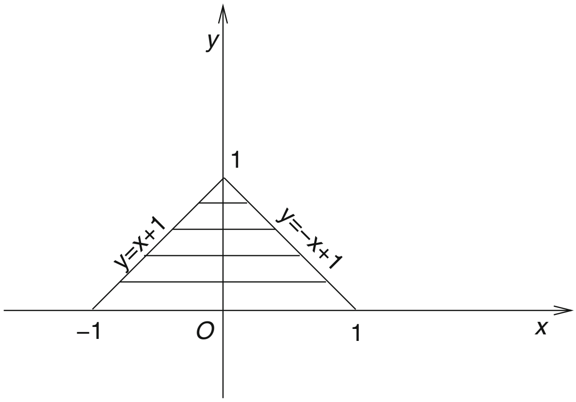
$$\begin{aligned}
\int\!\!\int_B e^{-y}\,\mathrm{d}x\mathrm{d}y &= \int_0^1 \mathrm{d}x \int_{x-1}^{x+1} e^{-y}\,\mathrm{d}y \\
&= \int_0^1 \left[e^{-y}\right]_{x-1}^{x+1} \mathrm{d}x \\
&= \int_0^1 \left(e^{-(x+1)} - e^{-(x-1)}\right) \mathrm{d}x \\
&= \int_0^1 \left(e^{-1} - e\right) e^{-x}\mathrm{d}x \\
&= \left(e^{-1} - e\right)\left(1 - e^{-1}\right) .
\end{aligned}$$

Example H.2.3 To perform a double integral, it is convenient to divide the domain of integration in a suitable way. Consider the example of Fig. H.3, where $D = \{0 < y < 1,\ y - 1 < x < -y + 1\}$.

We compute the integral of a function $f(x, y)$, which is assumed to be sufficiently regular, on D:

$$\begin{aligned}
\int\!\!\int_D f(x, y)\,\mathrm{d}x\mathrm{d}y &= \int_0^1 \mathrm{d}y \int_{y-1}^{-y+1} f(x, y)\,\mathrm{d}x \\
&= \int_0^1 \mathrm{d}x \int_0^{-x+1} f(x, y)\,\mathrm{d}y + \int_{-1}^0 \mathrm{d}x \int_0^{x+1} f(x, y)\,\mathrm{d}y.
\end{aligned}$$

Fig. H.3 Region D



In the first passage the extremes of integration can be found by drawing the parallels to the x-axis and finding the intersections with the border of the domain D. In the second step the integral has been split in two parts and the extremes have been found by drawing the parallels to the y-axis.

H.3 Partial Derivatives with Respect to a Single Variable

Let $f : \mathbb{R}^2 \to \mathbb{R}$, $z = f(x, y)$. We call *partial derivative* of f with respect to the variable x and write $\frac{\partial f}{\partial x}$ the derivative of f obtained by considering the function as depending only by the variable x and considering the other variables as parameters. Analogously we can define the partial derivatives of a function with respect to the other variables.

Example H.3.1 (Partial derivatives)

1. $f(x, y) = x^2 y$:

$$\frac{\partial f}{\partial x} = 2xy\,, \quad \frac{\partial f}{\partial y} = x^2\,;$$

2. $f(x, y) = \log(xy)$:

$$\frac{\partial f}{\partial x} = \frac{1}{x}\,, \quad \frac{\partial f}{\partial y} = \frac{1}{y}\,.$$

H.4 Change of Variables

Let $\Psi : \mathbb{R}^2 \to \mathbb{R}^2$, $(x, y) = (\Psi_1(x, y), \Psi_2(x, y))$. We call *Jacobian* J_ψ of the function ψ the matrix

$$J_\Psi = \begin{pmatrix} \frac{\partial \Psi_1}{\partial x} & \frac{\partial \Psi_1}{\partial y} \\ \\ \frac{\partial \Psi_2}{\partial x} & \frac{\partial \Psi_2}{\partial y} \end{pmatrix}.$$

A change of coordinates in $\mathbb{R}^2$ is given by a function

$$\Psi : \mathbb{R}^2 \to \mathbb{R}^2$$

$$(u, v) \longmapsto (x, y)$$

with particular regularity properties (diffeomorphismus). To change the variables in an integral, we use then the following rule:

$$\int\int_A f(x, y)\, \mathrm{d}x\mathrm{d}y = \int\int_{\Psi^{-1}(A)} f(\Psi(u, v))\ |\det J_\Psi|\ \mathrm{d}x\mathrm{d}y$$

with the help of the following diagram:

$$\begin{tikzcd} \mathbb{R}^2_{(u,v)} \arrow[r, "\Psi"] \arrow[dr, "f\circ\Psi"'] & \mathbb{R}^2_{(x,y)} \arrow[d, "f"] \\ & \mathbb{R} \end{tikzcd}$$

Example H.4.1 In this example we consider the computation of the normalization constant for the standard normal distribution in 2 dimension. To this purpose we need

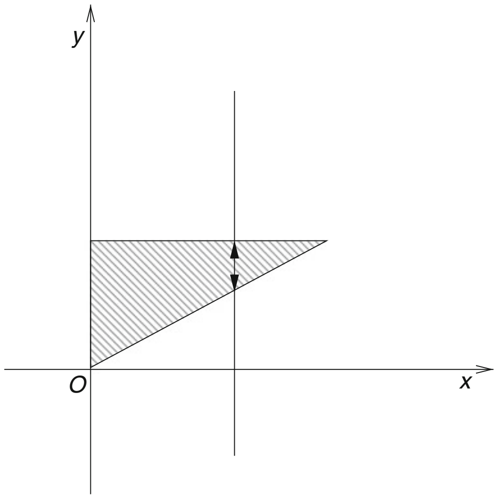


Fig. H.4 Extremes of integration as x varies

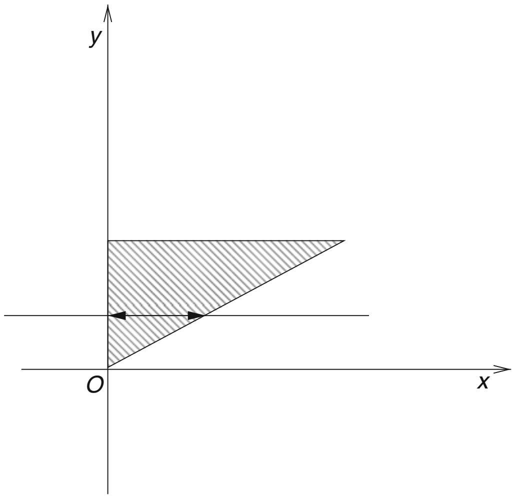


Fig. H.5 Extremes of integration as y varies

to use a change of variable (Figs. H.4 and H.5). Consider

$$\int\int_{\mathbb{R}^2} e^{-\frac{1}{2}(x^2+y^2)}\, \mathrm{d}x\mathrm{d}y\,.$$

To compute this integral, we use the *polar coordinates*:

$$x \;=\; \rho\,\cos\theta, \qquad y \;=\; \rho\,\sin\theta$$

$$(\theta,\,\rho) \overset{\Psi}{\mapsto} (x,\,y) = (\rho\,\cos\theta,\;\rho\,\sin\theta)\,.$$

The Jacobian of this transformation is given by

$$J_\Psi \;=\; \begin{pmatrix} \frac{\partial}{\partial\theta}\,\rho\,\cos\theta & \frac{\partial}{\partial\rho}\,\rho\,\cos\theta \\ \frac{\partial}{\partial\theta}\,\rho\,\sin\theta & \frac{\partial}{\partial\rho}\,\rho\,\sin\theta \end{pmatrix} \;=\; \begin{pmatrix} -\rho\,\sin\theta & \cos\theta \\ \rho\,\cos\theta & \sin\theta \end{pmatrix}.$$

The Jacobian determinant is then

$$\det J_\Psi \;=\; -\rho\,(\sin^2\theta + \cos^2\theta) \;=\; -\rho,$$

i.e.

$$|\det J_\Psi| \;=\; \rho\,.$$

It follows that

$$\begin{aligned}
\int\int_{\mathbb{R}^2} e^{-\frac{1}{2}(x^2+y^2)}\,\mathrm{d}x\mathrm{d}y &= \int_0^{+\infty} \mathrm{d}\rho \int_0^{2\pi} \rho\, e^{-\frac{1}{2}\rho^2}\,\mathrm{d}\theta \\
&= \int_0^{+\infty} \rho\, e^{-\frac{1}{2}\rho^2}\,\mathrm{d}\rho \int_0^{2\pi} \mathrm{d}\theta \\
&= 2\pi \underbrace{\left[-e^{-\frac{1}{2}\rho^2}\right]_0^{+\infty}}_{1} \\
&= 2\pi\,.
\end{aligned}$$

Hence

$$\begin{aligned}
\left(\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}\,\mathrm{d}x\right)^2 &= \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}\,\mathrm{d}x \int_{-\infty}^{+\infty} e^{-\frac{Y^2}{2}}\,\mathrm{d}y \\
&= \int\int_{\mathbb{R}^2} e^{-\frac{1}{2}(x^2+y^2)}\,\mathrm{d}x\mathrm{d}y \;=\; 2\pi\,.
\end{aligned}$$

Finally

$$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}\,\mathrm{d}x \;=\sqrt{2\pi}\,.$$

References

1. de Finetti, B.: Theory of Probability. A Critical Introduction Treatment, vol. 1, 2. Wiley, New York (1974, 1975)
2. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. 1. Wiley, New York (1957)
3. Foatà, D., Fuchs, A.: Calcul des probabilités, 2nd edn. Dunod, Paris (1998)
4. Gnedenko, B.: The Theory of Probability. Gordon and Breach Science Publishers, Amsterdam (1997)
5. Hogg, R.V., Tanis, E.A.: Probability and Statistical Inference. Prentice Hall, New York (2001)
6. Jacod, J., Protter, P.: Probability Essentials. Springer, Berlin (2003)
7. Kleinrock, L.: Queueing Systems 1. Wiley, New York (1975)
8. Kleinrock, L., Gail, R.: Queueing Systems: Problems and Solutions. Wiley, New York (1996)
9. Lee, P.M.: Bayesian Statistics: An Introduction. Edward Arnold, London (1994)
10. Lindley, D.V.: Introduction to Probability and Statistics, vol. 1, 2. Cambridge University Press, New York (1965)
11. Munkres, J.R.: Analysis On Manifolds. Advanced Books Classics. Westview Press, Boulder (1997)
12. Ross, S.M.: Introduction to Probability Models. Elsevier, New York (2010)

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8

Index

F. Biagini and M. Campanino, *Elements of Probability and Statistics*,
UNITEXT - La Matematica per il 3+2 98, DOI 10.1007/978-3-319-07254-8