Stochastic Processes

David Nualart
nualart@mat.ub.es

1 Stochastic Processes

1.1 Probability Spaces and Random Variables

In this section we recall the basic vocabulary and results of probability theory. A *probability space* associated with a *random experiment* is a triple $(\Omega, \mathcal{F}, P)$ where:

(i) Ω is the set of all possible outcomes of the random experiment, and it is called the *sample space*.

(ii) $\mathcal{F}$ is a family of subsets of Ω which has the structure of a σ-field:

a) $\emptyset \in \mathcal{F}$

b) If $A \in \mathcal{F}$, then its complement A^c also belongs to $\mathcal{F}$

c) $A_1, A_2, \ldots \in \mathcal{F} \Longrightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$

(iii) P is a function which associates a number $P(A)$ to each set $A \in \mathcal{F}$ with the following properties:

a) $0 \leq P(A) \leq 1$,

b) $P(\Omega) = 1$

c) For any sequence $A_1, A_2, \ldots$ of disjoints sets in $\mathcal{F}$ (that is, $A_i \cap A_j = \emptyset$ if $i \neq j$),

$$\boxed{P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)}$$

The elements of the σ-field $\mathcal{F}$ are called *events* and the mapping P is called a *probability measure.* In this way we have the following interpretation of this model:

$$\boxed{P(F)=\text{“probability that the event } F \text{ occurs”}}$$

The set $\emptyset$ is called the *empty event* and it has probability zero. Indeed, the additivity property (iii,c) implies

$$P(\emptyset) + P(\emptyset) + \cdots = P(\emptyset).$$

The set Ω is also called the *certain set* and by property (iii,b) it has probability one. Usually, there will be other events $A \subset \Omega$ such that $P(A) = 1$. If a statement holds for all ω in a set A with $P(A) = 1$, then it is customary to say that the statement is true *almost surely*, or that the statement holds for almost all $\omega \in \Omega$.

The axioms a), b) and c) lead to the following basic rules of the probability calculus:

$$
\begin{aligned}
P(A \cup B) &= P(A) + P(B) \text{ if } A \cap B = \emptyset \\
P(A^c) &= 1 - P(A) \\
A &\subset B \Longrightarrow P(A) \leq P(B).
\end{aligned}
$$

Example 1 Consider the experiment of flipping a coin once.

$$
\begin{aligned}
\Omega &= \{H, T\} \text{ (the possible outcomes are "Heads" and "Tails")} \\
\mathcal{F} &= \mathbb{P}(\Omega) \text{ ($\mathcal{F}$ contains all subsets of Ω)} \\
P(\{H\}) &= P(\{T\}) = \frac{1}{2}
\end{aligned}
$$

Example 2 Consider an experiment that consists of counting the number of traffic accidents at a given intersection during a specified time interval.

$$
\begin{aligned}
\Omega &= \{0, 1, 2, 3, \ldots\} \\
\mathcal{F} &= \mathbb{P}(\Omega) \text{ ($\mathcal{F}$ contains all subsets of Ω)} \\
P(\{k\}) &= e^{-\lambda} \frac{\lambda^k}{k!} \text{ (Poisson probability with parameter $\lambda > 0$)}
\end{aligned}
$$

Given an arbitrary family $\mathcal{U}$ of subsets of Ω, the smallest σ-field containing $\mathcal{U}$ is, by definition,

$$
\sigma(\mathcal{U}) = \cap \{\mathcal{G}, \mathcal{G} \text{ is a } \sigma\text{-field}, \mathcal{U} \subset \mathcal{G}\}.
$$

The σ-field $\sigma(\mathcal{U})$ is called the σ-field generated by $\mathcal{U}$. For instance, the σ-field generated by the open subsets (or rectangles) of $\mathbb{R}^n$ is called the Borel σ-field of $\mathbb{R}^n$ and it will be denoted by $\mathcal{B}_{\mathbb{R}^n}$.

Example 3 Consider a finite partition $\mathcal{P} = \{A_1, \ldots, A_n\}$ of Ω. The σ-field generated by $\mathcal{P}$ is formed by the unions $A_{i_1} \cup \cdots \cup A_{i_k}$ where $\{i_1, \ldots, i_k\}$ is an arbitrary subset of $\{1, \ldots, n\}$. Thus, the σ-field $\sigma(\mathcal{P})$ has 2^n elements.

Example 4 We pick a real number at random in the interval $[0, 2]$. $\Omega = [0, 2]$, $\mathcal{F}$ is the Borel σ-field of $[0, 2]$. The probability of an interval $[a, b] \subset [0, 2]$ is

$$P([a, b]) = \frac{b - a}{2}.$$

Example 5 Let an experiment consist of measuring the lifetime of an electric bulb. The sample space Ω is the set $[0, \infty)$ of nonnegative real numbers. $\mathcal{F}$ is the Borel σ-field of $[0, \infty)$. The probability that the lifetime is larger than a fixed value $t \geq 0$ is

$$P([t, \infty)) = e^{-\lambda t}.$$

A *random variable* is a mapping

$$\begin{array}{c} \Omega \xrightarrow{X} \mathbb{R} \\ \omega \rightarrow X(\omega) \end{array}$$

which is $\mathcal{F}$-measurable, that is, $X^{-1}(B) \in \mathcal{F}$, for any Borel set B in $\mathbb{R}$. The random variable X assigns a value $X(\omega)$ to each outcome ω in Ω. The measurability condition means that given two real numbers $a \leq b$, the set of all outcomes ω for which $a \leq X(\omega) \leq b$ is an event. We will denote this event by $\{a \leq X \leq b\}$ for short, instead of $\{\omega \in \Omega : a \leq X(\omega) \leq b\}$.

- A random variable defines a σ-field $\{X^{-1}(B), B \in \mathcal{B}_{\mathbb{R}}\} \subset \mathcal{F}$ called the σ-field generated by X.

- A random variable defines a probability measure on the Borel σ-field $\mathcal{B}_{\mathbb{R}}$ by $P_X = P \circ X^{-1}$, that is,

$$P_X(B) = P(X^{-1}(B)) = P(\{\omega : X(\omega) \in B\}).$$

  The probability measure P_X is called the *law* or the *distribution* of X.

We will say that a random variable X has a *probability density* f_X if $f_X(x)$ is a nonnegative function on $\mathbb{R}$, measurable with respect to the Borel σ-field and such that

$$P\{a < X < b\} = \int_a^b f_X(x)dx,$$

for all $a < b$. Notice that $\int_{-\infty}^{+\infty} f_X(x)dx = 1$. Random variables admitting a probability density are called *absolutely continuous*.

We say that a random variable X is *discrete* if it takes a finite or countable number of different values x_k. Discrete random variables do not have densities and their law is characterized by the *probability function*:

$$p_k = P(X = x_k).$$

Example 6 In the experiment of flipping a coin once, the random variable given by

$$X(H) = 1, X(T) = -1$$

represents the earning of a player who receives or loses an euro according as the outcome is heads or tails. This random variable is discrete with

$$P(X = 1) = P(X = -1) = \frac{1}{2}.$$

Example 7 If A is an event in a probability space, the random variable

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{if} \quad \omega \in A \\ 0 & \text{if} \quad \omega \notin A \end{cases}$$

is called the indicator function of A. Its probability law is called the *Bernoulli* distribution with parameter $p = P(A)$.

Example 8 We say that a random variable X has the *normal law* $N(m, \sigma^2)$ if

$$P(a < X < b) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_a^b e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

for all $a < b$.

Example 9 We say that a random variable X has the *binomial law* $B(n, p)$ if

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

for $k = 0, 1, \ldots, n$.

The function $F_X : \mathbb{R} \to [0, 1]$ defined by

$$\boxed{F_X(x) = P(X \le x) = P_X((-\infty, x])}$$

is called the *distribution function* of the random variable X.

- The distribution function F_X is non-decreasing, right continuous and with

$$\begin{aligned}\lim_{x\to-\infty} F_X(x) &= 0, \\ \lim_{x\to+\infty} F_X(x) &= 1.\end{aligned}$$

- If the random variable X is absolutely continuous with density f_X, then,

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy,$$

and if, in addition, the density is continuous, then $F_X'(x) = f_X(x)$.

The *mathematical expectation (*or *expected value)* of a random variable X is defined as the integral of X with respect to the probability measure P:

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

In particular, if X is a discrete variable that takes the values $\alpha_1, \alpha_2, \dots$ on the sets $A_1, A_2, \dots,$ then its expectation will be

$$E(X) = \alpha_1 P(A_1) + \alpha_1 P(A_1) + \cdots .$$

Notice that $E(\mathbf{1}_A) = P(A)$, so the notion of expectation is an extension of the notion of probability.

If X is a non-negative random variable it is possible to find discrete random variables X_n, $n = 1, 2, \dots$ such that

$$X_1(\omega) \le X_2(\omega) \le \cdots$$

and

$$\lim_{n\to\infty} X_n(\omega) = X(\omega)$$

for all ω. Then $E(X) = \lim_{n\to\infty} E(X_n) \le +\infty$, and this limit exists because the sequence $E(X_n)$ is non-decreasing. If X is an arbitrary random variable, its expectation is defined by

$$E(X) = E(X^+) - E(X^-),$$

where $X^+ = \max(X, 0)$, $X^- = -\min(X, 0)$, provided that both $E(X^+)$ and $E(X^-)$ are finite. Note that this is equivalent to say that $E(|X|) < \infty$, and in this case we will say that X is integrable.

A simple computational formula for the expectation of a non-negative random variable is as follows:

$$E(X) = \int_0^\infty P(X > t)dt.$$

In fact,

$$\begin{aligned} E(X) &= \int_\Omega X dP = \int_\Omega \left(\int_0^\infty \mathbf{1}_{\{X>t\}} dt \right) dP \\ &= \int_0^{+\infty} P(X > t)dt. \end{aligned}$$

The expectation of a random variable X can be computed by integrating the function x with respect to the probability law of X:

$$E(X) = \int_\Omega X(\omega)dP(\omega) = \int_{-\infty}^\infty x dP_X(x).$$

More generally, if $g : \mathbb{R} \to \mathbb{R}$ is a Borel measurable function and $E(|g(X)|) < \infty$, then the expectation of $g(X)$ can be computed by integrating the function g with respect to the probability law of X:

$$E(g(X)) = \int_\Omega g(X(\omega))dP(\omega) = \int_{-\infty}^\infty g(x)dP_X(x).$$

The integral $\int_{-\infty}^\infty g(x)dP_X(x)$ can be expressed in terms of the probability density or the probability function of X:

$$\int_{-\infty}^\infty g(x)dP_X(x) = \begin{cases} \int_{-\infty}^\infty g(x)f_X(x)dx, & f_X(x) \text{ is the density of } X \\ \sum_k g(x_k)P(X = x_k), & X \text{ is discrete} \end{cases}$$

Example 10 If X is a random variable with normal law $N(0,\sigma^2)$ and λ is a real number,

$$\begin{aligned} E(\exp(\lambda X)) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{\lambda x} e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{\sigma^2\lambda^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{(x-\sigma^2\lambda)^2}{2\sigma^2}} dx \\ &= e^{\frac{\sigma^2\lambda^2}{2}}. \end{aligned}$$

Example 11 If X is a random variable with Poisson distribution of parameter $\lambda > 0$, then

$$E(X) = \sum_{n=0}^{\infty} n \frac{e^{-\lambda}\lambda^n}{n!} = \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{e^{-\lambda}\lambda^{n-1}}{(n-1)!} = \lambda.$$

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

$$\sigma_X^2 = \mathrm{Var}(X) = E((X - E(X))^2) = E(X^2) - [E(X)]^2,$$

provided $E(X^2) < \infty$. The variance of X measures the deviation of X from its expected value. For instance, if X is a random variable with normal law $N(m,\sigma^2)$ we have

$$\begin{aligned} P(m - 1.96\sigma &\le X \le m + 1.96\sigma) = P(-1.96 \le \frac{X-m}{\sigma} \le 1.96) \\ &= \Phi(1.96) - \Phi(-1.96) = 0.95, \end{aligned}$$

where Φ is the distribution function of the law $N(0,1)$. That is, the probability that the random variable X takes values in the interval $[m-1.96\sigma, m+1.96\sigma]$ is equal to 0.95.

If X and Y are two random variables with $E(X^2) < \infty$ and $E(Y^2) < \infty$, then its covariance is defined by

$$\begin{aligned} \mathrm{cov}(X,Y) &= E\left[(X - E(X))(Y - E(Y))\right] \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

A random variable X is said to have a finite moment of order $p \ge 1$, provided $E(|X|^p) < \infty$. In this case, the pth moment of X is defined by

$$\boxed{m_p = E(X^p).}$$

The set of random variables with finite pth moment is denoted by $L^p(\Omega, \mathcal{F}, P)$.

The *characteristic function* of a random variable X is defined by

$$\varphi_X(t) = E(e^{itX}).$$

The moments of a random variable can be computed from the derivatives of the characteristic function at the origin:

$$m_n = \frac{1}{i^n}\varphi_X^{(n)}(t)|_{t=0},$$

for $n = 1, 2, 3, \ldots$.

We say that $X = (X_1, \ldots, X_n)$ is an n-dimensional *random vector* if its components are random variables. This is equivalent to say that X is a random variable with values in $\mathbb{R}^n$.

The mathematical expectation of an n-dimensional random vector X is, by definition, the vector

$$\boxed{E(X) = (E(X_1), \ldots, E(X_n))}$$

The *covariance matrix* of an n-dimensional random vector X is, by definition, the matrix $\Gamma_X = (\text{cov}(X_i, X_j))_{1 \le i,j \le n}$. This matrix is clearly symmetric. Moreover, it is non-negative definite, that means,

$$\sum_{i,j=1}^{n} \Gamma_X(i,j) a_i a_j \ge 0$$

for all real numbers $a_1, \ldots, a_n$. Indeed,

$$\sum_{i,j=1}^{n} \Gamma_X(i,j) a_i a_j = \sum_{i,j=1}^{n} a_i a_j \text{cov}(X_i, X_j) = \text{Var}(\sum_{i=1}^{n} a_i X_i) \ge 0$$

As in the case of real-valued random variables we introduce the law or distribution of an n-dimensional random vector X as the probability measure defined on the Borel σ-field of $\mathbb{R}^n$ by

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

We will say that a random vector X has a *probability density* f_X if $f_X(x)$ is a nonnegative function on $\mathbb{R}^n$, measurable with respect to the Borel σ-field and such that

$$P\{a_i < X_i < b_i, i = 1, \dots, n\} = \int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f_X(x_1, \dots, x_n) dx_1 \cdots dx_n,$$

for all $a_i < b_i$. Notice that

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_X(x_1, \dots, x_n) dx_1 \cdots dx_n = 1.$$

We say that an n-dimensional random vector X has a multidimensional *normal law* $N(m, \Gamma)$, where $m \in \mathbb{R}^n$, and Γ is a symmetric positive definite matrix, if X has the density function

$$f_X(x_1, \dots, x_n) = (2\pi \det \Gamma)^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_{i,j=1}^n (x_i - m_i)(x_j - m_j)\Gamma_{ij}^{-1}}.$$

In that case, we have, $m = E(X)$ and $\Gamma = \Gamma_X$.

If the matrix Γ is diagonal

$$\Gamma = \begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

then the density of X is the product of n one-dimensional normal densities:

$$f_X(x_1, \dots, x_n) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-m_i)^2}{2\sigma_i^2}} \right).$$

There exists degenerate normal distributions which have a singular covariance matrix Γ. These distributions do not have densities and the law of a random variable X with a (possibly degenerated) normal law $N(m, \Gamma)$ is determined by its characteristic function:

$$E\left(e^{it'X}\right) = \exp\left(it'm - \frac{1}{2}t'\Gamma t\right),$$

where $t \in \mathbb{R}^n$. In this formula t' denotes a row vector ($1 \times n$ matrix) and t denoted a column vector ($n \times 1$ matrix).

If X is an n-dimensional normal vector with law $N(m, \Gamma)$ and A is a matrix of order $m \times n$, then AX is an m-dimensional normal vector with law $N(Am, A\Gamma A')$.

We recall some basic inequalities of probability theory:

- Chebyshev's inequality: If $\lambda > 0$

$$P(|X| > \lambda) \leq \frac{1}{\lambda^p} E(|X|^p).$$

- Schwartz's inequality:

$$E(XY) \leq \sqrt{E(X^2)E(Y^2)}.$$

- Hölder's inequality:

$$E(XY) \leq [E(|X|^p)]^{\frac{1}{p}} [E(|Y|^q)]^{\frac{1}{q}},$$

  where $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$.

- Jensen's inequality: If $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function such that the random variables X and $\varphi(X)$ have finite expectation, then,

$$\varphi(E(X)) \leq E(\varphi(X)).$$

  In particular, for $\varphi(x) = |x|^p$, with $p \geq 1$, we obtain

$$|E(X)|^p \leq E(|X|^p).$$

We recall the different types of convergence for a sequence of random variables X_n, $n = 1, 2, 3, \ldots$:

Almost sure convergence: $X_n \xrightarrow{\text{a.s.}} X$, if

$$\lim_{n \to \infty} X_n(\omega) = X(\omega),$$

for all $\omega \notin N$, where $P(N) = 0$.

Convergence in probability: $X_n \xrightarrow{P} X$, if

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0,$$

for all $\varepsilon > 0$.

Convergence in mean of order $p \geq 1$: $X_n \xrightarrow{L^p} X$, if

$$\lim_{n\to\infty} E(|X_n - X|^p) = 0.$$

Convergence in law: $X_n \xrightarrow{\mathcal{L}} X$, if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

for any point x where the distribution function F_X is continuous.

- The convergence in mean of order p implies the convergence in probability. Indeed, applying Chebyshev's inequality yields

$$P(|X_n - X| > \varepsilon) \leq \frac{1}{\varepsilon^p} E(|X_n - X|^p).$$

- The almost sure convergence implies the convergence in probability. Conversely, the convergence in probability implies the existence of a subsequence which converges almost surely.

- The almost sure convergence implies the convergence in mean of order $p \geq 1$, if the random variables X_n are bounded in absolute value by a fixed nonnegative random variable Y possessing pth finite moment (*dominated convergence theorem*):

$$|X_n| \leq Y, \ E(Y^p) < \infty.$$

- The convergence in probability implies the convergence law, the reciprocal being also true when the limit is constant.

The independence is a basic notion in probability theory. Two events $A, B \in \mathcal{F}$ are said *independent* provided

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

Given an arbitrary collection of events $\{A_i, i \in I\}$, we say that the events of the collection are independent provided

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

for every finite subset of indexes $\{i_1, \ldots, i_k\} \subset I$.

A collection of classes of events $\{\mathcal{G}_i, i \in I\}$ is independent if any collection of events $\{A_i, i \in I\}$ such that $A_i \in \mathcal{G}_i$ for all $i \in I$, is independent.

A collection of random variables $\{X_i, i \in I\}$ is independent if the collection of σ-fields $\left\{X_i^{-1}(\mathcal{B}_{\mathbb{R}^n}), i \in I\right\}$ is independent. This means that

$$P(X_{i_1} \in B_{i_1}, \ldots, X_{i_k} \in B_{i_k}) = P(X_{i_1} \in B_{i_1}) \cdots P(X_{i_k} \in B_{i_k}),$$

for every finite set of indexes $\{i_1, \ldots, i_k\} \subset I$, and for all Borel sets B_j .

Suppose that X, Y are two independent random variables with finite expectation. Then the product XY has also finite expectation and

$$\boxed{E(XY) = E(X)E(Y)}.$$

More generally, if $X_1, \ldots, X_n$ are independent random variables,

$$E\left[g_1(X_1) \cdots g_n(X_n)\right] = E\left[g_1(X_1)\right] \cdots E\left[g_n(X_n)\right],$$

where g_i are measurable functions such that $E\left[|g_i(X_i)|\right] < \infty$.

The components of a random vector are independent if and only if the density or the probability function of the random vector is equal to the product of the marginal densities, or probability functions.

The *conditional probability* of an event A given another event B such that $P(B) > 0$ is defined by

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

We see that A and B are independent if and only if $P(A|B) = P(A)$. The conditional probability $P(A|B)$ represents the probability of the event A modified by the additional information that the event B has occurred occurred. The mapping

$$A \longmapsto P(A|B)$$

defines a new probability on the σ-field $\mathcal{F}$ concentrated on the set B. The mathematical expectation of an integrable random variable X with respect to this new probability will be the conditional expectation of X given B and it can be computed as follows:

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

The following are two two main limit theorems in probability theory.

Theorem 1 (Law of Large Numbers) *Let $\{X_n, n \geq 1\}$ be a sequence of independent, identically distributed random variables, such that $E(|X_1|) < \infty$. Then,*

$$\frac{X_1 + \cdots + X_n}{n} \overset{\text{a.s.}}{\to} m,$$

where $m = E(X_1)$.

Theorem 2 (Central Limit Theorem) *Let $\{X_n, n \geq 1\}$ be a sequence of independent, identically distributed random variables, such that $E(X_1^2) < \infty$. Set $m = E(X_1)$ and $\sigma^2 = \mathrm{Var}(X_1)$. Then,*

$$\frac{X_1 + \cdots + X_n - nm}{\sigma\sqrt{n}} \overset{\mathcal{L}}{\to} N(0, 1).$$

1.2 Stochastic Processes: Definitions and Examples

A stochastic process with state space S is a collection of random variables $\{X_t, t \in T\}$ defined on the same probability space $(\Omega, \mathcal{F}, P)$. The set T is called its *parameter set*. If $T = \mathbb{N} = \{0, 1, 2, \ldots\}$, the process is said to be a *discrete parameter process*. If T is not countable, the process is said to have a *continuous parameter*. In the latter case the usual examples are $T = \mathbb{R}_+ = [0, \infty)$ and $T = [a, b] \subset \mathbb{R}$. The index t represents time, and then one thinks of X_t as the "state" or the "position" of the process at time t. The state space is $\mathbb{R}$ in most usual examples, and then the process is said real-valued. There will be also examples where S is $\mathbb{N}$, the set of all integers, or a finite set.

For every fixed $\omega \in \Omega$, the mapping

$$\boxed{t \longrightarrow X_t(\omega)}$$

defined on the parameter set T, is called a realization, *trajectory*, sample path or sample function of the process.

Let $\{X_t, t \in T\}$ be a real-valued stochastic process and $\{t_1 < \cdots < t_n\} \subset T$, then the probability distribution $P_{t_1,\ldots,t_n} = P \circ (X_{t_1}, \ldots, X_{t_n})^{-1}$ of the random vector

$$(X_{t_1}, \ldots, X_{t_n}) : \Omega \longrightarrow \mathbb{R}^n.$$

is called a finite-dimensional marginal distribution of the process $\{X_t, t \in T\}$.

The following theorem, due to Kolmogorov, establishes the existence of a stochastic process associated with a given family of finite-dimensional distributions satisfying the *consistence condition*:

Theorem 3 *Consider a family of probability measures*

$$\{P_{t_1,\ldots,t_n},\ t_1 < \cdots < t_n, n \geq 1, t_i \in T\}$$

such that:

1. $P_{t_1,\ldots,t_n}$ *is a probability on* $\mathbb{R}^n$
2. *(Consistence condition): If* $\{t_{k_1} < \cdots < t_{k_m}\} \subset \{t_1 < \cdots < t_n\}$, *then* $P_{t_{k_1},\ldots,t_{k_m}}$ *is the marginal of* $P_{t_1,\ldots,t_n}$, *corresponding to the indexes* $k_1, \ldots, k_m$.

Then, there exists a real-valued stochastic process $\{X_t, t \geq 0\}$ *defined in some probability space* $(\Omega, \mathcal{F}, P)$ *which has the family* $\{P_{t_1,\ldots,t_n}\}$ *as finite-dimensional marginal distributions.*

A real-valued process $\{X_t, t \geq 0\}$ is called a second order process provided $E(X_t^2) < \infty$ for all $t \in T$. The *mean* and the *covariance function* of a second order process $\{X_t, t \geq 0\}$ are defined by

$$\begin{aligned} m_X(t) &= E(X_t) \\ \Gamma_X(s,t) &= \mathrm{Cov}(X_s, X_t) \\ &= E((X_s - m_X(s))(X_t - m_X(t)). \end{aligned}$$

The *variance* of the process $\{X_t, t \geq 0\}$ is defined by

$$\sigma_X^2(t) = \Gamma_X(t,t) = \mathrm{Var}(X_t).$$

Example 12 Let X and Y be independent random variables. Consider the stochastic process with parameter $t \in [0, \infty)$

$$X_t = tX + Y.$$

The sample paths of this process are lines with random coefficients. The finite-dimensional marginal distributions are given by

$$P(X_{t_1} \le x_1, \dots, X_{t_n} \le x_n) = \int_{\mathbb{R}} F_X\left(\min_{1 \le i \le n} \frac{x_i - y}{t_i}\right) P_Y(dy).$$

Example 13 Consider the stochastic process

$$X_t = A\cos(\varphi + \lambda t),$$

where A and φ are independent random variables such that $E(A) = 0$, $E(A^2) < \infty$ and φ is uniformly distributed on $[0, 2\pi]$. This is a second order process with

$$\begin{aligned} m_X(t) &= 0 \\ \Gamma_X(s,t) &= \frac{1}{2}E(A^2)\cos\lambda(t-s). \end{aligned}$$

Example 14 *Arrival process:* Consider the process of arrivals of customers at a store, and suppose the experiment is set up to measure the interarrival times. Suppose that the interarrival times are positive random variables $X_1, X_2, \dots$. Then, for each $t \in [0, \infty)$, we put $N_t = k$ if and only if the integer k is such that

$$X_1 + \cdots + X_k \le t < X_1 + \cdots + X_{k+1},$$

and we put $N_t = 0$ if $t < X_1$. Then N_t is the number of arrivals in the time interval $[0, t]$. Notice that for each $t \ge 0$, N_t is a random variable taking values in the set $S = \mathbb{N}$. Thus, $\{N_t, t \ge 0\}$ is a continuous time process with values in the state space $\mathbb{N}$. The sample paths of this process are non-decreasing, right continuous and they increase by jumps of size 1 at the points $X_1 + \cdots + X_k$. On the other hand, $N_t < \infty$ for all $t \ge 0$ if and only if

$$\sum_{k=1}^{\infty} X_k = \infty.$$

Example 15 Consider a discrete time stochastic process $\{X_n, n = 0, 1, 2, \ldots\}$ with a finite number of states $S = \{1, 2, 3\}$. The dynamics of the process is as follows. You move from state 1 to state 2 with probability 1. From state 3 you move either to 1 or to 2 with equal probability 1/2, and from 2 you jump to 3 with probability 1/3, otherwise stay at 2. This is an example of a *Markov chain.*

A real-valued stochastic process $\{X_t, t \in T\}$ is said to be *Gaussian or normal* if its finite-dimensional marginal distributions are multi-dimensional Gaussian laws. The mean $m_X(t)$ and the covariance function $\Gamma_X(s,t)$ of a Gaussian process determine its finite-dimensional marginal distributions. Conversely, suppose that we are given an arbitrary function $m : T \to \mathbb{R}$, and a symmetric function $\Gamma : T \times T \to \mathbb{R}$, which is nonnegative definite, that is

$$\sum_{i,j=1}^{n} \Gamma(t_i, t_j) a_i a_j \geq 0$$

for all $t_i \in T$, $a_i \in \mathbb{R}$, and $n \geq 1$. Then there exists a Gaussian process with mean m and covariance function Γ.

Example 16 Let X and Y be random variables with joint Gaussian distribution. Then the process $X_t = tX + Y$, $t \geq 0$, is Gaussian with mean and covariance functions

$$\begin{aligned} m_X(t) &= tE(X) + E(Y), \\ \Gamma_X(s,t) &= t^2 \mathrm{Var}(X) + 2t\mathrm{Cov}(X,Y) + \mathrm{Var}(Y). \end{aligned}$$

Example 17 *Gaussian white noise*: Consider a stochastic process $\{X_t, t \in T\}$ such that the random variables X_t are independent and with the same law $N(0, \sigma^2)$. Then, this process is Gaussian with mean and covariance functions

$$\begin{aligned} m_X(t) &= 0 \\ \Gamma_X(s,t) &= \begin{cases} 1 & \text{if } s = t \\ 0 & \text{if } s \neq t \end{cases} \end{aligned}$$

Definition 4 *A stochastic process* $\{X_t, t \in T\}$ *is equivalent to another stochastic process* $\{Y_t, t \in T\}$ *if for each* $t \in T$

$$P\{X_t = Y_t\} = 1.$$

We also say that $\{X_t, t \in T\}$ is a version of $\{Y_t, t \in T\}$. Two equivalent processes may have quite different sample paths.

Example 18 Let ξ be a nonnegative random variable with continuous distribution function. Set $T = [0, \infty)$. The processes

$$\begin{aligned} X_t &= 0 \\ Y_t &= \begin{cases} 0 & \text{if} \quad \xi \neq t \\ 1 & \text{if} \quad \xi = t \end{cases} \end{aligned}$$

are equivalent but their sample paths are different.

Definition 5 *Two stochastic processes $\{X_t, t \in T\}$ and $\{Y_t, t \in T\}$ are said to be indistinguishable if $X.(\omega) = Y.(\omega)$ for all $\omega \notin N$, with $P(N) = 0$.*

Two stochastic process which have right continuous sample paths and are equivalent, then they are indistinguishable.

Two discrete time stochastic processes which are equivalent, they are also indistinguishable.

Definition 6 *A real-valued stochastic process $\{X_t, t \in T\}$, where T is an interval of $\mathbb{R}$, is said to be continuous in probability if, for any $\varepsilon > 0$ and every $t \in T$*

$$\lim_{s \to t} P(|X_t - X_s| > \varepsilon) = 0.$$

Definition 7 *Fix $p \geq 1$. Let $\{X_t, t \in T\}$be a real-valued stochastic process, where T is an interval of $\mathbb{R}$, such that $E\left(|X_t|^p\right) < \infty$, for all $t \in T$. The process $\{X_t, t \geq 0\}$ is said to be continuous in mean of order p if*

$$\lim_{s \to t} E\left(|X_t - X_s|^p\right) = 0.$$

Continuity in mean of order p implies continuity in probability. However, the continuity in probability (or in mean of order p) does not necessarily implies that the sample paths of the process are continuous.

In order to show that a given stochastic process have continuous sample paths it is enough to have suitable estimations on the moments of the increments of the process. The following continuity criterion by Kolmogorov provides a sufficient condition of this type:

Proposition 8 (Kolmogorov continuity criterion) *Let* $\{X_t, t \in T\}$ *be a real-valued stochastic process and* T *is a finite interval. Suppose that there exist constants* $a > 1$ *and* $p > 0$ *such that*

$$E\left(|X_t - X_s|^p\right) \leq c_T |t-s|^\alpha \tag{1}$$

for all $s, t \in T$. *Then, there exists a version of the process* $\{X_t, t \in T\}$ *with continuous sample paths.*

Condition (1) also provides some information about the modulus of continuity of the sample paths of the process. That means, for a fixed $\omega \in \Omega$, which is the order of magnitude of $X_t(\omega) - X_s(\omega)$, in comparison $|t-s|$. More precisely, for each $\varepsilon > 0$ there exists a random variable G_ε such that, with probability one,

$$|X_t(\omega) - X_s(\omega)| \leq G_\varepsilon(\omega)|t-s|^{\frac{\alpha}{p}-\varepsilon}, \tag{2}$$

para todo $s, t \in T$. Moreover, $E(G_\varepsilon^p) < \infty$.

Exercises

1.1 Consider a random variable X taking the values $-2, 0, 2$ with probabilities 0.4, 0.3, 0.3 respectively. Compute the expected values of X, $3X^2 + 5$, e^{-X}.

1.2 The headway X between two vehicles at a fixed instant is a random variable with

$$P(X \leq t) = 1 - 0.6e^{-0.02t} - 0.4e^{-0.03t},$$

$t \geq 0$. Find the expected value and the variance of the headway.

1.3 Let Z be a random variable with law $N(0, \sigma^2)$. From the expression

$$E(e^{\lambda Z}) = e^{\frac{1}{2}\lambda^2\sigma^2},$$

deduce the following formulas for the moments of Z:

$$\begin{aligned} E(Z^{2k}) &= \frac{(2k)!}{2^k k!}\sigma^{2k}, \\ E(Z^{2k-1}) &= 0. \end{aligned}$$

1.4 Let Z be a random variable with Poisson distribution with parameter λ. Show that the characteristic function of Z is

$$\varphi_Z(t) = \exp\left[\lambda\left(e^{it} - 1\right)\right].$$

As an application compute $E(Z^2)$, $\mathrm{Var}(Z)$ y $E(Z^3)$.

1.5 Let $\{Y_n, n \geq 1\}$ be independent and identically distributed non-negative random variables. Put $Z_0 = 0$, and $Z_n = Y_1 + \cdots + Y_n$ for $n \geq 1$. We think of Z_n as the time of the nth arrival into a store. The stochastic process $\{Z_n, n \geq 0\}$ is called a *renewal process.* Let N_t be the number of arrivals during $(0, t]$.

a) Show that $P(N_t \geq n) = P(Z_n \leq t)$, for all $n \geq 1$ and $t \geq 0$.

b) Show that for almost all ω, $\lim_{t\to\infty} N_t(\omega) = \infty$.

c) Show that almost surely $\lim_{t\to\infty} \frac{Z_{N_t}}{N_t} = a$, where $a = E(Y_1)$.

d) Using the inequalities $Z_{N_t} \leq t < Z_{N_t+1}$ show that almost surely

$$\lim_{t\to\infty} \frac{N_t}{t} = \frac{1}{a}.$$

1.6 Let X and U be independent random variables, U is uniform in $[0, 2\pi]$, and the probability density of X is for $x > 0$

$$f_X(x) = 2x^3 e^{-1/2x^4}.$$

Show that the process

$$X_t = X^2 \cos(2\pi t + U)$$

is Gaussian and compute its mean and covariance functions.

2 Jump Processes

2.1 The Poisson Process

A random variable $T : \Omega \to (0, \infty)$ has *exponential distribution of parameter* $\lambda > 0$ if

$$P(T > t) = e^{-\lambda t}$$

for all $t \geq 0$. Then T has a density function

$$f_T(t) = \lambda e^{-\lambda t} \mathbf{1}_{(0,\infty)}(t).$$

The mean of T is given by $E(T) = \frac{1}{\lambda}$, and its variance is $\mathrm{Var}(T) = \frac{1}{\lambda^2}$. The exponential distribution plays a fundamental role in continuous-time Markov processes because of the following result.

Proposition 9 (Memoryless property) *A random variable $T : \Omega \to (0, \infty)$ has an exponential distribution if and only if it has the following memoryless property*

$$P(T > s + t | T > s) = P(T > t)$$

for all $s, t \geq 0$.

Proof. Suppose first that T has *exponential distribution of parameter* $\lambda > 0$. Then

$$\begin{aligned} P(T &> s + t | T > s) = \frac{P(T > s + t)}{P(T > s)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(T > t). \end{aligned}$$

The converse implication follows from the fact that the function $g(t) = P(T > t)$ satisfies

$$g(s + t) = g(s)g(t),$$

for all $s, t \geq 0$ and $g(0) = 1$. ■

A stochastic process $\{N_t, t \geq 0\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ is said to be a *Poisson process of rate* λ if it verifies the following properties:

i) $N_t = 0$,

ii) for any $n \geq 1$ and for any $0 \leq t_1 < \cdots < t_n$ the increments $N_{t_n} - N_{t_{n-1}}, \ldots, N_{t_2} - N_{t_1}$, are independent random variables,

iii) for any $0 \leq s < t$, the increment $N_t - N_s$ has a Poisson distribution with parameter $\lambda(t-s)$, that is,

$$P(N_t - N_s = k) = e^{-\lambda(t-s)} \frac{[\lambda(t-s)]^k}{k!},$$

$k = 0, 1, 2, \ldots$, where $\lambda > 0$ is a fixed constant.

Notice that conditions i) to iii) characterize the finite-dimensional marginal distributions of the process $\{N_t, t \geq 0\}$. Condition ii) means that the Poisson process has independent and stationary increments.

A concrete construction of a Poisson process can be done as follows. Consider a sequence $\{X_n, n \geq 1\}$ of independent random variables with exponential law of parameter λ. Set $T_0 = 0$ and for $n \geq 1$, $T_n = X_1 + \cdots + X_n$. Notice that $\lim_{n\to\infty} T_n = \infty$ almost surely, because by the strong law of large numbers

$$\lim_{n\to\infty} \frac{T_n}{n} = \frac{1}{\lambda}.$$

Let $\{N_t, t \geq 0\}$ be the arrival process associated with the interarrival times X_n. That is

$$N_t = \sum_{n=1}^{\infty} n \mathbf{1}_{\{T_n \leq t < T_{n+1}\}}. \tag{3}$$

Proposition 10 *The stochastic process $\{N_t, t \geq 0\}$ defined in (3) is a Poisson process with parameter $\lambda > 0$.*

Proof. Clearly $N_0 = 0$. We first show that N_t has a Poisson distribution with parameter λt. We have

$$\begin{aligned} P(N_t &= n) = P(T_n \leq t < T_{n+1}) = P(T_n \leq t < T_n + X_{n+1}) \\ &= \int_{\{x \leq t < x+y\}} f_{T_n}(x) \lambda e^{-\lambda y} dx dy \\ &= \int_0^t f_{T_n}(x) e^{-\lambda(t-x)} dx. \end{aligned}$$

The random variable T_n has a gamma distribution $\Gamma(n,\lambda)$ with density

$$f_{T_n}(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} \mathbf{1}_{(0,\infty)}(x).$$

Hence,

$$P(N_t = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}.$$

Now we will show that $N_{t+s} - N_s$ is independent of the random variables $\{N_r, r \leq s\}$ and it has a Poisson distribution of parameter λt. Consider the event

$$\{N_s = k\} = \{T_k \leq t < T_{k+1}\},$$

where $0 \leq s < t$ and $0 \leq k$. On this event the interarrival times of the process $\{N_t - N_s, t \geq s\}$ are

$$\widetilde{X}_1 = X_{k+1} - (s - T_k)$$

and

$$\widetilde{X}_n = X_{k+n}$$

for $n \geq 2$. Then, conditional on $\{N_s = k\}$ and on the values of the random variables $X_1, \ldots, X_k$, due to the memoryless property of the exponential law and independence of the X_n, the interarrival times $\left\{\widetilde{X}_n, n \geq 1\right\}$ are independent and with exponential distribution of parameter λ. Hence, $\{N_{t+s} - N_s, t \geq 0\}$ has the same distribution as $\{N_t, t \geq 0\}$, and it is independent of $\{N_r, r \leq s\}$. ■

Notice that $E(N_t) = \lambda t$. Thus λ is the expected number of arrivals in an interval of unit length, or in other words, λ is the *arrival rate*. On the other hand, the expect time until a new arrival is $\frac{1}{\lambda}$. Finally, $\mathrm{Var}(N_t) = \lambda t$.

We have seen that the sample paths of the Poisson process are discontinuous with jumps of size 1. However, the Poisson process is continuous in mean of order 2:

$$E\left[(N_t - N_s)^2\right] = \lambda(t-s) + [\lambda(t-s)]^2 \xrightarrow{s \to t} 0.$$

Notice that we cannot apply here the Kolmogorov continuity criterion.

The Poisson process with rate $\lambda > 0$ can also be characterized as an integer-valued process, starting from 0, with non-decreasing paths, with independent increments, and such that, as $h \downarrow 0$, uniformly in t,

$$\begin{aligned} P(X_{t+h} - X_t &= 0) = 1 - \lambda h + o(h), \\ P(X_{t+h} - X_t &= 1) = \lambda h + o(h). \end{aligned}$$

Example 1 An item has a random lifetime whose distribution is exponential with parameter $\lambda = 0.0002$ (time is measured in hours). The expected lifetime of an item is $\frac{1}{\lambda} = 5000$ hours, and the variance is $\frac{1}{\lambda^2} = 25 \times 10^6$ hours^2. When it fails, it is immediately replaced by an identical item; etc. If N_t is the number of failures in $[0, t]$, we may conclude that the process of failures $\{N_t, t \geq 0\}$ is a Poisson process with rate $\lambda > 0$.

Suppose that the cost of a replacement is β euros, and suppose that the discount rate of money is $\alpha > 0$. So, the present value of all future replacement costs is

$$C = \sum_{n=1}^{\infty} \beta e^{-\alpha T_n}.$$

Its expected value will be

$$E(C) = \sum_{n=1}^{\infty} \beta E(e^{-\alpha T_n}) = \frac{\beta \lambda}{\alpha}.$$

For $\beta = 800$, $\alpha = 0.24/(365 \times 24)$ we obtain $E(C) = 5840$ euros.

The following result is an interesting relation between the Poisson processes and uniform distributions. This result say that the jumps of a Poisson process are as randomly distributed as possible.

Proposition 11 *Let $\{N_t, t \geq 0\}$ be a Poisson process. Then, conditional on $\{N_t, t \geq 0\}$ having exactly n jumps in the interval $[s, s+t]$, the times at which jumps occur are uniformly and independently distributed on $[s, s+t]$.*

Proof. We will show this result only for $n = 1$. By stationarity of increments, it suffices to consider the case $s = 0$. Then, for $0 \leq u \leq t$

$$\begin{aligned} P(T_1 \leq u | N_t = 1) &= \frac{P(\{T_1 \leq u\} \cap \{N_t = 1\})}{P(N_t = 1)} \\ &= \frac{P(\{N_u = 1\} \cap \{N_t - N_u = 0\})}{P(N_t = 1)} \\ &= \frac{\lambda u e^{-\lambda u} e^{-\lambda(t-u)}}{\lambda t e^{-\lambda t}} = \frac{u}{t}. \end{aligned}$$

■

2.2 Superposition of Poisson Processes

Let $L = \{L_t, t \geq 0\}$ and $M = \{M_t, t \geq 0\}$ be two independent Poisson processes with respective rates λ and μ. The process $N_t = L_t + M_t$ is called the *superposition* of the processes $L$and M.

Proposition 12 *$N = \{N_t, t \geq 0\}$ is a Poisson process of rate $\lambda + \mu$.*

Proof. Clearly, the process N has independent increments and $N_0 = 0$. Then, it suffices to show that for each $0 \leq s < t$, the random variable $N_t - N_s$ has a Poisson distribution of parameter $(\lambda + \mu)(t - s)$:

$$\begin{aligned} P(N_t - N_s &= n) = \sum_{k=0}^{n} P(L_t - L_s = k, M_t - M_s = n - k) \\ &= \sum_{k=0}^{n} \frac{e^{-\lambda(t-s)} [\lambda(t-s)]^k}{k!} \frac{e^{-\mu(t-s)} [\mu(t-s)]^{n-k}}{(n-k)!} \\ &= \frac{e^{-(\lambda+\mu)(t-s)} [(\lambda+\mu)(t-s)]^n}{n!}. \end{aligned}$$

■

Poisson processes are unique in this regard.

2.3 Decomposition of Poisson Processes

Let $N = \{N_t, t \geq 0\}$ be a Poisson process with rate λ. Let $\{Y_n, n \geq 1\}$ be a sequence of independent Bernoulli random variables with parameter $p \in (0, 1)$, independent of N. That is,

$$\begin{aligned} P(Y_n &= 1) = p \\ P(Y_n &= 0) = 1 - p. \end{aligned}$$

Set $S_n = Y_1 + \cdots + Y_n$. We think of S_n as the number of successes at the first n trials, and we suppose that the nth trial is performed at the time T_n of the nth arrival. In this way, the number of successes obtained during the time interval $[0, t]$ is

$$M_t = S_{N_t},$$

and the number of failures is

$$L_t = N_t - S_{N_t}.$$

Proposition 13 *The processes* $L = \{L_t, t \geq 0\}$ *and* $M = \{M_t, t \geq 0\}$ *are independent Poisson processes with respective rates* λp *and* $\lambda(1-p)$

Proof. It is sufficient to show that for each $0 \leq s < t$, the event

$$A = \{M_t - M_s = m, L_t - L_s = k\}$$

is independent of the random variables $\{M_u, L_u; u \leq s\}$ and it has probability

$$P(A) = \frac{e^{-\lambda p(t-s)}\left[\lambda p(t-s)\right]^m}{m!}\frac{e^{-\lambda(1-p)(t-s)}\left[\lambda(1-p)(t-s)\right]^k}{k!}.$$

We have

$$A = \{N_t - N_s = m + k, S_{N_t} - S_{N_s} = m\}.$$

The σ-field generated by the random variables $\{M_u, L_u; u \leq s\}$ coincides with the σ-field $\mathcal{H}$ generated by$\{N_u, u \leq s; Y_1, \ldots, Y_{N_s}\}$. Clearly the event A is independent of $\mathcal{H}$. Finally, we have

$$\begin{aligned}
P(A) &= \sum_{n=0}^{\infty} P(A \cap \{N_s = n\}) \\
&= \sum_{n=0}^{\infty} P(N_s = n, N_t - N_s = m + k, S_{m+k+n} - S_n = m) \\
&= \sum_{n=0}^{\infty} P(N_s = n, N_t - N_s = m + k) P(S_{m+k+n} - S_n = m) \\
&= P(N_t - N_s = m + k) P(S_{m+k} = m) \\
&= \frac{e^{-\lambda(t-s)}\left[\lambda(t-s)\right]^{m+k}}{(m+k)!}\frac{(m+k)!}{m!k!} p^m (1-p)^k.
\end{aligned}$$

■

2.4 Compound Poisson Processes

Let $N = \{N_t, t \geq 0\}$ be a Poisson process with rate λ. Consider a sequence $\{Y_n, n \geq 1\}$ of independent and identically distributed random variables, which are also independent of the process N. Set $S_n = Y_1 + \cdots + Y_n$. Then, the process

$$Z_t = Y_1 + \cdots + Y_{N_t} = S_{N_t},$$

with $Z_t = 0$ if $N_t = 0$, is called a *compound Poisson process*.

Example 2 Arrivals of customers into a store form a Poisson process N. The amount of money spent by the nth customer is a random variable Y_n which is independent of all the arrival times (including his own) and all the amounts spent by others. The total amount spent by the first n customers is $S_n = Y_1 + \cdots + Y_n$ if $n \geq 1$, and we set $Y_0 = 0$. Since the number of customers arriving in $(0,t]$ is N_t, the sales to customers arriving in $(0,t]$ total $Z_t = S_{N_t}$.

Proposition 14 *The compound Poisson process has independent increments, and the law of an increment $Z_t - Z_s$ has characteristic function*

$$e^{\left(\varphi_{Y_1}(x)-1\right)\lambda(t-s)},$$

where $\varphi_{Y_1}(x)$ denotes the characteristic function of Y_1.

Proof. Fix $0 \leq s < t$. The σ-field generated by the random variables $\{Z_u, u \leq s\}$ is included into the σ-field $\mathcal{H}$ generated by$\{N_u, u \leq s; Y_1, \ldots, Y_{N_s}\}$. As a consequence, the increment $Z_t - Z_s = S_{N_t} - S_{N_s}$ is independent of the σ-field generated by the random variables $\{Z_u, u \leq s\}$. Let us compute the characteristic function of this increment

$$\begin{aligned}
E(e^{ix(Z_t - Z_s)}) &= \sum_{m,k=0}^{\infty} E(e^{ix(Z_t-Z_s)}\mathbf{1}_{\{N_s=m, N_t-N_s=k\}}) \\
&= \sum_{m,k=0}^{\infty} E(e^{ix(S_{m+k}-S_m)}\mathbf{1}_{\{N_s=m, N_t-N_s=k\}}) \\
&= \sum_{m,k=0}^{\infty} E(e^{ixS_k})P(N_s = m, N_t - N_s = k) \\
&= \sum_{k=0}^{\infty} \left[E(e^{ixY_1})\right]^k \frac{e^{-\lambda(t-s)}\left[\lambda(t-s)\right]^k}{k!} \\
&= \exp\left[\left(\varphi_{Y_1}(x) - 1\right)\lambda(t-s)\right].
\end{aligned}$$

■

If $E(Y_1) = \mu$, then $E(Z_t) = \mu\lambda t$.

2.5 Non-Stationary Poisson Processes

A non-stationary Poisson process is a stochastic process $\{N_t, t \geq 0\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ which verifies the following properties:

i) $N_t = 0$,

ii) for any $n \geq 1$ and for any $0 \leq t_1 < \cdots < t_n$ the increments $N_{t_n} - N_{t_{n-1}}, \ldots, N_{t_2} - N_{t_1}$, are independent random variables,

iii) for any $0 \leq t$, the variable N_t has a Poisson distribution with parameter $a(t)$, where a is a non-decreasing function.

Assume that a is continuous, and define the *time inverse* of a by

$$\tau(t) = \inf\{s : a(s) > t\},$$

$t \geq 0$. Then τ is also a non-decreasing function, and the process $M_t = N_{\tau(t)}$ is a stationary Poisson process with rate 1.

Example 3 The times when successive demands occur for an expensive "fad" item form a non-stationary Poisson process N with the arrival rate at time t being

$$\lambda(t) = 5625te^{-3t},$$

for $t \geq 0$. That is

$$a(t) = \int_0^t \lambda(s)ds = 625(1 - e^{-3t} - 3te^{-3t}).$$

The total demand N_∞ for this item throughout $[0, \infty)$ has the Poisson distribution with parameter 625. If, based on this analysis, the company has manufactured 700 such items, the probability of a shortage would be

$$P(N_\infty > 700) = \sum_{k=701}^{\infty} \frac{e^{-625}(625)^k}{k!}.$$

Using the normal distribution to approximate this, we have

$$P(N_\infty > 700) = P(\frac{N_\infty - 625}{\sqrt{625}} > 3) \backsimeq \int_3^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0.0013.$$

The time of sale of the last item is T_{700}. Since $T_{700} \leq t$ if and only if $N_t \geq 700$, its distribution is

$$P(T_{700} \leq t) = P(N_t \geq 700) = 1 - \sum_{k=0}^{699} \frac{e^{-625} a(t)^k}{k!}.$$

Exercises

2.1 Let $\{N_t, t \geq 0\}$ be a Poisson process with rate $\lambda = 15$. Compute

a) $P(N_6 = 9)$,

b) $P(N_6 = 9, N_{20} = 13, N_{56} = 27)$

c) $P(N_{20} = 13 | N_9 = 6)$

d) $P(N_9 = 6 | N_{20} = 13)$

2.2 Arrivals of customers into a store form a Poisson process N with rate $\lambda = 20$ per hour. Find the expected number of sales made during an eight-hour business day if the probability that a customer buys something is 0.30.

2.3 A department store has 3 doors. Arrivals at each door form Poisson processes with rates $\lambda_1 = 110$, $\lambda_2 = 90$, $\lambda_3 = 160$ customers per hour. 30% of all customers are male. The probability that a male customer buys something is 0.80, and the probability of a female customer buying something is 0.10. An average purchase is 4.50 euros.

a) What is the average worth of total sales made in a 10 hour day?

b) What is the probability that the third female customer to purchase anything arrives during the first 15 minutes? What is the expected time of her arrival?

3 Markov Chains

Let I be a countable set that will be the state space of the stochastic processes studied in this section. A probability distribution π on I is formed by numbers $0 \leq \pi_i \leq 1$ such that $\sum_{i \in I} \pi_i = 1$. If I is finite we label the states $1, 2, \ldots, N$; then π is an N-vector.

We say that a matrix $P = (p_{ij}, i, j \in I)$ is *stochastic* if $0 \leq p_{ij} \leq 1$ and for each $i \in I$ we have

$$\sum_{j \in I} p_{ij} = 1.$$

This means that the rows of the matrix form a probability distribution on I. If I has N elements, then P is a square $N \times N$ matrix, and in the general case P will be an infinite matrix.

A Markov chain will be a discrete time stochastic process with spate space I. The elements p_{ij} will represent the probability that $X_{n+1} = j$ conditional on $X_n = i$.

Examples of stochastic matrices:

$$P_1 = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}, \tag{4}$$

where $0 < \alpha, \beta < 1$,

$$P_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}. \tag{5}$$

Definition 15 *We say that a stochastic process $\{X_n, n \geq 0\}$ is a Markov chain with initial distribution π and transition matrix P if for each $n \geq 0$, $i_0, i_1, \ldots, i_{n+1} \in I$*

(i) $P(X_0 = i_0) = \pi_{i_0}$

(ii) $P(X_{n+1} = i_{n+1} | X_0 = i_0, \ldots, X_n = i_n) = p_{i_n i_{n+1}}$

Condition (ii) is a particular version of the so called Markov property. This property refers to the lack of memory of a given stochastic system.

Notice that conditions (i) and (ii) determine the finite dimensional marginal distributions of the stochastic process. In fact, we have

$$\begin{aligned} P(X_0 &= i_0, X_1 = i_1, \ldots, X_n = i_n) \\ &= P(X_0 = i_0)P(X_1 = i_1 | X_0 = i_0) \\ \cdots P(X_n &= i_n | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) \\ &= \pi_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}. \end{aligned}$$

The matrix P^2 defined by $(P^2)_{ik} = \sum_{j \in I} p_{ij} p_{jk}$ is again a stochastic matrix. Similarly P^n is a stochastic matrix and we will denote the identity by P^0. That is, $(P^0)_{ij} = \delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases}$. We write $p_{ij}^{(n)} = (P^n)_{ij}$ for the (i, j) entry in P^n. We also denote by πP the product of the vector π times the matrix P. With this notation we have

$$\boxed{P(X_n = j) = (\pi P^n)_j} \tag{6}$$

$$\boxed{P(X_{n+m} = j | X_m = i) = p_{ij}^{(n)}} \tag{7}$$

Proof of (6):

$$\begin{aligned} P(X_n &= j) = \sum_{i_0 \in I} \cdots \sum_{i_{n-1} \in I} P(X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = j) \\ &= \sum_{i_0 \in I} \cdots \sum_{i_{n-1} \in I} \pi_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} j} = (\pi P^n)_j \,. \end{aligned}$$

Proof of (7): Assume $P(X_m = i) > 0$. Then

$$\begin{aligned} P(X_{n+m} &= j | X_m = i) = \frac{P(X_{n+m} = j, X_m = i)}{P(X_m = i)} \\ &= \frac{\sum_{i_0 \in I} \cdots \sum_{i_{n+m-1} \in I} \pi_{i_0} p_{i_0 i_1} \cdots p_{i_{m-1} i} p_{i i_{m+1}} \cdots p_{i_{n+m-1} j}}{(\pi P^m)_i} \\ &= \sum_{i_{m+1} \in I} \cdots \sum_{i_{n+m-1} \in I} p_{i i_{m+1}} \cdots p_{i_{n+m-1} j} = p_{ij}^{(n)}. \end{aligned}$$

Property (7) says that the probability that the chain moves from state i to state j in n steps is the (i, j) entry of the nth power of the transition matrix P. We call $p_{ij}^{(n)}$ the n-step transition probability from i to j.

The following examples give some methods for calculating $p_{ij}^{(n)}$:

Example 1 Consider a two-states Markov chain with transition matrix P_1 given in (4). From the relation $P_1^{n+1} = P_1^n P_1$ we can write

$$p_{11}^{(n+1)} = p_{12}^{(n)}\beta + p_{11}^{(n)}(1-\alpha).$$

We also know that $p_{12}^{(n)} + p_{11}^{(n)} = 1$, so by eliminating $p_{12}^{(n)}$ we get a recurrence relation for $p_{11}^{(n)}$:

$$p_{11}^{(n+1)} = \beta + p_{11}^{(n)}(1-\alpha-\beta),\ p_{11}^{(0)} = 1.$$

This recurrence relation has a unique solution:

$$p_{11}^{(n)} = \begin{cases} \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n & \text{for} \quad \alpha+\beta > 0 \\ 1 & \text{for} \quad \alpha+\beta = 0 \end{cases}$$

Example 2 Consider the three-state chain with matrix P_2 given in (5). Let us compute $p_{11}^{(n)}$. First compute the eigenvalues of P_2: $1, \frac{i}{2}, -\frac{i}{2}$. From this we deduce that

$$p_{11}^{(n)} = a + b\left(\frac{i}{2}\right)^n + c\left(-\frac{i}{2}\right)^n.$$

The answer we want is real, and

$$\left(\pm\frac{i}{2}\right)^n = \left(\frac{1}{2}\right)^n (\cos\frac{n\pi}{2} \pm \sin\frac{n\pi}{2})$$

so it makes sense to rewrite $p_{11}^{(n)}$ in the form

$$p_{11}^{(n)} = \alpha + \left(\frac{1}{2}\right)^n \left(\beta\cos\frac{n\pi}{2} + \gamma\sin\frac{n\pi}{2}\right)$$

for constants α, β and γ. The first values of $p_{11}^{(n)}$ are easy to write down, so we get equations to solve for α, β and γ:

$$\begin{aligned} 1 &= p_{11}^{(0)} = \alpha + \beta \\ 0 &= p_{11}^{(1)} = \alpha + \frac{1}{2}\gamma \\ 0 &= p_{11}^{(2)} = \alpha - \frac{1}{4}\beta \end{aligned}$$

so $\alpha = 1/5$, $\beta = 4/5$, $\gamma = -2/5$ and

$$p_{11}^{(n)} = \frac{1}{5} + \left(\frac{1}{2}\right)^n \left(\frac{\beta}{5}\cos\frac{n\pi}{2} - \frac{2}{5}\sin\frac{n\pi}{2}\right).$$

Example 3 *Sums of independent random variables.* Let Y_1, Y_2, ... be independent and identically distributed discrete random variables such that

$$P(Y_1 = k) = p_k,$$

for $k = 0, 1, 2, \ldots$.Put $X_0 = 0$, and for $n \geq 1$, $X_n = Y_1 + \cdots + Y_n$. The space state of this process is $\{0, 1, 2, \ldots\}$. Then

$$\begin{aligned} P(X_{n+1} &= i_{n+1}|X_0 = i_0, \ldots, X_n = i_n) \\ &= P(Y_{n+1} = i_{n+1} - i_n|X_0 = i_0, \ldots, X_n = i_n) = p_{i_{n+1}-i_n} \end{aligned}$$

by the independence of Y_{n+1} and $X_0, \ldots, X_n$. Thus, $\{X_n, n \geq 0\}$ is a Markov chain whose transition probabilities are $p_{ij} = p_{j-i}$ if $i \leq j$, and $p_{ij} = 0$ if $i > j$. The initial distribution is $\pi_0 = 1$, $\pi_i = 0$ for $i > 0$. In matrix form

$$P = \begin{pmatrix} p_0 & p_1 & p_2 & p_3 & \cdot & \cdot & \cdot \\ & p_0 & p_1 & p_2 & \cdot & \cdot & \cdot \\ & & p_0 & p_1 & \cdot & \cdot & \cdot \\ & & & p_0 & \cdot & \cdot & \cdot \\ & & & & \cdot & & \\ & & & & & \cdot & \\ 0 & & & & & & \cdot \end{pmatrix}.$$

Let N_n be the number of successes in n Bernoulli trials, where the probability of a success in any one is p. This a particular case of a sum of independent random variables when the variables Y_n have Bernoulli distribution. The transition matrix is, in this case

$$P = \begin{pmatrix} 1-p & p & & & 0 \\ & 1-p & p & & \\ & & 1-p & p & \\ & & & \cdot & \cdot \\ 0 & & & & \cdot \end{pmatrix}.$$

Moreover, for $j = i, \ldots, n+i$, $p_{ij}^{(n)} = \binom{n}{j-i} p^{j-i}(1-p)^{n-j+i}$.

Example 4 *Remaining Lifetime.* Consider some piece of equipment which is now in use. When it fails, it is replaced immediately by an identical one. When that one fails it is again replaced by an identical one, and so on. Let p_k be the probability that a new item lasts for k units of time, $k = 1, 2, \ldots$. Let X_n be the remaining lifetime of the item in use at time n. That is

$$X_{n+1} = \mathbf{1}_{\{X_n \geq 1\}}(X_n - 1) + \mathbf{1}_{\{X_n = 0\}}(Z_{n+1} - 1)$$

where Z_{n+1} is the lifetime of the item installed at n. Since the lifetimes of successive items installed are independent, $\{X_n, n \geq 0\}$ is a Markov chain whose state space is $\{0, 1, 2, \ldots\}$. The transition probabilities are as follows. For $i \geq 1$

$$\begin{aligned} p_{ij} &= P(X_{n+1} = j | X_n = i) = P(X_n - 1 = j | X_n = i) \\ &= \begin{cases} 1 & \text{if} \quad j = i - 1 \\ 0 & \text{if} \quad j \neq i - 1 \end{cases} \end{aligned}$$

and for $i = 0$

$$\begin{aligned} p_{0j} &= P(X_{n+1} = j | X_n = 0) = P(Z_{n+1} - 1 = j | X_n = i) \\ &= p_{j+1}. \end{aligned}$$

Thus the transition matrix is

$$P = \begin{pmatrix} p_1 & p_2 & p_3 & \cdot & \cdot \\ 1 & 0 & 0 & \cdot & \\ & 1 & 0 & \cdot & \cdot \\ & & \cdot & \cdot & \cdot \\ 0 & & & \cdot & \cdot \end{pmatrix}.$$

Proposition 16 (Markov property) *Let $\{X_n, n \geq 0\}$ be a Markov chain with transition matrix P and initial distribution π. Then, conditional on $X_m = i$, $\{X_{n+m}, n \geq 0\}$ is a Markov chain with transition matrix P and initial distribution δ_i, independent of the random variables $X_0, \ldots, X_m$.*

Proof. Consider an event of the form

$$A = \{X_0 = i_0, \ldots, X_m = i_m\},$$

where $i_m = i$. We have to show that

$$\begin{aligned} & P\left(\{X_m = i_m, \ldots, X_{m+n} = i_{m+n}\} \cap A | X_m = i\right) \\ = \; & \delta_{i i_m} p_{i_m i_{m+1}} \cdots p_{i_{m+n-1} i_{m+n}} P(A | X_m = i). \end{aligned}$$

This follows from

$$\begin{aligned} & P\left(\{X_m = i_m, \ldots, X_{m+n} = i_{m+n}\} \cap A | X_m = i\right) \\ = \; & P\left(X_0 = i_0, \ldots, X_{m+n} = i \;\right) / P\left(X_m = i\right) \\ = \; & \delta_{i i_m} p_{i_m i_{m+1}} \cdots p_{i_{m+n-1} i_{m+n}} P\left(A \;\right) / P\left(X_m = i\right). \end{aligned}$$

■

3.1 Classification of the States

We say that i *leads to* j and we write $i \to j$ if

$$P(X_n = j \text{ for some } n \geq 0 | X_0 = i) > 0.$$

We say that i *communicates with* j and write $i \leftrightarrow j$ if $i \to j$ and $j \to i$.

For distinct states i and j the following statements are equivalent:

1. $i \to j$
2. $p_{i_0 i_1} \cdots p_{i_{n-1} i_n} > 0$ for some states $i_0, i_1, \ldots, i_n$ with $i_0 = i$ and $i_n = j$
3. $p_{ij}^{(n)} > 0$ for some $n \geq 0$.

The equivalence between 1. and 3. follows from

$$p_{ij}^{(n)} \leq P(X_n = j \text{ for some } n \geq 0 | X_0 = i) \leq \sum_{n=0}^{\infty} p_{ij}^{(n)},$$

and the equivalence between 2. and 3. follows from

$$p_{ij}^{(n)} = \sum_{i_1, \ldots, i_{n-1}} p_{i i_1} \cdots p_{i_{n-1} j}.$$

It is clear from 2. that $i \to j$ and $j \to k$ imply $i \to k$. Also $i \to i$ for any state i. So $\leftrightarrow$ is an equivalence relation on I, which partitions I into *communicating classes*. We say that a class C is *closed* if

$$i \in C, i \to j \;\text{ imply } j \in C.$$

A state i is *absorbing* if $\{i\}$ is a closed class. A chain of transition matrix P where I is a single class is called *irreducible*.

The class structure can be deduced from the diagram of transition probabilities. For the example:

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

the classes are $\{1, 2, 3\}$, $\{4\}$, and $\{5, 6\}$, with only$\{5, 6\}$ being closed.

3.2 Hitting Times and Absorption Probabilities

Let $\{X_n, n \geq 0\}$ be a Markov chain with transition matrix P. The *hitting time* of a subset A of I is the random variable H^A with values in $\{0, 1, 2, \ldots\} \cup \{\infty\}$ defined by

$$H^A = \inf\{n \geq 0 : X_n \in A\} \tag{8}$$

where we agree that the infimum of the empty set is ∞. The probability starting from i that the chain ever hits A is then

$$h_i^A = P(H^A < \infty | X_0 = i).$$

When A is a closed class, h_i^A is called the *absorption probability.* The mean time taken for the chain to reach A, if $P(H^A < \infty) = 1$, is given by

$$k_i^A = E(H^A | X_0 = i) = \sum_{n=0}^{\infty} n P(H^A = n).$$

The vector of hitting probabilities $h^A = (h_i^A, i \in I)$ satisfies the following system of linear equations

$$\left\{ \begin{array}{rl} h_i^A = 1 & \text{for} \quad i \in A \\ h_i^A = \sum_{j \in I} p_{ij} h_j^A & \text{for} \quad i \notin A \end{array} \right. . \tag{9}$$

In fact, if $X_0 = i \in A$, then $H^A = 0$ so $h_i^A = 1$. If $X_0 = i \notin A$, then $H^A \geq 1$, so by the Markov property

$$P(H^A < \infty | X_1 = j, X_0 = i) = P(H^A < \infty | X_0 = j) = h_j^A$$

and

$$\begin{aligned} h_i^A &= P(H^A < \infty | X_0 = i) \\ &= \sum_{j\in I} P(H^A < \infty, X_1 = j | X_0 = i) \\ &= \sum_{j\in I} P(H^A < \infty | X_1 = j) P(X_1 = j | X_0 = i) \\ &= \sum_{j\in I} p_{ij} h_j^A. \end{aligned}$$

The vector of mean hitting times $k^A = (k_i^A, i \in I)$ satisfies the following system of linear equations

$$\begin{cases} k_i^A = 0 & \text{for} \quad i \in A \\ k_i^A = 1 + \sum_{j\notin A} p_{ij} k_j^A & \text{for} \quad i \notin A \end{cases}. \tag{10}$$

In fact, if $X_0 = i \in A$, then $H^A = 0$ so $k_i^A = 0$. If $X_0 = i \notin A$, then $H^A \geq 1$, so by the Markov property

$$E(H^A | X_1 = j, X_0 = i) = 1 + E(H^A | X_0 = j)$$

and

$$\begin{aligned} k_i^A &= E(H^A | X_0 = i) \\ &= \sum_{j\in I} E(H^A \mathbf{1}_{\{X_1 = j\}} | X_0 = i) \\ &= \sum_{j\in I} E(H^A | X_1 = j, X_0 = i) P(X_1 = j | X_0 = i) \\ &= 1 + \sum_{j\notin A} p_{ij} k_j^A. \end{aligned}$$

Remark: The systems of equations (9) and (10) may have more than one solution. In this case, the vector of hitting probabilities h^A and the vector of mean hitting times k^A are the minimal non-negative solutions of these systems.

Example 5 Consider the chain with the matrix

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Questions: Starting from 1, what is the probability of absorption in 3? How long it take until the chain is absorbed in 0 or 3?

Set $h_i = P(\text{hit } 3|X_0 = i)$. Clearly $h_0 = 0$, $h_3 = 1$, and

$$\begin{aligned} h_1 &= \frac{1}{2}h_0 + \frac{1}{2}h_2 \\ h_2 &= \frac{1}{2}h_1 + \frac{1}{2}h_3 \end{aligned}$$

Hence, $h_2 = \frac{2}{3}$ and $h_1 = \frac{1}{3}$.

On the other hand, if $k_i = E(\text{hit } \{0,3\}|X_0 = i)$, then $k_0 = k_3 = 0$, and

$$\begin{aligned} k_1 &= 1 + \frac{1}{2}k_0 + \frac{1}{2}k_2 \\ k_2 &= 1 + \frac{1}{2}k_1 + \frac{1}{2}k_3 \end{aligned}$$

Hence, $k_1 = 2$ and $k_2 = 2$.

Example 6 *(Gambler's ruin)* Fix $0 < p = 1 - q < 1$, and consider the Markov chain with state space $\{0, 1, 2, \ldots\}$ and transition probabilities

$$\begin{aligned} p_{00} &= 1, \\ p_{i,i-1} &= q, p_{i,i+1} = p, \text{ for } i = 1, 2, \ldots \end{aligned}$$

Assume $X_0 = i$. Then X_n represents the fortune of a gambler after the nth play, if he starts with a capital of i euros, and at any time with probability p he earns 1 euro and with probability q he losses euro. What is the probability of ruin? Here 0 is an absorbing state and we are interested in the probability of hitting zero. Set $h_i = P(\text{hit } 0|X_0 = i)$. Then h satisfies

$$\begin{aligned} h_0 &= 1, \\ h_i &= ph_{i+1} + qh_{i-1}, \text{for } i = 1, 2, \ldots \end{aligned}$$

If $p \neq q$ this recurrence relation has a general solution

$$h_i = A + B\left(\frac{q}{p}\right)^i.$$

If $p < q$, then the restriction $0 \leq h_i \leq 1$ forces $B = 0$, so $h_i = 1$ for all i. If $p > q$, then since $h_0 = 1$ we get a family of solutions

$$h_i = \left(\frac{q}{p}\right)^i + A\left(1 - \left(\frac{q}{p}\right)^i\right);$$

for a non-negative solution we must have $A \geq 0$, to the minimal non-negative solution is $h_i = \left(\frac{q}{p}\right)^i$. Finally, if $p = q$ the recurrence relation has a general solution

$$h_i = A + Bi$$

and again the restriction $0 \leq h_i \leq 1$ forces $B = 0$, so $h_i = 1$ for all i.

The previous model represents the case a gambler plays against a casino. Suppose now that the opponent gambler has an initial fortune of $N - i$, N being the total capital of both gamblers. In that case the state-space is finite $I = \{0, 1, 2, \ldots, N\}$ and the transition matrix of the Markov chain is

$$P = \begin{pmatrix} 1 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ q & 0 & p & & & & \cdot \\ 0 & q & 0 & p & & & \cdot \\ \cdot & 0 & q & 0 & p & & \cdot \\ \cdot & & & \cdot & \cdot & \cdot & \\ \cdot & & & & q & 0 & p \\ 0 & \cdot & \cdot & \cdot & 0 & 0 & 1 \end{pmatrix}..$$

If we set $h_i = P(\text{hit } 0 | X_0 = i)$. Then h satisfies

$$\begin{aligned} h_0 &= 1, \\ h_i &= ph_{i+1} + qh_{i-1}, \text{for } i = 1, 2, \ldots, N-1 \\ h_N &= 0. \end{aligned}$$

The solution to these equations is

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

if $p \neq q$, and

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

if $p = q = \frac{1}{2}$. Notice that letting $N \to \infty$ we recover the results for the case of an infinite fortune.

Example 7 *(Birth-and-death chain)* Consider the Markov chain with state space $\{0, 1, 2, \ldots\}$ and transition probabilities

$$\begin{aligned} p_{00} &= 1, \\ p_{i,i-1} &= q_i, p_{i,i+1} = p_i, \text{ for } i = 1, 2, \ldots \end{aligned}$$

where, for $i = 1, 2, \ldots$, we have $0 < p_i = 1 - q_i < 1$. Assume $X_0 = i$. As in the preceding example, 0 is an absorbing state and we wish to calculate the absorption probability starting from i. Such a chain may serve as a model for the size of a population, recorded each time it changes, p_i being the probability that we get a birth before a death in a population of size i. Set $h_i = P(\text{hit } 0 | X_0 = i)$. Then h satisfies

$$\begin{aligned} h_0 &= 1, \\ h_i &= p_i h_{i+1} + q_i h_{i-1}, \text{for } i = 1, 2, \ldots \end{aligned}$$

This recurrence relation has variable coefficients so the usual technique fails. Consider $u_i = h_{i-1} - h_i,$, then $p_i u_{i+1} = q_i u_i$, so

$$u_{i+1} = \frac{q_i}{p_i} u_i = \frac{q_i q_{i-1} \cdots q_1}{p_i p_{i-1} \cdots p_1} u_1 = \gamma_i u_1,$$

where

$$\gamma_i = \frac{q_i q_{i-1} \cdots q_1}{p_i p_{i-1} \cdots p_1}.$$

Hence,

$$h_i = h_0 - (u_1 + \cdots + u_i) = 1 - A(\gamma_0 + \gamma_1 + \cdots + \gamma_{i-1}),$$

where $\gamma_0 = 1$ and $A = u_1$ is a constant to be determined. There are two different cases:

(i) If $\sum_{i=0}^{\infty} \gamma_i = \infty$, the restriction $0 \leq h_i \leq 1$ forces $A = 0$ and $h_i = 1$ for all i.

(ii) If $\sum_{i=0}^{\infty} \gamma_i < \infty$ we can take $A > 0$ as long as $1 - A(\gamma_0 + \gamma_1 + \cdots + \gamma_{i-1}) \geq 0$ for all i. Thus the minimal non-negative solution occurs when $A = \left(\sum_{i=0}^{\infty} \gamma_i\right)^{-1}$ and then

$$h_i = \frac{\sum_{j=i}^{\infty} \gamma_j}{\sum_{j=0}^{\infty} \gamma_j}.$$

In this case, for $i = 1, 2, \ldots$, we have $h_i < 1$, so the population survives with positive probability.

3.3 Stopping Times

Consider an non-decreasing family of σ-fields

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$$

in a probability space $(\Omega, \mathcal{F}, P)$. That is $\mathcal{F}_n \subset \mathcal{F}$ for all n.

A random variable $T : \Omega \to \{0, 1, 2, \ldots\} \cup \{\infty\}$ is called a *stopping time* if the event $\{T = n\}$ belongs to $\mathcal{F}_n$ for all $n = 0, 1, 2, \ldots$. Intuitively, $\mathcal{F}_n$ represents the information available at time n, and given this information you know when T occurs.

A discrete-time stochastic process $\{X_n, n \geq 0\}$ is said to be *adapted* to the family $\{\mathcal{F}_n, n \geq 0\}$ if X_n is $\mathcal{F}_n$-measurable for all n. In particular, $\{X_n, n \geq 0\}$ is always adapted to the family $\mathcal{F}_n^X$, where $\mathcal{F}_n^X$ as the σ-field generated by the random variables $\{X_0, \ldots, X_n\}$.

Example 8 Consider a discrete time stochastic process $\{X_n, n \geq 0\}$ adapted to $\{\mathcal{F}_n, n \geq 0\}$. Let A be a subset of the space state. Then the first hitting time H^A introduced in (8) is a stopping time because

$$\{H^A = n\} = \{X_0 \notin A, \ldots, X_{n-1} \notin A, X_n \in A\}.$$

The condition $\{T = n\} \in \mathcal{F}_n$ for all $n \geq 0$ is equivalent to $\{T \leq n\} \in \mathcal{F}_n$ for all $n \geq 0$. This is an immediate consequence of the relations

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

The extension of the notion of stopping time to continuous time will be inspired by this equivalence. Consider now a continuous parameter non-decreasing family of σ-fields $\{\mathcal{F}_t, t \geq 0\}$ in a probability space $(\Omega, \mathcal{F}, P)$. A random variable $T : \Omega \to [0, \infty]$ is called a *stopping time* if the event $\{T \leq t\}$ belongs to $\mathcal{F}_t$ for all $t \geq 0$.

Example 9 Consider a real-valued stochastic process $\{X_t, t \geq 0\}$ with continuous trajectories, and adapted to $\{\mathcal{F}_t, t \geq 0\}$. The *first passage time* for a level $a \in \mathbb{R}$ defined by

$$T_a := \inf\{t > 0 : X_t = a\}$$

is a stopping time because

$$\{T_a \leq t\} = \left\{ \sup_{0 \leq s \leq t} X_s \geq a \right\} = \left\{ \sup_{0 \leq s \leq t, s \in \mathbb{Q}} X_s \geq a \right\} \in \mathcal{F}_t.$$

Properties of stopping times:

1. If S and T are stopping times, so are $S \vee T$ and $S \wedge T$. In fact, this a consequence of the relations

$$\begin{aligned} \{S \vee T \leq t\} &= \{S \leq t\} \cup \{T \leq t\}, \\ \{S \wedge T \leq t\} &= \{S \leq t\} \cap \{T \leq t\}. \end{aligned}$$

2. Given a stopping time T, we can define the σ-field

$$\mathcal{F}_T = \{A : A \cap \{T \leq t\} \in \mathcal{F}_t, \text{ for all } t \geq 0\}.$$

   $\mathcal{F}_T$ is a σ-field because it contains the empty set, it is stable by complements due to

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

   and it is stable by countable intersections.

3. If S and T are stopping times such that $S \leq T$, then $\mathcal{F}_S \subset \mathcal{F}_T$. In fact, if $A \in \mathcal{F}_S$, then

$$A \cap \{T \leq t\} = [A \cap \{S \leq t\}] \cap \{T \leq t\} \in \mathcal{F}_t$$

   for all $t \geq 0$.

4. Let $\{X_t\}$ be an adapted stochastic process (with discrete or continuous parameter) and let T be a stopping time. If the parameter is continuous, we assume that the trajectories of the process $\{X_t\}$ are right-continuous. Then the random variable

$$X_T(\omega) = X_{T(\omega)}(\omega)$$

   is $\mathcal{F}_T$-measurable. In discrete time this property follows from the relation

$$\{X_T \in B\} \cap \{T = n\} = \{X_n \in B\} \cap \{T = n\} \in \mathcal{F}_n$$

   for any subset B of the state space (Borel set if the state space is $\mathbb{R}$).

Proposition 17 (Strong Markov property) *Let $\{X_n, n \geq 0\}$ be a Markov chain with transition matrix P and initial distribution π. Let T be a stopping time of $\{\mathcal{F}_n^X\}$. Then, conditional on $T < \infty$, and $X_T = i$, $\{X_{T+n}, n \geq 0\}$ is a Markov chain with transition matrix P and initial distribution δ_i, independent of $X_0, \ldots, X_T$.*

Proof. Let B be an event determined by the random variables $X_0, \ldots, X_T$. By the Markov property at time m

$$\begin{aligned}
& P\left(\{X_{T+1} = j_1, \ldots, X_{T+n} = j_n\} \cap B \cap \{T = m\} \cap \{X_T = i\}\right) \\
= \; & P\left(\{X_{m+1} = j_1, \ldots, X_{m+n} = j_n\} \cap B \cap \{T = m\} \cap \{X_T = i\}\right) \\
= \; & P(X_1 = j_1, \ldots, X_n = j_n | X_0 = i) \\
& \times P\left(B \cap \{T = m\} \cap \{X_T = i\}\right).
\end{aligned}$$

Now sum over $m = 0, 1, 2, \ldots$, and divide by $P(T < \infty, X_T = i)$ to obtain

$$\begin{aligned}
& P\left(\{X_{T+1} = j_1, \ldots, X_{T+n} = j_n\} \cap B | T < \infty, X_T = i\right) \\
= \; & p_{ij_1} p_{j_1 j_2} \cdots p_{j_{n-1}, j_n} P\left(B | T < \infty, X_T = i\right).
\end{aligned}$$

■

3.4 Recurrence and Transience

Let $\{X_n, n \geq 0\}$ be a Markov chain with transition matrix P . We say that a state i is *recurrent* if

$$P(X_n = i, \text{ for infinitely many } n) = 1.$$

We say that a state i is *transient* if

$$P(X_n = i, \text{ for infinitely many } n) = 0.$$

Thus a recurrent state is one to which you keep coming back and a transient state is one which you eventually leave for ever. We shall show that every state is either recurrent or transient.

We denote by T_i the *first passage time* to state i, which is a stopping time defined by

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

where $\inf \emptyset = \infty$. We define inductively the *rth passage time* $T_i^{(r)}$ to state i by

$$T_i^{(0)} = 0, T_i^{(1)} = T_i$$

and, for $r = 0, 1, 2, \ldots$,

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

The length of the rth excursion to i is then

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

Lemma 18 *For $r = 2, 3, \ldots$ conditional on $T_i^{(r-1)} < \infty$, $S_i^{(r)}$ is independent of $\{X_m : m \leq T_i^{(r-1)}\}$ and*

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

Proof. Apply the strong Markov property to the stopping time $T = T_i^{(r-1)}$. It is clear that $X_T = i$ on $T < \infty$. So, conditional to $T < \infty$, $\{X_{T+n}, n \geq 0\}$ is a Markov chain with transition matrix P and initial distribution δ_i, independent of $X_0, \ldots, X_T$. But

$$S_i^{(r)} = \inf\{n \geq 1 : X_{T+n} = i\},$$

so $S_i^{(r)}$ is the first passage time of $\{X_{T+n}, n \geq 0\}$ to state i. ■

We will make use of the notation $P(\cdot|X_0 = i) = P_i(\cdot)$ and $E(\cdot|X_0 = i) = E_i(\cdot)$.

The *number of visits* V_i to state i is defined by

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

and

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

We can also compute the distribution of V_i under P_i in terms of the *return probability*

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

Lemma 19 *For $r = 0, 1, 2, \ldots,$ we have $P_i(V_i > r) = f_i^r$.*

Proof. Observe that if $X_0 = i$ then $\{V_i > r\} = \{T_i^{(r)} < \infty\}$. When $r = 0$ the result is true. Suppose inductively that it is true for r, then

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

■

The next theorem provides two criteria for recurrence or transience, one in terms of the return probability, the other in terms of the n-step transition probabilities:

Theorem 20 *Every state i is recurrent or transient, and:*

(i) *if $P_i(T_i < \infty) = 1$, then i is recurrent and $\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$,*

(ii) *if $P_i(T_i < \infty) < 1$, then i is transient and $\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty$.*

Proof. If $P_i(T_i < \infty) = 1$, then, by Lemma 42,

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

so i is recurrent and

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = E_i(V_i) = \infty.$$

On the other hand, if $f_i = P_i(T_i < \infty) < 1$, then by Lemma 42,

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = E_i(V_i) = \sum_{r=0}^{\infty} P_i(V_i > r) = \sum_{r=0}^{\infty} f_i^r = \frac{1}{1 - f_1} < \infty$$

so $P_i(V_i = \infty) = 0$ and i is transient. ■

From this theorem we can solve completely the problem of recurrence or transience for Markov chains with finite state space. First we show that recurrence and transience are class properties.

Theorem 21 *Let C be a communicating class. Then either all states in C are transient or all are recurrent.*

Proof. Take any pair of states $i, j \in C$ and suppose that i is transient. There exists $n, m \geq 0$ with $p_{ij}^{(n)} > 0$ and $p_{ji}^{(m)} > 0$, and, for all $r \geq 0$

$$p_{ii}^{(n+r+m)} \geq p_{ij}^{(n)} p_{jj}^{(r)} p_{ji}^{(m)}$$

so

$$\sum_{r=0}^{\infty} p_{jj}^{(r)} \leq \frac{1}{p_{ij}^{(n)} p_{ji}^{(m)}} \sum_{r=0}^{\infty} p_{ii}^{(n+r+m)} < \infty.$$

Hence, j is also transient. ■

Theorem 22 *Every recurrent class is closed.*

Proof. Let C be a class which is not closed. Then there exists $i \in C$, $j \notin C$ and $m \geq 1$ with

$$P_i(X_m = j) > 0.$$

Since we have

$$P_i(\{X_m = j\} \cap \{X_n = i \text{ for infinitely many } n\}) = 0$$

this implies that

$$P_i(X_n = i \text{ for infinitely many } n) < 1$$

so i is not recurrent, and so neither is C. ■

Theorem 23 *Every finite closed class is recurrent.*

Proof. Suppose C is closed and finite and that $\{X_n, n \geq 0\}$ starts in C. Then for some $i \in C$ we have

$$\begin{aligned} 0 &< P(X_n = i \text{ for infinitely many } n) \\ &= P(X_n = i \text{ for some } n) P_i(X_n = i \text{ for infinitely many } n) \end{aligned}$$

by the strong Markov property. This shows that i is not transient, so C is recurrent. ■

So, any state in a non-closed class is transient. Finite closed classes are recurrent, but infinite closed classes may be transient.

Example 10 Consider the Markov chain with state space $I = \{1, 2, \ldots, 10\}$ and transition matrix

$$P = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & \frac{2}{3} & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & \frac{1}{3} \end{pmatrix}.$$

From the transition graph we see that $\{1, 3\}$, $\{2, 7, 9\}$ and $\{6\}$ are irreducible closed sets. These sets can be reached from the states $4, 5, 8$ and 10 (but not viceversa). The states $4, 5, 8$ and 10 are transient, state 6 is absorbing and states $1, 3, 2, 7, 9$ are recurrent.

Example 11 *(Random walk)* Let $\{X_n, n \geq 0\}$ be a Markov chain with state space $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$ and transition probabilities

$$p_{i,i-1} = q, p_{i,i+1} = p, \text{ for } i \in \mathbb{Z},$$

where $0 < p = 1 - q < 1$. This chain is called a random walk: If a person is at position i after step n, his next step leads him to either $i + 1$ or $i - 1$ with respective probabilities p and q. The Markov chain of Example 6 (Gambler's ruin) is a random walk on $\mathbb{N} = \{0, 1, 2, \ldots\}$ with 0 as absorbing state.

All states can be reached from each other, so the chain is irreducible. Suppose we start at i. It is clear that we cannot return to 0 after an odd number of steps, so $p_{ii}^{(2n+1)} = 0$ for all n. Any given sequence of steps of length $2n$ from i to i occurs with probability $p^n q^n$, there being n steps up and n steps down, and the number of such sequences is the number of ways of choosing the n steps up from$2n$. Thus

$$p_{ii}^{(2n)} = \binom{2n}{n} p^n q^n.$$

Using Stirling's formula $n! \backsim \sqrt{2\pi n}(n/e)^n$ we obtain

$$p_{ii}^{(2n)} = \frac{(2n)!}{(n!)^2} p^n q^n \backsim \frac{(4pq)^n}{\sqrt{2\pi n}}$$

as n tends to infinity.

In the symmetric case $p = q = \frac{1}{2}$, so

$$p_{ii}^{(2n)} \backsim \frac{1}{\sqrt{\pi n}}$$

and the random walk is recurrent because $\sum_{n=0}^{\infty} 1/\sqrt{n} = \infty$. On the other hand, if $p \neq q$, then $4pq = r < 1$, so

$$p_{ii}^{(2n)} \backsim \frac{r^n}{\sqrt{\pi n}}$$

and the random walk is transient because $\sum_{n=0}^{\infty} r^n/\sqrt{n} < \infty$.

Assuming $X_0 = 0$, we can write $X_n = Y_1 + \cdots + Y_n$, where the Y_i are independent random variables, identically distributed with $P(Y_1 = 1) = p$ and $P(Y_1 = -1) = 1 - p$. By the law of large numbers we know that

$$\frac{X_n}{n} \stackrel{\text{a.s.}}{\to} 2p - 1.$$

So $\lim_{n\to\infty} X_n = +\infty$ if $p > 1/2$ and $\lim_{n\to\infty} X_n = -\infty$ if $p < 1/2$. This also implies that the chain is transient if $p \neq q$.

3.5 Invariant Distributions

We say a distribution π on the state space I is *invariant* for a Markov chain with transition probability matrix P if

$$\pi P = \pi.$$

If the initial distribution π of a Markov chain is invariant, then the distribution of X_n is π for all n. In fact, from (6) this distribution is $\pi P^n = \pi$.

Let the state space I be finite. Invariant distributions are row eigenvectors of P with eigenvalue 1. The row sums of P are all 1, so the column vector of ones is an eigenvector with eigenvalue 1. So, P must have a row eigenvector with eigenvalue 1, that is, invariant distributions always exist. Invariant distributions are related to the limit behavior of P^n. Let us first show the following simple result.

Proposition 24 *Let the state space I be finite. Suppose for some $i \in I$ that*

$$p_{ij}^{(n)} \to \pi_j$$

as n tends to infinity, for all $j \in I$. Then $\pi = (\pi_j, j \in I)$ is an invariant distribution.

Proof. We have

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

and

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

where we have used the finiteness of I to justify the interchange of summation and limit operation. Hence π is an invariant distribution. ∎

This result is no longer true for infinite Markov chains.

Example 12 Consider the two-state Markov chain given in (4) with $\alpha \neq 0, 1$. The invariant distribution is $(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta})$. Notice that these are the rows of $\lim_{n \to \infty} P^n$.

Example 13 Consider the two-state Markov chain given in (5). To find an invariant distribution we must solve the linear equation $\pi P = \pi$. This equation together with the restriction

$$\pi_1 + \pi_2 + \pi_3 = 1$$

yields $\pi = (1/5, 2/5, 2/5)$.

Consider a Markov chain $\{X_n, n \geq 0\}$ irreducible and recurrent with transition matrix P. For a fixed state i, consider for each j the expected time spent in j between visits to i:

$$\gamma_j^i = E_i \sum_{n=0}^{T_i - 1} \mathbf{1}_{\{X_n = j\}}.$$

Clearly, $\gamma_i^i = 1$, and

$$\sum_{j \in I} \gamma_j^i = E_i\left(T_i\right) := m_i$$

is the *expected return time* in i. Notice that although the state i is recurrent, that is, $P_i(T_i < \infty) = 1$, m_i can be infinite. If the expected return time $m_i = E_i(T_i)$ is finite, then we say that i is *positive recurrent*. A recurrent state which fails to have this stronger property is called *null recurrent*. In a finite Markov chain all recurrent states are positive recurrent. If a state is positive recurrent, all states in its class are also positive recurrent.

The numbers $\{\gamma_j^i, j \in I\}$ form a measure on the state space I, with the following properties:

1. $0 < \gamma_j^i < \infty$
2. $\gamma^i P = \gamma^i$, that is γ^i is an invariant measure for the stochastic matrix P. Moreover, it is the unique invariant measure, up to a multiplicative constant.

Proposition 25 *Let $\{X_n, n \geq 0\}$ be an irreducible and recurrent Markov chain with transition matrix P. Let γ be an invariant measure for P. Then one of the following situations occurs:*

(i) $\gamma(I) = +\infty$. *Then the chain is null recurrent and there is no invariant distribution.*

(ii) $\gamma(I) < \infty$. *The chain is positive recurrent and there is a unique invariant distribution given by*

$$\pi_i = \frac{1}{m_i}.$$

Example 14 The symmetric random walk of Example 11 is irreducible and recurrent in the case $p = \frac{1}{2}$. The measure $\pi_i = 1$ for all i, is invariant. So we are in case (i) and there is no invariant distribution. The chain is null recurrent.

Example 15 Consider the asymmetric random walk on $\{0, 1, 2, \ldots\}$ with reflection at 0. The transition probabilities are

$$\begin{aligned} p_{i,i+1} &= p \\ p_{i-1,i} &= 1 - p = q \end{aligned}$$

for $i = 1, 2, \ldots$ and $p_{01} = 1$. So the transition matrix is

$$P = \begin{pmatrix} 0 & 1 & & & \\ q & 0 & p & & \\ & q & 0 & p & \\ & & & \cdot & \cdot \\ & & & & \cdot \end{pmatrix}.$$

This is an irreducible Markov chain. The invariant measure equation $\pi P = \pi$ leads to the recursive system

$$\pi_i = \pi_{i-1} p + \pi_{i+1} q$$

for $i = 2, 3, \ldots$ and

$$\begin{aligned} \pi_1 q &= \pi_0 \\ \pi_0 + \pi_2 q &= \pi_1. \end{aligned}$$

We obtain $\pi_2 q = \pi_1 p$, and recursively this implies $\pi_i p = \pi_{i+1} q$. Hence,

$$\pi_i = \pi_0 \frac{\pi_1 \pi_2 \cdots \pi_i}{\pi_0 \pi_1 \cdots \pi_{i-1}} = \pi_0 \frac{1}{q} \left(\frac{p}{q} \right)^{i-1}$$

for $i \geq 1$.If $p > q$, the chain is transient because, by the results obtained in the case of the random walk on the integers we have $\lim_{n \to \infty} X_n = +\infty$. Also, in that case there is no bounded invariant measure. If $p = \frac{1}{2}$, we already know, by the results in the case of the random walk in the integers, that the chain is recurrent. Notice that π_i must be constant, so again there is no invariant distribution and the chain is null recurrent. If $p < q$, the chain is positive recurrent and the unique invariant distribution is

$$\pi_i = \frac{1}{2q} \left(1 - \frac{p}{q} \right) \left(\frac{p}{q} \right)^{i-1},$$

if $i \geq 1$ and $\pi_0 = \frac{1}{2} \left(1 - \frac{p}{q} \right)$.

3.6 Convergence to Equilibrium

We shall study the limit behavior of the n-step transition probabilities $p_{ij}^{(n)}$ as $n \to \infty$. We have seen above that if the limit $\lim_{n\infty} p_{ij}^{(n)}$ exists as $n \to \infty$

for some i and for all j, and the state-space is finite, then the limit must be an invariant distribution. However, as the following example shows, the limit does not always exist.

Example 16 Consider the two-states Markov chain

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

Then, $P^{2n} = I$, and $P^{2n+1} = P$ for all n. Thus, $p_{ij}^{(n)}$ fails to converge for all i, j.

Let P be a transition matrix. Given a state $i \in I$ set

$$I(i) = \{n \geq 1 : p_{ii}^{(n)} > 0\}.$$

The greatest common divisor of $I(i)$ is called the period of the state i. In an irreducible chain all states have the same period. If a states i verifies $p_{ii}^{(n)} > 0$ for all sufficiently large n,then the period of i is one and this state is called *aperiodic.*

Proposition 26 *Suppose that the transition matrix P is irreducible, recurrent positive and aperiodic. Let $\pi_i = \frac{1}{m_i}$ be the unique invariant distribution. Then*

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$$

for all $i, j \in I$.

Proposition 27 *Suppose that the transition matrix P is irreducible, recurrent positive with period d aperiodic. Let $\pi_i = \frac{1}{m_i}$ be the unique invariant distribution. Then, for all $i, j \in I$, there exists $0 \leq r < d$, with $r = 0$ if $i = j$, and such that*

$$\begin{aligned} \lim_{n\to\infty} p_{ij}^{(nd+r)} &= d\pi_j \\ p_{ij}^{(nd+s)} &= 0 \end{aligned}$$

for all $s \neq r$, $0 \leq s < d$.

Finally let us state the following version of the *ergodic theorem.*

Proposition 28 *Let* $\{X_n, n \geq 0\}$ *be an irreducible and positive recurrent Markov chain with transition matrix* P*. Let* π *be the invariant distribution. Then for any function on* I *integrable with respect to* π *we have*

$$\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \stackrel{\text{a.s.}}{\to} \sum_{i\in I} f(i)\pi_i.$$

In particular,

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbf{1}_{\{X_k=j\}} \stackrel{\text{a.s.}}{\to} \pi_j.$$

3.7 Continuous Time Markov Chains

Let I be a countable set. The basic data for a continuous-time Markov chain on I are given in the form of a so called Q-matrix. By definition a Q-matrix on I is any matrix $Q = (q_{ij} : i, j \in I)$ which satisfies the following conditions:

(i) $q_{ii} \leq 0$ for all i,

(ii) $q_{ij} \geq 0$ for all $i \neq j$,

(iii) $\sum_{j\in I} q_{ij} = 0$ for all i.

In each row of Q we can choose the off-diagonal entries to be any non-negative real numbers, subject to the constraint the the off-diagonal row sum is finite:

$$q_i = \sum_{j\neq i} q_{ij} < \infty.$$

The diagonal entry q_{ii} is then $-q_i$, making the total row sum zero.

A Q-matrix Q defines a stochastic matrix $\Pi = (\pi_{ij} : i, j \in I)$ called the *jump matrix* defined as follows: If $q_i \neq 0$, the ith row of Π is

$$\begin{aligned} \pi_{ij} &= q_{ij}/q_i, \text{ if } j \neq i \text{ and} \\ \pi_{ii} &= 0. \end{aligned}$$

If $q_i = 0$ then $\pi_{ij} = \delta_{ij}$. From Π and the diagonal elements q_i we can recover the matrix Q.

Example 17 The Q-matrix

$$Q = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -1 & 0 \\ 2 & 1 & -3 \end{pmatrix} \tag{11}$$

has the associated jump matrix

$$\Pi = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 2/3 & 1/3 & 0 \end{pmatrix}.$$

Here is the construction of a continuous-time Markov chain with Q-matrix Q and initial distribution π. Consider a discrete-time Markov chain $\{Y_n, n \geq 0\}$ with initial distribution π, and transition matrix Π. The stochastic process $\{X_t, t \geq 0\}$ will visit successively the sates $Y_0, Y_1, Y_2, \ldots$, starting from $X_0 = Y_0$. Denote by $R_1, \ldots, R_n, \ldots$the holding times in these states. Then we want that conditional on $Y_0, \ldots, Y_{n-1}$, the holding times $R_1, \ldots, R_n$ are independent exponential random variables of parameters $q_{Y_0}, \ldots, q_{Y_{n-1}}$. In order to construct such process, let $\xi_1, \xi_2, \ldots$ be independent exponential random variables of parameter 1, independent of the chain $\{Y_n, n \geq 0\}$. Set

$$R_n = \frac{\xi_n}{q_{Y_{n-1}}},$$

$T_n = R_1 + \cdots + R_n$, and

$$X_t = \begin{cases} Y_n & \text{if} \quad T_n \leq t < T_{n+1} \\ \infty & \text{otherwise} \end{cases}.$$

If $q_{Y_n} = 0$, then we put $R_{n+1} = \infty$, and we have $X_t = Y_n$ for all $T_n \leq t$.

The random time

$$\zeta = \sup_n T_n = \sum_{n=0}^{\infty} R_n$$

is called the *explosion time*. We say that the Markov chain X_t is not explosive if $P(\zeta = \infty) = 1$. The following are sufficient conditions for no explosion:

1) I is finite,

2) $\sup_{i \in I} q_i < \infty$,

3) $X_0 = i$ and i is recurrent for the jump chain.

Set
$$P(t) = \sum_{k=0}^{\infty} \frac{t^k Q^k}{k!}.$$
Then, $\{P(t), t \geq 0\}$ is a continuous family of matrices that satisfies the following properties:

(i) S*semigroup property*:
$$P(s)P(t) = P(s+t).$$

(ii) For all $t \geq 0$, the matrix $P(t)$ is stochastic. In fact, if Q has zero row sums then so does Q^n for every n:
$$\sum_{k\in I} q_{ik}^{(n)} = \sum_{k\in I}\sum_{j\in I} q_{ij}^{(n-1)} q_{jk} = \sum_{j\in I} q_{ij}^{(n-1)} \sum_{k\in I} q_{jk} = 0.$$
So
$$\sum_{j\in I} p_{ij}(t) = 1 + \sum_{n=1}^{\infty} \frac{t^k}{n!} \sum_{j\in I} q_{ij}^{(n)} = 1.$$
Finally, since $P(t) = I + tQ + O(t^2)$ we get that $p_{ij}(t) \geq 0$ for all i, j and for t sufficiently small, and since $P(t) = P(t/n)^n$ it follows that $p_{ij}(t) \geq 0$ for all i, j and for all t.

(iii) *Forward equation:*
$$P'(t) = P(t)Q = QP(t),\ P(0) = I.$$
Furthermore, $Q = \frac{dP(t)}{dt}|_{t=0}$.

The family $\{P(t), t \geq 0\}$ is called the semigroup generated by Q.

Proposition 29 *Let $\{X_t, t \geq 0\}$ be a continuous-time Markov chain with Q-matrix Q. Then for all times $0 \leq t_0 \leq t_1 \leq \cdots \leq t_{n+1}$ and all states $i_0, i_1, \ldots, i_{n+1}$ we have*
$$P(X_{t_{n+1}} = i_{n+1} | X_{t_0} = i_0, \ldots, X_{t_n} = i_n) = p_{i_n i_{n+1}}(t_{n+1} - t_n).$$

A probability distribution (or, more generally, a measure) π on the state space I is said to be *invariant* for a continuous-time Markov chain $\{X_t, t \geq 0\}$ if $\pi P(t) = \pi$ for all $t \geq 0$. If we choose an invariant distribution π as initial distribution of the Markov chain $\{X_t, t \geq 0\}$, then the distribution of X_t is π for all $t \geq 0$.

If $\{X_t, t \geq 0\}$ is a continuous-time Markov chain irreducible and recurrent (that is, the associated jump matrix Π is recurrent) with Q-matrix Q, then, a measure π is invariant if and only if

$$\pi Q = 0, \tag{12}$$

and there is a unique (up to multiplication by constants) solution π which is strictly positive.

On the other hand, if we set $\alpha_i = \pi_i q_i$, then Equation (12) is equivalent to say that α is invariant for the jump matrix Π. In fact, we have $\alpha_i(\pi_{ij} - \delta_{ij}) = \pi_i q_{ij}$ for all i, j, so

$$(\alpha(\Pi - I))_j = \sum_{i \in I} \alpha_i(\pi_{ij} - \delta_{ij}) = \sum_{i \in I} \pi_i q_{ij} = 0.$$

Let $\{X_t, t \geq 0\}$ be an irreducible continuous-time Markov chain with Q-matrix Q. The following statements are equivalent:

(i) The jump chain Π is positive recurrent.

(ii) Q is non explosive and has an invariant distribution π.

Moreover, under these assumptions, we have

$$\lim_{t \to \infty} p_{ij}(t) = \pi_j = \frac{1}{q_j m_j},$$

where $m_i = E_i(T_i)$ is the expected return time to the state i.

Example 18 We calculate $p_{11}(t)$ for the continuous-time Markov chain with Q-matrix (11). Q has eigenvalues $0, -2, 4$. Then put

$$p_{11}(t) = a + be^{-2t} + ce^{-4t}$$

for some constants a, b, c. To determine the constants we use $p_{11}(0) = 1$, $p'_{11}(0) = q_{11} = -2$ and $p''_{11}(0) = q_{11}^{(2)} = 7$, so

$$p_{11}(t) = \frac{3}{8} + \frac{1}{4}e^{-2t} + \frac{3}{8}e^{-4t}.$$

Example 19 Let $\{N_t, t \geq 0\}$ be a Poisson process with rate $\lambda > 0$. Then, it is a continuous-time Markov chain with state-space $\{0, 1, 2, \ldots\}$ and Q-matrix

$$Q = \begin{pmatrix} -\lambda & \lambda & & & \\ & -\lambda & \lambda & & \\ & & -\lambda & \lambda & \\ & & & \ddots & \ddots \\ & & & & \ddots \end{pmatrix}$$

Example 20 (*Birth process*) A birth process i as generalization of the Poisson process in which the parameter λ is allowed to depend on the current state of the process. The data for a birth process consist of *birth rates* $0 \leq q_i < \infty$, where $i = 0, 1, 2, \ldots$. Then, a birth process $X = \{X_t, t \geq 0\}$ is a continuous time Markov chain with state-space $\{0, 1, 2, \ldots\}$, and Q-matrix

$$Q = \begin{pmatrix} -q_0 & q_0 & & & \\ & -q_1 & q_1 & & \\ & & -q_2 & q_2 & \\ & & & \ddots & \ddots \\ & & & & \ddots \end{pmatrix}.$$

That is, conditional on $X_0 = i$, the holding times $R_1, R_2, \ldots$ are independent exponential random variables of parameters $q_i, q_{i+1}, \ldots$, respectively, and the jump chain is given by $Y_n = i + n$.

Concerning the explosion time, two cases are possible:

(i) If $\sum_{j=0}^{\infty} \frac{1}{q_j} < \infty$, then $\zeta \overset{\text{a.s.}}{<} \infty$,

(ii) If $\sum_{j=0}^{\infty} \frac{1}{q_j} = \infty$, then $\zeta \overset{\text{a.s.}}{=} \infty$.

In fact, if $\sum_{j=0}^{\infty} \frac{1}{q_j} < \infty$, then, by monotone convergence

$$E_i(\sum_{n=1}^{\infty} R_n) = \sum_{n=1}^{\infty} \frac{1}{q_{i+n-1}} < \infty,$$

so $\sum_{n=1}^{\infty} R_n \overset{\text{a.s.}}{<} \infty$. If $\sum_{n=1}^{\infty} \frac{1}{q_{i+n-1}} = \infty$, then $\prod_{n=1}^{\infty} \left(1 + \frac{1}{q_{i+n-1}}\right) = \infty$ and by monotone convergence and independence

$$E\left[\exp\left(-\sum_{n=1}^{\infty} R_n\right)\right] = \prod_{n=1}^{\infty} E\left(e^{-R_n}\right) = \prod_{n=1}^{\infty} \left(1 + \frac{1}{q_{i+n-1}}\right)^{-1} = 0$$

so $\sum_{n=1}^{\infty} R_n \stackrel{\text{a.s.}}{=} \infty$.

Particular case (*Simple birth process*): Consider a population in which each individual gives birth after an exponential time of parameter λ, all independently. If i individuals are present then the first birth will occur after an exponential time of parameter $i\lambda$. Then we have $i+1$ individuals and, by the memoryless property, the process begins afresh. Then the size of the population performs a birth process with rates $q_i = \lambda i$, $i = 0, 2, \ldots$. Suppose $X_0 = 1$. Note that $\sum_{i=1}^{\infty} \frac{1}{i\lambda} = \infty$, so $\zeta \stackrel{\text{a.s.}}{=} \infty$ and there is no explosion in finite time. However, the mean population size growths exponentially:

$$E(X_t) = e^{\lambda t}.$$

Indeed, let T be the time of the first birth. Then if $\mu(t) = E(X_t)$

$$\begin{aligned} \mu(t) &= E(X_t \mathbf{1}_{\{T \le t\}}) + E(X_t \mathbf{1}_{\{T > t\}}) \\ &= \int_0^t \lambda e^{\lambda s} 2\mu(t-s) ds + e^{-\lambda t}. \end{aligned}$$

Setting $r = t - s$ we obtain $e^{\lambda t}\mu(t) \ = 2\lambda \int_0^t e^{\lambda r}\mu(r)dr$ and differentiating we get $\mu'(t) = \lambda\mu(t)$.

3.8 Birth and Death Processes

This is an important class of continuous time Markov chains with applications in biology, demography, and queueing theory. The random variable X_t will represent the size of a population at time t, and the state space is $I = \{0, 1, 2, \ldots\}$. A "birth" increases the size by one and a "death" decreases it by one. The parameters of this continuous Markov chain are:

1) p_i = probability that a birth occurs before a death if the size of the population is i. Note that $p_0 = 1$.

2) α_i = parameter of the exponential time until next birth or death.

Define $\lambda_i = p_i \alpha_i$ and $\mu_i = (1 - p_i)\alpha_i$, so these parameters can be thought as the respective time rates of births and deaths at an instant at which the population size is i.The parameters λ_i and μ_i are arbitrary non-negative numbers, except $\mu_0 = 0$ and $\lambda_i + \mu_i > 0$ for all $i \geq 1$. We have $p_i = \frac{\lambda_i}{\lambda_i + \mu_i}$.

As a consequence, the rates are $\lambda_0, \lambda_1+\mu_1, \lambda_2+\mu_2, \ldots$, and the jump matrix is, in the case $\lambda_0 > 0$

$$\Pi = \begin{pmatrix} 0 & 1 & & & \\ 1-p_i & 0 & p_i & & \\ & 1-p_i & 0 & p_i & \\ & & & \ddots & \end{pmatrix},$$

and the Q-matrix is

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 & & & \\ \mu_1 & -\lambda_1-\mu_1 & \lambda_1 & & \\ & \mu_2 & -\lambda_2-\mu_2 & \lambda_2 & \\ & & & \ddots & \end{pmatrix}.$$

The matrix Π is irreducible. Notice that $\frac{\sum_{n=0}^{\infty}\pi_{ii}^{(n)}}{\lambda_i+\mu_i}$ is the expected time spent in state i. A necessary and sufficient condition for non explosion is then:

$$\sum_{i=0}^{\infty} \frac{\sum_{n=0}^{\infty}\pi_{ii}^{(n)}}{\lambda_i+\mu_i} = \infty. \tag{13}$$

On the other hand, equation $\pi Q = 0$ satisfied by invariant measures leads to the system

$$\begin{aligned} \mu_1\pi_1 &= \lambda_0\pi_0, \\ \lambda_0\pi_0 + \mu_2\pi_2 &= (\lambda_1+\mu_1)\pi_1, \\ \lambda_{i-1}\pi_{i-1} + \mu_{i+1}\pi_{i+1} &= (\lambda_i+\mu_i)\pi_i, \end{aligned}$$

for $i = 2, 3, \ldots$. Solving these equations consecutively, we get

$$\pi_i = \frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\cdots\mu_i}\pi_0,$$

for $i = 1, 2, \ldots$ Hence, an invariant distribution exists if and only if

$$c = \sum_{i=1}^{\infty} \frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\cdots\mu_i} < \infty.$$

In that case the invariant distribution is

$$\pi_i = \begin{cases} \frac{1}{1+c} & \text{if} \quad i = 0 \\ \frac{1}{1+c} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \cdots \mu_i} & \text{if} \quad i \geq 1 \end{cases} . \tag{14}$$

Example 21 Suppose $\lambda_i = \lambda i$, $\mu_i = \mu i$ for some $\lambda, \mu > 0$. This corresponds to the case where each individual of the population has an exponential lifetime with parameter μ, and each individual can give birth to another after an exponential time with parameter λ. Notice that $\lambda_0 = 0$. So the jump chain is that of the Gambler's ruin problem (Example 6) with $p = \frac{\lambda}{\lambda+\mu}$, that is,

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

If $\lambda \leq \mu$ we know that with probability one the population with extinguish. If $\lambda > \mu$, condition (13) becomes

$$\sum_{i=1}^{\infty} \frac{\sum_{n=0}^{\infty} \pi_{ii}^{(n)}}{i(\lambda+\mu)} \geq \sum_{i=1}^{\infty} \frac{1}{i(\lambda+\mu)} = \infty,$$

and the chain is non explosive.

A variation of this example is the birth and death process with linear growth and immigration: $\lambda_i = \lambda i + a$, $\mu_i = \mu i$, for all $i \geq 0$, where $\lambda, \mu, a > 0$. Again there is no explosion because $\sum_{i=1}^{\infty} \frac{1}{i(\lambda+\mu)+a} = \infty$. Let us compute $\mu_i(t) := E_i(X_t)$ the average size of the population if the initial size is i.

The equality $P'(t) = P(t)Q$ leads to the following system of differential equations:

$$p'_{ij}(t) = \sum_{k=0}^{\infty} q_{ik} p_{kj}(t),$$

that is,

$$p'_{ij}(t) = p_{ij-1}(\lambda(j-1)+a) - p_{ij}(t)[(\lambda+\mu)j + a] + p_{ij+1}\mu(j+1),$$

for all $j \geq 1$, and $p'_{i0}(t) = -p_{i0}(t)a + p_{i1}\mu$.

Taking into account that $\mu_i(t) = \sum_{j=1}^{\infty} j p_{ij}(t)$, and using the above equations, we obtain

$$\begin{aligned}
\mu_i'(t) &= \sum_{j=1}^{\infty} j p_{ij}'(t) \\
&= \sum_{j=1}^{\infty} p_{ij}(t) \left((\lambda j + a)(j+1) - j[(\lambda + \mu) j + a] + \mu (j-1) j \right) \\
&= \sum_{j=1}^{\infty} p_{ij}(t) \left(a + j(\lambda - \mu) \right) \\
&= a + j(\lambda - \mu) \mu_i(t).
\end{aligned}$$

The solution to this differential equation with initial condition $\mu_i(0) = i$ yields

$$\begin{cases} \mu_i(t) = at + i & \text{if } \lambda = \mu \\ \mu_i(t) = \frac{a}{\lambda - \mu} \left(e^{(\lambda-\mu)t} - 1 \right) + i e^{(\lambda - \mu)t} & \text{if } \lambda \neq \mu \end{cases}.$$

Notice that in the case $\lambda \geq \mu$, $\mu_i(t)$ converges to infinity as t goes to infinity, and if $\lambda < \mu$, $\mu_i(t)$ converges to $\frac{a}{\mu - \lambda}$.

3.9 Queues

The basic mathematical model for queues is as follows. There is a line of customers waiting service. On arrival each customer must wait until a serves is free, giving priority to earlier arrivals. It is assumed that the times between arrivals are independent random variables with some distribution, and the times taken to serve customers are also independent random variables of another distribution. The quantity of interest is the continuous time stochastic process $\{X_t, t \geq 0\}$ recording the number of customers in the queue at time t. This is always taken to include both those being served and those waiting to be served.

3.9.1 M/M/1/∞ queue

Here M means memoryless, 1 stands for one server and infinite queues are permissible. Customers arrive according to a Poisson process with rate λ, and service times are exponential of parameter μ. Under these assumptions, $\{X_t, t \geq 0\}$ is a continuous time Markov chain with with state space

$\{0, 1, 2, \ldots\}$. Actually it is a special birth and death process with Q-matrix

$$Q = \begin{pmatrix} -\lambda & \lambda & & & \\ \mu & -\lambda-\mu & \lambda & & \\ & \mu & -\lambda-\mu & \lambda & \\ & & & \cdot & \cdot \\ & & & & \cdot \end{pmatrix}.$$

In fact, suppose at time 0 there are $i > 0$ customers in the queue. Denote by T the time taken to serve the first customer and by S the time of the next arrival. Then the first jump time R_1 is $R_1 = \inf(S, T)$, which is an exponential random variable of parameter $\lambda + \mu$. On the other hand, the events $\{T < S\}$ and $\{T > S\}$ are independent of $\inf(S, T)$ and have probabilities $\frac{\mu}{\lambda+\mu}$ and $\frac{\lambda}{\lambda+\mu}$, respectively. In fact,

$$\begin{aligned} P(T &< S, \inf(S,T) > x) = P(T < S, T > x) \\ &= \int_x^{\infty} e^{-\lambda t} \mu e^{-\mu t} dt = \frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)x} \\ &= P(T < S) P(\inf(S,T) > x). \end{aligned}$$

Thus, $X_{R_1} = i - 1$ with probability $\frac{\mu}{\lambda+\mu}$ and $X_{R_1} = i + 1$ with probability $\frac{\lambda}{\lambda+\mu}$.

On the other hand, if we condition on $\{T < S\}$, then $S - T$ is exponential of parameter λ, and similarly conditional on $\{T > S\}$, then $T - S$ is exponential of parameter μ. In fact,

$$\begin{aligned} P(S - T &> x | T < S) = \frac{\lambda+\mu}{\mu} P(S - T > x) \\ &= \frac{\lambda+\mu}{\mu} \int_0^{\infty} e^{-\lambda(t+x)} \mu e^{-\mu t} dt = e^{-\lambda x}. \end{aligned}$$

This means that conditional on $X_{R_1} = j$, the process$\{X_t, t \geq 0\}$ begins afresh from j at time R_1.

The holding times are independent and exponential with parameters λ, $\lambda + \mu$, $\lambda + \mu$, $\ldots$. Hence, the chain is non explosive. The jump matrix is

$$\Pi = \begin{pmatrix} 0 & 1 & & & \\ \frac{\mu}{\lambda+\mu} & 0 & \frac{\lambda}{\lambda+\mu} & & \\ & \frac{\mu}{\lambda+\mu} & 0 & \frac{\lambda}{\lambda+\mu} & \\ & & & \cdot & \cdot \\ & & & & \cdot \end{pmatrix}$$

is that of the random walk of Example 13 with $p = \frac{\lambda}{\lambda+\mu}$. If $\lambda > \mu$, the chain is transient and we have $\lim_{t\to\infty} X_t = +\infty$, so the queue growths without limit in the long term. If $\lambda = \mu$ the chain is null recurrent. If $\lambda < \mu$, the chain is positive recurrent and the unique invariant distribution is, using (14) or Example 13 :

$$\pi_j = (1-r)r^j,$$

where $r = \frac{\lambda}{\mu}$. The ratio r is called the *traffic intensity* for the system. The average numbers of customers in the queue in equilibrium is given by

$$E_\pi(X_t) = \sum_{i=1}^{\infty} P_\pi(X_t \geq i) = \sum_{i=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^i = \frac{\lambda}{\mu - \lambda}.$$

Also, the mean time to return to zero is given by

$$m_0 = \frac{1}{q_0 \pi_0} = \frac{\mu}{\lambda(\mu - \lambda)}$$

so the mean length of time that the server is continuously busy is given by

$$m_0 - \frac{1}{q_0} = \frac{1}{\mu - \lambda}.$$

3.9.2 M/M/1/m queue

This is a queueing system like the preceding one with a waiting room of finite capacity $m-1$. Hence, starting with less than m customers, the total number in the system cannot exceed m; that is, a customer arriving to find m or more customers in the system leaves and never returns (we do allow initial sizes of greater than m, however).

This is a birth and death process with $\lambda_i = \lambda$, for $i = 0, 1, \ldots, m-1$, $\lambda_i = 0$ for $i \geq m$ and $\mu_i = \mu$ for all $i \geq 1$. The states $m+1, m+2, \ldots$ are all transient and $C = \{0, 1, 2, \ldots, m\}$ is a recurrent class.

3.9.3 M/M/s/∞ queue

Arrivals from a Poisson process with rate λ, service times are exponential with parameter μ, there are s servers, and the waiting room is of infinite size. If i servers are occupied, the first service is completed at the minimum of i independent exponential times of parameter μ. The first service time is therefore exponential of parameter $i\mu$. The total service rate increases to a maximum of $s\mu$ when all servers are working. We emphasize that the queue size includes those customers who are currently being served. Then the queue size is a birth and death process with $\lambda_i = \lambda$ for all $i \geq 0$, $\mu_i = i\mu$, for $i = 1, 2, \ldots, s$ and $\mu_i = s\mu$ for $i \geq s+1$. The chain is positive recurrent when $\lambda < s\mu$. The invariant distribution is

$$\pi_i = \begin{cases} C\left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!} & \text{for} \quad i = 0, \ldots s \\ C\left(\frac{\lambda}{\mu}\right)^i \frac{1}{s^{s-i} i!} & \text{for} \quad i \geq s+1 \end{cases},$$

with some normalizing constant C.

In the particular case $s = \infty$ there are infinitely many servers so that no customer ever waits. In this case the normalizing constant is $C = e^{-\lambda/\mu}$ and π is a Poisson distribution of parameter λ/μ. This model applies to very large parking lots; then the customers are the vehicles, servers are the stalls, and the service times are the times vehicles remain parked.

This model also fits the short-run behavior of the size (measure in terms of the number of families living in them) of larger cities. The arrivals are the families moving into the city, and a service time is the amount of time spent there by a family before it moves out. In USA $1/\mu$ is about 4 years. If the city has a stable population size of about 1.000.000 families, then the arrival rate must be $\lambda = 250.000$ families per year. Approximating the Poisson distribution by a normal we obtain

$$P(997.000 \leq X_t \leq 1.003.000) \simeq 0.9974,$$

that is, the number of families in the city is about one million plus or minus three thousand.

3.9.4 Telephone exchange

A variation on the M/M/s queue is to turn away customers who cannot be served immediately. This queue might serve a simple model for a telephone exchange, when the maximum number of calls that can be connected at once is s: when the exchange is full, the additional calls are lost.

The invariant distribution of this irreducible chain is

$$\pi_i = \frac{(\lambda/\mu)^i \frac{1}{i!}}{\sum_{j=0}^{s} (\lambda/\mu)^j \frac{1}{j!}},$$

which is called the truncated Poisson distribution.

By the ergodic theorem, the long-run proportion of time that the exchange is full, and hence the long-run proportion of calls that are lost, is given by (*Erlang's formula*):

$$\pi_s = \frac{(\lambda/\mu)^s \frac{1}{s!}}{\sum_{j=0}^{s} (\lambda/\mu)^j \frac{1}{j!}}.$$

Exercises

3.1 Let $\{X_n, n \geq 0\}$ be a Markov chain with state space $\{1, 2, 3\}$ and transition matrix

$$P = \begin{pmatrix} 0 & 1/3 & 2/3 \\ 1/4 & 3/4 & 0 \\ 2/5 & 0 & 3/5 \end{pmatrix}$$

and initial distribution $\pi = (\frac{2}{5}, \frac{1}{5} \frac{2}{5})$. Compute the following probabilities:

(a) $P(X_1 = 2, X_2 = 2, X_3 = 2, X_4 = 1, X_5 = 3)$

(b) $P(X_5 = 2, X_6 = 2 | X_2 = 2)$

3.2 Classify the states of the Markov chains with the following transition matrices:

$$P_1 = \begin{pmatrix} 0.8 & 0 & 0.2 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0.3 & 0.4 & 0 & 0.3 \end{pmatrix}$$

$$P_2 = \begin{pmatrix} 0.8 & 0 & 0 & 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0.1 & 0 & 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0.3 & 0 & 0 & 0.7 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0.5 & 0 \end{pmatrix}$$

$$P_3 = \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}.$$

3.3 Find all invariant distributions of the transition matrix P_3 in exercise 2.2

3.4 In Example 4, instead of considering the remaining lifetime at time n, let us consider the age Y_n of the equipment in use ar time n. For any n, $Y_{n+1} = 0$ if the equipment failed during the period $n+1$, and $Y_{n+1} = Y_n + 1$ if it did not fail during the period $n+1$.

(a) Show that

$$P(Y_{n+1} = i+1 | Y_n = i) = \begin{cases} 1 - p_1 & \text{if} \quad i = 0 \\ \frac{1-p_1-\cdots-p_{i+1}}{1-p_1-\cdots-p_i} & \text{if} \quad i \geq 1 \end{cases}$$

(b) Show that $\{Y_n, n \geq 0\}$ is a Markov chain with state space $\{0, 1, 2, \ldots\}$ and find its transition matrix.

(c) Classify the states of this Markov chain.

3.5 Consider a system whose successive states form a Markov chain with state space $\{1, 2, 3\}$ and transition matrix

$$\Pi = \begin{pmatrix} 0 & 1 & 0 \\ 0.4 & 0 & 0.6 \\ 0 & 1 & 0 \end{pmatrix}.$$

Suppose the holding times in states $1, 2, 3$ are all independent and have exponential distributions with parameters $2, 5, 3$, respectively. Compute the Q-matrix of this Markov system.

3.6 Suppose the arrivals at a counter form a Poisson process with rate λ, and suppose each arrival is of type a or of type b with respective probabilities p and $1 - p = q$. Let X_t be the type of the last arrival before t. Show that $\{X_t, t \geq 0\}$ is a continuous Markov chain with state space $\{a, b\}$. Compute the distribution of the holding times, the Q-matrix and the transition semigroup $P(t)$.

3.7 Consider a queueing system M/M/1/2 with capacity 2. Compute the Q-matrix and the *potential matrix* $U^\alpha = (\alpha I - Q)^{-1}$, where $\alpha > 0$, for the case $\lambda = 2$, $\mu = 8$. Suppose a penalty cost of one dollar per unit time is applicable for each customer present in the system. Suppose the discount rate is α. Show that the expected total discounted cost if, starting at i, $(U^\alpha g)_i$, where $g = (0, 1, 2)$.

4 Martingales

We will first introduce the notion of conditional expectation of a random variable X with respect to a σ-field $\mathcal{B} \subset \mathcal{F}$ in a probability space $(\Omega, \mathcal{F}, P)$.

4.1 Conditional Expectation

Consider an integrable random variable X defined in a probability space $(\Omega, \mathcal{F}, P)$, and a σ-field $\mathcal{B} \subset \mathcal{F}$. We define the *conditional expectation* of X given $\mathcal{B}$ (denoted by $E(X|\mathcal{B})$) to be any integrable random variable Z that verifies the following two properties:

(i) Z is measurable with respect to $\mathcal{B}$.

(ii) For all $A \in \mathcal{B}$
$$E\left(Z\mathbf{1}_A\right) = E\left(X\mathbf{1}_A\right).$$

It can be proved that there exists a unique (up to modifications on sets of probability zero) random variable satisfying these properties. That is, if $\widetilde{Z}$ and Z verify the above properties, then $Z = \widetilde{Z}$, P-almost surely.

Property (ii) implies that for any bounded and $\mathcal{B}$-measurable random variable Y we have
$$E\left(E(X|\mathcal{B})Y\right) = E\left(XY\right). \tag{15}$$

Example 1 Consider the particular case where the σ-field $\mathcal{B}$ is generated by a finite partition $\{B_1, ..., B_m\}$. In this case, the conditional expectation $E(X|\mathcal{B})$ is a discrete random variable that takes the constant value $E(X|\mathcal{B}_j)$ on each set B_j :
$$E(X|\mathcal{B}) = \sum_{j=1}^{m} \frac{E\left(X\mathbf{1}_{B_j}\right)}{P(B_j)} \mathbf{1}_{B_j}.$$

Here are some rules for the computation of conditional expectations in the general case:

Rule 1 The conditional expectation is lineal:

$$\boxed{E(aX + bY|\mathcal{B}) = aE(X|\mathcal{B}) + bE(Y|\mathcal{B})}$$

Rule 2 A random variable and its conditional expectation have the same expectation:

$$\boxed{E\left(E(X|\mathcal{B})\right) = E(X)}$$

This follows from property (ii) taking $A = \Omega$.

Rule 3 If X and $\mathcal{B}$ are independent, then $E(X|\mathcal{B}) = E(X)$.

In fact, the constant $E(X)$ is clearly $\mathcal{B}$-measurable, and for all $A \in \mathcal{B}$ we have

$$E\left(X\mathbf{1}_A\right) = E\left(X\right)E(\mathbf{1}_A) = E(E(X)\mathbf{1}_A).$$

Rule 4 If X is $\mathcal{B}$-measurable, then $E(X|\mathcal{B}) = X$.

Rule 5 If Y is a bounded and $\mathcal{B}$-measurable random variable, then

$$E(YX|\mathcal{B}) = YE(X|\mathcal{B}).$$

In fact, the random variable $YE(X|\mathcal{B})$ is integrable and $\mathcal{B}$-measurable, and for all $A \in \mathcal{B}$ we have

$$E\left(E(X|\mathcal{B})Y\mathbf{1}_A\right) = E\left(E(XY\mathbf{1}_A|\mathcal{B})\right) = E(XY\mathbf{1}_A),$$

where the first equality follows from (15) and the second equality follows from Rule 3. This Rule means that $\mathcal{B}$-measurable random variables behave as constants and can be factorized out of the conditional expectation with respect to $\mathcal{B}$.

Rule 6 Given two σ-fields $\mathcal{C} \subset \mathcal{B}$, then

$$\boxed{E\left(E(X|\mathcal{B})|\mathcal{C}\right) = E\left(E(X|\mathcal{C})|\mathcal{B}\right) = E(X|\mathcal{C})}$$

Rule 7 Consider two random variable X y Z, such that Z is $\mathcal{B}$-measurable and X is independent of $\mathcal{B}$. Consider a measurable function $h(x, z)$ such that the composition $h(X, Z)$ is an integrable random variable. Then, we have

$$\boxed{E\left(h(X, Z)|\mathcal{B}\right) = E\left(h(X, z)\right)|_{z=Z}}$$

That is, we first compute the conditional expectation $E\left(h(X, z)\right)$ for any fixed value z of the random variable Z and, afterwards, we replace z by Z.

Conditional expectation has properties similar to those of ordinary expectation. For instance, the following monotone property holds:

$$X \leq Y \Rightarrow E(X|\mathcal{B}) \quad \leq E(Y|\mathcal{B}).$$

This implies $|E(X|\mathcal{B})\ | \leq E(|X|\ |\mathcal{B})$.

Jensen inequality also holds. That is, if φ is a convex function such that $E(|\varphi(X)|) < \infty$, then

$$\varphi\left(E(X|\mathcal{B})\right) \leq \ E(\varphi(X)|\mathcal{B}). \tag{16}$$

In particular, if we take $\varphi(x) = |x|^p$ with $p \geq 1$, we obtain

$$|E(X|\mathcal{B})|^p \leq \ E(|X|^p|\mathcal{B}),$$

hence, taking expectations, we deduce that if $E(|X|^p) < \infty$, then $E(|E(X|\mathcal{B})|^p) < \infty$ and

$$E(|E(X|\mathcal{B})|^p) \leq E(|X|^p). \tag{17}$$

We can define the conditional probability of an even $C \in \mathcal{F}$ given a σ-field $\mathcal{B}$ as

$$P(C|\mathcal{B}) = E(\mathbf{1}_C|\mathcal{B}).$$

Suppose that the σ-field $\mathcal{B}$ is generated by a finite collection of random variables $Y_1, \ldots, Y_m$. In this case, we will denote the conditional expectation of X given $\mathcal{B}$ by $E(X|Y_1, \ldots, Y_m)$. In this case this conditional expectation is the mean of the conditional distribution of X given $Y_1, \ldots, Y_m$.

The conditional distribution of X given $Y_1, \ldots, Y_m$ is a family of distributions $p(dx|y_1, \ldots, y_m)$ parameterized by the possible values $y_1, \ldots, y_m$ of the random variables $Y_1, \ldots, Y_m$, such that for all $a < b$

$$P(a \leq X \leq b|Y_1, \ldots, Y_m) = \int_a^b p(dx|Y_1, \ldots, Y_m).$$

Then, this implies that

$$E(X|Y_1, \ldots, Y_m) = \int_{\mathbb{R}} xp(dx|Y_1, \ldots, Y_m).$$

Notice that the conditional expectation $E(X|Y_1, \ldots, Y_m)$ is a function $g(Y_1, \ldots, Y_m)$ of the variables $Y_1, \ldots, Y_m$,where

$$g(y_1, \ldots, y_m) = \int_{\mathbb{R}} xp(dx|y_1, \ldots, y_m).$$

In particular, if the random variables $X, Y_1, \ldots, Y_m$ have a joint density $f(x, y_1, \ldots, y_m)$, then the conditional distribution has the density:

$$f(x|y_1, \ldots, y_m) = \frac{f(x, y_1, \ldots, y_m)}{\int_{-\infty}^{+\infty} f(x, y_1, \ldots, y_m) dy_1 \cdots dy_m},$$

and

$$E(X|Y_1, \ldots, Y_m) = \int_{-\infty}^{+\infty} x f(x|Y_1, \ldots, Y_m) dx.$$

The set of all square integrable random variables, denoted by $L^2(\Omega, \mathcal{F}, P)$, is a Hilbert space with the scalar product

$$\langle Z, Y \rangle = E(ZY).$$

Then, the set of square integrable and $\mathcal{B}$-measurable random variables, denoted by $L^2(\Omega, \mathcal{B}, P)$ is a closed subspace of $L^2(\Omega, \mathcal{F}, P)$.

Then, given a random variable X such that $E(X^2) < \infty$, the conditional expectation $E(X|\mathcal{B})$ is the projection of X on the subspace $L^2(\Omega, \mathcal{B}, P)$. In fact, we have:

(i) $E(X|\mathcal{B})$ belongs to $L^2(\Omega, \mathcal{B}, P)$ because it is a $\mathcal{B}$-measurable random variable and it is square integrable due to (17).

(ii) $X - E(X|\mathcal{B})$ is orthogonal to the subspace $L^2(\Omega, \mathcal{B}, P)$. In fact, for all $Z \in L^2(\Omega, \mathcal{B}, P)$ we have, using the Rule 5,

$$\begin{aligned} E\left[(X - E(X|\mathcal{B}))Z\right] &= E(XZ) - E(E(X|\mathcal{B})Z) \\ &= E(XZ) - E(E(XZ|\mathcal{B})) = 0. \end{aligned}$$

As a consequence, $E(X|\mathcal{B})$ is the random variable in $L^2(\Omega, \mathcal{B}, P)$ that minimizes the mean square error:

$$E\left[(X - E(X|\mathcal{B}))^2\right] = \min_{Y \in L^2(\Omega, \mathcal{B}, P)} E\left[(X - Y)^2\right]. \tag{18}$$

This follows from the relation

$$E\left[(X - Y)^2\right] = E\left[(X - E(X|\mathcal{B}))^2\right] + E\left[(E(X|\mathcal{B}) - Y)^2\right]$$

and it means that the conditional expectation is the optimal estimator of X given the σ-field $\mathcal{B}$.

4.2 Discrete Time Martingales

In this section we consider a probability space $(\Omega, \mathcal{F}, P)$ and a nondecreasing sequence of σ-fields

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_n \subset \cdots$$

contained in $\mathcal{F}$. A sequence of real random variables $M = \{M_n, n \geq 0\}$ is called a *martingale* with respect to the σ-fields $\{\mathcal{F}_n, n \geq 0\}$ if:

(i) For each $n \geq 0$, M_n is $\mathcal{F}_n$-measurable (that is, M is *adapted* to the filtration $\{\mathcal{F}_n, n \geq 0\}$).

(ii) For each $n \geq 0$, $E(|M_n|) < \infty$.

(iii) For each $n \geq 0$,

$$\boxed{E(M_{n+1}|\mathcal{F}_n) = M_n.}$$

The sequence $M = \{M_n, n \geq 0\}$ is called a *supermartingale (*or *submartingale)* if property (iii) is replaced by $E(M_{n+1}|\mathcal{F}_n) \leq M_n$ (or $E(M_{n+1}|\mathcal{F}_n) \geq M_n$).

Notice that the martingale property implies that $E(M_n) = E(M_0)$ for all n. On the other hand, condition (iii) can also be written as

$$E(\Delta M_n|\mathcal{F}_{n-1}) = 0,$$

for all $n \geq 1$, where $\Delta M_n = M_n - M_{n-1}$.

Example 2 Suppose that $\{\xi_n, n \geq 1\}$ are independent centered random variables. Set $M_0 = 0$ and $M_n = \xi_1 + ... + \xi_n$, for $n \geq 1$. Then M_n is a martingale with respect to the sequence of σ-fields $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$ for $n \geq 1$, and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. In fact,

$$\begin{aligned} E(M_{n+1}|\mathcal{F}_n) &= E(M_n + \xi_n|\mathcal{F}_n) \\ &= M_n + E(\xi_n|\mathcal{F}_n) \\ &= M_n + E(\xi_n) = M_n. \end{aligned}$$

Example 3 Suppose that $\{\xi_n, n \geq 1\}$ are independent random variable such that $P(\xi_n = 1) = p$, i $P(\xi_n = -1) = 1 - p$, on $0 < p < 1$. Then

$M_n = \left(\frac{1-p}{p}\right)^{\xi_1+...+\xi_n}$, $M_0 = 1$, is a martingale with respect to the sequence of σ-fields $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$ for $n \geq 1$, and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. In fact,

$$\begin{aligned}
E(M_{n+1}|\mathcal{F}_n) &= E\left(\left(\frac{1-p}{p}\right)^{\xi_1+...+\xi_{n+1}} |\mathcal{F}_n\right) \\
&= \left(\frac{1-p}{p}\right)^{\xi_1+...+\xi_n} E\left(\left(\frac{1-p}{p}\right)^{\xi_{n+1}} |\mathcal{F}_n\right) \\
&= M_n E\left(\frac{1-p}{p}\right)^{\xi_{n+1}} \\
&= M_n.
\end{aligned}$$

In the two previous examples, $\mathcal{F}_n = \sigma(M_0, \ldots, M_n)$, for all $n \geq 0$. That is, $\{\mathcal{F}_n\}$ is the filtration generated by the process $\{M_n\}$. Usually, when the filtration is not mentioned, we will take $\mathcal{F}_n = \sigma(M_0, \ldots, M_n)$, for all $n \geq 0$. This is always possible due to the following result:

Lemma 30 *Suppose $\{M_n, n \geq 0\}$ is a martingale with respect to a filtration $\{\mathcal{G}_n\}$. Let $\mathcal{F}_n = \sigma(M_0, \ldots, M_n) \subset \mathcal{G}_n$. Then $\{M_n, n \geq 0\}$ is a martingale with respect to a filtration $\{\mathcal{F}_n\}$.*

Proof. We have, by the Rule 6 of conditional expectations,

$$E(M_{n+1}|\mathcal{F}_n) = E(E(M_{n+1}|\mathcal{G}_n)|\mathcal{F}_n) = E(M|\mathcal{F}_n) = M_n.$$

■

Some elementary properties of martingales:

1. If $\{M_n\}$ is a martingale, then for all $m \geq n$ we have

$$\boxed{E(M_m|\mathcal{F}_n) = M_n.}$$

   In fact,

$$\begin{aligned}
M_n &= E(M_{n+1}|\mathcal{F}_n) = E(E(M_{n+2}|\mathcal{F}_{n+1})\ |\mathcal{F}_n) \\
&= E(M_{n+2}\ |\mathcal{F}_n) = ... = E(M_m|\mathcal{F}_n).
\end{aligned}$$

2. $\{M_n\}$ is a submartingale if and only if $\{-M_n\}$ is a supermartingale.

3. If $\{M_n\}$ is a martingale and φ is a convex function such that $E(|\varphi(M_n)|) < \infty$ for all $n \geq 0$, then $\{\varphi(M_n)\}$ is a submartingale. In fact, by Jensen's inequality for the conditional expectation we have

$$E\left(\varphi(M_{n+1})|\mathcal{F}_n\right) \geq \varphi\left(E(M_{n+1}|\mathcal{F}_n)\right) = \varphi(M_n).$$

   In particular, if $\{M_n\}$ is a martingale such that $E(|M_n|^p) < \infty$ for all $n \geq 0$ and for some $p \geq 1$, then $\{|M_n|^p\}$ is a submartingale.

4. If $\{M_n\}$ is a submartingale and φ is a convex and increasing function such that $E(|\varphi(M_n)|) < \infty$ for all $n \geq 0$, then $\{\varphi(M_n)\}$ is a submartingale. In fact, by Jensen's inequality for the conditional expectation we have

$$E\left(\varphi(M_{n+1})|\mathcal{F}_n\right) \geq \varphi\left(E(M_{n+1}|\mathcal{F}_n)\right) \geq \varphi(M_n).$$

   In particular, if $\{M_n\}$ is a submartingale , then $\{M_n^+\}$ and $\{M_n \wedge a\}$ are submartingales.

Suppose that $\{\mathcal{F}_n, n \geq 0\}$ is a given filtration. We say that $\{H_n, n \geq 1\}$ is a *predictable* sequence of random variables if for each $n \geq 1$, H_n is $\mathcal{F}_{n-1}$-measurable. We define the *martingale transform* of a martingale $\{M_n, n \geq 0\}$ by a predictable sequence $\{H_n, n \geq 1\}$ as the sequence

$$\boxed{(H \cdot M)_n = M_0 + \sum_{j=1}^n H_j \Delta M_j.}$$

Proposition 31 *If $\{M_n, n \geq 0\}$ is a (sub)martingale and $\{H_n, n \geq 1\}$ is a bounded (nonnegative) predictable sequence, then the martingale transform $\{(H \cdot M)_n\}$ is a (sub)martingale.*

Proof. Clearly, for each $n \geq 0$ the random variable $(H \cdot M)_n$ is $\mathcal{F}_n$-measurable and integrable. On the other hand, if $n \geq 0$ we have

$$\begin{aligned} E((H \cdot M)_{n+1} - (H \cdot M)_n|\mathcal{F}_n) &= E(\Delta M_{n+1}|\mathcal{F}_n) \\ &= H_{n+1}E(\Delta M_{n+1}|\mathcal{F}_n) = 0. \end{aligned}$$

■

We may think of H_n as the amount of money a gambler will bet at time n. Suppose that $\Delta M_n = M_n - M_{n-1}$ is the amount a gambler can win or lose at every step of the game if the bet is 1 Euro, and M_0 is the initial capital of the gambler. Then, M_n will be the fortune of the gambler at time n, and

$(H \cdot M)_n$ will be the fortune of the gambler is he uses the gambling system $\{H_n\}$. The fact that $\{M_n\}$ is a martingale means that the game is fair. So, the previous proposition tells us that if a game is fair, it is also fair regardless the gambling system $\{H_n\}$.

Suppose that $M_n = M_0 + \xi_1 + ... + \xi_n$, where $\{\xi_n, n \geq 1\}$ are independent random variable such that $P(\xi_n = 1) = P(\xi_n = -1) = \frac{1}{2}$. A famous gambling system is defined by $H_1 = 1$ and for $n \geq 2$, $H_n = 2H_{n-1}$ if $\xi_{n-1} = -1$, and $H_n = 1$ if $\xi_{n-1} = 1$. In other words, we double our bet when we lose, so that if we lose k times and then win, out net winnings will be

$$-1 - 2 - 4 - \cdots - 2^{k-1} + 2^k = 1.$$

This system seems to provide us with a "sure win", but this is not true due to the above proposition.

Example 4 Suppose that $S_n^0, S_n^1, \ldots, S_n^d$ are adapted and positive random variables which represent the prices at time n of $d+1$ financial assets. We assume that $S_n^0 = (1+r)^n$, where $r > 0$ is the interest rate, so the asset number 0 is non risky. We denote by $S_n = (S_n^0, S_n^1, ..., S_n^d)$ the vector of prices at time n.

In this context, a portfolio is a family of predictable sequences $\{\phi_n^i, n \geq 1\}$, $i = 0, \ldots, d$, such that ϕ_n^i represents the number of assets of type i at time n. We set $\phi_n = (\phi_n^0, \phi_n^1, ..., \phi_n^d)$. The value of the portfolio at time $n \geq 1$ is then

$$\boxed{V_n = \phi_n^0 S_n^0 + \phi_n^1 S_n^1 + ... + \phi_n^d S_n^d = \phi_n \cdot S_n.}$$

We say that a portfolio is self-financing if for all $n \geq 1$

$$\boxed{V_n = V_0 + \textstyle\sum_{j=1}^n \phi_j \cdot \Delta S_j,}$$

where V_0 denotes the initial capital of the portfolio. This condition is equivalent to

$$\boxed{\phi_n \cdot S_n = \phi_{n+1} \cdot S_n}$$

for all $n \geq 0$. Define the discounted prices by

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

Then the discounted value of the portfolio is

$$\widetilde{V}_n = (1+r)^{-n} V_n = \phi_n \cdot \widetilde{S}_n,$$

and the self-financing condition can be written as

$$\phi_n \cdot \widetilde{S}_n = \phi_{n+1} \cdot \widetilde{S}_n,$$

for $n \geq 0$, that is, $\widetilde{V}_{n+1} - \widetilde{V}_n = \phi_{n+1} \cdot (\widetilde{S}_{n+1} - \widetilde{S}_n)$ and summing in n we obtain

$$\widetilde{V}_n = V_0 + \sum_{j=1}^{n} \phi_j \cdot \Delta \widetilde{S}_j.$$

In particular, if $d = 1$, then $\widetilde{V}_n = (\phi^1 \cdot \widetilde{S})_n$ is the martingale transform of the sequence $\{\widetilde{S}_n\}$ by the predictable sequence $\{\phi_n^1\}$. As a consequence, if $\{\widetilde{S}_n\}$ is a martingale and $\{\phi_n^1\}$ is a bounded sequence, $\left\{\widetilde{V}_n\right\}$ will also be a martingale.

We say that a probability Q equivalent to P (that is, $P(A) = 0 \Leftrightarrow Q(A) = 0$), is a *risk-free probability*, if in the probability space $(\Omega, \mathcal{F}, Q)$ the sequence of discounted prices $\widetilde{S}_n$ is a martingale with respect to $\mathcal{F}_n$. Then, the sequence of values of any self-financing portfolio will also be a martingale with respect to $\mathcal{F}_n$, provided the β_n are bounded.

In the particular case of the binomial model (also called Ross-Cox-Rubinstein model), we assume that the random variables

$$T_n = \frac{S_n}{S_{n-1}} = 1 + \frac{\Delta S_n}{S_{n-1}},$$

$n = 1, ..., N$ are independent and take two different values $1 + a$, $1 + b$, $a < r < b$, with probabilities p, and $1 - p$, respectively. In this example, the risk-free probability will be

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

In fact, for each $n \geq 1$,

$$E(T_n) = (1 + a)p + (1 + b)(1 - p) = 1 + r,$$

and, therefore,

$$\begin{aligned} E(\widetilde{S}_n | \mathcal{F}_{n-1}) &= (1 + r)^{-n} E(S_n | \mathcal{F}_{n-1}) \\ &= (1 + r)^{-n} E(T_n S_{n-1} | \mathcal{F}_{n-1}) \\ &= (1 + r)^{-n} S_{n-1} E(T_n | \mathcal{F}_{n-1}) \\ &= (1 + r)^{-1} \widetilde{S}_{n-1} E(T_n) \\ &= \widetilde{S}_{n-1}. \end{aligned}$$

Consider a random variable $H \geq 0$, $\mathcal{F}_N$-measurable, which represents the payoff of a derivative at the maturity time N on the asset. For instance, for an European call option with strike price$K$, $H = (S_N - K)^+$. The derivative is replicable if there exists a self-financing portfolio such that $V_N = H$. We will take the value of the portfolio V_n as the price of the derivative at time n. In order to compute this price we make use of the martingale property of the sequence $\widetilde{V}_n$ and we obtain the following general formula for derivative prices:

$$\boxed{V_n = (1+r)^{-(N-n)} E_Q(H|\mathcal{F}_n).}$$

In fact,

$$\widetilde{V}_n = E_Q(\widetilde{V}_N|\mathcal{F}_n) = (1+r)^{-(N-n)} E_Q(H|\mathcal{F}_n).$$

In particular, for $n = 0$, if the σ-field $\mathcal{F}_0$ is trivial,

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

Example 5 Suppose that T is a stopping time. Then, the process

$$H_n = \mathbf{1}_{\{T \geq n\}}$$

is predictable. In fact, $\{T \geq n\} = \{T \leq n-1\} \in \mathcal{F}_{n-1}$. The martingale transform of $\{M_n\}$ by this sequence is

$$\begin{aligned}(H \cdot M)_n &= M_0 + \sum_{j=1}^{n} \mathbf{1}_{\{T \geq j\}} (M_j - M_{j-1}) \\ &= M_0 + \sum_{j=1}^{T \wedge n} (M_j - M_{j-1}) = M_{T \wedge n}.\end{aligned}$$

As a consequence, if $\{M_n\}$ is a (sub)martingale , the stopped process $\{M_{T\wedge n}\}$ will be a (sub)martingale.

Theorem 32 (Optional Stopping Theorem) *Suppose that $\{M_n\}$ is a submartingale and $S \leq T \leq m$ are two stopping times bounded by a fixed time m. Then*

$$E(M_T|\mathcal{F}_S) \geq M_S,$$

with equality in the martingale case.

This theorem implies that $E(M_T) \leq E(M_S)$.

Proof. We make the proof only in the martingale case. Notice first that M_T is integrable because

$$|M_T| \leq \sum_{n=0}^{m} |M_n|.$$

Consider the predictable process $H_n = \mathbf{1}_{\{S<n\leq T\}\cap A}$, where $A \in \mathcal{F}_S$. Notice that $\{H_n\}$ is predictable because

$$\{S < n \leq T\} \cap A = \{T < n\}^c \cap [\{S \leq n-1\} \cap A] \in \mathcal{F}_{n-1}.$$

Moreover, the random variables H_n are nonnegative and bounded by one. Therefore, by Proposition 31, $(H \cdot M)_n$ is a martingale. We have

$$\begin{aligned} (H \cdot M)_0 &= M_0, \\ (H \cdot M)_m &= M_0 + \mathbf{1}_A (M_T - M_S). \end{aligned}$$

The martingale property of $(H \cdot M)_n$ implies that $E((H \cdot M)_0) = E((H \cdot M)_m)$. Hence,

$$E\left(\mathbf{1}_A (M_T - M_S)\right) = 0$$

for all $A \in \mathcal{F}_S$ and this implies that $E(M_T|\mathcal{F}_S) = M_S$, because M_S is $\mathcal{F}_S$-measurable. ■

Theorem 33 (Doob's Maximal Inequality) *Suppose that $\{M_n\}$ is a submartingale and $\lambda > 0$. Then*

$$P(\sup_{0\leq n\leq N} M_n \geq \lambda) \leq \frac{1}{\lambda} E(M_N \mathbf{1}_{\left\{\sup_{0\leq n\leq N} M_n \geq \lambda\right\}}).$$

Proof. Consider the stopping time

$$T = \inf\{n \geq 0 : M_n \geq \lambda\} \wedge N.$$

Then, by the Optional Stopping Theorem,

$$\begin{aligned} E(M_N) &\geq E(M_T) = E(M_T \mathbf{1}_{\left\{\sup_{0\leq n\leq N} M_n \geq \lambda\right\}}) \\ &\quad + E(M_T \mathbf{1}_{\left\{\sup_{0\leq n\leq N} M_n < \lambda\right\}}) \\ &\geq \lambda P(\sup_{0\leq n\leq N} M_n \geq \lambda) + E(M_N \mathbf{1}_{\left\{\sup_{0\leq n\leq N} M_n < \lambda\right\}}). \end{aligned}$$

■

As a consequence, if $\{M_n\}$ is a martingale and $p \geq 1$, applying Doob's maximal inequality to the submartingale $\{|M_n|^p\}$ we obtain

$$P(\sup_{0\leq n\leq N} |M_n| \geq \lambda) \leq \frac{1}{\lambda^p} E(|M_N|^p),$$

which is a generalization of Chebyshev inequality.

Theorem 34 (The Martingale Convergence Theorem) *If $\{M_n\}$ is a submartingale such that $\sup_n E(M_n^+) < \infty$, then*

$$\boxed{M_n \xrightarrow{\text{a.s.}} M}$$

where M is an integrable random variable.

As a consequence, any nonnegative martingale converges almost surely. However, the convergence may not be in the mean.

Example 6 Suppose that $\{\xi_n, n \geq 1\}$ are independent random variables with distribution $N(0, \sigma^2)$. Set $M_0 = 1$, and

$$M_n = \exp\left(\sum_{j=1}^{n} \xi_j - \frac{n}{2}\sigma^2\right).$$

Then, $\{M_n\}$ is a nonnegative martingale such that $M_n \xrightarrow{\text{a.s.}} 0$, by the strong law of large numbers, but $E(M_n) = 1$ for all n.

Example 7 (*Branching processes*) Suppose that $\{\xi_i^n, n \geq 1, i \geq 0\}$ are nonnegative independent identically distributed random variables. Define a sequence $\{Z_n\}$ by $Z_0 = 1$ and for $n \geq 1$

$$Z_{n+1} = \begin{cases} \xi_1^{n+1} + \cdots + \xi_{Z_n}^{n+1} & \text{if} \quad Z_n > 0 \\ 0 & \text{if} \quad Z_n = 0 \end{cases}$$

The process $\{Z_n\}$ is called a Galton-Watson process. The random variable Z_n is the number of people in the nth generation and each member of a generation gives birth independently to an identically distributed number of children. $p_k = P(\xi_i^n = k)$ is called the *offspring distribution*. Set $\mu = E(\xi_i^n)$. The process Z_n/μ^n is a martingale. In fact,

$$\begin{aligned}
E(Z_{n+1}|\mathcal{F}_n) &= \sum_{k=1}^{\infty} E(Z_{n+1}\mathbf{1}_{\{Z_n=k\}}|\mathcal{F}_n) \\
&= \sum_{k=1}^{\infty} E(\left(\xi_1^{n+1} + \cdots + \xi_k^{n+1}\right)\mathbf{1}_{\{Z_n=k\}}|\mathcal{F}_n) \\
&= \sum_{k=1}^{\infty} \mathbf{1}_{\{Z_n=k\}} E(\xi_1^{n+1} + \cdots + \xi_k^{n+1}|\mathcal{F}_n) \\
&= \sum_{k=1}^{\infty} \mathbf{1}_{\{Z_n=k\}} E(\xi_1^{n+1} + \cdots + \xi_k^{n+1}) \\
&= \sum_{k=1}^{\infty} \mathbf{1}_{\{Z_n=k\}} k\mu = \mu Z_n.
\end{aligned}$$

This implies that $E(Z_n) = \mu^n$. On the other hand, Z_n/μ^n is a nonnegative martingale, so it converges almost surely to a limit. This limit is zero if $\mu \leq 1$ and ξ_i^n is not identically one. Actually, in this case $Z_n = 0$ for all n sufficiently large. This is intuitive: If each individual on the average gives birth to less than one child, the species dies out. If $\mu > 1$ the limit of Z_n/μ^n has a change of being nonzero. In this case $\rho = P(Z_n = 0 \text{ for some } n) < 1$ is the unique solution of $\varphi(\rho) = \rho$, where $\varphi(s) = \sum_{k=0}^{\infty} p_k s^k$ is the generating function of the spring distribution.

The following result established the convergence of the martingale in mean of order p in the case $p > 1$.

Theorem 35 *If $\{M_n\}$ is a martingale such that $\sup_n E(|M_n|^p) < \infty$, for some $p > 1$, then*

$$\boxed{M_n \to M}$$

almost surely and in mean of order p. Moreover, $M_n = E(M|\mathcal{F}_n)$ for all n.

Example 8 Consider the symmetric random walk $\{S_n, n \geq 0\}$. That is, $S_0 = 0$ and for $n \geq 1$, $S_n = \xi_1 + ... + \xi_n$where $\{\xi_n, n \geq 1\}$ are independent random variables such that $P(\xi_n = 1) = P(\xi_n = -1) = \frac{1}{2}$. Then $\{S_n\}$ is a martingale. Set

$$T = \inf\{n \geq 0 : S_n \notin (a, b)\},$$

where $b < 0 < a$. We are going to show that $E(T) = |ab|$.

In fact, we know that $\{S_{T \wedge n}\}$ is a martingale. So, $E(S_{T \wedge n}) = E(S_0) = 0$. This martingale converges almost surely and in mean of order one because it is uniformly bounded. We know that $P(T < \infty) = 1$, because the random walk is recurrent. Hence,

$$S_{T \wedge n} \overset{\text{a.s.}, L^1}{\longrightarrow} S_T \in \{b, a\}.$$

Therefore, $E(S_T) = 0 = aP(S_T = a) + bP(S_T = b)$. From these equation we obtain the absorption probabilities:

$$\begin{aligned} P(S_T &= a) = \frac{-b}{a - b}, \\ P(S_T &= b) = \frac{a}{a - b}. \end{aligned}$$

Now $\{S_n^2 - n\}$ is also a martingale, and by the same argument, this implies $E(S_T^2) = E(T)$, which leads to $E(T) = -ab$.

4.3 Continuous Time Martingales

Consider an nondecreasing family of σ-fields $\{\mathcal{F}_t, t \geq 0\}$. A real continuous time process $M = \{M_t, t \geq 0\}$ is called a *martingale* with respect to the σ-fields $\{\mathcal{F}_t, t \geq 0\}$ if:

(i) For each $t \geq 0$, M_t is $\mathcal{F}_t$-measurable (that is, M is *adapted* to the filtration $\{\mathcal{F}_t, t \geq 0\}$).

(ii) For each $t \geq 0$, $E(|M_t|) < \infty$.

(iii) For each $s \leq t$, $E(M_t | \mathcal{F}_s) = M_s$.

Property (iii) can also be written as follows:

$$E(M_t - M_s | \mathcal{F}_s) = 0$$

Notice that if the time runs in a finite interval $[0, T]$, then we have

$$M_t = E(M_T|\mathcal{F}_t),$$

and this implies that the terminal random variable M_T determines the martingale.

In a similar way we can introduce the notions of continuous time submartingale and supermartingale.

As in the discrete time the expectation of a martingale is constant:

$$E(M_t) = E(M_0).$$

Also, most of the properties of discrete time martingales hold in continuous time. For instance, we have the following version of Doob's maximal inequality:

Proposition 36 *Let $\{M_t, 0 \le t \le T\}$ be a martingale with continuous trajectories. Then, for all $p \ge 1$ and all $\lambda > 0$ we have*

$$P\left(\sup_{0\le t\le T} |M_t| > \lambda\right) \le \frac{1}{\lambda^p} E(|M_T|^p).$$

This inequality allows to estimate moments of $\sup_{0\le t\le T} |M_t|$. For instance, we have

$$E\left(\sup_{0\le t\le T} |M_t|^2\right) \le 4E(|M_T|^2).$$

Exercises

4.1 Let X and Y be two independent random variables such that $P(X = 1) = P(Y = 1) = p$, and $P(X = 0) = P(Y = 0) = 1 - p$. Set $Z = \mathbf{1}_{\{X+Y=0\}}$. Compute $E(X|Z)$ and $E(Y|Z)$. Are these random variables still independent?

4.2 Let $\{Y_n\}_{n\ge1}$ be a sequence of independent random variable uniformly distributed in $[-1, 1]$. Set $S_0 = 0$ and $S_n = Y_1 + \cdots + Y_n$ if $n \ge 1$. Check whether the following sequences are martingales:

$$\begin{aligned} M_n^{(1)} &= \sum_{k=1}^{n} S_{k-1}^2 Y_k,\ n \ge 1,\ M_0^{(1)} = 0 \\ M_n^{(2)} &= S_n^2 - \frac{n}{3},\ M_0^{(2)} = 0. \end{aligned}$$

4.3 Consider a sequence of independent random variables $\{X_n\}_{n\geq 1}$ with laws $N(0, \sigma^2)$. Define

$$Y_n = \exp\left(a \sum_{k=1}^{n} X_k - n\sigma^2\right),$$

where a is a real parameter and $Y_0 = 1$. For which values of a the sequence Y_n is a martingale?

4.4 Let $Y_1, Y_2, \ldots$ be nonnegative independent and identically distributed random variables with $E(Y_n) = 1$. Show that $X_0 = 1$, $X_n = Y_1 Y_2 \cdots Y_n$ defines a martingale. Show that the almost sure limit of X_n is zero if $P(Y_n = 1) < 1$ (Apply the strong law of large numbers to $\log Y_n$).

4.5 Let S_n be the total assets of an insurance company at the end of the year n. In year n, premiums totaling $c > 0$ are received and claims ξ_n are paid where ξ_n has the normal distribution $N(\mu, \sigma^2)$ and $\mu < c$. The company is ruined if assets drop to 0 or less. Show that

$$P(\text{ruin}) \leq \exp(-2(c - \mu)S_0/\sigma^2).$$

4.6 Let S_n be an asymmetric random walk with $p > 1/2$, and let $T = \inf\{n : S_n = 1\}$. Show that $S_n - (p-q)n$ is a martingale. Use this property to check that $E(T) = 1/(2p-1)$. Using the fact that $(S_n - (p-q)n)^2 - \sigma^2 n$ is a martingale, where $\sigma^2 = 1 - (p-q)^2$, show that $\text{Var}(T) = \left(1 - (p-q)^2\right)/(p-q)^3$.

5 Stochastic Calculus

5.1 Brownian motion

In 1827 Robert Brown observed the complex and erratic motion of grains of pollen suspended in a liquid. It was later discovered that such irregular motion comes from the extremely large number of collisions of the suspended pollen grains with the molecules of the liquid. In the 20's Norbert Wiener presented a mathematical model for this motion based on the theory of stochastic processes. The position of a particle at each time $t \geq 0$ is a three dimensional random vector B_t.

The mathematical definition of a Brownian motion is the following:

Definition 37 *A stochastic process $\{B_t, t \geq 0\}$ is called a Brownian motion if it satisfies the following conditions:*

i) $B_0 = 0$

ii) For all $0 \leq t_1 < \cdots < t_n$ the increments $B_{t_n} - B_{t_{n-1}}, \ldots, B_{t_2} - B_{t_1}$, are independent random variables.

iii) If $0 \leq s < t$, the increment $B_t - B_s$ has the normal distribution $N(0, t - s)$

iv) The process $\{B_t\}$ has continuous trajectories.

Remarks:

1) Brownian motion is a Gaussian process. In fact, the probability distribution of a random vector $(B_{t_1}, \ldots, B_{t_n})$, for $0 < t_1 < \cdots < t_n$, is normal, because this vector is a linear transformation of the vector $\left(B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}\right)$ which has a joint normal distribution, because its components are independent and normal.

2) The mean and autocovariance functions of the Brownian motion are:

$$\begin{aligned} E(B_t) &= 0 \\ E(B_s B_t) &= E(B_s(B_t - B_s + B_s)) \\ &= E(B_s(B_t - B_s)) + E(B_s^2) = s = \min(s, t) \end{aligned}$$

if $s \leq t$. It is easy to show that a Gaussian process with zero mean and autocovariance function $\Gamma_X(s,t) = \min(s,t)$, satisfies the above conditions i), ii) and iii).

3) The autocovariance function $\Gamma_X(s,t) = \min(s,t)$ is nonnegative definite because it can be written as

$$\min(s,t) = \int_0^\infty \mathbf{1}_{[0,s]}(r)\mathbf{1}_{[0,t]}(r)dr,$$

so

$$\begin{aligned} \sum_{i,j=1}^n a_i a_j \min(t_i, t_j) &= \sum_{i,j=1}^n a_i a_j \int_0^\infty \mathbf{1}_{[0,t_i]}(r)\mathbf{1}_{[0,t_j]}(r)dr \\ &= \int_0^\infty \left[\sum_{i=1}^n a_i \mathbf{1}_{[0,t_i]}(r)\right]^2 dr \geq 0. \end{aligned}$$

Therefore, by Kolmogorov's theorem there exists a Gaussian process with zero mean and covariance function $\Gamma_X(s,t) = \min(s,t)$. On the other hand, Kolmogorov's continuity criterion allows to choose a version of this process with continuous trajectories. Indeed, the increment $B_t - B_s$ has the normal distribution $N(0, t-s)$, and this implies that for any natural number k we have

$$E\left[(B_t - B_s)^{2k}\right] = \frac{(2k)!}{2^k k!}(t-s)^k. \tag{19}$$

So, choosing $k = 2$, it is enough because

$$E\left[(B_t - B_s)^4\right] = 3(t-s)^2.$$

4) In the definition of the Brownian motion we have assumed that the probability space $(\Omega, \mathcal{F}, P)$ is arbitrary. The mapping

$$\begin{aligned} \Omega &\rightarrow C([0,\infty), \mathbb{R}) \\ \omega &\rightarrow B_{\cdot}(\omega) \end{aligned}$$

induces a probability measure $P_B = P \circ B^{-1}$, called the *Wiener measure,* on the space of continuous functions $C = C([0,\infty),\mathbb{R})$ equipped with its Borel σ-field $\mathcal{B}_C$. Then we can take as canonical probability space for the Brownian motion the space $(C, \mathcal{B}_C, P_B)$. In this canonical space, the random variables are the evaluation maps: $X_t(\omega) = \omega(t)$.

5.1.1 Regularity of the trajectories

From the Kolmogorov's continuity criterium and using (19) we get that for all $\varepsilon > 0$ there exists a random variable $G_{\varepsilon,T}$ such that

$$|B_t - B_s| \leq G_{\varepsilon,T}|t-s|^{\frac{1}{2}-\varepsilon},$$

for all $s,t \in [0,T]$. That is, the trajectories of the Brownian motion are Hölder continuous of order $\frac{1}{2}-\varepsilon$ for all $\varepsilon > 0$. Intuitively, this means that

$$\Delta B_t = B_{t+\Delta t} - B_t \backsimeq (\Delta t)^{\frac{1}{2}}.$$

This approximation is exact in mean: $E\left[(\Delta B_t)^2\right] = \Delta t$.

5.1.2 Quadratic variation

Fix a time interval $[0,t]$ and consider a subdivision π of this interval

$$0 = t_0 < t_1 < \cdots < t_n = t.$$

The norm of the subdivision π is defined by $|\pi| = \max_k \Delta t_k$, where $\Delta t_k = t_k - t_{k-1}$. Set $\Delta B_k = B_{t_k} - B_{t_{k-1}}$. Then, if $t_j = \frac{jt}{n}$ we have

$$\sum_{k=1}^{n} |\Delta B_k| \backsimeq n\left(\frac{t}{n}\right)^{\frac{1}{2}} \longrightarrow \infty,$$

whereas

$$\sum_{k=1}^{n} (\Delta B_k)^2 \backsimeq n\frac{t}{n} = t.$$

These properties can be formalized as follows. First, we will show that $\sum_{k=1}^{n}(\Delta B_k)^2$ converges in mean square to the length of the interval as the norm of the subdivision tends to zero:

$$
\begin{aligned}
E\left[\left(\sum_{k=1}^{n}(\Delta B_k)^2 - t\right)^2\right] &= E\left[\left(\sum_{k=1}^{n}\left[(\Delta B_k)^2 - \Delta t_k\ \right]\right)^2\right] \\
&= \sum_{k=1}^{n} E\left(\left[(\Delta B_k)^2 - \Delta t_k\right]^2\right) \\
&= \sum_{k=1}^{n}\left[3\,(\Delta t_k)^2 - 2\,(\Delta t_k)^2 + (\Delta t_k)^2\right] \\
&= 2\sum_{k=1}^{n}(\Delta t_k)^2 \leq 2t|\pi| \xrightarrow{|\pi|\to 0} 0.
\end{aligned}
$$

On the other hand, the total variation, defined by

$$
V = \sup_{\pi} \sum_{k=1}^{n} |\Delta B_k|
$$

is infinite with probability one. In fact, using the continuity of the trajectories of the Brownian motion, we have

$$
\sum_{k=1}^{n}(\Delta B_k)^2 \leq \sup_k |\Delta B_k| \left(\sum_{k=1}^{n}|\Delta B_k|\right) \leq V \sup_k |\Delta B_k| \xrightarrow{|\pi|\to 0} 0 \qquad (20)
$$

if $V < \infty$, which contradicts the fact that $\sum_{k=1}^{n}(\Delta B_k)^2$ converges in mean square to t as $|\pi| \to 0$.

5.1.3 Self-similarity

For any $a > 0$ the process

$$
\left\{a^{-1/2}B_{at}, t \geq 0\right\}
$$

is a Brownian motion. In fact, this process verifies easily properties (i) to (iv).

5.1.4 Stochastic Processes Related to Brownian Motion

1.- *Brownian bridge*: Consider the process

$$X_t = B_t - tB_1,$$

$t \in [0,1]$. It is a centered normal process with autocovariance function

$$E(X_t X_s) = \min(s,t) - st,$$

which verifies $X_0 = 0$, $X_1 = 0$.

2.- *Brownian motion with drift:* Consider the process

$$X_t = \sigma B_t + \mu t,$$

$t \geq 0$, where $\sigma > 0$ and $\mu \in \mathbb{R}$ are constants. It is a Gaussian process with

$$\begin{aligned} E(X_t) &= \mu t, \\ \Gamma_X(s,t) &= \sigma^2 \min(s,t). \end{aligned}$$

3.- *Geometric Brownian motion:* It is the stochastic process proposed by Black, Scholes and Merton as model for the curve of prices of financial assets. By definition this process is given by

$$X_t = e^{\sigma B_t + \mu t},$$

$t \geq 0$, where $\sigma > 0$ and $\mu \in \mathbb{R}$ are constants. That is, this process is the exponential of a Brownian motion with drift. This process is not Gaussian, and the probability distribution of X_t is lognormal.

5.1.5 Simulation of the Brownian Motion

Brownian motion can be regarded as the limit of a symmetric random walk. Indeed, fix a time interval $[0,T]$. Consider n independent are identically distributed random variables $\xi_1, \ldots, \xi_n$ with zero mean and variance $\frac{T}{n}$. Define the partial sums

$$R_k = \xi_1 + \cdots + \xi_k, \; k = 1, \ldots, n.$$

By the *Central Limit Theorem* the sequence R_n converges, as n tends to infinity, to the normal distribution $N(0,T)$.

Consider the continuous stochastic process $S_n(t)$ defined by linear interpolation from the values

$$S_n(\frac{kT}{n}) = R_k \; k = 0, \dots, n.$$

Then, a functional version of the Central Limit Theorem, known as *Donsker Invariance Principle,* says that the sequence of stochastic processes $S_n(t)$ converges in law to the Brownian motion on $[0,T]$. This means that for any continuous and bounded function $\varphi : C([0,T]) \to \mathbb{R}$, we have

$$E(\varphi(S_n)) \to E(\varphi(B)),$$

as n tends to infinity.

The trajectories of the Brownian motion can also be simulated by means of Fourier series with random coefficients. Suppose that $\{e_n, n \geq 0\}$ is an orthonormal basis of the Hilbert space $L^2([0,T])$. Suppose that $\{Z_n, n \geq 0\}$ are independent random variables with law $N(0,1)$. Then, the random series

$$\sum_{n=0}^{\infty} Z_n \int_0^t e_n(s) ds$$

converges uniformly on $[0,T]$, for almost all ω, to a Brownian motion $\{B_t, t \in [0,T]\}$, that is,

$$\sup_{0 \leq t \leq T} \left| \sum_{n=0}^{N} Z_n \int_0^t e_n(s) ds - B_t \right| \overset{\text{a.s.}}{\to} 0.$$

This convergence also holds in mean square. Notice that

$$\begin{aligned} & E\left[\left(\sum_{n=0}^{N} Z_n \int_0^t e_n(r) dr \right) \left(\sum_{n=0}^{N} Z_n \int_0^s e_n(r) dr \right) \right] \\ = & \sum_{n=0}^{N} \left(\int_0^t e_n(r) dr \right) \left(\int_0^s e_n(r) dr \right) \\ = & \sum_{n=0}^{N} \left\langle \mathbf{1}_{[0,t]}, e_n \right\rangle_{L^2([0,T])} \left\langle \mathbf{1}_{[0,s]}, e_n \right\rangle_{L^2([0,T])} \overset{N \to \infty}{\to} \left\langle \mathbf{1}_{[0,t]}, \mathbf{1}_{[0,s]} \right\rangle_{L^2([0,T])} = s \wedge t. \end{aligned}$$

In particular, if we take the basis formed by trigonometric functions, $e_n(t) = \frac{1}{\sqrt{\pi}} \cos(nt/2)$, for $n \geq 1$, and $e_0(t) = \frac{1}{\sqrt{2\pi}}$, on the interval $[0, 2\pi]$, we obtain the Paley-Wiener representation of Brownian motion:

$$B_t = Z_0 \frac{t}{\sqrt{2\pi}} + \frac{2}{\sqrt{\pi}} \sum_{n=1}^{\infty} Z_n \frac{\sin(nt/2)}{n}, \ t \in [0, 2\pi].$$

In order to use this formula to get a simulation of Brownian motion, we have to choose some number M of trigonometric functions and a number N of discretization points:

$$Z_0 \frac{t_j}{\sqrt{2\pi}} + \frac{2}{\sqrt{\pi}} \sum_{n=1}^{M} Z_n \frac{\sin(nt_j/2)}{n},$$

where $t_j = \frac{2\pi j}{N}$, $j = 0, 1, \ldots, N$.

5.2 Martingales Related with Brownian Motion

Consider a Brownian motion $\{B_t, t \geq 0\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$. For any time t, we define the σ-field $\mathcal{F}_t$ generated by the random variables $\{B_s, s \leq t\}$ and the events in $\mathcal{F}$ of probability zero. That is, $\mathcal{F}_t$ is the smallest σ-field that contains the sets of the form

$$\{B_s \in A\} \cup N,$$

where $0 \leq s \leq t$, A is a Borel subset of $\mathbb{R}$, and $N \in \mathcal{F}$ is such that $P(N) = 0$. Notice that $\mathcal{F}_s \subset \mathcal{F}_t$ if $s \leq t$, that is , $\{\mathcal{F}_t, t \geq 0\}$ is a non-decreasing family of σ-fields. We say that $\{\mathcal{F}_t, t \geq 0\}$ is a *filtration* in the probability space $(\Omega, \mathcal{F}, P)$.

We say that a stochastic process $\{u_t, u \geq 0\}$ is *adapted* (to the filtration $\mathcal{F}_t$) if for all t the random variable u_t is $\mathcal{F}_t$-measurable.

The inclusion of the events of probability zero in each σ-field $\mathcal{F}_t$ has the following important consequences:

a) Any version of an adapted process is adapted.

b) The family of σ-fields is right-continuous: For all $t \geq 0$

$$\cap_{s>t} \mathcal{F}_s = \mathcal{F}_t.$$

If B_t is a Brownian motion and $\mathcal{F}_t$ is the filtration generated by B_t, then, the processes

$$\begin{aligned}&B_t\\&B_t^2 - t\\&\exp(aB_t - \frac{a^2 t}{2})\end{aligned}$$

are martingales. In fact, B_t is a martingale because

$$E(B_t - B_s|\mathcal{F}_s) = E(B_t - B_s) = 0.$$

For $B_t^2 - t$, we can write, using the properties of the conditional expectation, for $s < t$

$$\begin{aligned}E(B_t^2|\mathcal{F}_s) &= E((B_t - B_s + B_s)^2|\mathcal{F}_s)\\&= E((B_t - B_s)^2|\mathcal{F}_s) + 2E((B_t - B_s)B_s|\mathcal{F}_s)\\&\quad + E(B_s^2|\mathcal{F}_s)\\&= E(B_t - B_s)^2 + 2B_s E((B_t - B_s)|\mathcal{F}_s) + B_s^2\\&= t - s + B_s^2.\end{aligned}$$

Finally, for $\exp(aB_t - \frac{a^2 t}{2})$ we have

$$\begin{aligned}E(e^{aB_t - \frac{a^2 t}{2}}|\mathcal{F}_s) &= e^{aB_s} E(e^{a(B_t - B_s) - \frac{a^2 t}{2}}|\mathcal{F}_s)\\&= e^{aB_s} E(e^{a(B_t - B_s) - \frac{a^2 t}{2}})\\&= e^{aB_s} e^{\frac{a^2(t-s)}{2} - \frac{a^2 t}{2}} = e^{aB_s - \frac{a^2 s}{2}}.\end{aligned}$$

As an application of the martingale property of this process we will compute the probability distribution of the arrival time of the Brownian motion to some fixed level.

Example 1 Let B_t be a Brownian motion and $\mathcal{F}_t$ the filtration generated by B_t. Consider the stopping time

$$\tau_a = \inf\{t \geq 0 : B_t = a\},$$

where $a > 0$. The process $M_t = e^{\lambda B_t - \frac{\lambda^2 t}{2}}$ is a martingale such that

$$E(M_t) = E(M_0) = 1.$$

By the Optional Stopping Theorem we obtain

$$E(M_{\tau_a \wedge N}) = 1,$$

for all $N \geq 1$. Notice that

$$M_{\tau_a \wedge N} = \exp\left(\lambda B_{\tau_a \wedge N} - \frac{\lambda^2 (\tau_a \wedge N)}{2}\right) \leq e^{a\lambda}.$$

On the other hand,

$$\begin{array}{ll} \lim_{N\to\infty} M_{\tau_a \wedge N} = M_{\tau_a} & \text{if} \quad \tau_a < \infty \\ \lim_{N\to\infty} M_{\tau_a \wedge N} = 0 & \text{if} \quad \tau_a = \infty \end{array}$$

and the dominated convergence theorem implies

$$E\left(\mathbf{1}_{\{\tau_a < \infty\}} M_{\tau_a}\right) = 1,$$

that is,

$$E\left(\mathbf{1}_{\{\tau_a < \infty\}} \exp\left(-\frac{\lambda^2 \tau_a}{2}\right)\right) = e^{-\lambda a}.$$

Letting $\lambda \downarrow 0$ we obtain

$$P(\tau_a < \infty) = 1,$$

and, consequently,

$$E\left(\exp\left(-\frac{\lambda^2 \tau_a}{2}\right)\right) = e^{-\lambda a}.$$

With the change of variables $\frac{\lambda^2}{2} = \alpha$, we get

$$E(\exp(-\alpha \tau_a)) = e^{-\sqrt{2\alpha}a}. \tag{21}$$

From this expression we can compute the distribution function of the random variable τ_a:

$$P(\tau_a \leq t) = \int_0^t \frac{a e^{-a^2/2s}}{\sqrt{2\pi s^3}} ds.$$

On the other hand, the expectation of τ_a can be obtained by computing the derivative of (21) with respect to the variable a:

$$E(\tau_a \exp(-\alpha \tau_a)) = \frac{a e^{-\sqrt{2\alpha}a}}{\sqrt{2\alpha}},$$

and letting $\alpha \downarrow 0$ we obtain $E(\tau_a) = +\infty$.

5.3 Stochastic Integrals

We want to define stochastic integrals of the form.

$$\boxed{\int_0^T u_t dB_t}.$$

One possibility is to interpret this integral as a path-wise *Riemann Stieltjes* integral. That means, if we consider a sequence of partitions of an interval $[0, T]$:

$$\tau_n : \ 0 = t_0^n < t_1^n < \cdots < t_{k_n-1}^n < t_{k_n}^n = T$$

and intermediate points:

$$\sigma_n : \quad t_i^n \le s_i^n \le t_{i+1}^n \ , \ i = 0, \ldots, k_n - 1,$$

such that $\sup_i (t_i^n - t_{i-1}^n) \stackrel{n\to\infty}{\longrightarrow} 0$, then, given two functions f and g on the interval $[0, T]$, *the Riemann Stieltjes integral* $\int_0^T f dg$ is defined as

$$\lim_{n\to\infty} \sum_{i=1}^{n} f(s_{i-1}) \Delta g_i$$

provided this limit exists and it is independent of the sequences τ_n and σ_n, where $\Delta g_i = g(t_i) - g(t_{i-1})$.

The Riemann Stieltjes integral $\int_0^T f dg$ exists if f is continuous and g has bounded variation, that is,

$$\sup_\tau \sum_i |\Delta g_i| < \infty.$$

In particular, if g is continuously differentiable, $\int_0^T f dg = \int_0^T f(t) g'(t) dt$.

We know that the trajectories of Brownian motion have infinite variation on any finite interval. So, we cannot use this result to define the path-wise Riemann-Stieltjes integral $\int_0^T u_t(\omega) dB_t(\omega)$ for a continuous process u.

Note, however, that if u has continuously differentiable trajectories, then the path-wise Riemann Stieltjes integral $\int_0^T u_t(\omega) dB_t(\omega)$ exists and it can be computed integrating by parts:

$$\boxed{\int_0^T u_t dB_t = u_T B_T - \int_0^T u_t' B_t dt.}$$

We are going to construct the integral $\int_0^T u_t dB_t$ by means of a global probabilistic approach.

Denote by $L^2_{a,T}$ the space of stochastic processes

$$u = \{u_t, t \in [0,T]\}$$

such that:

a) u is adapted and measurable (the mapping $(s,\omega) \longrightarrow u_s(\omega)$ is measurable on the product space $[0,T] \times \Omega$ with respect to the product σ-field $\mathcal{B}_{[0,T]} \times \mathcal{F}$).

b) $E\left(\int_0^T u_t^2 dt\right) < \infty$.

Under condition a) it can be proved that there is a version of u which is *progressively measurable*. This condition means that for all $t \in [0,T]$, the mapping $(s,\omega) \longrightarrow u_s(\omega)$ on $[0,t] \times \Omega$ is measurable with respect to the product σ-field $\mathcal{B}_{[0,t]} \times \mathcal{F}_t$. This condition is slightly strongly that being adapted and measurable, and it is needed to guarantee that random variables of the form $\int_0^t u_s ds$ are $\mathcal{F}_t$-measurable.

Condition b) means that the moment of second order of the process is integrable on the time interval $[0,T]$. In fact, by Fubini's theorem we deduce

$$E\left(\int_0^T u_t^2 dt\right) = \int_0^T E\left(u_t^2\right) dt.$$

Also, condition b) means that the process u as a function of the two variables (t,ω) belongs to the Hilbert space. $L^2([0,T] \times \Omega)$.

We will define the *stochastic integral* $\int_0^T u_t dB_t$ for processes u in $L^2_{a,T}$ as the limit in mean square of the integral of simple processes. By definition a process u in $L^2_{a,T}$ is a *simple process* if it is of the form

$$u_t = \sum_{j=1}^n \phi_j \mathbf{1}_{(t_{j-1},t_j]}(t), \tag{22}$$

where $0 = t_0 < t_1 < \cdots < t_n = T$ and ϕ_j are square integrable $\mathcal{F}_{t_{j-1}}$-measurable random variables.

Given a simple process u of the form (22) we define the stochastic integral of u with respect to the Brownian motion B as

$$\int_0^T u_t dB_t = \sum_{j=1}^n \phi_j \left(B_{t_j} - B_{t_{j-1}} \right). \tag{23}$$

The stochastic integral defined on the space $\mathcal{E}$ of simple processes possesses the following *isometry property*:

$$\boxed{E\left[\left(\int_0^T u_t dB_t\right)^2\right] = E\left(\int_0^T u_t^2 dt\right)} \tag{24}$$

Proof of the isometry property: Set $\Delta B_j = B_{t_j} - B_{t_{j-1}}$. Then

$$E\left(\phi_i \phi_j \Delta B_i \Delta B_j\right) = \begin{cases} 0 & \text{if } \ i \neq j \\ E\left(\phi_j^2\right)(t_j - t_{j-1}) & \text{if } \ i = j \end{cases}$$

because if $i < j$ the random variables $\phi_i \phi_j \Delta B_i$ and ΔB_j are independent and if $i = j$ the random variables ϕ_i^2 and $(\Delta B_i)^2$ are independent. So, we obtain

$$\begin{aligned} E\left(\int_0^T u_t dB_t\right)^2 &= \sum_{i,j=1}^n E\left(\phi_i \phi_j \Delta B_i \Delta B_j\right) = \sum_{i=1}^n E\left(\phi_i^2\right)(t_i - t_{i-1}) \\ &= E\left(\int_0^T u_t^2 dt\right). \end{aligned}$$

□

The extension of the stochastic integral to processes in the class $L_{a,T}^2$ is based on the following approximation result:

Lemma 38 *If u is a process in $L_{a,T}^2$ then, there exists a sequence of simple processes $u^{(n)}$ such that*

$$\lim_{n \to \infty} E\left(\int_0^T \left| u_t - u_t^{(n)} \right|^2 dt\right) = 0. \tag{25}$$

Proof: The proof of this Lemma will be done in two steps:

1. Suppose first that the process u is continuous in mean square. In this case, we can choose the approximating sequence

$$u_t^{(n)} = \sum_{j=1}^{n} u_{t_{j-1}} \mathbf{1}_{(t_{j-1},t_j]}(t),$$

where $t_j = \frac{jT}{n}$. The continuity in mean square of u implies that

$$E\left(\int_0^T \left|u_t - u_t^{(n)}\right|^2 dt\right) \leq T \sup_{|t-s|\leq T/n} E\left(|u_t - u_s|^2\right),$$

which converges to zero as n tends to infinity.

2. Suppose now that u is an arbitrary process in the class $L^2_{a,T}$. Then, we need to show that there exists a sequence $v^{(n)}$, of processes in $L^2_{a,T}$, continuous in mean square and such that

$$\lim_{n\to\infty} E\left(\int_0^T \left|u_t - v_t^{(n)}\right|^2 dt\right) = 0. \tag{26}$$

In order to find this sequence we set

$$v_t^{(n)} = n\int_{t-\frac{1}{n}}^{t} u_s ds = n\left(\int_0^t u_s ds - \int_0^t u_{s-\frac{1}{n}} ds\right),$$

with the convention $u(s) = 0$ if $s < 0$. These processes are continuous in mean square (actually, they have continuous trajectories) and they belong to the class $L^2_{a,T}$. Furthermore, (26) holds because or each (t,ω) we have

$$\int_0^T \left|u(t,\omega) - v^{(n)}(t,\omega)\right|^2 dt \overset{n\to\infty}{\longrightarrow} 0$$

(this is a consequence of the fact that $v_t^{(n)}$ is defined from u_t by means of a convolution with the kernel $n1_{[-\frac{1}{n},0]}$) and we can apply the dominated convergence theorem on the product space $[0,T]\times\Omega$ because

$$\int_0^T \left|v^{(n)}(t,\omega)\right|^2 dt \leq \int_0^T |u(t,\omega)|^2 dt.$$

□

Definition 39 *The stochastic integral of a process u in the $L^2_{a,T}$ is defined as the following limit in mean square*

$$\int_0^T u_t dB_t = \lim_{n\to\infty} \int_0^T u_t^{(n)} dB_t, \tag{27}$$

where $u^{(n)}$ is an approximating sequence of simple processes that satisfy (25).

Notice that the limit (27) exists because the sequence of random variables $\int_0^T u_t^{(n)} dB_t$ is Cauchy in the space $L^2(\Omega)$, due to the isometry property (24):

$$\begin{aligned} E\left[\left(\int_0^T u_t^{(n)} dB_t - \int_0^T u_t^{(m)} dB_t\right)^2\right] &= E\left(\int_0^T \left(u_t^{(n)} - u_t^{(m)}\right)^2 dt\right) \\ &\le 2E\left(\int_0^T \left(u_t^{(n)} - u_t\right)^2 dt\right) \\ &\quad +2E\left(\int_0^T \left(u_t - u_t^{(m)}\right)^2 dt\right) \overset{n,m\to\infty}{\longrightarrow} 0. \end{aligned}$$

On the other hand, the limit (27) does not depend on the approximating sequence $u^{(n)}$.

Properties of the stochastic integral:

1.- Isometry:

$$E\left[\left(\int_0^T u_t dB_t\right)^2\right] = E\left(\int_0^T u_t^2 dt\right).$$

2.- Mean zero:

$$E\left[\left(\int_0^T u_t dB_t\right)\right] = 0.$$

3.- Linearity:

$$\int_0^T (au_t + bv_t)\, dB_t = a\int_0^T u_t dB_t + b\int_0^T v_t dB_t.$$

4.- Local property: $\int_0^T u_t dB_t = 0$ almost surely, on the set $G = \left\{\int_0^T u_t^2 dt = 0\right\}$.

Local property holds because on the set G the approximating sequence

$$u_t^{(n,m,N)} = \sum_{j=1}^{n} m \left[\int_{\frac{(j-1)T}{n} - \frac{1}{m}}^{\frac{(j-1)T}{n}} u_s ds \right] \mathbf{1}_{\left(\frac{(j-1)T}{n}, \frac{jT}{n} \right]}(t)$$

vanishes.

Example 1

$$\int_0^T B_t dB_t = \frac{1}{2} B_T^2 - \frac{1}{2} T.$$

The process B_t being continuous in mean square, we can choose as approximating sequence

$$u_t^{(n)} = \sum_{j=1}^{n} B_{t_{j-1}} \mathbf{1}_{(t_{j-1}, t_j]}(t),$$

where $t_j = \frac{jT}{n}$, and we obtain

$$\begin{aligned} \int_0^T B_t dB_t &= \lim_{n\to\infty} \sum_{j=1}^{n} B_{t_{j-1}} \left(B_{t_j} - B_{t_{j-1}} \right) \\ &= \frac{1}{2} \lim_{n\to\infty} \left(B_{t_j}^2 - B_{t_{j-1}}^2 \right) - \frac{1}{2} \lim_{n\to\infty} \sum_{j=1}^{n} \left(B_{t_j} - B_{t_{j-1}} \right)^2 \\ &= \frac{1}{2} B_T^2 - \frac{1}{2} T. \qquad (28) \end{aligned}$$

If x_t is a continuously differentiable function such that $x_0 = 0$, we know that

$$\int_0^T x_t dx_t = \int_0^T x_t x_t' dt = \frac{1}{2} x_T^2.$$

Notice that in the case of Brownian motion an additional term appears: $-\frac{1}{2}T$.

Example 2 Consider a deterministic function g such that $\int_0^T g(s)^2 ds < \infty$. The stochastic integral $\int_0^T g_s dB_s$ is a normal random variable with law

$$N(0, \int_0^T g(s)^2 ds).$$

5.4 Indefinite Stochastic Integrals

Consider a stochastic process u in the space $L^2_{a,T}$. Then, for any $t \in [0,T]$ the process $u\mathbf{1}_{[0,t]}$ also belongs to $L^2_{a,T}$, and we can define its stochastic integral:

$$\int_0^t u_s dB_s := \int_0^T u_s \mathbf{1}_{[0,t]}(s) dB_s.$$

In this way we have constructed a new stochastic process

$$\left\{ \int_0^t u_s dB_s, 0 \leq t \leq T \right\}$$

which is the indefinite integral of u with respect to B.

Properties of the indefinite integrals:

1. Additivity: For any $a \leq b \leq c$ we have

$$\int_a^b u_s dB_s + \int_b^c u_s dB_s = \int_a^c u_s dB_s.$$

2. Factorization: If $a < b$ and A is an event of the σ-field $\mathcal{F}_a$ then,

$$\int_a^b \mathbf{1}_A u_s dB_s = \mathbf{1}_A \int_a^b u_s dB_s.$$

   Actually, this property holds replacing $\mathbf{1}_A$ by any bounded and $\mathcal{F}_a$-measurable random variable.

3. Martingale property: The indefinite stochastic integral $M_t = \int_0^t u_s dB_s$ of a process $u \in L^2_{a,T}$ is a martingale with respect to the filtration $\mathcal{F}_t$.

   Proof of the martingale property: Consider a sequence $u^{(n)}$ of simple processes such that

$$\lim_{n \to \infty} E\left(\int_0^T \left| u_t - u_t^{(n)} \right|^2 dt \right) = 0.$$

   Set $M_n(t) = \int_0^t u_s^{(n)} dB_s$. If ϕ_j is the value of the simple stochastic process $u^{(n)}$ on each interval $(t_{j-1}, t_j]$, $j = 1, \ldots, n$ and $s \leq t_k \leq$

$t_{m-1} \leq t$ we have

$$\begin{aligned}
& E\left(M_n(t) - M_n(s)|\mathcal{F}_s\right) \\
= \; & E\left(\phi_k(B_{t_k} - B_s) + \sum_{j=k+1}^{m-1} \phi_j \Delta B_j + \phi_m(B_t - B_{t_{m-1}})|\mathcal{F}_s\right) \\
= \; & E\left(\phi_k(B_{t_k} - B_s)|\mathcal{F}_s\right) + \sum_{j=k+1}^{m-1} E\left(E\left(\phi_j \Delta B_j|\mathcal{F}_{t_{j-1}}\right)|\mathcal{F}_s\right) \\
& + E\left(E\left(\phi_m(B_t - B_{t_{m-1}})|\mathcal{F}_{t_{m-1}}\right)|\mathcal{F}_s\right) \\
= \; & \phi_k E\left((B_{t_k} - B_s)|\mathcal{F}_s\right) + \sum_{j=k+1}^{m-1} E\left(\phi_j E\left(\Delta B_j|\mathcal{F}_{t_{j-1}}\right)|\mathcal{F}_s\right) \\
& + E\left(\phi_m E\left((B_t - B_{t_{m-1}})|\mathcal{F}_{t_{m-1}}\right)|\mathcal{F}_s\right) \\
= \; & 0.
\end{aligned}$$

Finally, the result follows from the fact that the convergence in mean square $M_n(t) \to M_t$ implies the convergence in mean square of the conditional expectations. □

4. Continuity: Suppose that u belongs to the space $L^2_{a,T}$. Then, the stochastic integral $M_t = \int_0^t u_s dB_s$ has a version with continuous trajectories.

Proof of continuity: With the same notation as above, the process M_n is a martingale with continuous trajectories. Then, Doob's maximal inequality applied to the continuous martingale $M_n - M_m$ with $p = 2$ yields

$$\begin{aligned}
P\left(\sup_{0 \leq t \leq T} |M_n(t) - M_m(t)| > \lambda\right) & \leq \frac{1}{\lambda^2} E\left(|M_n(T) - M_m(T)|^2\right) \\
& = \frac{1}{\lambda^2} E\left(\int_0^T \left|u_t^{(n)} - u_t^{(m)}\right|^2 dt\right) \stackrel{n,m\to\infty}{\longrightarrow} 0.
\end{aligned}$$

We can choose an increasing sequence of natural numbers $n_k, k = 1, 2, \ldots$ such that

$$P\left(\sup_{0 \leq t \leq T} |M_{n_{k+1}}(t) - M_{n_k}(t)| > 2^{-k}\right) \leq 2^{-k}.$$

The events $A_k := \left\{\sup_{0\le t\le T} |M_{n_{k+1}}(t) - M_{n_k}(t)| > 2^{-k}\right\}$ verify

$$\sum_{k=1}^{\infty} P(A_k) < \infty.$$

Hence, Borel-Cantelli lemma implies that $P(\limsup_k A_k) = 0$, or

$$P(\liminf_k A_k^c) = 1.$$

That means, there exists a set N of probability zero such that for all $\omega \notin N$ there exists $k_1(\omega)$ such that for all $k \ge k_1(\omega)$

$$\sup_{0\le t\le T} |M_{n_{k+1}}(t,\omega) - M_{n_k}(t,\omega)| \le 2^{-k}.$$

As a consequence, if $\omega \notin N$, the sequence $M_{n_k}(t,\omega)$ is uniformly convergent on $[0,T]$ to a continuous function $J_t(\omega)$. On the other hand, we know that for any $t \in [0,T]$, $M_{n_k}(t)$ converges in mean square to $\int_0^t u_s dB_s$. So, $J_t(\omega) = \int_0^t u_s dB_s$ almost surely, for all $t \in [0,T]$, and we have proved that the indefinite stochastic integral possesses a continuous version. □

5. Maximal inequality for the indefinite integral: $M_t = \int_0^t u_s dB_s$ of a processes $u \in L^2_{a,T}$: For all $\lambda > 0$,

$$P\left(\sup_{0\le t\le T} |M_t| > \lambda\right) \le \frac{1}{\lambda^2} E\left(\int_0^T u_t^2 dt\right).$$

6. Stochastic integration up to a stopping time: If u belongs to the space $L^2_{a,T}$ and τ is a stopping time bounded by T, then the process $u\mathbf{1}_{[0,\tau]}$ also belongs to $L^2_{a,T}$ and we have:

$$\int_0^T u_t \mathbf{1}_{[0,\tau]}(t) dB_t = \int_0^\tau u_t \, dB_t. \tag{29}$$

Proof of (29)**:** The proof will be done in two steps:

(a) Suppose first that the process u has the form $u_t = F1_{(a,b]}(t)$, where $0 \le a < b \le T$ and $F \in L^2(\Omega, F_a, P)$. The stopping times

$\tau_n = \sum_{i=1}^{2^n} t_n^i 1_{A_n^i}$, where $t_n^i = \frac{iT}{2^n}$, $A_n^i = \left\{ \frac{(i-1)T}{2^n} \leq \tau < \frac{iT}{2^n} \right\}$ form a nonincreasing sequence which converges to τ. For any n we have

$$\begin{aligned}
\int_0^{\tau_n} u_t \, dB_t &= F(B_{b\wedge\tau_n} - B_{a\wedge\tau_n}) \\
\int_0^T u_t \mathbf{1}_{[0,\tau_n]}(t) dB_t &= \sum_{i=1}^{2^n} \mathbf{1}_{A_n^i} \int_0^T \mathbf{1}_{(a\wedge t_n^i, b\wedge t_n^i]}(t) F dB_t \\
&= \sum_{i=1}^{2^n} \mathbf{1}_{A_n^i} F(B_{b\wedge t_n^i} - B_{a\wedge t_n^i}) \\
&= F(B_{b\wedge\tau_n} - B_{a\wedge\tau_n}).
\end{aligned}$$

Taking the limit as n tends to infinity we deduce the equality in the case of a simple process.

(b) In the general case, it suffices to approximate the process u by simple processes. The convergence of the right-hand side of (29) follows from Doob's maximal inequality. □

The martingale $M_t = \int_0^t u_s dB_s$ has a nonzero quadratic variation, like the Brownian motion:

Proposition 40 (Quadratic variation) *Let u be a process in $L^2_{a,T}$. Then,*

$$\sum_{j=1}^{n} \left(\int_{t_{j-1}}^{t_j} u_s dB_s \right)^2 \overset{L^1(\Omega)}{\longrightarrow} \int_0^t u_s^2 ds$$

as n tends to infinity, where $t_j = \frac{jt}{n}$.

5.5 Extensions of the Stochastic Integral

Itô's stochastic integral $\int_0^T u_s dB_s$ can be defined for classes of processes larger than $L^2_{a,T}$.

A) First, we can replace the filtration $\mathcal{F}_t$ by a largest one $\mathcal{H}_t$ such that the Brownian motion B_t is a martingale with respect to $\mathcal{H}_t$. In fact, we only need the property

$$E(B_t - B_s | \mathcal{H}_s) = 0.$$

Notice that this martingale property also implies that $E((B_t - B_s)^2 |\mathcal{H}_s) = 0$, because

$$
\begin{aligned}
E((B_t - B_s)^2 |\mathcal{H}_s) &= E(2\int_s^t B_r dB_r + t - s|\mathcal{H}_s) \\
&= 2\lim_n E(\sum_{i=1}^n B_{t_t}(B_{t_t} - B_{t_{t-1}})|\mathcal{H}_s) + t - s \\
&= t - s.
\end{aligned}
$$

Example 3 Let $\{B_t, t \geq 0\}$ be a d-dimensional Brownian motion. That is, the components $\{B_k(t), t \geq 0\}$, $k = 1, \ldots, d$ are independent Brownian motions. Denote by $\mathcal{F}_t^{(d)}$ the filtration generated by B_t and the sets of probability zero. Then, each component $B_k(t)$ is a martingale with respect to $\mathcal{F}_t^{(d)}$, but $\mathcal{F}_t^{(d)}$ is not the filtration generated by $B_k(t)$. The above extension allows us to define stochastic integrals of the form

$$
\begin{aligned}
&\int_0^T B_2(s) dB_1^1(s), \\
&\int_0^T \sin(B_1^2(s) + B_1^2(s)) dB_2(s).
\end{aligned}
$$

B) The second extension consists in replacing property $E\left(\int_0^T u_t^2 dt\right) < \infty$ by the weaker assumption:

b') $P\left(\int_0^T u_t^2 dt < \infty\right) = 1$.

We denote by $L_{a,T}$the space of processes that verify properties a) and b'). Stochastic integral is extended to the space $L_{a,T}$ by means of a localization argument.

Suppose that u belongs to $L_{a,T}$. For each $n \geq 1$ we define the stopping time

$$
\tau_n = \inf\left\{t \geq 0 : \int_0^t u_s^2 ds = n\right\}, \tag{30}
$$

where, by convention, $\tau_n = T$ if $\int_0^T u_s^2 ds < n$. In this way we obtain a nondecreasing sequence of stopping times such that $\tau_n \uparrow T$. Furthermore,

$$t < \tau_n \Longleftrightarrow \int_0^t u_s^2 ds < n.$$

Set

$$u_t^{(n)} = u_t \mathbf{1}_{[0,\tau_n]}(t).$$

The process $u^{(n)} = \{u_t^{(n)}, 0 \leq t \leq T\}$ belongs to $L^2_{a,T}$ since $E\left(\int_0^T \left(u_s^{(n)}\right)^2 ds\right) \leq n$. If $n \leq m$, on the set $\{t \leq \tau_n\}$ we have

$$\int_0^t u_s^{(n)} dB_s = \int_0^t u_s^{(m)} dB_s$$

because by (29) we can write

$$\int_0^t u_s^{(n)} dB_s = \int_0^t u_s^{(m)} \mathbf{1}_{[0,\tau_n]}(s) dB_s = \int_0^{t \wedge \tau_n} u_s^{(m)} dB_s.$$

As a consequence, there exists an adapted and continuous process denoted by $\int_0^t u_s dB_s$ such that for any $n \geq 1$,

$$\int_0^t u_s^{(n)} dB_s = \int_0^t u_s dB_s$$

if $t \leq \tau_n$.

The stochastic integral of processes in the space $L_{a,T}$ is linear and has continuous trajectories. However, it may have infinite expectation and variance. Instead of the isometry property, there is a continuity property in probability as it follows from the next proposition:

Proposition 41 *Suppose that $u \in L_{a,T}$. For all $K, \delta > 0$ we have :*

$$P\left(\left|\int_0^T u_s dB_s\right| \geq K\right) \leq P\left(\int_0^T u_s^2 ds \geq \delta\right) + \frac{\delta}{K^2}.$$

Proof. Consider the stopping time defined by

$$\tau = \inf\left\{t \geq 0 : \int_0^t u_s^2 ds = \delta\right\},$$

with the convention that $\tau = T$ if $\int_0^T u_s^2 ds < \delta$.We have

$$\begin{aligned} P\left(\left|\int_0^T u_s dB_s\right| \geq K\right) &\leq P\left(\int_0^T u_s^2 ds \geq \delta\right) \\ &\quad + P\left(\left|\int_0^T u_s dB_s\right| \geq K, \int_0^T u_s^2 ds < \delta\right), \end{aligned}$$

and on the other hand,

$$\begin{aligned} P\left(\left|\int_0^T u_s dB_s\right| \geq K, \int_0^T u_s^2 ds < \delta\right) &= P\left(\left|\int_0^T u_s dB_s\right| \geq K, \tau = T\right) \\ &= P\left(\left|\int_0^\tau u_s dB_s\right| \geq K, \tau = T\right) \\ &\leq \frac{1}{K^2} E\left(\left|\int_0^\tau u_s dB_s\right|^2\right) \\ &= \frac{1}{K^2} E\left(\int_0^\tau u_s^2 ds\right) \leq \frac{\delta}{K^2}. \end{aligned}$$

■

As a consequence of the above proposition, if $u^{(n)}$ is a sequence of processes in the space $L_{a,T}$ which converges to $u \in L_{a,T}$ in probability:

$$P\left(\left|\int_0^T \left(u_s^{(n)} - u_s\right)^2 ds\right| > \epsilon\right) \overset{n\to\infty}{\to} 0, \text{ for all } \epsilon > 0$$

then,

$$\int_0^T u_s^{(n)} dB_s \overset{P}{\to} \int_0^T u_s dB_s.$$

5.6 Itô's Formula

Itô's formula is the stochastic version of the chain rule of the ordinary differential calculus. Consider the following example

$$\int_0^t B_s dB_s = \frac{1}{2} B_t^2 - \frac{1}{2} t,$$

that can be written as

$$B_t^2 = \int_0^t 2 B_s dB_s + t,$$

or in differential notation

$$d\left(B_t^2\right) = 2B_t dB_t + dt.$$

Formally, this follows from Taylor development of B_t^2 as a function of t, with the convention $(dB_t)^2 = dt$.

The stochastic process B_t^2 can be expressed as the sum of an indefinite stochastic integral $\int_0^t 2B_s dB_s$, plus a differentiable function. More generally, we will see that any process of the form $f(B_t)$, where f is twice continuously differentiable, can be expressed as the sum of an indefinite stochastic integral, plus a process with differentiable trajectories. This leads to the definition of *Itô processes*.

Denote by $L_{a,T}^1$ the space of processes v which satisfy properties a) and

b") $P\left(\int_0^T |v_t|\, dt < \infty\right) = 1.$

Definition 42 *A continuous and adapted stochastic process* $\{X_t, 0 \le t \le T\}$ *is called an Itô process if it can be expressed in the form*

$$X_t = X_0 + \int_0^t u_s dB_s + \int_0^t v_s ds, \tag{31}$$

where u *belongs to the space* $L_{a,T}$ *and* v *belongs to the space* $L_{a,T}^1$.

In differential notation we will write

$$dX_t = u_t dB_t + v_t dt.$$

Theorem 43 (Itô's formula) *Suppose that* X *is an Itô process of the form (31). Let* $f(t,x)$ *be a function twice differentiable with respect to the variable* x *and once differentiable with respect to* t*, with continuous partial derivatives* $\frac{\partial f}{\partial x}$, $\frac{\partial^2 f}{\partial x^2}$, *and* $\frac{\partial f}{\partial t}$ *(we say that* f *is of class* $C^{1,2}$*). Then, the process* $Y_t = f(t, X_t)$ *is again an Itô process with the representation*

$$\begin{aligned} Y_t &= f(0,X_0) + \int_0^t \frac{\partial f}{\partial t}(s,X_s)ds + \int_0^t \frac{\partial f}{\partial x}(s,X_s)u_s dB_s \\ &\quad + \int_0^t \frac{\partial f}{\partial x}(s,X_s)v_s ds + \frac{1}{2}\int_0^t \frac{\partial^2 f}{\partial x^2}(s,X_s)u_s^2 ds. \end{aligned}$$

1.- In differential notation Itô's formula can be written as

$$df(t,X_t) = \frac{\partial f}{\partial t}(t,X_t)dt + \frac{\partial f}{\partial x}(t,X_t)dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(s,X_s)\left(dX_t\right)^2,$$

where $(dX_t)^2$ is computed using the product rule

$\times$	dB_t	dt
dB_t	dt	0
dt	0	0

2.- The process Y_t is an Itô process with the representation

$$Y_t = Y_0 + \int_0^t \widetilde{u}_s dB_s + \int_0^t \widetilde{v}_s ds,$$

where

$$\begin{aligned} Y_0 &= f(0,X_0), \\ \widetilde{u}_t &= \frac{\partial f}{\partial x}(t,X_t)u_t, \\ \widetilde{v}_t &= \frac{\partial f}{\partial t}(t,X_t) + \frac{\partial f}{\partial x}(t,X_t)v_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t,X_t)u_t^2. \end{aligned}$$

Notice that $\widetilde{u}_t \in L_{a,T}$ and $\widetilde{v}_t \in L^1_{a,T}$ due to the continuity of X.

3.- In the particular case $u_t = 1$, $v_t = 0$, $X_0 = 0$, the process X_t is the Brownian motion B_t, and Itô's formula has the following simple version

$$\begin{aligned} f(t,B_t) &= f(0,0) + \int_0^t \frac{\partial f}{\partial x}(s,B_s)dB_s + \int_0^t \frac{\partial f}{\partial t}(s,B_s)ds \\ &\quad + \frac{1}{2}\int_0^t \frac{\partial^2 f}{\partial x^2}(s,B_s)ds. \end{aligned}$$

4.- In the particular case where f does not depend on the time we obtain the formula

$$f(X_t) = f(0) + \int_0^t f'(X_s)u_s dB_s + \int_0^t f'(X_s)v_s ds + \frac{1}{2}\int_0^t f''(X_s)u_s^2 ds.$$

Itô's formula follows from Taylor development up to the second order. We will explain the heuristic ideas that lead to Itô's formula using Taylor development. Suppose $v = 0$. Fix $t > 0$ and consider the times $t_j = \frac{jt}{n}$. Taylor's formula up to the second order gives

$$\begin{aligned} f(X_t) - f(0) &= \sum_{j=1}^{n} \left[f(X_{t_j}) - f(X_{t_{j-1}}) \right] \\ &= \sum_{j=}^{n} f'(X_{t_{j-1}})\Delta X_j + \frac{1}{2}\sum_{j=1}^{n} f''(\overline{X}_j)\left(\Delta X_j\right)^2, \end{aligned} \tag{32}$$

where $\Delta X_j = X_{t_j} - X_{t_{j-1}}$ and $\overline{X}_j$ is an intermediate value between $X_{t_{j-1}}$ and X_{t_j}.

The first summand in the above expression converges in probability to $\int_0^t f'(X_s)u_s dB_s$, whereas the second summand converges in probability to $\frac{1}{2}\int_0^t f''(X_s)u_s^2 ds$.

Some examples of application of Itô's formula:

Example 4 If $f(x) = x^2$ and $X_t = B_t$, we obtain

$$B_t^2 = 2\int_0^t B_s dB_s + t,$$

because $f'(x) = 2x$ and $f''(x) = 2$.

Example 5 If $f(x) = x^3$ and $X_t = B_t$, we obtain

$$B_t^3 = 3\int_0^t B_s^2 dB_s + 3\int_0^t B_s ds,$$

because $f'(x) = 3x^2$ and $f''(x) = 6x$. More generally, if $n \geq 2$ is a natural number,

$$B_t^n = n\int_0^t B_s^{n-1} dB_s + \frac{n(n-1)}{2}\int_0^t B_s^{n-2} ds.$$

Example 6 If $f(t,x) = e^{ax - \frac{a^2}{2}t}$, $X_t = B_t$, and $Y_t = e^{aB_t - \frac{a^2}{2}t}$, we obtain

$$Y_t = 1 + a \int_0^t Y_s dB_s$$

because

$$\frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} = 0. \tag{33}$$

This important example leads to the following remarks:

1.- If a function $f(t,x)$ of class $C^{1,2}$ satisfies the equality (33), then, the stochastic process $f(t, B_t)$ will be an indefinite integral plus a constant term. Therefore, $f(t, B_t$ will be a martingale provided f satisfies:

$$E\left[\int_0^t \left(\frac{\partial f}{\partial x}(s, B_s)\right)^2 ds\right] < \infty$$

for all $t \geq 0$.

2.- The solution of the stochastic differential equation

$$dY_t = aY_t dB_t$$

is not $Y_t = e^{aB_t}$, but $Y_t = e^{aB_t - \frac{a^2}{2}t}$.

Example 7 Suppose that $f(t)$ is a continuously differentiable function on $[0,T]$. Itô's formula applied to the function $f(t)x$ yields

$$f(t)B_t = \int_0^t f_s dB_s + \int_0^t B_s f_s' ds$$

and we obtain the integration by parts formula

$$\int_0^t f_s dB_s = f(t)B_t - \int_0^t B_s f_s' ds.$$

We are going to present a multidimensional version of Itô's formula. Suppose that $B_t = (B_t^1, B_t^2, \ldots, B_t^m)$ is an m-dimensional Brownian motion. Consider an n-dimensional Itô process of the form

$$\begin{cases} X_t^1 = X_0^1 + \int_0^t u_s^{11} dB_s^1 + \cdots + \int_0^t u_s^{1m} dB_s^m + \int_0^t v_s^1 ds \\ \vdots \\ X_t^n = X_0^n + \int_0^t u_s^{n1} dB_s^1 + \cdots + \int_0^t u_s^{nm} dB_s^m + \int_0^t v_s^n ds \end{cases}.$$

In differential notation we can write

$$dX_t^i = \sum_{l=1}^m u_t^{il} dB_t^l + v_t^i dt$$

or

$$dX_t = u_t dB_t + v_t dt.$$

where v_t is an n-dimensional process and u_t is a process with values in the set of $n \times m$ matrices and we assume that the components of u belong to $L_{a,T}$ and those of v belong to $L_{a,T}^1$.

Then, if $f : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}^p$ is a function of class $C^{1,2}$, the process $Y_t = f(t, X_t)$ is again an Itô process with the representation

$$\begin{aligned} dY_t^k &= \frac{\partial f_k}{\partial t}(t, X_t) dt + \sum_{i=1}^n \frac{\partial f_k}{\partial x_i}(t, X_t) dX_t^i \\ &\quad + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 f_k}{\partial x_i \partial x_j}(t, X_t) dX_t^i dX_t^j. \end{aligned}$$

The product of differentials $dX_t^i dX_t^j$ is computed by means of the product rules:

$$\begin{aligned} dB_t^i dB_t^j &= \begin{cases} 0 & \text{if } i \neq j \\ dt & \text{if } i = j \end{cases} \\ dB_t^i dt &= 0 \\ (dt)^2 &= 0. \end{aligned}$$

In this way we obtain

$$dX_t^i dX_t^j = \left(\sum_{k=1}^m u_t^{ik} u_t^{jk} \right) dt = (u_t u_t')_{ij} \, dt.$$

As a consequence we can deduce the following *integration by parts formula*: Suppose that X_t and Y_t are Itô processes. Then,

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s + \int_0^t dX_s dY_s.$$

5.7 Itô's Integral Representation

Consider a process u in the space $L^2_{a,T}$. We know that the indefinite stochastic integral

$$X_t = \int_0^t u_s dB_s$$

is a martingale with respect to the filtration $\mathcal{F}_t$. The aim of this subsection is to show that any square integrable martingale is of this form. We start with the integral representation of square integrable random variables.

Theorem 44 (Itô's integral representation) *Consider a random variable F in $L^2(\Omega, \mathcal{F}_T, P)$. Then, there exists a unique process u in the space $L^2_{a,T}$ such that*

$$F = E(F) + \int_0^T u_s dB_s.$$

Proof. Suppose first that F is of the form

$$F = \exp\left(\int_0^T h_s dB_s - \frac{1}{2}\int_0^T h_s^2 ds\right), \tag{34}$$

where h is a deterministic function such that $\int_0^T h_s^2 ds < \infty$. Define

$$Y_t = \exp\left(\int_0^t h_s dB_s - \frac{1}{2}\int_0^t h_s^2 ds\right).$$

By Itô's formula applied to the function $f(x) = e^x$ and the process $X_t = \int_0^t h_s dB_s - \frac{1}{2}\int_0^t h_s^2 ds$, we obtain

$$\begin{aligned} dY_t &= Y_t\left(h(t)dB_t - \frac{1}{2}h^2(t)dt\right) + \frac{1}{2}Y_t(h(t)dB_t)^2 \\ &= Y_t h(t) dB_t, \end{aligned}$$

that is,

$$Y_t = 1 + \int_0^t Y_s h(s) dB_s.$$

Hence,

$$F = Y_T = 1 + \int_0^T Y_s h(s) dB_s$$

and we get the desired representation because $E(F) = 1$,

$$\begin{aligned}
E\left(\int_0^T Y_s^2 h^2(s) ds\right) &= \int_0^T E\left(Y_s^2\right) h^2(s) ds \\
&= \int_0^T e^{\int_0^t h_s^2 ds} h^2(s) ds \\
&\leq \exp\left(\int_0^T h_s^2 ds\right) \int_0^T h_s^2 ds < \infty.
\end{aligned}$$

By linearity, the representation holds for linear combinations of exponentials of the form (34). In the general case, any random variable F in $L^2(\Omega, F_T, P)$ can be approximated in mean square by a sequence F_n of linear combinations of exponentials of the form (34). Then, we have

$$F_n = E(F_n) + \int_0^T u_s^{(n)} dB_s.$$

By the isometry of the stochastic integral

$$\begin{aligned}
E\left[(F_n - F_m)^2\right] &= E\left[\left(E(F_n - F_m) + \int_0^T \left(u_s^{(n)} - u_s^{(m)}\right) dB_s\right)^2\right] \\
&= (E(F_n - F_m))^2 + E\left[\left(\int_0^T \left(u_s^{(n)} - u_s^{(m)}\right) dB_s\right)^2\right] \\
&\geq E\left[\int_0^T \left(u_s^{(n)} - u_s^{(m)}\right)^2 ds\right].
\end{aligned}$$

The sequence F_n is a Cauchy sequence in $L^2(\Omega, F_T, P)$. Hence,

$$E\left[(F_n - F_m)^2\right] \overset{n,m\to\infty}{\longrightarrow} 0$$

and, therefore,

$$E\left[\int_0^T \left(u_s^{(n)} - u_s^{(m)}\right)^2 ds\right] \overset{n,m\to\infty}{\longrightarrow} 0.$$

This means that $u^{(n)}$ is a Cauchy sequence in $L^2([0,T]\times\Omega)$. Consequently, it will converge to a process u in $L^2([0,T]\times\Omega)$. We can show that the process u, as an element of $L^2([0,T]\times\Omega)$ has a version which is adapted, because there exists a subsequence $u^{(n)}(t,\omega)$ which converges to $u(t,\omega)$ for almost all (t,ω). So, $u \in L^2_{a,T}$. Applying again the isometry property, and taking into account that $E(F_n)$ converges to $E(F)$, we obtain

$$\begin{aligned} F &= \lim_{n\to\infty} F_n = \lim_{n\to\infty}\left(E(F_n) + \int_0^T u_s^{(n)} dB_s\right) \\ &= E(F) + \int_0^T u_s dB_s. \end{aligned}$$

Finally, uniqueness also follows from the isometry property: Suppose that $u^{(1)}$ and $u^{(2)}$ are processes in $L^2_{a,T}$ such that

$$F = E(F) + \int_0^T u_s^{(1)} dB_s = E(F) + \int_0^T u_s^{(2)} dB_s.$$

Then

$$0 = E\left[\left(\int_0^T \left(u_s^{(1)} - u_s^{(2)}\right) dB_s\right)^2\right] = E\left[\int_0^T \left(u_s^{(1)} - u_s^{(2)}\right)^2 ds\right]$$

and, hence, $u_s^{(1)}(t,\omega) = u_s^{(2)}(t,\omega)$ for almost all (t,ω). ■

Theorem 45 (Martingale representation theorem) *Suppose that $\{M_t, t \in [0,T]\}$ is a martingale with respect to the $\mathcal{F}_t$, such that $E(M_T^2) < \infty$. Then there exists a unique stochastic process u in the space$L^2_{a,T}$ such that*

$$M_t = E(M_0) + \int_0^t u_s dB_s$$

for all $t \in [0,T]$.

Proof. Applying Itô's representation theorem to the random variable $F = M_T$ we obtain a unique process $u \in L^2_T$ such that

$$M_T = E(M_T) + \int_0^T u_s dB_s = E(M_0) + \int_0^T u_s dB_s.$$

Suppose $0 \leq t \leq T$. We obtain

$$\begin{aligned} M_t &= E[M_T|\mathcal{F}_t] = E(M_0) + E[\int_0^T u_s dB_s|\mathcal{F}_t] \\ &= E(M_0) + \int_0^t u_s dB_s. \end{aligned}$$

■

Example 8 We want to find the integral representation of $F = B_T^3$. By Itô's formula

$$B_T^3 = \int_0^T 3B_t^2 dB_t + 3\int_0^T B_t dt,$$

and integrating by parts

$$\int_0^T B_t dt = TB_T - \int_0^T t dB_t = \int_0^T (T - t) dB_t.$$

So, we obtain the representation

$$B_T^3 = \int_0^T 3\left[B_t^2 + (T - t)\right] dB_t.$$

5.8 Girsanov Theorem

Girsanov theorem says that a Brownian motion with drift $B_t + \lambda t$ can be seen as a Brownian motion without drift, with a change of probability. We first discuss changes of probability by means of densities.

Suppose that $L \geq 0$ is a nonnegative random variable on a probability space $(\Omega, \mathcal{F}, P)$ such that $E(L) = 1$. Then,

$$\boxed{Q(A) = E(\mathbf{1}_A L)}$$

defines a new probability. In fact, Q is a σ-additive measure such that

$$Q(\Omega) = E(L) = 1.$$

We say that L is the *density* of Q with respect to P and we write

$$\frac{dQ}{dP} = L.$$

The expectation of a random variable X in the probability space $(\Omega, \mathcal{F}, Q)$ is computed by the formula

$$E_Q(X) = E(XL).$$

The probability Q is absolutely continuous with respect to P, that means,

$$P(A) = 0 \Longrightarrow Q(A) = 0.$$

If L is strictly positive, then the probabilities P and Q are *equivalent* (that is, mutually absolutely continuous), that means,

$$P(A) = 0 \iff Q(A) = 0.$$

The next example is a simple version of Girsanov theorem.

Example 9 Let X be a random variable with distribution $N(m, \sigma^2)$. Consider the random variable

$$L = e^{-\frac{m}{\sigma^2}X + \frac{m^2}{2\sigma^2}}.$$

which satisfies $E(L) = 1$. Suppose that Q has density L with respect to P. On the probability space $(\Omega, \mathcal{F}, Q)$ the variable X has the characteristic function:

$$\begin{aligned} E_Q(e^{itX}) &= E(e^{itX}L) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2} - \frac{mx}{\sigma^2} + \frac{m^2}{2\sigma^2} + itx} dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2} + itx} dx = e^{-\frac{\sigma^2 t^2}{2}}, \end{aligned}$$

so, X has distribution $N(0, \sigma^2)$.

Let $\{B_t, t \in [0,T]\}$ be a Brownian motion. Fix a real number λ and consider the martingale

$$\boxed{L_t = \exp\left(-\lambda B_t - \frac{\lambda^2}{2}t\right).} \tag{35}$$

We know that the process $\{L_t, t \in [0,T]\}$ is a positive martingale with expectation 1 which satisfies the linear stochastic differential equation

$$L_t = 1 - \int_0^t \lambda L_s dB_s.$$

The random variable L_T is a density in the probability space $(\Omega, \mathcal{F}_T, P)$ which defines a probability Q given by

$$Q(A) = E\left(\mathbf{1}_A L_T\right),$$

for all $A \in \mathcal{F}_T$.

The martingale property of the process L_t implies that, for any $t \in [0,T]$, in the space $(\Omega, \mathcal{F}_t, P)$, the probability Q has density L_t with respect to P. In fact, if A belongs to the σ-field $\mathcal{F}_t$ we have

$$\begin{aligned} Q(A) &= E\left(\mathbf{1}_A L_T\right) = E\left(E\left(\mathbf{1}_A L_T | \mathcal{F}_t\right)\right) \\ &= E\left(\mathbf{1}_A E\left(L_T | \mathcal{F}_t\right)\right) \\ &= E\left(\mathbf{1}_A L_t\right). \end{aligned}$$

Theorem 46 (Girsanov theorem) *In the probability space $(\Omega, \mathcal{F}_T, Q)$ the stochastic process*

$$\boxed{W_t = B_t + \lambda t,}$$

is a Brownian motion.

In order to prove Girsanov theorem we need the following technical result:

Lemma 47 *Suppose that X is a real random variable and $\mathcal{G}$ is a σ-field such that*

$$E\left(e^{iuX} | \mathcal{G}\right) = e^{-\frac{u^2\sigma^2}{2}}.$$

Then, the random variable X is independent of the σ-field $\mathcal{G}$ and it has the normal distribution $N(0, \sigma^2)$.

Proof. For any $A \in \mathcal{G}$ we have

$$E\left(\mathbf{1}_A e^{iuX}\right) = P(A) e^{-\frac{u^2\sigma^2}{2}} .$$

Thus, choosing $A = \Omega$ we obtain that the characteristic function of X is that of a normal distribution $N(0, \sigma^2)$. On the other hand, for any $A \in G$, the characteristic function of X with respect to the conditional probability given A is again that of a normal distribution $N(0, \sigma^2)$:

$$E_A\left(e^{iuX}\right) = e^{-\frac{u^2\sigma^2}{2}} .$$

That is, the law of X given A is again a normal distribution $N(0, \sigma^2)$:

$$P_A(X \leq x) = \Phi(x/\sigma) ,$$

where Φ is the distribution function of the law $N(0,1)$. Thus,

$$P\left((X \leq x) \cap A\right) = P(A)\Phi(x/\sigma) = P(A)P(X \leq x),$$

and this implies the independence of X and $\mathcal{G}$. ■

Proof of Girsanov theorem. It is enough to show that in the probability space $(\Omega, \mathcal{F}_T, Q)$, for all $s < t \leq T$ the increment $W_t - W_s$ is independent of $\mathcal{F}_s$ and has the normal distribution $N(0, t-s)$.

Taking into account the previous lemma, these properties follow from the following relation, for all $s < t$, $A \in \mathcal{F}_s$, $u \in \mathbb{R}$,

$$E_Q\left(\mathbf{1}_A e^{iu(W_t - W_s)}\right) = Q(A) e^{-\frac{u^2}{2}(t-s)}. \tag{36}$$

In order to show (36) we write

$$\begin{aligned} E_Q\left(\mathbf{1}_A e^{iu(W_t - W_s)}\right) &= E\left(\mathbf{1}_A e^{iu(W_t - W_s)} L_t\right) \\ &= E\left(\mathbf{1}_A e^{iu(B_t - B_s) + iu\lambda(t-s) - \lambda(B_t - B_s) - \frac{\lambda^2}{2}(t-s)} L_s\right) \\ &= E(\mathbf{1}_A L_s) E\left(e^{(iu-\lambda)(B_t - B_s)}\right) e^{iu\lambda(t-s) - \frac{\lambda^2}{2}(t-s)} \\ &= Q(A) e^{\frac{(iu-\lambda)^2}{2}(t-s) + iu\lambda(t-s) - \frac{\lambda^2}{2}(t-s)} \\ &= Q(A) e^{-\frac{u^2}{2}(t-s)}. \end{aligned}$$

■

Girsanov theorem admits the following generalization:

Theorem 48 *Let* $\{\theta_t, t \in [0,T]\}$ *be an adapted stochastic process such that it satisfies the following Novikov condition:*

$$E\left(\exp\left(\frac{1}{2}\int_0^T \theta_t^2 dt\right)\right) < \infty. \tag{37}$$

Then, the process

$$W_t = B_t + \int_0^t \theta_s ds$$

is a Brownian motion with respect to the probability Q *defined by*

$$Q(A) = E\left(\mathbf{1}_A L_T\right),$$

where

$$L_t = \exp\left(-\int_0^t \theta_s dB_s - \frac{1}{2}\int_0^t \theta_s^2 ds\right).$$

Notice that again L_t satisfies the linear stochastic differential equation

$$L_t = 1 - \int_0^t \theta_s L_s dB_s.$$

For the process L_t to be a density we need $E(L_t) = 1$, and condition (37) ensures this property.

As an application of Girsanov theorem we will compute the probability distribution of the hitting time of a level a by a Brownian motion with drift.

Let $\{B_t, t \geq 0\}$ be a Brownian motion. Fix a real number λ, and define

$$L_t = \exp\left(-\lambda B_t - \frac{\lambda^2}{2}t\right).$$

Let Q be the probability on each σ-field $\mathcal{F}_t$ such that for all $t > 0$

$$\frac{dQ}{dP}|_{\mathcal{F}_t} = L_t.$$

By Girsanov theorem, for all $T > 0$, in the probability space $(\Omega, \mathcal{F}_T, Q)$ the process $B_t + \lambda t := \widetilde{B}_t$ is a Brownian motion in the time interval $[0,T]$. That is, in this space B_t is a Brownian motion with drift $-\lambda t$. Set

$$\tau_a = \inf\{t \geq 0, B_t = a\},$$

where $a \neq 0$. For any $t \geq 0$ the event $\{\tau_a \leq t\}$ belongs to the σ-field $\mathcal{F}_{\tau_a \wedge t}$ because for any $s \geq 0$

$$\begin{aligned} \{\tau_a &\leq t\} \cap \{\tau_a \wedge t \leq s\} = \{\tau_a \leq t\} \cap \{\tau_a \leq s\} \\ &= \{\tau_a \leq t \wedge s\} \in \mathcal{F}_{s \wedge t} \subset \mathcal{F}_s. \end{aligned}$$

Consequently, using the Optional Stopping Theorem we obtain

$$\begin{aligned} Q\{\tau_a &\leq t\} = E\left(\mathbf{1}_{\{\tau_a \leq t\}} L_t\right) = E\left(\mathbf{1}_{\{\tau_a \leq t\}} E(L_t | \mathcal{F}_{\tau_a \wedge t})\right) \\ &= E\left(\mathbf{1}_{\{\tau_a \leq t\}} L_{\tau_a \wedge t}\right) = E\left(\mathbf{1}_{\{\tau_a \leq t\}} L_{\tau_a}\right) \\ &= E\left(\mathbf{1}_{\{\tau_a \leq t\}} e^{-\lambda a - \frac{1}{2}\lambda^2 \tau_a}\right) \\ &= \int_0^t e^{-\lambda a - \frac{1}{2}\lambda^2 s} f(s) ds, \end{aligned}$$

where f is the density of the random variable τ_a. We know that

$$f(s) = \frac{|a|}{\sqrt{2\pi s^3}} e^{-\frac{a^2}{2s}}.$$

Hence, with respect to Q the random variable τ_a has the probability density

$$\frac{|a|}{\sqrt{2\pi s^3}} e^{-\frac{(a+\lambda s)^2}{2s}}, \ s > 0.$$

Letting, $t \uparrow \infty$ we obtain

$$Q\{\tau_a < \infty\} = e^{-\lambda a} E\left(e^{-\frac{1}{2}\lambda^2 \tau_a}\right) = e^{-\lambda a - |\lambda a|}.$$

If $\lambda = 0$ (Brownian motion without drift), the probability to reach the level is one. If $-\lambda a > 0$ (the drift $-\lambda$ and the level a have the same sign) this probability is also one. If $-\lambda a < 0$ (the drift $-\lambda$ and the level a have opposite sign) this probability is $e^{-2\lambda a}$.

5.9 Application of Stochastic Calculus to Hedging and Pricing of Derivatives

The model suggested by *Black and Scholes* to describe the behavior of prices is a continuous-time model with one risky asset (a share with price S_t at time t) and a risk-less asset (with price S_t^0 at time t). We suppose that $S_t^0 = e^{rt}$ where $r > 0$ is the instantaneous interest rate and S_t is given by the geometric Brownian motion:

$$\boxed{S_t = S_0 e^{\mu t - \frac{\sigma^2}{2}t + \sigma B_t},}$$

where S_0 is the initial price, μ is the rate of growth of the price ($E(S_t) = S_0 e^{\mu t}$), and σ is called the *volatility*. We know that S_t satisfies the linear stochastic differential equation

$$dS_t = \sigma S_t dB_t + \mu S_t dt$$

or

$$\frac{dS_t}{S_t} = \sigma dB_t + \mu dt.$$

This model has the following properties:

a) The trajectories $t \to S_t$ are continuous.

b) For any $s < t$, the relative increment $\frac{S_t - S_s}{S_s}$ is independent of the σ-field generated by $\{S_u, 0 \le u \le s\}$.

c) The law of $\frac{S_t}{S_s}$ is lognormal with parameters $\left(\mu - \frac{\sigma^2}{2}\right)(t-s)$, $\sigma^2(t-s)$.

Fix a time interval $[0, T]$. A *portfolio* or *trading strategy* is a stochastic process

$$\phi = \{(\alpha_t, \beta_t), 0 \le t \le T\}$$

such that the components are measurable and adapted processes such that

$$\begin{aligned} \int_0^T |\alpha_t| \, dt &< \infty, \\ \int_0^T (\beta_t)^2 \, dt &< \infty. \end{aligned}$$

The component α_t is que quantity of non-risky and the component β_t is que quantity of shares in the portfolio. The value of the portfolio at time t is then

$$V_t(\phi) = \alpha_t e^{rt} + \beta_t S_t.$$

We say that the portfolio ϕ is *self-financing* if its value is an Itô process with differential

$$dV_t(\phi) = r\alpha_t e^{rt} dt + \beta_t dS_t.$$

The discounted prices are defined by

$$\widetilde{S}_t = e^{-rt} S_t = S_0 \exp\left((\mu - r)\, t - \frac{\sigma^2}{2} t + \sigma B_t \right).$$

Then, the discounted value of a portfolio will be

$$\widetilde{V}_t(\phi) = e^{-rt} V_t(\phi) = \alpha_t + \beta_t \widetilde{S}_t.$$

Notice that

$$\begin{aligned} d\widetilde{V}_t(\phi) &= -re^{-rt} V_t(\phi) dt + e^{-rt} dV_t(\phi) \\ &= -r\beta_t \widetilde{S}_t dt + e^{-rt} \beta_t dS_t \\ &= \beta_t d\widetilde{S}_t. \end{aligned}$$

By Girsanov theorem there exists a probability Q such that on the probability space $(\Omega, \mathcal{F}_T, Q)$ such that the process

$$W_t = B_t + \frac{\mu - r}{\sigma} t$$

is a Brownian motion. Notice that in terms of the process W_t the Black and Scholes model is

$$S_t = S_0 \exp\left(rt - \frac{\sigma^2}{2} t + \sigma W_t \right),$$

and the discounted prices form a martingale:

$$\widetilde{S}_t = e^{-rt} S_t = S_0 \exp\left(-\frac{\sigma^2}{2} t + \sigma W_t \right).$$

This means that Q is a non-risky probability.

The discounted value of a self-financing portfolio ϕ will be

$$\widetilde{V}_t(\phi) = V_0(\phi) + \int_0^t \beta_u d\widetilde{S}_u,$$

and it is a martingale with respect to Q provided

$$\int_0^T E(\beta_u^2 \widetilde{S}_u^2) du < \infty. \tag{38}$$

Notice that a self-financing portfolio satisfying (38) cannot be an arbitrage (that is, $V_0(\phi) = 0$, $V_T(\phi) \geq 0$, and $P(\ V_T(\phi) > 0) > 0$), because using the martingale property we obtain

$$E_Q\left(\widetilde{V}_T(\theta)\right) = V_0(\theta) = 0,$$

so $V_T(\phi) = 0$, Q-almost surely, which contradicts the fact that $P(\ V_T(\phi) > 0) > 0$.

Consider a derivative which produces a payoff at maturity time T equal to an $\mathcal{F}_T$-measurable nonnegative random variable h. Suppose also that $E_Q(h^2) < \infty$.

- In the case of an European call option with exercise price equal to K, we have
$$h = (S_T - K)^+.$$
- In the case of an European put option with exercise price equal to K, we have
$$h = (K - S_T)^+.$$

A self-financing portfolio ϕ satisfying (38) *replicates* the derivative if $V_T(\theta) = h$. We say that a derivative is replicable if there exists such a portfolio.

The price of a replicable derivative with payoff h at time $t \leq T$ is given by

$$\boxed{V_t(\phi) = E_Q(e^{-r(T-t)}h|\mathcal{F}_t),} \tag{39}$$

if ϕ replicates h, which follows from the martingale property of $\widetilde{V}_t(\phi)$ with respect to Q:

$$E_Q(e^{-rT}h|\mathcal{F}_t) = E_Q(\widetilde{V}_T(\phi)|\mathcal{F}_t) = \widetilde{V}_t(\phi) = e^{-rt}V_t(\phi).$$

In particular,

$$\boxed{V_0(\theta) = E_Q(e^{-rT}h).}$$

In the Black and Scholes model, any derivative satisfying $E_Q(h^2) < \infty$ is replicable. That means, the Black and Scholes model is *complete.* This is a consequence of the integral representation theorem. In fact, consider the square integrable martingale

$$M_t = E_Q\left(e^{-rT}h|\mathcal{F}_t\right).$$

We know that there exists an adapted and measurable stochastic process K_t verifying $\int_0^T E_Q(K_s^2)ds < \infty$ such that

$$M_t = M_0 + \int_0^t K_s dW_s.$$

Define the self-financing portfolio $\phi_t = (\alpha_t, \beta_t)$ by

$$\begin{aligned} \beta_t &= \frac{K_t}{\sigma \widetilde{S}_t}, \\ \alpha_t &= M_t - \beta_t \widetilde{S}_t. \end{aligned}$$

The discounted value of this portfolio is

$$\widetilde{V}_t(\phi) = \alpha_t + \beta_t \widetilde{S}_t = M_t,$$

so, its final value will be

$$V_T(\phi) = e^{rT} \widetilde{V}_T(\phi) = e^{rT} M_T = h.$$

On the other hand, it is a self-financing portfolio because

$$\begin{aligned} dV_t(\phi) &= re^{rt}\widetilde{V}_t(\phi)dt + e^{rt}d\widetilde{V}_t(\phi) \\ &= re^{rt}M_t dt + e^{rt}dM_t \\ &= re^{rt}M_t dt + e^{rt}K_t dW_t \\ &= re^{rt}\alpha_t dt - re^{rt}\beta_t \widetilde{S}_t dt + \sigma e^{rt}\beta_t \widetilde{S}_t dW_t \\ &= re^{rt}\alpha_t dt - re^{rt}\beta_t \widetilde{S}_t dt + e^{rt}\beta_t d\widetilde{S}_t \\ &= re^{rt}\alpha_t dt + \beta_t \, dS_t. \end{aligned}$$

Consider the particular case $h = g(S_T)$. The value of this derivative at time t will be

$$\begin{aligned} V_t &= E_Q\left(e^{-r(T-t)} g(S_T) | \mathcal{F}_t\right) \\ &= e^{-r(T-t)} E_Q\left(g(S_t e^{r(T-t)} e^{\sigma(W_T - W_t) - \sigma^2/2(T-t)}) | \mathcal{F}_t\right). \end{aligned}$$

Hence,

$$V_t = F(t, S_t), \tag{40}$$

where

$$F(t,x) = e^{-r(T-t)} E_Q\left(g(xe^{r(T-t)} e^{\sigma(W_T - W_t) - \sigma^2/2(T-t)}) \right). \quad (41)$$

Under general hypotheses on g (for instance, if g has linear growth, is continuous and piece-wise differentiable) which include the cases

$$\begin{aligned} g(x) &= (x-K)^+, \\ g(x) &= (K-x)^+, \end{aligned}$$

the function $F(t,x)$ is of class $C^{1,2}$. Then, applying Itô's formula to (40) we obtain

$$\begin{aligned} V_t &= V_0 + \int_0^t \sigma \frac{\partial F}{\partial x}(u, S_u) S_u dW_u + \int_0^t r \frac{\partial F}{\partial x}(u, S_u) S_u du \\ &\quad + \int_0^t \frac{\partial F}{\partial u}(u, S_u) du + \frac{1}{2} \int_0^t \frac{\partial^2 F}{\partial x^2}(u, S_u)\ \sigma^2 S_u^2 du. \end{aligned}$$

On the other hand, we know that V_t is an Itô process with the representation

$$V_t = V_0 + \int_0^t \sigma \beta_u S_u dW_u + \int_0^t rV_u du.$$

Comparing these expressions, and taking into account the uniqueness of the representation of an Itô process, we deduce the equations

$$\begin{aligned} \beta_t &= \frac{\partial F}{\partial x}(t, S_t), \\ rF(t, S_t) &= \frac{\partial F}{\partial t}(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 F}{\partial x^2}(t, S_t) \\ &\quad + rS_t \frac{\partial F}{\partial x}(t, S_t). \end{aligned}$$

The support of the probability distribution of the random variable S_t is $[0, \infty)$. Therefore, the above equalities lead to the following partial differential equation for the function $F(t,x)$

$$\begin{aligned} \frac{\partial F}{\partial t}(t,x) + rx\frac{\partial F}{\partial x}(t,x) + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}(t,x)\ \sigma^2\ x^2 &= rF(t,x), \\ F(T,x) &= g(x). \end{aligned} \quad (42)$$

The replicating portfolio is given by

$$\begin{aligned}\beta_t &= \frac{\partial F}{\partial x}(t, S_t),\\ \alpha_t &= e^{-rt}\left(F(t, S_t) - \beta_t S_t\right).\end{aligned}$$

Formula (41) can be written as

$$\begin{aligned}F(t,x) &= e^{-r(T-t)} E_Q\left(g(xe^{r(T-t)} e^{\sigma(W_T - W_t) - \sigma^2/2(T-t)})\right)\\ &= e^{-r\theta} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(xe^{r\theta - \frac{\sigma^2}{2}\theta + \sigma\sqrt{\theta}y}) e^{-y^2/2} dy,\end{aligned}$$

where $\theta = T - t$. In the particular case of an European call option with exercise price K and maturity T, $g(x) = (x-K)^+$, and we get

$$\begin{aligned}F(t,x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} \left(xe^{-\frac{\sigma^2}{2}\theta + \sigma\sqrt{\theta}y} - Ke^{-r\theta}\right)^+ dy\\ &= \boxed{x\Phi(d_+) - Ke^{-r(T-t)}\Phi(d_-),}\end{aligned}$$

where

$$\begin{aligned}d_- &= \frac{\log\frac{x}{K} + \left(r - \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}},\\ d_+ &= \frac{\log\frac{x}{K} + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}\end{aligned}$$

The price at time t of the option is

$$C_t = F(t, S_t),$$

and the replicating portfolio will be given by

$$\beta_t = \frac{\partial F}{\partial x}(t, S_t) = \Phi(d_+).$$

Exercises

5.1 Let B_t be a Brownian motion. Fix a time $t_0 \geq 0$. Show that the process

$$\left\{ \widetilde{B}_t = B_{t_0+t} - B_{t_0},\ t \geq 0 \right\}$$

is a Brownian motion.

5.2 Let B_t be a two-dimensional Brownian motion. Given $\rho > 0$, compute: $P(|B_t| < \rho)$.

5.3 Let B_t be a n-dimensional Brownian motion. Consider an orthogonal $n \times n$ matrix U (that is, $UU' = I$). Show that the process

$$\widetilde{B}_t = UB_t$$

is a Brownian motion.

5.4 Compute the mean and autocovariance function of the geometric Brownian motion. Is it a Gaussian process?

5.5 Let B_t be a Brownian motion. Find the law of B_t conditioned by B_{t_1}, B_{t_2}, and (B_{t_1}, B_{t_2}) assuming $t_1 < t < t_2$.

5.6 Check if the following processes are martingales, where B_t is a Brownian motion:

$$\begin{aligned}
X_t &= B_t^3 - 3tB_t \\
X_t &= t^2 B_t - 2\int_0^s sB_s ds \\
X_t &= e^{t/2}\cos B_t \\
X_t &= e^{t/2}\sin B_t \\
X_t &= (B_t + t)\exp(-B_t - \frac{1}{2}t) \\
X_t &= B_1(t)B_2(t).
\end{aligned}$$

In the last example, B_1 and B_2 are independent Brownian motions.

5.7 Find the stochastic integral representation on the time interval $[0, T]$ of the following random variables:

$$
\begin{aligned}
F &= B_T \\
F &= B_T^2 \\
F &= e^{B_T} \\
F &= \int_0^T B_t dt \\
F &= B_T^3 \\
F &= \sin B_T \\
F &= \int_0^T t B_t^2 dt
\end{aligned}
$$

5.8 Let $p(t,x) = 1/\sqrt{1-t}\exp(-x^2/2(1-t))$, for $0 \leq t < 1$, $x \in \mathbb{R}$, and $p(1,x) = 0$. Define $M_t = p(t, B_t)$, where $\{B_t, 0 \leq t \leq 1\}$ is a Brownian motion.

1. a) Show that for each $0 \leq t < 1$, $M_t = M_0 + \int_0^t \frac{\partial p}{\partial x}(s, B_s) dB_s$.

   b) Set $H_t = \frac{\partial p}{\partial x}(t, B_t)$. Show that $\int_0^1 H_t^2 dt < \infty$ almost surely, but $E\left(\int_0^1 H_t^2 dt\right) = \infty$.

5.9 The price of a financial asset follows the Black-Scholes model:

$$\frac{dS_t}{S_t} = 3dt + 2dB_t$$

with initial condition $S_0 = 100$. Suppose $r = 1$.

a) Give an explicit expression for S_t in terms of t and B_t.

b) Fix a maturity time T. Find a risk-less probability by means of Girsanov theorem.

c) Compute the price at time zero of a derivative whose payoff at time T is S_T^2.

6 Stochastic Differential Equations

Consider a Brownian motion $\{B_t, t \geq 0\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$. Suppose that $\{\mathcal{F}_t, t \geq 0\}$ is a filtration such that B_t is $\mathcal{F}_t$-adapted and for any $0 \leq s < t$, the increment $B_t - B_s$ is independent of $\mathcal{F}_s$.

We aim to solve *stochastic differential equations* of the form

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t \tag{43}$$

with an initial condition X_0, which is a random variable independent of the Brownian motion B_t.

The coefficients $b(t,x)$ and $\sigma(t,x)$ are called, respectively, *drift and diffusion coefficient.* If the drift vanishes, then we have (43) is the ordinary differential equation:

$$\frac{dX_t}{dt} = b(t, X_t).$$

For instance, in the linear case $b(t,x) = b(t)x$, the solution of this equation is

$$X_t = X_0 e^{\int_0^t b(s)ds}.$$

The stochastic differential equation (43) has the following heuristic interpretation. The increment $\Delta X_t = X_{t+\Delta t} - X_t$ can be approximatively decomposed into the sum of $b(t, X_t)\Delta t$ plus the term $\sigma(t, X_t)\Delta B_t$ which is interpreted as a random impulse. The approximate distribution of this increment will the the normal distribution with mean $b(t, X_t)\Delta t$ and variance $\sigma(t, X_t)^2 \Delta t$.

A formal meaning of Equation (**??**) is obtained by rewriting it in integral form, using stochastic integrals:

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s. \tag{44}$$

That is, the solution will be an Itô process $\{X_t, t \geq 0\}$. The solutions of stochastic differential equations are called *diffusion processes.*

The main result on the existence and uniqueness of solutions is the following.

Theorem 49 *Fix a time interval* $[0, T]$. *Suppose that the coefficients of Equation (??) satisfy the following Lipschitz and linear growth properties:*

$$|b(t,x) - b(t,y)| \leq D_1|x-y| \quad (45)$$

$$|\sigma(t,x) - \sigma(t,y)| \leq D_2|x-y| \quad (46)$$

$$|b(t,x)| \leq C_1(1+|x|) \quad (47)$$

$$|\sigma(t,x)| \leq C_2(1+|x|), \quad (48)$$

for all $x, y \in \mathbb{R}$, $t \in [0, T]$. *Suppose that* X_0 *is a random variable independent of the Brownian motion* $\{B_t, 0 \leq t \leq T\}$ *and such that* $E(X_0^2) < \infty$. *Then, there exists a unique continuous and adapted stochastic process* $\{X_t, t \in [0, T]\}$ *such that*

$$E\left(\int_0^T |X_s|^2 ds\right) < \infty,$$

which satisfies Equation (44).

Remarks:

1.- This result is also true in higher dimensions, when B_t is an m-dimensional Brownian motion, the process X_t is n-dimensional, and the coefficients are functions $b : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$, $\sigma : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$.

2.- The linear growth condition (47,48) ensures that the solution does not explode in the time interval $[0, T]$. For example, the deterministic differential equation

$$\frac{dX_t}{dt} = X_t^2, \ X_0 = 1, 0 \leq t \leq 1,$$

has the unique solution

$$X_t = \frac{1}{1-t}, \ 0 \leq t < 1,$$

which diverges at time $t = 1$.

3.- Lipschitz condition (45,46) ensures that the solution is unique. For example, the deterministic differential equation

$$\frac{dX_t}{dt} = 3X_t^{2/3}, \ X_0 = 0,$$

has infinitely many solutions because for each $a > 0$, the function

$$X_t = \begin{cases} 0 & \text{if } \ t \leq a \\ (t-a)^3 & \text{if } \ t > a \end{cases}$$

is a solution. In this example, the coefficient $b(x) = 3x^{2/3}$ does not satisfy the Lipschitz condition because the derivative of b is not bounded.

4.- If the coefficients $b(t,x)$ and $\sigma(t,x)$ are differentiable in the variable x, the Lipschitz condition means that the partial derivatives $\frac{\partial b}{\partial x}$ and $\frac{\partial \sigma}{\partial x}$ are bounded by the constants D_1 and D_2, respectively.

6.1 Explicit solutions of stochastic differential equations

Itô's formula allows us to find explicit solutions to some particular stochastic differential equations. Let us see some examples.

A) *Linear equations.* The geometric Brownian motion

$$X_t = X_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t}$$

solves the linear stochastic differential equation

$$dX_t = \mu X_t dt + \sigma X_t dB_t.$$

More generally, the solution of the homogeneous linear stochastic differential equation

$$dX_t = \ b(t) X_t dt + \sigma(t) X_t dB_t$$

where $b(t)$ and $\sigma(t)$ are continuous functions, is

$$\boxed{X_t = X_0 \exp\left[\int_0^t \left(b(s) - \tfrac{1}{2}\sigma^2(s)\right) ds + \int_0^t \sigma(s) dB_s\right].}$$

Application: Consider the Black-Scholes model for the prices of a financial asset, with time-dependent coefficients $\mu(t)$ y $\sigma(t) > 0$:

$$dS_t = S_t(\mu(t)dt + \sigma(t)dB_t).$$

The solution to this equation is

$$S_t = S_0 \exp\left[\int_0^t \left(\mu(s) - \frac{1}{2}\sigma^2(s)\right) ds + \int_0^t \sigma(s) dB_s\right].$$

If the interest rate $r(t)$ is also a continuous function of the time, there exists a risk-free probability under which the process

$$W_t = B_t + \int_0^t \frac{\mu(s) - r(s)}{\sigma(s)} ds$$

is a Brownian motion and the discounted prices $\widetilde{S}_t = S_t e^{-\int_0^t r(s)ds}$ are martingales:

$$\widetilde{S}_t = S_0 \exp\left(\int_0^t \sigma(s) dW_s - \frac{1}{2}\int_0^t \sigma^2(s) ds\right).$$

In this way we can deduce a generalization of the Black-Scholes formula for the price of an European call option, where the parameters σ^2 and r are replace by

$$\begin{aligned} \Sigma^2 &= \frac{1}{T-t}\int_t^T \sigma^2(s) ds, \\ R &= \frac{1}{T-t}\int_t^T r(s) ds. \end{aligned}$$

B) *Ornstein-Uhlenbeck process.* Consider the stochastic differential equation

$$\begin{aligned} dX_t &= a\,(m - X_t)\,dt + \sigma dB_t \\ X_0 &= x, \end{aligned}$$

where $a, \sigma > 0$ and m is a real number. This is a nonhomogeneous linear equation and to solve it we will make use of the method of variation of constants. The solution of the homogeneous equation

$$\begin{aligned} dx_t &= -a x_t dt \\ x_0 &= x \end{aligned}$$

is $x_t = xe^{-at}$. Then we make the change of variables $X_t = Y_t e^{-at}$, that is, $Y_t = X_t e^{at}$. The process Y_t satisfies

$$\begin{aligned} dY_t &= aX_t e^{at} dt + e^{at} dX_t \\ &= ame^{at} dt + \sigma e^{at} dB_t. \end{aligned}$$

Thus,

$$Y_t = x + m(e^{at} - 1) + \sigma \int_0^t e^{as} dB_s,$$

which implies

$$\boxed{X_t = m + (x-m)e^{-at} + \sigma e^{-at} \int_0^t e^{as} dB_s.}$$

The stochastic process X_t is Gaussian. Its mean and covariance function are:

$$\begin{aligned} E(X_t) &= m + (x-m)e^{-at}, \\ \operatorname{Cov}(X_t, X_s) &= \sigma^2 e^{-a(t+s)} E\left[\left(\int_0^t e^{ar} dB_r\right)\left(\int_0^s e^{ar} dB_r\right)\right] \\ &= \sigma^2 e^{-a(t+s)} \int_0^{t \wedge s} e^{2ar} dr \\ &= \frac{\sigma^2}{2a}\left(e^{-a|t-s|} - e^{-a(t+s)}\right). \end{aligned}$$

The law of X_t is the normal distribution

$$N(m + (x-m)e^{-at}, \frac{\sigma^2}{2a}\left(1 - e^{-2at}\right)$$

and it converges, as t tends to infinity to the normal law

$$\nu = N(m, \frac{\sigma^2}{2a}).$$

This distribution is called invariant or *stationary.* Suppose that the initial condition X_0 has distribution ν, then, for each $t > 0$ the law of X_t will be also ν. In fact,

$$X_t = m + (X_0 - m)e^{-at} + \sigma e^{-at} \int_0^t e^{as} dB_s$$

and, therefore

$$\begin{aligned} E(X_t) &= m + (E(X_0) - m)e^{-at} = m, \\ \mathrm{Var}X_t &= e^{-2at}\mathrm{Var}X_0 + \sigma^2 e^{-2at} E\left[\left(\int_0^t e^{as} dB_s\right)^2\right] = \frac{\sigma^2}{2a}. \end{aligned}$$

Examples of application of this model:

1. Vasicek model for rate interest $r(t)$:

$$dr(t) = a(b - r(t\ ))dt + \sigma\ dB_t, \tag{49}$$

where a, b are σ constants. Suppose we are working under the risk-free probability. Then the price of a bond with maturity T is given by

$$P(t,T) = E\left(e^{-\int_t^T r(s)ds} | \mathcal{F}_t\right) \ . \tag{50}$$

Formula (50) follows from the property $P(T,T) = 1$ and the fact that the discounted price of the bond $e^{-\int_0^t r(s)ds}P(t,T)$ is a martingale. Solving the stochastic differential equation (49) between the times t and s, $s \geq t$, we obtain

$$r(s) = r(t)e^{-a(s-t)} + b(1 - e^{-a(s-t)}) + \sigma e^{-as} \int_t^s e^{ar} dB_r.$$

From this expression we deduce that the law of $\int_t^T r(s)ds$ conditioned by $\mathcal{F}_t$ is normal with mean

$$(r(t) - b)\,\frac{1 - e^{-a(T-t)}}{a} \ + b(T - t) \tag{51}$$

and variance

$$-\frac{\sigma^2}{2a^3}\left(1 - e^{-a(T-t)}\right)^2 + \frac{\sigma^2}{a^2}\left((T-t) - \frac{1 - e^{-a(T-t)}}{a}\right). \tag{52}$$

This leads to the following formula for the price of bonds:

$$P(t,T) = A(t,T)e^{-B(t,T)r(t)}, \tag{53}$$

where

$$
\begin{aligned}
B(t,T) &= \frac{1-e^{-a(T-t)}}{a}, \\
A(t,T) &= \exp\left[\frac{(B(t,T)-T+t)\,(a^2b-\sigma^2/2)}{a^2} - \frac{\sigma^2 B(t,T)^2}{4a}\right].
\end{aligned}
$$

2. Black-Scholes model with stochastic volatility. We assume that the volatility $\sigma(t) = f(Y_t)$ is a function of an Ornstein-Uhlenbeck process, that is,

$$
\begin{aligned}
dS_t &= S_t(\mu dt + f(Y_t)dB_t) \\
dY_t &= a\,(m - Y_t)\,dt + \beta dW_t,
\end{aligned}
$$

where B_t and W_t Brownian motions which may be correlated:

$$E(B_t W_s) = \rho(s \wedge t).$$

C) Consider the stochastic differential equation

$$dX_t = f(t, X_t)dt + c(t)X_t dB_t,\ X_0 = x, \tag{54}$$

where $f(t,x)$ and $c(t)$ are deterministic continuous functions, such that f satisfies the required Lipschitz and linear growth conditions in the variable x. This equation can be solved by the following procedure:

a) Set $X_t = F_t Y_t$, where

$$F_t = \exp\left(\int_0^t c(s)dB_s - \frac{1}{2}\int_0^t c^2(s)ds\right),$$

is a solution to Equation (54) if $f = 0$ and $x = 1$. Then $Y_t = F_t^{-1}X_t$ satisfies

$$dY_t = F_t^{-1} f(t, F_t Y_t)dt,\ Y_0 = x. \tag{55}$$

b) Equation (55) is an ordinary differential equation with random coefficients, which can be solved by the usual methods.

For instance, suppose that $f(t,x) = f(t)x$. In that case, Equation (55) reads

$$dY_t = f(t)Y_t dt,$$

and

$$Y_t = x \exp\left(\int_0^t f(s)ds\right).$$

Hence,

$$\boxed{X_t = x \exp\left(\int_0^t f(s)ds + \int_0^t c(s)dB_s - \frac{1}{2}\int_0^t c^2(s)ds\right).}$$

D) *General linear stochastic differential equations.* Consider the equation

$$dX_t = (a(t) + b(t)X_t)\,dt + (c(t) + d(t)X_t)\,dB_t,$$

with initial condition $X_0 = x$, where a, b, c and d are continuous functions. Using the method of variation of constants, we propose a solution of the form

$$X_t = U_t V_t \tag{56}$$

where

$$dU_t = b(t)U_t dt + d(t)U_t dB_t$$

and

$$dV_t = \alpha(t)dt + \beta(t)dB_t,$$

with $U_0 = 1$ and $V_0 = x$. We know that

$$U_t = \exp\left(\int_0^t b(s)ds + \int_0^t d(s)dB_s - \frac{1}{2}\int_0^t d^2(s)ds\right).$$

On the other hand, differentiating (56) yields

$$\begin{aligned} a(t) &= U_t\alpha(t) + \beta(t)d(t)U_t \\ c(t) &= U_t\beta(t) \end{aligned}$$

that is,

$$\begin{aligned} \beta(t) &= c(t)\,U_t^{-1} \\ \alpha(t) &= [a(t) - c(t)d(t)]\,U_t^{-1}. \end{aligned}$$

Finally,

$$\boxed{X_t = U_t\left(x + \int_0^t [a(s) - c(s)d(s)]\,U_s^{-1}ds + \int_0^t c(s)\,U_s^{-1}dB_s\right)}$$

The stochastic differential equation

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s,$$

is written using the Itô's stochastic integral, and it can be transformed into a stochastic differential equation in the Stratonovich sense, using the formula that relates both types of integrals. In this way we obtain

$$X_t = X_0 + \int_0^t b(s, X_s)ds - \int_0^t \frac{1}{2}\left(\sigma\sigma'\right)(s, X_s)ds + \int_0^t \sigma(s, X_s) \circ dB_s,$$

because the Itô's decomposition of the process $\sigma(s, X_s)$ is

$$\begin{aligned} \sigma(t, X_t) &= \sigma(0, X_0) + \int_0^t \left(\sigma' b - \frac{1}{2}\sigma''\sigma^2\right)(s, X_s)ds \\ &\quad + \int_0^t \left(\sigma\sigma'\right)(s, X_s)dB_s. \end{aligned}$$

Yamada and Watanabe proved in 1971 that Lipschitz condition on the diffusion coefficient could be weakened in the following way. Suppose that the coefficients b and σ do not depend on time, the drift b is Lipschitz, and the diffusion coefficient σ satisfies the Hölder condition

$$|\sigma(x) - \sigma(y)| \leq D|x - y|^\alpha,$$

where $\alpha \geq \frac{1}{2}$. In that case, there exists a unique solution.

For example, the equation

$$\begin{cases} dX_t = |X_t|^r dB_t \\ X_0 = 0 \end{cases}$$

has a unique solution if $r \geq 1/2$.

Example 1 The Cox-Ingersoll-Ross model for interest rates:

$$dr(t) = a(b - r(t))dt + \sigma\sqrt{r(t)}dW_t.$$

,

6.2 Numerical approximations

Many stochastic differential equations cannot be solved explicitly. For this reason, it is convenient to develop numerical methods that provide approximated simulations of these equations.

Consider the stochastic differential equation

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \tag{57}$$

with initial condition $X_0 = x$.

Fix a time interval $[0, T]$ and consider the partition

$$t_i = \frac{iT}{n},\ i = 0, 1, \ldots, n.$$

The length of each subinterval is $\delta_n = \frac{T}{n}$.

Euler's method consists is the following recursive scheme:

$$X^{(n)}(t_i) = X^{(n)}(t_{i-1}) + b(X^{(n)}(t_{i-1}))\delta_n + \sigma(X^{(n)}(t_{i-1}))\Delta B_i,$$

$i = 1, \ldots, n$, where $\Delta B_i = B_{t_i} - B_{t_{i-1}}$. The initial value is $X_0^{(n)} = x$. Inside the interval (t_{i-1}, t_i) the value of the process $X^{(n)}$ is obtained by linear interpolation. The process $X^{(n)}$ is a function of the Brownian motion and we can measure the error that we make if we replace X by $X^{(n)}$:

$$e_n = \sqrt{E\left[\left(X_T - X_T^{(n)}\right)^2\right]}.$$

It holds that e_n is of the order $\delta_n^{1/2}$, that is,

$$\boxed{e_n \leq c\delta_n^{1/2}}$$

if $n \geq n_0$.

In order to simulate a trajectory of the solution using Euler's method, it suffices to simulate the values of n independent random variables $\xi_1, \ldots, \xi_n$ with distribution $N(0, 1)$, and replace ΔB_i by $\sqrt{\delta_n}\xi_i$.

Euler's method can be improved by adding a correction term. This leads to *Milstein's method.* Let us explain how this correction is obtained.

The exact value of the increment of the solution between two consecutive points of the partition is

$$X(t_i) = X(t_{i-1}) + \int_{t_{i-1}}^{t_i} b(X_s)ds + \int_{t_{i-1}}^{t_i} \sigma(X_s)dB_s. \tag{58}$$

Euler's method is based on the approximation of these exact values by

$$\begin{aligned} \int_{t_{i-1}}^{t_i} b(X_s)ds &\approx b(X(t_{i-1}))\delta_n, \\ \int_{t_{i-1}}^{t_i} \sigma(X_s)dB_s &\approx \sigma(X(t_{i-1}))\Delta B_i. \end{aligned}$$

In Milstein's method we apply Itô's formula to the processes $b(X_s)$ and $\sigma(X_s)$ that appear in (58), in order to improve the approximation. In this way we obtain

$$\begin{aligned} & X(t_i) - X(t_{i-1}) \\ = & \int_{t_{i-1}}^{t_i} \left[b(X(t_{i-1})) + \int_{t_{i-1}}^{s} \left(bb' + \frac{1}{2}b''\sigma^2 \right)(X_r)dr + \int_{t_{i-1}}^{s} (\sigma b')(X_r)dB_r \right] ds \\ & + \int_{t_{i-1}}^{t_i} \left[\sigma(X(t_{i-1})) + \int_{t_{i-1}}^{s} \left(b\sigma' + \frac{1}{2}\sigma''\sigma^2 \right)(X_r)dr + \int_{t_{i-1}}^{s} (\sigma\sigma')(X_r)dB_r \right] dB_s \\ = & \ b(X(t_{i-1}))\delta_n + \sigma(X(t_{i-1}))\Delta B_i + R_i. \end{aligned}$$

The dominant term is the residual R_i is the double stochastic integral

$$\int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^{s} (\sigma\sigma')(X_r)dB_r \right) dB_s,$$

and one can show that the other terms are of lower order and can be neglected. This double stochastic integral can also be approximated by

$$(\sigma\sigma')(X(t_{i-1})) \int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^{s} dB_r \right) dB_s.$$

The rules of Itô stochastic calculus lead to

$$\begin{aligned} \int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^{s} dB_r \right) dB_s &= \int_{t_{i-1}}^{t_i} \left(B_s - B_{t_{i-1}} \right) dB_s \\ &= \frac{1}{2}\left(B_{t_i}^2 - B_{t_{i-1}}^2 \right) - B_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}} \right) - \delta_n \\ &= \frac{1}{2}\left[(\Delta B_i)^2 - \delta_n \right]. \end{aligned}$$

In conclusion, Milstein's method consists in the following recursive scheme:

$$\begin{aligned} X^{(n)}(t_i) &= X^{(n)}(t_{i-1}) + b(X^{(n)}(t_{i-1}))\delta_n + \sigma(X^{(n)}(t_{i-1}))\ \Delta B_i \\ &\quad + \frac{1}{2}\left(\sigma\sigma'\right)(X^{(n)}(t_{i-1}))\left[(\Delta B_i)^2 - \delta_n\right]. \end{aligned}$$

One can shoe that the error e_n is of order δ_n, that is,

$$\boxed{e_n \leq c\delta_n}$$

if $n \geq n_0$.

6.3 Markov property of diffusion processes

Consider an n-dimensional diffusion process $\{X_t, t \geq 0\}$ which satisfies the stochastic differential equation

$$dX_t = \ b(t, X_t)dt + \ \sigma(t, X_t)dB_t, \tag{59}$$

where B is an m-dimensional Brownian motion and the coefficients b and σ are functions which satisfy the conditions of Theorem 49.

We will show that such a diffusion process satisfy the *Markov property*, which says that the future values of the process depend only on its present value, and not on the past values of the process (if the present value is known).

Definition 50 *We say that an n-dimensional stochastic process $\{X_t, t \geq 0\}$ is a Markov process if for every $s < t$ we have*

$$E(f(X_t)|X_r, r \leq s) = E(f(X_t)|X_s), \tag{60}$$

for any bounded Borel function f on $\mathbb{R}^n$.

In particular, property (60) says that for every Borel set $C \in \mathcal{B}_{\mathbb{R}^n}$ we have

$$P(X_t \in C|X_r, r \leq s) = P(X_t \in C|X_s).$$

The probability law of Markov processes is described by the so-called *transition probabilities:*

$$P(C, t, x, s) = P(X_t \in C|X_s = x),$$

where $0 \leq s < t$, $C \in \mathcal{B}_{\mathbb{R}^n}$ and $x \in \mathbb{R}^n$. That is, $P(\cdot, t, x, s)$ is the probability law of X_t conditioned by $X_s = x$. If this conditional distribution has a density, we will denote it by $p(y, t, x, s)$.

Therefore, the fact that X_t is a Markov process with transition probabilities $P(\cdot, t, x, s)$, means that for all $0 \leq s < t$, $C \in \mathcal{B}_{\mathbb{R}^n}$ we have

$$P(X_t \in C | X_r, r \leq s) = P(C, t, X_s, s).$$

For example, the real-valued Brownian motion B_t is a Markov process with transition probabilities given by

$$p(y, t, x, s) = \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{(x-y)^2}{2(t-s)}}.$$

In fact,

$$\begin{aligned} P(B_t & \in C | \mathcal{F}_s) = P(B_t - B_s + B_s \in C | \mathcal{F}_s) \\ & = P(B_t - B_s + x \in C)|_{x = B_s}, \end{aligned}$$

because $B_t - B_s$ is independent of $\mathcal{F}_s$. Hence, $P(\cdot, t, x, s)$ is the normal distribution $N(x, t-s)$.

We will denote by $\{X_t^{s,x}, t \geq s\}$ the solution of the stochastic differential equation (59) on the time interval $[s, \infty)$ and with initial condition $X_s^{s,x} = x$. If $s = 0$, we will write $X_t^{0,x} = X_t^x$.

One can show that there exists a continuous version (in all the parameters s, t, x) of the process

$$\{X_t^{s,x}, 0 \leq s \leq t, x \in \mathbb{R}^n\}.$$

On the other hand, for every $0 \leq s \leq t$ we have the property:

$$\boxed{X_t^x = X_t^{s, X_s^x}} \tag{61}$$

In fact, X_t^x for $t \geq s$ satisfies the stochastic differential equation

$$X_t^x = X_s^x + \int_s^t b(u, X_u^x) du + \int_s^t \sigma(u, X_u^x) dB_u.$$

On the other hand, $X_t^{s,y}$ satisfies

$$X_t^{s,y} = y + \int_s^t b(u, X_u^{s,y})du + \int_s^t \sigma(u, X_u^{s,y})dB_u$$

and substituting y by X_s^x we obtain that the processes X_t^x and X_t^{s,X_s^x} are solutions of the same equation on the time interval $[s, \infty)$ with initial condition X_s^x. The uniqueness of solutions allow us to conclude.

Theorem 51 (Markov property of diffusion processes) *Let f be a bounded Borel function on $\mathbb{R}^n$. Then, for every $0 \leq s < t$ we have*

$$E\left[f(X_t)|\mathcal{F}_s\right] = E\left[f(X_t^{s,x})\right]|_{x=X_s}.$$

Proof. Using (61) and the properties of conditional expectation we obtain

$$E\left[f(X_t)|\mathcal{F}_s\right] = E\left[f(X_t^{s,X_s})|\mathcal{F}_s\right] = E\left[f(X_t^{s,x})\right]|_{x=X_s},$$

because the process $\{X_t^{s,x}, t \geq s, x \in R^n\}$ is independent of $\mathcal{F}_s$ and the random variable X_s is $\mathcal{F}_s$-measurable. ■

This theorem says that diffusion processes possess the Markov property and their transition probabilities are given by

$$P(C, t, x, s) = P(X_t^{s,x} \in C).$$

Moreover, if a diffusion process is time homogeneous (that is, the coefficients do no depend on time), then the Markov property can be written as

$$E\left[f(X_t)|\mathcal{F}_s\right] = E\left[f(X_{t-s}^x)\right]|_{x=X_s}.$$

Example 2 Let us compute the transition probabilities of the Ornstein-Uhlenbeck process. To do this we have to solve the stochastic differential equation

$$dX_t = a\left(m - X_t\right)dt + \sigma dB_t$$

in the time interval $[s, \infty)$ with initial condition x. The solution is

$$X_t^{s,x} = m + (x - m)e^{-a(t-s)} + \sigma e^{-at}\int_s^t e^{ar}dB_r$$

and, therefore, $P(\cdot, t, x, s) = \mathcal{L}(X_t^{s,x})$ is a normal distribution with parameters

$$\begin{aligned} E\left(X_t^{s,x}\right) &= m + (x-m)e^{-a(t-s)}, \\ \mathrm{Var} X_t^{s,x} &= \sigma^2 e^{-2at} \int_s^t e^{2ar} dr = \frac{\sigma^2}{2a}(1 - e^{-2a(t-s)}). \end{aligned}$$

6.4 Feynman-Kac Formula

Consider an n-dimensional diffusion process $\{X_t, t \geq 0\}$ which satisfies the stochastic differential equation

$$dX_t = \ b(t, X_t)dt + \ \sigma(t, X_t)dB_t,$$

where B is an m-dimensional Brownian motion. Suppose that the coefficients b and σ satisfy the hypotheses of Theorem 49 and $X_0 = x_0$ is constant.

We can associate to this diffusion process a second order differential operator, depending on time, that will be denoted by A_s. This operator is called the *generator* of the diffusion process and it is given by

$$\boxed{A_s f(x) = \sum_{i=1}^n b_i(s, x)\frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^n \left(\sigma\sigma^T\right)_{i,j}(s, x)\frac{\partial^2 f}{\partial x_i \partial x_j}.}$$

In this expression f is a function on $[0, \infty) \times \mathbb{R}^n$ of class $C^{1,2}$. The matrix $\left(\sigma\sigma^T\right)(s, x)$ is the symmetric and nonnegative definite matrix given by

$$\left(\sigma\sigma^T\right)_{i,j}(s, x) = \sum_{k=1}^n \sigma_{i,k}(s, x)\sigma_{j,k}(s, x).$$

The relation between the operator A_s and the diffusion process comes from Itô's formula: Let $f(t, x)$ be a function of class $C^{1,2}$, then, $f(t, X_t)$ is an Itô process with differential

$$\begin{aligned} df(t, X_t) &= \left(\frac{\partial f}{\partial t}(t, X_t) + A_t f(t, X_t)\right)dt \\ &\quad + \sum_{i=1}^n \sum_{j=1}^m \frac{\partial f}{\partial x_i}(t, X_t)\sigma_{i,j}(t, X_t)dB_t^j. \end{aligned} \tag{62}$$

As a consequence, if

$$E\left(\int_0^t \left|\frac{\partial f}{\partial x_i}(s, X_s)\sigma_{i,j}(s, X_s)\right|^2 ds\right) < \infty \tag{63}$$

for every $t > 0$ and every i, j, then the process

$$M_t = f(t, X_t) \ - \int_0^t \left(\frac{\partial f}{\partial s} + A_s f \right)(s, X_s) ds \tag{64}$$

is a martingale. A sufficient condition for (63) is that the partial derivatives $\frac{\partial f}{\partial x^i}$ have linear growth, that is,

$$\left| \frac{\partial f}{\partial x^i}(s, x) \right| \leq C(1 + |x|^N). \tag{65}$$

In particular, ir f satisfies the equation $\frac{\partial f}{\partial t} + A_t f = 0$ and (65) holds, then $f(t, X_t)$ is a martingale.

The martingale property of this process leads to a probabilistic interpretation of the solution of a parabolic equation with fixed terminal value. Indeed, if the function $f(t, x)$ satisfies

$$\left. \begin{array}{l} \frac{\partial f}{\partial t} + A_t f = 0 \\ f(T, x) = g(x) \end{array} \right\}$$

in $[0, T] \times \mathbb{R}^n$, then

$$\boxed{f(t, x) = E(g(X_T^{t,x}))} \tag{66}$$

almost surely with respect to the law of X_t. In fact, the martingale property of the process $f(t, X_t)$ implies

$$f(t, X_t) = E(f(T, X_T)|X_t) = E(g(X_T)|X_t) = E(g(X_T^{t,x}))|_{x=X_t}.$$

Consider a continuous function $q(x)$ bounded from below. Applying again Itô's formula one can show that, if f is of class $C^{1,2}$ and satisfies (65), then the process

$$M_t = e^{-\int_0^t q(X_s)ds} f(t, X_t) - \int_0^t e^{-\int_0^s q(X_r)dr} \left(\frac{\partial f}{\partial s} + A_s f - qf \right)(s, X_s) ds$$

is a martingale. In fact,

$$dM_t = e^{-\int_0^t q(X_s)ds} \sum_{i=1}^n \sum_{j=1}^m \frac{\partial f}{\partial x^i}(t, X_t) \sigma_{i,j}(t, X_t) dB_t^j.$$

If the function f satisfies the equation $\frac{\partial f}{\partial s}+A_sf-qf=0$ then,

$$e^{-\int_0^t q(X_s)ds}f(t,X_t) \tag{67}$$

will be a martingale.

Suppose that $f(t,x)$ satisfies

$$\left.\begin{array}{c}\frac{\partial f}{\partial t}+A_tf-qf=0\\ f(T,x)=g(x)\end{array}\right\}$$

on $[0,T]\times\mathbb{R}^n$. Then,

$$\boxed{f(t,x)=E\left(e^{-\int_t^T q(X_s^{t,x})ds}g(X_T^{t,x})\right).} \tag{68}$$

In fact, the martingale property of the process (67) implies

$$f(t,X_t)=E\left(e^{-\int_t^T q(X_s)ds}f(T,X_T)|\mathcal{F}_t\right).$$

Finally, Markov property yields

$$E\left(e^{-\int_t^T q(X_s)ds}f(T,X_T)|\mathcal{F}_t\right)=E\left(e^{-\int_t^T q(X_s^{t,x})ds}g(X_T^{t,x})\right)|_{x=X_t}.$$

Formulas (66) and (68) are called the *Feynman-Kac* formulas.

Example 3 Consider Black-Scholes model, under the risk-free probability,

$$dS_t=r(t,S_t)S_tdt+\sigma(t,S_t)S_tdW_t,$$

where the interest rate and the volatility depend on time and on the asset price. Consider a derivative with payoff $g(S_T)$ at the maturity time T. The value of the derivative at time t is given by the formula

$$\boxed{V_t=E\left(e^{-\int_t^T r(s,S_s)ds}g(S_T)|\mathcal{F}_t\right).}$$

Markov property implies

$$V_t=f(t,S_t),$$

where

$$f(t,x)=E\left(e^{-\int_t^T r(s,S_s^{t,s})ds}g(S_T^{t,x})\right).$$

Then, Feynman-Kac formula says that the function $f(t,x)$ satisfies the following parabolic partial differential equation with terminal value:

$$\begin{aligned} \frac{\partial f}{\partial t} + r(t,x)x\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2(t,x)x^2\frac{\partial^2 f}{\partial x^2} - r(t,x)f &= 0, \\ f(T,x) &= g(x) \end{aligned}$$

Exercices

6.1 Show that the following processes satisfy the indicated stochastic differential equations:

(i) The process $X_t = \frac{B_t}{1+t}$ satisfies

$$\begin{aligned} dX_t &= -\frac{1}{1+t}X_t dt + \frac{1}{1+t}dB_t, \\ X_0 &= 0 \end{aligned}$$

(ii) The process $X_t = \sin B_t$, with $B_0 = a \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ satisfies

$$dX_t = -\frac{1}{2}X_t dt + \sqrt{1 - X_t^2}dB_t,$$

for $t < T = \inf\{s > 0 : B_s \notin \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

(iii) $(X_1(t), X_2(t)) = (t, e^t B_t)$ satisfies

$$\begin{bmatrix} dX_1 \\ dX_2 \end{bmatrix} = \begin{bmatrix} 1 \\ X_2 \end{bmatrix} dt + \begin{bmatrix} 0 \\ e^{X_1} \end{bmatrix} dB_t$$

(iv) $(X_1(t), X_2(t)) = (\cosh(B_t), \sinh(B_t))$ satisfies

$$\begin{bmatrix} dX_1 \\ dX_2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} dt + \begin{bmatrix} X_2 \\ X_1 \end{bmatrix} dB_t.$$

(v) $X_t = (\cos(B_t), \sin(B_t))$ satisfies

$$dX_t = \frac{1}{2}X_t dt + X_t dB_t,$$

where $M = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. The components of the process X_t satisfy

$$X_1(t)^2 + X_2(t)^2 = 1,$$

and, as a consequence, X_t can be regarded as a Brownian motion on the circle of radius 1. 1.

(vi) The process $X_t = (x^{1/3} + \frac{1}{3}B_t)^3$, $x > 0$, satisfies

$$dX_t = \frac{1}{3}X_t^{1/3}dt + X_t^{2/3}dB_t.$$

6.2 Consider an n-dimensional Brownian motion B_t and constants α_i, $i = 1, \dots, n$. Solve the stochastic differential equation

$$dX_t = rX_t dt + X_t \sum_{k=1}^{n} \alpha_k dB_k(t).$$

6.3 Solve the following stochastic differential equations:

$$\begin{aligned} dX_t &= rdt + \alpha X_t dB_t,\ X_0 = x \\ dX_t &= \frac{1}{X_t}dt + \alpha X_t dB_t,\ X_0 = x > 0 \\ dX_t &= X_t^{\gamma}dt + \alpha X_t dB_t,\ X_0 = x > 0. \end{aligned}$$

For which values of the parameters α, γ the solution explodes?

6.4 Solve the following stochastic differential equations:

(i)

$$\begin{bmatrix} dX_1 \\ dX_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} dt + \begin{bmatrix} 1 & 0 \\ 0 & X_1 \end{bmatrix} \begin{bmatrix} dB_1 \\ dB_2 \end{bmatrix}.$$

(ii)

$$\left[\begin{array}{l} dX_1(t) = X_2(t)dt + \alpha dB_1(t) \\ dX_2(t) = X_1(t)dt + \beta dB_2(t) \end{array} \right.$$

6.5 The nonlinear stochastic differential equation

$$dX_t = rX_t(K - X_t)dt + \beta X_t dB_t,\ X_0 = x > 0$$

is used to model the growth of population of size X_t in a random and crowded environment. The constant $K > 0$ is called the carrying capacity of the environment, the constant $r \in \mathbb{R}$ is a measure of the quality of the environment and $\beta \in \mathbb{R}$ is a measure of the size of the noise in the system. Show that the unique solution to this equation is given by

$$X_t = \frac{\exp\left(\left(rK - \frac{1}{2}\beta^2\right)t + \beta B_t\right)}{x^{-1} + r\int_0^t \exp\left(\left(rK - \frac{1}{2}\beta^2\right)s + \beta B_s\right)ds}.$$

6.6 Find the generator of the following diffusion processes:

a) $dX_t = \mu X_t dt + \sigma dB_t$, (Ornstein-Uhlenbeck process) μ and r are constants

b) $dX_t = rX_t dt + \alpha X_t dB_t$, (geometric Brownian motion) α and r are constants

c) $dX_t = rdt + \alpha X_t dB_t$, α and r are constants

d) $dY_t = \begin{bmatrix} dt \\ dX_t \end{bmatrix}$ whereX_t is the process introduced in a)

e) $\begin{bmatrix} dX_1 \\ dX_2 \end{bmatrix} = \begin{bmatrix} 1 \\ X_2 \end{bmatrix} dt + \begin{bmatrix} 0 \\ e^{X_1} \end{bmatrix} dB_t$

f) $\begin{bmatrix} dX_1 \\ dX_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} dt + \begin{bmatrix} 1 & 0 \\ 0 & X_1 \end{bmatrix} \begin{bmatrix} dB_1 \\ dB_2 \end{bmatrix}$

h) $X(t) = (X_1, X_2, \ldots .X_n)$,where

$$dX_k(t) = r_k X_k dt + X_k \sum_{j=1}^{n} \alpha_{kj} dB_j, \ 1 \leq k \leq n$$

6.7 Find diffusion process whose generator are:

a) $Af(x) = f\prime(x) + f''(x), \ f \in C_0^2(\mathbb{R})$

b) $Af(x) = \frac{\partial f}{\partial t} + cx\frac{\partial f}{\partial x} + \frac{1}{2}\alpha^2 x^2 \frac{\partial^2 f}{\partial x^2}, \ f \in C_0^2(\mathbb{R}^2)$

c) $Af(x_1, x_2) = \ 2x_2\frac{\partial f}{\partial x_1} + \log(1 + x_1^2 + x_2^2)\frac{\partial f}{\partial x_2} + \frac{1}{2}\left(1 + x_1^2\right)\frac{\partial^2 f}{\partial x_1^2}$
$+x_1\frac{\partial^2 f}{\partial x_1 \partial x_2} + \frac{1}{2}\ \frac{\partial^2 f}{\partial x_2^2}, \ f \in C_0^2(\mathbb{R}^2)$

6.8 Show the formulas (51), (52) y (53).

References

1. E. Cinlar: *Introduction to Stochastic Processes.* Prentice-Hall, 1975.
2. I. Karatzas and S. E. Schreve: *Brownian Motion and Stochastic Calculus.* Springer-Verlag 1991.
3. F. C. Klebaner: *Introduction to Stochastic Calculus with Applications.*
4. D. Lamberton and B. Lapeyre: *Introduction to Stochastic Calculus Applied to Finance.* Chapman and Hall, 1996.
5. T. Mikosch: *Elementary Stochastic Calculus.* World Scientific 2000.
6. J. R. Norris: *Markov Chains.* Cambridge Series in Statistical and Prob. Math., 1997.
7. B. Øksendal: *Stochastic Differential Equations.* Springer-Verlag 1998