AN ELEMENTARY COURSE ON STOCHASTIC PROCESSES

Marta Sanz-Solé
Facultat de Matemàtiques
Universitat de Barcelona

March 26, 2010

Contents

1 Review of Basic Notions and Definitions in Probability Theory

This chapter consists of a short survey of the basic tools in Probability theory used throughout the course. Those readers wishing to have a deeper insight on the notions and results quoted here are referred to [3], [9], [12]. We do not present here any result on convergence of random variables. There are squeezed in other chapters when they are needed.

1.1 Probability Spaces

The nature of random experiences is such that one cannot predict exactly their result. The main futures of this type of phenomena are captured by a mathematical object called *probability space*. It consists of a triple $(\Omega, \mathcal{F}, P)$ where

(a) Ω is the set of possible outcomes of the experience. Is is termed the *sample space*.

(b) $\mathcal{F}$ is a subset of $\mathcal{P}(\Omega)$, the set of all subsets of Ω, with σ-field structure, that is,

1. $\Omega \in \mathcal{F}$,
2. If $A \in \mathcal{F}$ then also $A^c \in \mathcal{F}$,
3. $\mathcal{F}$ is closed under countable union. That means, if $(A_n, n \geq 1)$ is a sequence of sets in $\mathcal{F}$, then $\cup_{n\geq 1} A_n \in \mathcal{F}$.

$\mathcal{F}$ is called the set of *events*.

(c) $P : \Omega \to [0, 1]$ satisfies the properties:

1. $P(\Omega) = 1$,
2. for any sequence $(A_n, n \geq 1) \subset \mathcal{F}$ of disjoint sets,

$$P(\cup_{n\geq 1} A_n) = \sum_{n\geq 1} P(A_n).$$

This property is called σ-*additivity*.
The mapping P is called the *probability*. It tell us how likely each event A does occur.

The triple $(\Omega, \mathcal{F}, P)$ is a particular case of a finite measure space with total mass equal to one.
The following properties of P follow from the above axioms.

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

(ii) *additivity* If $(A_n, n \in \{1, \cdots, m\}) \subset \mathcal{F}$ is a finite family of disjoint events, then

$$P(\cup_{n=1}^{m} A_n) = \sum_{n=1}^{m} P(A_n),$$

(iii) $P(A^c) = 1 - P(A)$,

(iv) *monotony* If $A \subset B$, then $P(A) \leq P(B)$,

(v) *subadditivity* If $(A_n, n \in \{1, \cdots, m\}) \subset \mathcal{F}$ is a finite family of events, then

$$P(\cup_{n=1}^{m} A_n) \leq \sum_{n=1}^{m} P(A_n),$$

(vi) If $(A_n, n \geq 1)$ is an increasing sequence of events and $A = \cup_{n \geq 1} A_n$,

$$\lim_{n \to \infty} P(A_n) = P(A),$$

(vii) If $(A_n, n \geq 1)$ is a decreasing sequence of events and $A = \cap_{n \geq 1} A_n$,

$$\lim_{n \to \infty} P(A_n) = P(A),$$

Example 1.1 *Tossing two coins simultaneously once.*

The sample space is

$$\Omega = \{H_1H_2, H_1T_2, T_1H_2, T_1T_1\},$$

where H_i (respectively, T_i), denotes outcome *head* (respectively, tail) of coin i. We consider as possible events any subset of Ω, that is $\mathcal{F} = \mathcal{P}(\Omega)$.
We give probability $\frac{1}{4}$ to any outcome and then $P(A) = \frac{1}{4}\text{card}(A)$.
In this example, Ω is a finite set. Therefore $\mathcal{P}(\Omega)$ is finite as well. It is easy to check that for any finite subset of $\mathcal{P}(\Omega)$, the structure of σ–field is given equivalently by the properties

1. $\Omega \in \mathcal{F}$,
2. If $A \in \mathcal{F}$ then also $A^c \in \mathcal{F}$,
3. $\mathcal{F}$ is closed under finite unions. That means, if $(A_n, n = 1, \ldots, m)$ is a finite family of sets in $\mathcal{F}$, then $\cup_{n=1}^{m} A_n \in \mathcal{F}$.

In this case, we say that $\mathcal{F}$ is a *field*. Moreover, the σ-additivity is equivalent to additivity.

For finite sample spaces, there is a standard way to define a probability, as follows. Let

$$\Omega = \{\omega_1, \ldots, \omega_r\}.$$

Consider positive real numbers $p_1, \ldots, p_r$ such that $\sum_{i=1}^r p_r = 1$. Then, for any $A \subset \Omega$, define

$$P(A) = \sum_{i:\omega_i \in A} p_i.$$

Notice that P is well defined on every element of $\mathcal{P}(\Omega)$ and satisfies the axioms of a probability (see (c) before). The particular choice $p_i = \frac{1}{r}$, for any $i = 1, \ldots, r$, leads to the formula

$$P(A) = \frac{\operatorname{card} A}{\operatorname{card} \Omega},$$

which is the *classical* definition of probability given by Laplace about 200 years ago.
This model is the *finite and uniform* probability space. The previous Example 1.1 belongs to this class of models. Notice that, a similar idea for associating uniform probability to an infinite sample path does not work.

1.2 Conditional Probability and Independence

Let $B \in \mathcal{F}$ be an event of strictly positive probability. The *conditional probability* given B is the mapping $P(\cdot/B) : \mathcal{F} \to [0, 1]$ defined by

$$P(A/B) = \frac{P(A \cap B)}{P(B)}.$$

It is easy to check that $P(\cdot/B)$ satisfies the axioms of a probability. Actually $(\Omega, \mathcal{F}, P(\cdot/B))$ is the probability space obtained by modification of the initial one, after having incorporated the new information provided by B.
The next statements provide useful formulas for computations.

(A) If $A_1, \cdots, A_n \in \mathcal{F}$ and if $P(A_1 \cap \cdots \cap A_{n-1}) > 0$, then

$$P(A_1 \cap \cdots \cap A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2) \cdots P(A_n/A_1 \cap \cdots \cap A_{n-1}).$$

(B) If $(A_n, n \geq 1) \subset \mathcal{F}$ is a countable partition of Ω, and $A \in \mathcal{F}$, then

$$P(A) = \sum_{n \geq 1} P(A \cap A_n) = \sum_{n \geq 1} P(A/A_n)P(A_n).$$

Definition 1.1 *Events of a family* $(A_i, i \in I)$ *are said to be independent if for each finite subset* $(i_1, \ldots, i_k) \subset I$,

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}).$$

In particular, two events A, B are independent if $P(A \cap B) = P(A)P(B)$. Noice that if A, B are independent, then $P(A/B) = P(A)$. That means, knowing B does not modify the probability of the event A.

A notion related with independence is the following:

Definition 1.2 *The events* A, B *are conditionally independent given an event* C *if*

$$P(A \cap B/C) = P(A/C)P(B/C).$$

As we shall see in forthcoming chapters, this notion is crucial in the formulation of the Markov property.

1.3 Random Variables

It is usual to associate numerical values to the outcomes of random phenomena. Formally, this can be described by a mapping $X : \Omega \to \mathbb{R}$. The properties of X should be such that the original structure of the probability space is transferred to a *numerical* probability space.
More precisely, let $\mathcal{B}$ be the Borel σ–field on $\mathbb{R}$ that is, the minimal σ–field containing the open sets of $\mathbb{R}$ with respect to the Euclidean topology. Then, a random variable is a mapping $X : \Omega \to \mathbb{R}$ such that for any $B \in \mathcal{B}$, the set $X^{-1}(B)$ is in $\mathcal{F}$ (it is an event).
With this definition, the map $P_X : \mathcal{B} \to [0, 1]$ defined by

$$P_X(B) = P(X^{-1}(B)),$$

is a probability on $\mathcal{B}$. It is termed the *law of* X. The triple $(\mathbb{R}, \mathcal{B}, P_X)$ is a probability space; it can be thought as the numerical replica of $(\Omega, \mathcal{F}, P)$ given by X.
A notion related with the law of a random variable is that of the *distribution function*, defined as follows:

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

Thus, F_X describes the values of the law of X for a particular class of sets in $\mathcal{B}$, $B = (0, x]$. In the sequel, we shall write F instead of F_X.
The distribution function has the following properties:

(1) F is non-decreasing,

(2) F is right-continuous,

(3) $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

Using the machinery of measure theory, it is proved that any probability measure on $(\mathbb{R}, \mathcal{B})$ is characterized by its distribution function. That is, given a real function F with values on $[0, 1]$, satisfying the properties (1)-(3) before, there exists a unique probability measure μ on $\mathcal{B}$ such that

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

Among the set of all random variables, there are two special important classes. We next give a description and several examples.

Definition 1.3 *A random variable X is discrete if it takes a countable number of values.*

Such random variables have a representation like $X = \sum_{n\in\mathbb{N}} a_n 1_{A_n}$, where the a_n are different to each other and $A_n = \{X = a_n\}$ are disjoint events. The above somehow abstract writing says that, on observations $\omega \in A_n$ the random variable X takes the values a_n.

Definition 1.4 *1. A random variable X is continuous if its distribution function is continuous.*

2. A random variable X is absolutely continuous if its distribution function can be written as

$$F(x) = \int_{-\infty}^{x} f(y)dy,$$

where f is a positive, Riemann integrable function, such that $\int_{-\infty}^{\infty} f(y)dy = 1$.

The function f is called the density of F and, by extension, the density of X as well. Clearly, an absolutely continuous random variable is continuous.

Example 1.2 *Fix $\lambda > 0$. A Poisson random variable with parameter λ is a random variable taking values on $\mathbb{Z}_+$ such that*

$$P\{X = k\} = \exp(-\lambda)\,\frac{\lambda^k}{k!}. \tag{1.1}$$

Poisson random variables are used for modelling rare events.
Clearly, (1.1) gives the distribution function of X and therefore its law.

Later on, we shall see that Poisson random variables underpine an important example of counting process -the Poisson process.
Other famous examples of discrete random variables are: Bernoulli, Binomial, Hypergeometric, etc.

Example 1.3 *Fix $\lambda, p > 0$. A Gamma random variable with parameters λ, p is an absolutely continuous random variable with density function*

$$f(x) = \frac{1}{\Gamma(p)} \lambda^p x^{p-1} \exp(-\lambda x) 1_{\{x \geq 0\}}, \tag{1.2}$$

where Γ is the Euler Gamma function.

We denote this law by $\Gamma(\lambda, p)$ For $p \in \mathbb{N}$, $\Gamma(p) = (p-1)!$.
Let $p = 1$ in the previous example. Then

$$f(x) = \lambda \exp(-\lambda x) 1_{\{x \geq 0\}}. \tag{1.3}$$

This is the density of an *exponential* random variable with parameter λ, denoted by $\exp(\lambda)$
Using characteristic functions (Fourier transform), it can be checked that the sum of n independent exponential random variables with parameter λ has distribution $\Gamma(\lambda, n)$.
Other important examples of absolutely continuous random variables are: normal, uniform, chi-square, etc.

1.4 Mathematical Expectation

One of the most important notion in Probability is that of expected value, *mean value* or *mathematical expectation* of a random variable. As it is suggested by its name, this is a real number associated to the random variable, giving a sort of average of all possible values $X(\omega)$, $\omega \in \Omega$.

Definition 1.5 *The mathematical expectation of a random variable X is the Lebesgue integral*

$$E(X) = \int_\Omega X dP.$$

Not every measurable function is Lebesgue integrable. Therefore, the notion of mathematical expectation may not be defined for an arbitrary random variable. We are not going to discuss the issue of existence; instead, we shall concentrate on how to compute mathematical expectations of random variables of the two types described in Section 1.3. Here are some results.

(1) *Discrete random variables.* The mathematical expectation of $X = \sum_{n\in\mathbb{N}} a_n 1_{A_n}$ exists if and only if

$$\sum_{n\in\mathbb{N}} |a_n| P(A_n) < \infty,$$

and in this case,

$$E(X) = \sum_{n\in\mathbb{N}} a_n P(A_n). \tag{1.4}$$

(2) *Absolutely continuous random variables.* The mathematical expectation of an absolutely continuous random variable X exists if and only if

$$\int_{-\infty}^{\infty} |x| f(x) dx < \infty,$$

and in this case,

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx. \tag{1.5}$$

A brief explanation for the validity of the preceding formulas follows. First, by the image measure theorem (a sort of change of variables formula),

$$E(X) = \int_{\mathbb{R}} x dP_X(x).$$

Then, the formulas are obtained taking into account the particular form of the measure P_X in the discrete and absolutely continuous case, respectively. Actually, (1.4) and (1.5) extend to random variables of the form $g(X)$, where $g : \mathbb{R}^n \to \mathbb{R}$. More specifically, we have

$$E\left(g(X)\right) = \sum_{n\in\mathbb{N}} g(a_n) P(X = a_n),$$
$$E\left(g(X)\right) = \int_{-\infty}^{\infty} g(x) f(x) dx.$$

The following spaces of random variables play an important role in probability theory. For any $p \in [1, \infty)$, $L^p(\Omega)$ is the space of random variables satisfying $E(|X|^p) < \infty$. Actually, in these spaces we identify random variables which coincide on events of probability one.

1.5 Random Vectors. Independence

A m-dimensional *random vector* is a mapping $X : \Omega \to \mathbb{R}^m$, $X = (X_1, \ldots, X_m)$ such that each component X_i, $i = 1, \ldots, m$, is a random variable.

Similarly as in the one-dimensional case, a random vector induces a probability measure on the σ-field of Borel sets of $\mathbb{R}^m$, $\mathcal{B}(\mathbb{R}^m)$, by the formula $P_X(B) = P(X^{-1}(B))$, $B \in \mathcal{B}(\mathbb{R}^m)$, called the *law* of X. The values of $P_X(B)$ are characterized by the *distribution function* of X.
The distribution function of a random vector is given as follows:

$$F_X : \mathbb{R}^m \to [0,1], \quad F_X(x) = P\{\omega : X(\omega) \leq x\}. \tag{1.6}$$

In (1.6), the symbol $\leq$ means the partial order in $\mathbb{R}^m$ defined coordinate wise. Hence, if $x = (x_1, \ldots, x_m)$,

$$F(x) = P\{\omega : X_1(\omega) \leq x_1, \ldots, X_m(\omega) \leq x_m\}.$$

Definitions 1.3, 1.4 can be extended to the multidimensional case with the following notion of density:
A function $f : \mathbb{R}^m \to [0, \infty)$ is a probability density on $\mathbb{R}^m$ if it is Riemann integrable and

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_m) dx_1 \ldots dx_m = 1.$$

A *multinomial random vector* -the one which gives the number of occurences of each one of the m possible outcomes after n independent repetitions of a random experience, like throwing a dice n times- is an example of discrete random vector. Denote by $A_1, \ldots, A_m$ the outcomes and $p_1, \ldots, p_m$ their respective probabilities. Notice that $\sum_{i=1}^m p_1 = 1$. Then, the law of X is concentrated in the set $\mathcal{M} = \{(n_1, \ldots, n_m) \in \mathbb{Z}_+^m : n_1 + \cdots + n_m = n\}$ and

$$P\{X_1 = n_1, \ldots, X_m = n_m\} = \frac{n!}{n_1! \ldots n_m!} p_1^{n_1} \ldots p_m^{n_m},$$

where $X_1, \ldots, X_m$ denotes the number of outcomes of $A_1, \ldots, A_m$, respectively.
An important example of absolutely continuous multidimensional probability distribution is the multidimensional Gaussian or normal law $N(\mu, \Lambda)$. Here $\mu \in \mathbb{R}^m$ and Λ is a m-dimensional symmetric, positive definite matrix. Its density is given by

$$f(x_1, \ldots, x_m) = (2\pi \det \Lambda)^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \sum_{i,j=1}^{n} (x_i - \mu_i)(\Lambda^{-1})_{i,j}(x_j - \mu_j) \right).$$

Definition 1.6 *The random variables $X_1, \ldots, X_m$, are said to be independent if for any choice of Borel sets $B_1, \ldots, B_n$, the events $X_1^{-1}(B_1)$, ..., $X_m^{-1}(B_m)$, are independent.*

The independence of the random variables $X_1, \ldots, X_m$ is equivalent to the fact that the distribution function of the random vector $X = (X_1, \ldots, X_m)$ is given by

$$F_X(x) = \Pi_{i=1}^m F_{X_i}(x_i),$$

for any $x = (x_i, \ldots, x_m)$. Moreover, if the random variables X_i, $i = 1, \ldots, m$, have finite expectation then the product $X_1 \times \cdots \times X_m$ has also finite expectation and

$$E(X_1 \times \cdots \times X_m) = \Pi_{i=1}^m E(X_i). \tag{1.7}$$

Throughout the course we shall be mainly interested in families of random variables which are not independent. For random variables X and Y with finite second order moments, that is with $X, Y \in L^2(\Omega)$, a parameter capturing the degree of dependence is the *covariance*, defined by

$$\mathrm{Cov}(X, Y) = E\left((X - EX)(Y - EY)\right).$$

Notice that, as a consequence of (1.7), if X and Y are independent, $\mathrm{Cov}(X, Y) = 0$.

2 Introduction to Stochastic Processes

This chapter is devoted to introduce the notion of stochastic processes and some general definitions related with this notion. For a more complete account on the topic, we refer the reader to [13]. Let us start with a definition.

Definition 2.1 *A stochastic process with state space S is a family $\{X_i, i \in I\}$ of random variables $X_i : \Omega \to S$ indexed by a set I.*

For a successful progress in the analysis of such object, one needs to put some structure on the index set I and on the state space S. In this course, we shall mainly deal with the particular cases: $I = \mathbb{N}, \mathbb{Z}_+, \mathbb{R}_+$ and S either a countable set or a subset of $\mathbb{R}$.
The basic problem statisticians are interested in, is the analysis of the probability law (mostly described by some parameters) of characters exhibited by populations. For a fixed character described by a random variable X, they use a finite number of independent copies of X -a sample of X. For many purposes, it is interesting to have samples of any size and therefore to consider sequences $X_n, n \geq 1$. It is important here to insist on the word *copies*, meaning that the circumstances around the different outcomes of X do not change. It is a static world. Hence, they deal with stochastic processes $\{X_n, n \geq 1\}$ consisting of independent and identically distributed random variables.
However, this is not the setting we are interested in here. Instead, we would like to give stochastic models for phenomena of the real world which evolve as time passes by. Stochasticity is a choice in front of a complete knowledge and extreme complexity. Evolution, in contrast with statics, is what we observe in most phenomena in Physics, Chemistry, Biology, Economics, Life Sciences, etc.
Stochastic processes are well suited for modeling stochastic evolution phenomena. The interesting cases correspond to families of random variables X_i which are not independent. In fact, the famous classes of stochastic processes are described by means of types of dependence between the variables of the process.

2.1 The Law of a Stochastic Process

The probabilistic features of a stochastic process are gathered in the joint distributions of their variables, as given in the next definition.

Definition 2.2 *The finite-dimensional joint distributions of the process $\{X_i, i \in I\}$ consists of the multi-dimensional probability laws of any finite*

family of random vectors $X_{i_1}, \ldots, X_{i_m}$, *where* $i_1, \ldots, i_m \in I$ *and* $m \geq 1$ *is arbitrary.*

Let us give an important example.

Example 2.1 *A stochastic process* $\{X_t, t \geq 0\}$ *is said to be Gaussian if its finite-dimensional joint distributions are Gaussian laws.*
Remember that in this case, the law of the random vector $(X_{t_1}, \ldots, X_{t_m})$ *is characterized by two parameters:*

$$\mu(t_1, \ldots, t_m) = E\left(X_{t_1}, \ldots, X_{t_m}\right) = \left(E(X_{t_1}), \ldots, E(X_{t_m})\right)$$
$$\Lambda(t_1, \ldots, t_m) = \Big(\mathrm{Cov}(X_{t_i}, X_{t_j})\Big)_{1 \leq i,j \leq m}.$$

In the sequel we shall assume that $I \subset \mathbb{R}_+$ and $S \subset \mathbb{R}$, either countable or uncountable, and denote by $\mathbb{R}^I$ the set of real-valued functions defined on I. A stochastic process $\{X_t, t \geq 0\}$ can be viewed as a random vector

$$X : \Omega \to \mathbb{R}^I.$$

Putting the appropriate σ-field of events in $\mathbb{R}^I$, say $\mathcal{B}(\mathbb{R}^I)$, one can define, as for random variables, the law of the process as the mapping

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

Mathematical results from measure theory tell us that P_X is defined by means of a procedure of extension of measures on cylinder sets given by the family of all possible finite-dimensional joint distributions. This is a deep result.
In Example 2.1, we have defined a class of stochastic processes by means of the type of its finite-dimensional joint distributions. But, does such an object exist? In other words, could one define stochastic processes giving only its finite-dimensional joint distributions? Roughly speaking, the answer is yes, adding some extra condition. The precise statement is a famous result by Kolmogorov that we now quote.

Theorem 2.1 *Consider a family*

$$\{P_{t_1,\ldots,t_n}, t_1 < \ldots < t_n, n \geq 1, t_i \in I\} \tag{2.1}$$

where:

1. $P_{t_1,\ldots,t_n}$ *is a probability on* $\mathbb{R}^n$,

2. *if* $\{t_{i_1} < \ldots < t_{i_m}\} \subset \{t_1 < \ldots < t_n\}$, *the probability law* $P_{t_{i_1}\ldots t_{i_m}}$ *is the marginal distribution of* $P_{t_1\ldots t_n}$.

There exists a stochastic process $\{X_t, t \in I\}$ *defined in some probability space, such that its finite-dimensional joint distributions are given by (2.1). That is, the law of the random vector* $(X_{t_1}, \ldots, X_{t_n})$ *is* $P_{t_1,\ldots,t_n}$.

One can apply this theorem to Example 2.1 to show the existence of Gaussian processes, as follows.
Let $K : I \times I \to \mathbb{R}$ be a symmetric, positive definite function. That means:

- for any $s, t \in I$, $K(t, s) = K(s, t)$;
- for any natural number n and arbitrary $t_1, \ldots, t_n \in I$, and $x_1, \ldots, x_n \in \mathbb{R}$,
$$\sum_{i,j=1}^{n} K(t_i, t_j) x_i x_j > 0.$$

Then there exists a Gaussian process $\{X_t, t \geq 0\}$ such that $E(X_t) = 0$ for any $t \in I$ and $\mathrm{Cov}\,(X_{t_i}, X_{t_j}) = K(t_i, t_j)$, for any $t_i, t_j \in I$.
To prove this result, fix $t_1, \ldots, t_n \in I$ and set $\mu = (0, \ldots, 0) \in \mathbb{R}^n$, $\Lambda = (K(t_i, t_j))_{1 \leq i,j \leq n}$ and
$$P_{t_1,\ldots,t_n} = N(0, \Lambda).$$
We denote by $(X_{t_1}, \ldots, X_{t_n})$ a random vector with law $P_{t_1,\ldots,t_n}$. For any subset $\{t_{i_1}, \ldots, t_{i_m}\}$ of $\{t_1, \ldots, t_n\}$, it holds that

$$A(X_{t_1}, \ldots, X_{t_n}) = (X_{t_{i_1}}, \ldots, X_{t_{i_m}}),$$

with
$$A = \begin{pmatrix} \delta_{t_1,t_{i_1}} & \cdots & \delta_{t_n,t_{i_1}} \\ \cdots & \cdots & \cdots \\ \delta_{t_1,t_{i_m}} & \cdots & \delta_{t_n,t_{i_m}} \end{pmatrix},$$
where $\delta_{s,t}$ denotes the Kronecker Delta function.
By the properties of Gaussian vectors, the random vector $(X_{t_{i_1}}, \ldots, X_{t_{i_m}})$ has an m-dimensional normal distribution, zero mean, and covariance matrix $A\Lambda A^t$. By the definition of A, it is trivial to check that

$$A\Lambda A^t = (K(t_{i_l}, t_{i_k}))_{1 \leq l,k \leq m}.$$

Hence, the assumptions of Theorem 2.1 hold true and the result follows.

2.2 Sample Paths

In the previous discussion, stochastic processes are considered as random vectors. In the context of modeling, what matters are observed values of the process. Observations correspond to fixed values of $\omega \in \Omega$. This new point of view leads to the next definition.

Definition 2.3 *The sample paths of a stochastic process $\{X_t, t \in I\}$ are the family of functions indexed by $\omega \in \Omega$, $X(\omega) : I \to S$, defined by $X(\omega)(t) = X_t(\omega)$.*

Example 2.2 *Consider random arrivals of customers at a store. We set our clock at zero and measure the time between two consecutive arrivals. They are random variables $X_1, X_2, \ldots$. Set $S_0 = 0$ and $S_n = \sum_{j=1}^{n} X_j$, $n \geq 1$. S_n is the time of the n-th arrival. The process we would like to introduce is N_t, giving the number of customers who have visited the store during the time interval $[0, t]$, $t \geq 0$.*
Clearly, $N_0 = 0$ and for $t > 0$, $N_t = k$ if and only if

$$S_k \leq t < S_{k+1}.$$

The stochastic process $\{N_t, t \geq 0\}$ takes values on $\mathbb{Z}_+$. Its sample paths are increasing right continuous functions, with jumps at the random times S_n, $n \geq 1$, of size one.
Later on in the course, we are going to study these kind of processes. For instance, the Poisson process is obtained when assuming the random variables $X_1, X_2, \ldots$ to be independent and identically distributed with exponential law. We shall see that each random variable N_t has a Poisson distribution with parameter λt (see Example 1.2).
The preceding example is a particular case of a *counting process*. Sample paths of counting processes are always increasing right continuous functions, their jumps are natural numbers.

Example 2.3 *Evolution of prices of risky assets can be described by real-valued stochastic processes $\{X_t, t \geq 0\}$ with continuous, although very rough, sample paths. They are generalizations of the Brownian motion.*
The Brownian motion, also called Wiener process, is a Gaussian process $\{B_t, t \geq 0\}$ with the following parameters:

$$E(B_t) = 0$$
$$E(B_s B_t) = s \wedge t,$$

This defines the finite dimensional distributions and therefore the existence of the process via Kolmogorov's theorem (see Theorem 2.1).

Before giving a heuristic motivation for the preceding definition of Brownian motion we introduce two further notions.

A stochastic process $\{X_t, t \in I\}$ has *independent increments* if for any $t_1 < t_2 < \ldots < t_k$ the random variables $X_{t_2} - X_{t_1}, \ldots, X_{t_k} - X_{t_{k-1}}$ are independent.

A stochastic process $\{X_t, t \in I\}$ has *stationary increments* if for any $t_1 < t_2$, the law of the random variable $X_{t_2} - X_{t_1}$ is the same as that of $X_{t_2 - t_1}$.

Brownian motion is termed after Robert Brown, an British botanist who observed and reported in 1827 the irregular movements of pollen particles suspended in a liquid. Assume that, when starting the observation, the pollen particle is at position $x = 0$. Denote by B_t the position of (one coordinate) of the particle at time $t > 0$. By physical reasons, the trajectories must be continuous functions and because of the erratic movement, it seems reasonable to say that $\{B_t, t \geq 0\}$ is a stochastic process. It also seems reasonable to assume that the change in position of the particle during the time interval $[t, t + s]$ is independent of its previous positions at times $\tau < t$ and therefore, to assume that the process has independent increments. The fact that such an increment must be stationary is explained by kinetic theory, assuming that the temperature during the experience remains constant.
The model for the law of B_t has been given by Einstein in 1905. More precisely, Einstein's definition of Brownian motion is that of a stochastic processes with independent and stationary increments such that the law of an increment $B_t - B_s$, $s < t$ is Gaussian, zero mean and $E(B_t - B_s)^2 = t - s$. This definition is equivalent to the one given before.

3 Discrete Time Martingales

In this chapter, we study a very popular class of stochastic processes: martingales and their relatives, submartingales and supermartingales. We shall keep to the simplest case of discrete time processes. As references for the topics we are reporting here, we cite [3], [4].
From the theoretical point of view, martingales represent a first generalization of sequences of independent random variables. In fact, if $\{X_n, n \geq 1\}$ denotes such a sequence, then $\{S_n = \sum_{i=1}^{n} X_i, n \geq 1\}$ provides an example of martingale. From an applied perspective, martingales are on the basis of modeling games and gambler's strategies. More recently, they are showing its performance in the analysis of financial markets.

3.1 Conditional Expectation

The mathematical tool for studying martingales is the notion of conditional expectation and its properties. Roughly speaking, a conditional expectation of a random variable is the mean value with respect to a modified probability after having incorporated some *a priori* information. The simplest case corresponds to conditioning with respect to an event $B \in \mathcal{F}$. In this case, the conditional expectation is the mathematical expectation computed on the modified probability space $(\Omega, \mathcal{F}, P(\cdot/B))$.
However, in general, additional information cannot be described so easily. Assuming that we know about some events $B_1, \ldots, B_n$ we also know about those that can be derived from them, like unions, intersections, complementaries. This explains the election of a σ-field to keep known information and to deal with it.
In the sequel, we denote by $\mathcal{G}$ an arbitrary σ-field included in $\mathcal{F}$ and by X a random variable with finite expectation ($X \in L^1(\Omega)$). Our final aim is to give a definition of the conditional expectation of X given $\mathcal{G}$. However, in order to motivate this notion, we shall start with more simple situations.

Conditional expectation given an event

Let $B \in \mathcal{F}$ be such that $P(B) \neq 0$. The conditional expectation of X given B is the real number defined by the formula

$$E(X/B) = \frac{1}{P(B)} E(\mathbf{1}_B X). \tag{3.1}$$

It immediately follows that

- $E(X/\Omega) = E(X)$,

- $E(1_A/B) = P(A/B)$.

With the definition (3.1), the conditional expectation coincides with the expectation with respect to the conditional probability $P(\cdot/B)$. We check this fact with a discrete random variable $X = \sum_{i=1}^{\infty} a_i 1_{A_i}$. Indeed,

$$\begin{aligned} E(X/B) &= \frac{1}{P(B)} E\left(\sum_{i=1}^{\infty} a_i 1_{A_i \cap B}\right) = \sum_{i=1}^{\infty} a_i \frac{P(A_i \cap B)}{P(B)} \\ &= \sum_{i=1}^{\infty} a_i P(A_i/B). \end{aligned}$$

Conditional expectation given a discrete random variable

Let $Y = \sum_{i=1}^{\infty} y_i 1_{A_i}$, $A_i = \{Y = y_i\}$. The conditional expectation of X given Y is the random variable defined by

$$E(X/Y) = \sum_{i=1}^{\infty} E(X/Y = y_i) 1_{A_i}. \tag{3.2}$$

Notice that, knowing Y means knowing all the events that can be described in terms of Y. Since Y is discrete, they can be described in terms of the basic events $\{Y = y_i\}$. This may explain the formula (3.2).
The following properties hold:

(a) $E\left(E(X/Y)\right) = E(X)$;

(b) if the random variables X and Y are independent, then $E(X/Y) = E(X)$.

For the proof of (a) we notice that, since $E(X/Y)$ is a discrete random variable

$$\begin{aligned} E\left(E(X/Y)\right) &= \sum_{i=1}^{\infty} E(X/Y = y_i) P(Y = y_i) \\ &= E\left(X \sum_{i=1}^{\infty} 1_{\{Y=y_i\}}\right) = E(X). \end{aligned}$$

Let us now prove (b). The independence of X and Y yields

$$\begin{aligned} E(X/Y) &= \sum_{i=1}^{\infty} \frac{E(X 1_{\{Y=y_i\}})}{P(Y = y_i)} 1_{A_i} \\ &= \sum_{i=1}^{\infty} E(X) 1_{A_i} = E(X). \end{aligned}$$

The next proposition states two properties of the conditional expectation that motivates the Definition 3.1

Proposition 3.1 *1. The random variable $Z := E(X/Y)$ is $\sigma(Y)$-measurable; that is, for any Borel set $B \in \mathcal{B}$, $Z^{-1}(B) \in \sigma(Y)$,*

2. for any $A \in \sigma(Y)$, $E\left(1_A E(X/Y)\right) = E(1_A X)$.

Proof: Set $c_i = E(X/\{Y = y_i\})$ and let $B \in \mathcal{B}$. Then

$$Z^{-1}(B) = \cup_{i:c_i \in B} \{Y = y_i\} \in \sigma(Y),$$

proving the first property.
To prove the second one, it suffices to take $A = \{Y = y_k\}$. In this case

$$\begin{aligned} E\left(1_{\{Y=y_k\}} E(X/Y)\right) &= E\left(1_{\{Y=y_k\}} E(X/Y = y_k)\right) \\ &= E\left(1_{\{Y=y_k\}} \frac{E(X 1_{\{Y=y_k\}})}{P(Y = y_k)}\right) = E(X 1_{\{Y=y_k\}}). \end{aligned}$$

□

Conditional expectation given a σ-field

Definition 3.1 *The conditional expectation of X given $\mathcal{G}$ is a random variable Z satisfying the properties*

1. *Z is $\mathcal{G}$-measurable; that is, for any Borel set $B \in \mathcal{B}$, $Z^{-1}(B) \in \mathcal{G}$,*
2. *for any $G \in \mathcal{G}$,*
$$E(Z 1_G) = E(X 1_G).$$

We will denote the conditional expectation Z by $E(X/\mathcal{G})$.
Notice that the conditional expectation is not a number but a random variable. There is nothing strange in this, since conditioning depends on the observations.
Condition (1) tell us that events that can be described by means of $E(X/\mathcal{G})$ are in $\mathcal{G}$. Whereas condition (2) tell us that on events in $\mathcal{G}$ the random variables X and $E(X/\mathcal{G})$ have the same mean value.

The existence of $E(X/\mathcal{G})$ is not a trivial issue. You should trust mathematicians and believe that there is a theorem in measure theory -the Radon-Nikodym Theorem- which ensures its existence.
Before stating properties of the conditional expectation, we are going to explain how to compute it in two particular situations.

Example 3.1 *Let $\mathcal{G}$ be the σ-field (actually, the field) generated by a finite partition $G_1, \ldots, G_m$. Then*

$$E(X/\mathcal{G}) = \sum_{j=1}^{m} \frac{E(X1_{G_j})}{P(G_j)} 1_{G_j}. \tag{3.3}$$

Formula (3.3) tell us that, on each generator of $\mathcal{G}$, the conditional expectation is constant; this constant is weighted by the *mass* of the generator ($P(G_j)$). It can be checked using Definition 3.1. Indeed, it suffices to consider G in the set of generators of $\mathcal{G}$, for instance let us fix $G := G_k$. Then

$$\begin{aligned} E\left(1_{G_k} E(X/\mathcal{G})\right) &= E\left(1_{G_k} \sum_{j=1}^{m} \frac{E(X1_{G_j})}{P(G_j)} 1_{G_j}\right) \\ &= E\left(1_{G_k} \frac{E(X1_{G_k})}{P(G_k)}\right) = E\left(1_{G_k} X\right). \end{aligned}$$

Example 3.2 *Let $\mathcal{G}$ be the σ-field generated by random variables $Y_1, \ldots, Y_m$, that is, the σ-field generated by events of the form $Y_1^{-1}(B_1), \ldots, Y_1^{-1}(B_m)$, with $B_1, \ldots, B_m$ arbitrary Borel sets. Assume in addition that the joint distribution of the random vector $(X, Y_1, \ldots, Y_m)$ has a density f. Then*

$$E(X/Y_1, \ldots, Y_m) = \int_{-\infty}^{\infty} x f(x/Y_1, \ldots, Y_m) dx, \tag{3.4}$$

with

$$f(x/y_1, \ldots, y_m) = \frac{f(x, y_1, \ldots, y_m)}{\int_{-\infty}^{\infty} f(x, y_1, \ldots, y_m) dx}. \tag{3.5}$$

In (3.5), we recognize the conditional density of X given $Y_1 = y_1, \ldots, Y_m = y_m$. Hence, in (3.4) we first compute the conditional expectation $E(X/Y_1 = y_1, \ldots, Y_m = y_m)$ and finally, replace the real values $y_1, \ldots, y_m$ by the random variables $Y_1, \ldots, Y_m$.

We now list some important properties of the conditional expectation.

(a) Linearity: for any random variables X, Y and real numbers a, b

$$E(aX + bY/\mathcal{G}) = aE(X/\mathcal{G}) + bE(Y/\mathcal{G}).$$

(b) Monotony: If $X \leq Y$ then $E(X/\mathcal{G}) \leq E(Y/\mathcal{G})$.

(c) The mean value of a random variable is the same as that of its conditional expectation: $E(E(X/\mathcal{G})) = E(X)$.

(d) If X is a $\mathcal{G}$-measurable random variable, then $E(X/\mathcal{G}) = X$

(e) Let X be independent of $\mathcal{G}$, meaning that any set of the form $X^{-1}(B)$, $B \in \mathcal{B}$ is independent of $\mathcal{G}$. Then $E(X/\mathcal{G}) = E(X)$.

(f) Factorization: If Y is a bounded, $\mathcal{G}$-measurable random variable,

$$E(YX/\mathcal{G}) = YE(X/\mathcal{G}).$$

(g) If $\mathcal{G}_i$, $i = 1, 2$ are σ-fields with $\mathcal{G}_1 \subset \mathcal{G}_2$,

$$E(E(X/\mathcal{G}_1)/\mathcal{G}_2) = E(E(X/\mathcal{G}_2)/\mathcal{G}_1) = E(X/\mathcal{G}_1).$$

(h) Assume that X is a random variable independent of $\mathcal{G}$ and Z another $\mathcal{G}$-measurable random variable. For any measurable function $h(x, z)$ such that the random variable $h(X, Z)$ is in $L^1(\Omega)$,

$$E(h(X, Z)/\mathcal{G}) = E(h(X, z))|_{Z=z}.$$

We give some proofs.
Property (a) follows from the definition of the conditional expectation and the linearity of the operator E. Indeed, the candidate $aE(X/\mathcal{G})+bE(Y/\mathcal{G})$ is $\mathcal{G}$-measurable. By property 2 of the conditional expectation and the linearity of E,

$$\begin{aligned} E\left(1_G[aE(X/\mathcal{G}) + bE(Y/\mathcal{G})]\right) &= aE(1_G X) + bE(1_G Y) \\ &= E(1_G[aX + bY]). \end{aligned}$$

Property (b) is a consequence of the monotony property of the operator E and a result in measure theory telling that, for $\mathcal{G}$-measurable random variables Z_1 and Z_2, satisfying

$$E(Z_1 1_G) \leq E(Z_2 1_G),$$

for any $G \in \mathcal{G}$, we have $Z_1 \leq Z_2$. Indeed, for any $G \in \mathcal{G}$ we have $E(1_G X) \leq E(1_G Y)$. Then, by property 2 of the conditional expectation,

$$E(1_G E(X/\mathcal{G})) = E(1_G X) \leq E(1_G Y) = E(1_G E(Y/\mathcal{G})).$$

By applying the above mentioned property to $Z_1 = E(X/\mathcal{G})$, $Z_2 = E(Y/\mathcal{G})$, we get the result.

Taking $G = \Omega$ in condition (2) above, we prove (c). Property (d) is obvious. Constant random variables are measurable with respect to any σ-field. Therfore $E(X)$ is $\mathcal{G}$-measurable. Assuming that X is independent of $\mathcal{G}$, yields

$$E(X1_G) = E(X)E(1_G) = E(E(X)1_G).$$

This proves (e).
For the proof of (f), we first consider the case $Y = 1_{\tilde{G}}$, $\tilde{G} \in \mathcal{G}$. Claiming (f) means that we propose as candidate for $E(YX/\mathcal{G}) = 1_{\tilde{G}}E(X/\mathcal{G})$. Clearly $1_{\tilde{G}}E(X/\mathcal{G})$ is $\mathcal{G}$-measurable. Moreover,

$$E\left(1_G 1_{\tilde{G}} E(X/\mathcal{G})\right) = E\left(1_{G\cap\tilde{G}} E(X/\mathcal{G})\right) = E\left(1_{G\cap\tilde{G}} X\right).$$

The validity of the property extends by linearity to simple random variables. Then, by monotone convergence to positive random variables and, finally, to random variables in $L^1(\Omega)$, by the usual decomposition $X = X^+ - X^-$.
For the proof of (g), we notice that since $E(X/\mathcal{G}_1)$ is $\mathcal{G}_1$-measurable, it is $\mathcal{G}_2$-measurable as well. Then, by the very definition of the conditional expectation,

$$E(E(X/\mathcal{G}_1)/\mathcal{G}_2) = E(X/\mathcal{G}_1).$$

Next, we prove that $E(X/\mathcal{G}_1) = E(E(X/\mathcal{G}_2)/\mathcal{G}_1)$. For this, we fix $G \in \mathcal{G}_1$ and apply the definition of the conditional expectation. This yields

$$E\left(1_G E(E(X/\mathcal{G}_2)/\mathcal{G}_1)\right) = E\left(1_G E(X/\mathcal{G}_2)\right) = E\left(1_G X\right).$$

Properety (h) is very intuitive: Since X is independent of $\mathcal{G}$ in does not enter the game of conditioning. Moreover, the measurability of Z means that by conditioning one can suppose it is a constant.

3.2 Martingales, Submartingales, Supermartingales

An increasing sequence of sub σ-fields of $\mathcal{F}$,

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_{n-1} \subset \mathcal{F}_n \subset \cdots,$$

is termed *a filtration.*
Given a stochastic process $\{X_n, n \geq 0\}$, there is a *natural* way to define a filtration associated to it, as follows. Set $\mathcal{F}_0$ the trivial σ-field generated by the constants and $\mathcal{F}_n$ the one generated by the random variables X_i, $0 \leq i \leq n$, for any $n \geq 1$.

Definition 3.2 *A stochastic process $\{X_n, n \geq 0\} \subset L^1(\Omega)$ is said to be a martingale with respect to $\{\mathcal{F}_n, n \geq 0\}$ if*

(i) X_n *is* $\mathcal{F}_n$*-measurable, for any* $n \geq 0$*,*

(ii) $E(X_{n+1}/\mathcal{F}_n) = X_n$*.*

Stochastic processes satisfying condition (i) are called *adapted* to the filtration $\{\mathcal{F}_n, n \geq 0\}$.
Replacing in condition (ii) the equality sign by $\geq$ (respectively, $\leq$) gives the definition of *submartingale* (respectively, *supermartingale*).
By property (d) of the conditional expectation, condition (ii) can be equivalently written as

$$E(X_{n+1} - X_n/\mathcal{F}_n) = 0.$$

In this form, we can attach a meaning to the martingale property in the following way. Assume that X_n gives the capital at time n owned by a gambler. Then $X_{n+1} - X_n$ is the amount he wins at the $n+1$-th game. The martingale condition means that the game is fair. Similarly, a submartingale is a favorable game and a supermartingale a non-favorable one.

Lemma 3.1 *Let* $\{X_n, n \geq 0\}$ *be a martingale (respectively, a submartingale, a supermartingale). Then,* $E(X_n) = E(X_0)$ *(respectively* $E(X_n) \geq E(X_0)$*,* $E(X_n) \leq E(X_0)$*), for any* $n \geq 1$*.*

The result follows immediately from property (c) of the conditional expectation.

Example 3.3 *Let* $\xi = \{\xi_n, n \geq 1\}$ *be a sequence of independent random variables with* $E(\xi_n) = 0$*. Set*

$$X_0 = 0 \tag{3.6}$$

$$X_n = \sum_{i=1}^{n} \xi_i. \tag{3.7}$$

The stochastic process $X = \{X_n, n \geq 0\}$ *is a martingale with respect to the natural filtration* $(\mathcal{F}_n, n \geq 0)$ *associated with* $\{\xi_n, n \geq 1\}$*.*

Indeed, owing to property (e) of the conditional expectation,

$$\begin{aligned} E(X_{n+1} - X_n/\mathcal{F}_n) &= E(\xi_{n+1}/\mathcal{F}_n) \\ &= E(\xi_{n+1}) = 0. \end{aligned}$$

If in this example, we assume in addition that the random variables ξ_n are identically distributed, with common mean μ, by the previous computations we see that X is a martingale (respectively, a submartingale, a supermartingale) if $\mu = 0$ (respectively, $\mu > 0$, $\mu < 0$.
Notice that $(\mathcal{F}_n, n \geq 0)$ coincides with the natural filtration associated with $\{X_n, n \geq 0\}$ as well.

Example 3.4 *Let $\xi = \{\xi_n, n \geq 1\}$ be a sequence of i.i.d. positive random variables. Fix a positive random variable X_0 and set $X_n = X_0 \cdot \xi_1 \cdots \xi_n$, $n \geq 1$. If $\mu := E\xi_1 = 1$ (respectively, $\mu > 1$, $\mu < 1$), then $X = \{X_n, n \geq 0\}$ is a martingale (respectively, a submartingale, a supermartingales) with respect to the natural filtration $(\mathcal{F}_n, n \geq 0)$ associated with $\{X_n, n \geq 0\}$.*

Notice that $(\mathcal{F}_n, n \geq 0)$ coincides with the filtration generated by $X_0, \xi_n, n \geq 1$.
To see that X defines a martingale, we apply property (f) and then (e) of the conditional expectation to obtain

$$\begin{aligned} E(X_{n+1} - X_n/\mathcal{F}_n) &= E((\xi_{n+1} - 1)X_0\xi_1 \cdots \xi_n/\mathcal{F}_n) \\ &= X_0\xi_1 \cdots \xi_n E(\xi_{n+1} - 1/\mathcal{F}_n) \\ &= X_0\xi_1 \cdots \xi_n E(\xi_{n+1} - 1) \\ &= X_n E(\xi_{n+1} - 1). \end{aligned}$$

This example is applied in modeling stock prices. In fact, the following particular cases appear in the financial literature.

1. *Discrete Black-Scholes model.* $\xi_n = \exp(Z)$, with $Z =^{(d)} \mathcal{N}(\mu, \sigma^2)$.
2. *Binomial model.* $\xi_n = (1 + a)\exp(-r)$, with probability p and $\xi_n = (1+a)^{-1}\exp(-r)$, with probability $1-p$. Here, the parameter r means the interest rate by which we discount future rewards. At time $n \geq 1$, the price would have the form $X_0(1 + a)^k \exp(-nr)$, $k \leq n$.

In applications, we shall often deal with two filtrations associated in some way to the process. Property (g) of the conditional expectation tells us when the martingale property is preserved. More precisely, we have the following result.

Proposition 3.2 *Let $X = \{X_n, n \geq 0\}$ be a martingale with respect to a filtration $\{\mathcal{F}_n, n \geq 0\}$. Assume that the natural filtration of X, $\{\mathcal{G}_n, n \geq 0\}$ satisfies $\mathcal{G}_n \subset \mathcal{F}_n$, for any $n \geq 0$. Then, X is also a martingale with respect to $\{\mathcal{G}_n, n \geq 0\}$.*

Proof: The process X is clearly adapted to $\{\mathcal{G}_n, n \geq 0\}$. By property (g) of the conditional expectation,

$$E(X_{n+1}/\mathcal{G}_n) = E(E(X_{n+1}/\mathcal{F}_n)/\mathcal{G}_n) = E(X_n/\mathcal{G}_n) = X_n,$$

where in the last equality we have applied property (d). □

3.3 Martingale Transforms

Let $\{\xi_n, n \geq 1\}$ be a i.i.d. sequence such that $P(\xi_n = 1) = p$, $P(\xi_n = -1) = 1 - p$, representing the amount a gambler wins by flipping a coin. Assume he starts with a capital $X_0 = C_0 > 0$. His capital at the n-th flipping would be $X_n = X_0 + \sum_{i=1}^{n} \xi_i$. Suppose he decides to bet, that means, in view of the evolution of the game, he brings some amount of money H_i at each flipping. For example, he could decide to bet H_i on heads. His capital at time n would be now given by

$$W_n = X_0 + \sum_{i=1}^{n} H_i \xi_i.$$

Notice that H_n depends on what happened at the flippings $i = 1, \cdots, n-1$. This example lead to the following notions.

Definition 3.3 *A stochastic process $H = \{H_n, n \geq 1\}$ is predictable with respect to a filtration $\{\mathcal{F}_n, n \geq 0\}$ if H_n is $\mathcal{F}_{n-1}$-measurable, for any $n \geq 1$.*

If $\{\mathcal{F}_n, n \geq 0\}$ is the natural filtration associated with some stochastic process $X = \{X_n, n \geq 0\}$, predictability means that H_n is described by knowledge on $X_0, \cdots, X_{n-1}$, that is, on the past of X.

Definition 3.4 *Fix a filtration $\{\mathcal{F}_n, n \geq 0\}$ for further reference. Let $H = \{H_n, n \geq 1\}$ be a predictable process and $X = \{X_n, n \geq 0\}$ be a martingale. The martingale transform of X by H is the stochastic process denoted by $\{(H \cdot X)_n, n \geq 0\}$ defined as*

$$(H \cdot X)_0 = X_0$$
$$(H \cdot X)_n = X_0 + \sum_{i=1}^{n} H_i (X_i - X_{i-1}),\ n \geq 1.$$

In the sequel, to simplify the notation we shall write $\Delta_i X = X_i - X_{i-1}$.
A martingale transform is in fact an integral operator: The integrand is the process H and the integrator X. One of its most important properties is that the martingale property is preserved under some special conditions, as is made explicit in the next proposition.

Proposition 3.3 *Let $X = \{X_n, n \geq 0\}$ be a martingale (respectively, a submartingale, a supermartingale) and $H = \{H_n, n \geq 1\}$ be a bounded positive predictable process. Then, the martingale transform process $\{(H \cdot X)_n, n \geq 0\}$ is a martingale (respectively, a submartingale, a supermartingale).*

Proof: Since $\{X_n, n \geq 0\} \subset L^1(\Omega)$ and H is a bounded sequence, we have $\{(H \cdot X)_n, n \geq 0\} \subset L^1(\Omega)$. Clearly, $(H \cdot X)$ is adapted to the reference filtration $\{\mathcal{F}_n, n \geq 0\}$. It remains to prove the martingale property. For this, we apply property (f) of the conditional expectation, yielding

$$\begin{aligned} E\left((H \cdot X)_{n+1} - (H \cdot X)_n / \mathcal{F}_n\right) &= E\left(H_{n+1}\Delta_{n+1}X / \mathcal{F}_n\right) \\ &= H_{n+1}E\left(\Delta_{n+1}X / \mathcal{F}_n\right). \end{aligned}$$

The conclusion follows from the properties of X.

Remark The hypothesis of H being positive in the preceding proposition is only necessary to state the result for sub and supermartingales.

With a bit more sophisticated technique, the boundedness of H in the previous proposition can be removed.

Proposition 3.4 *Let $X = \{X_n, 0 \leq n \leq n_0\}$ be a martingale and $H = \{H_n, 1 \leq n \leq n_0\}$ be a predictable process. Assume that $E\left((H \cdot X)_{n_0}^-\right) < \infty$. Then, the martingale transform process $\{(H \cdot X)_n, 0 \leq n \leq n_0\}$ is a martingale.*

Proof: We go to the bounded case by means of a stopping procedure. We define a random variable

$$\tau_k : \Omega \to \{0, 1, \cdots, n_0\}$$

by

$$\begin{aligned} \tau_k &= 0, \text{if } H_1 > k \\ &= \sup\{i : |H_i| \leq k\} \wedge n_0. \end{aligned}$$

Set $Z = (H \cdot X)$. Then, the sequence

$$Z_{n \wedge \tau_k} = X_0 + \sum_{i=1}^{n} H_i 1_{\{\tau_k \geq i\}} \Delta_i X$$

is a martingale. In fact, the random variables $H_i 1_{\{\tau_k \geq i\}}$ are bounded and $\mathcal{F}_{i-1}$-measurable, since

$$\{\tau_k \geq i\} = \{|H_1| \leq k, \cdots, |H_i| \leq k\} \in \mathcal{F}_{i-1}.$$

Hence,

$$E\left(Z_{(n+1) \wedge \tau_k} / \mathcal{F}_n\right) = Z_{n \wedge \tau_k}.$$

Notice that $\lim_{k\to\infty}\tau_k = n_0$. The result follows letting $k \to \infty$ in the preceding inequality. However, for making the argument rigourous, one needs the variable Z_n to be integrable. For this, it suffices to prove recursively that $E(Z_n^-)$ and $E(Z_n^+)$ are finite.
Jensen's inequality applied to the convex function $\varphi(x) = x^-$, yields

$$\begin{aligned} Z_n^- 1_{(\tau_k \geq n+1)} &= Z_{n\wedge\tau_k}^- 1_{(\tau_k \geq n+1)} \\ &= \left(E\left(Z_{(n+1)\wedge\tau_k}/\mathcal{F}_n\right)\right)^- 1_{(\tau_k \geq n+1)} \\ &\leq E\left(Z_{(n+1)\wedge\tau_k}^-/\mathcal{F}_n\right) 1_{(\tau_k \geq n+1)} \\ &= E\left(Z_{(n+1)\wedge\tau_k}^- 1_{(\tau_k \geq n+1)}/\mathcal{F}_n\right) \\ &= E\left(Z_{(n+1)}^- 1_{(\tau_k \geq n+1)}/\mathcal{F}_n\right). \end{aligned}$$

Taking expectations in both sides of the preceeding inequality yields

$$E\left(Z_n^- 1_{(\tau_k \geq n+1)}\right) \leq E\left(Z_{(n+1)}^- 1_{(\tau_k \geq n+1)}\right).$$

Since the random variables involved are positive, one can let k tend to infinity to obtain

$$E\left(Z_n^-\right) \leq E\left(Z_{(n+1)}^-\right) \leq E\left(Z_{n_0}^-\right),$$

where in the last inequality we have used that $(Z_n^-, n \geq 0)$ is a submartingale. The boundedness of the positive part is proved as follows:

$$\begin{aligned} E(Z_n^+) = E\left(\liminf_{k\to\infty} Z_{n\wedge\tau_k}^+\right) &\leq \liminf_{k\to\infty} E\left(Z_{n\wedge\tau_k}^+\right) \\ &= \liminf_{k\to\infty}\left(E\left(Z_{n\wedge\tau_k}\right) + E\left(Z_{n\wedge\tau_k}^-\right)\right) \\ &= E(Z_0) + \liminf_{k\to\infty} E\left(Z_{n\wedge\tau_k}^-\right) \\ &\leq |E(Z_0)| + \sum_{i=1}^{n_0} E(Z_i^-) < \infty. \end{aligned}$$

□

Remark 3.1 *The random variable τ_k defined in the proof of the previous theorem is an example of stopping time. We shall introduce this notion in the next section.*

3.4 Stopping Times and Martingales

The purpose of this section is to prove that when observing a martingale at random times, the martingale property is preserved. We have to make precise what *random times* are allowed.

Definition 3.5 *A random variable $T : \Omega \to \mathbb{Z}_+ \cup \{\infty\}$ is a stopping time with respect to a given filtration $(\mathcal{F}_n, n \geq 0\}$ if, for any $n \in \mathbb{Z}_+$, the event $\{T = n\}$ belongs to $\mathcal{F}_n$.*

The above definition is equivalent to say that for any $n \in \mathbb{Z}_+$ the event $\{T \leq n\}$ belongs to $\mathcal{F}_n$. Indeed, this follows trivially from the set of equalities

$$\begin{aligned} \{T \leq n\} &= \cup_{i=1}^{n} \{T = i\} \\ \{T = n\} &= \{T \leq n\} \cap (\{T \leq n-1\})^c \,. \end{aligned}$$

Let us mention some of the basic properties of stopping times. In the sequel we assume that the reference filtration is always the same.

1. A constant random variable is a stopping time. In fact, if $T = c$, a.s., the event $\{T = n\}$ is either Ω or the empty set.

2. Any linear combination of stopping times is also a stopping time.

3. The supremum and the infimum of two stopping times is a stopping time. More generally, let T_j, $j \geq 1$, be a sequence of stopping times. Then the random variables $S = \sup_{j\geq 1} T_j$ and $I = \inf_{j\geq 1} T_j$ are stopping times.

   Proof: It is a consequence of the following set of equalities and the structure of a σ-field.

$$\begin{aligned} \{S \leq n\} &= \cap_{j\geq 1} \{T_j \leq n\} \\ \{I > n\} &= \cap_{j\geq 1} \{T_j > n\}. \end{aligned}$$

The random variable τ_k of the proof of Proposition 3.4 is a stopping time. Indeed,

$$\{\tau_k = 0\} = \{H_1 > k\} = \{H_1 \leq k\}^c \in \mathcal{F}_0.$$

and, for $1 \leq n \leq n_0 - 1$,

$$\{\tau_k = n\} = \{|H_1| \leq k, \ldots, |H_n| \leq k, |H_{n+1}| > k\} \in \mathcal{F}_n.$$

Example 3.5 *Hitting times. Let $B \in \mathcal{B}$. The hitting time of a stochastic process $X = \{X_n, n \geq 0\}$ to B is defined as*

$$T_B = \inf \{n \geq 0 : X_n \in B\},$$

if this last set is nonempty, and $T_B = \infty$, otherwise.
We check that T_B is a stopping time with respect to the natural filtration associated with X. Indeed, for any $n \geq 1$,

$$\{T_B = n\} = \{X_0 \notin B, \ldots, X_{n-1} \notin B, X_n \in B\} \in \mathcal{F}_n,$$

while for $n = 0$,

$$\{T_B = 0\} = \{X_0 \in B\} \in \mathcal{F}_0.$$

Given a stochastic process $\{Z_n, n \geq 0\}$ and a stopping time T, the random variable Z_T is defined as

$$Z_T(\omega) = Z_{T(\omega)}(\omega).$$

Example 3.6 *In connection with the proof of Proposition 3.4, we give an example of predictable process.*
Let T be a stopping time. Set $H_n = 1_{\{T \geq n\}}$, $n \geq 1$. This defines a predictable process. Indeed

$$\{T \geq n\} = \{T \leq n-1\}^c \in \mathcal{F}_{n-1}.$$

For any stochastic process $X = \{X_n, n \geq 0\}$, we have

$$\begin{aligned}(H \cdot X)_n &= X_0 + \sum_{i=1}^{n} 1_{\{T \geq i\}} \Delta_i X \\ &= X_0 + \sum_{i=1}^{T \wedge n} \Delta_i X = X_{T \wedge n}.\end{aligned}$$

By Proposition 3.3, assuming that X is a martingale (respectively, a submartingale, supermartingale), the stochastic process $\{X_{T \wedge n}, n \geq 0\}$ is again a martingale (respectively, a submartingale, supermartingale).

Consider an $(\mathcal{F}_n, n \geq 0)$-adapted stochastic process $X = \{X_n, n \geq 0\}$; we may wonder what kind of measurability the process has when observed at random times. To give an answer to this question, we introduce a σ-algebra associated with the stopping time T.

Definition 3.6 *The σ-field of events prior to T, $\mathcal{F}_T$ is given by*

$$\mathcal{F}_T = \{A \in \mathcal{F} : A \cap \{T \leq n\} \in \mathcal{F}_n, \quad \textit{for all} \quad n \geq 0\}.$$

Let us check that $\mathcal{F}_T$ is actually a σ-field. Indeed, the equalities

$$A^c \cap \{T \leq n\} = (A \cup \{T > n\})^c = ((A \cap \{T \leq n\}) \cup \{T > n\})^c,$$

shows that, if $A \in \mathcal{F}_T$ then $A^c \in \mathcal{F}_T$.
On the other hand, by its very definition, $\mathcal{F}_T$ is closed by countable intersections.
Let us now prove that the random variable X_T is $\mathcal{F}_T$-measurable. For this, we fix a Borel set $B \in \mathcal{B}$ and check that $\{X_T \in B\} \cap \{T \leq n\} \in \mathcal{F}_n$. Indeed,

$$\begin{aligned}\{X_T \in B\} \cap \{T \leq n\} &= \cup_{i=0}^n (\{X_T \in B\} \cap \{T = i\}) \\ &= \cup_{i=0}^n (\{X_i \in B\} \cap \{T = i\}) \in \mathcal{F}_n.\end{aligned}$$

Theorem 3.1 (Stopping Theorem) *Let $\{X_n, n \geq 0\}$ be a martingale (respectively, a submartingale) and S, T be two stopping times satisfying $S \leq T \leq c$, for some $c \in \mathbb{N}$. Then*

$$E(X_T/\mathcal{F}_S) = X_S,$$

(respectively, $E(X_T/\mathcal{F}_S) \geq X_S$).

Proof: Since T is bounded, $|X_T| \leq \sum_{n=0}^c |X_n| \in L^1(\Omega)$.
We will prove that, for any $A \in \mathcal{F}_S$,

$$E\left(1_A(X_T - X_S)\right) = 0. \tag{3.8}$$

For this, we consider the predictable bounded process defined by $H_n = 1_{\{S<n\leq T\}\cap A}$, $n \geq 1$ and notice that

$$\begin{aligned}(H \cdot X)_0 &= X_0, \\ (H \cdot X)_c &= X_0 + 1_A(X_T - X_S)\end{aligned}$$

The martingale property of $\{(H \cdot X)_m, m \geq 0\}$ yields

$$E\left((H \cdot X)_0\right) = E\left((H \cdot X)_c\right) = E\left(X_0 + 1_A(X_T - X_S)\right)$$

(see Lemma 3.1). Consequently, we obtain (3.8). □

3.5 The Snell Envelope

Let $\{Z_n, 0 \leq n \leq N\}$ be a sequence of integrable positive random variables, defined on a probability space $(\Omega, \mathcal{F}, P)$, adapted to some filtration $\{\mathcal{F}_n, 0 \leq n \leq N\}$. The *Snell envelope* is a sequence of random variables $\{U_n, 0 \leq n \leq N\}$ defined recursively as follows:

$$U_N = Z_N$$
$$U_n = \max(Z_n, E(U_{n+1}/\mathcal{F}_n)),\ n = 0, \ldots, N-1. \tag{3.9}$$
$$\tag{3.10}$$

We shall see in the next chapter that such notion plays an important role in option pricing. For the moment, keeping at a theoretical framework, let us state an optimal property of the Snell envelope.

Proposition 3.5 *The Snell envelope is the smallest supermartingale such that $U_n \geq Z_n$, for any $0 \leq n \leq N$.*

Proof: The supermartingale property is obvious, since by its very definition

$$U_n \geq E(U_{n+1}/\mathcal{F}_n),\ n = 0, \ldots, N-1.$$

Let $\{T_n, 0 \leq n \leq N\}$ be another supermartingale satisfying $T_n \geq Z_n$ for any $0 \leq n \leq N$. We have $T_N \geq Z_N = U_N$. Assuming that $T_m \geq U_m$, for any $n \leq m \leq N$, we obtain

$$T_{n-1} \geq E(T_n/\mathcal{F}_{n-1}) \geq E(U_n/\mathcal{F}_{n-1}).$$

Hence,

$$T_{n-1} \geq \max\left(Z_{n-1}, E(U_n/\mathcal{F}_{n-1})\right) = U_{n-1}.$$

This finishes the proof. □

The next result is related with the martingale property of the stopped Snell envelope.

Proposition 3.6 *The random variable*

$$\nu_0 = \inf\{n \geq 0 : U_n = Z_n\} \wedge N \tag{3.11}$$

is a stopping time with respect to the filtration $(\mathcal{F}_n, 0 \leq n \leq N)$, and the process

$$\{U_{n \wedge \nu_0}, 0 \leq n \leq N\}$$

is a martingale with respect to the same filtration.

Proof: First, we prove that ν_0 is a stopping time. Indeed,

$$\{\nu_0 = 0\} = \{U_0 = Z_0\} \in \mathcal{F}_0.$$

Moreover,

$$\{\nu_0 = k\} = \{U_0 > Z_0, \ldots, U_{k-1} > Z_{k-1}, U_k = Z_k\} \in \mathcal{F}_k,$$

for any $k \geq 1$.
Let us next prove the martingale property. By definition

$$U_{n\wedge\nu_0} = U_0 + \sum_{j=1}^{n} 1_{\{\nu_0 \geq j\}} \Delta_j U.$$

Thus, for any $0 \leq n \leq N-1$,

$$U_{(n+1)\wedge\nu_0} - U_{n\wedge\nu_0} = 1_{\{\nu_0 \geq n+1\}}(U_{n+1} - U_n).$$

On the set $\{\nu_0 \geq n+1\}$, we have that $U_n > Z_n$ and consequently,

$$U_n = \max(Z_n, E(U_{n+1}/\mathcal{F}_n)) = E(U_{n+1}/\mathcal{F}_n).$$

Therefore,

$$\begin{aligned} E\left(U_{(n+1)\wedge\nu_0} - U_{n\wedge\nu_0}/\mathcal{F}_n\right) &= E\left(1_{\{\nu_0 \geq n+1\}}(U_{n+1} - U_n)/\mathcal{F}_n\right) \\ &= E\left(1_{\{\nu_0 \geq n+1\}}(U_{n+1} - E(U_{n+1}/\mathcal{F}_n))/\mathcal{F}_n\right) \\ &= 1_{\{\nu_0 \geq n+1\}} E\left(U_{n+1} - E(U_{n+1}/\mathcal{F}_n)/\mathcal{F}_n\right) \\ &= 0. \end{aligned}$$

□

3.6 Optimal Stopping

We keep in this section the same notation as in the previous one. We shall define the notion of optimal stopping time and explain some relation with the Snell envelope.

Definition 3.7 *A stopping time T with respect to the filtration $(\mathcal{F}_n, 0 \leq n \leq N)$ is optimal for the adapted sequence $(Z_n, 0 \leq n \leq N)$ if*

$$E\left(Z_T/\mathcal{F}_0\right) = \sup_{\nu \in \mathcal{T}_{0,N}} E\left(Z_\nu/\mathcal{F}_0\right),$$

where $\mathcal{T}_{0,N}$ is the family of all stopping times taking values in $\{0, 1, 2, \ldots, N\}$.

Proposition 3.7 *The stopping time ν_0 defined in (3.11) is optimal. In addition,*

$$U_0 = E(Z_{\nu_0}/\mathcal{F}_0) = \sup_{\nu \in \mathcal{T}_{0,N}} E(Z_\nu/\mathcal{F}_0).$$

Proof: We already know by Proposition 3.6 that $\{U_{n\wedge\nu_0}, 0 \leq n \leq N\}$ is a martingale. Thus,

$$U_0 = U_{0\wedge\nu_0} = E\left(U_{N\wedge\nu_0}/\mathcal{F}_0\right) = E\left(U_{\nu_0}/\mathcal{F}_0\right) = E\left(Z_{\nu_0}/\mathcal{F}_0\right).$$

Moreover, for any stopping time $\nu \in \mathcal{T}_{0,N}$, the process $\{U_{n\wedge\nu}, 0 \leq n \leq N\}$ is a supermartingale. Hence,

$$U_0 = U_{0\wedge\nu} \geq E\left(U_{N\wedge\nu}/\mathcal{F}_0\right) \geq E\left(Z_\nu/\mathcal{F}_0\right).$$

□

It is easy to check the following extension of the previous result: Fix $n \in \{0, 1, \ldots, N-1\}$. Denote by $\mathcal{T}_{n,N}$ the set of all stopping times taking values on $\{n, n+1, \ldots, N\}$. Then

$$U_n = \sup_{\nu \in \mathcal{T}_{n,N}} E\left(Z_\nu/\mathcal{F}_n\right) = E\left(Z_{\nu_n}/\mathcal{F}_n\right),$$

where $\nu_n = \inf\{j \geq n : U_j = Z_j\}$.

In terms of the Snell envelope, optimal stopping times are characterized as follows.

Theorem 3.2 *A stopping time ν is optimal for the sequence $(Z_n, 0 \leq n \leq N)$ if and only if $Z_\nu = U_\nu$ and the process*

$$\{U_{n\wedge\nu}, 0 \leq n \leq N\}$$

is a martingale.

Proof: Let us first prove that, under the stated conditions, ν is optimal. Clearly, $U_0 = E(U_\nu/\mathcal{F}_0) = E(Z_\nu/\mathcal{F}_0)$ and

$$U_0 = E(Z_\nu/\mathcal{F}_0) = \sup_{\sigma \in \mathcal{T}_{0,N}} E(Z_\sigma/\mathcal{F}_0).$$

Thus, ν is optimal.
Conversely, assume that ν is optimal. Owing to Proposition 3.7, since $Z_\nu \leq U_\nu$ and $\{U_{\nu\wedge n}, 0 \leq n \leq N\}$ is a supermartingale, we obtain

$$U_0 = \sup_{\sigma \in \mathcal{T}_{0,N}} E(Z_\sigma/\mathcal{F}_0) = E(Z_\nu/\mathcal{F}_0) \leq E(U_\nu/\mathcal{F}_0) \leq U_0. \tag{3.12}$$

Thus, we have an equality in (3.12)

$$E(Z_\nu/\mathcal{F}_0) = E(U_\nu/\mathcal{F}_0),$$

and $E(Z_\nu) = E(U_\nu)$.This yields $Z_\nu = U_\nu$, because $Z_\nu \leq U_\nu$.
By the supermartingale property of $\{U_{n\wedge\nu}, 0 \leq n \leq N\}$, we have

$$U_0 = U_{0\wedge\nu} \geq E(U_{n\wedge\nu}/\mathcal{F}_0) \geq E(U_{N\wedge\nu}/\mathcal{F}_0) = E(U_\nu/\mathcal{F}_0). \tag{3.13}$$

The fact that in (3.12) we have equalities, implies $U_0 = E(U_\nu/\mathcal{F}_0)$. Consequently, in (3.13) we also have equalities. Taking expectations yields

$$E(U_{n\wedge\nu}) = E(U_\nu) = E(E(U_\nu/\mathcal{F}_n)).$$

By the supermartingale property,

$$U_{n\wedge\nu} \geq E(U_{N\wedge\nu}/\mathcal{F}_n) = E(U_\nu/\mathcal{F}_n).$$

This implies that $\{U_{n\wedge\nu}, 0 \leq n \leq N\}$ is a martingale and ends the proof of the Theorem. □

Remark 3.2 *The stopping time ν_0 is the smallest optimal time. Indeed, if ν_1 is another optimal time, by the preceding theorem $Z_{\nu_1} = U_{\nu_1}$ and, by the definition of ν_0, $\nu_0 \leq \nu_1$.*

According to Theorem 3.2, in order to give an optimal stopping time we have to find the first time when the sequence $(U_n, 0 \leq n \leq N)$ fails to be a martingale. A useful tool to solve this question is the decomposition of a generic supermartingale given in the next proposition.

Proposition 3.8 (Doob Decomposition) *Let $(U_n, 0 \leq n \leq N)$ be a supermartingale. For any $0 \leq n \leq N$,*

$$U_n = M_n - A_n,$$

where the sequence $M = (M_n, 0 \leq n \leq N)$ is a martingale, $A = (A_n, 0 \leq n \leq N)$ is increasing, predictable and $A_0 = 0$.
There is a unique decomposition of this type.

Proof: The sequences M and A are constructed as follows. Set $M_0 = U_0$, and for $n \geq 1$,

$$M_n = U_0 + \sum_{j=1}^{n} \left(U_j - E(U_j/\mathcal{F}_{j-1})\right),$$
$$A_n = M_n - U_n.$$

It is very easy to check that the sequence M defined before is a martingale. It is also clear that $A_0 = 0$. Moreover,

$$\begin{aligned} A_n - A_{n-1} &= M_n - M_{n-1} - (U_n - U_{n-1}) \\ &= U_n - E(U_n/\mathcal{F}_{n-1}) - (U_n - U_{n-1}) \\ &= U_{n-1} - E(U_n/\mathcal{F}_{n-1}) \geq 0. \end{aligned}$$

Thus, A is increasing.
From the previous equalities, we have

$$A_n = \sum_{j=1}^{n} \left(U_{j-1} - E(U_j/\mathcal{F}_{j-1})\right),$$

and from this expression, it is obvious that A is a predictable sequence.
Let us prove the uniqueness of this decomposition. Assume that we have two sequences M' and A' with the same properties as M and A, respectively, such that

$$U_n = M_n - A_n = M'_n - A'_n.$$

Since $A_0 = A'_0$, we have that $M_0 = M'_0 = U_0$. Consider the algebraic relation

$$M_n - A_n - (M_{n-1} - A_{n-1}) = M'_n - A'_n - (M'_{n-1} - A'_{n-1}),$$

and apply the conditional expectation operator with respect to $\mathcal{F}_{n-1}$. We obtain

$$A_{n-1} - A_n = A'_{n-1} - A'_n.$$

Since $A_0 = A'_0 = 0 = 0$, this implies the identity of the sequences A and A' and therefore the same holds true for M and M'. □

We can now give the optimal stopping time of a sequence $Z = (Z_n, 0 \leq n \leq N)$ via the Doob decomposition of its Snell envelope.

Proposition 3.9 *Let $(A_n, 0 \leq n \leq N)$ be the increasing sequence in the Doob decomposition of the Snell envelope of Z, which we denote by $(U_n, 0 \leq n \leq N)$. The stopping time defined by*

$$\nu_m = \begin{cases} N, & \text{if } A_N = 0, \\ \inf\{n \geq 0 : A_{n+1} \neq 0\}, & \text{if } A_N > 0, \end{cases} \tag{3.14}$$

is optimal.
Moreover, ν_m is the largest optimal stopping time for Z.

Proof: Let us first check that ν_m is a stopping time. Indeed,

$$\{\nu_m = n\} = \cap_{j\leq n}\{A_j = 0\} \cap \{A_{n+1} > 0\} \in \mathcal{F}_n.$$

Let us now prove that $\{U_{\nu_m\wedge n}, 0 \leq n \leq N\}$ is a martingale. Indeed, $U_n = M_n - A_n$ and $A_j = 0$ if $j \leq \nu_m$; thus, $U_{\nu_m\wedge n} = M_{\nu_m\wedge n}$, and the statement follows.
We now check that $U_{\nu_m} = Z_{\nu_m}$. In fact, by the definition of U,

$$\begin{aligned} U_{\nu_m} &= \sum_{j=0}^{N-1} 1_{\{\nu_m=j\}} U_j + 1_{\{\nu_m=N\}} U_N \\ &= \sum_{j=0}^{N-1} 1_{\{\nu_m=j\}} \max(Z_j, E(U_{j+1}/\mathcal{F}_j)) + 1_{\{\nu_m=N\}} Z_N. \end{aligned} \tag{3.15}$$

By the Doob decomposition,

$$E(U_{j+1}/\mathcal{F}_j) = E(M_{j+1} - A_{j+1}/\mathcal{F}_j) = M_j - A_{j+1}.$$

On the set $(\nu_m = j)$, we have $A_j = 0$ and $A_{j+1} > 0$. Therefore, on $(\nu_m = j)$, $M_j = U_j$ and

$$E(U_{j+1}/\mathcal{F}_j) = M_j - A_{j+1} = U_j - A_{j+1} < U_j.$$

Thus, on $(\nu_m = j)$, $U_j = \max(Z_j, E(U_{j+1}/\mathcal{F}_j)) = Z_j$. Plugging this equality in (3.15), we obtain the announced result. According to Theorem 3.2, the stopping time ν_m is optimal.
We finally prove that ν_m is the largest stopping time satisfying the optimality property. For this, consider a stopping time ν such that $P(\nu > \nu_m) > 0$. Then

$$\begin{aligned} E(U_\nu) &= E(M_\nu) - E(A_\nu) \\ &= E(M_0) - E(A_\nu) = E(U_0) - E(A_\nu) \\ &< E(U_0). \end{aligned}$$

This inequality tell us that the sequence $(U_{\nu\wedge n}, 0 \leq n \leq N)$ is not a martingale and therefore, ν is not optimal. □

3.7 Convergence Results

We end this chapter on discrete time martingale theory with a result on convergence of a matingale sequence when $n \to \infty$. For its proof, we need a refinement of Chebychev's inequality for martingales, as follows.

Proposition 3.10 *(Doob-Kolmogorov inequality) Let $\{S_n, n \geq 1\}$ be a martingale with respect to a filtration $(\mathcal{F}_n, n \geq 1\}$. Then, for any $\varepsilon > 0$,*

$$P\left(\max_{1\leq i\leq n} |S_n| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} E\left(S_n^2\right).$$

Proof: We first consider a decomposition of the set Ω into disjoint subsets, as follows. Set

$$\begin{aligned} A_0 &= \Omega, \\ A_k &= \{|S_i| < \varepsilon, \text{for all}\, i \leq k\}, \\ B_k &= A_{k-1} \cap \{|S_k| \geq \varepsilon\}. \end{aligned}$$

It holds that

$$\Omega = A_n \cup \left(\cup_{i=1}^n B_i\right).$$

Hence, by the linear property of the mathematical expectation,

$$E\left(S_n^2\right) = \sum_{i=1}^n E\left(S_n^2 1_{B_i}\right) + E\left(S_n^2 1_{A_n}\right) \geq \sum_{i=1}^n E\left(S_n^2 1_{B_i}\right).$$

We next notice that

$$\begin{aligned} E\left(S_n^2 1_{B_i}\right) &= E\left((S_n - S_i + S_i)^2 1_{B_i}\right) \\ &= \alpha + \beta + \gamma, \end{aligned}$$

with

$$\begin{aligned} \alpha &= E\left((S_n - S_i)^2 1_{B_i}\right), \\ \beta &= 2E\left((S_n - S_i) S_i 1_{B_i}\right), \\ \gamma &= E\left(S_i^2 1_{B_i}\right). \end{aligned}$$

We are going to give a lower bound for each one of these terms. Clearly, $\alpha \geq 0$ and $\gamma \geq \varepsilon^2 P(B_i)$.

Using property (c) of the conditional expectation and that $S_i 1_{B_i}$ is $\mathcal{F}_i$ measurable yield

$$\begin{aligned} E\left((S_n - S_i)S_i 1_{B_i}\right) &= E\left[S_i 1_{B_i} E\left(S_n - S_i/\mathcal{F}_i\right)\right] \\ &= 0. \end{aligned}$$

Consequently,

$$E\left(S_n^2\right) \geq \sum_{i=1}^{n} \varepsilon^2 P(B_i) \geq \varepsilon^2 P\left(\max_{1\leq i\leq n} |S_i| \geq \varepsilon\right),$$

proving the proposition. □

The next theorem gives the behaviour of a martingale for large values of n.

Theorem 3.3 *Let $(X_n, n \geq 0)$ be a martingale bounded in L^2, that is, satisfying $\sup_n E(X_n^2) < M < \infty$. There exists a L^2–valued random variable Y such that*

$$\lim_{n\to\infty} X_n = Y,$$

almost surely and in L^2.

Proof: First, we prove that $\{E\left(X_n^2\right), n \geq 0\}$ is an increasing sequence. Indeed, using property (c) of the conditional expectation we obtain

$$E\left(X_m(X_{m+n} - X_m)\right) = E\left[X_m E\left(X_{m+n} - X_m/\mathcal{F}_m\right)\right] = 0.$$

This implies

$$E\left(X_{m+n}^2\right) = E\left(X_m^2\right) + E\left(X_{m+n} - X_m\right)^2.$$

We set

$$M := \lim_{n\to\infty} E\left(X_n^2\right).$$

Next, we show that $(X_n, n \geq 0)$ is a Cauchy sequence, a.s. For this, we introduce the set

$$\begin{aligned} C = \Big\{ &\text{for all } \varepsilon > 0, \text{there exists} \, m \geq 1, \\ &\text{such that} \, |X_{m+i} - X_m| < \varepsilon, \text{for all} \, i \geq 1 \Big\} \end{aligned}$$

and show that $P(C) = 1$.
By definition,

$$C = \cap_{\varepsilon>0} \cup_{m\geq 1} \left\{|X_{m+i} - X_m| < \varepsilon, \text{for all} \, i \geq 1\right\}.$$

Otherwise stated,

$$C^c = \cup_{\varepsilon>0} \cap_{m\geq 1} A_m(\varepsilon),$$

where $A_m(\varepsilon) = \{|X_{m+i} - X_m| \geq \varepsilon, \text{for some}\, i \geq 1\}$. Since $A_m(\varepsilon)$ decreases in ε,

$$P(C^c) \leq \lim_{\varepsilon\to 0} \lim_{m\to\infty} P\left(A_m(\varepsilon)\right).$$

We next prove that for each $\varepsilon > 0$,

$$\lim_{m\to\infty} P\left(A_m(\varepsilon)\right) = 0,$$

using Doob-Kolomogorov's inequality.
Set $Y_n = X_{m+n} - X_m$. The σ-fields of the natural filtration associated with $Y_n, n \geq 1$, say $\mathcal{G}_n$, are included in $\mathcal{F}_{m+n}$, for each n. Thus,

$$\begin{aligned} E\left(Y_{n+1}/\mathcal{G}_n\right) &= E\left(E\left(Y_{n+1}/\mathcal{F}_{m+n}\right)/\mathcal{G}_n\right) \\ &= E\left(Y_n/\mathcal{G}_n\right) = Y_n. \end{aligned}$$

Hence, $\{Y_n, n \geq 1\}$ is a martingale with respect to $(\mathcal{G}_n, n \geq 1)$ and by applying Proposition 3.10 we obtain

$$\begin{aligned} P\left(A_m(\varepsilon)\right) &= P\left(|X_{m+i} - X_m| \geq \varepsilon, \text{for some } i \geq 1\right) \\ &= \lim_{n\to\infty} P\left(\sup_{1\leq i\leq n} |X_{m+i} - X_m| \geq \varepsilon\right) \\ &\leq \frac{1}{\varepsilon^2} \lim_{n\to\infty} E\left(\left(X_{m+n} - X_m\right)^2\right) \\ &= \frac{1}{\varepsilon^2} \lim_{n\to\infty} \left(E\left(X_{m+n}^2\right) - E\left(X_m^2\right)\right). \end{aligned}$$

Thus,

$$P\left(A_m(\varepsilon)\right) \leq \frac{1}{\varepsilon^2}\left(M - E\left(X_m^2\right)\right).$$

From this, it follows that

$$\lim_{m\to\infty} P\left(A_m(\varepsilon)\right) = 0.$$

finishing the proof of the a.s. convergence.
The proof of the L^2 convergence follows from Fatou's lemma. Indeed,

$$\begin{aligned} E\left(\left(X_n - Y\right)^2\right) &= E\left(\liminf_{m\to\infty} (X_n - X_m)^2\right) \\ &\leq \liminf_{m\to\infty} E\left(\left(X_n - X_m\right)^2\right) \\ = M - E\left(X_n^2\right). \end{aligned}$$

This last expression tends to zero as $n \to \infty$. This finishes the proof of the theorem. □

4 Applications of Martingale Theory to Finance

In this chapter, we shall apply martingale theory to some mathematical models for financial markets. We follow the approach of [4].

4.1 Basic Notions and Definitions

We consider a fixed probability framework consisting of a probability space $(\Omega, \mathcal{F}, P)$ and a *finite* filtration, that is, a family of σ-algebras $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_N = \mathcal{F}$, where $\mathcal{F}_0$ is the σ-field consisting of sets A with either $P(A) = 0$ or $P(A) = 1$.

Definition 4.1 *A finite mathematical market is a sequence of $d+1$-dimensional random vectors $\{(S_n^0, S_n^1, \cdots, S_n^d), 0 \leq n \leq N\}$ such that each S_n^i is positive ($0 \leq n \leq N$, $0 \leq i \leq d$) is $\mathcal{F}_n$-measurable.*

The random variable S_n^i represent the value at time n of some financial asset labeled by i. The condition about measurability tell us that the value of assets at time n may be known on the basis of what has been the evolution until time n. That means, there is no insight into the future. The value of N represents a fixed time horizon.
The 0-asset is assumed to be riskless and with initial value 1, i.e. $S_0^0 = 1$. That means, its associated return over a unity of time is constant and equal to r. Thus $S_n^0 = (1+r)^n$. It is deterministic.

Definition 4.2 *A portfolio, or a trading strategy, in the market $\{(S_n^0, \cdots, S_n^d), n \geq 0\}$ is a sequence of $d+1$-dimensional random vectors $\{(\Phi_n^0, \Phi_n^1, \cdots, \Phi_n^d)\}$, with Φ_n^0 constant and Φ_n^i is $\mathcal{F}_{n-1}$-measurable for each $n \geq 1$ and $1 \leq i \leq d$.*

Here, the random variables Φ_n^i mean the number of shares of asset i at time n. The measurability condition in definition 4.2 means that the composition of the portfolio a time n is decided taking into account the evolution until time $n-1$.

Definition 4.3 *The value of a portfolio at time n is defined by*

$$V_n(\Phi) = \sum_{j=0}^{d} \Phi_n^j S_n^j = \Phi_n \cdot S_n, \tag{4.1}$$

where the symbol "$\cdot$" means the scalar product on $\mathbb{R}^{d+1}$.

Definition 4.4 *A portfolio is self-financing if for any* $n \in \{0, 1, \cdots, N-1\}$,

$$\Phi_n \cdot S_n = \Phi_{n+1} \cdot S_n. \tag{4.2}$$

The meaning of this notion is as follows: A time $n+1$, the investor fixes his strategy Φ_{n+1} by readjusting the composition of the portfolio in such a way that neither extra money is needed, nor extra money is left.
With this property, the value has the structure of a *martingale transform*. More precisely,

Proposition 4.1 *A portfolio is self-financing if and only if*

$$V_n(\Phi) = V_0(\Phi) + \sum_{i=1}^{n} \Phi_i \cdot \Delta_i S. \tag{4.3}$$

Proof: Assume that the portfolio is self-financing. Then $\Delta_i \Phi \cdot S_{i-1} = 0$, for any $i = 1, \cdots, N$. Simple computations yield

$$\begin{aligned} \Delta_i V(\Phi) &= \Phi_i \cdot S_i - \Phi_{i-1} \cdot S_{i-1} \\ &= \Phi_i \cdot S_i - \Phi_i \cdot S_{i-1} + \Phi_i \cdot S_{i-1} - \Phi_{i-1} \cdot S_{i-1} \\ &= \Phi_i \cdot \Delta_i S + \Delta_i \Phi \cdot S_{i-1}. \end{aligned} \tag{4.4}$$

Consequently, we should have $\Delta_i V(\Phi) = \Phi_i \cdot \Delta_i S$. This is equivalent to (4.3). Conversely, assume that (4.3) holds. As just mentioned, this is equivalent to $\Delta_i V(\Phi) = \Phi_i \cdot \Delta_i S$. Owing to (4.4) we must have $\Delta_i \Phi \cdot S_{i-1} = 0$ and this is equivalent to the self-financing property. □

The next proposition provides a way to obtain a self-financing portfolio.

Proposition 4.2 *Fix a real number x (initial capital) and a predictable family of random vectors $\{(\Phi_n^1, \cdots, \Phi_n^d), 0 \le n \le N\}$. There exists a unique family of predictable random vectors, $\tilde{\Phi} = \{\tilde{\Phi}_n^0, 0 \le n \le N\}$, such that $\Phi = \{(\tilde{\Phi}_n^0, \Phi_n^1, \ldots, \Phi_n^d), 0 \le n \le N, 0 \le i \le d\}$ is self-financing and with initial value $V_0(\Phi) = x$.*

Proof: Set

$$\tilde{\Phi}_0^0 = x - \Phi_0^1 S_0^1 - \cdots - \Phi_0^d S_0^d.$$

We define recursively $\tilde{\Phi}_i^0$ by means of the formula (4.2). Then

$$V_n(\Phi) = \tilde{\Phi}_{n+1}^0 S_n^0 + \Phi_{n+1}^1 S_n^1 + \cdots + \Phi_{n+1}^d S_n^d.$$

Since $S_n^0 = (1+r)^{-n}$, this equation gives the value of $\tilde{\Phi}_{n+1}^0$. Predictability follows easily. □

Definition 4.5 *For a given market* $\{S_n, 0 \leq n \leq N\}$*, the normalized market* $\{\tilde{S}_n, 0 \leq n \leq N\}$ *is defined by setting*

$$\tilde{S}_n = (1+r)^{-n} S_n.$$

Obviously, for a normalized market, $\tilde{S}_n^0 = 1$ at any time n. Moreover, the normalized value of the portfolio is

$$\tilde{V}_n(\Phi) = (1+r)^{-n} V_n(\Phi) = \Phi_n \cdot \tilde{S}_n.$$

The self-financing property reads

$$\Phi_n \cdot \tilde{S}_n = \Phi_{n+1} \cdot \tilde{S}_n,$$

for each $n = 0, 1, \ldots, N-1$. Equivalently,

$$\tilde{V}_{n+1}(\Phi) - \tilde{V}_n(\Phi) = \Phi_{n+1} \cdot (\tilde{S}_{n+1} - \tilde{S}_n).$$

Clearly, $\tilde{V}_0(\Phi) = V_0(\Phi)$ and summing up both terms of this identity yields

$$\tilde{V}_n(\Phi) = V_0(\Phi) + \sum_{i=1}^{n} \Phi_i \cdot \Delta_i \tilde{S},$$

for $n = 1, \ldots, N$.

4.2 Admisible Strategies and Arbitrage

In the definition of a trading strategy, we allow the values of the process Φ to be negative. However, we are interested in markets with positive value. This leads to the following definition.

Definition 4.6 *A trading strategy is admissible if it is self-financing and* $V_n(\Phi) \geq 0$ *for any* $n \in \{0, 1, \ldots, N\}$.

We now introduce the notion of *arbitrage* which is a sort of possibility of riskless profit.

Definition 4.7 *An arbitrage strategy is an admissible strategy with zero initial value (*$V_0(\Phi) = 0$*,* $V_N(\Phi) \geq 0$*) and* $P\{V_N(\Phi) > 0\} > 0$.

Definition 4.8 *A market is viable if there is no arbitrage opportunity.*

Viability of financial markets can be characterized in probabilistic terms using the martingale property, as is shown in Theorem 4.1. Let's first give an additional definition.

Definition 4.9 *A probability Q defined on the σ-field $\mathcal{F}$ is neutral if*

1. *Q is equivalent to P, that is, $P(A) = 0$ if and only if $Q(A) = 0$, for any $A \in \mathcal{F}$.*

2. *On the new probability space $(\Omega, \mathcal{F}, Q)$, the discounted prices $\{\tilde{S}_n^i, 0 \leq n \leq N\}$ are martingales for each $i = 1, \ldots, d$.*

Theorem 4.1 *The following statements concerning a finite, admisible market are equivalent.*

(a) The market is viable

(b) There exists a neutral probability.

Proof: Asume first (b). There exists a self-financing strategy Φ with $V_0(\Phi) = 0$, $V_N(\Phi) \geq 0$. The sequence $\{\tilde{V}_n(\Phi), 0 \leq n \leq N\}$ is a martingale transform in the probability space $(\Omega, \mathcal{F}, Q)$. By assumption, $(\tilde{V}_N(\Phi))^- = 0$. Hence, the hypotheses of Proposition 3.4 are satisfied and consequently, $\{\tilde{V}_n(\Phi), 0 \leq n \leq N\}$ is a martingale on $(\Omega, \mathcal{F}, Q)$, null at $n = 0$. Thus,

$$E_Q(\tilde{V}_N(\Phi)) = E_Q(\tilde{V}_0(\Phi)) = 0.$$

Since $V_N(\Phi) \geq 0$, this implies $V_N(\Phi) = 0$, Q-a.s. and, by the equivalence of P and Q, we conclude $V_N(\Phi) = 0$, P-a.s.
Conversely, let us assume that there is no arbitrage strategy and prove the existence of a neutral probability. This part of the proof is rather difficult. For the sake of illustration, we shall fix the particular framework which consists of a *finite* sample space Ω, $\mathcal{F} = \mathcal{P}(\Omega)$ and $P(\{\omega\}) > 0$, for any $\omega \in \Omega$.
Let $\mathcal{C}$ be the set of positive random variables with mean value equal to one. Denote by M the cardinal of Ω. The set $\mathcal{C}$ is a convex, compact subset of $\mathbb{R}^M$. Denote by Γ the set of random variables of the form $V_N(\Phi)$, where Φ is a self-financing strategy, $\Phi_0 = 0$. The set Γ is also a subset of $\mathbb{R}^M$. Since we are assuming that there is no arbitrage, $\mathcal{C}$ and Γ are disjoint. By the convex sets separation theorem, there exists a linear map $L : \mathbb{R}^M \to \mathbb{R}$ such that $L > 0$ on $\mathcal{C}$ and $L \equiv 0$ on Γ. Hence, there exists a random variable Y such that

(i) $\sum_\omega X(\omega)Y(\omega) > 0$, for each $X \in \mathcal{C}$,

(ii) $\sum_\omega V_N(\Phi)(\omega)Y(\omega) = 0$, for each self-financing strategy with $\Phi_0 = 0$.

Property (i) implies $Y(\omega) > 0$, for any $\omega \in \Omega$. Set

$$Q(\{\omega\}) = \frac{Y(\omega)}{\Lambda},$$

with $\Lambda = \sum_\omega Y(\omega)$.
The probabilities P and Q are equivalent. Let $(\Phi_n^1, \ldots, \Phi_n^d)$, $n \geq 1$ be a sequence of predictable processes. There exists a self-financing trading strategy with null initial value such that $(\Phi_n^1, \ldots, \Phi_n^d)$ corresponds to the number of assets at time n.
By virtue of (ii),

$$E_Q(\tilde{V}_N(\Phi)) = \frac{1}{\Lambda} \sum_\omega \tilde{V}_N(\Phi)(\omega) Y(\omega) = 0,$$

and consequently $E_Q\left(\sum_{i=1}^N \Phi_i \cdot \Delta_i \tilde{S}\right) = 0$. Set $\Phi_i^j = 0$ for $j \neq j_0$ and $i \neq i_0$, and $\Phi_{i_0}^{j_0} = 1_A$, with $A \in \mathcal{F}_{i_0-1}$, we obtain

$$E_Q\left(\Delta_{i_0} \tilde{S}^{j_0} 1_A\right) = 0.$$

Thus,

$$E_Q\left(\Delta_{i_0} \tilde{S}^{j_0} / \mathcal{F}_{i_0-1}\right) = 0.$$

This proves that the discounted prices process, $\{\tilde{S}_n^j, 0 \leq n \leq N\}$, are martingales with respect to Q. □

4.3 Options. Notions and Definitions

A *derivative* is a contract on some assets of the financial market. Here, we shall deal with a special case of derivatives: the options. An *option* gives the holder the right, but not the obligation, to buy or sell a certain amount of a financial asset at a certain date, at a certain price.
An option is defined by the following ingredients

(i) Its type: a *call* is an option to buy, a *put* is an option to sell

(ii) The kind and quantity of assets concerned, for example, stocks, bonds, currency, etc.

(iii) The maturity time or expiration date

(iv) The exercise price, which fixes the price at which the transaction is done when the option is exercised.

If options are traded by established markets, their prices are fixed by the market.

Example 4.1 *An European call option on a stock is defined by the price of the stock at any time, S_t, the expiration date $T > 0$ and the exercise price K. The option is exercised at T.*

Assume that $S_T > K$. In this case, the holder makes a profit by exercising the option, because he will buy the option at a price K but sell at price S_T. If $S_T \leq K$, he does not make any profit by exercising the option at the maturity time. The value of the option at T is given by

$$(S_T - K)^+ := \max(S_T - K, 0).$$

In a similar way, for put options, the value at time T is $(K - S_T)^+$.
Comming back to call options, if the holder exercises the option, the writer has to generate the amount $(S_T - K)^+$. This yields to the following questions:

1. How much should the potential option holder pay for the asset at time $t = 0$, when the deal starts? This problem is called *option pricing*.

2. How should the writer design an strategy assuring that it will be possible to generate the amount $(S_T - K)^+$, to avoid losing money? This problem is called *hedging* the option.

In these examples of options, the contingent claim H by the holder ($H := (S_T - K)^+$), depends only of S_T, and therefore it is measurable with respect to $\mathcal{F}_T$, the σ-field generated by $S_0, \ldots, S_T$. There are more complex options. For instance, *Asian options* fix as value of the contingent claim

$$H = \left(\frac{1}{T+1} \sum_{n=0}^{T} (S_n - K) \right)^+ .$$

Notice that H is still $\mathcal{F}_T$-measurable.

4.4 Complete Markets. Option Pricing

We introduce and study in this section the notion of completeness, which allows basically to develop a simple theory.

Definition 4.10 *A random variable $H \geq 0$ is an attainable contingent claim if there exists an admissible trading strategy Φ such that $V_T(\Phi) = H$.*

We shall say that such strategy replicates the option.

Definition 4.11 *A viable market (see Definition 4.8) is complete if every contingent claim is attainable.*

Notice that in a viable, complete market, hedging is always possible. In this case, we shall fix as price of the option the initial value $V_0(\Phi)$ of a replicating strategy and as contingent claim $V_T(\Phi)$. By the next proposition, this makes sense.

Proposition 4.3 *Fix a contingent claim H at time T in a viable market. Then H characterizes the value sequence $\{V_n(\Phi), 0 \leq n \leq T\}$ of any replicating portfolio.*

Proof: Let Φ_1, Φ_2 be two self-financing strategies, with

$$V_T(\Phi_1) = V_T(\Phi_2) = H$$

but such that the sequences $\{V_n(\Phi_1), 0 \leq n \leq T\}$, $\{V_n(\Phi_2), 0 \leq n \leq T\}$ do not coincide. Set $n = \inf\{k = 0, \ldots, T : V_k(\Phi_1) \neq V_k(\Phi_2)\}$.
Assume first that $n = 0$ and $V_0(\Phi_1) < V_0(\Phi_2)$. Let Ψ be a self-financing strategy with null initial value and $\Psi_n^j = \Phi_{1,n}^j - \Phi_{2,n}^j$, $j = 1, \ldots, d$, $n = 0, \ldots, T$. Then,

$$\begin{aligned} V_T(\Psi) &= V_T(\Psi - \Phi_1 + \Phi_2) + V_T(\Phi_1) - V_T(\Phi_2) \\ &= (1+r)^N V_0(\Psi - \Phi_1 + \Phi_2). \end{aligned}$$

Indeed, $\Psi - \Phi_1 + \Phi_2$ is a riskless portfolio. This yields

$$V_T(\Psi) = (1+r)^T (V_0(-\Phi_1) + V_0(\Phi_2)) > 0.$$

Hence, there is arbitrage, contradicting the fact that the market is viable (there is no arbitrage opportunity).
Assume now $n \geq 1$ and thus, $V_0(\Phi_1) = V_0(\Phi_2)$. We may assume $P(A) > 0$, where

$$A = \{V_n(\Phi_1) < V_n(\Phi_2)\}.$$

Let us define a self-financing strategy Ψ as follows:

(i) If either $\omega \in A^c$ or $\omega \in A$ but $k \leq n$, set $\Psi_k(\omega) = \Phi_{2,k}(\omega) - \Phi_{1,k}(\omega)$.

(ii) If $\omega \in A$ and $k > n$, set $\Psi_k^0(\omega) = \tilde{V}_n(\Phi_2) - \tilde{V}_n(\Phi_1)$, and $\Psi_k^i = 0$, for $i = 1, \ldots, d$.

Clearly, Ψ is predictable. Moreover, for either $k < n$ or for $\omega \in A^c$, the self-financing equation holds, because of the same properties of Φ_i, $i = 1, 2$. For $\omega \in A$ and $k > n$, $\Psi_{k+1} = \Psi_k$. Finally, for $k = n$ and $\omega \in A$,

$$\Psi_n \cdot S_n = (\Phi_{2,n} - \Phi_{1,n}) \cdot S_n = V_n(\Phi_2) - V_n(\Phi_1),$$
$$\Psi_{n+1} \cdot S_n = \left(\tilde{V}_n(\Phi_2) - \tilde{V}_n(\Phi_1)\right) S_n^0 = V_n(\Phi_2) - V_n(\Phi_1).$$

On the set A, the portfolio Ψ has a null initial value and final value $V_T(\Phi_2) - V_T(\Phi_1) = 0$, while on the set A^c, $\left(\tilde{V}_n(\Phi_2) - \tilde{V}_T(\Phi_1)\right) S_n^0 > 0$. Thus, Ψ is an arbitrage trading strategy, contradicting the assumption that the market is viable.

□

Complete markets are characterized by the following theorem.

Theorem 4.2 *A viable market is complete if and only if there exists a unique probability Q equivalent to P such that the discounted prices are martingales, that is, Q is neutral (see Definition 4.9).*

Proof: Assume first that the market is complete. Let P_i, $i = 1, 2$ be two equivalent probabilities such that the discounted prices are martingales with respect to both probabilities. Let $H = 1_A$, $A \in \mathcal{F}$ be a contingent claim. By completeness, there exists a self-financing strategy Φ such that $V_N(\Phi) = 1_A$. The sequence $\{\tilde{V}_n(\Phi), 0 \leq n \leq T\}$ is a martingale with respect to both P_i, $i = 1, 2$. Consequently,

$$E_{P_1}(\tilde{V}_N(\Phi)) = E_{P_1}(\tilde{V}_0(\Phi)) = V_0(\Phi),$$
$$E_{P_2}(\tilde{V}_N(\Phi)) = E_{P_2}(\tilde{V}_0(\Phi)) = V_0(\Phi),$$

which yields $P_1(A) = P_2(A)$. Since $A \in \mathcal{F}$ is arbitrary, we conclude $P_1 = P_2$.

Conversely, assume that there exists a unique probability Q equivalent to P such that the discounted prices are martingales, but that the market is not complete. We are going to obtain a contradiction. For the sake of simplicity, we shall assume that Ω is finite, $\mathcal{F} = \mathcal{P}(\Omega)$ and $P(\{\omega\}) > 0$.
Let H be a non replicable contingent claim. Let $\mathcal{V}$ be the vector subspace of $L^2(\Omega, \mathcal{F}, Q)$ consisting of random variables of the form

$$c + \sum_{n=1}^{T} \Phi_n \cdot \Delta_n \tilde{S},\ c \in \mathbb{R},\ \Phi_n, 0 \leq n \leq T,\ \textit{predictable}.$$

Since H is not replicable, $\frac{H}{S_T^0} = H(1+r)^{-T} \notin \mathcal{V}$. Hence $\mathcal{V}$ is a strict subspace of $L^2(\Omega, \mathcal{F}, Q)$. Let X be a non null random variable orthogonal to $\mathcal{V}$.
Set

$$P^*(\{\omega\}) = \left(1 + \frac{X(\omega)}{2\|X\|_\infty}\right) Q(\{\omega\}).$$

The following properties hold:

1. Since $1 \in \mathcal{V}$, $E_{P^*}(X) = 0$ and therefore P^* defines a probability.
2. P^* and Q do not coincide, because X is not identically zero.
3. Using that X is orthogonal to $\mathcal{V}$ in the space $L^2(\Omega, \mathcal{F}, Q)$ and that the sequence $\{\tilde{S}_n^j, 0 \leq n \leq T\}$ is a martingale with respect to Q yields,

$$\begin{aligned} E_{P*}\left(\sum_{n=1}^{T} \Phi_n \cdot \Delta_n \tilde{S}\right) &= E_Q\left(\sum_{n=1}^{T} \Phi_n \cdot \Delta_n \tilde{S}\right) + E_Q\left(\frac{X}{\|X\|_\infty} \sum_{n=1}^{T} \Phi_n \cdot \Delta_n \tilde{S}\right) \\ &= E_Q\left(\sum_{n=1}^{T} \Phi_n \cdot \Delta_n \tilde{S}\right) = 0, \end{aligned}$$

   Therefore, $\{\tilde{S}_n, 0 \leq n \leq T\}$ is a martingale with respect to P^*.

□

In a viable and complete market option pricing and hedging is always possible. Indeed, let H be a contingent claim, that is, a random variable $H \geq 0$. There exists a self-financing portfolio Φ such that $V_T(\Phi) = H$. Moreover, with respect to an equivalent probability Q, the sequence of discounted values $\{\tilde{V}_n(\Phi), 0 \leq n \leq T\}$ is a martingale. Then,

$$(1+r)^{-T} E_Q(H) = E_Q(\tilde{V}_T(\Phi)) = V_0(\Phi),$$

which says that the initial value of this portfolio is determined by H. Then, for a contingent claim H, we price the asset by

$$V_0(\Phi) = (1+r)^{-T} E_Q(H).$$

Moreover, by the martingale property, we also have

$$(1+r)^{-n} V_n(\Phi) = E_Q\left(H(1+r)^{-T}/\mathcal{F}_n\right).$$

Thus,

$$V_n(\Phi) = (1+r)^{n-T} E_Q(H/\mathcal{F}_n),$$

which says that, at any time n, the value of an admissible strategy replicating H is completely determined by H. It seems natural to denote $V_n(\Phi)$ the price of the option at time n (it only depends on H and not on Φ).

4.5 Cox, Ross and Rubinstein model

We introduce here an example which is the discrete version of the Black and Scholes model.
Assume that the market consists of a single asset: $S = \{(S_n^0, S_n^1), n \geq 1\}$, $S_n^0 = (1+r)^n$, $S_n^1 = S_n$. We assume that the value of this asset is $S_0 > 0$ and

$$S_{n+1} = \begin{cases} either\ S_n(1+a) \\ or\ S_n(1+b), \end{cases}$$

with $-1 < a < b$, meaning that the relative price change between n and $n+1$ is either a or b. Equivalently, setting

$$T_n = \frac{S_n}{S_{n-1}},\ 1 \leq n \leq N,$$

$T_n \in \{1+a, 1+b\}$.
Set $\Omega = \{1+a, 1+b\}^N$, the set of possible values of the vector $(T_1, \ldots, T_N)$, $\mathcal{F} = \mathcal{P}(\Omega)$ and P the law of this vector. Remember that P is determined by the probability function $P\{T_1 = x_1, \ldots, T_N = x_N\}$, for any $(x_1, \ldots, x_N) \in \Omega$.
If the market has no arbitrage strategies, then $r \in (a, b)$. Indeed, Assume for instance $r \leq a$. Assume we borrow S_0 at $t = 0$ and at time N we give back S_0 and sell the asset. We make a profit of $S_N - S_0(1+r)^N \geq 0$. Indeed, $S_N \geq S_0(1+a)^N$. Moreover, with strictly positive probability $S_N - S_0(1+a)^N > 0$, hence we have arbitrage. If $r \geq b$, we also reach a similar conclusion.

Assume $r \in (a, b)$ and set $p = \frac{r-a}{b-a}$. The following statements are equivalent:

(A) The discounted price sequence $\{\tilde{S}_n, 0 \leq n \leq N\}$ is a martingale with respect to P

(B) The random variables $T_1, \ldots, T_N$ are independent, with the same distribution and

$$P\{T_1 = 1+a\} = 1-p$$
$$P\{T_1 = 1+b\} = p.$$

Therefore, the financial market is complete and the neutral probability is given by the probability P such that (B) holds.
The discounted price sequence $\{\tilde{S}_n, 0 \leq n \leq N\}$ is a martingale with respect to P if and only if $E(T_{n+1}/\mathcal{F}_n) = r + 1$.
Indeed, $E_P(\tilde{S}_{n+1}/\mathcal{F}_n) = \tilde{S}_n$ is equivalent to $E_P(\frac{\tilde{S}_{n+1}}{\tilde{S}_n}/\mathcal{F}_n) = 1$. Since $\tilde{S}_n$ is $\mathcal{F}_n$-measurable, this last equality is equivalent to $E(T_{n+1}/\mathcal{F}_n) = r + 1$.

Let us prove the above-mentioned equivalence between (A) and (B). Assume first that (B) holds. Then

$$\begin{aligned} E(T_{n+1}/\mathcal{F}_n) &= E(T_{n+1}) = p(1+b) + (1-p)(1+a) \\ &= 1 + r. \end{aligned}$$

This proves (A).
Assume now (A). From the identities

$$\begin{aligned} &(1+a)P(T_{n+1} = 1 + a/\mathcal{F}_n) + (1+b)P(T_{n+1} = 1 + b/\mathcal{F}_n) \\ &= E(T_{n+1}/\mathcal{F}_n) = r + 1, \\ &P(T_{n+1} = 1 + a/\mathcal{F}_n) + P(T_{n+1} = 1 + b/\mathcal{F}_n) = 1, \end{aligned}$$

we obtain

$$\begin{aligned} P(T_{n+1} = 1 + a/\mathcal{F}_n) &= 1 - p, \\ P(T_{n+1} = 1 + b/\mathcal{F}_n) &= p, \end{aligned}$$

This shows (B).

Value of an European call and put

Denote by C_n (respectively, P_n) the value at time n of an European call (respectively, put) of one asset in the Cox, Ross and Rubinstein model with strike (exercise) price K and maturity time N. By the definition of T_n before, we have

$$\begin{aligned} C_n &= (1+r)^{-(N-n)} E\left((S_N - K)^+/\mathcal{F}_n\right) \\ &= (1+r)^{-(N-n)} E\left(\left(S_n \prod_{i=n+1}^{N} T_i - K\right)^+ /\mathcal{F}_n\right). \end{aligned}$$

The random variable S_n is $\mathcal{F}_n$- measurable and $\prod_{i=n+1}^{N} T_i$ is independent of $\mathcal{F}_n$. Applying property (h) of the conditional expectation, yields $C_n = c(n, S_n)$, with

$$\begin{aligned} c(n,x) &= (1+r)^{-(N-n)} E\left(\left(x \prod_{i=n+1}^{N} T_i - K\right)^+\right) \\ &= (1+r)^{-(N-n)} \sum_{j=0}^{N-n} \frac{(N-n)!}{(N-n-j)!j!} p^{N-n-j}(1-p)^j \\ &\quad \times \left(x(1+a)^j(1+b)^{N-n-j} - K\right)^+. \end{aligned}$$

For P_n we have

$$P_n = (1+r)^{-(N-n)} E\left((K - S_N)^+ / \mathcal{F}_n\right).$$

The following simple relation between C_n and P_n allows to compute P_n explicitly:

$$\begin{aligned} C_n - P_n &= (1+r)^{-(N-n)} E\left((S_N - K)^+ - (K - S_N)^+ / \mathcal{F}_n\right) \\ &= (1+r)^{-(N-n)} E\left((S_N - K)/\mathcal{F}_n\right) \\ &= S_n - K(1+r)^{-(N-n)}. \end{aligned}$$

Let us now compute a replicating strategy (remember that in this model, the market is complete). Such portfolio should satisfy

$$\Phi_n^0 (1+r)^n + \Phi_n^1 S_n = c(n, S_n).$$

Substituting S_n by its two possible values yields the two next equations

$$\begin{aligned} \Phi_n^0 (1+r)^n + \Phi_n^1 S_{n-1}(1+a) &= c(n, S_{n-1}(1+a)), \\ \Phi_n^0 (1+r)^n + \Phi_n^1 S_{n-1}(1+b) &= c(n, S_{n-1}(1+b)). \end{aligned}$$

Thus,

$$\Phi_n^1 = \frac{c(n, S_{n-1}(1+b)) - c(n, S_{n-1}(1+a))}{S_{n-1}(b-a)}.$$

4.6 American Options

In Section 4.3, we described an European option. In this section, we shall introduce the notion of American options and we shall apply the results of Chapter 3 on the Snell envelope (see Section 3.5 and 3.6) to price and hedge such type of options.

The difference between an European and an American option is that, in the later the option can be exercised at any time $0, 1, \ldots, N$, before the maturity time N. Otherwise, we shall keep all the notions and definitions given in Section 4.3.

Assume for simplicity that there is a single stock S^1. Then, instead of the contingent claim denoted by H in Section 4.3, we shall have a finite sequence $Z_n, n = 0, 1, \ldots, N$, defined as follows:

For a *Call American Option*,

$$Z_n = (S_n^1 - K)^+.$$

For a *Put American Option*,

$$Z_n = (K - S_n^1)^+.$$

Here K denotes the exercise price and S_n^1, the price of the stock at time n.

Pricing an American option

We shall assume that the market is viable and complete.
Denote by U_n the price of the option at time n. We fix the value of U_n by means of a backward recursive argument.

- For $n = N$, we naturally set $U_N = Z_N$.
- For $n = N - 1$, there are two possibilities. Assume that we exercise the option at $n = N - 1$. In this case, we earn Z_{N-1}. Assume we do not exercise the option; that is, we should exercise the option at time N. We are now in the same situation than for an European call option with exercise price Z_N. The value at $N - 1$ will be

  $$(1 + r)^{-1} E_Q(Z_N / \mathcal{F}_{N-1}).$$

  We are free to choose anyone of these two possibilities. Therefore,

  $$U_{N-1} = \max \left(Z_{N-1}, (1 + r)^{-1} E_Q(Z_N / \mathcal{F}_{N-1}) \right).$$

- Let us now consider an arbitrary time n. There are two possibilities. Either we exercise the option at this time, and the profit will be Z_n, or we do exercise later. In this case, it is like having an European option with contingent claim U_{n+1} and a price at time $n - 1$ given by

  $$(1 + r)^{-1} E_Q(U_{n+1} / \mathcal{F}_n).$$

  Thus,

  $$U_n = \max \left(Z_n, (1 + r)^{-1} E_Q(U_{n+1} / \mathcal{F}_n) \right).$$

Summarising the previous arguments, we fix the price of an American call option as follows:

$$U_n = \begin{cases} Z_N, & \text{if } n = N, \\ \max \left(Z_n, (1 + r)^{-1} E_Q(U_{n+1} / \mathcal{F}_n) \right) & \text{if } n = 0, \ldots, N - 1. \end{cases}$$

Set $\tilde{Z}_n = (1 + r)^{-n} Z_n$, $\tilde{U}_n = (1 + r)^{-n} U_n$, for any $n = 0, 1, \ldots, N$. Then

$$\tilde{U}_n = \begin{cases} \tilde{Z}_N, & \text{if } n = N, \\ \max \left(\tilde{Z}_n, E_Q(\tilde{U}_{n+1} / \mathcal{F}_n) \right) & \text{if } n = 0, \ldots, N - 1. \end{cases}$$

We see that $(\tilde{U}_n, n = 0, \ldots, N)$ is the Snell envelope of $(\tilde{Z}_n, n = 0, \ldots, N)$. This yields the following expression for the price (see Proposition 3.7 and its extension):

$$\tilde{U}_n = \sup_{\nu \in \mathcal{T}_{n,N}} E_Q(\tilde{Z}_\nu / \mathcal{F}_n),$$
$$U_n = \sup_{\nu \in \mathcal{T}_{n,N}} E_Q((1+r)^{-(\nu-n)} Z_\nu / \mathcal{F}_n).$$

The expression $E_Q((1+r)^{-(\nu-n)} Z_\nu / \mathcal{F}_n)$ represents the price of the European option at time n in case that the owner decides to exercise it at the stopping time $\nu \geq n$.

Hedging the American Option

Consider an initial capital U_0 identical to the price of the stock at the initial time. Since the market is complete, there exists a self-financing strategy Φ such that $V_N(\Phi) = M_N$, where $\tilde{U}_n = \tilde{M}_n - \tilde{A}_n$ is the Doob decomposition of the supermartingale $(\tilde{U}_n, 0 \leq n \leq N)$ and

$$M_n = (1+r)^n \tilde{M}_n$$
$$A_n = (1+r)^n \tilde{A}_n$$

Both sequences $(\tilde{M}_n, 0 \leq n \leq N)$ and $(\tilde{V}_n(\Phi), 0 \leq n \leq N)$ are martingales and coincide at N; therefore they must coincide. The initial value of this financial strategy is $V_0(\Phi) = M_0 = U_0$. This is the price for the American option. Moreover, $U_n = M_n - A_n = V_n(\Phi) - A_n$, which yields

$$V_n(\Phi) \geq U_n \geq Z_n.$$

This means that the financial strategy Φ hedges the profit of the American option at any time.

What is the optimal date to exercise the option? The date of exercise can be chosen among the set $\mathcal{T}_{0,N}$ of all stopping times. However, from the point of view of the buyer of the option, there is no point in exercising at a time n such that $U_n > Z_n$, because he would trade an asset worth U_n (the price of the option) for an amount Z_n (by exercising the option). Hence, we are looking for a stopping time τ such that $U_\tau = Z_\tau$. On the other hand, using the notation of Proposition 3.9, we would like to have $\tau \leq \nu_m$, since if $\tau > \nu_m$, $A_\tau > 0$, and then $V_\tau(\Phi) = U_r + A_r > U_r \geq Z_r$. That is, $V_\tau(\Phi) > Z_r$. That is, extra wealth is generated.
A stopping time τ such that $U_\tau = Z_\tau$ and $\tau \leq \nu_m$ is optimal. Indeed, the sequence $(U_{n \wedge \tau} = U_{n \wedge \tau \wedge \tau_m}, 0 \leq n \leq N)$ is a martingale, because $(U_{n \wedge \tau_m}, 0 \leq$

$n \leq N$) is also a martingale (see Proposition 3.9) and we can apply Theorem 3.2.
Thus, we have proved that the optimal dates to exercise are the optimal stopping times.
Notice that, exercising at an optimal stopping time τ gives

$$V_\tau(\Phi) = U_\tau + A_\tau = U_\tau = Z_\tau,$$

because $A_\tau = 0$. Hence the hedging is exact.

Comparing the Values of American and European Options

Let $(U_n, 0 \leq n \leq N)$ be the sequence of values of an American option with associated benefits $(Z_n, 0 \leq n \leq N)$, and $(C_n, 0 \leq n \leq N)$ the ones corresponding to the pricing of an European option with exercise benefit Z_N. We have the following facts

- For any $0 \leq n \leq N$, $U_n \geq C_n$.
- In addition, if $C_n \geq Z_n$ for any $0 \leq n \leq N$, then

$$U_n = C_n.$$

That means: In general, the price of American options is higher than that of European ones.
To prove these facts, consider the neutral probability Q such that $(\tilde{U}_n, 0 \leq n \leq N)$ is a supermartingale. Then

$$\tilde{U}_n \geq E_Q(\tilde{U}_N/\mathcal{F}_n) = E_Q(\tilde{Z}_N/\mathcal{F}_n) = \tilde{C}_n.$$

Assume now that $C_n \geq Z_n$ for each $0 \leq n \leq N$. The process $(\tilde{C}_n, 0 \leq n \leq N)$ is a martingale (and therefore, a supermartingale) satisfying $\tilde{C}_n \geq \tilde{Z}_n$, $0 \leq n \leq N$. Owing to the properties of the Snell envelope, $\tilde{C}_n \geq \tilde{U}_n$, $0 \leq n \leq N$, yielding $C_n = U_n$ for each $0 \leq n \leq N$.

This finishes the proof of the above statements.

Example 4.2 *Consider a call option with exercise price K. We clearly have*

$$\begin{aligned}
C_n &= (1+r)^{-(N-n)} E_Q\left((S_N - K)^+/\mathcal{F}_n\right) \\
&\geq (1+r)^{-(N-n)} E_Q(S_N - K/\mathcal{F}_n) \\
&= (1+r)^n E_Q(\tilde{S}_N/\mathcal{F}_n) - (1+r)^{-(N-n)} K \\
&= (1+r)^n \tilde{S}_n - (1+r)^{-(N-n)} K = S_n - (1+r)^{-(N-n)} K \\
&\geq S_n - K.
\end{aligned}$$

Since $C_n \geq 0$, for any $0 \leq n \leq N$, the preceding inequality yields $C_n \geq (S_n - K)^+ = Z_n$. By the previous discussion, we have that, in this case the prices of American and European options coincide.

For a put, the value of an American option will be in general higher than for an European option. Remember that $Z_n = (K - S_n)^+$.
We can compute the price of the American option in the binomial model as follows: $U_n = u_n(S_n)$, where

$$u_n(x) = \begin{cases} (K-x)^+ & \text{if } n = N, \\ \max\left((K-x)^+, (1+r)^{-1} f_{n+1}(x)\right), & \text{if } n = 0, \ldots, N-1, \end{cases}$$

with

$$f_{n+1}(x) = (1-p)u_{n+1}\left(x(1+a)\right) + p u_{n+1}\left(x(1+b)\right),$$

and $p = \frac{r-a}{b-a}$.

5 Discrete Time Markov Chains

This chapter is devoted to study sequences of random variables $\{X_n, n \geq 0\}$, taking values on countable sets, such that, at any step $n \geq 1$, the random variables corresponding to the *future*, $\{X_m, m \geq n\}$, and those corresponding to the *past*, $\{X_m, 0 \leq m \leq n\}$ are conditionally independent given X_n. This is similar to say that the information about the past of the evolution is captured by what happened at the last step. This kind of dependence is called the *Markov property*.
As for martingales, Markov property was introduced as an attempt to go beyond sequences of independent random variables and to extend classical results of Probability Theory, like the law of large numbers.
Along this chapter, unless otherwise specified, random variables take their values on a countable set denoted by I. We call *states* the elements of I.

5.1 The Markov Property

The definition of a Markov chain needs two ingredients:

(a) A probability ν on I,

(b) A matrix $\Pi = (p_{i,j})_{i,j \in I}$, such that $\sum_{j \in I} p_{i,j} = 1$, for any $i \in I$.

In the sequel, any matrix like the one described in (b) will be called a *stochastic matrix*.

Definition 5.1 *A stochastic process $\{X_n, n \geq 0\}$ is a homogeneous Markov chain with initial distribution ν and transition probability matrix Π if the following two properties hold:*

(1) $P(X_0 = i) = \nu_i$, for any $i \in I$

(2) for any $n \geq 1$, and $i_0, i_1, \ldots, i_{n-1}, i, j \in I$,

$$P(X_{n+1} = j / X_0 = i_0, X_1 = i_1, \ldots, X_n = i) = p_{i,j}. \tag{5.1}$$

Condition (1) determines the probability distribution of the *initial* random variable X_0. Condition (2) tells us that, from the knowledge of values of the process at time $n = 0, 1, \ldots, n$, we only keep the one at n, since the dependence on $i_0, \ldots, i_{n-1}$ is not visible at the right hand-side of 5.1. Moreover, $p_{i,j}$ does not depend on n; this is why we put the word *homogeneous* in the definition, meaning stationary in time.

Example 5.1 *Consider a sequence of independent, identically distributed random variables* $\{Y_n, n \geq 0\}$ *taking values on* $\mathbb{Z}$*. That is, infinitely many independent copies of a random variable* Y*. Set*

$$X_n = \sum_{i=0}^{n} Y_i,\ n \geq 0.$$

The stochastic process $X = \{X_n, n \geq 0\}$ *is a homogeneous Markov chain with initial distribution the probability law of the random variable* Y_0 *and transition probability matrix* Π *given by* $p_{i,j} = P(Y = j - i)$*,* $i, j \in \mathbb{Z}$*.*

Indeed, fix $n \geq 1$. By the definition of X_n, the formula for conditional probabilities and the independence of the random variables Y_n , we have

$$\begin{aligned}
&P(X_{n+1} = j/X_0 = i_0, X_i = i_1, \ldots, X_n = i) \\
&\quad = \frac{P(X_{n+1} = j, X_0 = i_0, X_1 = i_1, \ldots, X_n = i)}{P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i)} \\
&\quad = \frac{P(Y_{n+1} = j - i, X_0 = i_0, Y_1 = i_1 - i_0, \ldots, Y_n = i - i_{n-1})}{P(X_0 = i_0, Y_1 = i_1 - i_0, \ldots, Y_n = i - i_{n-1})} \\
&\quad = P(Y_{n+1} = j - i).
\end{aligned}$$

With similar, but simpler computations, we show that

$$P(X_{n+1} = j/X_n = i) = P(Y_{n+1} = j - i).$$

We see in this example that the elements of the transition probability matrix are

$$p_{i,j} = P(X_{n+1} = j/X_n = i),$$

which justifies its name.
We shall see later that this property holds for every homogeneous Markov chain and therefore, that condition (2) in Definition 5.1 can be written as

$$P(X_{n+1} = j/X_0 = i_0, X_1 = i_1, \ldots, X_n = i) = P(X_{n+1} = j/X_n = i). \quad (5.2)$$

Assume in the previous example that Y takes values on the set $\{-1, 1\}$ with probabilities $1 - p$ and p, respectively, $p \in]0, 1[$. Then the process X is the *Bernoulli random walk* on $\mathbb{Z}$. The values of this process give the position of a walker that starts at some integer position and moves either forward or backward according to the result of coin tossing.
The initial distribution is the law of Y and the transition probability matrix is given by

$$\Pi = \begin{pmatrix} \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & 1-p & 0 & p & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \quad (5.3)$$

Conditions (1) and (2) in Definition 5.1 determine the finite dimensional joint distributions of the process $\{X_n, n \geq 0\}$ (see Definition 2.2). This is a consequence of formula (A) in section 1.2. Indeed, fix $i_0, i_1, \ldots, i_{n-1}, i_n$, then

$$\begin{aligned} & P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) \\ & \quad P(X_0 = i_0) P(X_1 = i_1 / X_0 = i_0) \times \ldots \\ & \quad \times P(X_n = i_n / X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}) \\ & \quad = \eta_{i_0} p_{i_0, i_1} \cdots p_{i_{n-1}, i_n}. \end{aligned} \tag{5.4}$$

As a consequence, we obtain the probability law of each random variable X_n, $n \geq 1$. More precisely, we have the following formula

$$P(X_n = j) = (\nu \Pi^n)_j, \tag{5.5}$$

Indeed, owing to property (B) in section 1.2 and (5.4) we have

$$\begin{aligned} P(X_n = j) = & \sum_{i_0 \in I, \cdots, i_{n-1} \in I} \eta_{i_0} p_{i_0, i_1} \cdots p_{i_{n-1}, j} \\ & (\nu \Pi^n)_j. \end{aligned}$$

We can now prove a useful formula for the transition probabilities in m steps. For any $m \geq 0$, $n \geq 1$, $i, j \in I$, we define

$$p_{i,j}^{(n)} = P(X_{n+m} = j / X_m = i).$$

Then,

$$p_{i,j}^{(n)} = (\Pi^n)_{i,j}. \tag{5.6}$$

That means, assuming that at some time m, the chain visits the state i, the probability that after n steps, that is, at time $m + n$, the chain visits the state j, is the element indexed by i, j of the n-th power of the matrix Π.
Let us prove (5.6). We apply property (B) in section 1.2 and (5.4) to obtain

$$\begin{aligned} p_{i,j}^{(n)} = P(X_{n+m} = j / X_m = i) &= \frac{P(X_{n+m} = j, X_m = i)}{P(X_m = i)} \\ &= \frac{\sum_{i_0, \ldots, i_{m-1}, i_{m+1}, \cdots, i_{n+m-1} \in I} \eta_{i_0} p_{i_0, i_1} \cdots p_{i_{m-1}, i} p_{i, i_{m+1}} \cdots p_{i_{n+m-1}, j}}{(\nu P^m)_i} \\ &= \sum_{i_{m+1}, \cdots, i_{n+m-1} \in I} p_{i, i_{m+1}} \cdots p_{i_{n+m-1}, j} \\ &= (\Pi^n)_{i,j}. \end{aligned}$$

The next proposition is a statement about the Markov property we have mentioned at the beginning of the chapter.

Proposition 5.1 *(Markov Property) Let $\{X_n, n \geq 0\}$ be a HMC with initial distribution ν and transition probability matrix Π. Fix $m \geq 1$. Then, conditionally on $(X_m = i)$, the process $\{X_{n+m}, n \geq 0\}$ is a HMC, with transition probability matrix Π and initial distribution $\delta_{\{i\}}$, independent of the random variables $X_0, \cdots, X_m$.*

Proof: For any $A \in \mathcal{F}$, set $P_i(A) = P(A/X_m = i)$. Clearly,

$$P_i(X_m = j) = \delta_{i,j},$$

We next prove that, for any $n \geq 0$ and $i_1, \cdots, i_n, i_{n+1} \in I$,

$$P_i(X_{n+m+1} = i_{n+1}/X_m = i, X_{m+1} = i_1, \ldots, X_{m+n} = i_n) = p_{i_n, i_{n+1}}. \quad (5.7)$$

Indeed, using the definition of the conditional probability, we have

$$\begin{aligned} &P_i(X_{n+m+1} = i_{n+1}/X_m = i, X_{m+1} = i_1, \ldots, X_{m+n} = i_n) \\ &= \frac{P(X_{n+m+1} = i_{n+1}, X_{m+n} = i_n, \ldots, X_{m+1} = i_1, X_m = i)}{P(X_m = i, X_{m+1} = i_1, \ldots, X_{m+n} = i_n)} \\ &= \frac{p_{i,i_1} \cdots p_{i_n,i_{n+1}}}{p_{i,i_1} \cdots p_{i_{n-1},i_n}} = p_{i_n,i_{n+1}}. \end{aligned}$$

This proves the first statement concerning the process $\{X_{n+m}, n \geq 0\}$.
Let $D = \{X_0 = i_0, \ldots, X_m = i_m\}$. To prove the conditional independence, we have to check that for any set of the type

$$B = \{X_{n_j+m} = i_{n_j+m}, n_j \geq 1, j = 1, \ldots, k\},$$

the following property holds:

$$P_i(B \cap D) = P_i(B) P_i(D). \quad (5.8)$$

To simplify the notation and give an idea of the proof, we asume that $B = \{X_{n+m} = i_{n+m}\}$. If $i_m \neq i$, both terms of the previous equality are zero. Assume $i_m = i$. Then, by (5.5),

$$\begin{aligned} P_i(B \cap D) &= \frac{P(B \cap D)}{P(X_m = i)} \\ &= \frac{1}{P(X_m = i)} \sum_{i_{m+1}, \ldots i_{n+m-1}} \pi_{i_0} p_{i_0,i_1} \ldots p_{i_{m-1},i} p_{i,i_{m+1}} \ldots p_{i_{m+n-1},i_{n+m}}, \end{aligned}$$

while

$$\begin{aligned} P_i(D) &= \frac{1}{P(X_m = i)} \pi_{i_0} p_{i_0,i_1} \ldots p_{i_{m-1},i}, \\ P_i(B) &= \sum_{i_{m+1}, \ldots i_{n+m-1}} p_{i,i_{m+1}} \ldots p_{i_{m+n-1},i_{n+m}}. \end{aligned}$$

This finishes the proof of (5.8) in this particular case.

□

There is a stronger version of the previous proposition. In fact, the deterministic time m can be replaced by a stopping time (see Definition 3.5). This is called the *strong Markov property.*

Using an argument based on the *total probability principle* and then similar ideas as those of the proof of Proposition 5.1, we can prove the following Proposition. In the statement, T is a stopping time with respect to the natural filtration generated by X (see section 3.2).

Proposition 5.2 *(Strong Markov Property) Let $X = \{X_n, n \geq 0\}$ be a HMC with initial distribution ν and transition probability matrix Π. Conditionally on $T < \infty$ and $X_T = i$, the process $\{X_{n+T}, n \geq 0\}$ is a HMC, with transition probability matrix Π and initial distribution $\delta_{\{i\}}$, independent of the random variables $X_0, \cdots, X_T$.*

Proof: Set

$$P^*(\cdot) = P\left(\cdot / T < \infty, X_T = i\right).$$

Clearly

$$P^*(X_T = j) = \delta_{i,j}.$$

Fix arbitrary states $i, i_1, \ldots, i_n, i_{n+1}$. The definition of the conditional probability and the total probability principle yields

$$P^*\left(X_{T+n+1} = i_{n+1} / X_T = i, X_{T+1} = i_1, \ldots, X_{T+n} = i_n\right)$$

$$= \frac{P\left(X_{T+n+1} = i_{n+1}, X_T = i, X_{T+1} = i_1, \ldots, X_{T+n} = i_n, T < \infty\right)}{P\left(X_T = i, X_{T+1} = i_1, \ldots, X_{T+n} = i_n, T < \infty\right)}$$

$$= \frac{\sum_{m=1}^{\infty} P\left(T = m, X_{m+n+1} = i_{n+1}, X_m = i, X_{m+1} = i_1, \ldots, X_{m+n} = i_n\right)}{\sum_{m=1}^{\infty} P\left(T = m, X_m = i, X_{m+1} = i_1, \ldots, X_{m+n} = i_n\right)}.$$

By multiplying and dividing by $P(X_m = i)$, using that $(T = m) \in \mathcal{F}_m$ and owing to the previous Proposition 5.1, the previous expression is equal to

$$\frac{\sum_{m=1}^{\infty} P_i(T = m) P_i\left(X_{m+n+1} = i_{n+1}, X_m = i, X_{m+1} = i_1, \ldots, X_{m+n} = i_n\right)}{\sum_{m=1}^{\infty} P_i(T = m) P_i\left(X_m = i, X_{m+1} = i_1, \ldots, X_{m+n} = i_n\right)}$$

$$= \frac{\sum_{m=1}^{\infty} P_i(T = m) P_i(X_m = i) p_{i,i_1} \ldots p_{i_n,i_{n+1}}}{\sum_{m=1}^{\infty} P_i(T = m) P_i(X_m = i) p_{i,i_1} \ldots p_{i_{n_1},i_n}}$$

$$= p_{i_n,i_{n+1}}.$$

This proves the first statement.
The proof of the statement about independence is carried out in a similar way than the corresponding one in Proposition 5.1, by considering, as before, all the possible values of the stopping time T.

□

5.2 A Brief Analysis of the States of a Markov Chain

In this section, we study some possibilities for the Markov chain to wandering through the state space.

Definition 5.2 *(a) Given two states $i, j \in I$, we say that j is accessible from i if there exists a non-negative integer k such that $p_{i,j}^{(k)} > 0$.*
We shall write $i \to j$ for j being accessible from i.

(b) Two states $i, j \in I$ communicate if j is accessible from i and i is accessible from j. For two communicating states, $i, j \in I$, we shall write $i \leftrightarrow j$.

(c) A Markov chain is called irreducible if all states communicate.

The Bernoulli random walk on $\mathbb{Z}$ is an example of irreducible Markov chain. Communication between states establishes an equivalence relation on the set of states. Each equivalence class contains states that communicate to each other.

The next definition deals with the idea of the intensity that states are successively visited by the dynamical system given by the Markov chain.

Definition 5.3 *(a) A state $i \in I$ is termed recurrent if*

$$P\left(\limsup_n \{X_n = i\}\right) = 1.$$

Otherwise stated, the probability that $X_n = i$ for infinitely many values of n is one.

(b) A state $i \in I$ is termed transient if

$$P\left(\limsup_n \{X_n = i\}\right) = 0.$$

This can also be stated saying that

$$P\left(\liminf_n \{X_n \neq i\}\right) = 1.$$

Recurrent states are those which are being visited infinitely many times, while transient states are those which eventually are left forever, almost surely. From the above definition it does not clearly follow that both notions, recurrence and transience, are opposite to each other. However, we shall see later that this is actually the case.
In the analysis of these notions, the following definitions will be useful.

Definition 5.4 *(a) For a state $i \in I$, the first passage time is given by*

$$T_i = \inf\{n \geq 1 : X_n = i\},$$

with the usual convention that, if the above set is empty, $T_i = \infty$.

(b) The r–th passage time, $r \geq 2$ is defined by

$$T_i^{(r)} = \inf\{n \geq T_i^{(r-1)} + 1 : X_n = i\},$$

where $T_i^{(1)} = T_i$, and $T_i^{(0)} = 0$, by convention.

(c) The length of the r–th excursion to i is the stopping time defined by

$$S_i^{(r)} = \begin{cases} T_i^{(r)} - T_i^{(r-1)} & \text{if } T_i^{(r-1)} < \infty \\ 0 & \text{otherwise.} \end{cases}$$

By the strong Markov property, the r–th excursion is independent of the Markov chain until the $r-1$–th visit to i. More precisely, we have the following result.

Proposition 5.3 *Fix $r \geq 2$ and $i \in I$. Conditionally to $(T_i^{(r-1)} < \infty)$, the r–th excursion to i, $S_i^{(r)}$, is independent of the random variables $\{X_m, m \leq T_i^{(r-1)}\}$. Moreover, its probability law is the same as that of T_i, that is,*

$$P\left(S_i^{(r)} = n / T_i^{(r-1)} < \infty\right) = P\left(T_i = n / X_0 = i\right).$$

Proof: Set $T = T_i^{(r-1)}$. By definition, on $(T_i^{(r-1)} < \infty)$, $X_T = i$. By the strong Markov property, the process $\{X_{T+n}, n \geq 0\}$ is a Markov chain, with the same probability transitions matrix that the initial one and initial distribution given by $\delta_{\{i\}}$, independent of the random variables $X_0, \ldots, X_T$. This proves the first statement.
Concerning the second one, we notice that for the new chain $\{X_{T+n}, n \geq 0\}$, $S_i^{(r)}$ is the first passage time to i.

□

In the sequel, we shall use the notation $P_i(\cdot)$ for the conditional probability $P(\cdot / X_0 = i)$ and E_i for the conditional expectation with respect to P_i.
Since recurrence and transience are related to the number of passages to a state, it makes sense to study the random variable that describes this number. For this, we introduce the following notation:
The number of visits to $i \in I$ is

$$V_i = \sum_{n=0}^{\infty} 1_{\{X_n = i\}}.$$

Clearly,

$$E_i(V_i) = \sum_{n=0}^{\infty} P_i(X_n = i) = \sum_{n=0}^{\infty} p_{i,i}^{(n)}.$$

We also introduce the following notation:

$$f_i = P_i(T_i < \infty).$$

The quantity f_i is the probability that, starting from i, the chain visits i some time in the future. Notice that

1. The state i is recurrent if and only if $P_i(V_i = \infty) = 1$,
2. The state i is transient if and only if $P_i(V_i < \infty) = 1$.

With the quantities f_i, we can obtain the distribution function of the random variable V_i, as follows.

Proposition 5.4 *For $r = 0, 1, \ldots,$ $P_i(V_i > r) = f_i^r$.*

Proof: On the set $(X_0 = i)$, we have $(V_i > r) = (T_i^r < \infty)$. We will use this fact to prove the proposition recursively on r.
For $r = 0$, $T_i^{(0)} = 0$ and $P_i(V_i > 0) = 1$. Therefore the formula holds true. Assume that it is true for any integer less or equal to r. Then,

$$\begin{aligned} P_i(V_i > r+1) &= P_i\left(T_i^{(r+1)} < \infty\right) \\ &= P_i\left(T_i^{(r)} < \infty, S_i^{(r+1)} < \infty\right) \\ &= P_i\left(S_i^{(r+1)} < \infty / T_i^{(r)} < \infty\right) \quad P_i\left(T_i^{(r)} < \infty\right) \\ &= f_i f_i^r = f_i^{r+1}. \end{aligned}$$

□

With this proposition, we can give a characterization of recurrence and transicence.

Theorem 5.1 *1. If $P_i(T_i < \infty) = 1$, then the state i is recurrent and $\sum_{n=0}^{\infty} p_{i,i}^{(n)} = \infty$;*

2. If $P_i(T_i < \infty) < 1$, then the state i is transient and $\sum_{n=0}^{\infty} p_{i,i}^{(n)} < \infty$. Thus, each state is either recurrent or transient.

Proof: Assume first $P_i(T_i < \infty) = 1$. By Proposition 5.4 and the sequential continuity property of the probability,

$$\begin{aligned} P_i(V_i = \infty) &= \lim_{r\to\infty} P_i(V_i > r) \\ &= \lim_{r\to\infty} \left(P_i(T_i < \infty)\right)^r = 1. \end{aligned}$$

In this case,

$$\sum_{n=0}^{\infty} p_{i,i}^{(n)} = E_i(V_i) = \infty,$$

proving (1).
Assume now that $P_i(T_i < \infty) = f_i < 1$. Since V_i is a discrete random variable, its expectation can be computed by the formula

$$E_i(V_i) = \sum_{n=0}^{\infty} P_i(V_i > n).$$

Hence, owing to Proposition 5.4, one reaches

$$\begin{aligned} \sum_{n=0}^{\infty} p_{i,i}^{(n)} = E_i(V_i) &= \sum_{n=0}^{\infty} P_i(V_i > n) \\ &= \sum_{n=0}^{\infty} f_i^n = \frac{1}{1-f_i} < \infty. \end{aligned}$$

Since $E_i(V_i) < \infty$, $P_i(V_i = \infty) = 0$ and i is transient.

□

An easy consequence of the preceding theorem is that *recurrence* and *transcience* is a class property, in the following sense

Corollary 5.1 *All states in an equivalence class with respect to the equivalence relation given by the notion of communicating are either transient or recurrent.*

Proof: Let $C \subset I$ denote a generic equivalence class. Fix $i, j \in C$ and assume that i is transient. Let $n, m \geq 0$ be such that $p_{i,j}^{(n)} > 0$ and $p_{j,i}^{(m)} > 0$. Then for any $k \geq 0$,

$$p_{i,i}^{(n+m+k)} \geq p_{i,j}^{(n)} p_{j,j}^{(k)} p_{j,i}^{(m)}.$$

Consequently,

$$\sum_{n=0}^{\infty} p_{j,j}^{(k)} \leq \frac{1}{p_{i,j}^{(n)} p_{j,i}^{(m)}} \sum_{k=0}^{\infty} p_{i,i}^{(n+m+k)} < \infty.$$

By the previous theorem, j should be transient.

□

We finish this section stating two important properties without proofs.
A set $A \subset I$ is *closed* for a Markov chain $X = \{X_n, n \geq 1\}$ if, whenever the chain hits the set, no exit is possible. We also say that A is an absorbing set. This term is specially used when A reduces to a single element $i \in I$. That is, we say that i is an *absorbing state*.

1. Every recurrent class is closed.
2. Every finite closed equivalence class (with respect to communication) is recurrent.

5.3 Hitting Times

We recall the notion introduced in Example 3.5: For a stochastic process $X = \{X_n, n \geq 0\}$ and a set $A \subset I$, the hitting time of A by X is defined by

$$T_A = \inf\{n \geq 0 : X_n \in A\}.$$

By convention, if the set $\{n \geq 0 : X_n \in A\}$ is empty, $T_A = \infty$.
Hitting times are stopping times with respect to the natural filtration generated by the process

One of the most important applications of hitting times is related to the notion of *absorption* that we are now going to define.
Hitting times of absorbing sets are called absorbing times. For practical reasons, it is important to compute the probability that such absorption takes place at a finite time. More precisely, for $A \subset I$, set

$$h_i^A = P(T_A < \infty / X_0 = i).$$

The quantities $h_i^A, i \in I$, satisfy the linear system of equations

$$h_i^A = 1,\ i \in A \tag{5.9}$$

$$h_i^A = \sum_{j \in I} p_{i,j} h_j^A,\ i \notin A. \tag{5.10}$$

Indeed. The first equation in (5.9) is obvious, because if $X_0 = i \in A$, then $T_A = 0$ and consequently, $h_i^A = 1$.
Assume that $X_0 = i \notin A$. Then $T_A \geq 1$. By the Markov property,

$$P(T_A < \infty / X_1 = j, X_0 = i) = P(T_A < \infty / X_1 = j) = h_j^A.$$

By the principle of total probabilities, we have

$$\begin{aligned} h_i^A &= P(T_A < \infty / X_0 = i) \\ &= \sum_{j \in I} P(T_A < \infty, X_1 = j / X_0 = i) \\ &= \sum_{j \in I} \frac{P(T_A < \infty, X_1 = j, X_0 = i)}{P(X_0 = i))} \times \frac{P(X_1 = j, X_0 = i)}{P(X_1 = j, X_0 = i)} \\ &= \sum_{j \in I} P(T_A < \infty / X_1 = j) P(X_1 = j / X_0 = i) = \sum_{j \in I} p_{i,j} h_j^A, \end{aligned}$$

proving the second equation of (5.9).
A related interesting quantity is the mean value of the absorption time

$$k_i^A = E(T_A / X_0 = i).$$

Since T_A is a discrete random variable,

$$k_i^A = \sum_{n=0}^{\infty} n P(T_A = n / X_0 = i).$$

Following similar arguments as those leading to (5.9) we have that $k_i^A, i \in I$, satisfy the linear system of equations

$$k_i^A = 0,\ i \in A \tag{5.11}$$

$$k_i^A = 1 + \sum_{j \notin A} p_{i,j} k_j^A,\ i \notin A. \tag{5.12}$$

In fact, the first equation in (5.11) is trivial, because if $X_0 = i \in A$, then $T_A = 0$. The second one is proved as follows

$$\begin{aligned} k_i^A &= \sum_{n=0}^{\infty} nP(T_A = n/X_0 = i) \\ &= \sum_{n=0}^{\infty} n \sum_{j\in I} \frac{P(T_A = n, X_1 = j, X_0 = i)}{P(X_1 = j, X_0 = i)} \times \frac{P(X_1 = j, X_0 = i)}{P(X_0 = i)} \\ &= \sum_{j\in I} E(T_A/X_1 = j, X_0 = i))p_{i,j} \\ &= \sum_{j\in A} E(T_A/X_1 = j, X_0 = i))p_{i,j} + \sum_{j\notin A} E(T_A/X_1 = j, X_0 = i)p_{i,j} \\ &= 1 + \sum_{j\notin A} k_j^A p_{i,j}. \end{aligned}$$

To illustrate the preceding notions, let us consider the Markov chain associated to the gambler's ruin problem, that means a random walk with two absorbing states $0, N$, N being the total fortune of both gamblers. The state space is $I = \{0, 1, 2, \ldots, N\}$ and the transition probability matrix

$$\Pi = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & 1-p & 0 & p & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \tag{5.13}$$

We wish to solve the system (5.9) when $A = \{0\}$. For the sake of simplicity, we write h_i instead of h_i^A and $q = 1 - p$. Because of the particular form of Π, (5.9) reads

$$h_0 = 1 \tag{5.14}$$

$$h_i = ph_{i+1} + qh_{i-1},\ i = 1, 2, \ldots, N-1 \tag{5.15}$$

$$h_N = 0. \tag{5.16}$$

Notice that the second equation in (5.14) can be written equivalently

$$ph_{i+2} - h_{i+1} + qh_i = 0,\ i = 0, \ldots, N-2.$$

We solve these equations by trying solutions of the form $h_i = \lambda^i$, with $\lambda \in \mathbb{R} - \{0\}$, which amount to find values of λ satisfying

$$p\lambda^{i+2} - \lambda^{i+1} + q\lambda^i = 0,\ i = 0, \ldots, N-2. \tag{5.17}$$

There are two possibilities.
Case 1: $p \neq q$. There are two solution of (5.17), and there are $\lambda_1 = 1$, $\lambda_2 = \frac{q}{p}$. Then, a general solution to (5.14) is given by

$$h_i = A\lambda_1^i + B\lambda_2^i = A + B\left(\frac{q}{p}\right)^i, \tag{5.18}$$

$i = 1, \ldots, N-1$, with the boundary conditions

$$h_0 = A + B = 1$$
$$h_N = A + B\left(\frac{q}{p}\right)^N = 0,$$

yielding

$$B = \frac{1}{1 - \left(\frac{q}{p}\right)^N},$$
$$A = 1 - B.$$

Substituting these values of A and B into (5.18) yields

$$h_i = \frac{\left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^N},$$

$i = 1, \ldots, N-1$

Case 2: $p = q = \frac{1}{2}$. There is a unique solution of (5.17) and it is $\lambda = \frac{1}{2p} = 1$. A general solution to (5.14) is given by

$$h_i = A\lambda^i + Bi\lambda^i = A + Bi,$$

with the boundary conditions $h_0 = 1$, $h_N = 0$, which implies $A = 1$, $B = -\frac{1}{N}$. Consequently,

$$h_i = \frac{N - i}{N},$$

$i = 1, \ldots, N-1$.

5.4 Stationary Distributions

In this section, we develop a notion related with the limit behaviour of a Markov chain. In the next definition, $\nu = (\nu_i, i \in I)$ is a probability on I and Π a stochastic matrix indexed by I.
In matrix calculus, vectors will be written as column matrices; the notation ν^t means the transpose of ν, that means, a matrix consisting of one single row.

Definition 5.5 *A probability ν on is said to be invariant or stationary for* Π *if the following identity holds:*

$$\nu^t = \nu^t \Pi. \tag{5.19}$$

It is very easy to check recursively that the identity (5.19) is equivalent to

$$\nu^t = \nu^t \Pi^n.$$

Along this section, we shall denote by $\{X_n, n \geq 0\}$ a Markov chain with transition probability matrix Π.
Assume that the law of X_0 is a stationary distribution ν. Then, each random variable X_n of the process has the same distribution and this is ν. Indeed, by (5.5)

$$P(X_n = j) = (\nu^t \Pi^n)_j = \nu^t.$$

Assume that I is finite, say $I = \{1, 2, \ldots, N\}$. The equation (5.19) defining invariance is equivalent to

$$\nu^t (I - \Pi) = 0 \tag{5.20}$$

I meaning the N-dimensional identity matrix. Since Π is a stochastic matrix, we clearly have $\det(I - \Pi) = 0$. Hence the linear system (5.20) has at least a solution. Actually the null vector is a solution. But this is not the solution we are looking for, since we are interested in solutions defining a probability. Notice also that (5.20) tell us that ν must be an eigenvector of Π with one as eigenvalue. Assume that Π is symmetric. Then, since it is a stochastic matrix, one can easily check that $\mu = (1, 1, \ldots, 1)$ is one of these eigenvectors. Of course μ does not define a probability, but it suffices to normalize by N to get one. More precisely, in the finite case,

$$\nu = \left(\frac{1}{N}, \ldots, \frac{1}{N}\right)$$

provides an invariant probability for any stochastic matrix Π.
Using a more sophisticated argument based on compactness one can prove the existence of invariant probability as follows.
Let v be a probability on I. For any $n \geq 1$, set

$$v_n^t = \frac{1}{n} \sum_{k=0}^{n-1} v^t \Pi^k.$$

One can easily prove that v_n defines a probability on I. Moreover,

$$v_n^t - v_n^t \Pi = \frac{1}{n}(v^t - v^t \Pi^n).$$

Then, an invariant probability is obtained by the limit (which exists by compactness, since I is finite) of some subsequence $(v_{n_k}, k \geq 1)$ of $(v_n, n \geq 1)$.
The previous arguments cannot be extended to countable sets I.
There is no uniqueness of invariant probabilities. For example, if Π is the identity matrix, any probability ν is invariant. Moreover, given two invariant probabilities ν_i, $i = 1, 2$, any linear convex combination $\lambda\nu_1 + (1 - \lambda)\nu_2$, $\lambda \in [0, 1]$, is also an invariant probability.

Example 5.2 *Consider the random walk with absorbing states $0, N$. The transition probability matrix is given by (5.13). The finite dimensional linear system $\nu^t = \nu^t\Pi$ can be written coordinatewise as follows:*

$$\begin{aligned}
\nu_0 &= \nu_0 + \nu_1 q \\
\nu_1 &= \nu_2 q \\
\nu_j &= \nu_{j-1}p + \nu_{j+1}q,\ j = 2, \ldots, N-2 \\
\nu_{N-1} &= \nu_{N-2}p \\
\nu_N &= \nu_{N-1}p + \nu_N.
\end{aligned}$$

From the last equation, we obtain $\nu_{N-1} = 0$. Substituting this value in the equation before the last one, we obtain $\nu_{N-2} = 0$. Proceeding further in the same way yields

$$\nu_1 = \nu_2 = \cdots = \nu_{N-1} = 0.$$

Finally, from the first and last equation, we obtain that ν_0 and ν_N can be any numbers in $[0, 1]$ satisfying $\nu_0 + \ u_N = 1$.
Consequently, there exist -but there is no uniqueness- invariant probabilities, and they are given by

$$(\lambda, 0, \ldots, 0, 1 - \lambda),$$

with $\lambda \in [0, 1]$.

5.5 Limiting Distributions

The existence of invariant probability is related to the existence of limiting distributions. This section is devoted to study more closely this fact. The main result (see Theorem 5.2) concerns the particular case I is finite.
Let I be countable. Assume that there exists $i \in I$ such that for any $j \in I$, the limits

$$\lim_{n\to\infty} p_{i,j}^{(n)}$$

exist and do not depend on i. That is

$$\lim_{n\to\infty} p_{i,j}^{(n)} = \pi_j. \tag{5.21}$$

Then, the limit vector $\pi = (\pi_j, j \in I)$ defines an invariant probability. Moreover, if the assumption (5.21) holds for any $i \in I$, then this is the unique invariant probability.
Indeed, let us first prove that π is a probability on I. By its very definition

$$\sum_{j\in I} \pi_j = \sum_{j\in I} \lim_{n\to\infty} p_{i,j}^{(n)} = \lim_{n\to\infty} \sum_{j\in I} p_{i,j}^{(n)} = 1.$$

In addition,

$$\begin{aligned}\pi_j &= \lim_{n\to\infty} p_{i,j}^{(n+1)} = \lim_{n\to\infty} \sum_{k\in I} p_{i,k}^{(n)} p_{k,j} \\ &= \sum_{k\in I} \lim_{n\to\infty} p_{i,k}^{(n)} p_{k,j} = \sum_{k\in I} \pi_k p_{k,j},\end{aligned}$$

proving the invariance.
Let us now prove uniqueness Let $\tilde{\pi}$ be an invariant probability for Π. That is

$$\tilde{\pi}_j = \sum_{k\in I} \tilde{\pi}_k p_{k,j}^{(n)}.$$

Taking limits as $n \to \infty$ in both sides of the preceding equality, we have

$$\tilde{\pi}_j = \sum_{k\in I} \tilde{\pi}_k (\lim_{n\to\infty}) p_{k,j}^{(n)} = \pi_j \sum_{k\in I} \tilde{\pi}_k = \pi_j.$$

We remark that in the above arguments, we could exchange sums and limits by monotone convergence.
There are simple examples of Markov chains for which there is no limiting distribution. In fact, consider the *deterministic* Markov chain whose associated transition probability matrix is given by

$$\Pi = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Simple computations show that $\Pi^{2n} = I$, while $\Pi^{2n+1} = \Pi$. Hence, the sequence $p_{i,j}^{(n)}$, for fixed i, j takes either value 0 or 1 and therefore, cannot converge.

The next theorem gives the existence of limiting distributions for Markov chains with a *finite* number of states. Let us first introduce some notation: A Markov chain is called *regular* if its associated transition probability matrix satisfies the following property
(R) There exists a natural number n_0 such that

$$\min_{i,j\in I} p_{i,j}^{(n_0)} > 0. \tag{5.22}$$

Theorem 5.2 *1. Assume that the Markov chain is finite and regular. There exists a probability $\pi = (\pi_j, j \in I)$, with $\pi_j \in (0,1]$ for any $j \in I$ such that*

$$\lim_{n\to\infty} p_{i,j}^n = \pi_j, \text{ for all } \quad i \in I. \tag{5.23}$$

and this probability is stationary for Π.

2. Conversely, if there exists a probability $\pi = (\pi_j, j \in I)$, with $\pi_j \in (0,1)$ satisfying (5.23) then **(R)** *holds.*

Proof: The probability π is given in a constructive way as follows. For any $j \in I$, we set

$$m_j^{(n)} = \min_{i\in I} p_{i,j}^{(n)}, \quad M_j^{(n)} = \max_{i\in I} p_{i,j}^{(n)}.$$

We clearly have

$$p_{i,j}^{(n+1)} = \sum_{k\in I} p_{i,k} p_{k,j}^{(n)}. \tag{5.24}$$

Consequently,

$$\begin{aligned} m_j^{(n+1)} &= \min_{i\in I} p_{i,j}^{(n+1)} = \min_{i\in I} \sum_{k\in I} p_{i,k} p_{k,j}^{(n)} \\ &\geq \min_{i\in I} \sum_{k\in I} p_{i,k} \min_{k\in I} p_{k,j}^{(n)} = \min_{k\in I} p_{k,j}^{(n)} = m_j^{(n)}. \end{aligned}$$

Similarly,

$$M_j^{(n)} \geq M_j^{(n+1)},\ n \geq 1. \tag{5.25}$$

To prove (5.23), it suffices to establish that

$$\lim_{n\to\infty} (M_j^{(n)} - m_j^{(n)}) = 0,\ j \in I. \tag{5.26}$$

Indeed, we have just checked that $(m_j^{(n)}, n \geq 1)$ increases and $(M_j^{(n)}, n \geq 1)$ decreases. Since both sequences are bounded, their respective limits as $n \to \infty$ do exist. Let us call them m_j and M_j, respectively. If (5.26) holds then necessarily, $m_j = M_j$. But, by definition

$$m_j^{(n)} \leq p_{i,j}^{(n)} \leq M_j^{(n)},$$

and this implies (5.23) with $\pi_j = m_j = M_j$, $j \in I$.
To prove (5.26), set $\epsilon = \min_{i,j\in I} p_{i,j}^{(n_0)}$, a strictly positive number by assumption (**R**). For any $n \geq 1$,

$$\begin{aligned} p_{i,j}^{(n_0+n)} &= \sum_{k\in I} p_{i,k}^{(n_0)} p_{k,j}^{(n)} = \sum_{k\in I} (p_{i,k}^{(n_0)} - \epsilon p_{j,k}^{(n)}) p_{k,j}^{(n)} \\ &\quad \epsilon \sum_{k\in I} p_{j,k}^{(n)} p_{k,j}^{(n)} \\ &= \sum_{k\in I} (p_{i,k}^{(n_0)} - \epsilon p_{j,k}^{(n)}) p_{k,j}^{(n)} + \epsilon p_{j,j}^{(2n)}. \end{aligned}$$

Since $\epsilon = \min_{i,j\in I} p_{i,j}^{(n_0)} \leq p_{i,k}^{(n_0)}$ and $p_{j,k}^{(n)} \leq 1$, we have that $p_{i,k}^{(n_0)} - \epsilon p_{j,k}^{(n)} \geq 0$. Therefore,

$$\begin{aligned} p_{i,j}^{(n_0+n)} &\geq m_j^{(n)} \sum_{k\in I} (p_{i,k}^{(n_0)} - \epsilon p_{j,k}^{(n)}) + \epsilon p_{j,j}^{(2n)} \\ &= m_j^{(n)}(1-\epsilon) + \epsilon p_{j,j}^{(2n)}. \end{aligned}$$

Therefore,

$$m_j^{(n_0+n)} \geq m_j^{(n)}(1-\epsilon) + \epsilon p_{j,j}^{(2n)}.$$

With similar arguments, we can also prove that

$$M_j^{(n_0+n)} \leq M_j^{(n)}(1-\epsilon) + \epsilon p_{j,j}^{(2n)}.$$

Combining both inequalities yields

$$M_j^{(n_0+n)} - m_j^{(n_0+n)} \leq (M_j^{(n)} - m_j^{(n)})(1-\epsilon),$$

and by iteration

$$0 \leq M_j^{(kn_0+n)} - m_j^{(kn_0+n)} \leq (M_j^{(n)} - m_j^{(n)})(1-\epsilon)^k.$$

The last expression tends to zero as $k \to \infty$, because $\epsilon > 0$. Thus, we have proved the existence of a subsequence of $(M_j^{(n)} - m_j^{(n)}, n \geq 1)$ converging to zero. But this sequence is monotone; hence (5.26) holds true.
For any $n \geq n_0$, $m_j^{(n)} \geq m_j^{(n_0)} = \epsilon > 0$. This implies

$$\pi_j := \lim_{n\to\infty} m_j^{(n)} \geq \epsilon > 0.$$

From (5.20), it follows that

$$1 = \sum_{j\in I} p_{i,j}^{(n)} \to_{n\to\infty} \sum_{j\in I} \pi_j.$$

The fact that a limiting distribution is invariant has already been proved in the preceding section. Consequently, the proof of the first part of the theorem is complete.

Let us now prove part 2 of the statement. From (5.23) is follows that, for any $j \in I$ there exist n_j and for any $n \geq n_j$, $p_{i,j}^{(n)} > 0$, for any $i \in I$. Hence, $\min_{i\in I} p_{i,j}^{(n_j)} > 0$. Set $n_0 = \max(n_j, j \in I)$. Clearly, this yields property (**R**). □

After having proved the theorem, a natural question to ask is whether there are simple conditions on the process ensuring property (**R**). The next Proposition gives an answer to this question.

Proposition 5.5 *Let $\{X_n, n \geq 0\}$ be a finite irreducible Markov chain. Assume that there exists $h \in I$ such that $p_{h,h} > 0$. Then (**R**) is satisfied.*

Proof: Fix $i, j \in I$ and let $n(i,j) \geq 0$ be such that $p_{i,j}^{(n(i,j))} > 0$. Set $m = \max_{i,j \in I} n(i,j)$. The matrix Π^{2m+1} has all its entries strictly positive. Indeed,

$$p_{i,j}^{(2m+1)} \geq p_{i,h}^{n(i,h)} p_{h,h} \cdot \cdots \cdot p_{h,h} p_{h,j}^{n(h,j)} > 0,$$

where in the second term of the preceding inequality, we have written $2m + 1 - n(i,h) - n(h,j)$ many factors $p_{h,h}$. □

6 Brownian motion

In example 2.3 of Chapter 2, we have introduced an example of Gaussian continuous time stochastic process. In this chapter we will study some of the most important properties of this process in view of the introduction to Itô's stochastic calculus developed later in this course.

6.1 Study of The Probability Law

Let us start by recalling the definition.

Definition 6.1 *The Brownian motion or Wiener process is a Gaussian stochastic process* $\{B_t, t \geq 0\}$ *such that*

$$E(B_t) = 0,$$
$$E(B_s B_t) = s \wedge t.$$

As mentioned in Chapter 2, the existence of this processes is ensured by Kolmogorov's theorem. For this, se have to make clear that the proposal we made for the covariance is correct. This means the following. A covariance function of a stochastic processes is a mapping

$$(s,t) \to \Gamma(s,t)$$

which is required to be nonnegative definite. That means, for any $t_i, t_j \geq 0$ and any real numbers a_i, a_j, $i,j = 1,\ldots,m$,

$$\sum_{i,j=1}^{m} a_i a_j \Gamma(t_i,t_j) \geq 0.$$

It is not difficult to check that the function $\Gamma(s,t) = s \wedge t$ possesses this property. Indeed, notice first that

$$s \wedge t = \int_0^\infty 1_{[0,s]}(r)\, 1_{[0,t]}(r)\, dr.$$

Hence,

$$\begin{aligned}\sum_{i,j=1}^{m} a_i a_j t_i \wedge t_j &= \sum_{i,j=1}^{m} a_i a_j \int_0^\infty 1_{[0,t_i]}(r)\, 1_{[0,t_j]}(r)\, dr \\ &= \int_0^\infty \left(\sum_{i=1}^{m} a_i 1_{[0,t_i]}(r)\right)^2 dr \geq 0.\end{aligned}$$

Since $E(B_0^2) = 0$, the random variable B_0 is zero almost surely.
Using the formula on transformations of densities of random vectors by smooth functions, it is possible to obtain the density of the random vector

$$B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_m} - B_{t_{m-1}},$$

for any choice $0 \leq t_1 \leq \cdots \leq t_m$ and to see that it corresponds to that of m independent random variables, Gaussian, centered, with variance $t_i - t_{i-1}$, $i = 1, \ldots, m$. Hence, Brownian motion has independent and stationary increments.
Each random variable B_t, $t > 0$, of the Brownian motion has a density and it is

$$p_t(x) = \frac{1}{\sqrt{2\pi t}} \exp(-\frac{x^2}{2t}),$$

while for $t = 0$, its "density" is a Dirac mass at zero, $\delta_{\{0\}}$.
Differentiating $p_t(x)$ once with respect to t, and then twice with respect to x easily yields

$$\begin{aligned} \frac{\partial}{\partial t} p_t(x) &= \frac{1}{2} \frac{\partial^2}{\partial x^2} p_t(x) \\ p_0(x) &= \delta_{\{0\}}. \end{aligned}$$

This is the heat equation on $\mathbb{R}$ with initial condition $p_0(x) = \delta_{\{0\}}$. That means, as time evolves, the density of the random variables of the Brownian motion behaves like a diffusive physical phenomenon.

6.2 Sample Paths

Brownian motion was introduced as a model for erratic trajectories of particles. Thus, one expects to be able to prove that its sample paths are almost surely continuous, but with brusque changes in directions of its trajectories. This section is devoted to give some elements towards a good understanding of these properties.

The continuity of the sample paths of Brownian motion can be proved by different methods. It can be obtained as a by-product of an explicit construction. It can also be proved using Kolmogorov's continuity criterion, a result which allows to catch the roughness of sample paths from $L^p(\Omega)$ estimates of increments of the process at different times.

Brownian motion as limit of a random walk
Let $\{\xi_j, j \in \mathbb{N}\}$ be a sequence of independent, identically distributed random variables, with mean zero and variance $\sigma^2 > 0$. Consider the sequence of

partial sums defined by $S_0 = 0$, $S_n = \sum_{j=1}^{n} \xi_j$. We have already seen in previous chapters that the sequence $\{S_n, n \geq 0\}$ is a Markov chain, and also a martingale.
Let us consider the continuous time stochastic process defined by linear interpolation of $\{S_n, n \geq 0\}$, as follows. For any $t \geq 0$, let $[t]$ denote its integer value. Then set

$$Y_t = S_{[t]} + (t - [t])\xi_{[t]+1}, \tag{6.1}$$

for any $t \geq 0$.
The next step is to scale the sample paths of $\{Y_t, t \geq 0\}$. By analogy with the scaling in the statement of the central limit theorem, we set

$$B_t^{(n)} = \frac{1}{\sigma\sqrt{n}} Y_{nt}, \tag{6.2}$$

$t \geq 0$.
A famous result in probability theory -Donsker theorem- tell us that the sequence of processes $B_t^{(n)}, t \geq 0\}$, $n \geq 1$, converges in law to the Brownian motion. The reference sample space is the set of continuous functions vanishing at zero. Hence, proving the statement, we obtain continuity of the sample paths of the limit.
Donsker theorem is the infinite dimensional version of the above mentioned central limit theorem. Considering $s = \frac{k}{n}$, $t = \frac{k+1}{n}$, the increment $B_t^{(n)} - B_s^{(n)} = \frac{1}{\sigma\sqrt{n}}\xi_{k+1}$ is a random variable, with mean zero and variance $t - s$. Hence $B_t^{(n)}$ is not that far from the Brownian motion, and this is what Donsker's theorem proves.

Kolmogorov's continuity criterion: Application to the Brownian motion

Proposition 6.1 *Let $\{X_t, t \geq 0\}$ be a stochastic process satisfying the following property: For some positive real numbers α, β and C,*

$$E\left(|X_t - X_s|^\alpha\right) \leq C|t - s|^{1+\beta}.$$

Then almost surely, the sample paths of the process are γ-Hölder continuous with $\gamma \leq \frac{\beta}{\alpha}$.

The law of the random variable $B_t - B_s$ is $\mathrm{N}(0, t - s)$. Thus, it is possible to compute the moments, and we have

$$E\left((B_t - B_s)^{2k}\right) = \frac{(2k)!}{2^k k!}(t - s)^k,$$

for any $k \in \mathbb{N}$. Therefore, Proposition 6.1 yields that almost surely, the sample paths of the Brownian motion are γ–Hölder continuous with $\gamma \in (0, \frac{1}{2})$.

Quadratic variation

The notion of quadratic variation provides a measure of the roughness of a function. Existence of variations of different orders are also important in procedures of approximation via a Taylor expansion and also in the development of infinitesimal calculus. We will study here the existence of quadratic variation, i.e. variation of order two, for the Brownian motion. As shall be explained in more detail in the next chapter, this fact provides the explanation to the fact that rules of Itô's stochastic calculus are different from those of the classical differential deterministic calculus.

Fix a finite interval $[0, T]$ and consider the sequence of partitions given by the points $(t_j^n = \frac{jT}{n}, j = 0, 1, \ldots, n)$. Set $\Delta_k B = B_{t_k^n} - B_{t_{k-1}^n}$.

Proposition 6.2 *The sequence* $\{\sum_{k=1}^n (\Delta_k B)^2, n \geq 1\}$ *converges in* $L^2(\Omega)$ *to the deterministic random variable* T*. That is,*

$$\lim_{n\to\infty} E\left[\left(\sum_{k=1}^n (\Delta_k B)^2 - T\right)^2\right] = 0.$$

Proof: For the sake of simplicity, we shall omit the dependence on n. Set $\Delta_k t = t_k - t_{k-1}$. Notice that the random variables $(\Delta_k B)^2 - \Delta_k t$, $k = 1, \ldots, n$, are independent and centered. Thus,

$$\begin{aligned} E\left[\left(\sum_{k=1}^n (\Delta_k B)^2 - T\right)^2\right] &= E\left[\left(\sum_{k=1}^n \Big((\Delta_k B)^2 - \Delta_k t\Big)\right)^2\right] \\ &= \sum_{k=1}^n E\left[\Big((\Delta_k B)^2 - \Delta_k t\Big)^2\right] \\ &= \sum_{k=1}^n \Big[3(\Delta_k t)^2 - 2(\Delta_k t)^2 + (\Delta_k t)^2\Big] \\ &= 2\sum_{k=1}^n (\Delta_k t)^2 \leq 2\frac{T}{n}, \end{aligned}$$

which clearly tends to zero as n tends to infinity. □

This proposition, together with the continuity of the sample paths of Brownian motion yields

$$\sup_n \sum_{k=1}^n |\Delta_k B| = \infty, \text{a.}s.$$

Indeed, assume that $V := \sup_n \sum_{k=1}^n |\Delta_k B| < \infty$. Then

$$\begin{aligned}\sum_{k=1}^n (\Delta_k B)^2 &\leq \sup_k |\Delta_k B| \left(\sum_{k=1}^n |\Delta_k B| \right) \\ &\leq V \sup_k |\Delta_k B|.\end{aligned}$$

We obtain $\lim_{n\to\infty} \sum_{k=1}^n (\Delta_k B)^2 = 0$. a.s., which contradicts the result proved in Proposition 6.2.

6.3 The Martingale Property of Brownian Motion

We start this section by extending the definition of martingale given in Section 3.2 to continuous time stochastic processes. First, we introduce the appropriate notion of filtration, as follows.
A family $\{\mathcal{F}_t, t \geq 0\}$ of sub σ–fields of $\mathcal{F}$ is termed a *filtration* if

1. $\mathcal{F}_0$ contains all the sets of $\mathcal{F}$ of null probability,
2. For any $0 \leq s \leq t$, $\mathcal{F}_s \subset \mathcal{F}_t$.

If in addition

$$\cap_{s>t} \mathcal{F}_s = \mathcal{F}_t,$$

for any $t \geq 0$, the filtration is said to be *right-continuous*.

Definition 6.2 *A stochastic process $\{X_t, t \geq 0\}$ is a martingale with respect to the filtration $\{\mathcal{F}_t, t \geq 0\}$ if each variable belongs to $L^1(\Omega)$ and moreover*

1. *X_t is $\mathcal{F}_t$–measurable for any $t \geq 0$,*
2. *for any $0 \leq s \leq t$, $E(X_t/\mathcal{F}_s) = X_s$.*

If the equality in (2) is replaced by $\leq$ (respectively, $\geq$), we have a supermartingale (respectively, a submartingale).
As for discrete parameter sets, given a stochastic process $\{X_t, t \geq 0\}$, there is a natural way to define a filtration by considering

$$\mathcal{F}_t = \sigma(B_s, 0 \leq s \leq t),\ t \geq 0.$$

To ensure that the above property (1) holds, one needs to complete the σ-field. In general, there is no reason to expect right-continuity. However, for the Brownian motion, the natural filtration possesses this property.
In example 3.3, we have seen that the sequence of partial sums of independent centered random variables with zero mean is a martingale with respect to

the natural filtration. The reason for this to be true relies on the fact that the increments of the partial sums are independent random variables. In continuous time, the property of independent increments allows to prove the martingale property with respect to the natural filtration as well.
Indeed, fix $0 \leq s \leq t$. Owing to the property (c) of the conditional expectation

$$E(X_t - X_s/\mathcal{F}_s) = E(X_t - X_s) = 0.$$

Hence, a Brownian motion possess the martingale property with respect to the natural filtration.

Other examples of martingales with respect to the same filtration, related with the Brownian motion are

1. $\{B_t^2 - t, t \geq 0\}$,
2. $\{\exp\left(aB_t - \frac{a^2 t}{2}\right), t \geq 0\}$.

Indeed, for the first example, let us consider $0 \leq s \leq t$. Then,

$$\begin{aligned} E\left(B_t^2/\mathcal{F}_s\right) &= E\left((B_t - B_s + B_s)^2/\mathcal{F}_s\right) \\ &= E\left((B_t - B_s)^2/\mathcal{F}_s\right) + 2E\left((B_t - B_s)B_s/\mathcal{F}_s\right) \\ &\quad + E\left(B_s^2/\mathcal{F}_s\right). \end{aligned}$$

Since $B_t - B_s$ is independent of $\mathcal{F}_s$, owing to the properties of the conditional expectation, we have

$$\begin{aligned} E\left((B_t - B_s)^2/\mathcal{F}_s\right) &= E\left((B_t - B_s)^2\right) = t - s, \\ E\left((B_t - B_s)B_s/\mathcal{F}_s\right) &= B_s E\left(B_t - B_s/\mathcal{F}_s\right) = 0, \\ E\left(B_s^2/\mathcal{F}_s\right) &= B_s^2. \end{aligned}$$

Consequently,

$$E\left(B_t^2 - B_s^2/\mathcal{F}_s\right) = t - s.$$

For the second example, we also use the property of independent increments, as follows:

$$\begin{aligned} E\left(\exp\left(aB_t - \frac{a^2 t}{2}\right)/\mathcal{F}_s\right) &= \exp(aB_s) E\left(\exp\left(a(B_t - B_s) - \frac{a^2 t}{s}\right)/\mathcal{F}_s\right) \\ &= \exp(aB_s) E\left(\exp\left(a(B_t - B_s) - \frac{a^2 t}{s}\right)\right). \end{aligned}$$

Using the density of the random variable $B_t - B_s$ one can easily check that

$$E\left(\exp\left(a(B_t - B_s) - \frac{a^2 t}{s}\right)\right) = \exp\left(\frac{a^2(t-s)}{2} - \frac{a^2 t}{2}\right).$$

Therefore, we obtain

$$E\left(\exp\left(aB_t - \frac{a^2 t}{2}\right)/\mathcal{F}_s\right) = \exp\left(aB_s - \frac{a^2 s}{2}\right).$$

6.4 Markov Property

For any $0 \leq s \leq t$, $x \in \mathbb{R}$ and $A \in \mathcal{B}(\mathbb{R})$, we set

$$p(s,t,x,A) = \frac{1}{(2\pi(t-s))^{\frac{1}{2}}} \int_A \exp\left(-\frac{|x-y|^2}{2(t-s)}\right) dy. \tag{6.3}$$

Actually, $p(s,t,x,A)$ is the probability that a random variable, Normal, with mean x and variance $t-s$ take values on a fixed set A.
Let us prove the following identity:

$$P\{B_t \in A \,/\, \mathcal{F}_s\} = p(s,t,B_s,A), \tag{6.4}$$

which means that, conditionally to the past of the Brownian motion until time s, the law of B_t at a future time t only depends on B_s.
Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded measurable function. Then, since B_s is $\mathcal{F}_s$–measurable and $B_t - B_s$ independent of $\mathcal{F}_s$, we obtain

$$\begin{aligned} E\left(f(B_t)/\,\mathcal{F}_s\right) &= E\left(f\left(B_s + (B_t - B_s)\right)/\,\mathcal{F}_s\right) \\ &= E\left(f(x + B_t - B_s)\right)\Big|_{x=B_s}. \end{aligned}$$

The random variable $x + B_t - B_s$ is $\mathrm{N}(x, t-s)$. Thus,

$$E\left(f(x + B_t - B_s)\right) = \int_{\mathbb{R}} f(y) p(s,t,x,dy),$$

and consequently,

$$E\left(f(B_t)/\,\mathcal{F}_s\right) = \int_{\mathbb{R}} f(y) p(s,t,B_s,dy).$$

This yields (6.4) by taking $f = 1_A$.
Going back to (6.3), we notice that the function $x \to p(s,t,x,A)$ is measurable, and the mapping $A \to p(s,t,x,A)$ is a probability.

Let us prove the additional property, called Chapman-Kolmogorov equation: For any $0 \leq s \leq u \leq t$,

$$p(s,t,x,A) = \int_{\mathbb{R}} p(u,t,y,A)p(s,u,x,dy). \tag{6.5}$$

We recall that the sum of two independent Normal random variables, is again Normal, with mean the sum of the respective means, and variance the sum of the respective variances. This is expressed in mathematical terms by the fact that

$$\begin{aligned} f_{\mathrm{N}(x,\sigma_1)} * f_{\mathrm{N}(y,\sigma_2)} &= \int_{\mathbb{R}} f_{\mathrm{N}(x,\sigma_1)}(y) f_{\mathrm{N}(y,\sigma_2)}(\cdot - y)dy \\ &= f_{\mathrm{N}(x+y,\sigma_1+\sigma_2)}. \end{aligned}$$

Using this fact, we obtain

$$\begin{aligned} \int_{\mathbb{R}} p(u,t,y,A)p(s,u,x,dy) &= \int_A dz \left(f_{\mathrm{N}(x,u-s)} * f_{\mathrm{N}(0,t-u)} \right)(z) \\ &= \int_A dz f_{\mathrm{N}(x,t-s)}(z) = p(s,t,x,A). \end{aligned}$$

proving (6.5).
This equation is the time continuous analogue of the property own by the transition probability matrices of a Markov chain. That is,

$$\Pi^{(m+n)} = \Pi^{(m)}\Pi^{(n)},$$

meaning that evolutions in $m+n$ steps are done by concatenating m-step and n-step evolutions. In (6.5) $m+n$ is replaced by the real time $t-s$, m by $t-u$, and n by $u-s$, respectively.
We are now prepared to give the extension of the Markov property introduced in Chapter 5 for continuous time stochastic processes.
Consider a mapping

$$p : \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R} \times \mathcal{B}(\mathbb{R}) \to \mathbb{R}_+,$$

satisfying the properties

(i) for any fixed $s,t \in \mathbb{R}_+$, $A \in \mathcal{B}(\mathbb{R})$,

$$x \to p(s,t,x,A)$$

is $\mathcal{B}(\mathbb{R})$–measurable,

(ii) for any fixed $s, t \in \mathbb{R}_+$, $x \in \mathbb{R}$,

$$A \to p(s, t, x, A)$$

is a probability

(iii) Equation (6.5) holds.

Such a function p is termed a Markovian transition function.
Let us also fix a probability μ on $\mathcal{B}(\mathbb{R})$.

Definition 6.3 *A real valued stochastic process $\{X_t, t \in \mathbb{R}_+\}$ is a Markov process with initial law μ and transition probability function p if*

(a) the law of X_0 is μ,

(b) for any $0 \leq s \leq t$,

$$P\{X_t \in A / \mathcal{F}_s\} = p(s, t, X_s, A).$$

Therefore, we have proved that the Brownian motion is a Markov process with initial law a Dirac delta function at 0 and transition probability function p the one defined in (6.3).

7 Basic Notions on Itô's Calculus

Itô's calculus has been developed in the 50' by Kyoshi Itô in an attempt to give rigourous meaning to some differential equations driven by the Brownian motion appearing in the study of some problems related with continuous time Markov processes. Roughly speaking, one could say that Itô's calculus is an analogue of the classical Newton and Leibniz calculus for stochastic processes. In fact, in deterministic mathematical analysis, there are several extensions of the Riemann integral $\int f(x)dx$. For example, if g is an increasing bounded function (or the difference of two of this class of functions), Lebesgue-Stieltjes integral gives a precise meaning to the integral $\int f(x)g(dx)$, for some set of functions f. However, before Itô's development, no theory allowing nowhere differentiable integrators g was known. Brownian motion, introduced in the preceding chapter, is an example of stochastic process whose sample paths, although continuous, are nowhere differentiable. Therefore, in the framework of Lebesgue-Stieltjes theory it is not possible to give a rigourous meaning to stochastic integration with respect to Brownian motion.
There are many motivations coming from a variety of sciences for considering stochastic differential equations driven by a Brownian motion. Such an object is defined as

$$\begin{aligned} dX_t &= \sigma(t, X_t)dB_t + b(t, X_t)dt, \\ X_0 &= x_0, \end{aligned}$$

or in integral form,

$$X_t = x_0 + \int_0^t \sigma(s, X_s)dB_s + \int_0^t b(s, X_s)ds. \tag{7.1}$$

The first notion to be introduced is that of *stochastic integral*. In fact, in (7.1) the integral $\int_0^t b(s, X_s)ds$ can be defined pathwise, but this is not the case for $\int_0^t \sigma(s, X_s)dB_s$, because of the roughness of the paths of the stochastic integral we have just mentioned. More explicitely, it is not possible to fix $\omega \in \Omega$, then to consider the path $\sigma(s, X_s(\omega))$, and finally to integrate with respect to $B_s(\omega)$.

7.1 Itô's Integral

Along this section, we will consider a Brownian motion $B = \{B_t, t \geq 0\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$. We also will consider the natural filtration $(\mathcal{F}_t, t \geq 0)$ associated with B.
We fix a finite time horizon T and define $L^2_{a,T}$ the set of stochastic processes $u = \{u_t, t \in [0, T]\}$ satisfying the following conditions:

(i) u is adapted and jointly measurable in (t,ω), with respect to the product σ-field $\mathcal{B}([0,T])\times\mathcal{F}$.

(ii) $\int_0^T E(u_t^2)dt<\infty$.

The notation $L^2_{a,T}$ gathers the two properties -adaptedness and square integrability- described before.
Consider first the subset of $L^2_{a,T}$ consisting of *step processes*. That is, stochastic processes which can be written as

$$u_t=\sum_{j=1}^{n}u_j\mathbf{1}_{]t_{j-1},t_j[}(t), \tag{7.2}$$

with $0=t_0\leq t_1\leq\cdots\leq t_n=T$ and where u_j, $j=1,\ldots,n$, are $\mathcal{F}_{t_{j-1}}$– measurable square integrable random variables. We shall denote by $\mathcal{E}$ the set of these processes.
For step processes, the Itô stochastic integral is defined by the very natural formula

$$\int_0^T u_t dB_t=\sum_{j=1}^{n}u_j(B_{t_j}-B_{t_{j-1}}). \tag{7.3}$$

Notice that $\int_0^T u_t dB_t$ is a random variable. Of course, we would like to be able to consider more general integrands than step processes. Therefore, we must attempt to extend the definition (7.3). For this, we have to use tools provided by *Functional Analysis* based upon a very natural idea: If we are able to prove that (7.3) gives a continuous functional between two metric spaces, then the stochastic integral defined for the very particular class of step stochastic processes could be extended to a more general class given by the closure of this set with respect to a suitable norm.
The idea of continuity is made precise by the
Isometry property:

$$E\left(\int_0^T u_t dB_t\right)^2=E\left(\int_0^T u_t^2dt\right). \tag{7.4}$$

Let us prove (7.4) for *step processes*. Clearly

$$\begin{aligned}E\left(\int_0^T u_t dB_t\right)^2&=\sum_{j=1}^{n}E\left(u_j^2(\Delta_jB)^2\right)\\&\quad+2\sum_{j<k}E(u_ju_k(\Delta_jB)(\Delta_kB)).\end{aligned}$$

The measurability property of the random variables u_j, $j = 1, \ldots, n$, implies that the random variables u_j^2 are independent of $(\Delta_j B)^2$. Hence, the contribution of the first term in the right hand-side of the preceding identity is equal to

$$\sum_{j=1}^{n} E(u_j^2)(t_j - t_{j-1}) = \int_0^T E(u_t^2)dt.$$

For the second term, we notice that for fixed j and k, $j < k$, the random variables $u_j u_k \Delta_j B$ are independent of $\Delta_k B$. Therefore,

$$E(u_j u_k(\Delta_j B)(\Delta_k B)) = E(u_j u_k(\Delta_j B))E(\Delta_k B) = 0.$$

Thus, we have (7.4).
This property tell us that the stochastic integral is a continuous functional defined on $\mathcal{E}$, endowed with the norm of $L^2(\Omega \times [0,T])$, taking values on the set $L^2(\Omega)$ of square integrable random variables.
The next step consists of identifying a bigger set than $\mathcal{E}$ of random processes such that $\mathcal{E}$ is dense in the norm $L^2(\Omega \times [0,T])$. This is actually the set denoted before by $L^2_{a,T}$. Indeed, one can prove -and this is a crucial fact in Itô's theory- that for any $u \in L^2_{a,T}$, there exists a sequence $(u^n, n \geq 1) \subset \mathcal{E}$ such that

$$\lim_{n\to\infty} \int_0^t E\left((u_t^n - u_t)^2 dt\right) = 0.$$

Owing to this fact, we can give the following definition.

Definition 7.1 *The Itô stochastic integral of a process $u \in L^2_{a,T}$ is*

$$\int_0^T u_t dB_t := L^2(\Omega) - \lim_{n\to\infty} \int_0^T u_t^n dB_t. \tag{7.5}$$

In order this definition to make sense, one needs to be sure that if the process u is approximated by two different sequences, say $u^{n,1}$ and $u^{n,2}$, the definition of the stochastic integral, using either $u^{n,1}$ or $u^{n,2}$ coincide. This is proved using the isometry property. Indeed

$$\begin{aligned}
&E\left(\int_0^T u_t^{n,1} dB_t - \int_0^T u_t^{n,2} dB_t\right)^2 = \int_0^T E\left(u_t^{n,1} - u_t^{n,2}\right)^2 dt \\
&\leq 2\int_0^T E\left(u_t^{n,1} - u_t\right)^2 dt + 2\int_0^T E\left(u_t^{n,2} - u_t\right)^2 dt \\
&\to 0,
\end{aligned}$$

By its very definition, the stochastic integral defined in Definition 7.1 satisfies the isometry property as well. Moreover,

- stochastic integrals are centered random variables:

$$E\left(\int_0^T u_t dB_t\right) = 0,$$

- stochastic integration is a linear operator:

$$\int_0^T (au_t + bv_t)\, dB_t = a\int_0^T u_t dB_t + b\int_0^T v_t dB_t.$$

To prove these facts, we first consider processes in $\mathcal{E}$, in this case the proof is very easy, and then we extend their validity by a density argument.

We end this section with an interesting example.

Example 7.1 *For the Brownian motion B, the following formula holds:*

$$\int_0^T B_t dB_t = \frac{1}{2}\left(B_T^2 - T\right).$$

Let us remark that we would rather expect $\int_0^T B_t dB_t = \frac{1}{2}B_T^2$, by analogy with rules of deterministic calculus.
To prove this identity, we define a particular sequence of approximating step processes, as follows. Consider the partition of $[0,T]$ given by $t_j = \frac{jT}{n}$ and set

$$u_t^n = \sum_{j=1}^n B_{t_{j-1}}\ \mathbf{1}_{]t_{j-1},t_j]}.$$

We have

$$\int_0^T E\left(u_t^n - B_t\right)^2 dt = \sum_{j=1}^n \int_{t_{j-1}}^{t_j} E\left(B_{t_{j-1}} - B_t\right)^2 dt$$
$$\leq \frac{T}{n}\sum_{j=1}^n \int_{t_{j-1}}^{t_j} dt = \frac{T^2}{n}.$$

Therefore, $u^n, n \geq 1$ is an approximating sequence of B in the norm of $L^2(\Omega \times [0,T])$. According to Definition 7.1,

$$\int_0^T B_t dB_t = \lim_{n\to\infty} \sum_{j=1}^n B_{t_{j-1}}\left(B_{t_j} - B_{t_{j-1}}\right),$$

in the $L^2(\Omega)$ norm.

Clearly,

$$\sum_{j=1}^{n} B_{t_{j-1}} \left(B_{t_j} - B_{t_{j-1}} \right) = \frac{1}{2} \sum_{j=1}^{n} \left(B_{t_j}^2 - B_{t_{j-1}}^2 \right) - \frac{1}{2} \sum_{j=1}^{n} \left(B_{t_j} - B_{t_{j-1}} \right)^2 = \frac{1}{2} B_T^2 - \frac{1}{2} \sum_{j=1}^{n} \left(B_{t_j} - B_{t_{j-1}} \right)^2 .$$

We conclude by using Proposition 6.2.

7.2 The Itô Integral as a Stochastic Process

The indefinite Itô stochastic integral of a process $u \in L^2_{a,T}$ is defined as follows:

$$\int_0^t u_s dB_s := \int_0^T u_s 1_{[0,t]}(s) dB_s, \tag{7.6}$$

$t \in [0, T]$.
For this definition to make sense, we need that for any $t \in [0, T]$, the process $\{u_s 1_{[0,t]}(s), s \in [0, T]\}$ belongs to $L^2_{a,T}$. This is clearly true.
Obviously, properties of the integral mentioned in the previous section, like zero mean, isometry, linearity, also hold for the indefinite integral.
The rest of the section is devoted to the study of important properties of the stochastic process given by an indefinite Itô integral.

Proposition 7.1 *The process* $\{I_t = \int_0^t u_s dB_s, t \in [0, T]\}$ *is a martingale.*

Proof: We first establish the martingale property for any approximating sequence

$$I_t^n = \int_0^t u_s^n dB_s, t \in [0, T],$$

where u^n converges to u in $L^2(\Omega \times [0, T])$. This suffices to prove the Proposition, since $L^2(\Omega)$–limits of martingales are again martingales.
Let u_t^n, $t \in [0, T]$, be defined by the right hand-side of (7.2). Fix $0 \leq s \leq t \leq T$ and assume that $s \leq t_k \leq t_l \leq t$. Then

$$I_t^n - I_s^n = u_k(B_{t_k} - B_s) + \sum_{j=k+1}^{l} u_j \Delta_j B + u_l(B_t - B_{t_l}).$$

Using properties (g) and (f), respectively, of the conditional expectation yields

$$\begin{aligned} E\left(I_t^n - I_s^n/\mathcal{F}_s\right) &= E\left(u_k(B_{t_k} - B_s)/\mathcal{F}_s\right) + \sum_{j=k+1}^{l} E\left(E\left(u_j \Delta_j B/\mathcal{F}_{t_{j-1}}\right)/\mathcal{F}_s\right) \\ &\quad + E\left(u_l E\left(B_t - B_{t_l}/\mathcal{F}_{t_{l-1}}\right)/\mathcal{F}_s\right) \\ &= 0. \end{aligned}$$

This finishes the proof of the proposition. □

A proof not very different as that of Proposition 6.2 yields

Proposition 7.2 *For any process $u \in L^2_{a,T}$,*

$$L^1(\Omega) - \lim_{n\to\infty} \sum_{j=1}^{n} \left(\int_{t_{j-1}}^{t_j} u_s dB_s \right)^2 = \int_0^t u_s^2 ds.$$

That means, the quadratic variation of the indefinite stochastic integral is given by the process $\{\int_0^t u_s^2 ds, t \in [0,T]\}$.

The isometry property of the stochastic integral can be extended in the following sense. Let $p \in [2, \infty[$. Then,

$$E\left(\int_0^t u_s dB_s \right)^p \leq C(p) E\left(\int_0^t u_s^2 ds \right)^{\frac{p}{2}}. \tag{7.7}$$

Here $C(p)$ is a positive constant depending on p. This is Burkholder's inequality.
A combination of Burkholder's inequality and Kolmogorov's continuity criterion allows to deduce the continuity of the sample paths of the indefinite stochastic integral. Indeed, assume that $\int_0^T E\left(u_r\right)^{\frac{p}{2}} dr < \infty$, for any $p \in [2, \infty[$. Using first (7.7) and then Hölder's inequality (be smart!) implies

$$\begin{aligned} E\left(\int_s^t u_r dB_r \right)^p &\leq C(p) E\left(\int_s^t u_r^2 dr \right)^{\frac{p}{2}} \\ &\leq C(p)|t-s|^{\frac{p}{2}-1} \int_s^t E\left(u_r\right)^{\frac{p}{2}} dr \\ &\leq C(p)|t-s|^{\frac{p}{2}-1}. \end{aligned}$$

Since $p \geq 2$ is arbitrary, with Proposition 6.1 we have that the sample paths of $\int_0^t u_s dB_s, t \in [0,T]$ are γ–Hölder continuous with $\gamma \in]0, \frac{1}{2}[$.

7.3 Remarks on Extensions of The Stochastic Integral

Consider stochastic processes satisfying

$$P\left\{\int_0^T u_t^2 dt < \infty\right\} = 1. \tag{7.8}$$

Clearly (7.8) holds for processes in $L^2_{a,T}$.
One can construct a stochastic integral for processes u satisfying (7.8) by means of a stopping procedure, as follows.
For any natural number n, define the positive random variable

$$\tau_n = \inf\left\{t \geq 0 : \int_0^t u_s^2 ds = n\right\} \tag{7.9}$$

One can prove that τ_n is a *stopping time* with respect to the filtration $(\mathcal{F}_t, t \geq 0)$. Than means $(\tau_n \leq t) \in \mathcal{F}_t$, for any $t \geq 0$.
By virtue of assumption (7.8), the sequence of random variables $(\tau_n, n \geq 1)$ increases to the deterministic random variable T. In addition, it holds that if $t \leq \tau_n$, then $\int_0^t u_s^2 ds \leq n$.
Define

$$u_t^{(n)} = u_t \mathbf{1}_{[0,\tau_n]}(t),$$

$n \geq 1$. This process belongs to $L^2_{a,T}$. Moreover, if $m \geq n$, on the set $(t \leq \tau_n)$ both processes $u^{(n)}$ and $u^{(m)}$ coincide. By the *local property of the stochastic integral*, one has that

$$\int_0^t u_s^{(n)} dB_s = \int_0^t u_s^{(m)} dB_s.$$

Fix $t \leq T$. Because of the convergence of τ_n to T, there exists n_0 such that $t < \tau_n$ for any $n \geq n_0$ and, by the above discussion, it makes sense to define

$$\int_0^t u_s dB_s = \int_0^t u_s^{(n_0)} dB_s.$$

A second aspect to be mentioned in this section, concerns the underlying filtration. One can follow the constructions given so far for the Itô's stochastic integral by replacing the natural filtration generated by the Brownian motion by a new filtration $(\mathcal{G}_t, t \geq 0)$ satisfying $E\left(B_t - B_s / \mathcal{G}_s\right) = 0$. The corresponding stochastic integral possess all the properties mentioned before with the obvious changes.

7.4 A Change of Variables Formula: Itô's Formula

Like in Example 7.1, we can prove the following formula, valid for any $t \geq 0$:

$$B_t^2 = 2\int_0^t B_s dB_s + t. \tag{7.10}$$

If the sample paths of $\{B_t, t \geq 0\}$ were sufficiently smooth -for example, of bounded variation- and you were asked for a formula for B_t^2, you most likely would answer

$$B_t^2 = 2\int_0^t B_s dB_s, \tag{7.11}$$

relying on the rules of classical calculus.
The difference between the formula we have rigorously established, (7.10), and the spontaneous answer (7.11) is the term t. Where does it come from? Consider the very naïve decomposition of B_t^2 associated with any partition of $[0,t]$ defined by $0 = t_0 \leq t_1 \leq \cdots \leq t_n = t$,

$$\begin{aligned} B_t^2 &= \sum_{j=0}^{n-1} \left(B_{t_{j+1}}^2 - B_{t_j}^2 \right) \\ &= 2\sum_{j=0}^{n-1} B_{t_j} \left(B_{t_{j+1}} - B_{t_j} \right) + \sum_{j=0}^{n-1} \left(B_{t_{j+1}} - B_{t_j} \right)^2, \end{aligned} \tag{7.12}$$

where we have used that $B_0 = 0$.
Consider a sequence of partitions of $[0,t]$ whose mesh tends to zero and let us compute the limit of the right hand-side of (7.12). By the result proved in Proposition 6.2, we infer that

$$\sum_{j=0}^{n-1} \left(B_{t_{j+1}} - B_{t_j} \right)^2 \to t,$$

in the convergence of $L^2(\Omega)$. This gives the extra contribution in the development of B_t^2 in comparison with the classical calculus approach.
Notice that, if B were of bounded variation then, we could argue as follows:

$$\begin{aligned} \sum_{j=0}^{n-1} \left(B_{t_{j+1}} - B_{t_j} \right)^2 &\leq \sup_{0 \leq j \leq n-1} |B_{t_{j+1}} - B_{t_j}| \\ &\quad \times \sum_{j=0}^{n-1} |B_{t_{j+1}} - B_{t_j}|. \end{aligned}$$

By the continuity of the sample paths of the Brownian motion, the first factor in the right hand-side of the preceding inequality tends to zero as the mesh

of the partition tends to zero, while the second factor remains finite, by the property of bounded variation.
Summarising. Differential calculus with respect to the Brownian motion should take into account second order differential terms. Roughly speaking

$$(dB_t)^2 = dt.$$

We can attach a precise meaning to this heuristic formula by means of Proposition 6.2.

7.4.1 One dimensional Itô's formula

In this section, we shall extend the formula (7.10) and write an expression for $f(t, B_t)$ for a class of functions f which include $f(x) = x^2$.

Definition 7.2 *Let $\{v_t, t \in [0,T]\}$ be a stochastic process, adapted, whose sample paths are almost surely Lebesgue integrable, that is $\int_0^T |v_t| dt < \infty$, a.s.. Consider a stochastic process $\{u_t, t \in [0,T]\}$ belonging to $L^2_{a,T}$ and a random variable X_0. The stochastic process defined by*

$$X_t = X_0 + \int_0^t u_s dB_s + \int_0^t v_s ds, \tag{7.13}$$

$t \in [0,T]$ is termed an Itô process.

An alternative writing of (7.13) in *differential form* is

$$dX_t = u_t dB_t + v_t dt.$$

We are now going to state a version of the Itô formula.

Theorem 7.1 *Let $f : [0,T] \times \mathbb{R} \to \mathbb{R}$ be a function in $\mathcal{C}^{1,2}$ and X be an Itô process with decomposition given in (7.13). The following formula holds true:*

$$\begin{aligned} f(t, X_t) = f(0, X_0) &+ \int_0^t \partial_s f(s, X_s) ds + \int_0^t \partial_x f(s, X_s) u_s dB_s \\ &+ \int_0^t \partial_x f(s, X_s) v_s ds + \frac{1}{2} \int_0^t \partial^2_{xx} f(s, X_s) u_s^2 ds. \end{aligned} \tag{7.14}$$

An idea of the proof. Consider a sequence of partitions of $[0,T]$, for example the one defined by $t_j^n = \frac{jt}{n}$. In the sequel, we avoid mentioning the superscript n for the sake of simplicity. We can write

$$\begin{aligned} f(t,X_t) - f(0,X_0) &= \sum_{j=0}^{n-1} \Big[f(t_{j+1}, X_{t_{j+1}}) - f(t_j, X_{t_j}) \Big] \\ &= \sum_{j=0}^{n-1} \Big[f(t_{j+1}, X_{t_j}) - f(t_j, X_{t_j}) \Big] \qquad (7.15) \\ &\quad + \Big[f(t_{j+1}, X_{t_{j+1}}) - f(t_{j+1}, X_{t_j}) \Big] \\ &= \sum_{j=0}^{n-1} \Big[\partial_s f(\bar{t}_j, X_{t_j})(t_{j+1} - t_j) \Big] \qquad (7.16) \\ &\quad + \Big[\partial_x f(t_{j+1}, X_{t_j})(X_{t_{j+1}} - X_{t_j}) \Big] \\ &\quad + \frac{1}{2} \sum_{j=0}^{n-1} \partial^2_{xx} f(t_{j+1}, \bar{X}_j)(X_{t_{j+1}} - X_{t_j})^2. \qquad (7.17) \end{aligned}$$

with $\bar{t} \in]t_j, t_{j+1}[$ and $\bar{X}_j$ an intermediate (random) point on the segment determined by X_{t_j} and $X_{t_{j+1}}$.
In fact, this follows from a Taylor expansion of the function f up to the first order in the variable s, and up to the second order in the variable x. The asymmetry in the orders is due to the existence of quadratic variation of the processes involved. The expresion (7.15) is the analogue of (7.12). The former is much simpler for two reasons. Firstly, there is no s-variable; secondly, f is a polynomial of second degree, and therefore it has an exact Taylor expansion. But both formulas have the same structure.
When passing to the limit as $n \to \infty$, we obtain

$$\begin{aligned} &\sum_{j=0}^{n-1} \partial_s f(t_j, X_{t_j})(t_{j+1} - t_j) \to \int_0^t \partial_s f(s, X_s) ds \\ &\sum_{j=0}^{n-1} \partial_x f(t_j, X_{t_j})(X_{t_{j+1}} - X_{t_j}) \to \int_0^t \partial_x f(s, X_s) u_s dB_s \\ &\qquad + \int_0^t \partial_x f(s, X_s) v_s ds \\ &\sum_{j=0}^{n-1} \partial_{xx} f(t_j, \bar{X}_j)(X_{t_{j+1}} - X_{t_j})^2 \to \int_0^t \partial^2_{xx} f(s, X_s) u_s^2 ds, \end{aligned}$$

in the convergence of probability.

Itô's formula (7.14) can be written in the formal simple *differential form*

$$df(t, X_t) = \partial_t f(t, X_t)dt + \partial_x f(t, X_t)dX_t + \frac{1}{2}\partial^2_{xx} f(t, X_t)(dX_t)^2, \qquad (7.18)$$

where $(dX_t)^2$ is computed using the formal rule of composition

$$\begin{aligned} &dB_t \times dB_t = dt, \\ &dB_t \times dt = dt \times dB_t = 0, \\ &dt \times dt = 0. \end{aligned}$$

Consider in Theorem 7.1 the particular case where $f : \mathbb{R} \to \mathbb{R}$ is a function in $\mathcal{C}^2$. Then formula (7.14) becomes

$$\begin{aligned} f(X_t) = f(X_0) &+ \int_0^t f'(X_s)u_s dB_s + \int_0^t f'(X_s)v_s ds \\ &+ \frac{1}{2}\int_0^t f''(X_s)u_s^2 ds. \end{aligned} \qquad (7.19)$$

Example 7.2 *Consider the function*

$$f(t, x) = e^{\mu t - \frac{\sigma^2}{2}t + \sigma x},$$

with $\mu, \sigma \in \mathbb{R}$.
Applying formula (7.14) to $X_t := B_t$ *-a Brownian motion- yields*

$$f(t, B_t) = 1 + \mu \int_0^t f(s, B_s)ds + \sigma \int_0^t f(s, B_s)dB_s.$$

Hence, the process $\{Y_t = f(t, B_t), t \geq 0\}$ *satisfies the equation*

$$Y_t = 1 + \mu \int_0^t Y_s ds + \sigma \int_0^t Y_s dB_s.$$

The equivalent differential form of this identity is the linear stochastic differential equation

$$\begin{aligned} dY_t &= \mu Y_t dt + \sigma Y_t dB_t, \\ Y_0 &= 1. \end{aligned} \qquad (7.20)$$

Black and Scholes proposes as model of a market with a single risky asset with initial value $S_0 = 1$*, the process* $S_t = Y_t$*. We have seen that such a process is in fact the solution to a linear stochastic differential equation (see (7.20)). The section 7.5 will be devoted to a further analysis of this example.*

7.4.2 Multidimensional Version of Itô's Formula

Consider a m-dimensional Brownian motion $\{(B_t^1, \cdots, B_t^m), t \geq 0\}$ and n real-valued Itô processes, as follows:

$$dX_t^i = \sum_{l=1}^{m} u_t^{i,l} dB_t^l + v_t^i dt,$$

$i = 1, \cdot, n$. We assume that each one of the processes $u_t^{i,l}$ belong to $L_{a,T}^2$ and that $\int_0^T |v_t^i| dt < \infty$.
Consider also a function $f : [0, \infty) \times \mathbb{R}^n \mapsto \mathbb{R}$ of class $\mathcal{C}^{1,2}$. Using similar ideas as for the proof of Theorem 7.1, one can establish the following formula

$$f(t, X_t) = f(0, X_0) + \int_0^t \partial_s f(s, X_s) ds + \sum_{i=1}^{n} \int_0^t \partial f_{x_i}(s, X_s) dX_s^i + \frac{1}{2} \sum_{i,j=1}^{n} \int_0^t \partial_{x_i, x_j}(s, X_s) dX_s^i dX_s^j, \tag{7.21}$$

where in order to compute $dX_s^i dX_s^j$, we have to apply the following rules

$$\begin{aligned} dB_s^i dB_t^j &= \delta_{i,j} ds, \\ dB_s^i ds &= 0, \\ (ds)^2 &= 0. \end{aligned} \tag{7.22}$$

We remark that the identity (7.22) is a consequence of the independence of the components of the Brownian motion.

Example 7.3 *Consider the particular case $m = 1$, $n = 2$ and $f(x, y) = xy$. That is, f does not depend on t and we have denoted a generic point of $\mathbb{R}$ by (x, y). Then the above formula (7.21) yields*

$$X_t^1 X_t^2 = X_0^1 X_0^2 + \int_0^t X_s^1 dX_s^2 + \int_0^t X_s^2 dX_s^1 + \int_0^t \left[\left(u_s^1 \right)^2 + \left(u_s^2 \right)^2 \right] ds. \tag{7.23}$$

7.5 An Application of Stochastic Calculus: The Black and Scholes Model

In this section, we shall consider the mathematical market consisting of a single risky asset in continuous time, obtained by a limiting procedure of the discrete time Cox, Ross and Rubinstein model. Denoting by S_0 its initial

value, the value at any time t is the continuous time stochastic process defined by

$$S_t = S_0 e^{\mu t - \frac{\sigma^2}{2}t + \sigma B_t},$$

while for the non risky asset

$$S_t^0 = e^{rt},$$

where $r > 0$ denotes the instantaneous interest rate.

Definition 7.3 *A portfolio or a trading strategy in the market* $\{(S_t^0, S_t), t \geq 0\}$ *is a stochastic process* $\Phi = \{(\alpha_t, \beta_t), 0 \leq t \leq T\}$ *satisfying the following conditions:*

1. α *and* β *are measurable and adapted processes,*

2. $$\int_0^T |\alpha_t| dt < \infty,$$
$$\int_0^T \beta_t^2 dt < \infty.$$

Recall that the term *portfolio* refers to the number of shares of each asset at a given time t.

Definition 7.4 *The value of the portfolio at time* t *is the stochastic process*

$$V_t(\Phi) = \alpha_t e^{rt} + \beta_t S_t.$$

Definition 7.5 *A portfolio* Φ *is said to be self-financing if its Itô differential is given by*

$$dV_t(\Phi) = r\alpha_t e^{rt} dt + \beta_t dS_t. \tag{7.24}$$

Notice that, the conditions required on the processes α and β imply that (7.24) is well defined. In fact, the integral form of (7.24) is

$$V_t(\Phi) = V_0(\Phi) + \int_0^t \alpha_s dS_0 + \int_0^t \beta_s dS_s,$$

and the integrals of the right hand-side of this expression make sense for the class of processes under consideration.

The condition of being *self-financing* is really a restriction on the model. Indeed, equation (7.24) would be the one derived from the definition of $V_t(\Phi)$ if

$$\begin{aligned} d(\alpha_t e^{rt}) &= \alpha_t d(e^{rt}), \\ d(\beta_t S_t) &= \beta_t dS_t. \end{aligned}$$

This is of course true if α_t, β_t were constant real numbers.
As in the discrete case, we can define the *normalized* market of *discounted prices* by

$$\tilde{S}_t = e^{-rt} S_t = S_0 \exp\left((\mu - r)t - \frac{\sigma^2}{2} t + \sigma B_t \right).$$

Then, the discounted value of the portfolio is

$$\tilde{V}_t(\Phi) = e^{-rt} V_t(\Phi) = \alpha_t + \beta_t \tilde{S}_t.$$

When taking the Itô differential, we have

$$\begin{aligned} d\tilde{V}_t(\Phi) &= -re^{-rt} V_t(\Phi) dt + e^{-rt} dV_t(\Phi) \\ &= -r\beta_t \tilde{S}_t dt + e^{-rt} \beta_t dS_t \\ &= \beta_t d\tilde{S}_t. \end{aligned}$$

7.5.1 Viability of the Black and Scholes Model

Remember the notion of a neutral probability given in Definition 4.9 and its relationship with arbitrage free strategies.
In this setting, the existence of a neutral probability follows from an important result in stochastic calculus given by Girsanov's Theorem. We shall report now on it.

Given a Brownian motion $\{B_t, 0 \leq t \leq T\}$ and a real number λ, the stochastic process

$$L_t = \exp\left(-\lambda B_t - \frac{\lambda^2}{2} t \right),$$

$t \geq 0$, is a martingale.
Indeed, this follows easily applying Itô's formula, since clearly,

$$L_t = 1 - \lambda \int_0^t L_s dB_s.$$

By construction, L is positive and $E(L_t) = 1$, for any $t \geq 0$.

Consider the probability space $(\Omega, \mathcal{F}_T, P)$. For any $A \in \mathcal{F}_T$ define

$$Q(A) = E\left(1_A L_T\right).$$

Obviously, $Q(\Omega) = 1$ and Q defines a new probability on $(\Omega, \mathcal{F}_T)$. Moreover, if instead of considering the probability space $(\Omega, \mathcal{F}_T, P)$, we take $(\Omega, \mathcal{F}_t, P)$, $0 \leq t \leq T$, the restriction of Q on $\mathcal{F}_t$ can be computed as follows:

$$\begin{gathered} Q(A) = E\left(1_A L_T\right) = E\left(E\left(1_A L_T / \mathcal{F}_t\right)\right) \\ E\left(1_A E\left(L_T / \mathcal{F}_t\right)\right) = E\left(1_A L_t\right), \end{gathered}$$

for any $A \in \mathcal{F}_t$, where we have used the martingale property of the process L_t. Clearly P and Q are equivalent in the (usual) sense that $P(A) = 0$ if and only if $Q(A) = 0$.

Theorem 7.2 *(Girsanov's Theorem) Fix $\lambda \in \mathbb{R}$ and consider the translation of the sample paths of the Brownian motion defined by*

$$W_t = B_t + \lambda t,$$

for any $t \in [0, T]$. On the probability space $(\Omega, \mathcal{F}_T, Q)$, the stochastic process $\{W_t, t \in [0, T]\}$ is a Brownian motion.

Let us go back to the Black and Scholes model. Define

$$W_t = B_t + \frac{\mu - r}{\sigma} t,$$

$t \in [0, T]$.
The value of the parameter $\lambda := \frac{\mu - r}{\sigma}$ is chosen in such a way that

$$S_t = S_0 \exp\left(rt - \frac{\sigma^2}{2} t + \sigma W_t\right)$$

and therefore

$$\tilde{S}_t = S_0 \exp\left(-\frac{\sigma^2}{2} t + \sigma W_t\right). \tag{7.25}$$

That is, *in the probability space $(\Omega, \mathcal{F}_T, Q)$, the discounted prices are a martingale*. This property yields that the Black and Scholes model is *viable*.

7.5.2 Pricing in the Black and Scholes Model

From the expression

$$\tilde{V}_t(\Phi) = V_0(\Phi) + \int_0^t \beta_u d\tilde{S}_u,$$

we infer that the value of the portfolio is a martingale with respect to the probability Q provided that

$$\int_0^T E(\beta_u^2 \tilde{S}_u^2) du < \infty.$$

Assume this condition in the sequel. Then, for an European call option with maturity time T and exercise price K, any replicable strategy Φ must satisfy $V_T(\Phi) = (S_T - K)^+$. The price of this replicable strategy is

$$V_t(\Phi) = E_Q\left(e^{-r(T-t)}(S_T - K)^+ / \mathcal{F}_t\right),$$

because of the martingale property mentioned before. This formula is valid for any $t \in [0, T]$. In particular, for $t = 0$, this gives the price of the option:

$$V_0(\Phi) = E_Q\left(e^{-rT}(S_T - K)^+\right).$$

7.5.3 Completeness of the Black and Scholes Model

Like its discrete time analogue, the Black and Scholes model is also *complete*. The proof of this fact relies on a result of representation of random variables in $L^2(\Omega, \mathcal{F}_T, P)$ in terms of a stochastic integral with respect to the Brownian motion, as follows.

Theorem 7.3 *Let $F \in L^2(\Omega, \mathcal{F}_T, P)$. There exists a unique stochastic process $\{u_t, t \in [0, T]\}$ belonging to $L^2_{a,T}$ such that*

$$F = E(F) + \int_0^T u_s dB_s. \tag{7.26}$$

From this result, it is easy to obtain a representation for *square integrable martingales*. More precisely, consider a martingale $\{M_t, t \in [0, T]\}$ such that each random variable M_t satisfies $E(M_t)^2 < \infty$. Then, there exists a unique stochastic process $\{u_t, t \in [0, T]\}0$ in $L^2_{a,T}$ such that

$$M_t = E(M_0) + \int_0^t u_s dB_s. \tag{7.27}$$

Indeed, we can apply (7.26) to $F := M_T$ and then, take the conditional expectation with respect to $\mathcal{F}_t$ to obtain

$$\begin{aligned} M_t = E(M_T/\mathcal{F}_t) &= E(M_0) + E\left(\int_0^T u_s dB_s/\mathcal{F}_t\right) \\ &= E(M_0) + \int_0^t u_s dB_s, \end{aligned}$$

where in the last equality we have applied the martingale property of the stochastic integral.
For example, from the Itô formula, we obtain that the process u in the integral representation of $F := B_T^2$ is given by $u_s = 2B_s$, since $E(B_T^2) = T$ (see equation (7.10)).
With these results, let us prove the completeness of Black and Scholes model for an European call option.
First, let H be the payoff of a derivative with maturity time T and assume that $H \in L^2(\Omega, \mathcal{F}_T, Q)$. Later on, $H = (S_T - K)^+$.
Consider the square integrable martingale on the probability space $(\Omega, \mathcal{F}_T, Q)$

$$M_t = E_Q\left(e^{-rT} H/\mathcal{F}_t\right),$$

and its integral representation

$$M_t = M_0 + \int_0^t u_s dW_s.$$

Consider the pair of stochastic processes:

$$\begin{aligned} \beta_t &= \frac{u_t}{\sigma \tilde{S}_t}, \\ \alpha_t &= M_t - \beta_t \tilde{S}_t. \end{aligned}$$

Let $\Phi_t = (\alpha_t, \beta_t)$. We check that $V_t(\Phi) = H$ and therefore H is attainable.
Indeed, the discounted value of this trading strategy is

$$\tilde{V}_t(\Phi) = \alpha_t + \beta_t \tilde{S}_t = M_t,$$

and its value at time T,

$$V_T(\Phi) = e^{rT} \tilde{V}_T(\Phi) = e^{rT} M_T = H.$$

Let us next check that Φ is self-financing (see (7.24). Indeed, it is clear that

$$\begin{aligned} dV_t(\Phi) &= re^{rt}\tilde{V}_t(\Phi)dt + e^{rt}d\tilde{V}_t(\Phi) \\ &= re^{rt}M_t dt + e^{rt}dM_t \\ &= re^{rt}M_t dt + e^{rt}u_t dW_t \\ &= re^{rt}\left(\alpha_t dt + \beta_t \tilde{S}_t dt\right) + \sigma e^{rt}\beta_t \tilde{S}_t dW_t \\ &= re^{rt}\left(\alpha_t dt + \beta_t \tilde{S}_t dt\right) + e^{rt}\beta_t d\tilde{S}_t \\ &= re^{rt}\alpha_t dt + \beta_t dS_t. \end{aligned}$$

7.5.4 Computing the Replicating Strategy

Let $H = g(S_T)$. By writing

$$\begin{aligned} V_t := V_t(\Phi) &= E_Q\left(e^{-r(T-t)}g(S_T)/\mathcal{F}_t\right) \\ &= e^{-r(T-t)}E_Q\left(g(S_t e^{r(T-t)}e^{\sigma(W_T-W_t)-\frac{\sigma^2}{2}(T-t)})/\mathcal{F}_t\right), \end{aligned} \tag{7.28}$$

we obtain

$$V_t = F(t, S_t),$$

with

$$F(t,x) = e^{-r(T-t)}E_Q\left(g(xe^{r(T-t)}e^{\sigma(W_T-W_t)-\frac{\sigma^2}{2}(T-t)})/\mathcal{F}_t\right).$$

This is a useful expression. In fact, assume that $F(t,x)$ is a $\mathcal{C}^{1,2}$ function. This condition shall depend, of course on the function g. Recall that, in terms of the new Brownian motion W_t,

$$S_t = S_0 e^{(r-\frac{\sigma^2}{2})t+\sigma W_t},$$

which implies

$$dS_t = rS_t dt + \sigma S_t dW_t.$$

By applying Itô's formula, we obtain

$$\begin{aligned} V_t = V_0 &+ \int_0^t \sigma\frac{\partial F}{\partial x}(s, S_s)S_s dW_s + \int_0^t r\frac{\partial F}{\partial x}(s, S_s)S_s ds \\ &+ \int_0^t \frac{\partial F}{\partial s}(s, S_s)ds + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}(s, S_s)\sigma^2 S_s^2 ds. \end{aligned} \tag{7.29}$$

On the other hand,

$$\begin{aligned} dV_t &= re^{rt}\tilde{V}_t dt + e^{rt}d\tilde{V}_t = rV_t dt + e^{rt}dM_t \\ &= rV_t dt + e^{rt}u_t dW_t \\ &= rV_t dt + e^{rt}\sigma\beta_t\tilde{S}_t dW_t \\ &= rV_t dt + \sigma\beta_t S_t dW_t. \end{aligned}$$

This yields the expression of V_t as an Itô process,

$$V_t = V_0 + \int_0^t \sigma\beta_s S_s dW_s + \int_0^t rV_s ds. \tag{7.30}$$

Comparing the expressions (7.29) and (7.30), and by virtue of the uniqueness of the Itô representation, we obtain the crucial relations

$$\beta_t = \frac{\partial F}{\partial x}(t, S_t), \tag{7.31}$$

$$\begin{aligned} rV_t = rF(t, S_t) = &\frac{\partial F}{\partial x}(t, S_t)S_t + \frac{\partial F}{\partial t}(t, S_t) \\ &+ \frac{1}{2}\frac{\partial^2 F}{\partial x^2}(t, S_t)\sigma^2 S_t^2. \end{aligned} \tag{7.32}$$

Since S_t is positive, we see that the function $F(t, x)$ satisfies the partial differential equation on $(0, \infty)$

$$\begin{aligned} \frac{\partial F}{\partial x}(t, x)x + \frac{\partial F}{\partial t}(t, x) + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}(t, x)\sigma^2 x^2 = rF(t, S_t), \\ F(T, x) = g(x). \end{aligned} \tag{7.33}$$

This is Black-Scholes-Merton partial differential equation.
We have thus obtained the following formulas for the replicating portfolio:

$$\begin{aligned} \beta_t &= \frac{\partial F}{\partial x}(t, S_t), \\ \alpha_t &= e^{-rt}\left(F(t, S_t) - \beta_t S_t\right). \end{aligned}$$

But this is not enough, because we need the explicit expression for $F(t, x)$. For this reason, we go back to formula (7.28), we set $\theta = T - t$ and we obtain

$$F(t, x) = \frac{1}{\sqrt{2\pi}}e^{-r\theta}\int_{\mathbb{R}} e^{-\frac{y^2}{2}} g\left(xe^{r\theta - \frac{\sigma^2}{2}\theta + \sigma\sqrt{\theta}y}\right) dy.$$

In the particular case $g(x) = (x - K)^+$, and consequently, $H = (S_T - K)^+$ (European call option), this formula yields

$$F(t,x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{y^2}{2}} \left(x e^{-\frac{\sigma^2}{2}\theta + \sigma\sqrt{\theta} y} - K e^{-r(T-t)} \right)^+ dy$$
$$= x\Phi(d_+) - K e^{-r(T-t)} \Phi(d_-), \tag{7.34}$$

where Φ denotes the distribution function of a standard unidimensional Gaussian law and

$$d_+ = \frac{\log \frac{x}{K} + \left(r + \frac{\sigma^2}{2} \right)(T-t)}{\sigma\sqrt{T-t}},$$
$$d_- = \frac{\log \frac{x}{K} + \left(r - \frac{\sigma^2}{2} \right)(T-t)}{\sigma\sqrt{T-t}}.$$

The relation (7.34) provides a formula for pricing European call options in the Black and Scholes model:

$$V_t(\alpha, \beta) = F(t, S_t),$$

with the replicating portfolio

$$\beta_t = \frac{\partial F}{\partial x}(t, S_t) = \Phi(d_+),$$
$$\alpha_t = e^{-rt}(F(t, S_t) - \beta_t S_t).$$

References

[1] P. Baldi: Calcolo delle probabilità e statistica. Mc Grw-Hill Libri Italia, 1992.

[2] H.-H. Kuo: Introduction to Stochastic Integration. Universitext. Springer, 2006.

[3] J. Jacod, P. Protter: Probability Essentials. Universitext, Springer Verlag, 2000.

[4] D. Lamberton, B. Lapeyre: Introduction to Stochastic Calculus Applied to Finance. Chapman & HALL/CRC, 2000.

[5] T. Mickosch: Elementary stochastic calculus. World Scientific Pub. 2000.

[6] D. Nualart: Models probabilistes dels mercats financers. Lecture Notes (in catalan, available at the home page of the author).

[7] J.R. Norris: Markov chains. Cambridge University Press, 1997.

[8] B. Oksendal: Stochastic differential equations: An introduction with applications. 6th Ed. Springer Verlag, 2003.

[9] M. Sanz-Solé: Probabilitats. Col.lecció UB, 28. Edicions Universitat de Barcelona, 1999.

[10] S. E. Shreve: Stochastic calculus for finance I, the binomial asset pricing model. Springer Finance Textbook. Springer 2004

[11] S. E. Shreve: Stochastic calculus for finance II, continuous-time models. Springer Finance Textbook. Springer 2004.

[12] A. N. Shiryayev: Probability. Graduate Texts in Mathematics. Springer Verlag, 1984.

[13] P. Todorovic: An introduction to stochastic processes and their applications. Springer, 1992.