Chapter 10

Basic Probabilistic Tools for Finance

In this chapter, the reader will find a short summary of the basic probability tools useful for understanding of the following chapters. A more detailed version including proofs can be found in Janssen and Manca (2006).

We will focus our attention on stochastic processes in discrete time and continuous time defined by sequences of random variables.

10.1. The sample space

In order to model finance problems, the basic concrete notion in probability theory is that of the *random experiment*, that is to say an experiment for which we cannot predict in advance the outcome. With each random experiment, we can associate the *elementary events* ω, which often represent the time evolution of the values of an asset on a stock exchange on a time interval $[0,T]$. The set of all these events Ω is called the *sample space*. Some other subsets of Ω will represent possible *events*.

Let us consider the following examples.

Example 10.1 A bank is to invest in some shares, so the bank looks at the history of the value of different shares. In this case, the sample space is the set of all non-negative real numbers $\mathbb{R}^+$.

To be useful, the set of all possible events must have some properties of stability so that we can generate new events such as:

(i) *the complement* A^c : $A^c = \{\omega \in \Omega : \omega \notin A\}$; (10.1)

(ii) *the union* $A \cup B$: $A \cup B = \{\omega : \omega \in A \text{ or } \omega \in B\}$; (10.2)

(iii) *the intersection* $A \cap B$: $A \cap B = \{\omega : \omega \in A, \omega \in B\}$. (10.3)

More generally, if $(A_n, n \geq 1)$ represents a sequence of events, we can also consider the following events:

$$\bigcup_{n \geq 1} A_n, \ \bigcap_{n \geq 1} A_n \tag{10.4}$$

representing respectively the *union* and the *intersection* of all the events of the given sequence. The first of these two events occurs if and only if at least one of these events occurs and the second if and only if all the events of the given sequence occur. The set Ω is called the *certain event* and the set $\varnothing$ the *empty event*. Two events *A* and *B* are said to be *disjoint* or *mutually exclusive* if and only if

$$A \cap B = \varnothing . \tag{10.5}$$

Event A *implies* event *B* if and only if

$$A \subset B . \tag{10.6}$$

In Example 10.1, the event "the value of the share is between \$50 and \$80" is given by the set [50,80].

10.2. Probability space

Given a sample space Ω, the set of all possible events will be noted by $\Im$, assumed to have the structure of an σ*-field* or an σ*-algebra.*

Definition 10.2 The family $\Im$ of subsets of Ω is called a σ-field or a σ-algebra if and only if the following conditions are satisfied:

(i) $\Omega, \varnothing$ *belong* to $\Im$ *;*

(ii) Ω *is stable under a denumerable intersection:*

$$A_n \in \Im, \forall n \geq 1 \Rightarrow \bigcap_{n \geq 1} A_n \in \Im, \tag{10.7}$$

(iii) $\Im$ *is stable for the complement set operation:*

$$A \in \Im \Rightarrow A^c \in \Im, \tag{10.8}$$

(*with* $A^c = \Omega - A$).

Using the well-known de Morgan's laws of set theory, it is easy to prove that a σ*-algebra* $\Im$ is also stable under a denumerable union:

$$A_n \in \Im, \forall n \geq 1 \Rightarrow \bigcup_{n \geq 1} A_n \in \Im. \tag{10.9}$$

Any couple $(\Omega, \Im)$ where $\Im$ is an σ-algebra is called a *measurable space*.

The next definition concerning the concept of *probability measure* or simply *probability* is an idealization of the concept of the *frequency* of an event.

Let us consider a random experiment called *E* with which the couple $(\Omega, \Im)$ is associated; if set *A* belongs to $\Im$ and if we can repeat experiment *E n* times under the same environmental conditions, we can count how many times *A* occurs. If *n*(*A*) represents the number of occurrences, the *frequency* of the event *A* is defined as

$$f(A) = \frac{n(A)}{n}. \tag{10.10}$$

In general, this number tends to become stable for large values of *n*.

The notion of frequency satisfies the following elementary properties:

(i) $\left(A, B \in \Im, A \cap B = \varnothing\right) \Rightarrow f(A \cup B) = f(A) + f(B),$ (10.11)

(ii) $f(\Omega) = 1,$ (10.12)

(iii) $A, B \in \Im, \Rightarrow f(A \cup B) = f(A) + f(B) - f(A \cap B),$ (10.13)

(iv) $A \in \Im \Rightarrow f(A^c) = 1 - f(A).$ (10.14)

In order to have a useful mathematical model for the theoretical idealization of the notion of frequency, we now introduce the following definition.

Definition 10.3

a) *The triplet* $(\Omega, \Im, P)$ *is called a probability space if* Ω *is a non-void set of elements,* $\Im$ *a* σ*-algebra of subsets of* Ω *and P an application from* $\Im$ *to* [0,1] *such that:*

$$(A_n, n \geq 1), A_n \in \Im, n \geq 1 : \left(i \neq j \Rightarrow A_i \cap A_j = \phi\right)$$

(i) $$\Rightarrow P\left(\bigcup_{n\geq 1} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \ (\sigma\text{-}additivity\, of\ P), \tag{10.15}$$

(ii) $$P(\Omega) = 1. \tag{10.16}$$

b) *The application P satisfying conditions* (10.15) *and* (10.16) *is called a probability measure or simply probability.*

Remark 10.1

Relation (10.17) assigns the value 1 for the probability of the entire sample space Ω. There may exist events A' which are strictly subsets of Ω such that

$$P\left(A'\right) = 1. \tag{10.17}$$

In this case, we say that *A* is *almost sure* or that the statement defining *A* is true *almost surely* (in short a.s.) or holds for almost all ω.

From axioms (10.15) and (10.16), we can deduce the following properties.

Property 10.1

(i) If $A, B \in \Im$, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{10.18}$$

(ii) If $A \in \Im$, then

$$P(A^c) = 1 - P(A). \tag{10.19}$$

(iii) $$P(\varnothing) = 0. \tag{10.20}$$

(iv) If $(B_n, n \geq 1)$ is a sequence of disjoint elements of $\Im$ forming a partition of Ω, then for all *A* belonging to $\Im$,

$$P(A) = \sum_{n=1}^{\infty} P(A \cap B_n). \tag{10.21}$$

(v) *Continuity property of P*: if $(A_n, n \geq 1)$ is an increasing (decreasing) sequence of elements of $\Im$, then

$$P\left(\bigcup_{n\geq 1} A_n\right) = \lim_n P(A_n); \left(P\left(\bigcap_{n\geq 1} A_n\right) = \lim_n P(A_n)\right). \tag{10.22}$$

(vi) *Boole's inequality* asserts that if $(A_n, n \geq 1)$ is a sequence of events, then

$$P\left(\bigcup_{n\geq 1} A_n\right) \leq \sum_{n\geq 1} P(A_n). \tag{10.23}$$

Example 10.2

a) *The discrete case*

When the sample space Ω is *finite* or *denumerable*, we can set

$$\Omega = \{\omega_1, ..., \omega_j, ...\} \tag{10.24}$$

and select for $\Im$ the set of all the subsets of Ω, represented by 2^{Ω}.

Any probability measure *P* can be defined with the following sequence:

$$(p_j, j \geq 1),\ p_j \geq 0, j \geq 1,\ \sum_{j\geq 1} p_j = 1 \tag{10.25}$$

so that

$$P(\{w_j\}) = p_j, j \geq 1. \tag{10.26}$$

On the probability space $(\Omega, 2^{\Omega}, P)$, the probability assigned for an arbitrary event $A = \left\{\omega_{k_1}, ..., \omega_{k_l}\right\}, k_j \geq 1, j = 1, ..., l, k_i \neq k_j$ if $i \neq j$ is given by

$$P(A) = \sum_{j=1}^{l} p_{k_j}. \tag{10.27}$$

b) *The continuous case*

Let Ω be the real set $\mathbb{R}$; it can be proven (Halmos (1974)) that there exists a minimal σ-algebra generated by the set of intervals:

$$\beta = \left\{(a,b), [a,b], [a,b), (a,b], a,b \in \mathbb{R}, a \leq b\right\}. \tag{10.28}$$

It is called the *Borel σ-algebra* represented by β and the elements of β are called *Borel sets*.

Given a probability measure *P* on (Ω, β), we can define the real function *F*, called the distribution function related to *P*, as follows.

Definition 10.4 *The function F from* $\mathbb{R}$ *to* $[0,1]$ *defined by:*

$$P\left((-\infty, x]\right) = F(x), x \in \mathbb{R} \tag{10.29}$$

is called the distribution function related to the probability measure P.

From this definition and the basic properties of *P*, we easily deduce that:

$$\begin{aligned} &P\left((a,b]\right) = F(b) - F(a),\ P\left((a,b)\right) = F(b-) - F(a), \\ &P\left([a,b)\right) = F(b-) - F(a-),\ P\left([a,b]\right) = F(b) - F(a-). \end{aligned} \tag{10.30}$$

Moreover, from (10.29), any function *F* from $\mathbb{R}$ to $[0,1]$ is a distribution function (in short d.f.) if and only if it is a non-decreasing function satisfying the following conditions:

– *F* is right continuous at every point x_0,

$$\lim_{x \uparrow x_0} F(x) = F(x_0), \tag{10.31}$$

– and moreover

$$\lim_{x\to+\infty} F(x)=1, \lim_{x\to-\infty} F(x)=0\,. \tag{10.32}$$

If function F is derivable on $\mathbb{R}$ with f as the derivative, we have

$$F(x)=\int_{-\infty}^{x} f(y)dy, x\in\mathbb{R}. \tag{10.33}$$

Function f is called the density function associated with the d.f. F and in the case of the existence of such a Lebesgue integrable function on $\mathbb{R}$, F is said to be *absolutely continuous*.

From the definition of the concept of integral, we can give the intuitive interpretation of f as follows; given the small positive real number Δx, we have:

$$P\left(\{x, x+\Delta x\}\right)\approx f(x)\Delta x\,. \tag{10.34}$$

Using the Lebesgue Stieltjes integral, it can be seen that it is possible to define a probability measure P on $(\mathbb{R},\beta)$ starting from a d.f. F on $\mathbb{R}$ by the following definition of P:

$$P(A)=\int_A dF(x), \forall A\in\Im. \tag{10.35}$$

In the absolutely continuous case, we obtain

$$P(A)=\int_A f(y)dy. \tag{10.36}$$

10.3. Random variables

Let us suppose the probability space $(\Omega,\Im,P)$ and the measurable space (E,ψ) are given.

Definition 10.5 *A random variable* (*in short r.v.*) *with values in E is an application X from* Ω *to E such that*

$$\forall B\in\psi : X^{-1}(B)\in\Im, \tag{10.37}$$

where $X^{-1}(B)$ is called the inverse image of the set B defined by

$$X^{-1}(B) = \{\omega : X(\omega) \in B\}, X^{-1}(B) \in \Im. \tag{10.38}$$

Particular cases

a) If $(E, \psi) = (\mathbb{R}, \beta)$, X is called a *real random variable*.

b) If $(E, \psi) = (\overline{\mathbb{R}}, \overline{\beta})$, where $\overline{\mathbb{R}}$ is the *extended real line* defined by $\mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$ and $\overline{\beta}$ the *extended Borel σ-field* of $\overline{\mathbb{R}}$, that is, the minimal σ-field containing all the elements of β and the extended intervals

$$\begin{aligned} &[-\infty, a), (-\infty, a], [-\infty, a], (-\infty, a), \\ &[a, +\infty), (a, +\infty], [a, +\infty], (a, +\infty), \ a \in \mathbb{R}, \end{aligned} \tag{10.39}$$

X is called a *real extended value random variable*.

c) If $E = \mathbb{R}^n (n > 1)$ with the product σ-field $\beta^{(n)}$ of β, X is called an *n-dimensional real random variable*.

d) If $E = \overline{\mathbb{R}}^{(n)}$ (n>1) with the product σ-field $\overline{\beta}^{(n)}$ of $\overline{\beta}$, X is called an *extended n-dimensional real random variable.*

A r.v. X is called *discrete* or *continuous* according to the fact that X takes a value in a set at most denumerable or non-denumerable.

Remark 10.2 In *measure theory*, the only difference is that condition (10.17) is no longer required and in this case the definition of a r.v. given above gives the notion of a *measurable function*. In particular, a measurable function from $(\mathbb{R}, \beta)$ to $(\mathbb{R}, \beta)$ is called a *Borel function*.

Let X be a real r.v. and let us consider, for any real x, the following subset of Ω: $\{\omega : X(\omega) \le x\}$.

Given that, from relation (10.38),

$$\{\omega : X(\omega) \le x\} = X^{-1}((-\infty, x]), \tag{10.40}$$

it is clear from relation (10.37) that this set belongs to the σ-algebra $\Im$.

Conversely, it can be proved that the condition

$$\{\omega : X(\omega) \le x\} \in \Im, \tag{10.41}$$

valid for every x belonging to a dense subset of $\mathbb{R}$, is sufficient for X being a real r.v. defined on Ω.

The probability measure P on $(\Omega, \Im)$ induces a probability measure μ on $(\mathbb{R}, \beta)$ defined as

$$\forall B \in \beta : \mu(B) = P\left(\{\omega : X(\omega) \in B\}\right). \tag{10.42}$$

We say that μ is the induced probability measure on $(\mathbb{R}, \beta)$, called the *probability distribution* of the r.v. *X*.

Introducing the distribution function related to μ, we obtain the next definition.

Definition 10.6 *The distribution function of the r.v. X, represented by* F_X *, is the function from* $\mathbb{R} \to [0,1]$ *defined by*

$$F_X(x) = \mu\left((-\infty, x]\right) = P\left(\{\omega : X(\omega) \le x\}\right). \tag{10.43}$$

In short, we write

$$F_X(x) = P\left(X \le x\right). \tag{10.44}$$

This last definition can be extended to the multi-dimensional case with r.v. *X* being an *n*-dimensional real vector: $X = (X_1, ..., X_n)$, a measurable application from $(\Omega, \Im, P)$ to $(\mathbb{R}^n, \beta^n)$.

Definition 10.7 *The distribution function of the r.v.* $X = (X_1, ..., X_n)$*, represented by* F_X *, is the function from* $\mathbb{R}^n$ *to* $[0,1]$ *defined by*

$$F_X(x_1, ..., x_n) = P\left(\{\omega : X_1(\omega) \le x_1, ..., X_n(\omega) \le x_n\}\right). \tag{10.45}$$

In short, we write

$$F_X(x_1, ..., x_n) = P(X_1 \le x_1, ..., X_n \le x_n). \tag{10.46}$$

Each component X_i (i=1,…,n) is itself a one-dimensional real r.v. whose d.f., called the *marginal d.f.*, is given by

$$F_{X_i}(x_i) = F_X(+\infty,...,+\infty,x_i,+\infty,...,+\infty). \tag{10.47}$$

The concept of random variable is *stable* under many mathematical operations; thus, any Borel function of a r.v. X is also an r.v.

Moreover, if X and Y are two r.v., so are

$$\inf\{X,Y\}, \sup\{X,Y\}, X+Y, X-Y, X\cdot Y, \frac{X}{Y}, \tag{10.48}$$

provided, in the last case, that Y does not vanish.

Concerning the convergence properties, we must mention the property that, if $(X_n, n \geq 1)$ is a *convergent* sequence of r.v. – that is, for all $\omega \in \Omega$, the sequence $(X_n(\omega))$ converges to $X(\omega)$ – then the limit X is also a r.v. on Ω. This convergence, which may be called the *sure convergence*, can be weakened to give the concept of an a.s. *convergence* of the given sequence.

Definition 10.8 *The sequence* $(X_n(\omega))$ *converges a.s. to* $X(\omega)$ *if*

$$P\left(\{\omega : \lim X_n(\omega) = X(\omega)\}\right) = 1. \tag{10.49}$$

This last notion means that the possible set where the given sequence does not converge is a *null set*, that is, a set N belonging to $\Im$ such that

$$P(N) = 0. \tag{10.50}$$

In general, let us note that, given a null set, it is not true that every subset of it belongs to $\Im$ but of course if it belongs to $\Im$, it is clearly a null set (see relation (10.26)).

To avoid unnecessary complications, we will assume from now on that any considered probability space is *complete*, This means that all the subsets of a null set also belong to $\Im$ and thus that their probability is zero.

10.4. Expectation and independence

Let us consider a complete measurable space $(\Omega, \Im, \mu)$ and a real measurable variable X defined on Ω. Using the concept of an integral, it is possible to define the *expectation* of X represented by

$$E(X) = \int_{\Omega} X dP \left(= \int X dP \right), \tag{10.51}$$

provided that this integral exists. The calculation of the integral

$$\int_{\Omega} X dP \left(= \int X \, dP \right) \tag{10.52}$$

can be done using the induced measure μ on $(\mathbb{R}, \beta)$, defined by relation (10.42) and then using the d.f. F of X.

Indeed, we can write

$$E(X) \left(= \int_{\Omega} X dP \right) = \int_{R} X \, d\mu, \tag{10.53}$$

and if F_X is the d.f. of X, it can be shown that

$$E(X) = \int_{R} x dF_X(x), \tag{10.54}$$

this last integral being a Lebesgue Stieltjes integral.

Moreover, if F_X is absolutely continuous with f_X as the density, we obtain

$$E(X) = \int_{-\infty}^{+\infty} x f_x(x) dx. \tag{10.55}$$

If g is a Borel function, we also have (see for example Chung (2000), Royden (1963), Loeve (1963))

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) dF_X \tag{10.56}$$

and with a density for X

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx . \tag{10.57}$$

The most important properties of the expectation are given in the next proposition.

Proposition 10.1

(i) *Linearity property of the expectation: if X and Y are two integrable r.v. and a, b two real numbers, then the r.v. aX+bY is also integrable and*

$$E(aX + bY) = aE(X) + bE(Y). \tag{10.58}$$

(ii) *If* $(A_n, n \geq 1)$ *is a partition of* Ω, *then*

$$E(X) = \sum_{n=1}^{\infty} \int_{A_n} X dP . \tag{10.59}$$

(iii) *The expectation of a non-negative r.v. is non-negative.*

(iv) *If X and Y are integrable r.v., then*

$$X \leq Y \Rightarrow E(X) \leq E(Y). \tag{10.60}$$

(v) *If X is integrable, then so is* $|X|$ *and*

$$|E(X)| \leq E|X| . \tag{10.61}$$

(vi) *Dominated convergence theorem (Lebesgue): if* $(X_n, n \geq 1)$ *is a sequence of r.v. converging a.s. to the integrable r.v. X, then all r.v.* X_n *are integrable and moreover*

$$\lim E(X_n) = E(\lim X_n) \ (= E(X)) . \tag{10.62}$$

(vii) *Monotone convergence theorem (Lebesgue): if* $(X_n, n \geq 1)$ *is a non-decreasing sequence of non-negative r.v, then relation* (10.62) *is still true provided that* $+\infty$ *is a possible value for each member.*

(viii) *If the sequence of integrable r.v.* $(X_n, n \geq 1)$ *is such that*

$$\sum_{n=1}^{\infty} E\left(\left|X_n\right|\right) < \infty, \tag{10.63}$$

then the random series $\sum_{n=1}^{\infty} X_n$ *converges absolutely a.s. and moreover*

$$\sum_{n=1}^{\infty} E(X_n) = E\left(\sum_{n=1}^{\infty} X_n\right) \ (= E(X)), \tag{10.64}$$

where the r.v. is defined as the sum of the convergent series.

Given a r.v. *X*, *moments* are special cases of expectation.

Definition 10.8 *If a is a real number and r a positive real number, then the expectation*

$$E\left(\left|X - a\right|^r\right) \tag{10.65}$$

is called the absolute moment of X of order r, centered on a.

The moments are said to be centered moments of order *r* if $a=E(X)$. In particular, for $r=2$, we obtain the *variance* of *X* represented by σ^2 $(\text{var}(X))$,

$$\sigma^2 = E\left(\left|X - m\right|^2\right). \tag{10.66}$$

Remark 10.3 From the linearity of the expectation (see relation (10.58)), it is easy to prove that

$$\sigma^2 = E(X^2) - (E(X))^2, \tag{10.67}$$

and so

$$\sigma^2 \leq E(X^2), \tag{10.68}$$

and more generally, it can be proven that the variance is the smallest moment of order 2 regardless of what *a* is.

The last fundamental concept we will now introduce in this section is that of *stochastic independence* or, more simply, *independence*.

Definition 10.9 *The events* $A_1,...,A_n,(n>1)$ *are stochastically independent or independent if and only if*

$$\forall m=2,...,n,\forall n_k=1,...,n:n_1\neq n_2\neq\cdots\neq n_k:P\left(\bigcap_{k=1}^{m}A_{n_k}\right)=\prod_{k=1}^{m}P(A_{n_k}).\quad(10.69)$$

For *n=2*, relation (10.69) reduces to

$$P(A_1\cap A_2)=P(A_1)P(A_2). \tag{10.70}$$

Let us note that piecewise independence of the events $A_1,...,A_n,(n>1)$ does not necessarily imply the independence of these sets and thus does not imply the stochastic independence of these *n* events.

Definition 10.10

(i) *The n real r.vs.* $X_1,X_2,\ldots,X_n$ *defined on the probability space* $(\Omega,\Im,P)$ *are said to be stochastically independent, or simply independent, if and only if for any Borel sets* $B_1,B_2,\ldots,B_n$, *we have*

$$P\left(\bigcap_{k=1}^{n}\{\omega:X_k(\omega)\in B_k\}\right)=\prod_{k=1}^{n}P\left(\{\omega:X_k(\omega)\in B_k\}\right). \tag{10.71}$$

(ii) *For an infinite family of r.vs., independence means that the members of every finite subfamily are independent. It is clear that if* $X_1,X_2,\ldots,X_n$ *are independent, so are the r.vs.* $X_{i_1},...,X_{i_k}$ *with*

$$i_1\neq\cdots\neq i_k,\ i_k=1,...,n,k=2,...,n.$$

From relation (10.71), we find that

$$P(X_1\leq x_1,...,X_n\leq x_n)=P(X_1\leq x_1)\cdots P(X_n\leq x_n),\forall(x_1,...,x_n)\in\mathbb{R}^n. \tag{10.72}$$

If the functions $F_X,F_{X_1},...,F_{X_n}$ are the distribution functions of r.v. $X=(X_1,...,X_n),X_1,...,X_n$, we can write the preceding relation under the form

$$F_X(x_1,....,x_n)=F_{X_1}(x_1)\cdot\cdots\cdot F_{X_n}(x_n),\forall(x_1,...,x_n)\in\mathbb{R}^n. \tag{10.73}$$

It can be shown that this last condition is also sufficient for the independence of $X=(X_1,...,X_n), X_1,...,X_n$. If these d.f. have densities f_X, $f_{X_1},...,f_{X_n}$, relation (10.73) is equivalent to

$$f_X(x_1,\ldots,x_n)=f_{X_1}(x_1)\cdots f_{X_n}(x_n),\forall(x_1,...,x_n)\in\mathbb{R}^n. \tag{10.74}$$

In case of the integrability of *n* real r.vs. $X_1,X_2,\ldots,X_n$, a direct consequence of relation (10.72) is that we have a very important property for the expectation of the product of *n* independent r.vs.:

$$E\left(\prod_{k=1}^{n}X_k\right)=\prod_{k=1}^{n}E(X_k). \tag{10.75}$$

The notion of independence gives the possibility to prove the result called the *strong law of large numbers* which says that if $(X_n,n\geq 1)$ is a sequence of integrable independent and identically distributed r.vs., then

$$\frac{1}{n}\sum_{k=1}^{n}X_k\xrightarrow{a.s.}E(X). \tag{10.76}$$

The next section will present the most useful distribution functions for stochastic modeling.

10.5. Main distribution probabilities

Here we shall restrict ourselves to presenting the principal distribution probabilities related to real random variables.

10.5.1. *The binomial distribution*

Let us consider a random experiment *E* such that only two results are possible: a "success" (*S*) with probability *p* and a "failure" (*F*) with probability *q*=1-*p*. If *n* independent trials are made in exactly the same experimental environment, the total number of trials in which the event *S* occurs may be represented by a r.v. *X* whose distribution $(p_i,i=0,...,n)$ with

$$p_i=P(X=i),\ i=1,...,n \tag{10.77}$$

is called a *binomial distribution* with parameters (*n*,*p*).

From the basic axioms of probability theory previously stated, it is easy to prove that

$$p_i = \binom{n}{i} p^i q^{n-i},\ i = 0,...,n, \tag{10.78}$$

a result from which we get

$$E(X) = np, \text{var}(X) = npq. \tag{10.79}$$

The *characteristic function* and the *generating function,* when they exist, of X respectively defined by

$$\begin{aligned} \varphi_X(t) &= E(e^{itX}), \\ g_X(t) &= E(e^{tX}) \end{aligned} \tag{10.80}$$

are given by

$$\begin{aligned} \varphi_X(t) &= (pe^{it} + q)^n, \\ g_X(t) &= (pe^t + q)^n. \end{aligned} \tag{10.81}$$

This distribution is currently used in the financial model of Cox, Ross and Rubinstein (1979), developed in Chapter 5.

10.5.2. *The Poisson distribution*

If X is an r.v. with values in $\mathbb{N}$ such that the probability distribution is given by

$$P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}, i = 0,1,... \tag{10.82}$$

where λ is a strictly positive constant, then X is called a *Poisson variable with parameter* λ. This is one of the most important distributions for all applications. For example, if we consider an insurance company looking at the total number of claims in one year, this variable may often be considered as a Poisson variable.

Basic parameters of this Poisson distribution are given here:

$$E(X)=\lambda,\ \text{var}(X)=\lambda,$$
$$\varphi_X(t)=e^{\lambda(e^{it}-1)},\ g_X(t)=e^{\lambda(e^{t}-1)}. \tag{10.83}$$

A remarkable result is that the Poisson distribution is the limit of a binomial distribution of parameters (*n*,*p*) if n tends to $+\infty$ and *p* to 0, so that *np* converges to λ.

The Poisson distribution is often used for the occurrence of rare events, for example, in credit risk presented in Chapter 19.

10.5.3. *The normal (or Laplace Gauss) distribution*

The real r.v. *X* has a normal (or Laplace Gauss) distribution of parameters $(\mu,\sigma^2), \mu\in\mathbb{R}, \sigma^2>0$, if its density function is given by

$$f_X(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x\in\mathbb{R}. \tag{10.84}$$

From now on, we will use the notation $X \prec N(\mu,\sigma^2)$.

The main parameters of this distribution are

$$E(X)=\mu,\ \text{var}(X)=\sigma^2,$$
$$\varphi_X(t)=\exp\left(i\mu t-\frac{\sigma^2t^2}{2}\right),\ g_X(t)=\exp\left(\mu t+\frac{\sigma^2t^2}{2}\right). \tag{10.85}$$

If $\mu=0,\ \sigma^2=1$, the distribution of *X* is called a *reduced* or *standard normal distribution*. In fact, if *X* has a normal distribution $(\mu,\sigma^2), \mu\in R, \sigma^2>0$, then from (10.85), the reduced r.v. *Y* defined by

$$Y=\frac{X-\mu}{\sigma} \tag{10.86}$$

has a standard normal distribution with mean 0 and variance 1.

Let Φ be the distribution function of the standard normal distribution; it is possible to express the distribution function of any normal r.v. X with parameters $(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0$ as follows:

$$F_X(x) = P(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right). \tag{10.87}$$

Also, from the numerical point of view, it suffices to know numerical values for the standard distribution.

From relation (10.87), we also deduce that

$$f_X(x) = \frac{1}{\sigma}\Phi'\left(\frac{x-\mu}{\sigma}\right), \tag{10.88}$$

where of course from (10.84)

$$\Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{10.89}$$

From the definition of Φ, we have

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy, \; x \in \mathbb{R} \tag{10.90}$$

and so

$$\Phi(-x) = 1 - \Phi(x), x > 0, \tag{10.91}$$

and consequently, for X normally distributed with parameters (0,1), we obtain

$$P\left(|X| \le x\right) = \Phi(x) - \Phi(-x) = 2\Phi(x) - 1, \; x > 0. \tag{10.92}$$

In particular, let us mention the following numerical results:

$$\begin{aligned} &P\left(\left|X-m\right| \le \frac{2}{3}\sigma\right) = 0.4972(\approx 50\%), \\ &P\left(\left|X-m\right| \le \sigma\right) = 0.6826(\approx 68\%), \\ &P\left(\left|X-m\right| \le 2\sigma\right) = 0.9544(\approx 95\%), \\ &P\left(\left|X-m\right| \le 3\sigma\right) = 0.9974(\approx 99\%). \end{aligned} \tag{10.93}$$

Remark 10.4 *Numerical calculation of the d.f.* Φ

For applications in finance, for example the Black-Scholes (1973) model for option pricing (see Chapter 5), we will need the following numerical approximation method for calculating Φ *with seven decimal places* instead of the four given by the standard statistical tables:

$$\begin{aligned} &1)\ x > 0\ : \\ &\Phi(x) \approx 1 - \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} (b_1 c + \cdots + b_5 c^5), \\ &c = \frac{1}{1+px}, \\ &p = 0.2316419,\ b_1 = 0.319381530, \\ &b_2 = -0.356563782,\ b_3 = 1.781477937, \\ &b_4 = -1.821255978,\ b_5 = 1.330274429, \\ &2)\ x < 0\ : \\ &\Phi(x) = 1 - \Phi(-x). \end{aligned} \tag{10.94}$$

The normal distribution is one of the most commonly used distributions, by virtue of the *central limit theorem* which says that if $(X_n, n \ge 1)$ is a sequence of independent identically distributed (in short IID) r.vs. with mean *m* and variance σ^2, then the sequence of r.vs. defined by

$$\frac{S_n - nm}{\sigma\sqrt{n}} \tag{10.95}$$

with

$$S_n = X_1 + \cdots + X_n, \quad n > 0 \tag{10.96}$$

converges in law to a standard normal distribution.

This means that the sequence of the distribution functions of the variables defined by (10.93) converges to Φ.

This theorem was used by the Nobel Prize winner H. Markowitz (1959) to justify that the return of a diversified portfolio of assets has a normal distribution. As a particular case of the central limit theorem, let us mention *de Moivre's theorem*, starting with

$$X_n = \begin{cases} 1, \text{ with prob. } p, \\ 0, \text{ with prob. } 1-p, \end{cases} \tag{10.97}$$

so that, for each n, the r.v. defined by relation (10.94) has a binomial distribution with parameters (n,p).

By applying the central limit theorem, we obtain the following result:

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow[n\to+\infty]{law} N(0,1), \tag{10.98}$$

called de Moivre's result.

10.5.4. *The log-normal distribution*

Though the normal distribution is the most frequently used, it is nevertheless true that it could not be used for example to model the time evolution of a financial asset like a share or a bond, as the minimal value of these assets is 0 and so the support of their d.f. is half of the real line $[0,+\infty)$. One possible solution is to consider the *truncated normal distribution*, defined by setting all the probability mass of the normal distribution on the negative half-real line on the positive one; however, then all the interesting properties of the normal distribution are lost.

Also, in order to have a better approach to some financial market data, we have to introduce the *log-normal distribution*. The real non-negative r.v. X has a

lognormal distribution with parameters μ,σ – which we will write as $X \prec LN(\mu,\sigma)$ – if the r.v. log*X* has a normal distribution with parameters μ,σ^2. Consequently, the density function of *X* is given by

$$f_X(x)=\begin{cases} 0, x \le 0, \\ \dfrac{1}{x\sqrt{2\pi}\sigma} e^{-\frac{(\log x-\mu)^2}{2\sigma^2}}, x>0. \end{cases} \tag{10.99}$$

Indeed, we can write

$$P(X \le x) = P(\log X \le \log x), \tag{10.100}$$

and so

$$F_X(x)=\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\log x} e^{\frac{(t-\mu)^2}{2\sigma^2}} dt \left(= \Phi\left(\frac{\log x-\mu}{\sigma}\right)\right), \tag{10.101}$$

and after the change of variable $t=\log x$, we obtain relation (10.99).

Let us note that relation (10.101) is the most useful for the calculation of the d.f. of *X* with the help of the normal d.f.

For the density function, we can also write

$$f_X(x)=\frac{1}{\sigma x}\Phi\left(\frac{\log x-\mu}{\sigma}\right). \tag{10.102}$$

The basic parameters of this distribution are given by

$$\begin{aligned} &E(X)=e^{\mu+\frac{\sigma^2}{2}}, \\ &\operatorname{var}(X)=e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right), \\ &E\left(X^r\right)=e^{r\left(\mu+\frac{\sigma^2}{2}\right)}. \end{aligned} \tag{10.103}$$

Let us mention that the lognormal distribution has no generating function and that the characteristic function has no explicit form. When $\sigma < 0.3$, some authors recommend a normal approximation with parameters (μ, σ^2).

The normal distribution is *stable* under the addition of independent r.vs.; this property means that the sum of *n* independent normal r.vs. is still normal. That is no longer the case with the lognormal distribution which is stable under *multiplication*, which means that for two independent lognormal r.vs. X_1,X_2, we have

$$X_i \prec LN(\mu_i, \sigma_i), i = 1,2 \Rightarrow X_1 \times X_2 \prec LN\left(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right). \quad (10.104)$$

10.5.5. *The negative exponential distribution*

The non-negative r.v. *X* has a *negative exponential distribution* (or simply *exponential distribution*) of parameter λ if its density function is given by

$$f_X(x) = \lambda e^{-\lambda x}, x \geq 0, \quad (10.105)$$

where λ is a strictly positive real number.

By integration, we obtain the explicit form of the exponential distribution function

$$F_X(x) = 1 - e^{-\lambda x}, x \geq 0. \quad (10.106)$$

Of course, F_X is zero for negative values of *x*.

The basic parameters are

$$E(X) = \frac{1}{\lambda}, \ \text{var}\, X = \frac{1}{\lambda^2},$$
$$\varphi_X(t) = \frac{1}{1 - i\frac{t}{\lambda}}, \ g_X(t) = \frac{1}{1 - \frac{t}{\lambda}}, t < \lambda. \quad (10.107)$$

In fact, this distribution is the first to be used in reliability theory.

10.5.6. *The multidimensional normal distribution*

Let us consider an *n*-dimensional real r.v. X represented as a column vector of its *n* components $X = (X_1,..., X_n)'$. Its d.f. is given by:

$$F_X(x_1,..., x_n) = P(X_1 \le x_1,..., X_n \le x_n). \tag{10.108}$$

If the density function of X exists, the relations between the d.f. and the density function are:

$$\begin{aligned} f_X(x_1,..., x_n) &= \frac{\partial^n F_X}{\partial x_1 ... \partial x_n}(x_1,..., x_n), \\ F_X(x_1,..., x_n) &= \int_{-\infty}^{x_1} ... \int_{-\infty}^{x_n} f_X(\xi_1,..., \xi_n) d\xi_1,..., d\xi_n. \end{aligned} \tag{10.109}$$

For the principal parameters we will use the following notation:

$$\begin{aligned} &E(X_k) = \mu_k, k = 1,..., n, \\ &E((X_k - \mu_k)(X_l - \mu_l)) = \sigma_{kl}, k, l = 1,..., n, \\ &E((X_k - \mu_k))^2 = \sigma_k^2, k = 1,..., n, \\ &\rho_{kl} = \frac{E((X_k - \mu_k)(X_l - \mu_l))}{\sqrt{E((X_k - \mu_k)^2) E((X_k - \mu_k)^2)}} \left(= \frac{\sigma_{kl}}{\sigma_k \sigma_l} \right), k, l = 1,..., n. \end{aligned} \tag{10.110}$$

The parameters σ_{kl} are called the *covariances* between the r.v. X_k and X_l, and the parameters ρ_{kl}, the *correlation coefficients* between the r.v. X_k and X_l.

It is well known that the correlation coefficient ρ_{kl} measures a certain linear dependence between the two r.v. X_k and X_l. More precisely, if it is equal to 0, there is no such dependence and the two variables are called *uncorrelated*; for the values +1 and –1 this dependence is certain.

With matrix notation, the following $n \times n$ matrix

$$\mathbf{\Sigma}_X = \left[\sigma_{ij} \right] \tag{10.111}$$

is called the *variance-covariance* matrix of X.

The *characteristic function* of X is defined as:

$$\varphi_X(t_1,...,t_n) = E\left(e^{i(t_1X_1+..+t_nX_n)_1}\right)\left(= E\left(e^{i\mathbf{t}'X}\right)\right). \tag{10.112}$$

Let $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ be an n-dimensional real vector and an $n \times n$ positive definite matrix, respectively. The n-dimensional real r.v. X has a *non-degenerated n-dimensional normal distribution* with parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ if its density function is given by:

$$f_X(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det \boldsymbol{\Sigma}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \mathbf{x} \in \mathbb{R}^n. \tag{10.113}$$

Then, it can be shown by integration that parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ are indeed respectively the *mean vector* and the *variance-covariance matrix* of X.

As usual, we will use the notation: $X \prec N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The characteristic function of X is given by:

$$\varphi_X(\mathbf{t}) = e^{i\boldsymbol{\mu}'\mathbf{t}-\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}. \tag{10.114}$$

The main fundamental properties of the n-dimensional normal distribution are:

– every subset of k r.vs. of the set $\{X_1,\ldots,X_n\}$ also has a k-dimensional distribution which is also normal;

– the multi-dimensional normal distribution is *stable* under linear transformations of X;

– the multi-dimensional normal distribution is *stable* for addition of r.vs., which means that if $X_k \prec N_n(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), k = 1,...,m$ and if these m random vectors are independent, then

$$X_1 + \cdots + X_m \prec N_n(\boldsymbol{\mu}_1 + \cdots + \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_1 + \cdots + \boldsymbol{\Sigma}_m). \tag{10.115}$$

Particular case: *the two-dimensional normal distribution*

In this case, we have:

$$\begin{aligned}&\boldsymbol{\mu}=(\mu_1,\mu_2)',\boldsymbol{\Sigma}=\begin{bmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{21} & \sigma_2^2\end{bmatrix},\rho=\frac{\sigma_{12}}{\sigma_1\sigma_2},\\ &\boldsymbol{\Sigma}^{-1}=\begin{bmatrix}\dfrac{1}{\sigma_1^2} & -\dfrac{\rho}{\varsigma_1\sigma_2}\\ -\dfrac{\rho}{\varsigma_1\sigma_2} & \dfrac{1}{\sigma_2^2}\end{bmatrix},\det\boldsymbol{\Sigma}=\sigma_1\sigma_2\sqrt{1-\rho^2}.\end{aligned} \tag{10.116}$$

From the first main fundamental properties of the *n*-dimensional normal distribution given above, we have:

$$X_k \prec N_1(\mu_k,\sigma_k^2), k=1,2. \tag{10.117}$$

For the special degenerated case of $|\rho|=1$, it can be proved that:

$$\begin{aligned}&\rho=1:\ \frac{X_2-\mu_2}{\sigma_2}=\frac{X_1-\mu_1}{\sigma_1},\\ &\rho=-1:\ \frac{X_2-\mu_2}{\sigma_2}=-\frac{X_1-\mu_1}{\sigma_1},\end{aligned} \tag{10.118}$$

meaning that in this case, all the probability mass in the plane lies on a straight line so the two r.vs. X_1,X_2 are perfectly dependent with probability 1.

To conclude this section, let us recall the well-known property stating that two independent r.vs. are uncorrelated, but the converse is not true except for the normal distribution.

10.6. Conditioning

Let us begin to briefly recall the concept of *conditional probability*. Let $(\Omega,\Im,P)$ be a probability space and let *A*, *B* be elements of $\Im$, and let us observe the number of occurrences of event *A* whenever *B* has already been observed in a sequence of *n* trials of our experiment. We shall call this number $n(A|B)$.

In terms of the frequency of events defined by relation (10.11), we have:

$$n\left(A\middle|B\right)=\frac{n(A\cap B)}{n(B)}, \tag{10.119}$$

provided that $n(B)$ is not 0.

Dividing by n the two members of relation (10.119), we obtain:

$$\frac{n\left(A\middle|B\right)}{n}=\frac{\frac{n(A\cap B)}{n}}{\frac{n(B)}{n}}. \tag{10.120}$$

In terms of frequencies, we obtain:

$$f\left(A\middle|B\right)=\frac{f(A\cap B)}{f(B)}. \tag{10.121}$$

From the experimental interpretation of the concept of probability of an event seen in section 10.2, we can now define the *conditional probability of A given B* as:

$$P\left(A\middle|B\right)=\frac{P(A\cap B)}{P(B)}, P(B)>0. \tag{10.122}$$

If events A and B are independent, from relation (10.122), we obtain:

$$P\left(A\middle|B\right)=P(A), \tag{10.123}$$

meaning that, in the case of independence, the conditional probability of set A does not depend on the given set B.

As the independence of sets A and B is equivalent to the independence of sets A and B^c, we also have:

$$P\left(A\middle|B^c\right)=P(A). \tag{10.124}$$

The notion of conditional probability is very useful for calculating probabilities of a product of *dependent* events A and B not satisfying relation (4.39). Indeed, from relations (10.122) and (10.124), we can write:

$$P(A \cap B) = P(A)P\left(A \middle| B\right) = P(B)P\left(B \middle| A\right). \tag{10.125}$$

More generally, for n events $A_1,\ldots,A_n$, we obtain the "theorem of compound probability":

$$P\left(\bigcap_{k=1}^{n} A_k\right) = P(A_1)P\left(A_2 \middle| A_1\right)\cdots P\left(A_n \middle| A_1 \cap A_2 \ldots \cap A_{n-1}\right), \tag{10.126}$$

a relation expanding relation (10.125).

$$P\left(\bigcap_{k=1}^{n} A_k\right) = P(A_1)\ldots P(A_n) \tag{10.127}$$

is true in the case of the independence of the n considered events.

If event B is fixed and of strictly positive probability, relation (10.122) provides the way of defining a *new* probability measure on $(\Omega, \Im)$ denoted P_B as follows:

$$P_B(A) = \frac{P(A \cap B)}{P(B)}, \forall A \in \Im. \tag{10.128}$$

P_B is in fact a probability measure as it is easy to verify that it satisfies conditions (10.16) and (10.17), and so P_B is called the *conditional probability measure given B.*

The integral with respect to this measure is called the conditional expectation E_B relative to P_B.

From relation (10.128) and since $P_B(B)=1$, we thus obtain for any integrable r.v. Y:

$$E_B(Y) = \int_{\Omega} Y(\omega) dP_B = \frac{1}{P(B)} \int_{B} Y(\omega) dP. \tag{10.129}$$

We can now extend this definition to *arbitrary* sub-σ-algebras instead of the simple case of $\left\{\varnothing, B, B^c, \Omega\right\}$ using an extension of property (10.129) as a definition with the help of the Radon Nikodym theorem (Halmos (1974)).

Definition 10.11 *If* $\Im_1$ *is a sub-*σ*-algebra of* $\Im$*, the conditional expectation of the integrable r.v. Y given* $\Im_1$*, denoted by* $E_{\Im_1}(Y)$ *or* $E\left(Y\middle|\Im_1\right)$*, is any r.v. of the equivalence class such that:*

(i) $E_{\Im_1}(Y)$ is $\Im_1$*-measurable,*

(ii) $\int_B E_{\Im_1}(Y)(\omega)dP = \int_B Y(\omega)dP, B \in \Im_1.$ (10.130)

In fact, the class of equivalence contains all the r.vs. a.s. equally satisfying relation (10.130).

Remark 10.5 Taking $B = \Omega$ in relation (10.130), we obtain:

$$E(E_{\Im_1}Y)) = E(Y). \tag{10.131}$$

Particular cases

(i) $\Im_1$ *is generated by r.v. X.*

This case means that $\Im_1$ is the sub-σ-algebra of $\Im$ generated by all the inverse images of *X*, and we will use as notation:

$$E_{\Im_1}(Y) = E\left(Y\middle|X\right), \tag{10.132}$$

where this conditional expectation is called the *conditional expectation of Y given X.*

(ii) $\Im_1$ *is generated by n r.vs.* $X_1,...,X_n$.

This case means that $\Im_1$ is the sub-σ-algebra of $\Im$ generated by all the inverse images of $X_1,...,X_n$ and we will use as notation:

$$E_{\Im_1}(Y) = E(Y\middle|X_1,...,X_n), \tag{10.133}$$

where this conditional expectation is called the *conditional expectation of Y given* $X_1,...,X_n$.

In this latter case, it can be shown (Loeve (1977)) that there exists a version $\varphi(X_1,...,X_n)$ of the conditional expectation so that φ is a Borel function from $\mathbb{R}^n$ to $\mathbb{R}$, and as such it follows that $E(Y\middle|X_1,...,X_n)$ is constant on each set belonging to $\Im_1$ for which $X_1(\omega) = x_1,...,X_n(\omega) = x_n$, for instance.

This justifies the abuse of notation

$$E\left(Y\left|X_1(\omega)=x_1,...,X_n(\omega)=x_n\right.\right)=\varphi(x_1,...,x_n) \tag{10.134}$$

representing the value of this conditional expectation on all the ωs belonging to the set $\left\{\omega: X_1(\omega)=x_1,...,X_n(\omega)=x_n\right\}$.

Taking $B=\Omega$ in relation (10.130), we obtain:

$$\begin{aligned}&E(Y)\\&=\int_{R_n} E\left(Y\left|X_1(\omega)=x_1,...,X_n(\omega)=x_n\right.\right)dP(X_1(\omega)\le x_1,...,X_n(\omega)\le x_n)\end{aligned} \tag{10.135}$$

a result often used in the sequel to evaluate the mean of an r.v. using its conditional expectation with respect to some given event.

(iii) If $\Im_1=\{\varnothing,\Omega\}$, we obtain $E(Y|\Im_1)=E(Y)$ and if $\Im_1=\{\varnothing,B,B^c,\Omega\}$, then $E(Y|\Im_1)=E(Y|B)$ on B and $E(Y|\Im_1)=E(Y|B^c)$ on B^c.

(iv) Taking r.v. Y as the indicator of the event A, that is to say:

$$1_{A(\omega)}=\begin{cases}1,\omega\in A,\\0,\omega\notin A,\end{cases} \tag{10.136}$$

the conditional expectation becomes the *conditional probability of A given* $\Im_1$ denoted as follows:

$$P(A|\Im_1)=E(1_A(\omega)|\Im_1) \tag{10.137}$$

and then relation (10.130) becomes:

$$\int_B P\left(A\left|\Im_1(\omega)\right.\right)dP=P(A\cap B), B\in\Im_1. \tag{10.138}$$

Letting $B=\Omega$ in this final relation, we obtain:

$$E\left(P\left(A|\Im_1\right)\right)=P(A), \tag{10.139}$$

a property extending the *theorem of total probability* .

If, moreover, A is independent of $\Im_1$, that is to say, if for all B belonging to $\Im_1$:

$$P(A \cap B) = P(A)P(B), \tag{10.140}$$

then we see from relation (10.137) that:

$$P\left(A \middle| \Im_1\right)(\omega) = P(A), \omega \in \Omega. \tag{10.141}$$

Similarly, if r.v. Y is independent of $\Im_1$, that is to say if for each event B belonging to $\Im_1$ and each set A belonging to the σ-algebra generated by the inverse images of Y, denoted by $\sigma(Y)$, relation (10.140) is true, then from relation (10.130), we have:

$$E\left(Y \middle| \Im_1\right) = E(Y). \tag{10.142}$$

Indeed, from relation (10.140), we can write that:

$$\begin{aligned} \int_B E_{\Im_1}(Y)(\omega)dP &= \int_B Y(\omega)dP, B \in \Im_1, \\ &= E\left(Y1_B\right), \\ &= E(Y)P(B), \\ &= \int_B E(Y)dP, \end{aligned} \tag{10.143}$$

and so, relation (10.142) is proved.

In particular, if $\Im_1$ is generated by the r.vs. $X_1,\ldots,X_n$, then the independence between Y and $\Im_1$ implies that:

$$E\left(Y \middle| X_1, ..., X_n\right) = E(Y). \tag{10.144}$$

Relations (10.142) and (10.144) allow us to have a better understanding of the *intuitive meaning of conditioning* and its importance in finance.

Under independence assumptions, conditioning has absolutely no impact, for example, on the expectation or the probability; on the contrary, dependence implies that the results with or without conditioning will be different, meaning that we can interpret conditioning as given *additional information* useful to obtain more precise results in the case of dependence of an asset.

The properties of expectation, stated in section 10.4, are also properties of conditional expectation, true a.s., but there are supplementary properties which are very important in stochastic modeling. They are given in the next proposition.

Proposition 10.2 (Supplementary properties of conditional expectation) *On the probability space* $(\Omega, \Im, P)$ *, we have the following properties:*

(i) *If r.v. X is* $\Im_1$*-measurable, then*

$$E(X \mid \Im_1) = X, a.s.. \tag{10.145}$$

(ii) *If X is a r.v. and Y* $\Im_1$*-measurable, then*

$$E(XY \mid \Im_1) = YE(X \mid \Im_1), a.s.. \tag{10.146}$$

This property means that $\Im_1$*-measurable r.vs. are like constants for the classical expectation.*

(iii) *Since from relation (10.145) we have* $E_{\Im}(Y) = Y$*, taking* $Y = E_{\Im_1}(Y)$ *, we see that:*

$$E_{\Im}(E_{\Im_1}(Y)) = E_{\Im_1}(Y) \tag{10.147}$$

and of course since:

$$E_{\Im_1}(E_{\Im}(Y)) = E_{\Im_1}(Y), \tag{10.148}$$

combining these last two relations, we obtain:

$$E_{\Im}(E_{\Im_1}(Y)) = E_{\Im_1}E_{\Im}((Y)) = E_{\Im_1}(Y). \tag{10.149}$$

This last result may be generalized as follows.

Proposition 10.3 (Smoothing property of conditional expectation) *Let* $\Im_1$*,* $\Im_2$ *be two sub-*σ*-algebras of* $\Im$ *such that* $\Im_1 \subset \Im_2$*; then it is true that:*

$$E_{\Im_2}(E_{\Im_1}(Y)) = E_{\Im_1}(E_{\Im_2}(Y)) = E_{\Im_1}(Y), \tag{10.150}$$

a property called the smoothing property in Loeve (*1977*).

A particular case of relation (10.150) is for example:

$$E\left(E\left(Y\middle|X_1,...,X_n\right)\middle|X_1\right)=E\left(E\left(Y\middle|X_1\right)\middle|X_1,...,X_n\right)=E\left(Y\middle|X_1\right). \quad (10.151)$$

This type of property is very useful for calculating probabilities using conditioning and will often be used in the following chapters.

Here is an example illustrating sums of a random number of r.vs. with the *Wald identities*.

Example 10.3 (Wald's identities) Let $(X_n, n \geq 1)$ be a sequence of IID real r.v.s and N a non-negative r.v. with integer values independent of the given sequence. The r.v. defined by:

$$S_N = \sum_{n=1}^{N} X_n \quad (10.152)$$

is called a *sum of a random number of random variables* and the problem to be solved is the calculation of the mean and the variance of this sum assuming that the r.vs. X_n have a variance.

From relation (10.150), we have:

$$E(S_N) = E\left(E\left(S_N\middle|N\right)\right) \quad (10.153)$$

and as, from the independence assumptions:

$$E\left(S_N\middle|N\right) = NE(X), \quad (10.154)$$

we also have:

$$E(S_N) = E(N)E(X), \quad (10.155)$$

called the *first Wald's identity*.

For the variance of S_N, it is possible to show that (see for example Janssen and Manca (2007))

$$\text{var}(S_N) = E(N)\,\text{var}(X) + \text{var}(N)(E(X))^2 \quad (10.156)$$

called the *second Wald's identity*.

In the particular case of an *n*-dimensional real r.v. $X=(X_1,\ldots,X_n)$, we can now introduce the very useful definition of the *conditional distribution function of X given* $\mathfrak{I}_1$ defined as follows:

$$\begin{aligned} F_{\mathfrak{I}_1}(x_1,\ldots,x_n,\omega) &= P\left(X_1 \le x_1,\ldots,X_n \le x_n \,\middle|\, \mathfrak{I}_1\right) \\ &= Q\left(\left\{\omega' : X_1(\omega') \le x_1,\ldots,X_n(\omega') \le x_n\right\},\omega\right). \end{aligned} \tag{10.157}$$

Another useful definition concerns an extension of the concept of the independence of random variables for the definition of *conditional independence of the n variables* $X_1,\ldots,X_n$. For all $(x_1,\ldots,x_n)$ belonging to $\mathbb{R}^n$, we have the following identity:

$$F_{\mathfrak{I}_1}(x_1,\ldots,x_n,\omega) = \prod_{k=1}^{n} F_{\mathfrak{I}}(x_k,\omega), \tag{10.158}$$

where of course we have:

$$F_{\mathfrak{I}}(x_k,\omega) = P\left(X_k \le x_k \,\middle|\, \mathfrak{I}_1\right) \tag{10.159}$$

according to definition (10.157) with $n=1$.

Example 10.4 On the probability space $(\Omega,\mathfrak{I},P)$, let (X,Y) be a two-dimensional real r.v. whose d.f. is given by

$$F(x,y) = P(X \le x, Y \le y). \tag{10.160}$$

As $\mathbb{R}^2$ is a complete separable metric space, there exist regular conditional probabilities given the sub-σ-algebras $\sigma(X)$ or $\sigma(Y)$, and so the related conditional d.f. denoted by:

$$F_{X|Y}\left(x \,\middle|\, \{\omega : Y=y\}\right),\ F_{Y|X}\left(y \,\middle|\, \{\omega : X=x\}\right) \tag{10.161}$$

also exists.

If, moreover, the d.f. F has a density f, we can also introduce the concept of *conditional density* for functions $F_{X|Y}$, $F_{Y|X}$ and F_X, giving at the same time an intuitive interpretation of conditioning in this special case.

We know that for every fixed (x,y):

$$f(x,y)\Delta x\Delta y + o(x,y,\Delta x,\Delta y) = P(x < X \le x+\Delta x, y < Y \le y+\Delta y), \quad (10.162)$$

where $o(x,y,\Delta x,\Delta y) \to 0$ for $(\Delta x,\Delta y) \to (0,0)$, and similarly for the marginal density function of X:

$$f_X(x)\Delta x + \bar{o}(x,\Delta x) = P(x < X \le x+\Delta x), \quad (10.163)$$

where $\bar{o}(x,\Delta x) \to 0$ for $\Delta x \to 0$ with of course:

$$f_X(x) = \int_R f(x,y)dy\,. \quad (10.164)$$

Using formula (10.122), we thus obtain:

$$P\left(y < Y \le y+\Delta y \middle| x < X \le x+\Delta x\right) = \frac{f(x,y)\Delta x\Delta y + o(x,y,\Delta x,\Delta y)}{f_X(x)\Delta x + \bar{o}(x,\Delta x)}. \quad (10.165)$$

Letting Δx tend to 0, we obtain:

$$\lim_{\Delta x\to 0} P\left(y < Y \le y+\Delta y \middle| x < X \le x+\Delta x\right) = \frac{f(x,y)}{f_X(x)}\Delta y\,. \quad (10.166)$$

This relation shows that the function $f_{Y|X}$ defined by:

$$f_{Y|X}\left(y\middle|x\right) = \frac{f(x,y)}{f_X(x)} \quad (10.167)$$

is the *conditional density of Y, given X*. Similarly, the *conditional density of X, given Y* is given by:

$$f_{X|Y}\left(x\middle|y\right) = \frac{f(x,y)}{f_Y(y)}\,. \quad (10.168)$$

Consequently, for any Borel subsets A and B of $\mathbb{R}$, we have:

$$\begin{aligned} P\left(X \in A \middle| Y(\omega) = y\right) &= \int_A f_{X|Y}\left(x \middle| y\right) dx = \frac{1}{f_Y(y)} \int_A f(x,y) dx, \\ P((X,Y) \in A \cap B) &= \int_{A \cap B} f(x,y) dx dy = \int_B \left(\int_A f_{X|Y}\left(x \middle| y\right) dx \right) f_Y(y) dy. \end{aligned} \tag{10.169}$$

The last equalities show that the density of (X,Y) can also be characterized by one marginal d.f. and the associated conditional density, as from relations (10.166) and (10.169):

$$f = f_X \times f_{Y|X} = f_Y \times f_{X|Y}. \tag{10.170}$$

It is possible that *conditional means* exist; if so, they are given by the following relations:

$$E\left(X \middle| Y = y\right) = \int_{\mathbb{R}} f\left(x \middle| y\right) dx,\ E\left(Y \middle| X = x\right) = \int_{\mathbb{R}} f\left(y \middle| x\right) dy. \tag{10.171}$$

The conditional mean of X (respectively Y) given $Y=y$ (respectively $X=x$) can be seen as a function of the real variable y (respectively x) called *the regression curve of X* (*respectively Y*) *given Y* (*respectively X*).

The two regression curves will generally not coincide and not be straight lines except if the two r.vs. X and Y are independent because, in this case, we obtain from relations (10.166) and (10.168) that:

$$f_{X|Y} = f_X,\ f_{Y|X} = f_Y \tag{10.172}$$

and so:

$$E\left(X \middle| Y\right) = E(X),\ E\left(Y \middle| X\right) = E(Y), \tag{10.173}$$

proving that the two regression curves are straight lines parallel to the axes passing through the "center of gravity" $(E(X), E(Y))$ of the probability mass in $\mathbb{R}^2$.

In the special case of a non-degenerated normal distribution for (X,Y) with vector mean (m_1,m_2) and variance covariance matrix:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}, \tag{10.174}$$

it can be shown that the two conditional distributions are also normal with parameters:

$$\begin{aligned} Y \,|\, X \prec N_2 \left(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1), \sigma_2^2 (1 - \rho^2) \right), \\ X \,|\, Y \prec N_2 \left(\mu_1 + \frac{1}{\rho} \frac{\sigma_1}{\sigma_2}(y - \mu_2), \sigma_2^2 (1 - \rho^2) \right). \end{aligned} \tag{10.175}$$

Thus, the two regression curves are linear.

10.7. Stochastic processes

In this section, we shall always consider a *complete* probability space $(\Omega, \Im, \mathrm{P})$ with a *filtration F*.

Let us recall that a probability space $(\Omega, \Im, \mathrm{P})$ is *complete* if every subset of an event of probability 0 is measurable, i.e. in the σ-algebra $\Im$, and so also of probability 0.

Definition 10.12 *F is a filtration on the considered basic probability space if F is a family of* $(\Im_t, t \in T)$ *of sub-*σ*-algebras of* $\Im$*, the index set T being either the natural set* $\{0,1,...,n,...\}$ *or the positive half real line* $[0, \infty)$ *such that:*

$$\begin{aligned} &\text{(i)} \quad s < t \Rightarrow \Im_s \subset \Im_t, \\ &\text{(ii)} \quad \Im_t = \bigcap_{u > t} \Im_u, \\ &\text{(iii)} \quad \Im_0 \textit{ contains all subsets with probability } 0. \end{aligned} \tag{10.176}$$

Assumption (ii) is called the *right continuity property* of filtration *F*.

Any filtration satisfying these three assumptions is called a filtration *satisfying the usual assumptions.*

The concept of filtration can be interpreted as a family of amounts of information so that $\Im_t$ gives all the observable events at time *t*.

Definition 10.13 *The quadruplet* $\left(\left(\Omega,\Im,\mathrm{P},\left(\Im_t,t\in T\right)\right)\right)$ *is called a filtered probability space.*

Definition 10.14 *A r.v.* $\tau:\Omega\mapsto T$ *is a stopping time if:*

$$\forall t\in T:\ \left\{\omega:\tau(\omega)\le t\right\}\in\Im_t. \tag{10.177}$$

The interpretation is the following: the available information at time *t* allows for the possibility to observe the event given in (10.177) and to decide for example if the future observations will be stopped after time *t*, or not.

We have the following proposition:

Proposition 10.4 *The r.v.* τ *is a stopping time if and only if*

$$\left\{\omega:\tau(\omega)<t\right\}\in\Im_t,\ \forall t\in T. \tag{10.178}$$

Definition 10.5 *A stochastic process (or simply process) with values in the measurable space* $\left(E,\aleph\right)$ *is a family of r.vs.:*

$$\left\{X_t,t\in T\right\} \tag{10.179}$$

where for all t:

$$X_t:\Omega\mapsto E,\ \left(\Im,\aleph\right)\text{-measurable.}$$

This means, in particular, that for every subset *B* of the σ-algebra $\aleph$, the set

$$X_t^{-1}\left(B\right)=\left\{\omega:X_t(\omega)\in B\right\} \tag{10.180}$$

belongs to the σ-algebra $\Im$.

Remark 10.6 If $\left(E,\aleph\right)=\left(\mathbb{R},\beta\right)$, the process is called a *real stochastic process* with values in $\mathbb{R}$; if $\left(E,\aleph\right)=\left(\mathbb{R}^n,\beta^n\right)$, it is called a *real multidimensional process* with values in $\mathbb{R}^n$.

If *T* is the natural set $\left\{0,1,...,n,...\right\}$, the process *X* is called a *discrete time stochastic process* or a *random sequence*; if *T* is the positive half of the real line $\left[0,\infty\right)$, the process *X* is called a *continuous time stochastic process.*

Definition 10.16 The stochastic process *x* is adapted to the filtration *f* if, for all *t*, the r.v. X_t is $\Im_t$-measurable. This means that, for all $t \in T$:

$$X_t^{-1}(B) = \{\omega : X_t(\omega) \in B\} \in \Im_t, \forall B \in \aleph. \tag{10.181}$$

Definition 10.17 *Two processes x and y are indistinguishable if a.s., for all* $t \in T$*:*

$$X_t = Y_t. \tag{10.182}$$

This means that:

$$\mathrm{P}(X_t = Y_t, \forall t \in T) = 1. \tag{10.183}$$

Definition 10.18 *The process X (or Y) is a modification of the process Y (or X) if a.s., for all* $t \in T$:

$$X_t = Y_t, \ a.s. \tag{10.184}$$

This means that:

$$\mathrm{P}(X_t = Y_t., \forall t \in T) = 1 \tag{10.185}$$

for all $t \in T$.

Definition 10.19 *For every stochastic process x, the function from t to e,*

$$t \mapsto X_t(\omega) \tag{10.186}$$

defined for each $\omega \in \Omega$*, is called a trajectory or sample path of the process.*

It must be clearly understood that the "modern" study of stochastic processes is concerned with the study of the properties of these trajectories.

For example, we can affirm that if two processes *X* and *Y* are indistinguishable, then there exists a set *N* belonging to $\Im$ of probability 0 such that:

$$\forall \omega \notin N : X_t(\omega) = Y_t(\omega), \forall t \in T. \tag{10.187}$$

In other words, for each ω element of the set $\Omega - N$, the two functions $t \mapsto X_t(\omega)$ and $t \mapsto Y_t(\omega)$ are equal.

As the basic probability space is complete, the neglected set *N* belongs to $\Im_t$, for all $t \in T$.

Definition 10.20 *A real stochastic process x is càdlàg if a.s. the trajectories of x are right continuous and have left limits at every point t.*

Definition 10.21 *If x is a real stochastic process and a set* $\Lambda \in \beta$, *then the r.v. defined by:*

$$T(\omega) = \inf\left\{t > 0 : X_t(\omega) \in \Lambda\right\} \tag{10.188}$$

is called the hitting time of Λ *by process X.*

It is easily shown that the *properties* of stopping and hitting times are as follows (see Protter (1990)):

(i) if *X* is càdlàg, adapted and $\Lambda \in \beta$, then the hitting time related to Λ is a stopping time;

(ii) if *S* and *T* are two stopping times, then the following r.v.:

$$S \wedge T \left(= \min\left\{S,T\right\}\right),\ S \vee T \left(= \max\left\{S,T\right\}\right),\ S+T,\ \alpha S (\alpha > 1) \tag{10.189}$$

are also stopping times.

Definition 10.22 *If T is a stopping time, the* σ*-algebra* $\Im_T$ *defined by:*

$$\Im_T = \left\{\Lambda \in \Im : \Lambda \bigcap \left\{\omega : T(\omega) \le t\right\} \in \Im_t, \forall t \ge 0\right\} \tag{10.190}$$

is called the stopping time σ*-algebra.*

In fact, the σ-algebra $\Im_T$ represents the information of all observable sets up to stopping time *T*. We can also say that $\Im_T$ is the smallest stopping time containing all the events related to the r.v. $X_{T(\omega)}(\omega)$ for all the adapted càdlàg processes *X* or generated by these r.v.

We also have for two stopping times *S* and *T*:

(i) $S \le T$ a.s. $\Rightarrow \Im_S \subset \Im_T$, (10.191)

(ii) $\Im_S \bigcap \Im_T, = \Im_{S \wedge T}$. (10.192)

10.8. Martingales

In this section, we shall briefly present some topics related to the most well-known category of stochastic processes called *martingales*.

Let X be a real stochastic process defined on the filtered complete probability space $\left(\Omega, \Im, P, \left(\Im_t, t \in T\right)\right)$.

Definition 10.23 *The process x is called a* $(\Im_t)$*-martingale if:*

$$\text{(i)}\ \forall t \geq 0, \exists E\left(X_t\right), \tag{10.193}$$

$$\text{(ii)}\ s < t \Rightarrow E\left(X_t \mid \Im_s\right) = X_s, \text{a.s.} \tag{10.194}$$

The latter equality is called the *martingale property* or the *martingale equality*.

Definition 10.24 *The process X is called a super-martingale (respectively sub-martingale) if:*

$$\text{(i)}\ \forall t \geq 0, \exists E\left(X_t\right), \tag{10.195}$$

$$\text{(ii)}\ s < t \Rightarrow E\left(X_t \mid \Im_s\right) \leq (\geq) X_s, \text{a.s.} \tag{10.196}$$

The martingale concept is interesting; indeed, as the best estimator at time s ($s>t$) for the value of X_t, as given by the conditional expectation appearing in relation (8.2), the martingale equality means that *the best predicted value* is simply the observed value of the process at the time of predicting s.

In finance the martingale is frequently used (see Janssen and Skiadas (1995)) to model the concept of an *efficient financial market*.

Definition 10.25 *The martingale X is closed if:*

$$\begin{aligned} &\exists\ Y\ : \\ &\text{(i)} \mathrm{E}\left(|Y|\right) < \infty, \\ &\text{(ii)} \forall t \in [0, \infty) : \mathrm{E}\left(Y \mid \Im_t\right) = X_t, \text{a.s.}. \end{aligned} \tag{10.197}$$

It is possible to prove the following result (see for example Protter (1990)).

Proposition 10.5

(i) *If X is a supermartingale, then the function* $t \mapsto E\left(X_t\right)$ *is right continuous if and only if there exists a unique modification Y of X such that Y is càdlàg.*

(ii) *If X is a martingale then, up to a modification, the function* $t \mapsto E\left(X_t\right)$ *is right continuous.*

It follows that every martingale, such that the function $t \mapsto E\left(X_t\right)$ is right continuous, is càdlàg.

The two most important results about martingales are *the martingale convergence theorem* and the *optional sampling (or Doob's) theorem.*

Before giving these results, we still need a final technical definition.

Definition 10.25 (Meyer (1966)) *A family* $\left(\xi_u, u \in A\right)$ *where A is an infinite index set is uniformly integrable if:*

$$\limsup_{n\to\infty} \sup_{\alpha} \int_{\{\omega;|\xi_\alpha(\omega)|\geq n\}} \left|\xi_\alpha(\omega)\right| d\mathrm{P}(\omega) = 0 . \tag{10.198}$$

Proposition 10.6 *Let x be a super-martingale in such a way that the function* $t \mapsto E\left(X_t\right)$ *is right continuous such that:*

$$\sup_{t\in[0,\infty)} E\left(\left|X_t\right|\right) < \infty ; \tag{10.199}$$

then, there exists a r.v. Y such that:

$$\begin{aligned} &\text{(i)} E\left(\left|Y\right|\right), \\ &\text{(ii)} Y = \lim_{t\to\infty} X_t , \text{ a.s.} \end{aligned} \tag{10.200}$$

Moreover, if X is a martingale closed by r.v. Z, then r.v. Y also closes X and:

$$Y = E\left(Z \middle| \Im_\infty\right), \tag{10.201}$$

where

$$\Im_\infty = \sigma\left(\bigcup_{0\le t<\infty} \Im_t \right). \tag{10.202}$$

With the aid of the concept of uniform integrability, we can obtain the following corollary.

Corollary 10.1

(i) *Let X be a right continuous martingale and uniformly integrable; then the following limit:*

$$Y = \lim_{t\to\infty} X_t \tag{10.203}$$

exists a.s.; moreover $Y \in L^1$ *and the r.v. Y closes the martingale X.*

(ii) *Let X be a right continuous martingale; then* $X = (X_t, t \ge 0)$ *is uniformly integrable if and only if*

$$Y = \lim_{t\to\infty} X_t \tag{10.204}$$

exists a.s., $Y \in L^1$, *and* $\left(X_t, t \in [0,\infty] \right)$ *is a martingale with, a.s.:*

$$X_\infty = Y. \tag{10.205}$$

Now, an interesting question is: what happens if we observe a martingale *X* at two stopping times *S,T* (*S*<*T*, a.s.)? The solution is given by the *optional sampling theorem*, also called *Doob's theorem.*

Proposition 10.7 (The optional sampling theorem or Doob's theorem) *Let X be a right continuous martingale closed by* X_∞ *and let S and T be two stopping times so that a.s. S < T; then the r.v.* $X_S, X_T \in L^1$ *and:*

$$X_S = E\left(X_T \middle| \Im_S \right), \text{ a.s.} \tag{10.206}$$

This important theorem means that if we restrict the random observation time set to $\{S,T\}$, then the restriction of the martingale to this set is *still* a martingale provided that *S* and *T* are two stopping times with of course *S*<*T*, a.s.

This result is interesting for the concept of *stopped process.*

Definition 10.26 *Let X be a stochastic process and T a stopping time. The stopped stochastic process* X^T *is defined by:*

$$X^T = \left(X_t^T, t \in [0, \infty] \right) \tag{10.207}$$

where:

$$\begin{aligned} &X_t^T(\omega) = X_{t \wedge T}(\omega), \\ &\text{with } t \wedge T = \inf\{t, T\}. \end{aligned} \tag{10.208}$$

From this definition, it follows that if process *X* is adapted and càdlàg, then so is the stopped process X^T. This is due to the fact that $t \wedge T$ is also a stopping time and moreover:

$$X_t^T = X_t 1_{\{t<T\}} + X_T 1_{\{t \geq T\}}. \tag{10.209}$$

This leads to the last result we want to mention.

Proposition 10.8 *Let x be a right continuous uniformly integrable martingale; then the stopped process* $X^T = \left(X_{t \wedge T}, t \in [0, \infty] \right)$ *has the same properties with respect to the filtration* $\left(\Im_t, t \in [0, \infty] \right)$.

10.9. Brownian motion

There are many types of stochastic process and some of them will be extensively studied in the following chapters, such as renewal processes, random walks, Markov chains, semi-Markov and Markov processes and their main extensions.

Figure 10.1 shows a typical sample path for models in finance.

To obtain such trajectories, it is necessary to introduce a specific stochastic process called the Brownian motion.

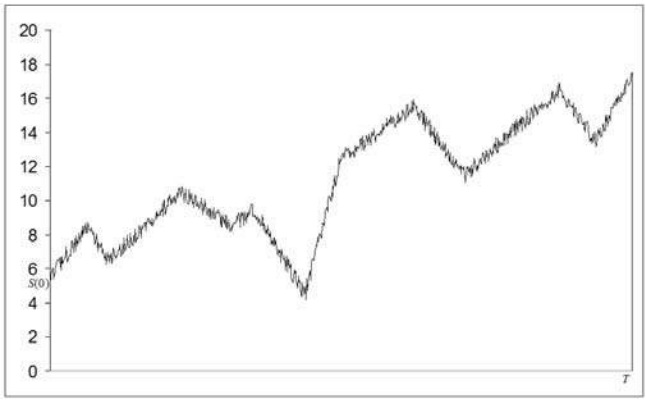

Figure 10.1. *Sample of a Brownian motion*

We will work on a basic complete filtered probability space satisfying the usual assumptions and noted $\left(\Omega, \Im, P, \left(\Im_t, t \in [0,\infty)\right)\right)$.

Definition 10.27 *The real stochastic process* $B = \left(B_t, t \in [0,\infty)\right)$ *will be called a Brownian motion or Brownian or Wiener process with trend* μ *and variance* σ^2 *provided that:*

(i) *B is adapted to the basic filtration,*

(ii) *B has independent increments, i.e. that:*

$$\forall s,t\ (0 \le s < t)\ :\ P\left(B_t - B_s \in A \middle| \Im_S\right) = P\left(B_t - B_s \in A\right), \quad \forall \textit{Borel set } B, \tag{10.210}$$

(iii) *B* has *stationary increments, i.e.*:

$$\forall s,t\ (0 \le s < t)\colon\ B_t - B_s \textit{ has a normal distribution } N(\mu(t-s), \sigma^2(t-s)), \tag{10.211}$$

(iv) $P\left(B_0 = x\right) = 1,\ (x \in \mathbb{R}).$ (10.212)

If, moreover, we have:

$$\mu = 0,\ \sigma^2 = 1,\ x = 0, \tag{10.213}$$

then the Brownian motion is said to be *standard*.

Let us now give the most important properties of the standard Brownian motion.

Property 10.2 *If B is a Brownian motion, then there exists a modification of B, the process B*, such that B* has, a.s., continuous trajectories.*

Property10.3 *If B is a standard Brownian motion, then B is a martingale*.

Property 10.4 *If B is a standard Brownian motion, then the process Q where*

$$Q = \left(B_t^2 - t, t \in [0, \infty) \right) \tag{10.214}$$

is a martingale.

Remark 10.8 It can also be proved that both Properties 10.3 and 10.4 characterize a standard Brownian motion.

Property 10.5 *If B is a standard Brownian motion, then for almost all* ω*, the trajectory* $\omega \mapsto B_t(\omega)$ *is not of bounded variation on every closed interval* $[a, b]$.

This explains why it is necessary for models in finance and in insurance to define a new type of integral, called the *Itô* or *stochastic integral*, if we want to integrate with respect to *B* (see for example Protter (1990)). This will be done in section 13.3.