

# 1

# Modelling Risks

A risk can be described as an event that may or may not take place, and that brings about some adverse financial consequences. It is thus natural that the modelling of risks uses probability theory. The basics of probability theory are briefly reviewed in this first chapter, with special emphasis on multivariate tools, such as random vectors and related quantities. The material introduced here will be extensively used throughout the book.

## 1.1 INTRODUCTION

Much of our life is based on the belief that the future is largely unpredictable. We express this belief by the use of words such as ‘random’ or ‘probability’ and we aim to assign quantitative meanings to such usage. The branch of mathematics dealing with uncertainty and randomness is called probability theory. Together with statistics, it forms the basis of actuarial science.

In a broad sense, insurance refers to the business of transferring (totally or partially) the economic impact of unforeseen mishaps. The central notion in actuarial mathematics is the notion of risk. A risk can be described as an event that may or may not take place, and that brings about some adverse financial consequences. It is thus natural that the modelling of risks uses probability theory, with the concepts of random events and random variables playing a central role.

This first chapter aims to lay the mathematical foundations for the modelling of insurance risks. We begin by describing the classical axiomatic construction of probability theory. Probability spaces are carefully defined. Subsequent sections deal with random variables, distribution functions, quantile functions, mathematical expectations, etc. Emphasis is put on mutual independence and random vectors.

We also list several transforms, such as the hazard rate, the mean-excess function, the Laplace transform, the moment generating function as well as the probability generating function. These transforms will be used in the next chapters to characterize partial order relations defined on sets of distribution functions.

The final sections of this chapter are devoted to very particular dependence structures, extreme in some sense to be specified later on: comonotonicity and mutual exclusivity. The former corresponds to perfect positive dependence: all the random variables can be written as

non-decreasing transformations of the same underlying random variable. They thus ‘move in the same direction’, are ‘common monotonic’ – hence the name. On the other hand, mutual exclusivity can be seen as a very strong negative dependence concept. In this case, just a single random variable can be positive (and the others then have to be equal to zero). These two structures will be widely used in later chapters.

An excellent introduction to probability theory can be found in Chow and Teicher (2003). The book by Barlow and Proschan (1975) contains many results involving reliability concepts. A detailed account of comonotonicity can be found in Dhaene *et al.* (2002a,b).

## **1.2 THE PROBABILISTIC DESCRIPTION OF RISKS**

### **1.2.1 Probability space**

In probability theory the starting point is a ‘probability space’. The usual phrase at the beginning of a stochastic model is (or should be): ‘Let  $(\Omega, \mathcal{F}, \text{Pr})$  be a probability space . . .’. Such a general approach to probability plays a fundamental role in the theory, and it is not our intention to recall all definitions and axioms, which can easily be found in any textbook on probability theory. We shall confine ourselves to concepts and results used in this book. The three ingredients of a probability space are a universe  $\Omega$ , a sigma-algebra  $\mathcal{F}$  and a probability measure  $\text{Pr}$ . We briefly review each of these notions in this section.

### **1.2.2 Experiment and universe**

Many everyday statements for actuaries take the form ‘the probability of  $A$  is  $p$ ’, where  $A$  is some event (such as ‘the total losses exceed the threshold € 1 million’ or ‘the number of claims reported by a given policyholder is less than 2’) and  $p$  is a real number between 0 and 1. The occurrence or non-occurrence of  $A$  depends upon the chain of circumstances under consideration. Such a particular chain is called an experiment in probability; the result of an experiment is called its outcome and the set of all outcomes (called the universe) is denoted by  $\Omega$ .

The word ‘experiment’ is used here in a very general sense to describe virtually any process of which all possible outcomes can be specified in advance and of which the actual outcome will be one of those specified. The basic feature of an experiment is that its outcome is not definitely known by the actuary beforehand.

### **1.2.3 Random events**

Random events are subsets of the universe  $\Omega$  associated with a given experiment. A random event is the mathematical formalization of an event described in words. It is random since we cannot predict with certainty whether it will be realized or not during the experiment. For instance, if we are interested in the number of claims made by a policyholder belonging to an automobile portfolio in one year, the experiment consists in observing the driving behaviour of this individual during an annual period, and the universe  $\Omega$  is simply the set

$\mathbb{N} \equiv \{0, 1, 2, \dots\}$  of the non-negative integers. The random event  $A =$  ‘the policyholder makes at most one claim’ is identified with the pair  $\{0, 1\} \subset \Omega$ .

As usual, we use  $A \cup B$  and  $A \cap B$  to represent the union and the intersection of any two subsets  $A$  and  $B$  of  $\Omega$ , respectively. The union of sets is defined to be the set that contains the points that belong to at least one of the sets. The intersection of sets is defined to be the set that contains the points that are common to all the sets. These set operations correspond to the words ‘or’ and ‘and’ between sentences:  $A \cup B$  is the event realized if  $A$  or  $B$  is realized and  $A \cap B$  is the event realized if  $A$  and  $B$  are simultaneously realized during the experiment. We also define the difference between sets  $A$  and  $B$ , denoted as  $A \setminus B$ , as the set of elements in  $A$  but not in  $B$ . Finally,  $\overline{A}$  is the complement of the event  $A$ , defined as  $\Omega \setminus A$ ; it is the set of points of  $\Omega$  that do not belong to  $A$ . This corresponds to negation:  $\overline{A}$  is realized if  $A$  is not realized during the experiment. In particular,  $\overline{\Omega} = \emptyset$  where  $\emptyset$  is the empty set.

### 1.2.4 Sigma-algebra

For technical reasons, it is useful to consider a certain family  $\mathcal{F}$  of random events, that is, of subsets of  $\Omega$ . In practice,  $\mathcal{F}$  can be chosen so that this limitation is not restrictive in the sense that virtually every subset of interest is sufficiently regular to belong to  $\mathcal{F}$ . The family  $\mathcal{F}$  has to be closed under standard operations on sets; indeed, given two events  $A$  and  $B$  in  $\mathcal{F}$ , we want that  $A \cup B$ ,  $A \cap B$  and  $\overline{A}$  are still events (i.e., still belong to  $\mathcal{F}$ ). Technically speaking, this will be the case if  $\mathcal{F}$  is a sigma-algebra, as defined below.

**Definition 1.2.1.** A family  $\mathcal{F}$  of subsets of the universe  $\Omega$  is called a sigma-algebra if it fulfils the three following properties:

- P1**  $\Omega \in \mathcal{F}$ ;
- P2**  $A \in \mathcal{F} \Rightarrow \overline{A} \in \mathcal{F}$ ;
- P3**  $A_1, A_2, A_3, \dots \in \mathcal{F} \Rightarrow \cup_{i \geq 1} A_i \in \mathcal{F}$ .

∇

The properties P1–P3 are quite natural. Indeed, P1 means that  $\Omega$  itself is an event (it is the event which is always realized). P2 means that if  $A$  is an event, the complement of  $A$  is also an event. P3 means that the event consisting of the realization of at least one of the  $A_i$  is also an event.

### 1.2.5 Probability measure

Once the universe  $\Omega$  has been equipped with a sigma-algebra  $\mathcal{F}$  of random events, a probability measure  $\text{Pr}$  can be defined on  $\mathcal{F}$ . The knowledge of  $\text{Pr}$  allows us to discuss the likelihoods of the occurrence of events in  $\mathcal{F}$ . To be specific,  $\text{Pr}$  assigns to each random event  $A$  its probability  $\text{Pr}[A]$ ;  $\text{Pr}[A]$  is the likelihood of realization of  $A$ . The probability of  $A$  is a numerical measure of the likelihood that the actual outcome of the experiment will be an element of  $A$ .

**Definition 1.2.2.** A probability measure  $\Pr$  maps  $\mathcal{F}$  to  $[0, 1]$ , with  $\Pr[\Omega] = 1$ , and is such that given  $A_1, A_2, A_3, \dots \in \mathcal{F}$  which are pairwise disjoint, that is, such that  $A_i \cap A_j = \emptyset$  if  $i \neq j$ ,

$$\Pr \left[ \bigcup_{i \geq 1} A_i \right] = \sum_{i \geq 1} \Pr[A_i];$$

this technical property is usually referred to as the sigma-additivity of  $\Pr$ .  $\nabla$

The properties assigned to  $\Pr$  in Definition 1.2.2 naturally follow from empirical evidence: if we were allowed to repeat an experiment a large number of times, keeping the initial conditions as equal as possible, the proportion of times that an event  $A$  occurs would behave according to Definition 1.2.2. Note that  $\Pr[A]$  is then the mathematical idealization of the proportion of times  $A$  occurs.

We can associate a probability space  $(\Omega, \mathcal{F}, \Pr)$  with any experiment, and all the questions associated with the experiment can be reformulated in terms of this space. It may seem reasonable to ask for the numerical value of the probability  $\Pr[A]$  of some event  $A$ . This value is deduced from empirical observations (claims statistics recorded by the insurance company in the past, for instance) and is often derived from a parametric model.

## 1.3 INDEPENDENCE FOR EVENTS AND CONDITIONAL PROBABILITIES

### 1.3.1 Independent events

Independence is a crucial concept in probability theory. It aims to formalize the intuitive notion of ‘not influencing each other’ for random events: we would like to give a precise meaning to the fact that the realization of an event does not decrease nor increase the probability that the other event occurs. The following definition offers a mathematically unambiguous meaning of mutual independence for a pair of random events. Nevertheless, we will have to wait until formula (1.3) to get an intuitive meaning for this concept.

**Definition 1.3.1.** Two events  $A$  and  $B$  are said to be independent if the probability of their intersection factors into the product of their respective probabilities, that is, if  $\Pr[A \cap B] = \Pr[A]\Pr[B]$ .  $\nabla$

Definition 1.3.1 is extended to more than two events as follows.

**Definition 1.3.2.** The events in a family  $\mathcal{A}$  of events are independent if for every finite sequence  $A_1, A_2, \dots, A_k$  of events in  $\mathcal{A}$ ,

$$\Pr \left[ \bigcap_{i=1}^k A_i \right] = \prod_{i=1}^k \Pr[A_i]. \quad (1.1)$$

$\nabla$

The concept of independence is very important in assigning probabilities to events. For instance, if two or more events are regarded as being physically independent, in the sense that the occurrence or non-occurrence of some of them has no influence on the occurrence or non-occurrence of the others, then this condition is translated into mathematical terms through the assignment of probabilities satisfying (1.1).

### 1.3.2 Conditional probability

Independence is the exception rather than the rule. In any given experiment, it is often necessary to consider the probability of an event  $A$  when additional information about the outcome of the experiment has been obtained from the occurrence of some other event  $B$ . This corresponds to intuitive statements of the form ‘if  $B$  occurs then the probability of  $A$  is  $p$ ’, where  $B$  might be ‘March is rainy’ and  $A$  ‘the claim frequency in motor insurance increases by 5%’. This is called the conditional probability of  $A$  given  $B$ .

**Definition 1.3.3.** If  $\Pr[B] > 0$  then the conditional probability  $\Pr[A|B]$  of  $A$  given  $B$  is defined to be

$$\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]}. \quad (1.2)$$

▽

The definition of conditional probabilities through (1.2) is in line with empirical evidence. Repeating a given experiment a large number of times,  $\Pr[A|B]$  is the mathematical idealization of the proportion of times  $A$  occurs in those experiments where  $B$  occurred, hence the ratio (1.2).

Let us now justify the definition of conditional probabilities by means of the ratio (1.2). As mentioned earlier, conditional probabilities correspond to situations where additional information is available; this information is reflected by the fact that an event  $B$  is realized, and implies that only events compatible with  $B$  have a positive probability (hence the numerator  $\Pr[A \cap B]$  in (1.2)) and that  $B$  is given probability one (being the new universe, hence the denominator  $\Pr[B]$  in (1.2)).

With Definition 1.3.3, it is easy to see from Definition 1.3.1 that  $A$  and  $B$  are independent if, and only if,

$$\Pr[A|B] = \Pr[A|\bar{B}] = \Pr[A]. \quad (1.3)$$

Note that this interpretation of independence is much more intuitive than Definition 1.3.1: indeed, the identity expresses the natural idea that the realization or not of  $B$  does not increase nor decrease the probability that  $A$  occurs.

## 1.4 RANDOM VARIABLES AND RANDOM VECTORS

### 1.4.1 Random variables

Actuaries are often not interested in an experiment itself but rather in some consequences of its random outcome. For instance, they are more concerned with the amounts the insurance company will have to pay than with the particular circumstances which give rise to the

claims. Such consequences, when real-valued, may be thought of as functions mapping  $\Omega$  into the real line  $\mathbb{R}$ .

Such functions are called random variables provided they satisfy certain desirable properties, precisely stated in the following definition.

**Definition 1.4.1.** A random variable (rv)  $X$  is a measurable function mapping  $\Omega$  to the real numbers, that is,  $X : \Omega \rightarrow \mathbb{R}$  is such that  $X^{-1}((-\infty, x]) \in \mathcal{F}$  for any  $x \in \mathbb{R}$ , where  $X^{-1}((-\infty, x]) = \{\omega \in \Omega | X(\omega) \leq x\}$ . ∇

Henceforth, rvs are denoted by capital letters: for example  $X$ . They are mathematical formalizations of random outcomes given by numerical values. An example of an rv is the amount of a claim associated with the occurrence of an automobile accident. The rv  $X$  can be represented as in Figure 1.1:  $X$  has a specified value  $X(\omega)$  at every possible outcome  $\omega$  in the universe  $\Omega$ .

In words, the measurability condition  $X^{-1}((-\infty, x]) \in \mathcal{F}$  involved in Definition 1.4.1 ensures that the actuary can make statements such as ‘ $X$  is less than or equal to  $x$ ’ and quantify their likelihood.

Of course, some rvs assume values in subsets of  $\mathbb{R}$  rather than in the whole real line. The set of all the possible values for an rv  $X$  is called the support of  $X$  and is formally defined in Definition 1.5.10.

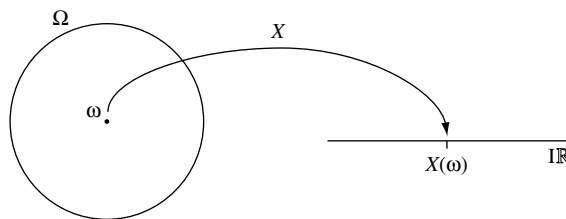
### 1.4.2 Random vectors

In this work, we will be mainly concerned with the impact of a possible dependence among risks. For this purpose, we have to consider rvs simultaneously rather than separately. Mathematically speaking, this means that random vectors are involved: the outcomes of most experiments that will be considered in this book will be  $n$ -tuples of real numbers. The  $n$ -dimensional Euclidean space of all  $n$ -tuples of real numbers will be denoted by  $\mathbb{R}^n$ , that is,  $\mathbb{R}^n$  consists of the points  $\mathbf{x} = (x_1, x_2, \dots, x_n)^t$  where  $x_i \in \mathbb{R}$ ,  $i = 1, 2, \dots, n$ . By convention, all vectors will be written in bold and will be considered as column vectors, with the superscript ‘ $t$ ’ for transposition.

Let us now formally define the concept of random vectors.

**Definition 1.4.2.** An  $n$ -dimensional random vector  $\mathbf{X}$  is a measurable function mapping the universe  $\Omega$  to  $\mathbb{R}^n$ , that is,  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$  satisfies

$$\mathbf{X}^{-1}((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_n]) \in \mathcal{F}$$



**Figure 1.1** The random variable  $X$

for any  $x_1, x_2, \dots, x_n \in \mathbb{R}$ , where

$$\begin{aligned} & \mathbf{X}^{-1}((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_n]) \\ &= \left\{ \omega \in \Omega \mid \mathbf{X}(\omega) \in (-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_n]) \right\}. \end{aligned}$$

▽

Again, the measurability condition allows the actuary to consider the event ‘each  $X_i$  is less than or equal to the threshold  $x_i$ ,  $i = 1, \dots, n$ ’.

Such a random vector  $\mathbf{X} = (X_1, X_2, \dots, X_n)^t$  is in fact a collection of  $n$  univariate rvs,  $X_1, X_2, \dots, X_n$ , say, defined on the same probability space  $(\Omega, \mathcal{F}, \Pr)$ . Random vectors are denoted by bold capital letters:  $\mathbf{X}$ , for example. A bold lower-case letter  $\mathbf{x}$  means a point  $(x_1, x_2, \dots, x_n)^t$  in  $\mathbb{R}^n$ , and  $\mathbb{R}^n$  is endowed with the usual componentwise order, that is, given  $\mathbf{x}$  and  $\mathbf{y}$  in  $\mathbb{R}^n$ ,  $\mathbf{x} \leq \mathbf{y}$  ( $\mathbf{x} < \mathbf{y}$ ) means that  $x_i \leq y_i$  ( $x_i < y_i$ ) for  $i = 1, 2, \dots, n$ . In an abuse of notation, we will often denote

$$[\mathbf{X} \leq \mathbf{x}] = \mathbf{X}^{-1}((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_n])$$

and interpret the event  $[\mathbf{X} \leq \mathbf{x}]$  componentwise, that is,

$$[\mathbf{X} \leq \mathbf{x}] = [X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n] = \bigcap_{k=1}^n [X_k \leq x_k].$$

### 1.4.3 Risks and losses

In a broad sense, insurance refers to the business of transferring (totally or partially) the economic impact of unforeseen mishaps. The central notion in actuarial mathematics is the notion of risk. A risk can be described as an event that may or may not take place (thus, a random event), and that brings about some adverse financial consequences. It always contains an element of uncertainty: either the moment of its occurrence (as in life insurance), or the occurrence itself, or the nature and severity of its consequences (as in automobile insurance).

The actuary models an insurance risk by an rv which represents the random amount of money the insurance company will have to pay out to indemnify the policyholder and/or the third party for the consequences of the occurrence of the insured peril. From the remarks above, the rvs modelling the insurance risks may generally be assumed to be non-negative. This leads to the following formal definition.

**Definition 1.4.3.** A risk  $X$  is a non-negative rv representing the random amount of money paid by an insurance company to indemnify a policyholder, a beneficiary and/or a third-party in execution of an insurance contract. ▽

In return for providing coverage, the insurer will receive premiums. The insurer will often be interested in the total cash flow associated with a policy. The loss (over a certain reference period) is defined as the (discounted value of the) payments to be made by the insurer minus the (discounted value of the) premiums to be paid by the insured.

**Definition 1.4.4.** Given a risk  $X$  covered by an insurance company in return of a premium payment  $p$  ( $p$  is the discounted value of premiums to be paid), the associated loss  $L$  is defined as  $L = X - p$ .  $\nabla$

**Remark 1.4.5.** In many actuarial textbooks, the premium  $p$  is assumed to be a known amount of money, fixed by the policy conditions. The insurance business thus consists of replacing the random consequences of the insured peril by a deterministic premium amount. For one-year policies with a single premium payment (at policy issue), the premium reduces to a fixed amount  $p$ . There are, however, many situations where the premium  $p$  itself is an rv. In life insurance, for instance,  $p$  will often be a non-trivial rv depending on the remaining lifetime of the insured. Also, in automobile insurance, the implementation of merit-rating systems (such as bonus–malus mechanisms) makes the premium paid by the policyholder contingent on the claims reported in the past.  $\nabla$

## 1.5 DISTRIBUTION FUNCTIONS

### 1.5.1 Univariate distribution functions

#### 1.5.1.1 Definition

In many cases, neither the universe  $\Omega$  nor the function  $X$  need be given explicitly. The practitioner only has to know the probability law governing  $X$  or, in other words, its distribution. This means that he is interested in the probabilities that  $X$  takes values in appropriate subsets of the real line.

To each rv  $X$  is associated a function  $F_X$  called the distribution function of  $X$ , describing the stochastic behaviour of  $X$ . Of course,  $F_X$  does not indicate the actual outcome of  $X$ , but how the possible values of  $X$  are distributed (hence its name).

**Definition 1.5.1.** The distribution function (df) of the rv  $X$ , denoted by  $F_X$ , is defined as

$$F_X(x) = \Pr[X^{-1}((-\infty, x])] \equiv \Pr[X \leq x], \quad x \in \mathbb{R}.$$

In words,  $F_X(x)$  represents the probability that the rv  $X$  assumes a value that is less than or equal to  $x$ .  $\nabla$

If  $X$  is the total monetary amount of claims generated by some policyholder,  $F_X(x)$  is the probability that this policyholder produces a total claim amount of at most  $x$ . The df  $F_X$  corresponds to an estimated physical probability distribution or a well-chosen subjective probability distribution.

**Remark 1.5.2.** Each rv  $X$  induces a probability measure  $\mathbb{P}_X$  on  $(\Omega, \mathcal{F})$ , defined for  $A \in \mathcal{F}$  as

$$\mathbb{P}_X[A] = \Pr[X^{-1}(A)] \equiv \Pr[X \in A].$$

In order to describe an rv  $X$ , one would need to know  $\mathbb{P}_X[B]$  for all possible  $B \in \mathcal{F}$ . However, it turns out that it suffices to know the value of  $\Pr[X \in B]$  for sets  $B$  of the form  $(-\infty, x]$ ,  $x \in \mathbb{R}$ . The probability distribution of an rv  $X$  is then uniquely determined by its df  $F_X$ .  $\nabla$

### 1.5.1.2 Characterization

Let us now examine the set of properties satisfied by all dfs. This allows us to characterize the set of all possible dfs.

**Property 1.5.3**

Any df  $F_X$  maps the real line  $\mathbb{R}$  to the unit interval  $[0, 1]$  and possesses the following properties:

**P1**  $F_X$  is non-decreasing.

**P2**  $F_X$  is right-continuous, that is,

$$\lim_{\Delta x \rightarrow 0^+} F_X(x + \Delta x) = F_X(x)$$

holds for any  $x \in \mathbb{R}$ ; the limit

$$F_X(x-) \equiv \lim_{\Delta x \rightarrow 0^+} F_X(x - \Delta x) = \Pr[X < x]$$

is thus well defined.

**P3**  $F_X$  satisfies  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  and  $\lim_{x \rightarrow +\infty} F_X(x) = 1$ .

P1–P3 are direct consequences of Definition 1.5.1.

**Example 1.5.4.** The knowledge of  $F_X$  provides the actuary with the complete description of the stochastic behaviour of the rv  $X$ . For instance, let us consider the graph of  $F_X$  depicted in Figure 1.2. Since  $F_X(0) = 0$ ,  $X$  cannot assume negative values. Considering  $x_1$ ,  $F_X(x_1)$  gives the probability of  $X$  being smaller than  $x_1$ . Since  $F_X$  is continuous at  $x_1$ ,

$$F_X(x_1) = F_X(x_1-) \Leftrightarrow \Pr[X \leq x_1] = \Pr[X < x_1].$$

Flat parts of the graph of  $F_X$  indicate forbidden values for  $X$ ; for instance,  $X$  cannot assume a value between  $x_2$  and  $x_3$  since

$$\Pr[x_2 < X \leq x_3] = F_X(x_3) - F_X(x_2) = 0.$$

Discontinuity jumps in  $F_X$  indicate atoms (i.e., points receiving a positive probability mass); for instance,

$$\Pr[X = x_4] = F_X(x_4) - F_X(x_4-).$$

∇

In general, we have

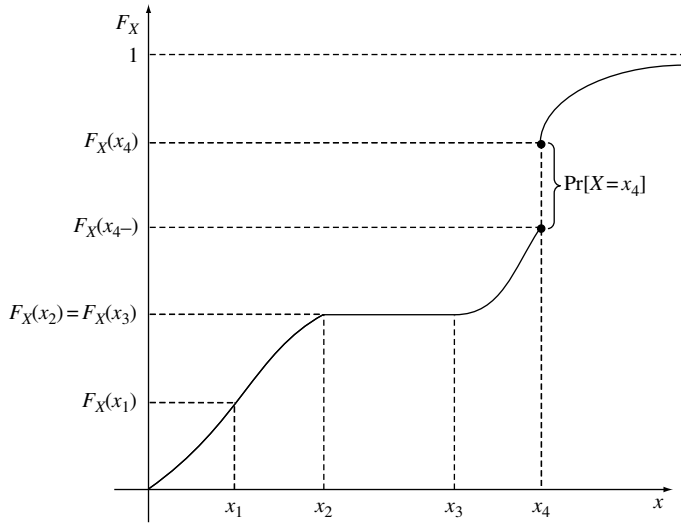
$$\Pr[a < X \leq b] = F_X(b) - F_X(a),$$

$$\Pr[a \leq X \leq b] = F_X(b) - F_X(a-),$$

$$\Pr[a < X < b] = F_X(b-) - F_X(a),$$

$$\Pr[a \leq X < b] = F_X(b-) - F_X(a-).$$

In these relations we may have  $a = -\infty$  or  $b = +\infty$ .



**Figure 1.2** Graph of a df  $F_X$

**Remark 1.5.5.** Actuaries are often more interested in the df of an rv than in the rv itself. For two rvs  $X$  and  $Y$  which are equal in distribution, that is,  $F_X \equiv F_Y$ , we will write  $X =_d Y$ .  $\nabla$

## 1.5.2 Multivariate distribution functions

### 1.5.2.1 Definition

Suppose that  $X_1, X_2, \dots, X_n$  are  $n$  rvs defined on the same probability space  $(\Omega, \mathcal{F}, \Pr)$ . Their marginal dfs  $F_1, F_2, \dots, F_n$  contain all the information about their associated probabilities. But how can the actuary encapsulate information about their properties relative to each other? As explained above, the key idea is to think of  $X_1, X_2, \dots, X_n$  as being components of a random vector  $\mathbf{X} = (X_1, X_2, \dots, X_n)'$  taking values in  $\mathbb{R}^n$  rather than being unrelated rvs each taking values in  $\mathbb{R}$ .

As was the case for rvs, each random vector  $\mathbf{X}$  possesses a df  $F_{\mathbf{X}}$  that describes its stochastic behaviour.

**Definition 1.5.6.** The df of the random vector  $\mathbf{X}$ , denoted by  $F_{\mathbf{X}}$ , is defined as

$$\begin{aligned} F_{\mathbf{X}}(x_1, x_2, \dots, x_n) &= \Pr[\mathbf{X}^{-1}((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_n])] \\ &= \Pr[X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n], \end{aligned}$$

$x_1, x_2, \dots, x_n \in \mathbb{R}$ .  $\nabla$

The value  $F_{\mathbf{X}}(x_1, x_2, \dots, x_n)$  represents the probability that simultaneously  $X_1$  assumes a value that is less than or equal to  $x_1$ ,  $X_2$  assumes a value that is less than or equal to

$x_2, \dots, X_n$  assumes a value that is less than or equal to  $x_n$ ; a more compact way to express this is

$$F_X(\mathbf{x}) = \Pr[\mathbf{X} \leq \mathbf{x}], \quad \mathbf{x} \in \mathbb{R}^n.$$

Even if the df  $F_X$  does not tell us what is the actual value of  $\mathbf{X}$ , it thoroughly describes the range of possible values for  $\mathbf{X}$  and the probabilities assigned to each of them.

### 1.5.2.2 Characterization

The next result establishes the properties that any multivariate df has to fulfil.

#### Property 1.5.7

A multivariate df  $F_X$  is a function mapping  $\mathbb{R}^n$  to  $[0, 1]$  such that:

**P1**  $F_X$  is non-decreasing on  $\mathbb{R}^n$ ;

**P2**  $F_X$  is right-continuous on  $\mathbb{R}^n$ ;

**P3**  $F_X$  satisfies

- (i)  $\lim_{x_j \rightarrow -\infty} F_X(x_1, x_2, \dots, x_n) = 0$  for  $j = 1, 2, \dots, n$ ;
- (ii)  $\lim_{x_1, x_2, \dots, x_n \rightarrow +\infty} F_X(x_1, x_2, \dots, x_n) = 1$ ;
- (iii) for all  $(\alpha_1, \alpha_2, \dots, \alpha_n), (\beta_1, \beta_2, \dots, \beta_n) \in \mathbb{R}^n$ , with  $\alpha_i \leq \beta_i$  for  $i = 1, 2, \dots, n$ , defining

$$\begin{aligned} \Delta_{\alpha_i, \beta_i} F_X(\mathbf{x}) &= F_X(x_1, \dots, x_{i-1}, \beta_i, x_{i+1}, \dots, x_n) \\ &\quad - F_X(x_1, \dots, x_{i-1}, \alpha_i, x_{i+1}, \dots, x_n), \end{aligned}$$

then

$$\Delta_{\alpha_1, \beta_1} \Delta_{\alpha_2, \beta_2} \dots \Delta_{\alpha_n, \beta_n} F_X(\mathbf{x}) \geq 0.$$

**Remark 1.5.8.** Note that condition P3 (iii) ensures that

$$\Pr[\boldsymbol{\alpha} \leq \mathbf{X} \leq \boldsymbol{\beta}] \geq 0 \text{ for any } \boldsymbol{\alpha} \leq \boldsymbol{\beta} \in \mathbb{R}^n.$$

We observe that when  $F_X$  is differentiable, condition P3(iii) is equivalent to

$$\frac{\partial^n}{\partial x_1 \partial x_2 \dots \partial x_n} F_X \geq 0 \text{ on } \mathbb{R}^n.$$

▽

### 1.5.3 Tail functions

In addition to the df, we also introduce a tail function (tf), often called a survival function in biostatistics, and defined as follows:

$$\bar{F}_X(\mathbf{x}) = 1 - F_X(\mathbf{x}) = \Pr[\mathbf{X} > \mathbf{x}], \quad \mathbf{x} \in \mathbb{R}.$$

In words,  $\bar{F}_X(x)$  represents the probability that  $X$  assumes a value larger than  $x$ . If  $X$  is the random future lifetime of a policyholder, then  $\bar{F}_X(x)$  is the probability that the policyholder survives up to age  $x$ . If  $X$  is the total amount of claims produced by a given policyholder then  $\bar{F}_X(x)$  is the probability that the corresponding policy generates a loss larger than  $x$ .

From Definition 1.5.1, we immediately deduce that  $\bar{F}_X$  is non-increasing, right-continuous and such that

$$\lim_{x \rightarrow -\infty} \bar{F}_X(x) = 1 \text{ and } \lim_{x \rightarrow +\infty} \bar{F}_X(x) = 0.$$

We also define

$$\bar{F}_X(x-) = 1 - F_X(x-) = \Pr[X \geq x], \quad x \in \mathbb{R}.$$

Note that this function is non-increasing and left-continuous.

In addition to the multivariate df, we also introduce a multivariate tf  $\bar{F}_X$  defined as

$$\bar{F}_X(\mathbf{x}) = \Pr[\mathbf{X} > \mathbf{x}], \quad \mathbf{x} \in \mathbb{R}^n.$$

Of course, the simple identity  $F_X \equiv 1 - \bar{F}_X$  does not hold in general.

### 1.5.4 Support

As we can see from the graph of Figure 1.2, the points corresponding to jump discontinuities in the df (such as  $x_4$ ) receive positive probability masses; this yields the following definition.

**Definition 1.5.9.** The point  $a$  is an atom of  $X$  if it is a discontinuity point of the df  $F_X$ , that is,  $F_X(a-) \neq F_X(a)$ . Then  $\Pr[X = a] > 0$  and the mass at the point  $a$  equals the jump of  $F_X$  at  $a$ .  $\nabla$

The set of all the possible outcomes for an rv  $X$  is called its support and is precisely defined next.

**Definition 1.5.10.** The support  $\mathcal{S}_X$  of an rv  $X$  with df  $F_X$  is defined as the set of all the points  $x \in \mathbb{R}$  where  $F_X$  is strictly increasing. Similarly, the support  $\mathcal{S}_X$  of a random vector  $\mathbf{X}$  is defined as the subset of  $\mathbb{R}^n$  consisting of all the points  $\mathbf{x}$  such that  $F_X$  is strictly increasing at  $\mathbf{x}$ .  $\nabla$

### 1.5.5 Discrete random variables

According to the structure of their support, rvs can be classified in different categories. A discrete rv  $X$  assumes only a finite (or countable) number of values,  $x_1, x_2, x_3, \dots$ , say. The support  $\mathcal{S}_X$  of  $X$  thus contains a finite or countable number of elements;  $\mathcal{S}_X = \{x_1, x_2, x_3, \dots\}$ . The df of a discrete rv has jump discontinuities at the values  $x_1, x_2, x_3, \dots$  and is constant in between. The (discrete) probability density function (pdf) is defined as

$$f_X(x_i) = \Pr[X = x_i], \quad i = 1, 2, 3, \dots,$$

and  $f_X(x) = 0$  for  $x \neq x_1, x_2, x_3, \dots$ . Of course, any discrete pdf  $f_X$  has to satisfy  $\sum_i f_X(x_i) = 1$ .

The most important subclass of non-negative discrete rvs is the integer case, in which  $x_i = i$  for  $i \in \mathcal{J} \subseteq \mathbb{N} = \{0, 1, 2, \dots\}$ . The number of claims produced by a given policyholder during a certain reference period is of this type. The discrete probability models used in this book are summarized in Table 1.1.

### 1.5.6 Continuous random variables

An rv  $X$  is called continuous if its support is an interval, a union of intervals or the real (half-) line and the associated df  $F_X$  may be represented as

$$F_X(x) = \int_{-\infty}^x f_X(y) dy, \quad x \in \mathbb{R}, \tag{1.4}$$

for some integrable function  $f_X: \mathbb{R} \rightarrow \mathbb{R}^+$ ;  $f_X$  is called the continuous probability density function (pdf) of  $X$ .

**Remark 1.5.11.** It is worth mentioning that rvs with df of the form (1.4) are called absolutely continuous in probability theory. Continuous rvs refer to rvs with a continuous df (i.e., without atoms). In this book, we will use the term continuous rvs for rvs with a df of the form (1.4). ▽

The function  $f_X$  involved in (1.4) has a physical interpretation: if we plot  $f_X$  in the two-dimensional cartesian coordinates  $(x, y)$  as in Figure 1.3, the area bounded by the plot of  $f_X$ , the horizontal axis and two vertical lines crossing the horizontal axis at  $a$  and  $b$  ( $a < b$ ) determines the value of the probability that  $X$  assumes values in  $(a, b)$ .

We obviously deduce from Definition 1.5.1 together with (1.4) that the pdf  $f_X$  satisfies

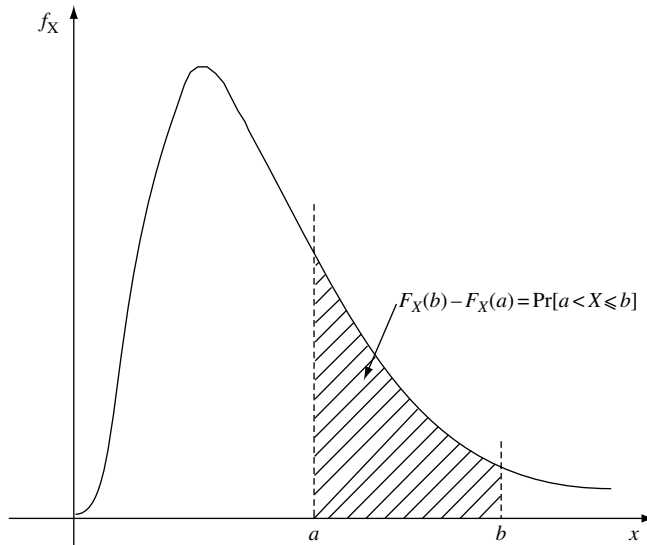
$$\int_{-\infty}^{+\infty} f_X(y) dy = 1.$$

Note that the df  $F_X$  of a continuous rv has derivative  $f_X$ . In other words, the continuous pdf  $f_X$  involved in (1.4) satisfies

$$f_X(x) = \lim_{\Delta x \rightarrow 0} \frac{F_X(x + \Delta x) - F_X(x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{\Pr[x < X \leq x + \Delta x]}{\Delta x}$$

**Table 1.1** Standard discrete probability models

Probability distribution	Notation	Parametric space	Support	Pdf
Bernoulli	$Ber(q)$	$[0, 1]$	0, 1	$q^k (1 - q)^{1-k}$
Binomial	$Bin(m, q)$	$\{1, 2, \dots\} \times [0, 1]$	$\{0, 1, \dots, m\}$	$\binom{m}{k} q^k (1 - q)^{m-k}$
Geometric	$Geo(q)$	$[0, 1]$	$\mathbb{N}$	$q(1 - q)^k$
Negative binomial	$\mathcal{N}Bin(\alpha, q)$	$(0, +\infty) \times (0, 1)$	$\mathbb{N}$	$\binom{\alpha + k - 1}{k} q^\alpha (1 - q)^k$
Poisson	$\mathcal{P}oi(\lambda)$	$\mathbb{R}^+$	$\mathbb{N}$	$\exp(-\lambda) \frac{\lambda^k}{k!}$



**Figure 1.3** Graph of the pdf  $f_X$  corresponding to a continuous rv  $X$

so that the approximation

$$\Pr[x < X \leq x + \Delta x] \approx f_X(x)\Delta x$$

is valid for small  $\Delta x$ . This yields the physical interpretation of the pdf:  $f_X(x)$  can be regarded as the likelihood that  $X \approx x$  (i.e., that  $X$  assumes a value in the neighbourhood of  $x$ ).

Continuous models used in this book are summarized in Table 1.2.

### 1.5.7 General random variables

A more general type of df is a combination of the discrete and (absolutely) continuous cases, being continuous apart from a countable set of exception points  $x_1, x_2, x_3, \dots$  with positive

**Table 1.2** Standard continuous probability models

Probability distribution	Notation	Parametric space	Support	Probability density function
Normal	$\mathcal{N}or(\mu, \sigma^2)$	$\mathbb{R} \times \mathbb{R}^+$	$\mathbb{R}$	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)$
Lognormal	$\mathcal{L}N or(\mu, \sigma^2)$	$\mathbb{R} \times \mathbb{R}^+$	$\mathbb{R}^+$	$\frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(\ln(x) - \mu)^2\right)$
Exponential	$\mathcal{E}xp(\theta)$	$\mathbb{R}^+$	$\mathbb{R}^+$	$\theta \exp(-\theta x)$
Gamma	$\mathcal{G}am(\alpha, \tau)$	$\mathbb{R}^+ \times \mathbb{R}^+$	$\mathbb{R}^+$	$\frac{x^{\alpha-1} \tau^\alpha \exp(-x\tau)}{\Gamma(\alpha)}$
Pareto	$\mathcal{P}ar(\alpha, \theta)$	$\mathbb{R}^+ \times \mathbb{R}^+$	$\mathbb{R}^+$	$\frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}$
Beta	$\mathcal{B}et(\alpha, \beta)$	$\mathbb{R}^+ \times \mathbb{R}^+$	$[0, 1]$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$
Uniform	$\mathcal{U}ni(a, b)$	$\mathbb{R} \times \mathbb{R}$	$(a, b)$	$\frac{1}{b-a}$

probabilities of occurrence, causing jumps in the df at these points. Such a df  $F_X$  can be represented as

$$F_X(x) = (1 - p)F_X^{(c)}(x) + pF_X^{(d)}(x), \quad x \in \mathbb{R}, \tag{1.5}$$

for some  $p \in [0, 1]$ , where  $F_X^{(c)}$  is a continuous df and  $F_X^{(d)}$  is a discrete df.

**Example 1.5.12.** A mixed type rv frequently encountered in actuarial science is an insurance risk for which there is a probability mass in zero (the probability of non-occurrence of claims), while the claim amount given that a claim occurs is a continuous rv. For instance, we could assume that the claim amount  $X$  relating to some policy of the portfolio during a given reference period has a df  $F_X$  of the form

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1 - (1 - p)\exp(-\lambda x), & \text{if } x \geq 0. \end{cases}$$

Such an rv takes the value 0 (i.e., no claim reported by the policyholder) with the probability  $p$ . Given that  $X > 0$  (i.e., at least one claim has occurred), the claim amount is  $\mathcal{Exp}(\lambda)$  distributed. Hence,

$$F_X^{(d)}(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1, & \text{if } x \geq 0, \end{cases}$$

$$F_X^{(c)}(x) = 1 - \exp(-\lambda x), \quad x \geq 0.$$

▽

**Remark 1.5.13.** Note that, in general, it can be proven that every df  $F_X$  may be represented as a mixture of three different kinds of df. Specifically, the identity

$$F_X(x) = p_1 F_X^{(d)}(x) + p_2 F_X^{(c)}(x) + p_3 F_X^{(s)}(x),$$

holds for any  $x \in \mathbb{R}$  where  $p_i \geq 0$  for  $i = 1, 2, 3$  and  $p_1 + p_2 + p_3 = 1$ ,  $F_X^{(d)}$  is a discrete df,  $F_X^{(c)}$  is an absolutely continuous df and  $F_X^{(s)}$  is a singular continuous df (which is defined as a df that is a continuous function of  $x$  but  $\frac{d}{dx}F_X^{(s)}(x) = 0$  almost everywhere, that is,  $F_X^{(s)}$  is continuous but has its points of increase on a set of zero Lebesgue measure). In the remainder of this text, we will only consider dfs with  $p_3 = 0$ ; this particular case covers all the situations encountered by actuaries in practice. ▽

## 1.5.8 Quantile functions

### 1.5.8.1 Definition

There are basically two ways to define a generalized inverse for a df; they are both given in the next definition.

**Definition 1.5.14.** Given a df  $F_X$ , we define the inverse functions  $F_X^{-1}$  and  $F_X^{-1+}$  of  $F_X$  as

$$F_X^{-1}(p) = \inf \{x \in \mathbb{R} \mid F_X(x) \geq p\} = \sup \{x \in \mathbb{R} \mid F_X(x) < p\},$$

and

$$F_X^{-1+}(p) = \inf \{x \in \mathbb{R} \mid F_X(x) > p\} = \sup \{x \in \mathbb{R} \mid F_X(x) \leq p\},$$

for  $p \in [0, 1]$ , where, by convention,  $\inf \emptyset = +\infty$  and  $\sup \emptyset = -\infty$ . ∇

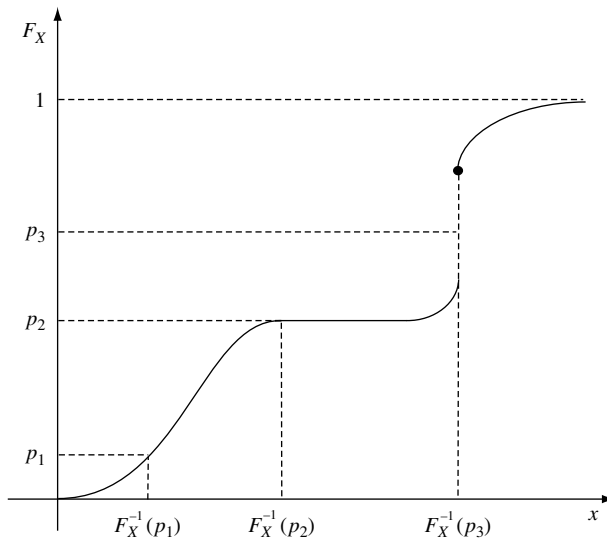
Given some probability level  $p$ ,  $F_X^{-1}(p)$  is the  $p$ th quantile of  $X$  (it is sometimes denoted by  $q_p$ ). To be specific,  $F_X^{-1}(p)$  is a threshold exceeded by  $X$  with probability at most  $1 - p$ . More generally, we adopt the same definitions for the inverses  $t^{-1}$  and  $t^{-1+}$  of any non-decreasing and right-continuous function  $t$ .

### 1.5.8.2 Properties

One can verify that  $F_X^{-1}$  and  $F_X^{-1+}$  are both non-decreasing, and that  $F_X^{-1}$  is left-continuous while  $F_X^{-1+}$  is right-continuous. We have that  $F_X^{-1}(p) = F_X^{-1+}(p)$  if, and only if,  $p$  does not correspond to a ‘flat part’ of  $F_X$  (i.e., a segment  $(x_2, x_3)$  on Figure 1.2 or a probability level  $p_2$  in Figure 1.4), or equivalently, if, and only if,  $F_X^{-1}$  is continuous at  $p$ . As  $F_X^{-1}$  is non-decreasing, it is continuous everywhere, except on an at most countable set of points.

Let us consider Figure 1.4 to illustrate the definition of  $F_X^{-1}$ . When  $F_X$  is one-to-one, as is the case for  $p_1$ ,  $F_X^{-1}(p_1)$  is the standard inverse for  $F_X$  evaluated at  $p_1$  (i.e., the unique  $x$ -value mapped to  $p_1$  by  $F_X$ ) and  $F_X^{-1}(p_1) = F_X^{-1+}(p_1)$ . Two other situations may be encountered, corresponding to  $p_2$  and  $p_3$ . Firstly,  $p_2$  corresponds to a flat part of the graph of  $F_X$ . In this case,  $F_X^{-1}(p_2)$  is the leftmost point of the interval and  $F_X^{-1+}(p_2)$  is the rightmost point of the interval. Note that in this case  $F_X^{-1}(p_2) \neq F_X^{-1+}(p_2)$ . Secondly,  $p_3$  is not a possible value for  $F_X$  (i.e., there is no  $x$ -value such that  $F_X(x) = p_3$ ). In this case,  $F_X^{-1}(p_3)$  is the smallest  $x$ -value mapped to a quantity at least equal to  $p_3$ , and  $F_X^{-1}(p_3) = F_X^{-1+}(p_3)$ .

The following lemma will be frequently used in this book.



**Figure 1.4** Inverse  $F_X^{-1}$  of the df  $F_X$  for different probability levels

**Lemma 1.5.15**

For any real number  $x$  and probability level  $p$ , the following equivalences hold:

- (i)  $F_X^{-1}(p) \leq x \Leftrightarrow p \leq F_X(x)$ ;
- (ii)  $x \leq F_X^{-1+}(p) \Leftrightarrow \Pr[X < x] = F_X(x-) \leq p$ .

*Proof.* We only prove (i); (ii) can be proven in a similar way. The  $\Rightarrow$  part of (i) is proven if we can show that

$$p > F_X(x) \Rightarrow x < F_X^{-1}(p).$$

Assume that  $p > F_X(x)$ . Then there exists an  $\epsilon > 0$  such that  $p > F_X(x + \epsilon)$ . From the sup-definition of  $F_X^{-1}(p)$  in Definition 1.5.14, we find that  $x + \epsilon \leq F_X^{-1}(p)$ , which implies that  $x < F_X^{-1}(p)$ .

We now prove the  $\Leftarrow$  part of (i). If  $p \leq F_X(x)$  then we find that  $p \leq F_X(x + \epsilon)$  for all  $\epsilon > 0$ . From the inf-definition of  $F_X^{-1}(p)$  we can conclude that  $F_X^{-1}(p) \leq x + \epsilon$  for all  $\epsilon > 0$ . Taking the limit for  $\epsilon \downarrow 0$ , we obtain  $F_X^{-1}(p) \leq x$ . □

The following property relates the inverse dfs of the rvs  $X$  and  $t(X)$ , for a continuous non-decreasing function  $t$ .

**Property 1.5.16**

Let  $X$  be an rv. For any  $0 < p < 1$ , the following equalities hold:

- (i) If  $t$  is non-decreasing and continuous then  $F_{t(X)}^{-1}(p) = t(F_X^{-1}(p))$ .
- (ii) If  $t$  is non-decreasing and continuous then  $F_{t(X)}^{-1+}(p) = t(F_X^{-1+}(p))$ .

*Proof.* We only prove (i); (ii) can be proven in a similar way. By application of Lemma 1.5.15, we find that the following equivalences hold for all real  $x$ :

$$\begin{aligned} F_{t(X)}^{-1}(p) \leq x &\Leftrightarrow p \leq F_{t(X)}(x) \\ &\Leftrightarrow p \leq F_X(t^{-1+}(x)) \\ &\Leftrightarrow F_X^{-1}(p) \leq t^{-1+}(x) \\ &\Leftrightarrow t(F_X^{-1}(p)) \leq x. \end{aligned}$$

Note that the above proof only holds if  $t^{-1+}$  is finite. But one can verify that the equivalences also hold if  $t^{-1+}(x) = \pm\infty$ . □

**Remark 1.5.17.** Property 1.5.16 allows us to define an inverse for the tf. The inverses of the df  $F_X$  and of the tf  $\bar{F}_X$  are related by

$$F_X^{-1}(p) = \bar{F}_X^{-1}(1 - p) \text{ and } F_X^{-1+}(p) = \bar{F}_X^{-1+}(1 - p),$$

for any probability level  $p$ . ∇

**Remark 1.5.18.** The continuity assumption put on the function  $t$  in Property 1.5.16 can be relaxed as follows: in (i) it is enough for  $t$  to be left-continuous, whereas in (ii) it is enough for  $t$  to be right-continuous.  $\nabla$

### 1.5.8.3 Probability integral transform theorem

The classical probability integral transform theorem emphasizes the central role of the law  $\mathcal{Uni}(0, 1)$  among continuous dfs. It is stated next.

**Property 1.5.19**

If an rv  $X$  has a continuous df  $F_X$ , then  $F_X(X) \sim \mathcal{Uni}(0, 1)$ .

*Proof.* This follows from Lemma 1.5.15(i) which ensures that for all  $0 < u < 1$ ,

$$\Pr[F_X(X) \geq u] = \Pr[X \geq F_X^{-1}(u)] = \bar{F}_X(F_X^{-1}(u)) = 1 - u,$$

from which we conclude that  $F_X(X) \sim \mathcal{Uni}(0, 1)$ .  $\square$

The probability integral transform theorem has an important ‘inverse’ which is sometimes referred to as the quantile transformation theorem and which is stated next.

**Property 1.5.20**

Let  $X$  be an rv with df  $F_X$ , not necessarily continuous. If  $U \sim \mathcal{Uni}(0, 1)$  then

$$\begin{aligned} X &= {}_d F_X^{-1}(U) = {}_d \bar{F}_X^{-1}(U) \\ &= {}_d F_X^{-1+}(U) = {}_d \bar{F}_X^{-1+}(U). \end{aligned} \tag{1.6}$$

*Proof.* We see from Lemma 1.5.15(i) that

$$\Pr[F_X^{-1}(U) \leq x] = \Pr[U \leq F_X(x)] = F_X(x);$$

the other statements have similar proofs.  $\square$

### 1.5.9 Independence for random variables

A fundamental concept in probability theory is the notion of independence. Roughly speaking, the rvs  $X_1, X_2, \dots, X_n$  are mutually independent when the behaviour of one of these rvs does not influence the others. Formally, the rvs  $X_1, X_2, \dots, X_n$  are mutually independent if, and only if, all the random events constructed with these rvs are independent. This results from the following definition.

**Definition 1.5.21.** The rvs  $X_1, X_2, \dots, X_n$  are independent if, and only if,

$$F_X(\mathbf{x}) = \prod_{i=1}^n F_{X_i}(x_i) \text{ holds for all } \mathbf{x} \in \mathbb{R}^n,$$

or equivalently, if, and only if,

$$\bar{F}_X(x) = \prod_{i=1}^n \bar{F}_{X_i}(x_i) \text{ holds for all } x \in \mathbb{R}^n.$$

∇

In words, the joint df (or tf) of a random vector  $X$  with independent components is thus the product of the marginal dfs (or tfs). Similarly, the joint (discrete or continuous) pdfs of independent random vectors factor into the product of the univariate pdfs.

**Example 1.5.22.** The random couple  $X = (X_1, X_2)$  with joint tf

$$\bar{F}_X(x_1, x_2) = \exp(-\lambda_1 x_1 - \lambda_2 x_2), \quad x_1, x_2 \geq 0, \quad \lambda_1, \lambda_2 > 0,$$

has independent components since  $\bar{F}_X(\mathbf{x}) = \bar{F}_{X_1}(x_1)\bar{F}_{X_2}(x_2)$ , where  $\bar{F}_{X_i}$  is the tf associated with  $\mathcal{E}xp(\lambda_i)$ ,  $i = 1, 2$ . ∇

**Remark 1.5.23.** It is worth mentioning that the mutual independence of Definition 1.5.21 is not equivalent to pairwise independence. In order to check this assertion, consider the random vector  $X = (X_1, X_2, X_3)$  with the distribution defined by  $\Pr[X = (1, 0, 0)] = \Pr[X = (0, 1, 0)] = \Pr[X = (0, 0, 1)] = \Pr[X = (1, 1, 1)] = \frac{1}{4}$ . Then, it is easy to see that  $X_1, X_2$  and  $X_3$  are pairwise independent. However,

$$\Pr[X = (1, 1, 1)] = \frac{1}{4} \neq \frac{1}{8} = \Pr[X_1 = 1]\Pr[X_2 = 1]\Pr[X_3 = 1],$$

and hence  $X_1, X_2$  and  $X_3$  are not mutually independent. ∇

## 1.6 MATHEMATICAL EXPECTATION

### 1.6.1 Construction

Given an rv  $X$ , we can define an important characteristic which is called the mean, or the expected value, and is denoted by  $\mathbb{E}[X]$ . The construction of the operator  $\mathbb{E}[\cdot]$  is briefly recalled next.

If  $X > 0$  and  $\Pr[X = +\infty] > 0$  (i.e.,  $X$  is defective) we put  $\mathbb{E}[X] = +\infty$ , while if  $\Pr[X < +\infty] = 1$  we define

$$\begin{aligned} \mathbb{E}[X] &= \lim_{n \rightarrow +\infty} \sum_{k=1}^{+\infty} \frac{k}{2^n} \Pr \left[ \frac{k}{2^n} < X \leq \frac{k+1}{2^n} \right] \\ &= \lim_{n \rightarrow +\infty} \sum_{k=1}^{+\infty} \frac{k}{2^n} \left( F_X \left( \frac{k+1}{2^n} \right) - F_X \left( \frac{k}{2^n} \right) \right). \end{aligned} \tag{1.7}$$

For an arbitrary rv  $X$ , let us define  $X_+ = \max\{X, 0\}$  and  $X_- = \max\{-X, 0\}$ . Since  $X_+$  and  $X_-$  are non-negative, their expectations can be obtained by (1.7), and if either  $\mathbb{E}[X_+] < +\infty$  or  $\mathbb{E}[X_-] < +\infty$  then

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

We say that the expectation of  $X$  is finite if both  $\mathbb{E}[X_+]$  and  $\mathbb{E}[X_-]$  are finite. Since  $|X| = X_+ + X_-$ , the finiteness of  $\mathbb{E}[X]$  is equivalent to  $\mathbb{E}[|X|] < +\infty$ .

**Example 1.6.1.** If  $X$  has a df of the form (1.4) then  $X$  has a finite expectation if, and only if,

$$\int_{-\infty}^{+\infty} |x|f_X(x)dx < +\infty$$

and it is easy to see from (1.7) that

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

∇

**Remark 1.6.2.** The expectation  $\mathbb{E}[X]$  of any non-negative rv  $X$  is thus defined but may be infinite. For instance, if  $X \sim \text{Par}(\alpha, \theta)$  with  $\alpha \leq 1$  then  $\mathbb{E}[X] = +\infty$ . ∇

**Example 1.6.3.** If  $X$  is discrete with support  $\{x_1, x_2, x_3, \dots\}$  and discrete pdf  $f_X$  then  $X$  has a finite expectation if, and only if

$$\sum_{j \geq 1} |x_j|f_X(x_j) < +\infty,$$

and (1.7) yields

$$\mathbb{E}[X] = \sum_{j \geq 1} x_j f_X(x_j).$$

∇

The representations of the mathematical expectation derived in Examples 1.6.1 and 1.6.3 can be used to compute the expectations associated with the standard probability models presented in Tables 1.1 and 1.2; Table 1.3 summarizes the results.

### 1.6.2 Riemann–Stieltjes integral

Let us assume that  $F_X$  is of the form (1.5) with

$$pF_X^{(d)}(t) = \sum_{d_n \leq t} (F_X(d_n) - F_X(d_n -)) = \sum_{d_n \leq t} \Pr[X = d_n],$$

**Table 1.3** Expectations associated with classical parametric models

Probability law	Expectation	Probability law	Expectation
$Poi(\lambda)$	$\lambda$	$Nor(\mu, \sigma^2)$	$\mu$
$Ber(q)$	$q$	$\mathcal{LN}or(\mu, \sigma^2)$	$\exp(\mu + \frac{\sigma^2}{2})$
$Bin(m, q)$	$mq$	$Exp(\theta)$	$1/\theta$
$Geo(q)$	$\frac{1-q}{q}$	$Gam(\alpha, \tau)$	$\frac{\alpha}{\tau}$
$\mathcal{NB}in(\alpha, q)$	$\frac{\alpha(1-q)}{q}$	$Par(\alpha, \theta)$	$\frac{\theta}{\alpha-1}$ if $\alpha > 1$
		$Bet(\alpha, \beta)$	$\frac{\alpha}{\alpha+\beta}$
		$Uni(a, b)$	$\frac{a+b}{2}$

where  $\{d_1, d_2, \dots\}$  denotes the set of discontinuity points and

$$(1 - p)F_X^{(c)}(t) = F_X(t) - pF_X^{(d)}(t) = \int_{-\infty}^t f_X^{(c)}(x)dx.$$

Then

$$\mathbb{E}[X] = \sum_{n \geq 1} d_n (F_X(d_n) - F_X(d_n -)) + \int_{-\infty}^{+\infty} x f_X^{(c)}(x) dx. \tag{1.8}$$

If we define the differential of  $F_X$ , denoted by  $dF_X$ , as

$$dF_X(x) = \begin{cases} F_X(d_n) - F_X(d_n -), & \text{if } x = d_n, \\ f_X^{(c)}(x), & \text{otherwise,} \end{cases}$$

we then have

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x dF_X(x).$$

This unified notation allows us to avoid tedious repetitions of statements like ‘the proof is given for continuous rvs; the discrete case is similar’. A very readable introduction to differentials and Riemann–Stieltjes integrals can be found in Carter and Van Brunt (2000).

**Example 1.6.4.** The rv  $X$  defined in Example 1.5.12 can be represented as

$$X = \begin{cases} 0, & \text{with probability } p, \\ Y, & \text{with probability } 1 - p, \end{cases}$$

with  $Y \sim \mathcal{Exp}(\lambda)$ . In such a case, actuaries often write  $X$  as the product  $IY$  where  $I$  and  $Y$  are independent, and  $I \sim \mathcal{Ber}(1 - p)$ . Then,

$$pF_X^{(d)}(s) = \begin{cases} 0, & \text{if } s < 0, \\ p, & \text{if } s \geq 0, \end{cases}$$

and

$$\mathbb{E}[X] = 0 \times p + (1 - p) \times \frac{1}{\lambda} = \frac{1 - p}{\lambda}.$$

▽

**Remark 1.6.5.** Let  $\mathbf{X}$  be an  $n$ -dimensional random vector and let  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  be a (measurable) function. Then,  $g(\mathbf{X})$  is a univariate rv so that we can consider its mathematical expectation as

$$\mathbb{E}[g(\mathbf{X})] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g(\mathbf{x}) dF_{\mathbf{X}}(\mathbf{x}),$$

with the notation of the Stieltjes integral.

▽

### 1.6.3 Law of large numbers

The importance of the mathematical expectation originates in the famous law of large numbers, relating this theoretical concept to the intuitive idea of averages in the long run. Specifically, given a sequence  $\{X_1, X_2, \dots\}$  of independent and identically distributed rvs with common expectation  $\mu$ , the sequence of arithmetic averages of the  $X_i$ , that is,

$$\left\{ \bar{X}^{(n)}, n = 1, 2, \dots \right\} \text{ with } \bar{X}^{(n)} = \frac{1}{n} \sum_{i=1}^n X_i,$$

converges to  $\mu$  in the following sense:

$$\Pr \left[ \lim_{n \rightarrow +\infty} \bar{X}^{(n)} = \mu \right] = 1. \quad (1.9)$$

This remarkable result plays a central role in risk management and explains the importance of expected values in actuarial science.

### 1.6.4 Alternative representations for the mathematical expectation in the continuous case

Let us prove that the mathematical expectation can be seen as an integrated right tail.

#### Property 1.6.6

Let  $X$  be a non-negative rv. Then

$$\mathbb{E}[X] = \int_0^{+\infty} \bar{F}_X(x) dx.$$

*Proof.* It suffices to invoke Fubini's theorem and to write

$$\begin{aligned} \mathbb{E}[X] &= \int_0^{+\infty} t dF_X(t) = \int_{t=0}^{+\infty} \int_{x=0}^t dx dF_X(t) \\ &= \int_{x=0}^{+\infty} \int_{t=x}^{+\infty} dF_X(t) dx = \int_0^{+\infty} \bar{F}_X(x) dx. \end{aligned}$$

□

**Remark 1.6.7.** It is worth mentioning that Property 1.6.6 can be generalized to higher dimensions as follows. Let us show that the product moment of the components of an  $n$ -dimensional non-negative random vector  $\mathbf{X}$  can be written as

$$\mathbb{E} \left[ \prod_{i=1}^n X_i \right] = \int_{x_1=0}^{+\infty} \int_{x_2=0}^{+\infty} \dots \int_{x_n=0}^{+\infty} \bar{F}_X(\mathbf{x}) dx_1 dx_2 \dots dx_n. \quad (1.10)$$

To see this, first write

$$\begin{aligned} &\int_{x_1=0}^{+\infty} \int_{x_2=0}^{+\infty} \dots \int_{x_n=0}^{+\infty} \bar{F}_X(\mathbf{x}) dx_1 dx_2 \dots dx_n \\ &= \int_{x_1=0}^{+\infty} \int_{x_2=0}^{+\infty} \dots \int_{x_n=0}^{+\infty} \int_{y_1=x_1}^{+\infty} \int_{y_2=x_2}^{+\infty} \dots \int_{y_n=x_n}^{+\infty} dF_X(\mathbf{y}) dx_1 dx_2 \dots dx_n. \end{aligned}$$

Then invoke Fubini's theorem to get

$$\begin{aligned} & \int_{y_1=0}^{+\infty} \int_{y_2=0}^{+\infty} \cdots \int_{y_n=0}^{+\infty} \int_{x_1=0}^{y_1} \int_{x_2=0}^{y_2} \cdots \int_{x_n=0}^{y_n} dx_1 dx_2 \cdots dx_n dF_X(\mathbf{y}) \\ &= \int_{y_1=0}^{+\infty} \int_{y_2=0}^{+\infty} \cdots \int_{y_n=0}^{+\infty} \left( \prod_{i=1}^n y_i \right) dF_X(\mathbf{y}) = \mathbb{E} \left[ \prod_{i=1}^n X_i \right], \end{aligned}$$

as required. □

### 1.6.5 Alternative representations for the mathematical expectation in the discrete case

Let us now establish a discrete analogue to Property 1.6.6.

#### Property 1.6.8

Let  $N$  be an integer-valued rv. Then

$$\mathbb{E}[N] = \sum_{k=0}^{+\infty} \Pr[N > k].$$

*Proof.* We argue as follows:

$$\begin{aligned} \mathbb{E}[N] &= \Pr[N = 1] + 2 \Pr[N = 2] + 3 \Pr[N = 3] + \dots \\ &= \Pr[N = 1] + \Pr[N = 2] + \Pr[N = 3] + \dots + \\ &\quad \Pr[N = 2] + \Pr[N = 3] + \dots + \\ &\quad \Pr[N = 3] + \dots \\ &= \Pr[N \geq 1] + \Pr[N \geq 2] + \Pr[N \geq 3] + \dots \\ &= \sum_{k=1}^{+\infty} \Pr[N \geq k] = \sum_{k=0}^{+\infty} \Pr[N > k]. \end{aligned}$$

□

### 1.6.6 Stochastic Taylor expansion

#### 1.6.6.1 Univariate case

Suppose we are interested in  $\mathbb{E}[g(X)]$  for some fixed non-linear function  $g$  and some rv  $X$  whose first few moments  $\mu_1, \mu_2, \dots, \mu_n$  are known. A convenient approximation of  $\mathbb{E}[g(X)]$  is based on a naive Taylor expansion of  $g$  around the origin yielding

$$\mathbb{E}[g(X)] \approx \sum_{k=0}^n \frac{g^{(k)}(0)}{k!} \mu_k. \tag{1.11}$$

However, there is no indication about the accuracy of (1.11). Massey and Whitt (1993), derived a probabilistic generalization of Taylor's theorem, suitably modified by Lin (1994).

They give the error when the actuary resorts to the approximation (1.11). In this book we will use some particular cases of their results that we recall now.

### Property 1.6.9

Given a risk  $X$ , assume that the inequalities  $0 < \mathbb{E}[X^s] < +\infty$  hold for some positive integer  $s$ . Let  $g$  be a real-valued function having an  $s$ th derivative  $g^{(s)} \geq 0$ . Then

$$\mathbb{E}[g(X)] = \sum_{k=0}^{s-1} \frac{g^{(k)}(0)}{k!} \mathbb{E}[X^k] + \int_0^{+\infty} \frac{\mathbb{E}[(X-t)_+^{s-1}]}{(s-1)!} g^{(s)}(t) dt. \quad (1.12)$$

*Proof.* Let us start from the Taylor expansion of  $g$  around the origin,

$$g(x) = \sum_{k=0}^{s-1} \frac{g^{(k)}(0)}{k!} x^k + \int_0^{+\infty} \frac{(x-t)_+^{s-1}}{(s-1)!} g^{(s)}(t) dt.$$

It suffices then to invoke Fubini's theorem to get the result.  $\square$

### Corollary 1.6.10

It is interesting to note that for  $s = 1$  and  $2$  we respectively get from (1.12) that

$$\mathbb{E}[g(X)] = g(0) + \int_0^{+\infty} \bar{F}_X(t) g'(t) dt, \quad (1.13)$$

$$\mathbb{E}[g(X)] = g(0) + g'(0)\mu_1 + \int_0^{+\infty} \mathbb{E}[(X-t)_+] g''(t) dt. \quad (1.14)$$

Note that (1.13) reduces to Property 1.6.6 when  $g(x) = x$ .

## 1.6.6.2 Bivariate case

Let us now extend the result of Property 1.6.9 to the bivariate case. The following property is taken from Denuit, Lefèvre and Mesfioui (1999) and will turn out to be useful in the next chapters.

### Property 1.6.11

Let  $\mathbf{X} = (X_1, X_2)$  be a pair of risks such that  $0 < \mathbb{E}[X_1^{s_1}] < +\infty$  and  $0 < \mathbb{E}[X_2^{s_2}] < +\infty$  for some positive integers  $s_1$  and  $s_2$ . Let  $g: \mathbb{R}^2 \rightarrow \mathbb{R}$  with derivatives  $\frac{\partial^{k_1+k_2}}{\partial x_1^{k_1} \partial x_2^{k_2}} g \geq 0$  for  $0 \leq k_1 \leq s_1$ ,  $0 \leq k_2 \leq s_2$ . Then

$$\begin{aligned} \mathbb{E}[g(\mathbf{X})] &= \sum_{i_1=0}^{s_1-1} \sum_{i_2=0}^{s_2-1} \frac{\partial^{i_1+i_2} g(0,0)}{\partial x_1^{i_1} \partial x_2^{i_2}} \frac{\mathbb{E}[X_1^{i_1} X_2^{i_2}]}{i_1! i_2!} \\ &+ \sum_{i_1=0}^{s_1-1} \int_0^{+\infty} \frac{\mathbb{E}[(X_2 - t_2)_+^{s_2-1} X_1^{i_1}]}{(s_2-1)! i_1!} \frac{\partial^{i_1+s_2} g(0, t_2)}{\partial x_1^{i_1} \partial x_2^{s_2}} dt_2 \\ &+ \sum_{i_2=0}^{s_2-1} \int_0^{+\infty} \frac{\mathbb{E}[(X_1 - t_1)_+^{s_1-1} X_2^{i_2}]}{(s_1-1)! i_2!} \frac{\partial^{s_1+i_2} g(t_1, 0)}{\partial x_1^{s_1} \partial x_2^{i_2}} dt_1 \\ &+ \int_0^{+\infty} \int_0^{+\infty} \frac{\mathbb{E}[(X_1 - t_1)_+^{s_1-1} (X_2 - t_2)_+^{s_2-1}]}{(s_1-1)! (s_2-1)!} \frac{\partial^{s_1+s_2} g(t_1, t_2)}{\partial x_1^{s_1} \partial x_2^{s_2}} dt_2 dt_1. \end{aligned}$$

*Proof.* By Taylor’s expansion of  $g$  viewed as a function of  $x_1$  around 0 (for fixed  $x_2$ ), we get

$$g(x_1, x_2) = \sum_{i_1=0}^{s_1-1} \frac{\partial^{i_1} g(0, x_2)}{\partial x_1^{i_1}} \frac{x_1^{i_1}}{i_1!} + \int_0^{x_1} \frac{(x_1 - t_1)^{s_1-1}}{(s_1 - 1)!} \frac{\partial^{s_1} g(t_1, x_2)}{\partial x_1^{s_1}} dt_1. \tag{1.15}$$

Then inserting

$$\frac{\partial^{i_1} g(0, x_2)}{\partial x_1^{i_1}} = \sum_{i_2=0}^{s_2-1} \frac{\partial^{i_1+i_2} g(0, 0)}{\partial x_1^{i_1} \partial x_2^{i_2}} \frac{x_2^{i_2}}{i_2!} + \int_0^{x_2} \frac{(x_2 - t_2)^{s_2-1}}{(s_2 - 1)!} \frac{\partial^{i_1+s_2} g(0, t_2)}{\partial x_1^{i_1} \partial x_2^{s_2}} dt_2,$$

and

$$\frac{\partial^{s_1} g(t_1, x_2)}{\partial x_1^{s_1}} = \sum_{i_2=0}^{s_2-1} \frac{\partial^{s_1+i_2} g(t_1, 0)}{\partial x_1^{s_1} \partial x_2^{i_2}} \frac{x_2^{i_2}}{i_2!} + \int_0^{x_2} \frac{(x_2 - t_2)^{s_2-1}}{(s_2 - 1)!} \frac{\partial^{s_1+s_2} g(t_1, t_2)}{\partial x_1^{s_1} \partial x_2^{s_2}} dt_2,$$

in (1.15) and using Fubini’s theorem yields the result. □

**Corollary 1.6.12**

It is interesting to note that for  $s_1 = s_2 = 1$  we get

$$\begin{aligned} \mathbb{E}[g(X)] &= g(0, 0) + \int_0^{+\infty} \Pr[X_2 > t_2] \frac{\partial g(0, t_2)}{\partial x_2} dt_2 \\ &\quad + \int_0^{+\infty} \Pr[X_1 > t_1] \frac{\partial g(t_1, 0)}{\partial x_1} dt_1 \\ &\quad + \int_0^{+\infty} \int_0^{+\infty} \Pr[X_1 > t_1, X_2 > t_2] \frac{\partial^2 g(t_1, t_2)}{\partial x_1 \partial x_2} dt_2 dt_1. \end{aligned}$$

Note that Corollary 1.6.12 reduces to (1.10) with  $n = 2$  when  $g(x) = x_1 x_2$  is considered.

**1.6.7 Variance and covariance**

The variance is the expected squared difference between an rv  $X$  and its mathematical expectation  $\mu$ . Specifically, the variance of  $X$ , denoted by  $\mathbb{V}[X]$ , is given by

$$\mathbb{V}[X] = \mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2] - \mu^2$$

since the expectation acts as a linear operator.

The variances associated with the standard probability distributions are gathered in Table 1.4.

Given two rvs  $X$  and  $Y$ , the covariance between these rvs is defined as

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

The value of the covariance indicates the extent to which  $X$  and  $Y$  ‘move together’ (hence the name). Nevertheless, we will see in Chapter 5 that the value of the covariance may not be a solid indicator of the strength of dependence existing between two rvs.

**Table 1.4** Variances associated with standard probability distributions

Law	Variance	Law	Variance
$Ber(q)$	$q(1-q)$	$\mathcal{LN}or(\mu, \sigma^2)$	$\exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$
$Bin(m, q)$	$mq(1-q)$	$\mathcal{Exp}(\theta)$	$\frac{1}{\theta^2}$
$Geo(q)$	$\frac{1-q}{q^2}$	$\mathcal{Gam}(\alpha, \tau)$	$\frac{\alpha}{\tau^2}$
$in(\alpha, q)$	$\frac{\alpha(1-q)}{q^2}$	$\mathcal{Par}(\alpha, \theta)$	$\frac{\alpha\theta^2}{(\alpha-2)(\alpha-1)^2}$ if $\alpha > 2$
$Poi(\lambda)$	$\lambda$	$\mathcal{Bet}(\alpha, \beta)$	$\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$
		$\mathcal{Uni}(a, b)$	$\frac{(b-a)^2}{12}$

When non-negative rvs are involved, the following result readily follows from Property 1.6.6 together with (1.10). But it remains valid for arbitrary rvs. The proof given here is taken from Drouet-Mari and Kotz (2001).

**Property 1.6.13**

Given two rvs  $X$  and  $Y$ , their covariance can be represented as

$$\begin{aligned}\mathbb{C}[X, Y] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left( \Pr[X > x, Y > y] - \bar{F}_X(x)\bar{F}_Y(y) \right) dx dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left( \Pr[X \leq x, Y \leq y] - F_X(x)F_Y(y) \right) dx dy.\end{aligned}$$

*Proof.* Let  $(X_1, Y_1)$  and  $(X_2, Y_2)$  be two independent copies of  $(X, Y)$ . Then,

$$\begin{aligned}2\mathbb{C}[X, Y] &= 2\mathbb{E}[X_1 Y_1 - \mathbb{E}[X_1]\mathbb{E}[Y_1]] \\ &= \mathbb{E}[(X_1 - X_2)(Y_1 - Y_2)] \\ &= \mathbb{E} \left[ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\mathbb{I}[u \leq X_1] - \mathbb{I}[u \leq X_2])(\mathbb{I}[v \leq Y_1] - \mathbb{I}[v \leq Y_2]) dudv \right].\end{aligned}$$

Assuming the finiteness of  $|\mathbb{E}[XY]|$ ,  $\mathbb{E}|X|$  and  $\mathbb{E}|Y|$ , we are allowed to exchange the expectations and integral signs, which gives

$$\begin{aligned}2\mathbb{C}[X, Y] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathbb{E} \left[ (\mathbb{I}[u \leq X_1] - \mathbb{I}[u \leq X_2])(\mathbb{I}[v \leq Y_1] - \mathbb{I}[v \leq Y_2]) \right] dudv \\ &= 2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left( \Pr[X \leq u, Y \leq v] - F_X(u)F_Y(v) \right) dudv.\end{aligned}$$

The proof of the other equality is similar.  $\square$

Property 1.6.13 is sometimes referred to as Höfdding's lemma, and can be traced back to Höfdding (1940). It will be useful in the next chapters.

## 1.7 TRANSFORMS

### 1.7.1 Stop-loss transform

#### 1.7.1.1 Definition

Given an rv  $X$ , the rv  $(X - t)_+$ , where  $\xi_+ = \max\{\xi, 0\}$ , represents the amount by which  $X$  exceeds the threshold  $t$ . In an actuarial context,  $t$  is often called the deductible or priority (think of stop-loss reinsurance agreements, for instance).

**Definition 1.7.1.** The function  $\pi_X(t) = \mathbb{E}[(X - t)_+]$  is called the stop-loss transform of  $X$ .  $\nabla$

See Kaas (1993) for details on the use and computation of stop-loss premiums.

#### 1.7.1.2 Properties

It is useful to gather together some properties of the stop-loss transform  $\pi_X$ .

##### Property 1.7.2

Assume that  $\mathbb{E}[|X|] < +\infty$ . The stop-loss transform  $\pi_X$  has the following properties:

- (i) it is decreasing and convex;
- (ii)  $\lim_{t \rightarrow +\infty} \pi_X(t) = 0$  and  $\lim_{t \rightarrow -\infty} \{\pi_X(t) + t\} = \mathbb{E}[X]$ .

*Proof.* (i) follows immediately from the representation

$$\pi_X(t) = \int_t^{+\infty} \bar{F}_X(\xi) d\xi. \quad (1.16)$$

This is a direct consequence of Property 1.6.6 since the tf of the rv  $(X - t)_+$  is  $\bar{F}_X(x + t)$ ,  $x \geq 0$ , and 0 otherwise.

Concerning (ii), the first limit is obvious from (1.16), while the second comes from

$$\lim_{t \rightarrow -\infty} \{\pi_X(t) + t\} = \lim_{t \rightarrow -\infty} \mathbb{E}[\max\{X, t\}] = \mathbb{E}[X].$$

□

#### 1.7.1.3 Characterization

The following property basically states that given a function fulfilling (i)–(ii) of Property 1.7.2, there exists an rv  $X$  for which the function gives the stop-loss premium.

##### Property 1.7.3

For every function  $g$  which satisfies (i)–(ii) of Property 1.7.2, there exists an rv  $X$  such that  $g = \pi_X$ . The df of  $X$  is given by

$$F_X(t) = 1 + g'_+(t),$$

where  $g'_+$  denotes the right-derivative of  $g$ .

*Proof.* If  $g$  is convex, then its right-derivative  $g'_+$  exists and is right-continuous and non-decreasing. Now

$$\lim_{t \rightarrow +\infty} g(t) = 0 \Rightarrow \lim_{t \rightarrow +\infty} g'_+(t) = 0,$$

and  $\lim_{t \rightarrow -\infty} \{g(t) + t\}$  can only exist if  $\lim_{t \rightarrow -\infty} g'_+(t) = -1$ . Hence, Property 1.5.3 ensures that  $1 + g'_+$  is a df,  $F_X$  say. Given  $U \sim \text{Uni}(0, 1)$ , it suffices to take  $X = F_X^{-1}(U)$  according to Property 1.5.20.  $\square$

## 1.7.2 Hazard rate

### 1.7.2.1 Definition

The  $tf$  assesses the likelihood of a large loss:  $\bar{F}_X(x)$  gives the probability of the loss  $X$  exceeding the value  $x$ . Large values of  $\bar{F}_X(x)$  for given  $x$  indicate heavy-tailed behaviour. As pointed out in Klugman, Panjer and Willmot (1998), a quantity that can help the actuary in evaluating tail weight is the hazard rate, whose definition is recalled next.

**Definition 1.7.4.** Given a non-negative rv  $X$  with df (1.4), the associated hazard rate function  $r_X$  is defined as

$$r_X(x) = \frac{f_X(x)}{\bar{F}_X(x)}, \quad x \geq 0.$$

$\nabla$

The hazard rate is referred to as the failure rate in reliability theory. It corresponds to the well-known force of mortality in life insurance.

### 1.7.2.2 Equivalent expression

It is easy to see that

$$r_X(x) = \lim_{\Delta x \rightarrow 0} \frac{\Pr[x < X \leq x + \Delta x | X > x]}{\Delta x}.$$

To check this formula, it suffices to write

$$\begin{aligned} \Pr[x < X \leq x + \Delta x | X > x] &= \frac{\Pr[x < X \leq x + \Delta x]}{\bar{F}_X(x)} \\ &= \frac{\bar{F}_X(x) - \bar{F}_X(x + \Delta x)}{\bar{F}_X(x)}, \end{aligned}$$

whence it follows that

$$\begin{aligned} \lim_{\Delta x \rightarrow 0} \frac{\Pr[x < X \leq x + \Delta x | X > x]}{\Delta x} &= \frac{1}{\bar{F}_X(x)} \lim_{\Delta x \rightarrow 0} \frac{\bar{F}_X(x) - \bar{F}_X(x + \Delta x)}{\Delta x} \\ &= -\frac{1}{\bar{F}_X(x)} \frac{d}{dx} \bar{F}_X(x). \end{aligned}$$

Thus,  $r_X(x)$  may be interpreted as the probability of ‘failure’ at  $x$  given ‘survival’ to  $x$ . Intuitively speaking, if  $r_X$  becomes small then the distribution is heavy-tailed. Conversely, if  $r_X$  becomes large then the distribution is light-tailed.

Note that

$$r_X(x) = -\frac{d}{dx} \ln \bar{F}_X(x), \tag{1.17}$$

and that integrating both sides over  $x$  from 0 to  $t$ , taking  $\bar{F}_X(0) = 1$  into account, gives

$$\bar{F}_X(x) = \exp\left(-\int_0^x r_X(\xi) d\xi\right), \quad x \geq 0. \tag{1.18}$$

Equation (1.18) shows that  $r_X$  uniquely characterizes the distribution.

**Example 1.7.5.** The hazard rate for the  $\mathcal{P}ar(\alpha, \theta)$  distribution is

$$r_X(x) = \frac{\alpha}{\theta + x}.$$

We see that  $r_X$  is strictly decreasing from  $r_X(0) = \alpha/\theta$  to  $r_X(+\infty) = 0$ . ∇

### 1.7.2.3 IFR and DFR distributions

If, as in the above example,  $r_X$  is decreasing then we say that  $X$  has a decreasing failure rate (DFR) distribution. On the other hand, if  $r_X$  is increasing then  $X$  is said to have an increasing failure rate (IFR) distribution. A DFR distribution has an heavier tail than an IFR one.

It is often difficult to examine  $r_X$  when  $\bar{F}_X$  is complicated. Let us now establish the following results, relating the IFR/DFR concepts to log-convexity and log-concavity (precisely defined in Definition 2.8.6).

#### Property 1.7.6

If  $f_X$  is log-convex (log-concave) then  $X$  has a DFR (IFR) distribution.

*Proof.* Starting from

$$\frac{1}{r_X(x)} = \frac{\bar{F}_X(x)}{f_X(x)} = \int_0^{+\infty} \frac{f_X(x+y)}{f_X(x)} dy,$$

we see that if  $f_X(x+y)/f_X(x)$  is an increasing function of  $x$  for any fixed  $y \geq 0$  (i.e.,  $f_X$  is log-convex) then  $1/r_X(x)$  is increasing in  $x$  and  $X$  has a DFR distribution. Similarly, if  $f_X$  is log-concave (i.e., has a Pólya frequency of order 2) then  $X$  has an IFR distribution. □

The sufficient conditions of Property 1.7.6 are often easy to check. Let us now give an equivalent condition for DFR/IFR in terms of the log-convexity/log-concavity of the tfs. This result immediately follows from (1.17).

#### Property 1.7.7

The rv  $X$  has a DFR (IFR) distribution if, and only if,  $\bar{F}_X$  is log-convex (log-concave).

### 1.7.3 Mean-excess function

#### 1.7.3.1 Definition

Another function that is useful in analysing the thickness of tails is the mean-excess loss, whose definition is recalled next.

**Definition 1.7.8.** Given a non-negative rv  $X$ , the associated mean-excess function (mef)  $e_X$  is defined as

$$e_X(x) = \mathbb{E}[X - x | X > x], \quad x > 0.$$

▽

The mef corresponds to the well-known expected remaining lifetime in life insurance. In reliability theory, when  $X$  is a non-negative rv,  $X$  can be thought of as the lifetime of a device and  $e_X(x)$  then expresses the conditional expected residual life of the device at time  $x$  given that the device is still alive at time  $x$ .

#### 1.7.3.2 Equivalent expressions

Intuitively, if  $e_X(x)$  is large for large  $x$ , then the distribution has a heavy tail since the expected loss  $X - x$  is large. Conversely, if  $e_X(x)$  is small for large  $x$ , then the distribution has a light tail. Clearly, if  $F_X(0) = 0$  then  $e_X(0) = \mathbb{E}[X]$ .

Now Property 1.6.6 allows us to write

$$e_X(x) = \int_0^{+\infty} \Pr[X - x > t | X > x] dt = \frac{1}{\bar{F}_X(x)} \int_0^{+\infty} \bar{F}_X(x+t) dt$$

so that (1.16) yields the following useful relationship between the mef and the stop-loss transform

$$e_X(x) = \frac{\pi_X(x)}{\bar{F}_X(x)} = \frac{1}{\frac{d}{dx} \ln \pi_X(x)}, \quad x \geq 0. \quad (1.19)$$

#### 1.7.3.3 Characterization

Clearly,  $e_X(t) \geq 0$ , but not every nonnegative function is an mef corresponding to some rv. The following property gives the characteristics of mefs.

##### Property 1.7.9

A function  $e_X$  is the mef of some continuous non-negative rv if, and only if,  $e_X$  satisfies the following properties:

- (i)  $0 \leq e_X(t) < \infty$  for all  $t \geq 0$ .
- (ii)  $e_X(0) > 0$ .
- (iii)  $e_X$  is continuous.

- (iv)  $e_X(t) + t$  is non-decreasing on  $\mathbb{R}^+$ .
- (v) When there exists a  $t_0$  such that  $e_X(t_0) = 0$ , then  $e_X(t) = 0$  for all  $t \geq t_0$ . Otherwise, when there does not exist such a  $t_0$  with  $e_X(t_0) = 0$ , then

$$\int_0^{+\infty} \frac{1}{e_X(t)} dt = +\infty.$$

### 1.7.3.4 Relationship between the mef and hazard rate

There is a close relationship between  $e_X$  and  $r_X$ . Provided the indicated limits exist, we can write

$$\lim_{x \rightarrow +\infty} e_X(x) = \lim_{x \rightarrow +\infty} \frac{\int_x^{+\infty} \bar{F}_X(t) dt}{\bar{F}_X(x)} = \lim_{x \rightarrow +\infty} \frac{\bar{F}_X(x)}{f_X(x)} = \lim_{x \rightarrow +\infty} \frac{1}{r_X(x)}.$$

This shows that the asymptotic behaviour of  $e_X$  is easily established from that of  $r_X$ , and vice versa.

### 1.7.3.5 IMRL and DMRL distributions

If  $e_X$  is non-decreasing then  $X$  is said to have an increasing mean residual lifetime (IMRL) distribution. Similarly, if  $e_X$  is non-increasing then  $X$  is said to have an decreasing mean residual lifetime (DMRL) distribution.

The following result shows that DFR implies IMRL and IFR implies DMRL.

**Proposition 1.7.10**

The following implications hold:

- (i)  $F_X$  IFR  $\Rightarrow F_X$  DMRL;
- (ii)  $F_X$  DFR  $\Rightarrow F_X$  IMRL.

*Proof.* We only prove (i); the reasoning for (ii) is similar. Since  $F_X$  is IFR, we know from Property 1.7.7 that  $\bar{F}_X$  is log-concave, that is,  $x \mapsto \bar{F}_X(x+y)/\bar{F}_X(x)$  is non-increasing for each fixed  $y \geq 0$ . Hence, for all  $t_1 \leq t_2$ , the inequality

$$\Pr[X - t_1 > y | X > t_1] \geq \Pr[X - t_2 > y | X > t_2]$$

is valid whatever the value of  $y$ . This allows us to write

$$\begin{aligned} e_X(t_1) &= \int_0^{+\infty} \Pr[X - t_1 > y | X > t_1] dy \\ &\geq \int_0^{+\infty} \Pr[X - t_2 > y | X > t_2] dy = e_X(t_2), \end{aligned}$$

which concludes the proof. □

## 1.7.4 Stationary renewal distribution

### 1.7.4.1 Definition

The stationary renewal distribution plays an important role in ruin theory (see Kaas *et al.* 2001, Section 4.7). Let us recall the definition of this concept.

**Definition 1.7.11.** For a non-negative rv  $X$  with finite mean, let  $X_{\{1\}}$  denote an rv with df

$$F_{X_{\{1\}}}(x) = \frac{1}{\mathbb{E}[X]} \int_0^x \bar{F}_X(y) dy = 1 - \frac{\pi_X(x)}{\mathbb{E}[X]}, \quad x \geq 0. \quad (1.20)$$

The df  $F_{X_{\{1\}}}$  is known as the stationary renewal distribution associated with  $X$ . ∇

### 1.7.4.2 Hazard rate associated with stationary renewal distribution

The failure rate  $r_{X_{\{1\}}}$  of  $X_{\{1\}}$  can be written as

$$r_{X_{\{1\}}}(x) = \frac{f_{X_{\{1\}}}(x)}{\bar{F}_{X_{\{1\}}}(x)} = \frac{\bar{F}_X(x)}{\pi_X(x)} = \frac{1}{e_X(x)}$$

by virtue of (1.19). We get from this relationship that

$$\begin{aligned} \bar{F}_X(x) &= e_X(0) f_{X_{\{1\}}}(x) \\ &= e_X(0) r_{X_{\{1\}}}(x) \bar{F}_{X_{\{1\}}}(x) \\ &= \frac{e_X(0)}{e_X(x)} \exp\left(-\int_0^x \frac{1}{e_X(t)} dt\right), \end{aligned}$$

which demonstrates that  $e_X$  uniquely characterizes the distribution.

## 1.7.5 Laplace transform

### 1.7.5.1 Definition

Laplace transforms are useful when positive rvs are being studied. Their definition is recalled next.

**Definition 1.7.12.** The Laplace transform  $L_X$  associated with a risk  $X$ , is given by

$$L_X(t) = \mathbb{E}[\exp(-tX)], \quad t > 0.$$

∇

The expressions for  $L_X$  for the classical continuous parametric models are given in Table 1.5.

**Table 1.5** Laplace transforms associated with standard parametric models

Probability law	$L_X(t)$
$Uni(a, b)$	$\frac{\exp(-at) - \exp(-bt)}{(b-a)t}$
$Bet(\alpha, \beta)$	No closed form available
$Nor(\mu, \sigma^2)$	$\exp(-\mu t + \frac{1}{2}\sigma^2 t^2)$
$Exp(\theta)$	$(1 + \frac{t}{\theta})^{-1}$
$Gam(\alpha, \tau)$	$(1 + \frac{t}{\tau})^{-\alpha}$
$\mathcal{LN}or(\mu, \sigma^2)$	No closed form available
$\mathcal{P}ar(\alpha, \theta)$	No closed form available

### 1.7.5.2 Completely monotone functions and Bernstein’s theorem

The theory of Laplace transforms makes extensive use of complete monotonicity. A function  $g : (0, +\infty) \rightarrow \mathbb{R}^+$  is said to be completely monotone if it satisfies  $(-1)^k g^{(k)} \geq 0$  for all  $k \geq 1$ , where  $g^{(k)}$  denotes the  $k$ th derivative of  $g$ . As  $x \rightarrow 0$ , the derivatives of any completely monotone function  $g$  approach finite or infinite limits denoted by  $g^{(k)}(0)$ . Typical examples of completely monotone functions are  $x \mapsto 1/x$  and  $x \mapsto \exp(-x)$ . It is easy to see that the Laplace transform of any non-negative rv  $X$  is completely monotone. A classical result from real analysis, known as Bernstein’s theorem, states that conversely every completely monotone function  $g$  such that  $g(0) = 1$  is the Laplace transform of some non-negative rv.

**Property 1.7.13**

Given a completely monotone function  $g$ , there exists a measure  $\mu$  on  $\mathbb{R}^+$ , not necessarily finite, such that

$$g(x) = \int_0^{+\infty} \exp(-tx) d\mu(t), \quad x \in \mathbb{R}^+. \tag{1.21}$$

For a proof of this result, see Theorem 1a of Feller (1966, p. 416).

### 1.7.5.3 Discrete Laplace transform: Probability generating function

Probability generating functions characterize integer-valued rvs. Their definition is recalled next.

**Definition 1.7.14.** The probability generating function (pgf) of the integer-valued rv  $N$  is defined as

$$\varphi_N(t) = \mathbb{E}[t^N] = L_N(-\ln t), \quad 0 < t < 1.$$

**Table 1.6** Probability generating functions associated with standard discrete probability models

Probability law	$\varphi_N(t)$
$Ber(q)$	$(1 - q + qt)$
$Bin(m, q)$	$(1 - q + qt)^m$
$Geo(q)$	$\frac{q}{1 - (1 - q)t}$
$\mathcal{NBin}(\alpha, q)$	$\left(\frac{q}{1 - (1 - q)t}\right)^\alpha$
$Poi(\lambda)$	$\exp(\lambda(t - 1))$

The pgfs associated with the classical integer-valued parametric models are given in Table 1.6.

## 1.7.6 Moment generating function

### 1.7.6.1 Definition

The moment generating function (mgf) is a widely used tool in many statistics texts, as it is in Kaas *et al.* (2001). These functions serve to prove statements about convolutions of distributions, and also about limits. Unlike Laplace transforms or risks, mgfs do not always exist. If the mgf of an rv exists in some neighbourhood of 0, it is called light-tailed.

**Definition 1.7.15.** The mgf of the risk  $X$ , denoted by  $M_X$ , is given by

$$M_X(t) = \mathbb{E}[\exp(tX)], \quad t > 0.$$

▽

It is interesting to mention that  $M_X$  characterizes the probability distribution of  $X$ , that is, the information contained in  $F_X$  and  $M_X$  is equivalent.

The mgfs associated with the classical continuous parametric models are given in Table 1.7.

**Table 1.7** Moment generating functions associated with standard parametric models

Probability laws	$M_X(t)$
$Uni(a, b)$	$\frac{\exp(bt) - \exp(at)}{(b - a)t}$
$Bet(\alpha, \beta)$	No closed form available
$Nor(\mu, \sigma^2)$	$\exp(\mu t + \frac{1}{2}\sigma^2 t^2)$
$Exp(\theta)$	$(1 - \frac{t}{\theta})^{-1}$ if $t < \theta$
$Gam(\alpha, \tau)$	$(1 - \frac{t}{\tau})^{-\alpha}$ if $t < \tau$

### 1.7.6.2 The mgf and thickness of tails

If  $h > 0$  exists such that  $M_X(t)$  exists and is finite for  $0 < t < h$ , then the Taylor expansion of the exponential function yields

$$M_X(t) = 1 + \sum_{n=1}^{+\infty} \frac{t^n}{n!} \mathbb{E}[X^n] \text{ for } 0 < t < h. \tag{1.22}$$

It is well known that if any moment of a distribution is infinite, the mgf does not exist. However, it is conceivable that there might exist distributions with moments of all orders and, yet, the mgf does not exist in any neighbourhood around 0. In fact, the  $\mathcal{LNor}(\mu, \sigma^2)$  distribution is one such example.

The set  $\mathcal{E} = \{t > 0 | M_X(t) < +\infty\}$  can be the positive real half-line, a finite interval or even the empty set. Let  $t_{\max} = \sup \mathcal{E}$ . If  $\mathcal{E} \neq \emptyset$ , we see that  $t \mapsto M_X(t)$  is a well-defined continuous and strictly increasing function of  $t \in [0, t_{\max})$ , with value 1 at the origin. If the mgf of  $X$  is finite for a value  $t_0 > 0$  then there exists a constant  $b > 0$  such that for all  $x \geq 0$ ,

$$\bar{F}_X(x) \leq b \exp(-t_0 x).$$

In other words,  $X$  has an exponentially bounded tail.

## 1.8 CONDITIONAL DISTRIBUTIONS

### 1.8.1 Conditional densities

Let  $\mathbf{X} = (X_1, X_2, \dots, X_n)'$  and  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)'$  be two random vectors, possibly of different dimensions, and let  $g(\mathbf{x}, \mathbf{y})$  be the value of their joint pdf at any points  $\mathbf{x} \in \mathbb{R}^n$  and  $\mathbf{y} \in \mathbb{R}^m$ . Let  $f_Y$  be the pdf of  $\mathbf{Y}$  and consider any point  $\mathbf{y} \in \mathbb{R}^m$  such that  $f_Y(\mathbf{y}) > 0$ . Then, the conditional pdf of  $\mathbf{X}$  given  $\mathbf{Y} = \mathbf{y}$ , denoted by  $f_{X|Y}(\mathbf{x}|\mathbf{y})$ , is defined at any point  $\mathbf{x} \in \mathbb{R}^n$  as

$$f_{X|Y}(\mathbf{x}|\mathbf{y}) = \frac{g(\mathbf{x}, \mathbf{y})}{f_Y(\mathbf{y})}. \tag{1.23}$$

The definition of the conditional pdf  $f_{X|Y}(\cdot|\mathbf{y})$  is irrelevant for any point  $\mathbf{y} \in \mathbb{R}^m$  such that  $f_Y(\mathbf{y}) = 0$  since these points form a set having zero probability.

The next result is an extension of Bayes' theorem.

#### Proposition 1.8.1

Let  $\mathbf{X} = (X_1, X_2, \dots, X_n)'$  and  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)'$  be two random vectors with respective pdfs  $f_X$  and  $f_Y$  and conditional pdfs  $f_{X|Y}(\mathbf{x}|\mathbf{y})$  and  $f_{Y|X}(\mathbf{y}|\mathbf{x})$  defined according to (1.23). Then, for any point  $\mathbf{x} \in \mathbb{R}^n$  such that  $f_X(\mathbf{x}) > 0$ ,

$$f_{Y|X}(\mathbf{y}|\mathbf{x}) = \frac{f_{X|Y}(\mathbf{x}|\mathbf{y})f_Y(\mathbf{y})}{\int_{\mathbf{t} \in \mathbb{R}^m} f_{X|Y}(\mathbf{x}|\mathbf{t})f_Y(\mathbf{t})d\mathbf{t}}.$$

As above, given a measurable function  $\Psi: \mathbb{R}^{n+m} \rightarrow \mathbb{R}$ , the conditional expectation of  $\Psi(\mathbf{X}, \mathbf{Y})$  given  $\mathbf{Y}$ , denoted by  $\mathbb{E}[\Psi(\mathbf{X}, \mathbf{Y})|\mathbf{Y}]$ , is defined as a function of the random vector  $\mathbf{Y}$  whose value  $\mathbb{E}[\Psi(\mathbf{X}, \mathbf{Y})|\mathbf{Y} = \mathbf{y}]$  when  $\mathbf{Y} = \mathbf{y}$  is given by

$$\mathbb{E}[\Psi(\mathbf{X}, \mathbf{Y})|\mathbf{Y} = \mathbf{y}] = \int_{\mathbf{x} \in \mathbb{R}^n} \Psi(\mathbf{x}, \mathbf{y})f_{X|Y}(\mathbf{x}|\mathbf{y})d\mathbf{x}.$$

### 1.8.2 Conditional independence

Let  $\mathbf{Y}$  be an  $m$ -dimensional random vector. The risks  $X_1, X_2, \dots, X_n$  are conditionally independent given  $\mathbf{Y}$  if the identity

$$f_{X|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) = \prod_{i=1}^n f_{X_i|\mathbf{Y}}(x_i|\mathbf{y})$$

holds for every  $\mathbf{x} \in \mathbb{R}^n$  and  $\mathbf{y} \in \mathbb{R}^m$ . In particular, if the risks  $X_1, X_2, \dots, X_n$  are conditionally independent given  $\mathbf{Y}$ , then

$$\mathbb{E}[X_i|X_j, j \neq i, \mathbf{Y}] = \mathbb{E}[X_i|\mathbf{Y}].$$

### 1.8.3 Conditional variance and covariance

Let  $\mathbf{Y}$  be an  $m$ -dimensional random vector. The conditional covariance of the risks  $X_1$  and  $X_2$  given  $\mathbf{Y}$  is the rv

$$\mathbb{C}[X_1, X_2|\mathbf{Y}] = \mathbb{E}\left[(X_1 - \mathbb{E}[X_1|\mathbf{Y}])(X_2 - \mathbb{E}[X_2|\mathbf{Y}]) \middle| \mathbf{Y}\right].$$

The conditional variance of  $X_1$  given  $\mathbf{Y}$  is the rv

$$\mathbb{V}[X_1|\mathbf{Y}] = \mathbb{C}[X_1, X_1|\mathbf{Y}].$$

The conditional variances and covariances have the following properties.

#### Property 1.8.2

- (i)  $\mathbb{C}[X_1, X_2] = \mathbb{E}\left[\mathbb{C}[X_1, X_2|\mathbf{Y}]\right] + \mathbb{C}\left[\mathbb{E}[X_1|\mathbf{Y}], \mathbb{E}[X_2|\mathbf{Y}]\right]$ .
- (ii)  $\mathbb{V}[X_1] = \mathbb{E}\left[\mathbb{V}[X_1|\mathbf{Y}]\right] + \mathbb{V}\left[\mathbb{E}[X_1|\mathbf{Y}]\right]$ .
- (iii) If  $X_1$  and  $X_2$  are conditionally independent given  $\mathbf{Y}$ ,  $\mathbb{C}[X_1, X_2|\mathbf{Y}] = 0$ .
- (iv) If  $t: \mathbb{R}^n \rightarrow \mathbb{R}$  is square-integrable then  $\mathbb{C}[t(\mathbf{Y}), X|\mathbf{Y}] = 0$ .

### 1.8.4 The multivariate normal distribution

In the univariate case, an rv  $X$  is said to have a normal distribution with mean  $\mu$  and variance  $\sigma^2$  if its pdf is of the form

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}Q_1(x; \mu, \sigma^2)\right), \quad x \in \mathbb{R},$$

with

$$Q_1(x; \mu, \sigma^2) = \frac{1}{\sigma^2}(x - \mu)^2 = (x - \mu)(\sigma^2)^{-1}(x - \mu)$$

where  $\mu \in \mathbb{R}$  and  $\sigma^2 > 0$ . The bivariate normal distribution introduced below is a natural extension of the univariate normal pdf.

**Definition 1.8.3.** (i) A random couple  $X = (X_1, X_2)'$  is said to have a non-singular bivariate normal distribution if its pdf is of the form

$$f_X(\mathbf{x}) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}Q_2(\mathbf{x}; \boldsymbol{\mu}, \Sigma)\right), \quad \mathbf{x} \in \mathbb{R}^2,$$

where

$$Q_2(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix},$$

$\sigma_i^2 > 0, i = 1, 2, |\sigma_{12}| < \sigma_1\sigma_2$ .

(ii)  $X$  is said to have a singular normal distribution function if there exist real numbers  $\sigma_1, \sigma_2, \mu_1$  and  $\mu_2$  such that  $X =_d (\sigma_1 Z + \mu_1, \sigma_2 Z + \mu_2)$ , where  $Z$  is  $\mathcal{N}or(0, 1)$  distributed and  $\sigma_i > 0, i = 1, 2$ . ∇

The extension of Definition 1.8.3 to higher dimensions is straightforward. Given an  $n \times n$  positive definite matrix  $\Sigma$  and a real vector  $\boldsymbol{\mu}$ , define  $Q_n(\mathbf{x}, \boldsymbol{\mu}, \Sigma) = (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})$ . The random vector  $X = (X_1, X_2, \dots, X_n)'$  is said to have a multivariate normal distribution if its pdf is of the form

$$f_X(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}Q_n(\mathbf{x}; \boldsymbol{\mu}, \Sigma)\right), \quad \mathbf{x} \in \mathbb{R}^n. \tag{1.24}$$

Henceforth, we denote the fact that the random vector  $X$  has multivariate normal distribution with pdf (1.24) by  $X \sim \mathcal{N}or_n(\boldsymbol{\mu}, \Sigma)$ . A good reference for the multivariate normal distribution is Tong (1990).

A convenient characterization of the multivariate normal distribution is as follows:  $X \sim \mathcal{N}or_n(\boldsymbol{\mu}, \Sigma)$  if, and only if, any rv of the form  $\sum_{i=1}^n \alpha_i X_i$  with  $\alpha \in \mathbb{R}^n$ , has the univariate normal distribution.

Let us now compute the conditional distribution associated with the multivariate normal distribution.

**Property 1.8.4**

Let  $X = (X_1, X_2)$  have the bivariate normal distribution with parameters  $\boldsymbol{\mu}$  and  $\Sigma$ . Let  $r = \sigma_{12}/\sigma_1\sigma_2$ . Then:

- (i) the marginal distribution of  $X_i$  is normal with parameters  $\mu_i$  and  $\sigma_i^2, i = 1, 2$ ;
- (ii) for  $|r| < 1$ , the conditional distribution of  $X_1$  given  $X_2 = x_2$  is normal with mean

$$\mu_1 + r \frac{\sigma_1}{\sigma_2} (x_2 - \mu_2)$$

and variance  $\sigma_1^2(1 - r^2)$ ;

- (iii)  $X_1$  and  $X_2$  are independent if, and only if,  $r = 0$ .

*Proof.* Since  $|\Sigma| = \sigma_1^2 \sigma_2^2 (1 - r^2)$ , the inverse of  $\Sigma$  exists if, and only if,  $|r| < 1$ . Straightforward calculation shows that

$$\Sigma^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - r^2)} \begin{pmatrix} \sigma_2^2 & -r\sigma_1\sigma_2 \\ -r\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}.$$

For  $|r| < 1$ , we can write

$$f_X(\mathbf{x}) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left(-\frac{1}{2(1-r^2)} \times \left(\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - 2r\left(\frac{x_1 - \mu_1}{\sigma_1}\right)\left(\frac{x_2 - \mu_2}{\sigma_2}\right) + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2\right)\right).$$

From the identity

$$\begin{aligned} &\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - 2r\left(\frac{x_1 - \mu_1}{\sigma_1}\right)\left(\frac{x_2 - \mu_2}{\sigma_2}\right) + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 \\ &= (1 - r^2)\left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 + \left(\frac{x_1 - \mu_1}{\sigma_1} - r\frac{x_2 - \mu_2}{\sigma_2}\right)^2 \end{aligned}$$

we get

$$f_X(\mathbf{x}) = f_2(x_2)f_{1|2}(x_1|x_2)$$

with

$$f_2(x_2) = \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_2^2}(x_2 - \mu_2)^2\right)$$

and

$$f_{1|2}(x_1|x_2) = \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-r^2}} \exp\left(-\frac{1}{2\sigma_1^2(1-r^2)} \times \left(x_1 - \left(\mu_1 + r\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)\right)\right)^2\right).$$

Combining these expressions yields the required results.  $\square$

From the above results, it now becomes clear that  $\mu$  and  $\Sigma$  satisfy

$$\mathbb{E}[\mathbf{X}] = \begin{pmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \end{pmatrix} = \boldsymbol{\mu}$$

and

$$\mathbb{C}[\mathbf{X}] = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^t] = \begin{pmatrix} \mathbb{V}[X_1] & \mathbb{C}[X_1, X_2] \\ \mathbb{C}[X_1, X_2] & \mathbb{V}[X_2] \end{pmatrix} = \boldsymbol{\Sigma}.$$

Thus,  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are the mean vector and the covariance matrix of the multivariate normal distribution. The multivariate normal distribution also has the following useful invariance property.

**Property 1.8.5**

Let  $C$  be a given  $n \times n$  matrix with real entries and let  $b$  be an  $n$ -dimensional real vector. If  $X \sim \mathcal{N}or_n(\mu, \Sigma)$  then  $Y = CX + b$  is  $\mathcal{N}or_n(C\mu + b, C\Sigma C')$ .

Property 1.8.5 enables us to obtain a bivariate normal vector  $X$  with any mean vector  $\mu$  and covariance matrix  $\Sigma$  through a transformation of two independent  $\mathcal{N}or(0, 1)$  rvs  $Z_1$  and  $Z_2$ . It suffices, indeed, to resort to the transformation  $X = CZ + \mu$  with

$$C = \begin{pmatrix} \sigma_1 & 0 \\ \frac{\sigma_{12}}{\sigma_1} & \sigma_2 \sqrt{1 - r^2} \end{pmatrix}$$

which is non-singular if, and only if,  $|r| < 1$ . Furthermore,  $C$  satisfies  $CC' = \Sigma$ .

Note that any other ‘square root’ of  $\Sigma$  does the job as well. Quite convenient is the lower triangular matrix that can be constructed using the Cholesky decomposition; it also works for dimensions higher than 2.

**1.8.5 The family of the elliptical distributions**

This section is devoted to elliptical distributions that can be seen as convenient extensions of multivariate normal distributions. A standard reference on the topic is Fang, Kotz and Ng (1990). The reading of Gupta and Varga (1993) is also instructive. This section is based on Valdez and Dhaene (2004). We refer the reader to Frahm, Junker and Szimayer (2003) for a discussion about the applicability of the elliptical distributions.

The characteristic function plays an important role in the theory of elliptical distributions. The characteristic function of  $X \sim \mathcal{N}or_n(\mu, \Sigma)$  is given by

$$\mathbb{E}[\exp(i\xi^t X)] = \exp(i\xi^t \mu) \exp\left(-\frac{1}{2} \xi^t \Sigma \xi\right), \quad \xi \in \mathbb{R}^n. \tag{1.25}$$

The class of multivariate elliptical distributions is a natural extension of the class of multivariate normal distributions, as can be seen from the next definition.

**Definition 1.8.6.** The random vector  $X$  is said to have an elliptical distribution with parameters  $\mu$  and  $\Sigma$  if its characteristic function can be expressed as

$$\mathbb{E}[\exp(i\xi^t X)] = \exp(i\xi^t \mu) \phi(\xi^t \Sigma \xi), \tag{1.26}$$

for some function  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  and where  $\Sigma$  is given by

$$\Sigma = AA^t \tag{1.27}$$

for some  $n \times m$  matrix  $A$ . We denote the fact that  $X$  has characteristic function (1.26) by  $X \sim \mathcal{E}ll_n(\mu, \Sigma, \phi)$ . ▽

In (1.25), the generator of the multivariate normal distribution is given by  $\phi(u) = \exp(-u/2)$ .

It is well known that the characteristic function of a random vector always exists and that there is a one-to-one correspondence between probability distributions and characteristic functions. Note, however, that not every function  $\phi$  can be used to construct a characteristic function of an elliptical distribution. Obviously, this function  $\phi$  should already fulfil the requirement  $\phi(0) = 1$ . Moreover, a necessary and sufficient condition for the function  $\phi$  to be a characteristic generator of an  $n$ -dimensional elliptical distribution is given in Theorem 2.2 of Fang, Kotz and Ng (1990).

Note that (1.27) guarantees that the matrix  $\Sigma$  is symmetric, positive definite and has positive elements on the main diagonal. Hence, denoting by  $\sigma_{kl}$  the elements of  $\Sigma$  for any  $k$  and  $l$ , one has that  $\sigma_{kl} = \sigma_{lk}$ , whereas  $\sigma_{kk} > 0$  (which is denoted by  $\sigma_k^2$ ).

It is interesting to note that in the one-dimensional case, the class of elliptical distributions consists mainly of the class of symmetric distributions which include the well-known normal and Student distributions.

We have seen above that an  $n$ -dimensional random vector  $\mathbf{X}$  is  $\mathcal{N}or_n(\boldsymbol{\mu}, \Sigma)$  distributed if, and only if, any linear combination  $\boldsymbol{\alpha}'\mathbf{X}$  of the  $X_k$  has a univariate normal distribution with mean  $\boldsymbol{\alpha}'\boldsymbol{\mu}$  and variance  $\boldsymbol{\alpha}'\Sigma\boldsymbol{\alpha}$ . It is straightforward to generalize this result to the case of multivariate elliptical distributions.

### Property 1.8.7

An  $n$ -dimensional random vector  $\mathbf{X}$  has the  $\mathcal{E}ll_n(\boldsymbol{\mu}, \Sigma, \phi)$  distribution if, and only if, for any vector  $\boldsymbol{\alpha} \in \mathbb{R}^n$ , one has

$$\boldsymbol{\alpha}'\mathbf{X} \sim \mathcal{E}ll_1(\boldsymbol{\alpha}'\boldsymbol{\mu}, \boldsymbol{\alpha}'\Sigma\boldsymbol{\alpha}, \phi).$$

From Property 1.8.7, we find in particular that for  $k = 1, 2, \dots, n$ ,

$$X_k \sim \mathcal{E}ll_1(\mu_k, \sigma_k^2, \phi). \quad (1.28)$$

Hence, the marginal components of a multivariate elliptical distribution have an elliptical distribution with the same characteristic generator.

Defining

$$S = \sum_{k=1}^n X_k = \mathbf{e}'\mathbf{X}$$

where  $\mathbf{e} = (1, 1, \dots, 1)'$ , it follows that

$$\mathbf{X} \sim \mathcal{E}ll_n(\boldsymbol{\mu}, \Sigma, \phi) \Rightarrow S \sim \mathcal{E}ll_1(\mathbf{e}'\boldsymbol{\mu}, \mathbf{e}'\Sigma\mathbf{e}, \phi) \quad (1.29)$$

where  $\mathbf{e}'\boldsymbol{\mu} = \sum_{k=1}^n \mu_k$  and  $\mathbf{e}'\Sigma\mathbf{e} = \sum_{k=1}^n \sum_{l=1}^n \sigma_{kl}$ .

In the following result, it is stated that any random vector with components that are linear combinations of the components of an elliptical distribution is again an elliptical distribution with the same characteristic generator.

**Property 1.8.8**

For any  $m \times n$  matrix  $\mathbf{B}$ , any vector  $\mathbf{c} \in \mathbb{R}^m$  and any random vector  $\mathbf{X} \sim \mathcal{E}ll_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ , we have that

$$\mathbf{B}\mathbf{X} + \mathbf{c} \sim \mathcal{E}ll_m(\mathbf{B}\boldsymbol{\mu} + \mathbf{c}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^t, \phi). \tag{1.30}$$

It is easy to see that Property 1.8.8 is a generalization of Property 1.8.7.

Suppose that for a random vector  $\mathbf{X}$ , the expectation  $\mathbb{E} \left[ \prod_{k=1}^n Y_k^{r_k} \right]$  exists for some set of non-negative integers  $r_1, r_2, \dots, r_n$ . Then this expectation can be found from the relation

$$\mathbb{E} \left[ \prod_{k=1}^n Y_k^{r_k} \right] = \frac{1}{i^{r_1+r_2+\dots+r_n}} \left( \frac{\partial^{r_1+r_2+\dots+r_n}}{\partial i_1^{r_1} \partial i_2^{r_2} \dots \partial i_n^{r_n}} \mathbb{E} [\exp(i\xi^t \mathbf{X})] \right) \Big|_{\xi=\mathbf{0}} \tag{1.31}$$

where  $\mathbf{0} = (0, 0, \dots, 0)^t$ .

The moments of  $\mathbf{X} \sim \mathcal{E}ll_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$  do not necessarily exist. However, from (1.26) and (1.31) we deduce that if  $\mathbb{E}[X_k]$  exists, then it will be given by

$$\mathbb{E}[X_k] = \mu_k \tag{1.32}$$

so that  $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu}$ , if the mean vector exists. Moreover, if  $\mathbb{C}[X_k, X_l]$  and/or  $\mathbb{V}[X_k]$  exist, then they will be given by

$$\mathbb{C}[X_k, X_l] = -2\phi'(0) \sigma_{kl} \tag{1.33}$$

and/or

$$\mathbb{V}[X_k] = -2\phi'(0) \sigma_k^2, \tag{1.34}$$

where  $\phi'$  denotes the first derivative of the characteristic generator. In short, if the covariance matrix of  $\mathbf{X}$  exists, then it is given by  $-2\phi'(0) \boldsymbol{\Sigma}$ . A necessary condition for this covariance matrix to exist is

$$|\phi'(0)| < \infty,$$

see Cambanis, Huang and Simons (1981).

The following result, due to Kelker (1970), shows that any multivariate elliptical distribution with mutually independent components must necessarily be multivariate normal.

**Property 1.8.9**

Let  $\mathbf{X} \sim \mathcal{E}ll_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$  with mutually independent components. Assume that the expectations and variances of the  $X_k$  exist and that  $\mathbb{V}[X_k] > 0$ . Then it follows that  $\mathbf{X}$  is multivariate normal.

*Proof.* Independence of the rvs and existence of their expectations imply that the covariances exist and are equal to 0. Hence, we find that  $\boldsymbol{\Sigma}$  is a diagonal matrix, and that

$$\phi(\boldsymbol{\xi}^t \boldsymbol{\xi}) = \prod_{k=1}^n \phi(\boldsymbol{\xi}_k^2)$$

holds for all  $n$ -dimensional vectors  $\xi$ . This equation is known as Hamel's equation, and its solution has the form

$$\phi(x) = \exp(-\alpha x),$$

for some positive constant  $\alpha$  satisfying  $\alpha = -\phi'(0)$ . To prove this, first note that

$$\phi(\xi' \xi) = \phi\left(\sum_{k=1}^n \xi_k^2\right) = \prod_{k=1}^n \phi(\xi_k^2)$$

or equivalently,

$$\phi(u_1 + \dots + u_n) = \phi(u_1) \dots \phi(u_n).$$

Let us now make the (unnecessary) assumption of differentiability of  $\phi$ .

Consider the partial derivative with respect to  $u_k$ , for some  $k = 1, 2, \dots, n$ . We have

$$\begin{aligned} \frac{\partial \phi}{\partial u_k} &= \lim_{h \rightarrow 0} \frac{\phi(u_1 + \dots + (u_k + h) + \dots + u_n) - \phi(u_1 + \dots + u_n)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\phi(u_1 + \dots + u_n) \phi(h) - \phi(u_1 + \dots + u_n)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\phi(u_1 + \dots + u_n) [\phi(h) - \phi(0)]}{h} \\ &= \phi(u_1) \dots \phi(u_n) \phi'(0). \end{aligned}$$

But the left-hand side is

$$\frac{\partial \phi}{\partial u_k} = \phi(u_1) \dots \phi'(u_k) \dots \phi(u_n) = \phi(u_1) \dots \phi(u_n) \frac{\phi'(u_k)}{\phi(u_k)}.$$

Thus, equating the two, we get

$$\frac{\phi'(u_k)}{\phi(u_k)} = \phi'(0)$$

which gives the desired solution  $\phi(x) = \exp(-\alpha x)$  with  $\alpha = -\phi'(0)$ . This leads to the characteristic generator of a multivariate normal.  $\square$

An elliptically distributed random vector  $X \sim \mathcal{E}ll_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$  does not necessarily have a multivariate density function  $f_X$ . A necessary condition for  $X$  to have a density is that  $\text{rank}(\boldsymbol{\Sigma}) = n$ . If  $X \sim \mathcal{E}ll_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$  has a density, then it will be of the form

$$f_X(\mathbf{x}) = \frac{c}{\sqrt{|\boldsymbol{\Sigma}|}} g((\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})) \tag{1.35}$$

for some non-negative function  $g(\cdot)$  satisfying the condition

$$0 < \int_0^\infty z^{n/2-1} g(z) dz < \infty \tag{1.36}$$

and a normalizing constant  $c$  given by

$$c = \frac{\Gamma(n/2)}{\pi^{n/2}} \left( \int_0^\infty z^{n/2-1} g(z) dz \right)^{-1}. \tag{1.37}$$

Also, the converse statement holds. Any non-negative function  $g(\cdot)$  satisfying (1.36) can be used to define an  $n$ -dimensional density of the form (1.35) for an elliptical distribution, with  $c$  given by (1.37). The function  $g(\cdot)$  is called the *density generator*. One sometimes writes  $X \sim \mathcal{E}ll_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$  for the  $n$ -dimensional elliptical distributions generated from the function  $g(\cdot)$ . A detailed proof of these results, using spherical transformations of rectangular coordinates, can be found in Landsman and Valdez (2002).

Note that for a given characteristic generator  $\phi$ , the density generator  $g$  and/or the normalizing constant  $c$  may depend on the dimension of the random vector  $X$ . Often one considers the class of elliptical distributions of dimensions  $1, 2, 3, \dots$ , all derived from the same characteristic generator  $\phi$ . If these distributions have a density, we will denote their respective density generators by  $g_n$ , where the subscript  $n$  denotes the dimension of the random vector  $X$ .

**Example 1.8.10.** One immediately finds that the density generators and the corresponding normalizing constants of the multivariate normal random vectors  $X \sim \mathcal{N}or_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  for  $n = 1, 2, \dots$  are given by

$$g_n(u) = \exp(-u/2) \tag{1.38}$$

and

$$c_n = (2\pi)^{-n/2}, \tag{1.39}$$

respectively. ∇

**Example 1.8.11.** As an example, let us consider the elliptical Student distribution  $X \sim \mathcal{E}ll_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g_n)$ , with

$$g_n(u) = \left( 1 + \frac{u}{m} \right)^{-(n+m)/2}.$$

We will denote this multivariate distribution (with  $m$  degrees of freedom) by  $t_n^{(m)}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Its multivariate density is given by

$$f_X(\mathbf{x}) = \frac{c_n}{\sqrt{|\boldsymbol{\Sigma}|}} \left( 1 + \frac{(\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})}{m} \right)^{-(n+m)/2}. \tag{1.40}$$

In order to determine the normalizing constant, first note from (1.37) that

$$\begin{aligned} c_n &= \frac{\Gamma(n/2)}{\pi^{n/2}} \left( \int_0^\infty z^{n/2-1} g(z) dz \right)^{-1} \\ &= \frac{\Gamma(n/2)}{\pi^{n/2}} \left( \int_0^\infty z^{n/2-1} \left( 1 + \frac{z}{m} \right)^{-(n+m)/2} dz \right)^{-1}. \end{aligned}$$

Performing the substitution  $u = 1 + (z/m)$ , we find that

$$\int_0^\infty z^{n/2-1} \left(1 + \frac{z}{m}\right)^{-(n+m)/2} dz = m^{n/2} \int_1^\infty (1-u^{-1})^{n/2-1} u^{-m/2-1} du.$$

Making one more substitution  $v = 1 - u^{-1}$ , we get

$$\int_0^\infty z^{n/2-1} \left(1 + \frac{z}{m}\right)^{-(n+m)/2} dz = m^{n/2} \frac{\Gamma(n/2)\Gamma(m/2)}{\Gamma((n+m)/2)},$$

from which we find that

$$c_n = \frac{\Gamma((n+m)/2)}{(m\pi)^{n/2} \Gamma(m/2)}. \tag{1.41}$$

From Property 1.8.7 and (1.28), we have that the marginals of the multivariate elliptical Student distribution are again Student distributions hence  $X_k \sim t_1^{(m)}(\mu_k, \sigma_k^2)$ . The results above lead to

$$f_{X_k}(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{(m\pi)^{1/2} \Gamma\left(\frac{m}{2}\right)} \frac{1}{\sigma_k} \left(1 + \frac{1}{m} \left(\frac{x - \mu_k}{\sigma_k}\right)^2\right)^{-(m+1)/2}, \quad k = 1, 2, \dots, n, \tag{1.42}$$

which is indeed the well-known density of a univariate Student rv with  $m$  degrees of freedom. Its mean is

$$\mathbb{E}[Y_k] = \mu_k, \tag{1.43}$$

and it can be verified that its variance is given by

$$\mathbb{V}[X_k] = \frac{m}{m-2} \sigma_k^2, \tag{1.44}$$

provided the degrees of freedom  $m > 2$ . Note that  $\frac{m}{m-2} = -2\phi'(0)$ , where  $\phi$  is the characteristic generator of the family of Student distributions with  $m$  degrees of freedom.  $\nabla$

In Table 1.8, we consider some well-known families of the class of multivariate elliptical distributions. Each family consists of all elliptical distributions constructed from one particular characteristic generator  $\phi(u)$ . For more details about these families of elliptical distributions, see Landsman and Valdez (2003) and the references therein.

An  $n$ -dimensional random vector  $\mathbf{Z}$  is said to have a multivariate standard normal distribution if all the  $Z_i$  are mutually independent  $\mathcal{N}or(0, 1)$  distributed. We will write this as  $\mathbf{Z} \sim \mathcal{N}or_n(\mathbf{0}_n, \mathbf{I}_n)$ , where  $\mathbf{I}_n$  denotes the  $n \times n$  identity matrix. The characteristic function of  $\mathbf{Z}$  is given by

$$\mathbb{E}[\exp(i\xi'\mathbf{Z})] = \exp\left(-\frac{1}{2}\xi'\xi\right). \tag{1.45}$$

Hence, from (1.45), we find that the characteristic generator of  $N_n(\mathbf{0}_n, \mathbf{I}_n)$  is given by  $\phi(u) = \exp(-u/2)$ . The class of multivariate spherical distributions is an extension of the class of standard multivariate normal distributions.

**Table 1.8** Some families of elliptical distributions with their characteristic generator and/or density generator

Family	Density $g_n(\cdot)$ or characteristic $\phi(\cdot)$ generator
Cauchy	$g_n(u) = (1 + u)^{-(n+1)/2}$
Exponential power	$g_n(u) = \exp(-r(u)^s), r, s > 0$
Laplace	$g_n(u) = \exp(- u )$
Logistic	$g_n(u) = \frac{\exp(-u)}{(1 + \exp(-u))^2}$
Normal	$g_n(u) = \exp(-u/2); \phi(u) = \exp(-u/2)$
Stable laws	$\phi(u) = \exp(-r(u)^{s/2}), 0 < s \leq 2, r > 0$
Student	$g_n(u) = (1 + \frac{u}{m})^{-(n+m)/2}, m > 0$ an integer

**Definition 1.8.12.** A random vector  $\mathbf{Z}$  is said to have an  $n$ -dimensional spherical distribution with characteristic generator  $\phi$  if  $\mathbf{Z} \sim \mathcal{E}ll_n(\mathbf{0}_n, \mathbf{I}_n, \phi)$ . ∇

We will often use the notation  $\mathcal{S}ph_n(\phi)$  for  $\mathcal{E}ll_n(\mathbf{0}_n, \mathbf{I}_n, \phi)$  in the case of spherical distributions. From the definition above, we find that the random vector  $\mathbf{Z} \sim \mathcal{S}ph_n(\phi)$  if, and only if,

$$\mathbb{E}[\exp(i\xi'\mathbf{Z})] = \phi(\xi'\xi). \tag{1.46}$$

Consider an  $m$ -dimensional random vector  $\mathbf{X}$  such that

$$\mathbf{X} =_d \boldsymbol{\mu} + \mathbf{A}\mathbf{Z} \tag{1.47}$$

for some vector  $\boldsymbol{\mu}$ , some  $n \times m$  matrix  $\mathbf{A}$  and some  $m$ -dimensional random vector  $\mathbf{Z} \sim \mathcal{S}ph_n(\phi)$ . Then it is straightforward to prove that  $\mathbf{X} \sim \mathcal{E}ll_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ , where  $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}'$ .

Observe that from the characteristic functions of  $\mathbf{Z}$  and  $\boldsymbol{\alpha}'\mathbf{Z}$ , one immediately finds the following result.

**Property 1.8.13**

$\mathbf{Z} \sim \mathcal{S}ph_n(\phi)$  if, and only if, for any  $n$ -dimensional vector  $\boldsymbol{\alpha}$ , one has

$$\frac{\boldsymbol{\alpha}'\mathbf{Z}}{\sqrt{\boldsymbol{\alpha}'\boldsymbol{\alpha}}} \sim \mathcal{S}ph_1(\phi). \tag{1.48}$$

As a special case of this result, we find that any component  $Z_i$  of  $\mathbf{Z}$  has a  $\mathcal{S}ph_1(\phi)$  distribution.

From the results concerning elliptical distributions, we find that if a spherical random vector  $\mathcal{S}ph_n(\phi)$  possesses a density  $f_{\mathbf{Z}}$ , then it has to be a pdf of the form

$$f_{\mathbf{Z}}(\mathbf{z}) = cg(\mathbf{z}'\mathbf{z}), \tag{1.49}$$

where the density generator  $g$  satisfies (1.36) and the normalizing constant  $c$  satisfies (1.37). Furthermore, the converse also holds: any non-negative function  $g(\cdot)$  satisfying (1.36) can be

used to define an  $n$ -dimensional pdf  $cg(z'z)$  of a spherical distribution with the normalizing constant  $c$  satisfying (1.37). One often writes  $\mathcal{Sph}_n(g)$  for the  $n$ -dimensional spherical distribution generated from the density generator  $g(\cdot)$ .

The following result explores the conditional distributions in the case of elliptical random vectors. It extends Property 1.8.4 to the class of elliptical laws. For a proof, see Valdez and Dhaene (2004).

**Property 1.8.14**

Let the random vector  $Y$  be  $\mathcal{Ell}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$  distributed with density generator  $g_n(\cdot)$ . Define  $\Lambda_\alpha$  and  $\Lambda_\beta$  to be linear combinations of the components of  $X$ , that is,  $\Lambda_\alpha = \boldsymbol{\alpha}'X$  and  $\Lambda_\beta = \boldsymbol{\beta}'X$ , for some  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta} \in \mathbb{R}^n$ . Then, we have that

$$(\Lambda_\alpha, \Lambda_\beta) \sim \mathcal{Ell}_2(\boldsymbol{\mu}_{\alpha\beta}, \boldsymbol{\Sigma}_{\alpha\beta}, \phi), \quad (1.50)$$

where

$$\boldsymbol{\mu}_{\alpha\beta} = \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}'\boldsymbol{\mu} \\ \boldsymbol{\beta}'\boldsymbol{\mu} \end{pmatrix}, \quad (1.51)$$

$$\begin{aligned} \boldsymbol{\Sigma}_{\alpha\beta} &= \begin{pmatrix} \sigma_\alpha^2 & r_{\alpha\beta}\sigma_\alpha\sigma_\beta \\ r_{\alpha\beta}\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha} & \boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\beta} \\ \boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\beta} & \boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta} \end{pmatrix}. \end{aligned} \quad (1.52)$$

Furthermore, given  $\Lambda_\beta = \lambda$ , the rv  $\Lambda_\alpha$  has the univariate elliptical distribution

$$\mathcal{Ell}_1\left(\mu_\alpha + r_{\alpha\beta}\frac{\sigma_\alpha}{\sigma_\beta}(\lambda - \mu_\beta), (1 - r_{\alpha\beta}^2)\sigma_\alpha^2, \phi_a\right), \quad (1.53)$$

for some characteristic generator  $\phi_a(\cdot)$  depending on  $a = (\lambda - \mu_\beta)^2 / \sigma_\beta^2$ .

From this result, it follows that the characteristic function of  $\Lambda_\alpha | \Lambda_\beta = \lambda$  is given by

$$\mathbb{E}[\exp(it\Lambda_\alpha) | \Lambda_\beta = \lambda] = \exp\left(i\mu_{\Lambda_\alpha|\Lambda_\beta=\lambda}t\right)\phi_a\left(\sigma_{\Lambda_\alpha|\Lambda_\beta=\lambda}^2t^2\right),$$

where

$$\mu_{\Lambda_\alpha|\Lambda_\beta=\lambda} = \mu_\alpha + r_{\alpha\beta}\frac{\sigma_\alpha}{\sigma_\beta}(\lambda - \mu_\beta)$$

and

$$\sigma_{\Lambda_\alpha|\Lambda_\beta=\lambda}^2 = (1 - r_{\alpha\beta}^2)\sigma_\alpha^2.$$

## 1.9 COMONOTONICITY

### 1.9.1 Definition

A standard way of modelling situations where individual rvs  $X_1, \dots, X_n$  are subject to the same external mechanism is to use a secondary mixing distribution. The uncertainty about the external mechanism is then described by a structure variable  $z$ , which is a realization of an rv  $Z$  and acts as a (random) parameter of the distribution of  $X$ . The aggregate claims can then be seen as a two-stage process: first, the external parameter  $Z = z$  is drawn from the df  $F_Z$  of  $Z$ . The claim amount of each individual risk  $X_i$  is then obtained as a realization from the conditional df of  $X_i$  given  $Z = z$ . This construction is known as a common mixture model and will be studied in detail in Chapter 7.

A special type of mixing model is the case where given  $Z = z$ , the claim amounts  $X_i$  are degenerate on  $x_i$ , where the  $x_i = x_i(z)$  are non-decreasing in  $z$ . Such a model is in a sense an extreme form of a mixing model, as in this case the external parameter  $Z = z$  completely determines the aggregate claims. In such a case, the risks  $X_1, \dots, X_n$  are said to be comonotonic. Comonotonicity is discussed in Kaas *et al.* (2001, Section 10.6). The definition of this concept is recalled next.

**Definition 1.9.1.** A random vector  $X$  is comonotonic if and only if there exist an rv  $Z$  and non-decreasing functions  $t_1, t_2, \dots, t_n$ , such that

$$X =_d (t_1(Z), t_2(Z), \dots, t_n(Z))'$$

▽

In this book, the notation  $(X_1^c, \dots, X_n^c)$  will be used to indicate a comonotonic random vector. The support of  $X^c$  is

$$\mathcal{S}_{X^c} = \{(F_{X_1}^{-1}(p), F_{X_2}^{-1}(p), \dots, F_{X_n}^{-1}(p)) \mid 0 < p < 1\}.$$

Note that this support is an ordered set, since  $s, t \in \mathcal{S}_{X^c}$  entails either  $s \leq t$  or  $s \geq t$  componentwise.

### 1.9.2 Comonotonicity and Fréchet upper bound

Fréchet spaces offer the natural framework for studying dependence. These spaces gather together all the probability distributions with fixed univariate marginals. Elements in a given Fréchet space only differ in their dependence structures, and not in their marginal behaviours.

**Definition 1.9.2.** Let  $F_1, F_2, \dots, F_n$  be univariate dfs. The Fréchet space  $\mathcal{R}_n(F_1, F_2, \dots, F_n)$  consists of all the  $n$ -dimensional (dfs  $F_X$  of) random vectors  $X$  possessing  $F_1, F_2, \dots, F_n$  as marginal dfs, that is,

$$F_i(x) = \Pr[X_i \leq x], \quad x \in \mathbb{R}, \quad i = 1, 2, \dots, n.$$

▽

The elements of  $\mathcal{R}_n(F_1, F_2, \dots, F_n)$  are bounded above by a special multivariate df, called the Fréchet upper bound, as shown in the next result.

**Property 1.9.3**

Define the Fréchet upper bound as

$$W_n(\mathbf{x}) = \min\{F_1(x_1), F_2(x_2), \dots, F_n(x_n)\}, \quad \mathbf{x} \in \mathbb{R}^n.$$

Then the inequality

$$F_X(\mathbf{x}) \leq W_n(\mathbf{x}) \tag{1.54}$$

holds for all  $\mathbf{x} \in \mathbb{R}^n$  and  $\mathbf{X} \in \mathcal{R}_n(F_1, F_2, \dots, F_n)$ .

*Proof.* This is obvious since  $\cap_{i=1}^n [X_i \leq x_i] \subseteq [X_j \leq x_j]$  for any  $j \in \{1, \dots, n\}$ . □

Note that  $W_n$  is an element of  $\mathcal{R}_n(F_1, F_2, \dots, F_n)$ . Indeed, given an rv  $U \sim \text{Uni}(0, 1)$ ,  $W_n$  is the df of

$$(F_1^{-1}(U), F_2^{-1}(U), \dots, F_n^{-1}(U)) \in \mathcal{R}_n(F_1, F_2, \dots, F_n),$$

since

$$\begin{aligned} \Pr [F_1^{-1}(U) \leq x_1, F_2^{-1}(U) \leq x_2, \dots, F_n^{-1}(U) \leq x_n] \\ = \Pr [U \leq \min\{F_1(x_1), F_2(x_2), \dots, F_n(x_n)\}] = W_n(\mathbf{x}). \end{aligned}$$

We have thus proven the next result, which relates comonotonicity to the Fréchet upper bound.

**Proposition 1.9.4**

A random vector  $\mathbf{X} \in \mathcal{R}_n(F_1, F_2, \dots, F_n)$  is comonotonic if, and only if, its multivariate df is  $W_n$ .

**Remark 1.9.5.** Early results about dependence are due to Höfdding (1940) and Fréchet (1951). Until recently, the work of Höfdding did not receive the attention it deserved, due primarily to the fact that his papers were published in relatively obscure German journals at the outbreak of World War II. Unaware of Höfdding's work, Fréchet independently rediscovered many of the same results, which has led to terms such as 'Fréchet spaces' and 'Fréchet bounds'.

Fréchet bounds have attracted a lot of interest in different fields of application. They have been extended in a number of ways during the last few decades. See Nelsen *et al.* (2004) for an illustration. ▽

## 1.10 MUTUAL EXCLUSIVITY

### 1.10.1 Definition

In this section we introduce, following Dhaene and Denuit (1999), a kind of opposite of comonotonicity, namely mutual exclusivity. Note that we restrict ourselves to risks, that is, to non-negative rvs in this section. We will work in Fréchet spaces  $\mathcal{R}_n^+(F_1, F_2, \dots, F_n)$ , where the  $F_i$  are such that  $F_1(0-) = F_2(0-) = \dots = F_n(0-) = 0$ .

Roughly speaking, the risks  $X_1, X_2, \dots, X_n$  are said to be mutually exclusive when at most one of them can be different from zero. This can be considered as a sort of dual notion of comonotonicity. Indeed, the knowledge that one risk assumes a positive value directly implies that all the others vanish.

**Definition 1.10.1.** The multivariate risk  $X$  in  $\mathcal{R}_n^+(F_1, F_2, \dots, F_n)$  is said to be mutually exclusive when

$$\Pr[X_i > 0, X_j > 0] = 0 \text{ for all } i \neq j.$$

∇

We observe that mutual exclusivity of  $X$  means that its multivariate pdf  $f_X$  is concentrated on the axes.

**Remark 1.10.2.** In the bivariate case, the concept of countermonotonicity has attracted a lot of interest. Let us recall that the bivariate risk is said to be countermonotonic if it is distributed as  $(t_1(Z), t_2(Z))$  for some rv  $Z$ , an increasing function  $t_1$  and a decreasing function  $t_2$ . Therefore, increasing the value of one component tends to decrease the value of the other. Countermonotonicity does not extend to higher dimensions. This is why mutual exclusivity has been used instead for dimensions higher than 2. ∇

### 1.10.2 Fréchet lower bound

The elements of  $\mathcal{R}_n(F_1, F_2, \dots, F_n)$  are bounded below by a special function, called the Fréchet lower bound, as shown in the next result.

#### Property 1.10.3

Let us define the Fréchet lower bound as

$$M_n(\mathbf{x}) = \max \left\{ \sum_{i=1}^n F_i(x_i) - (n-1), 0 \right\}, \quad \mathbf{x} \in \mathbb{R}^n.$$

Then the inequality

$$M_n(\mathbf{x}) \leq F_X(\mathbf{x}) \tag{1.55}$$

holds for all  $\mathbf{x} \in \mathbb{R}^n$  and  $X \in \mathcal{R}_n(F_1, F_2, \dots, F_n)$ .

*Proof.* Obviously,  $\Pr[\cup_{i=1}^n \bar{A}_i] \leq \sum_{i=1}^n \Pr[\bar{A}_i]$  for any choice of events  $A_1, A_2, \dots, A_n$ . Therefore,  $\Pr[\cap_{i=1}^n A_i] \geq \sum_{i=1}^n \Pr[A_i] - n + 1$ . Now take  $A_i = [X_i \leq x_i]$ ,  $i = 1, 2, \dots, n$ .  $\square$

**Remark 1.10.4.** In the bivariate case, the Fréchet lower bound  $M_2$  is an element of  $\mathcal{R}_2(F_1, F_2)$ . Specifically,  $M_2$  is the df of  $(F_1^{-1}(U), F_2^{-1}(1 - U))$ , where  $U \sim \text{Uni}(0, 1)$ . This is easily deduced from

$$\begin{aligned} \Pr[F_1^{-1}(U) \leq x_1, F_2^{-1}(1 - U) \leq x_2] &= \Pr[U \leq F_1(x_1), 1 - U \leq F_2(x_2)] \\ &= M_2(x_1, x_2). \end{aligned}$$

When  $n \geq 3$ , however,  $M_n$  is no longer always a df (it is just a signed measure), as shown by the following counterexample proposed by Tchen (1980): for  $n = 3$ , take  $X_1, X_2$  and  $X_3 \sim \text{Uni}(0, 1)$ ; then the ‘probability’ that  $X$  lies in  $[0.5, 1] \times [0.5, 1] \times [0.5, 1]$  is equal to

$$\begin{aligned} &1 - \Pr[X_1 < 0.5 \text{ or } X_2 < 0.5 \text{ or } X_3 < 0.5] \\ &= 1 - F_1(0.5) - F_2(0.5) - F_3(0.5) + F_{(X_1, X_2)}(0.5, 0.5) + F_{(X_1, X_3)}(0.5, 0.5) \\ &\quad + F_{(X_2, X_3)}(0.5, 0.5) - F_{(X_1, X_2, X_3)}(0.5, 0.5, 0.5) = -0.5 \end{aligned}$$

when the dependence structure is described by  $M_3$ . Hence,  $M_3$  cannot be a proper df.  $\nabla$

From inequalities (1.54) and (1.55) we can derive many useful results, such as the following. They provide bounds on probabilities involving the minimum and the maximum of a set of correlated risks.

**Corollary 1.10.5**

For any  $X \in \mathcal{R}_n(F_1, F_2, \dots, F_n)$ ,

$$\begin{aligned} &1 - \min\{F_1(x), F_2(x), \dots, F_n(x)\} \\ &\leq \Pr[\max\{X_1, X_2, \dots, X_n\} > x] \leq \min \left\{ 1, \sum_{i=1}^n (1 - F_i(x)) \right\}, \text{ for all } x \in \mathbb{R}, \end{aligned}$$

and

$$\begin{aligned} &\max\{F_1(x), F_2(x), \dots, F_n(x)\} \\ &\leq \Pr[\min\{X_1, X_2, \dots, X_n\} \leq x] \leq \min \left\{ 1, \sum_{i=1}^n F_i(x) \right\}, \text{ for all } x \in \mathbb{R}. \end{aligned}$$

These inequalities also provide useful bounds on the distribution of the largest and smallest claims in an insurance portfolio consisting of dependent risks. Therefore, they can be used to get bounds on the premium of an  $LCR(1)$  treaty. By the latter we mean a reinsurance agreement covering the largest claim occurring during a given reference period (one year, say). Of course, when the  $X_i$  are thought of as being time-until-death random variables, these inequalities also yield bounds on life insurance policies or annuities based on either a joint-life status or a last-survivor status. These bounds have been used by Dhaene, Vanneste and Wolthuis (2000) in order to find extremal joint-life and last-survivor statuses.

### 1.10.3 Existence of Fréchet lower bounds in Fréchet spaces

As shown in Remark 1.10.4, Fréchet lower bounds are not necessarily dfs. The following result provides us with necessary and sufficient conditions for  $M_n$  to be a df in  $\mathcal{R}_n(F_1, F_2, \dots, F_n)$ .

**Proposition 1.10.6**

A necessary and sufficient condition for  $M_n$  to be a proper df in  $\mathcal{R}_n(F_1, F_2, \dots, F_n)$  is that, for each  $x$  with  $0 < F_j(x_j) < 1$  for  $j = 1, 2, \dots, n$ , either

$$\sum_{j=1}^n F_j(x_j) \leq 1 \tag{1.56}$$

or

$$\sum_{j=1}^n \bar{F}_j(x_j) \leq 1 \tag{1.57}$$

holds true.

For a proof of this result, we refer the reader to Joe (1997, Theorem 3.7).

### 1.10.4 Fréchet lower bounds and maxima

Despite the fact that  $M_n$  is not always a proper df, Tchen (1980, Theorem 4) proved that there exists  $X \in \mathcal{R}_n(F_1, F_2, \dots, F_n)$  achieving the lower bound  $M_n$  when all the  $x_i$  are equal. This is formally stated in the next result.

**Proposition 1.10.7**

There exists  $X \in \mathcal{R}_n(F_1, F_2, \dots, F_n)$  such that

$$\Pr[\max\{X_1, X_2, \dots, X_n\} \leq x] = M_n(x, x, \dots, x)$$

for any  $x \in \mathbb{R}$ .

### 1.10.5 Mutual exclusivity and Fréchet lower bound

A Fréchet space does not always contain mutually exclusive risks. A necessary and sufficient condition is provided in the following result.

**Proposition 1.10.8**

A Fréchet space  $\mathcal{R}_n^+(F_1, F_2, \dots, F_n)$  contains mutually exclusive risks if, and only if, it satisfies

$$\sum_{i=1}^n q_i \leq 1 \text{ where } q_i = 1 - F_i(0), \quad i = 1, 2, \dots, n. \tag{1.58}$$

*Proof.* First, assume that  $\mathbf{X}$  is mutually exclusive and belongs to  $\mathcal{R}_n^+(F_1, F_2, \dots, F_n)$ . Define the indicator variables  $I_1, I_2, \dots, I_n$  as

$$I_i = \begin{cases} 1, & \text{if } X_i > 0, \\ 0, & \text{if } X_i = 0, \end{cases}$$

so that  $I_i \sim \text{Ber}(q_i)$ ,  $i = 1, 2, \dots, n$ . Note that since  $\mathbf{X}$  is mutually exclusive,

$$\begin{aligned} \Pr[I_1 = I_2 = \dots = I_n = 0] &= 1 - \Pr[I_1 = 1 \text{ or } I_2 = 1 \text{ or } \dots \text{ or } I_n = 1] \\ &= 1 - \sum_{i=1}^n q_i, \end{aligned}$$

so that (1.58) has to be fulfilled.

Conversely, assume that  $\mathcal{R}_n^+(F_1, F_2, \dots, F_n)$  satisfies (1.58). From Proposition 1.10.6, we know that  $M_n$  is a df in  $\mathcal{R}_n^+(F_1, F_2, \dots, F_n)$ . Consider  $\mathbf{X} \in \mathcal{R}_n^+(F_1, F_2, \dots, F_n)$  with df  $M_n$ . Then we find that

$$\Pr[X_i = 0, X_j = 0] = 1 - q_i - q_j \text{ for all } i \neq j,$$

whence it follows that

$$\Pr[X_i > 0, X_j > 0] = 0 \text{ for all } i \neq j,$$

which, in turn, means that  $\mathbf{X}$  is mutually exclusive. □

Let us prove the following characterization of mutual exclusivity, which relates this notion to the Fréchet lower bound (just as comonotonicity corresponds to the Fréchet upper bound). More precisely, we prove that when (1.58) is fulfilled, the multivariate df of the mutually exclusive risks in the Fréchet space  $\mathcal{R}_n^+(F_1, F_2, \dots, F_n)$  is given by the Fréchet lower bound  $M_n$ .

**Proposition 1.10.9**

Consider a Fréchet space  $\mathcal{R}_n^+(F_1, F_2, \dots, F_n)$  satisfying (1.58) and let  $\mathbf{X} \in \mathcal{R}_n^+(F_1, F_2, \dots, F_n)$ . Then,  $\mathbf{X}$  is mutually exclusive if, and only if,

$$F_{\mathbf{X}}(\mathbf{x}) = M_n(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^n.$$

*Proof.* Assume that  $\mathbf{X}$  is mutually exclusive. Defining the indicator variables  $I_i$  as in the proof of Proposition 1.10.8, we have for  $\mathbf{x} \geq \mathbf{0}$  that

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{x}) &= \sum_{i=1}^n \Pr[X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n | I_i = 1] \Pr[I_i = 1] \\ &\quad + \Pr[I_1 = I_2 = \dots = I_n = 0] \\ &= \sum_{i=1}^n \Pr[X_i \leq x_i | I_i = 1] q_i + 1 - \sum_{i=1}^n q_i \\ &= \sum_{i=1}^n (F_i(x_i) - F_i(0)) + 1 - \sum_{i=1}^n q_i \\ &= \sum_{i=1}^n F_i(x_i) + 1 - n = M_n(\mathbf{x}), \end{aligned}$$

which proves that the condition is necessary. That it is also sufficient follows from the second part of the proof of Proposition 1.10.8.  $\square$

Combining Propositions 1.10.8 and 1.10.9, we find that a Fréchet space  $\mathcal{R}_n^+(F_1, F_2, \dots, F_n)$  has the property that the Fréchet lower bound is the unique df of  $\mathcal{R}_n^+(F_1, F_2, \dots, F_n)$  with pdf concentrated on the axes if, and only if, it satisfies (1.58).

**Remark 1.10.10.** In view of Remarks 1.10.2 and 1.10.4 we have that the bivariate risk  $(X_1, X_2)$  is countermonotonic if, and only if, it has  $M_2$  as joint df. For instance, with unit uniform marginals,  $(U, 1 - U)$  with  $U \sim \mathcal{Uni}(0, 1)$  is countermonotonic but not mutually exclusive. Note that in this case, (1.58) is not satisfied.  $\nabla$

### 1.11 EXERCISES

**Exercise 1.11.1.** Show that

$$F_X(\mathbf{x}) = F_1(x_1)F_2(x_2) \left\{ 1 + \epsilon \bar{F}_1(x_1)\bar{F}_2(x_2) \right\}, \quad \mathbf{x} \in \mathbb{R}^2,$$

is a two-dimensional df whose marginals are  $F_1$  and  $F_2$  for  $0 < \epsilon < 1$ .

**Exercise 1.11.2.** Prove that the following chain of equivalences holds true for any  $\mathbf{x} \in \mathbb{R}^2$ : for any  $X \in \mathcal{R}_2(F_1, F_2)$ ,

- (i)  $F_X(x_1, x_2) = \min\{F_1(x_1), F_2(x_2)\} \Leftrightarrow \bar{F}_X(x_1, x_2) = \min\{\bar{F}_1(x_1), \bar{F}_2(x_2)\} \Leftrightarrow \Pr[X_1 \leq x_1, X_2 > x_2] = \max\{F_1(x_1) + \bar{F}_2(x_2) - 1, 0\}$ ;
- (ii)  $F_X(x_1, x_2) = \max\{F_1(x_1) + F_2(x_2) - 1, 0\} \Leftrightarrow \bar{F}_X(x_1, x_2) = \max\{\bar{F}_1(x_1) + \bar{F}_2(x_2) - 1, 0\} \Leftrightarrow \Pr[X_1 \leq x_1, X_2 > x_2] = \min\{F_1(x_1), \bar{F}_2(x_2)\}$ .

**Exercise 1.11.3.** Let  $(X_1, X_2) \in \mathcal{R}_2(F_1, F_2)$  be a random pair with continuous marginals. Prove that:

- (i) the random pair  $(X_1, X_2)$  has df  $W_2$  if, and only if,

$$(X_1, X_2) =_d (X_1, F_2^{-1}(F_1(X_1)));$$

- (ii) the random pair  $(X_1, X_2)$  has df  $M_2$  if, and only if,

$$(X_1, X_2) =_d (X_1, F_2^{-1}(\bar{F}_1(X_1))).$$

**Exercise 1.11.4.** Let  $X$  be the value of a share at a future time  $t, t \geq 0$ . Consider European options with expiration date  $t$ , exercise price  $d$  and the share as underlying asset. Let  $Y^{(c)}$  be the payoff of the call option at time  $t$ , that is,

$$Y^{(c)} = \max\{0, X - d\}.$$

Similarly, let  $Y^{(p)}$  be the payoff of the put option at time  $t$ , that is,

$$Y^{(p)} = \max\{0, d - X\}.$$

Show that  $(X, Y^{(c)})$  are comonotonic, while both  $(X, Y^{(p)})$  and  $(Y^{(p)}, Y^{(c)})$  have  $M_2$  as df.

**Exercise 1.11.5.** Let  $\mathbf{X} = (X_1, X_2, X_3) \in \mathcal{R}_3(F_1, F_2, F_3)$ , for some continuous univariate dfs  $F_1, F_2$  and  $F_3$ . Prove the following assertions:

- (i) If  $(X_1, X_2)$  and  $(X_2, X_3)$  are both comonotonic then so is  $(X_1, X_3)$  and  $F_{\mathbf{X}} \equiv W_3$ .
- (ii) If  $(X_1, X_2)$  is comonotonic and  $(X_2, X_3)$  is countermonotonic then  $(X_1, X_3)$  is countermonotonic and

$$F_{\mathbf{X}}(\mathbf{x}) = \max \left\{ 0, \min \{ F_1(x_1), F_2(x_2) \} + F_3(x_3) - 1 \right\}.$$

- (iii) If  $(X_1, X_2)$  and  $(X_2, X_3)$  are both countermonotonic then  $(X_1, X_3)$  is comonotonic and

$$F_{\mathbf{X}}(\mathbf{x}) = \max \left\{ 0, \min \{ F_1(x_1), F_3(x_3) \} + F_2(x_2) - 1 \right\}.$$

**Exercise 1.11.6.** Show that the inverse df  $F_{S^c}^{-1}$  of a sum  $S^c$  of comonotonic rvs  $X_1^c, X_2^c, \dots, X_n^c$  is given by

$$F_{S^c}^{-1}(p) = \sum_{i=1}^n F_{X_i^c}^{-1}(p), \quad 0 < p < 1. \quad (1.59)$$

**Exercise 1.11.7.** Assume in Exercise 1.11.6 that  $X_i^c \sim \mathcal{Exp}(1/b_i)$ ,  $i = 1, 2, \dots, n$ . Show that

$$F_{S^c}^{-1}(p) = -b_{\bullet} \ln(1-p),$$

where  $b_{\bullet} = \sum_{i=1}^n b_i$ , so that  $S^c \sim \mathcal{Exp}(1/b_{\bullet})$ . In words, the comonotonic sum of exponentially distributed rvs still has an exponential distribution.

**Exercise 1.11.8.** Let  $F_{i_1 i_2 \dots i_k}$  denote the marginal df of  $(X_{i_1}, X_{i_2}, \dots, X_{i_k})$ ,  $1 \leq i_1 < i_2 < \dots < i_k \leq n$ ,  $k = 1, 2, \dots, n$ , that is,

$$F_{i_1 i_2 \dots i_k}(x_{i_1}, x_{i_2}, \dots, x_{i_k}) = \Pr[X_{i_1} \leq x_{i_1}, X_{i_2} \leq x_{i_2}, \dots, X_{i_k} \leq x_{i_k}],$$

$x_{i_1}, x_{i_2}, \dots, x_{i_k} \in \mathbb{R}$ . Assume that all the marginals  $F_{i_1 i_2 \dots i_k}$  are compatible (i.e., that there exists at least one proper multivariate cdf  $F_{\mathbf{X}}$  with these marginals). For any  $\mathbf{x} \in \mathbb{R}^3$ , show that the inequalities

$$F_{\mathbf{X}}(\mathbf{x}) \leq F_U(\mathbf{x}) \equiv \min \left\{ F_{12}(x_1, x_2), F_{13}(x_1, x_3), F_{23}(x_2, x_3), \right. \\ \left. 1 - F_1(x_1) - F_2(x_2) - F_3(x_3) + F_{12}(x_1, x_2) + F_{13}(x_1, x_3) + F_{23}(x_2, x_3) \right\}$$

and

$$F_{\mathbf{X}}(\mathbf{x}) \geq F_L(\mathbf{x}) \equiv \max \left\{ 0, F_{12}(x_1, x_2) + F_{13}(x_1, x_3) - F_1(x_1), \right. \\ \left. F_{12}(x_1, x_2) + F_{23}(x_2, x_3) - F_2(x_2), F_{13}(x_1, x_3) + F_{23}(x_2, x_3) - F_3(x_3) \right\}$$

hold for any  $\mathbf{X} \in \mathcal{R}_3(F_1, F_2, F_3)$ .

**Exercise 1.11.9.** Suppose all the bivariate marginals of  $X \in \mathcal{R}_n(F_1, F_2, \dots, F_n)$  are Fréchet upper bounds, that is,

$$F_{i_1 i_2}(x_{i_1}, x_{i_2}) = \min \{F_{i_1}(x_{i_1}), F_{i_2}(x_{i_2})\} \text{ for all } 1 \leq i_1 < i_2 \leq n.$$

Show that  $F_X \equiv W_n$ , that is,  $X$  is comonotonic.

**Exercise 1.11.10.** Show that

$$F_{S_n}(x) = M_n(x, x, \dots, x) \tag{1.60}$$

if  $X$  is mutually exclusive.

**Exercise 1.11.11.** A franchise deductible divides the risk  $X$  in two parts  $X = X_1 + X_2$ , with the retained part given by

$$X_1 = \begin{cases} X, & \text{if } X < d, \\ 0, & \text{if } X \geq d, \end{cases}$$

and the insured part by

$$X_2 = \begin{cases} 0, & \text{if } X < d, \\ X, & \text{if } X \geq d. \end{cases}$$

Show that  $X_1$  and  $X_2$  are mutually exclusive.

