Chapter 1

Brownian Motions and Stochastic Integrals

1.1 Introduction

Systems in many branches of science and industry are often subject to various types of noise and uncertainty. For example, let us consider a simple model of an asset price. Suppose that at time t the asset price is $x(t)$. Consider a small subsequent time interval dt, during which $x(t)$ changes to $x(t) + dx(t)$. (We use the notation $d\cdot$ for the small change in any quantity over this time interval when we intend to consider it as an infinitesimal change.) By definition, the return of per unit of the asset price at time t is $dx(t)/x(t)$. How might we model this return?

To understand the modelling more easily, suppose that the asset is a bank deposit while the bank deposit interest rate is r. So $x(t)$ is the balance of the saving account at time t. Thus the return $dx(t)/x(t)$ of the saving at time t is rdt, that is

$$\frac{dx(t)}{x(t)} = rdt$$

or

$$\frac{dx(t)}{dt} = rx(t).$$

This ordinary differential equation can be solved exactly to give exponential growth in the value of the saving, i.e.

$$x(t) = x_0 e^{r(t-t_0)},$$

where x_0 is the initial deposit of the saving account at time t_0.

However asset prices do not move as money invested in a risk-free bank. It is often stated that asset prices must move randomly because of the *efficient market hypothesis* (see Chapter 10 for more details). The most common model decomposes the return $dx(t)/x(t)$ of the asset price into two parts. One is a predictable, deterministic and anticipated return akin to the return on money invested in a risk-free bank. It gives a contribution

$$\mu dt$$

to the return $dx(t)/x(t)$, where μ is a measure of the average rate of growth of the asset price, also known as the drift. The second contribution to $dx(t)/x(t)$ models the random change in the asset price in response to external effects, such as unexpected news. There are many external effects so by the well-known central limit theorem this second contribution can be represented by a random sample drawn from a normal distribution with mean zero and adds a term

$$\sigma dB(t)$$

to $dx(t)/x(t)$. Here σ is a number called the *volatility*, which measures the standard deviation of the returns. The quantity $dB(t)$ is the sample from a normal distribution with mean zero and variance dt. In other words, $dB(t)$ is the increment of a Brownian motion $B(t)$. Putting these contributions together, we obtain

$$\frac{dx(t)}{x(t)} = \mu dt + \sigma dB(t)$$

or

$$dx(t) = \mu x(t)dt + \sigma x(t)dB(t). \tag{1.1}$$

That is, in form of integration,

$$x(t) = x_0 + \int_{t_0}^{t} \mu x(u)du + \int_{t_0}^{t} \sigma x(u)dB(u). \tag{1.2}$$

The question is: what is the integration $\int_{t_0}^{t} \sigma x(u)dB(u)$? If the Brownian motion $B(t)$ were differentiable with its derivative $\dot{B}(t) = dB(t)/dt$, then the integral would have no problem at all as it could be done as the classical Lebesgue integral $\int_{t_0}^{t} \sigma x(u)\dot{B}(u)du$. Unfortunately, we shall see that the Brownian motion $B(t)$ is nowhere differentiable. Moreover, if the Brownian motion $B(t)$ were a process of finite variation, the integral $\int_{t_0}^{t} \sigma x(u)dB(u)$ could be regarded as the Lebesgue–Stieltjes one. However, we shall see that almost every sample path of the Brownian motion has infinite variation in any finite time interval. Hence the integral can not be defined in the ordinary way. It turns out that we need to make use of the stochastic nature of the Brownian motion in order to define the integral. This integral was first defined by one of the greatest Japanese mathematicians, K. Itô in 1949 and is now known as the *Itô stochastic integral.* Equation (1.1) is a linear stochastic differential equation (SDE) and is also known as the geometric Brownian motion which is a Nobel prize winning model in economics, namely the Black–Scholes model.

Let us now take one more step to see other random fluctuation. In their model (1.1), Black and Scholes assumed that the average rate of return μ and the volatility σ are constants. However, it has been proved by many authors that both of them, especially the volatility, are random processes in many situations. There is a strong evidence to indicate that the rate μ is a Markov jump process which can be modelled by a Markov chain. Of course, when the rate jumps, the volatility will jump accordingly. Taking these jumps into account, the classical model (1.1) has recently be generalised to form a new financial model

$$dx(t) = \mu(r(t))x(t)dt + \sigma(r(t))x(t)dB(t). \tag{1.3}$$

Here $r(t)$ is a Markov chain with a finite state space $\mathbb{S} = \{1, 2, \ldots, N\}$ and μ, σ are now mappings from $\mathbb{S}$ to $[0, \infty)$. So, if the Markov chain is initially in state $r(0) = r_0 \in \mathbb{S}$, then before its first jump from r_0 to $r_1 \in \mathbb{S}$ at its first (random) jump time τ_1, the underlying asset price obeys the following geometric Brownian motion

$$dx(t) = \mu(r_0)x(t)dt + \sigma(r_0)x(t)dB(t)$$

with initial value $x(t_0) = x_0$. During this period from t_0 to τ_1, the rate and volatility are $\mu(r_0)$ and $\sigma(r_0)$, respectively. At time τ_1, the Markov chain jumps to r_1 where it will stay till the next jump at time τ_2. During the period from τ_1 to τ_2, the underlying asset price obeys another geometric Brownian motion

$$dx(t) = \mu(r_1)x(t)dt + \sigma(r_1)x(t)dB(t)$$

with initial value $x(\tau_1)$ at time τ_1, and the rate and volatility have been switched to $\mu(r_1)$ and $\sigma(r_1)$ from $\mu(r_0)$ and $\sigma(r_0)$, respectively. The underlying asset price will continue to switch from one geometric Brownian motion to other according to the Markovian switching. This type of random fluctuation, namely the Markovian switching, is one of the key features we are going to address in this book. Equation (1.3) is known as the geometric Brownian motion with Markovian switching or the hybrid geometric Brownian motion.

The main aims of this chapter are to introduce the stochastic nature of Brownian motion and to define the stochastic integral with respect to Brownian motion. To make this book self-contained, we shall briefly review the basic notations of probability theory and stochastic processes. We then give the mathematical definition of Brownian motions and introduce their important properties. Making use of these properties, we proceed to define the stochastic integral with respect to Brownian motion and establish the well-known Itô formula. To cope with the Markovian switching, which is the key feature of this book, we shall also review the essential notations and properties of Markov chains and establish the generalised Itô formula under Markovian switching.

1.2 Basic Notations of Probability Theory

Probability theory is concerned with the mathematical analysis of the intuitive notion of "chance" or "randomness," which, like all notions, is born of experience. The quantitative idea of randomness first took form at the gaming tables, and probability theory began, Pascal and Fermat (1654), as a theory of games of chance. Since then, the notion of chance has found its way into almost all branches of knowledge.

A theory becomes mathematical when it sets up a mathematical models of the phenomena with which it is concerned, that is, when, to describe the phenomena, it uses a collection of well-defined symbols and operations on the symbols. Probability theory deals with mathematical models of trials whose outcomes depend on chance. All the possible outcomes—the elementary events—are grouped together to form a set Ω with typical element $\omega \in \Omega$. Not every subset of Ω is in general an observable or interesting event. So we only group these observable or interesting events together as a family $\mathcal{F}$ of subsets of Ω. For the purpose of probability theory, such a family $\mathcal{F}$ should have the following properties:

- $\emptyset \in \mathcal{F}$, where $\emptyset$ denotes the empty set.
- If $A \in \mathcal{F}$, then its complement $A^C = \Omega - A \in \mathcal{F}$.
- If $\{A_i\}_{1 \le i < \infty} \subset \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

A family $\mathcal{F}$ with these three properties is called a *σ-algebra*. The pair

$(\Omega, \mathcal{F})$ is called a *measurable space*, and the elements of $\mathcal{F}$ is henceforth called *$\mathcal{F}$-measurable sets* instead of events. If $\mathcal{C}$ is a family of subsets of Ω, then there exists a smallest σ-algebra $\sigma(\mathcal{C})$ on Ω which contains $\mathcal{C}$. This $\sigma(\mathcal{C})$ is called the *σ-algebra generated by $\mathcal{C}$*. If $\Omega = \mathbb{R}^n$ and $\mathcal{C}$ is the family of all open sets in $\mathbb{R}^n$, then $\mathcal{B}^n = \sigma(\mathcal{C})$ is called the *Borel σ-algebra* and the elements of $\mathcal{B}^n$ are called the *Borel sets.*

A real-valued function $X : \Omega \to \mathbb{R}$ is said to be *$\mathcal{F}$-measurable* if

$$\{\omega : X(\omega) \le a\} \in \mathcal{F} \quad \text{for all } a \in \mathbb{R}.$$

The function X is also called a real-valued ($\mathcal{F}$-measurable) *random variable.* An $\mathbb{R}^n$-valued function $X(\omega) = (X_1(\omega), \cdots, X_n(\omega))^T$ is said to be *$\mathcal{F}$-measurable* if all the elements X_i are $\mathcal{F}$-measurable. Similarly, an $n \times m$-matrix-valued function $X(\omega) = (X_{ij}(\omega))_{n \times m}$ is said to be *$\mathcal{F}$-measurable* if all the elements X_{ij} are $\mathcal{F}$-measurable. The *indicator function* I_A of a set $A \subset \Omega$ is defined by

$$I_A(\omega) = \begin{cases} 1 \text{ if } \omega \in A, \\ 0 \text{ if } \omega \notin A. \end{cases}$$

The indicator function I_A is $\mathcal{F}$-measurable if and only if A is an $\mathcal{F}$-measurable set, i.e. $A \in \mathcal{F}$. If the measurable space is $(\mathbb{R}^n, \mathcal{B}^n)$, a $\mathcal{B}^n$-measurable function is then called a *Borel measurable function.* More generally, let $(\Omega', \mathcal{F}')$ be another measurable space. A mapping $X : \Omega \to \Omega'$ is said to be *$(\mathcal{F}, \mathcal{F}')$-measurable* if

$$\{\omega : X(\omega) \in A'\} \in \mathcal{F} \quad \text{for all } A' \in \mathcal{F}'.$$

The mapping X is then called an Ω'-valued $(\mathcal{F}, \mathcal{F}')$-measurable (or simply, $\mathcal{F}$-measurable) random variable.

Let $X : \Omega \to \mathbb{R}^n$ be any function. The *σ-algebra $\sigma(X)$ generated by X* is the smallest σ-algebra on Ω containing all the sets $\{\omega : X(\omega) \in U\}$, $U \subset \mathbb{R}^n$ open. That is

$$\sigma(X) = \sigma(\{\omega : X(\omega) \in U\} : U \subset \mathbb{R}^n \text{ open}).$$

Clearly, X will then be $\sigma(X)$-measurable and $\sigma(X)$ is the smallest σ-algebra with this property. If X is $\mathcal{F}$-measurable, then $\sigma(X) \subset \mathcal{F}$, i.e. X generates a sub-σ-algebra of $\mathcal{F}$. If $\{X_i : i \in I\}$ is a collection of $\mathbb{R}^n$-valued functions, define

$$\sigma(X_i : i \in I) = \sigma\Big(\bigcup_{i \in I} \sigma(X_i)\Big)$$

which is called the *σ-algebra generated by* $\{X_i : i \in I\}$. It is the smallest σ-algebra with respect to which every X_i is measurable. The following result is useful. It is a special case of a result sometimes called the Doob–Dynkin lemma.

Lemma 1.1 *If $X, Y : \Omega \to \mathbb{R}^n$ are two given functions, then Y is $\sigma(X)$-measurable if and only if there exists a Borel measurable function $g : \mathbb{R}^n \to \mathbb{R}^n$ such that $Y = g(X)$.*

A *probability measure* $\mathbb{P}$ on a measurable space $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \to [0,1]$ such that

(i) $\mathbb{P}(\Omega) = 1$;
(ii) for any disjoint sequence $\{A_i\}_{i\geq 1} \subset \mathcal{F}$ (i.e. $A_i \cap A_j = \emptyset$ if $i \neq j$)

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space.* If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, we set

$$\bar{\mathcal{F}} = \{A \subset \Omega : \exists\, B, C \in \mathcal{F} \text{ such that } B \subset A \subset C,\ \mathbb{P}(B) = \mathbb{P}(C)\}.$$

Then $\bar{\mathcal{F}}$ is a σ-algebra and is called the *completion* of $\mathcal{F}$. If $\mathcal{F} = \bar{\mathcal{F}}$, the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be *complete.* If not, one can easily extend $\mathbb{P}$ to $\bar{\mathcal{F}}$ by defining $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C)$ for $A \in \bar{\mathcal{F}}$, where $B, C \in \mathcal{F}$ with the properties that $B \subset A \subset C$ and $\mathbb{P}(B) = \mathbb{P}(C)$. Now $(\Omega, \bar{\mathcal{F}}, \mathbb{P})$ is a complete probability space, called the *completion* of $(\Omega, \mathcal{F}, \mathbb{P})$.

In the remaining of this section, we let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If X is a real-valued random variable and is *integrable* with respect to the probability measure $\mathbb{P}$, then the number

$$\mathbb{E}X = \int_{\Omega} X(\omega) d\mathbb{P}(\omega)$$

is called the *expectation* of X (with respect to $\mathbb{P}$). The number

$$Var(X) = \mathbb{E}(X - \mathbb{E}X)^2$$

is called the *variance* of X (here and in the sequel of this section we assume that all integrals concerned exist). The number $\mathbb{E}|X|^p$ $(p > 0)$ is called the pth moment of X. If Y is another real-valued random variable,

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$$

is called the *covariance* of X and Y. If $Cov(X,Y) = 0$, X and Y are said to be *uncorrelated.* For an $\mathbb{R}^n$-valued random variable $X = (X_1, \cdots, X_n)^T$, define $\mathbb{E}X = (\mathbb{E}X_1, \cdots, \mathbb{E}X_n)^T$. For an $n \times m$-matrix-valued random variable $X = (X_{ij})_{n\times m}$, define $\mathbb{E}X = (\mathbb{E}X_{ij})_{n\times m}$. If X and Y are both $\mathbb{R}^n$-valued random variables, the symmetric non-negative definite $n \times n$ matrix

$$Cov(X,Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)^T]$$

is called their *covariance matrix.*

Let X be an $\mathbb{R}^n$-valued random variable. Then X induces a probability measure μ_X on the Borel measurable space $(\mathbb{R}^n, \mathcal{B}^n)$, defined by

$$\mu_X(B) = \mathbb{P}\{\omega : X(\omega) \in B\} \quad \text{for } B \in \mathcal{B}^n,$$

and μ_X is called the *distribution* of X. The expectation of X can now be expressed as

$$\mathbb{E}X = \int_{\mathbb{R}^n} x d\mu_X(x).$$

More generally, if $g : \mathbb{R}^n \to \mathbb{R}^m$ is Borel measurable, we then have the following *transformation formula*

$$\mathbb{E}g(X) = \int_{\mathbb{R}^n} g(x) d\mu_X(x).$$

For $p \in (0, \infty)$, let $L^p = L^p(\Omega; \mathbb{R}^n)$ be the family of $\mathbb{R}^n$-valued random variables X with $\mathbb{E}|X|^p < \infty$. In L^1, we have $|\mathbb{E}X| \le \mathbb{E}|X|$. Moreover, the following three inequalities are very useful:

- **Hölder's inequality**

$$|\mathbb{E}(X^T Y)| \le (\mathbb{E}|X|^p)^{1/p} \, (\mathbb{E}|Y|^q)^{1/q}$$

  if $p > 1, \ 1/p + 1/q = 1, \ X \in L^p, \ Y \in L^q$;
- **Minkovski's inequality**

$$(\mathbb{E}|X + Y|^p)^{1/p} \le (\mathbb{E}|X|^p)^{1/p} + (\mathbb{E}|Y|^p)^{1/p}$$

  if $p > 1, X, Y \in L^p$;
- **Chebyshev's inequality**

$$\mathbb{P}\{\omega : |X(\omega)| \ge c\} \le c^{-p}\mathbb{E}|X|^p$$

  if $c > 0, \ p > 0, \ X \in L^p$.

A simple application of Hölder's inequality implies

$$(\mathbb{E}|X|^r)^{1/r} \le (\mathbb{E}|X|^p)^{1/p}$$

if $0 < r < p < \infty$, $X \in L^p$.

Let X and X_k, $k \ge 1$, be $\mathbb{R}^n$-valued random variables. The following four convergence concepts are very important:

(a) If there exists a $\mathbb{P}$-null set $\Omega_0 \in \mathcal{F}$ such that for every $\omega \notin \Omega_0$, the sequence $\{X_k(\omega)\}$ converges to $X(\omega)$ in the usual sense in $\mathbb{R}^n$, then $\{X_k\}$ is said to converge to X *almost surely* or *with probability 1*, and we write $\lim_{k\to\infty} X_k = X$ a.s.
(b) If for every $\varepsilon > 0$, $\mathbb{P}\{\omega : |X_k(\omega) - X(\omega)| > \varepsilon\} \to 0$ as $k \to \infty$, then $\{X_k\}$ is said to converge to X *stochastically* or *in probability.*
(c) If X_k and X belong to L^p and $\mathbb{E}|X_k - X|^p \to 0$, then $\{X_k\}$ is said to converge to X *in pth moment* or *in* L^p.
(d) If for every real-valued continuous bounded function g defined on $\mathbb{R}^n$, $\lim_{k\to\infty} \mathbb{E}g(X_k) = \mathbb{E}g(X)$, then $\{X_k\}$ is said to converge to X *in distribution.*

These convergence concepts have the following relationship:

$$\begin{array}{rc} & \text{convergence in } L^p \\ & \Downarrow \\ \text{a.s. convergence} \Rightarrow & \text{convergence in probability} \\ & \Downarrow \\ & \text{convergence in distribution} \end{array}$$

Furthermore, a sequence converges in probability if and only if every subsequence of it contains an almost surely convergent subsequence. A sufficient condition for $\lim_{k\to\infty} X_k = X$ a.s. is the condition

$$\sum_{k=1}^{\infty} \mathbb{E}|X_k - X|^p < \infty \quad \text{for some } p > 0.$$

Let us now state two very important integration convergence theorems.

Theorem 1.1 *(Monotonic convergence theorem) If $\{X_k\}$ is an increasing sequence of non-negative random variables, then*

$$\lim_{k\to\infty} \mathbb{E}X_k = \mathbb{E}\Big(\lim_{k\to\infty} X_k\Big).$$

Theorem 1.2 ***(Dominated convergence theorem)*** *Let $p \geq 1$, $\{X_k\} \subset L^p(\Omega;\mathbb{R}^n)$ and $Y \in L^p(\Omega;\mathbb{R})$. Assume that $|X_k| \leq Y$ a.s. and $\{X_k\}$ converges to X in probability. Then $X \in L^p(\Omega;\mathbb{R}^n)$, $\{X_k\}$ converges to X in L^p, and*

$$\lim_{k\to\infty} \mathbb{E}X_k = \mathbb{E}X.$$

When Y is bounded, this theorem is also referred to as the ***bounded convergence theorem.***

Two sets $A, B \in \mathcal{F}$ are said to be *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Three sets $A, B, C \in \mathcal{F}$ are said to be *independent* if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \quad \mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C),$$

$$\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C) \quad \text{and} \quad \mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

Let I be an index set. A collection of sets $\{A_i : i \in I\} \subset \mathcal{F}$ is said to be *independent* if

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

holds for arbitrary distinct finite indices $i_1, \cdots, i_k \in I$. Two sub-σ-algebras $\mathcal{F}_1$ and $\mathcal{F}_2$ of $\mathcal{F}$ are said to be *independent* if

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2) \quad \text{for all } A_1 \in \mathcal{F}_1,\ A_2 \in \mathcal{F}_2.$$

A collection of sub-σ-algebras $\{\mathcal{F}_i : i \in I\}$ is said to be *independent* if for arbitrary distinct finite indices $i_1, \cdots, i_k \in I$,

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

holds for all $A_{i_1} \in \mathcal{F}_{i_1}, \cdots, A_{i_k} \in \mathcal{F}_{i_k}$. A family of random variables $\{X_i : i \in I\}$ (whose ranges may differ for different values of the index) is said to be *independent* if the σ-algebras $\sigma(X_i)$, $i \in I$ generated by them are independent. For example, two random variables $X : \Omega \to \mathbb{R}^n$ and $Y : \Omega \to \mathbb{R}^m$ are independent if and only if

$$\mathbb{P}\{\omega : X(\omega) \in A,\ Y(\omega) \in B\} = \mathbb{P}\{\omega : X(\omega) \in A\}\ \mathbb{P}\{\omega : Y(\omega) \in B\}$$

holds for all $A \in \mathcal{B}^n$, $B \in \mathcal{B}^m$. If X and Y are two independent real-valued integrable random variables, then XY is also integrable and

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

If $X, Y \in L^2(\Omega; \mathbb{R})$ are uncorrelated, then

$$Var(X+Y) = Var(X) + Var(Y).$$

If the X and Y are independent, they are uncorrelated; but the converse is not true. However, if (X, Y) has a joint normal distribution, then X and Y are independent if and only if they are uncorrelated.

Let $\{A_k\}$ be a sequence of sets in $\mathcal{F}$. The set of all those points which belong to almost all A_k (all but any finite number) is called the *inferior limit* of A_k, and is denoted by $\liminf_{k\to\infty} A_k$. Clearly,

$$\liminf_{k\to\infty} A_k = \bigcup_{i=1}^{\infty} \bigcap_{k=i}^{\infty} A_k.$$

The set of all those points which belong to infinitely many A_k is called the *superior limit* of A_k and is denoted by $\limsup_{k\to\infty} A_k$. It is easy to see

$$\limsup_{k\to\infty} A_k = \bigcap_{i=1}^{\infty} \bigcup_{k=i}^{\infty} A_k.$$

Moreover,

$$\liminf_{k\to\infty} A_k \subset \limsup_{k\to\infty} A_k.$$

With regard to their probabilities, we have the following well-known *Borel–Cantelli lemma.*

Lemma 1.2 ***(Borel–Cantelli's lemma)***
(1) If $\{A_k\} \subset \mathcal{F}$ and $\sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty$, then

$$\mathbb{P}\Big(\limsup_{k\to\infty} A_k\Big) = 0.$$

That is, there exists a set $\Omega_1 \in \mathcal{F}$ with $\mathbb{P}(\Omega_1) = 1$ and an integer-valued random variable k_1 such that for every $\omega \in \Omega_1$ we have $\omega \notin A_k$ whenever $k \ge k_1(\omega)$.

(2) If the sequence $\{A_k\} \subset \mathcal{F}$ is independent and $\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \infty$, then

$$\mathbb{P}\Big(\limsup_{k\to\infty} A_k\Big) = 1.$$

That is, there exists a set $\Omega_2 \in \mathcal{F}$ with $\mathbb{P}(\Omega_2) = 1$ such that for every $\omega \in \Omega_2$, there exists a sub-sequence $\{A_{k_i}\}$ such that the ω belongs to every A_{k_i}.

Let $A, B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$. The *conditional probability of A under condition B* is

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

However, we frequently encounter a family of conditions so we need the more general concept of *conditional expectation*. Let $X \in L^1(\Omega; \mathbb{R})$. Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra of $\mathcal{F}$ so $(\Omega, \mathcal{G})$ is a measurable space. In general, X is not $\mathcal{G}$-measurable. We now seek an integrable $\mathcal{G}$-measurable random variable Y such that it has the same values as X on the average in the sense that

$$\mathbb{E}(I_G Y) = \mathbb{E}(I_G X) \quad \text{i.e.} \quad \int_G Y(\omega) d\mathbb{P}(\omega) = \int_G X(\omega) d\mathbb{P}(\omega) \quad \forall G \in \mathcal{G}.$$

By the Radon–Nikodym theorem, there exists one such Y, almost surely unique. It is called the *conditional expectation of X under the condition $\mathcal{G}$*, and we write

$$Y = \mathbb{E}(X|\mathcal{G}).$$

If $\mathcal{G}$ is the σ-algebra generated by a random variable Y, we write

$$\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X|Y).$$

As an example, consider a collection of sets $\{A_k\} \subset \mathcal{F}$ with

$$\bigcup_k A_k = \Omega, \quad \mathbb{P}(A_k) > 0, \quad A_k \cap A_j = \emptyset \text{ if } k \neq j.$$

Let $\mathcal{G} = \sigma(\{A_k\})$, i.e. $\mathcal{G}$ is generated by $\{A_k\}$. Then $\mathbb{E}(X|\mathcal{G})$ is a *step function* on Ω given by

$$\mathbb{E}(X|\mathcal{G}) = \sum_k \frac{I_{A_k} \mathbb{E}(I_{A_k} X)}{\mathbb{P}(A_k)}.$$

In other words, if $\omega \in A_k$,

$$\mathbb{E}(X|\mathcal{G})(\omega) = \frac{\mathbb{E}(I_{A_k} X)}{\mathbb{P}(A_k)}.$$

It follows from the definition that

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X)$$

and

$$|\mathbb{E}(X|\mathcal{G})| \le \mathbb{E}(|X|\ |\mathcal{G}) \quad a.s.$$

Other important properties of the conditional expectation are as follows (all the equalities and inequalities shown hold almost surely):

- $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}X$ if $\mathcal{G} = \{\emptyset, \Omega\}$;
- $\mathbb{E}(X|\mathcal{G}) \ge 0$ if $X \ge 0$;
- $\mathbb{E}(X|\mathcal{G}) = X$ if X is $\mathcal{G}$-measurable;
- $\mathbb{E}(X|\mathcal{G}) = c$ if $X = c = const.$;
- $\mathbb{E}(aX + bY|\mathcal{G}) = a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G})$ if $a, b \in \mathbb{R}$;
- $\mathbb{E}(X|\mathcal{G}) \le \mathbb{E}(Y|\mathcal{G})$ if $X \le Y$;
- $\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G})$ if X is $\mathcal{G}$-measurable, in particular, $\mathbb{E}(\mathbb{E}(X|\mathcal{G})\ Y|\mathcal{G}) = \mathbb{E}(X|\mathcal{G})\ \mathbb{E}(Y|\mathcal{G})$;
- $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}X$ if $\sigma(X)$ and $\mathcal{G}$ are independent, in particular, $\mathbb{E}(X|Y) = \mathbb{E}X$ if X, Y are independent;
- $\mathbb{E}(\mathbb{E}(X|\mathcal{G}_2)|\mathcal{G}_1) = \mathbb{E}(X|\mathcal{G}_1)$ if $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$.

Finally, if $X = (X_1, \cdots, X_n)^T \in L^1(\Omega; \mathbb{R}^n)$, its *conditional expectation under* $\mathcal{G}$ is defined as

$$\mathbb{E}(X|\mathcal{G}) = (\mathbb{E}(X_1|\mathcal{G}), \cdots, \mathbb{E}(X_n|\mathcal{G}))^T.$$

1.3 Stochastic Processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *filtration* is a family $\{\mathcal{F}_t\}_{t\ge0}$ of increasing sub-σ-algebras of $\mathcal{F}$ (i.e. $\mathcal{F}_t \subset \mathcal{F}_s \subset \mathcal{F}$ for all $0 \le t < s < \infty$). The filtration is said to be *right continuous* if $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$ for all $t \ge 0$. When the probability space is complete, the filtration is said to satisfy the *usual conditions* if it is right continuous and $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets.

From now on, unless otherwise specified, we shall always work on a given complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ *with a filtration* $\{\mathcal{F}_t\}_{t\ge0}$ *satisfying the usual conditions. We also define* $\mathcal{F}_\infty = \sigma(\bigcup_{t\ge0} \mathcal{F}_t)$, *i.e. the* σ*-algebra generated by* $\bigcup_{t\ge0} \mathcal{F}_t$.

A family $\{X_t\}_{t\in I}$ of $\mathbb{R}^n$-valued random variables is called a *stochastic process* with *parameter set* (or *index set*) I and *state space* $\mathbb{R}^n$. The parameter set I is usually (as in this book) the half line $\mathbb{R}_+ = [0, \infty)$, but it may also be an interval $[a, b]$, the non-negative integers or even subsets of

$\mathbb{R}^n$. Note that for each fixed $t \in I$ we have a random variable

$$\Omega \ni \omega \to X_t(\omega) \in \mathbb{R}^n.$$

On the other hand, for each fixed $\omega \in \Omega$ we have a function

$$I \ni t \to X_t(\omega) \in \mathbb{R}^n$$

which is called a *sample path* of the process, and we shall write $X_.(\omega)$ for the path. Sometimes it is convenient to write $X(t, \omega)$ instead of $X_t(\omega)$, and the stochastic process may be regarded as a function of two variables (t, ω) from $I \times \Omega$ to $\mathbb{R}^n$. Similarly, one can define matrix-valued stochastic processes *etc.* We often write a stochastic process $\{X_t\}_{t \geq 0}$ as $\{X_t\}$, X_t or $X(t)$.

Let $\{X_t\}_{t \geq 0}$ be an $\mathbb{R}^n$-valued stochastic process. It is said to be *continuous* (resp. *right continuous*, *left continuous*) if for almost all $\omega \in \Omega$ function $X_t(\omega)$ is continuous (resp. right continuous, left continuous) on $t \geq 0$. It is said to be *cadlag* (*right continuous and left limit*) if it is right continuous and for almost all $\omega \in \Omega$ the left limit $\lim_{s \uparrow t} X_s(\omega)$ exists and is finite for all $t > 0$. It is said to be *integrable* if for every $t \geq 0$, X_t is an integrable random variable. It is said to be $\{\mathcal{F}_t\}$*-adapted* (or simply, *adapted*) if for every t, X_t is $\mathcal{F}_t$-measurable. It is said to be *measurable* if the stochastic process regarded as a function of two variables (t, ω) from $\mathbb{R}_+ \times \Omega$ to $\mathbb{R}^n$ is $\mathcal{B}(\mathbb{R}_+) \times \mathcal{F}$-measurable, where $\mathcal{B}(\mathbb{R}_+)$ is the family of all Borel sub-sets of $\mathbb{R}_+$. The stochastic process is said to be *progressively measurable* or *progressive* if for every $T \geq 0$, $\{X_t\}_{0 \leq t \leq T}$ regarded as a function of (t, ω) from $[0, T] \times \Omega$ to $\mathbb{R}^n$ is $\mathcal{B}([0, T]) \times \mathcal{F}_T$-measurable, where $\mathcal{B}([0, T])$ is the family of all Borel sub-sets of $[0, T]$. Let $\mathcal{O}$ (resp. $\mathcal{P}$) denote the smallest σ-algebra on $\mathbb{R}_+ \times \Omega$ with respect to which every cadlag adapted process (resp. left continuous process) is a measurable function of (t, ω). A stochastic process is said to be *optional* (resp. *predictable*) if the process regarded as a function of (t, ω) is $\mathcal{O}$-measurable (resp. $\mathcal{P}$-measurable). A real-valued stochastic process $\{A_t\}_{t \geq 0}$ is called an *increasing process* if for almost all $\omega \in \Omega$, $A_t(\omega)$ is non-negative nondecreasing right continuous on $t \geq 0$. It is called a *process of finite variation* if $A_t = \bar{A}_t - \hat{A}_t$ with $\{\bar{A}_t\}$ and $\{\hat{A}_t\}$ both increasing processes. It is obvious that the processes of finite variation are cadlag. Hence the adapted processes of finite variation are optional.

The relations among the various stochastic processes are summarised as follows:

continuous adapted		continuous adapted		adapted increasing
$\Downarrow$		$\Downarrow$		$\Downarrow$
left continuous adapted		cadlag adapted	$\Leftarrow$	adapted finite variation
$\Downarrow$		$\Downarrow$		
predictable	$\Rightarrow$	optional		
		$\Downarrow$		
		progressive	$\Rightarrow$	adapted
		$\Downarrow$		
		measurable		

Let $\{X_t\}_{t\geq 0}$ be a stochastic process. Another stochastic process $\{Y_t\}_{t\geq 0}$ is called a *version* or *modification* of $\{X_t\}$ if for all $t \geq 0$, $X_t = Y_t$ a.s. (i.e. $\mathbb{P}\{\omega : X_t(\omega) = Y_t(\omega)\} = 1$). Two stochastic processes $\{X_t\}_{t\geq 0}$ and $\{Y_t\}_{t\geq 0}$ are said to be *indistinguishable* if for almost all $\omega \in \Omega$, $X_t(\omega) = Y_t(\omega)$ for all $t \geq 0$ (i.e. $\mathbb{P}\{\omega : X_t(\omega) = Y_t(\omega) \text{ for all } t \geq 0\} = 1$).

A random variable $\tau : \Omega \to [0, \infty]$ (it may take the value ∞) is called an $\{\mathcal{F}_t\}$*-stopping time* (or simply, *stopping time*) if $\{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t$ for any $t \geq 0$. Let τ and ρ be two stopping times with $\tau \leq \rho$ a.s. We define

$$[[\tau,\ \rho[[= \{(t, \omega) \in \mathbb{R}_+ \times \Omega : \tau(\omega) \leq t < \rho(\omega)\}$$

and call it a *stochastic interval*. Similarly, we can define stochastic intervals $[[\tau,\ \rho]]$, $]]\tau,\ \rho]]$ and $]]\tau,\ \rho[[$. If τ is a stopping time, define

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t \text{ for all } t \geq 0\}$$

which is a sub-σ-algebra of $\mathcal{F}$. If τ and ρ are two stopping times with $\tau \leq \rho$ a.s., then $\mathcal{F}_\tau \subset \mathcal{F}_\rho$. The following two theorems are useful.

Theorem 1.3 *If $\{X_t\}_{t\geq 0}$ is a progressively measurable process and τ is a stopping time, then $X_\tau I_{\{\tau<\infty\}}$ is $\mathcal{F}_\tau$-measurable. In particular, if τ is finite, then X_τ is $\mathcal{F}_\tau$-measurable.*

Theorem 1.4 *Let $\{X_t\}_{t\geq 0}$ be an $\mathbb{R}^n$-valued cadlag $\{\mathcal{F}_t\}$-adapted process, and D an open subset of $\mathbb{R}^n$. Define*

$$\tau = \inf\{t \geq 0 : X_t \notin D\},$$

where we use the convention $\inf \emptyset = \infty$. *Then τ is an $\{\mathcal{F}_t\}$-stopping time, and is called the first exit time from D. Moreover, if ρ is a stopping time, then*

$$\theta = \inf\{t \geq \rho : X_t \notin D\}$$

is also an $\{\mathcal{F}_t\}$-stopping time, and is called the first exit time from D after ρ.

An $\mathbb{R}^n$-valued $\{\mathcal{F}_t\}$-adapted integrable process $\{M_t\}_{t\geq 0}$ is called a *martingale with respect to $\{\mathcal{F}_t\}$* (or simply, *martingale*) if

$$\mathbb{E}(M_t|\mathcal{F}_s) = M_s \quad \text{a.s. for all } 0 \leq s < t < \infty.$$

It should be pointed out that every martingale has a cadlag modification since we always assume that the filtration $\{\mathcal{F}_t\}$ is right continuous. Therefore we can always assume that any martingale is cadlag in the sequel. If $X = \{X_t\}_{t\geq 0}$ is a progressively measurable process and τ is a stopping time, then $X^\tau = \{X_{\tau\wedge t}\}_{t\geq 0}$ is called a *stopped process* of X. The following is the well-known Doob martingale stopping theorem.

Theorem 1.5 *Let $\{M_t\}_{t\geq 0}$ be an $\mathbb{R}^n$-valued martingale with respect to $\{\mathcal{F}_t\}$, and let θ, ρ be two finite stopping times. Then*

$$\mathbb{E}(M_\theta|\mathcal{F}_\rho) = M_{\theta\wedge\rho} \quad a.s.$$

In particular, if τ is a stopping time, then

$$\mathbb{E}(M_{\tau\wedge t}|\mathcal{F}_s) = M_{\tau\wedge s} \quad a.s.$$

holds for all $0 \leq s < t < \infty$. That is, the stopped process $M^\tau = \{M_{\tau\wedge t}\}$ is still a martingale with respect to the same filtration $\{\mathcal{F}_t\}$.

A stochastic process $X = \{X_t\}_{t\geq 0}$ is called *square-integrable* if $\mathbb{E}|X_t|^2 < \infty$ for every $t \geq 0$. If $M = \{M_t\}_{t\geq 0}$ is a real-valued square-integrable continuous martingale, then there exists a unique continuous integrable adapted increasing process denoted by $\{\langle M, M\rangle_t\}$ such that $\{M_t^2 - \langle M, M\rangle_t\}$ is a continuous martingale vanishing at $t = 0$. The process $\{\langle M, M\rangle_t\}$ is called the *quadratic variation* of M. In particular, for any finite stopping time τ,

$$\mathbb{E}M_\tau^2 = \mathbb{E}\langle M, M\rangle_\tau.$$

If $N = \{N_t\}_{t\geq 0}$ is another real-valued square-integrable continuous martingale, we define

$$\langle M, N\rangle_t = \frac{1}{2}\Big(\langle M+N, M+N\rangle_t - \langle M, M\rangle_t - \langle N, N\rangle_t\Big),$$

and call $\{\langle M, N\rangle_t\}$ the *joint quadratic variation* of M and N. It is useful to know that $\{\langle M, N\rangle_t\}$ is the unique continuous integrable adapted process of finite variation such that $\{M_tN_t - \langle M, N\rangle_t\}$ is a continuous martingale vanishing at $t = 0$. In particular, for any finite stopping time τ,

$$\mathbb{E}M_\tau N_\tau = \mathbb{E}\langle M, N\rangle_\tau.$$

A right continuous adapted process $M = \{M_t\}_{t\geq 0}$ is called a *local martingale* if there exists a nondecreasing sequence $\{\tau_k\}_{k\geq 1}$ of stopping times with $\tau_k \uparrow \infty$ a.s. such that every $\{M_{\tau_k\wedge t} - M_0\}_{t\geq 0}$ is a martingale. Every martingale is a local martingale (by Theorem 1.5), but the converse is not true. If $M = \{M_t\}_{t\geq 0}$ and $N = \{N_t\}_{t\geq 0}$ are two real-valued continuous local martingales, their *joint quadratic variation* $\{\langle M, N\rangle\}_{t\geq 0}$ is the unique continuous adapted process of finite variation such that $\{M_tN_t - \langle M, N\rangle_t\}_{t\geq 0}$ is a continuous local martingale vanishing at $t = 0$. When $M = N$, $\{\langle M, M\rangle\}_{t\geq 0}$ is called the *quadratic variation* of M. The following result is the useful strong law of large numbers.

Theorem 1.6 ***(Strong law of large numbers)*** *Let $M = \{M_t\}_{t\geq 0}$ be a real-valued continuous local martingale vanishing at $t = 0$. Then*

$$\lim_{t\to\infty} \langle M, M\rangle_t = \infty \quad a.s. \quad \Rightarrow \quad \lim_{t\to\infty} \frac{M_t}{\langle M, M\rangle_t} = 0 \quad a.s.$$

and also

$$\limsup_{t\to\infty} \frac{\langle M, M\rangle_t}{t} < \infty \quad a.s. \quad \Rightarrow \quad \lim_{t\to\infty} \frac{M_t}{t} = 0 \quad a.s.$$

More generally, if $A = \{A_t\}_{t\geq 0}$ is a continuous adapted increasing process such that

$$\lim_{t\to\infty} A_t = \infty \quad and \quad \int_0^\infty d\langle M, M\rangle_t (1 + A_t)^2 < \infty \quad a.s.$$

then

$$\lim_{t\to\infty} \frac{M_t}{A_t} = 0 \quad a.s.$$

A real-valued $\{\mathcal{F}_t\}$-adapted integrable process $\{M_t\}_{t\geq 0}$ is called a *supermartingale (with respect to $\{\mathcal{F}_t\}$)* if

$$\mathbb{E}(M_t|\mathcal{F}_s) \leq M_s \quad \text{a.s. for all } 0 \leq s < t < \infty.$$

It is called a *submartingale (with respect to $\{\mathcal{F}_t\}$)* if we replace the sign $\leq$ in the last formula with $\geq$. Clearly, $\{M_t\}$ is submartingale if and only if $\{-M_t\}$ is supermartingale. For a real-valued martingale $\{M_t\}$, $\{M_t^+ := \max(M_t, 0)\}$ and $\{M_t^- := \max(0, -M_t)\}$ are submartingales. For a supermartingale (resp. submartingale), $\mathbb{E}M_t$ is monotonically decreasing (resp. increasing). Moreover, if $p \geq 1$ and $\{M_t\}$ is an $\mathbb{R}^n$-valued martingale such that $M_t \in L^p(\Omega; \mathbb{R}^n)$, then $\{|M_t|^p\}$ is a non-negative submartingale. Moreover, Doob's stopping Theorem 1.5 holds for supermartingales and submartingales as well.

Theorem 1.7 ***(Doob's martingale convergence theorem)***
(i) Let $\{M_t\}_{t\geq 0}$ be a real-valued right-continuous supermartingale. If

$$\sup_{0\leq t<\infty} \mathbb{E}M_t^- < \infty,$$

then M_t converges almost surely to a random variable $M_\infty \in L^1(\Omega; \mathbb{R})$. In particular, this holds if M_t is non-negative.

(ii) Let $\{M_t\}_{t\geq 0}$ be a real-valued right-continuous supermartingale. Then $\{M_t\}_{t\geq 0}$ is uniformly integrable, i.e.

$$\lim_{c\to\infty} \Big[\sup_{t\geq 0} \mathbb{E}\Big(I_{\{|M_t|\geq c\}} |M_t| \Big) \Big] = 0$$

if and only if there exists a random variable $M_\infty \in L^1(\Omega; \mathbb{R})$ such that $M_t \to M_\infty$ a.s. and in L^1 as well.

(iii) Let $X \in L^1(\Omega; \mathbb{R})$. Then

$$\mathbb{E}(X|\mathcal{F}_t) \to \mathbb{E}(X|\mathcal{F}_\infty) \quad \text{as } t \to \infty$$

a.s. and in L^1 as well.

Theorem 1.8 ***(Supermartingale inequalities)*** *Let $\{M_t\}_{t\geq 0}$ be a real-valued supermartingale. Let $[a, b]$ be a bounded interval in $\mathbb{R}_+$. Then*

$$c\, P\Big\{\omega : \sup_{a\leq t\leq b} M_t(\omega) \geq c\Big\} \leq \mathbb{E}M_a + \mathbb{E}M_b^-$$

and

$$c\, P\Big\{\omega : \inf_{a\leq t\leq b} M_t(\omega) \leq -c\Big\} \leq \mathbb{E}M_b^-$$

hold for all $c > 0$.

For submartingales we have the following well-known Doob inequality.

Theorem 1.9 ***(Doob's submartingale inequalities)*** *Let $p > 1$. Let $\{M_t\}_{t\geq 0}$ be a real-valued non-negative submartingale such that $M_t \in$*

$L^p(\Omega;\mathbb{R})$. Let $[a,b]$ be a bounded interval in $\mathbb{R}_+$. Then

$$\mathbb{E}\Big(\sup_{a\le t\le b} M_t^p\Big) \le \Big(\frac{p}{p-1}\Big)^p \mathbb{E}M_b^p.$$

Theorem 1.10 ***(Non-negative semimartingale convergence theorem)*** *Let $\{A_t\}_{t\ge0}$ and $\{U_t\}_{t\ge0}$ be two continuous adapted increasing processes with $A_0 = U_0 = 0$ a.s. Let $\{M_t\}_{t\ge0}$ be a real-valued continuous local martingale with $M_0 = 0$ a.s. Let ξ be a non-negative $\mathcal{F}_0$-measurable random variable such that $\mathbb{E}\xi < \infty$. Define*

$$X_t = \xi + A_t - U_t + M_t \quad \text{for } t \ge 0.$$

If X_t is non-negative, then

$$\Big\{\lim_{t\to\infty} A_t < \infty\Big\} \subset \Big\{\lim_{t\to\infty} X_t \text{ exists and is finite}\Big\} \cap \Big\{\lim_{t\to\infty} U_t < \infty\Big\} \quad a.s.$$

where $B \subset D$ a.s. means $\mathbb{P}(B\cap D^c) = 0$. In particular, if $\lim_{t\to\infty} A_t < \infty$ a.s., then for almost all $\omega \in \Omega$

$$\lim_{t\to\infty} X_t(\omega) < \infty, \quad \lim_{t\to\infty} U_t(\omega) < \infty \quad \text{and} \quad -\infty < \lim_{t\to\infty} M_t(\omega) < \infty.$$

If we apply these results to an $\mathbb{R}^n$-valued martingale, we obtain the following Doob martingale inequalities.

Theorem 1.11 ***(Doob's martingale inequalities)*** *Let $\{M_t\}_{t\ge0}$ be an $\mathbb{R}^n$-valued martingale. Let $[a,b]$ be a bounded interval in $\mathbb{R}_+$.*

(i) If $p \ge 1$, $c > 0$ and $M_t \in L^p(\Omega;\mathbb{R}^n)$, then

$$\mathbb{P}\Big\{\omega : \sup_{a\le t\le b} |M_t(\omega)| \ge c\Big\} \le \frac{\mathbb{E}|M_b|^p}{c^p}.$$

(ii) If $p > 1$ and $M_t \in L^p(\Omega;\mathbb{R}^n)$, then

$$\mathbb{E}\Big(\sup_{a\le t\le b} |M_t|^p\Big) \le \Big(\frac{p}{p-1}\Big)^p \mathbb{E}|M_b|^p.$$

1.4 Brownian Motions

Brownian motion is at the heart of most models in practice. Its name comes from the Scottish botanist Robert Brown who, in around 1827, reported experimental observations involving the erratic behaviour of a pollen grain when bombarded by (relatively small and effectively invisible) water molecules. A mathematical theory for Brownian motion has since been developed, with famous names such as Albert Einstein and Norbert Weiner making significant contributions. To describe the motion mathematically it is natural to use the concept of a stochastic process $B_t(\omega)$, interpreted as the position of the pollen grain ω at time t. Let us now give the mathematical definition of Brownian motion.

Definition 1.12 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$. A (standard) one-dimensional Brownian motion is a real-valued continuous $\{\mathcal{F}_t\}$-adapted process $\{B_t\}_{t\geq 0}$ with the following properties:

(i) $B_0 = 0$ a.s.;
(ii) for $0 \leq s < t < \infty$, the increment $B_t - B_s$ is normally distributed with mean zero and variance $t - s$;
(iii) for $0 \leq s < t < \infty$, the increment $B_t - B_s$ is independent of $\mathcal{F}_s$.

We shall sometimes speak of a Brownian motion $\{B_t\}_{0\leq t\leq T}$ on $[0, T]$, for some $T > 0$, and the meaning of this terminology is apparent.

If $\{B_t\}_{t\geq 0}$ is a Brownian motion and $0 \leq t_0 < t_1 < \cdots < t_k < \infty$, then the increments $B_{t_i} - B_{t_{i-1}}$, $1 \leq i \leq k$ are independent, and we say that Brownian motion has *independent increments*. Moreover, the distribution of $B_{t_i} - B_{t_{i-1}}$ depends only on the difference $t_i - t_{i-1}$, and we say that Brownian motion has *stationary increments*.

The filtration $\{\mathcal{F}_t\}$ is a part of the definition of Brownian motion. However, we sometimes speak of a Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ without filtration. That is, $\{B_t\}_{t\geq 0}$ is a real-valued continuous process with properties (i) and (ii) but property (iii) is replaced by that it has independent increments. In this case, define $\mathcal{F}_t^B = \sigma(B_s : 0 \leq s \leq t)$ for $t \geq 0$, i.e. $\mathcal{F}_t^B$ is the σ-algebra generated by $\{B_s : 0 \leq s \leq t\}$. We call $\{\mathcal{F}_t^B\}_{t\geq 0}$ the *natural filtration* generated by $\{B_t\}$. Clearly, $\{B_t\}$ is a Brownian motion with respect to the natural filtration $\{\mathcal{F}_t^B\}$. Moreover, if $\{\mathcal{F}_t\}$ is a "larger" filtration in the sense that $\mathcal{F}_t^B \subset \mathcal{F}_t$ for $t \geq 0$, and $B_t - B_s$ is independent of $\mathcal{F}_s$ whenever $0 \leq s < t < \infty$, then $\{B_t\}$ is a Brownian motion with respect to the filtration $\{\mathcal{F}_t\}$.

In the definition we do not require the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ be complete and the filtration $\{\mathcal{F}_t\}$ satisfy the usual conditions. However, it is often necessary to work on a complete probability space with a filtration satisfying the usual conditions. Let $\{B_t\}_{t\geq 0}$ be a Brownian motion defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $(\Omega, \bar{\mathcal{F}}, \mathbb{P})$ be the completion of $(\Omega, \mathcal{F}, \mathbb{P})$. Clearly, $\{B_t\}$ is a Brownian motion on the complete probability space $(\Omega, \bar{\mathcal{F}}, \mathbb{P})$. Let $\mathcal{N}$ be the collection of $\mathbb{P}$-null sets, i.e. $\mathcal{N} = \{A \in \bar{\mathcal{F}} : \mathbb{P}(A) = 0\}$. For $t \geq 0$, define

$$\bar{\mathcal{F}}_t = \sigma(\mathcal{F}_t^B \cup \mathcal{N}).$$

We called $\{\bar{\mathcal{F}}_t\}$ the *augmentation under* $\mathbb{P}$ *of the natural filtration* $\{\mathcal{F}_t^B\}$ *generated by* $\{B_t\}$. It is known that the augmentation $\{\bar{\mathcal{F}}_t\}$ is a filtration on $(\Omega, \bar{\mathcal{F}}, \mathbb{P})$ satisfying the usual condition. Moreover, $\{B_t\}$ is a Brownian motion on $(\Omega, \bar{\mathcal{F}}, \mathbb{P})$ with respect to $\{\bar{\mathcal{F}}_t\}$. This shows that given a Brownian motion $\{B_t\}_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, one can construct a complete probability space with a filtration satisfying the usual conditions to work on.

However, throughout this book, unless otherwise specified, we would rather assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space with a filtration $\{\mathcal{F}_t\}$ satisfying the usual conditions, and the one-dimensional Brownian motion $\{B_t\}$ is defined on it.

In Section 1.1 we mentioned that the integral $\int_{t_0}^t \sigma x(u) dB(u)$ cannot be defined as the classical Lebesgue integral since for almost every $\omega \in \Omega$, the Brownian sample path $B_.(\omega)$ is nowhere differentiable. To begin, we note that Brownian motion has a remarkable scaling property: for any fixed $c \neq 0$,

$$X_t := \frac{B_{c^2 t}}{c}, \quad t \geq 0, \tag{1.4}$$

is a Brownian motion with respect to the filtration $\{\mathcal{F}_{c^2 t}\}$. Consider the quantity $|B_t|/t$. Since $B_0 = 0$, if B_t were differentiable then $|B_t|/t$ would converge to $|B_0'|$ as $t \to 0$. Let $t = 1/k^2$, where k is large, and set $c = k^2$ in (1.4). Since B_t and X_t have the same distributions we have

$$\mathbb{P}\Big\{\frac{|B_{1/k^4}|}{1/k^4} > k\Big\} = \mathbb{P}\Big\{\frac{|X_{1/k^4}|}{1/k^4} > k\Big\} = \mathbb{P}\Big\{\frac{|B_1|}{1/k^2} > k\Big\} = \mathbb{P}\Big\{|B_1| > \frac{1}{k}\Big\}.$$

Since B_1 is $N(0,1)$, we have

$$\lim_{k\to\infty} \mathbb{P}\Big\{\frac{|B_{1/k^4}|}{1/k^4} > k\Big\} = 1.$$

This shows that, with probability 1, B_t is not differentiable at $t = 0$. A similar argument can be used to show that B_t is nowhere differentiable, with probability 1.

Another way of examining the roughness of Brownian motion is to consider its variation. Recall that a continuously differentiable function,

$f \in C^1([0,T];\mathbb{R})$, has finite variation. In fact, let k be any large integer and set $\Delta = T/k$ and $t_j = j\Delta$ for $0 \le j \le k$. The mean value theorem says that

$$f(t_j) - f(t_{j-1}) = \Delta f'(\theta_j), \quad \text{for some } \theta_j \in (t_{j-1}, t_j).$$

Thus

$$\sum_{j=1}^{k} |f(t_j) - f(t_{j-1})| = \Delta \sum_{j=1}^{k} |f'(\theta_j)| \le T \max_{t\in[0,T]} |f'(t)|.$$

It follows that the variation of f obeys

$$\limsup_{k\to\infty} \sum_{j=1}^{k} |f(t_j) - f(t_{j-1})| \le T \max_{t\in[0,T]} |f'(t)| < \infty.$$

To see whether Brownian motion has a similar property we use the inequality

$$\sum_{j=1}^{k} (B_{t_j} - B_{t_{j-1}})^2 \le \Big(\max_{1\le j\le k} |B_{t_j} - B_{t_{j-1}}| \Big) \sum_{j=1}^{k} |B_{t_j} - B_{t_{j-1}}|. \qquad (1.5)$$

Note that the random variable $\sum_{j=1}^{k}(B_{t_j} - B_{t_{j-1}})^2$ has mean T and variance of $O(\Delta)$ (see Exercise 1.1). This implies

$$\lim_{k\to\infty} \sum_{j=1}^{k} (B_{t_j} - B_{t_{j-1}})^2 = T \quad a.s.$$

On the other hand, each $B_{t_j} - B_{t_{j-1}}$ has mean zero and variance Δ. It can then be shown that

$$\lim_{k\to\infty} \Big(\max_{1\le j\le k} |B_{t_j} - B_{t_{j-1}}| \Big) = 0 \quad a.s.$$

In order for inequality (1.5) to hold it must therefore be true that, with probability 1, $\sum_{j=1}^{k} |B_{t_j} - B_{t_{j-1}}|$ is unbounded as $k \to \infty$. We thus say that Brownian motion has infinite variation in any finite time interval.

Although Brownian motion is rough, it has many important properties, and some of them are summarised below:

(a) $\{B_t\}$ is a continuous square-integrable martingale and its quadratic variation $\langle B, B\rangle_t = t$ for all $t \ge 0$.

(b) The strong law of large numbers states that

$$\lim_{t\to\infty} \frac{B_t}{t} = 0 \quad a.s.$$

(c) For almost every $\omega \in \Omega$, the Brownian sample path $B_\cdot(\omega)$ is locally Hölder continuous with exponent δ if $\delta \in (0, \frac{1}{2})$. However, for almost every $\omega \in \Omega$, the Brownian sample path $B_\cdot(\omega)$ is nowhere Hölder continuous with exponent $\delta > \frac{1}{2}$.

Besides, we have the following well-known law of the iterated logarithm.

Theorem 1.13 *(Law of the Iterated Logarithm) For almost every $\omega \in \Omega$, we have*

$$\text{(i)} \ \limsup_{t\downarrow 0} \frac{B_t(\omega)}{\sqrt{2t\log\log(1/t)}} = 1, \quad \text{(ii)} \ \liminf_{t\downarrow 0} \frac{B_t(\omega)}{\sqrt{2t\log\log(1/t)}} = -1,$$

$$\text{(iii)} \ \limsup_{t\to\infty} \frac{B_t(\omega)}{\sqrt{2t\log\log t}} = 1, \quad \text{(iv)} \ \liminf_{t\to\infty} \frac{B_t(\omega)}{\sqrt{2t\log\log t}} = -1.$$

This theorem shows that for any $\varepsilon > 0$ there exists a positive random variable ρ_ε such that for almost every $\omega \in \Omega$, the Brownian sample path $B_\cdot(\omega)$ is within the interval $\pm(1+\varepsilon)\sqrt{2t\log\log t}$ whenever $t \geq \rho_\varepsilon(\omega)$, that is

$$-(1+\varepsilon)\sqrt{2t\log\log t} \leq B_t(\omega) \leq (1+\varepsilon)\sqrt{2t\log\log t} \quad \text{for all } t \geq \rho_\varepsilon(\omega).$$

On the other hand, the bounds $-(1-\varepsilon)\sqrt{2t\log\log t}$ and $(1-\varepsilon)\sqrt{2t\log\log t}$ (for $0 < \varepsilon < 1$) are exceeded in every t-neighbourhood of ∞ for every sample path.

Let us now define an n-dimensional Brownian motion.

Definition 1.14 An n-dimensional process $\{B_t = (B_t^1, \cdots, B_t^n)\}_{t\geq 0}$ is called an n-dimensional Brownian motion if every $\{B_t^i\}$ is a one-dimensional Brownian motion, and $\{B_t^1\}, \cdots, \{B_t^n\}$ are independent.

For an n-dimensional Brownian motion, we still have, for example,

$$\limsup_{t\to\infty} \frac{|B_t|}{\sqrt{2t\log\log t}} = 1 \quad a.s.$$

This is somewhat surprising because it means that the independent individual components of B_t are not simultaneously of the order $\sqrt{2t\log\log t}$, otherwise $\sqrt{n}$ instead of 1 would have appeared in the right-hand side of the above equality.

It is easy to see that an n-dimensional Brownian motion is an n-dimensional continuous martingale with the joint quadratic variations

$$\langle B^i, B^j\rangle_t = \delta_{ij}t \quad \text{for } 1 \leq i, j \leq n,$$

where δ_{ij} is the Dirac delta function, i.e.

$$\delta_{ij} = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j. \end{cases}$$

It turns out that this property characterises Brownian motion among continuous local martingales. This is described by the following well-known Lévy theorem.

Theorem 1.15 *Let $\{M_t = (M_t^1, \cdots, M_t^n)\}_{t\geq 0}$ be an n-dimensional continuous local martingale with respect to the filtration $\{\mathcal{F}_t\}$ and $M_0 = 0$ a.s. If*

$$\langle M^i, M^j\rangle_t = \delta_{ij}t \quad \textit{for } 1 \leq i, j \leq n,$$

then $\{M_t = (M_t^1, \cdots, M_t^n)\}_{t\geq 0}$ is an n-dimensional Brownian motion with respect to $\{\mathcal{F}_t\}$.

As an application of the Lévy theorem, one can show the following useful result.

Theorem 1.16 *Let $M = \{M_t\}_{t\geq 0}$ be a real-valued continuous local martingale such that $M_0 = 0$ and $\lim_{t\to\infty}\langle M, M\rangle_t = \infty$ a.s. For each $t \geq 0$, define the stopping time*

$$\tau_t = \inf\{s : \langle M, M\rangle_s > t\}.$$

Then $\{M_{\tau_t}\}_{t\geq 0}$ is a Brownian motion with respect to the filtration $\{\mathcal{F}_{\tau_t}\}_{t\geq 0}$.

1.5 Stochastic Integrals

In this section we shall define the stochastic integral

$$\int_0^t f(s)dB_s$$

with respect to an m-dimensional Brownian motion $\{B_t\}$ for a class of $n \times m$-matrix-valued stochastic processes $\{f(t)\}$. Since for almost all $\omega \in \Omega$, the Brownian sample path $B_.(\omega)$ is not only nowhere differentiable but also has infinite variation in any finite time interval, the integral can not be defined in the ordinary way. However, we can define the integral for a large class of stochastic processes by making use of the stochastic nature of Brownian motion. This integral was first defined by K. Itô in 1949 and is now known as *Itô stochastic integral.* We shall now start to define the stochastic integral step by step.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions. Let $B = \{B_t\}_{t\geq 0}$ be a one-dimensional Brownian motion defined on the probability space adapted to the filtration.

Definition 1.17 Let $0 \leq a < b < \infty$. Denote by $\mathcal{M}^2([a,b];\mathbb{R})$ the space of all real-valued measurable $\{\mathcal{F}_t\}$-adapted processes $f = \{f(t)\}_{a\leq t\leq b}$ such that

$$\|f\|_{a,b}^2 = \mathbb{E}\int_a^b |f(t)|^2 dt < \infty. \tag{1.6}$$

We identify f and $\bar{f}$ in $\mathcal{M}^2([a,b];\mathbb{R})$ if $\|f-\bar{f}\|_{a,b}^2 = 0$. In this case we say that f and $\bar{f}$ are equivalent and write $f = \bar{f}$.

Clearly, $\|\cdot\|_{a,b}$ defines a metric on $\mathcal{M}^2([a,b];\mathbb{R})$ and the space is complete under this metric. Let us point out that for every $f \in \mathcal{M}^2([a,b];\mathbb{R})$, there is a predictable $\bar{f} \in \mathcal{M}^2([a,b];\mathbb{R})$ such that $f = \bar{f}$. In fact, f has a progressively measurable modification $\hat{f}$ in $\mathcal{M}^2([a,b];\mathbb{R})$ and then we may take

$$\bar{f}(t) = \limsup_{h\downarrow 0} \frac{1}{h}\int_{t-h}^t \hat{f}(s)ds.$$

Thus, if necessary, we may assume that $f \in \mathcal{M}^2([a,b];\mathbb{R})$ is predictable without loss of generality. However, in this book we would rather follow the usual custom of not being very careful about the distinction between the equivalence processes.

For stochastic processes $f \in \mathcal{M}^2([a,b];\mathbb{R})$ we shall show how to define the Itô integral $\int_a^b f(t)dB_t$. The idea is natural: first define the integral $\int_a^b g(t)dB_t$ for a class of simple processes g. Then we show that each $f \in \mathcal{M}^2([a,b];\mathbb{R})$ can be approximated by such simple processes g's and we define the limit of $\int_a^b g(t)dB_t$ as the integral $\int_a^b f(t)dB_t$. Let us first introduce the concept of simple processes.

Definition 1.18 A real-valued stochastic process $g = \{g(t)\}_{a\le t\le b}$ is called a simple (or step) process if there exists a partition $a = t_0 < t_1 < \cdots < t_k = b$ of $[a, b]$, and bounded random variables $\xi_i,\ 0 \le i \le k-1$ such that ξ_i is $\mathcal{F}_{t_i}$-measurable and

$$g(t) = \xi_0 I_{[t_0,\ t_1]}(t) + \sum_{i=1}^{k-1} \xi_i I_{(t_i,\ t_{i+1}]}(t). \tag{1.7}$$

Denote by $\mathcal{M}_0([a,b];\mathbb{R})$ the family of all such processes.

It is obvious that $\mathcal{M}_0([a,b];\mathbb{R}) \subset \mathcal{M}^2([a,b];\mathbb{R})$. We now give the definition of the Itô integral for such simple processes.

Definition 1.19 **(Part 1 of the definition of Itô's integral)** For a simple process g with the form of (1.7) in $\mathcal{M}_0([a,b];\mathbb{R})$, define

$$\int_a^b g(t)dB_t = \sum_{i=0}^{k-1} \xi_i (B_{t_{i+1}} - B_{t_i}) \tag{1.8}$$

and call it the stochastic integral of g with respect to the Brownian motion $\{B_t\}$ or the Itô integral.

Clearly, the stochastic integral $\int_a^b g(t)dB_t$ is $\mathcal{F}_b$-measurable. We shall now show that it belongs to $L^2(\Omega;\mathbb{R})$.

Lemma 1.3 *If $g \in \mathcal{M}_0([a,b];\mathbb{R})$, then*

$$\mathbb{E}\int_a^b g(t)dB_t = 0, \tag{1.9}$$

$$\mathbb{E}\Big|\int_a^b g(t)dB_t\Big|^2 = \mathbb{E}\int_a^b |g(t)|^2 dt. \tag{1.10}$$

Proof. Since ξ_i is $\mathcal{F}_{t_i}$-measurable whereas $B_{t_{i+1}} - B_{t_i}$ is independent of $\mathcal{F}_{t_i}$,

$$\mathbb{E}\int_a^b g(t)dB_t = \sum_{i=0}^{k-1} \mathbb{E}\big[\xi_i (B_{t_{i+1}} - B_{t_i})\big] = \sum_{i=0}^{k-1} \mathbb{E}\xi_i\ \mathbb{E}(B_{t_{i+1}} - B_{t_i}) = 0.$$

Moreover, note that $B_{t_{j+1}} - B_{t_j}$ is independent of $\xi_i\xi_j(B_{t_{i+1}} - B_{t_i})$ if $i < j$.

Thus

$$\mathbb{E}\left|\int_a^b g(t)dB_t\right|^2 = \sum_{0\le i,j\le k-1} \mathbb{E}\big[\xi_i\xi_j(B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})\big]$$
$$= \sum_{i=0}^{k-1} \mathbb{E}\big[\xi_i^2(B_{t_{i+1}} - B_{t_i})^2\big]$$
$$= \sum_{i=0}^{k-1} \mathbb{E}\xi_i^2 \mathbb{E}(B_{t_{i+1}} - B_{t_i})^2$$
$$= \sum_{i=0}^{k-1} \mathbb{E}\xi_i^2 (t_{i+1} - t_i) = \mathbb{E}\int_a^b |g(t)|^2 dt$$

as required. □

Lemma 1.4 *Let $g_1, g_2 \in \mathcal{M}_0([a,b];\mathbb{R})$ and let c_1, c_2 be two real numbers. Then $c_1g_1 + c_2g_2 \in \mathcal{M}_0([a,b];\mathbb{R})$ and*

$$\int_a^b [c_1g_1(t) + c_2g_2(t)]dB_t = c_1\int_a^b g_1(t)dB_t + c_2\int_a^b g_2(t)dB_t.$$

The proof is left to the reader as an exercise. We shall now use the properties shown in Lemmas 5.4 and 5.5 to extend the integral definition from simple processes to processes in $\mathcal{M}^2([a,b];\mathbb{R})$. This is based on the following approximation result.

Lemma 1.5 *For any $f \in \mathcal{M}^2([a,b];\mathbb{R})$, there exists a sequence $\{g_n\}$ of simple processes such that*

$$\lim_{k\to\infty} \mathbb{E}\int_a^b |f(t) - g_n(t)|^2 dt = 0. \tag{1.11}$$

Proof. We divide the whole proof into three steps.

Step 1. We first claim that for any $f \in \mathcal{M}^2([a,b];\mathbb{R})$, there exists a sequence $\{\varphi_k\}_{k\ge1}$ of bounded processes in $\mathcal{M}^2([a,b];\mathbb{R})$ such that

$$\lim_{k\to\infty} \mathbb{E}\int_a^b |f(t) - \varphi_k(t)|^2 dt = 0. \tag{1.12}$$

In fact, for each k, put

$$\varphi_k(t) = [-k \vee f(t)] \wedge k.$$

Then (1.12) follows by the dominated convergence theorem (i.e. Theorem 1.2).

Step 2. We next claim that if $\varphi \in \mathcal{M}^2([a,b];\mathbb{R})$ is bounded, say $|\varphi| \le C = const.$, then there exists a sequence $\{\phi_k\}_{k\ge 1}$ of bounded continuous processes in $\mathcal{M}^2([a,b];\mathbb{R})$ such that

$$\lim_{k\to\infty} \mathbb{E}\int_a^b |\varphi(t) - \phi_k(t)|^2 dt = 0. \tag{1.13}$$

In fact, for each k, let $\rho_k : \mathbb{R} \to \mathbb{R}_+$ be a continuous function such that $\rho_k(s) = 0$ for $s \le -\frac{1}{k}$ and $s \ge 0$ and

$$\int_{-\infty}^{\infty} \rho_k(s)ds = 1.$$

Define

$$\phi_k(t) = \phi_k(t,\omega) = \int_a^b \rho_k(s-t)\varphi(s,\omega)ds.$$

Then for every ω, $\phi_k(\cdot,\omega)$ is continuous and $|\phi_k(t,\omega)| \le C$. Also ϕ_k is a measurable $\{\mathcal{F}_t\}$-adapted process. Moreover, for all $\omega \in \Omega$,

$$\lim_{k\to\infty} \int_a^b |\varphi(t,\omega) - \phi_k(t,\omega)|^2 dt = 0.$$

So (1.13) follows by the bounded convergence theorem.

Step 3. We now claim that if $\phi \in \mathcal{M}^2([a,b];\mathbb{R})$ is bounded and continuous, then there exists a sequence $\{g_k\}$ of simple processes such that

$$\lim_{k\to\infty} \mathbb{E}\int_a^b |\phi(t) - g_k(t)|^2 dt = 0. \tag{1.14}$$

In fact, for each k, let

$$\begin{aligned} g_k(t) &= \phi(a)\ I_{[a,\ a+(b-a)/k]}(t) \\ &\quad + \sum_{i=1}^{k-1} \phi(a + i(b-a)/k)\ I_{(a+i(b-a)/k,\ a+(i+1)(b-a)/k]}(t). \end{aligned}$$

Then $g_k \in \mathcal{M}_0([a,b];\mathbb{R})$, and for every ω,

$$\lim_{k\to\infty} \int_a^b |\phi(t,\omega) - g_k(t,\omega)|^2 dt = 0.$$

So (1.14) follows by the bounded convergence theorem once again. Finally, the conclusion of the lemma follows clearly from steps 1–3 and the proof is now complete. □

We can now explain how to define the Itô integral for a process $f \in \mathcal{M}^2([a,b];\mathbb{R})$. By Lemma 5.6, there is a sequence $\{g_k\}_{k\geq 1}$ of simple processes such that

$$\lim_{k\to\infty} \mathbb{E}\int_a^b |f(t) - g_k(t)|^2 dt = 0.$$

Thus, by Lemmas 1.3 and 1.4,

$$\mathbb{E}\Big|\int_a^b g_k(t)dB_t - \int_a^b g_j(t)dB_t\Big|^2 = \mathbb{E}\Big|\int_a^b [g_k(t) - g_k(t)]dB_t\Big|^2$$
$$= \mathbb{E}\int_a^b |g_k(t) - g_k(t)|^2 dt \to 0 \quad \text{as } k, j \to \infty.$$

In other words, $\{\int_a^b g_k(t)dB_t\}$ is a Cauchy sequence in $L^2(\Omega;\mathbb{R})$. So the limit exists and we define the limit as the stochastic integral. This leads to the following definition.

Definition 1.20 (Part 2 of the definition of Itô's integral) Let $f \in \mathcal{M}^2([a,b];\mathbb{R})$. The Itô integral of f with respect to $\{B_t\}$ is defined by

$$\int_a^b f(t)dB_t = \lim_{k\to\infty}\int_a^b g_k(t)dB_t \quad \text{in } L^2(\Omega;\mathbb{R}), \tag{1.15}$$

where $\{g_k\}$ is a sequence of simple processes such that

$$\lim_{k\to\infty} \mathbb{E}\int_a^b |f(t) - g_k(t)|^2 dt = 0. \tag{1.16}$$

The above definition is independent of the particular sequence $\{g_k\}$. Indeed, if $\{h_k\}$ is another sequence of simple processes converging to f in the sense that

$$\lim_{k\to\infty} \mathbb{E}\int_a^b |f(t) - h_k(t)|^2 dt = 0,$$

then the sequence $\{\varphi_k\}$, where $\varphi_{2k-1} = g_k$ and $\varphi_{2k} = h_k$, is also convergent to f in the same sense. Hence, by what we have proved, the sequence $\{\int_a^b \varphi_k(t)dB_t\}$ is convergent in $L^2(\Omega;\mathbb{R})$. It follows that the limits (in L^2) of $\int_a^b g_k(t)dB_t$ and of $\int_a^b h_k(t)dB_t$ are equal almost surely.

The stochastic integral has many nice properties. We first observe the following:

Theorem 1.21 *Let $f, g \in \mathcal{M}^2([a,b];\mathbb{R})$, and let α, β be two real numbers. Then*

(i) $\int_a^b f(t)dB_t$ *is* $\mathcal{F}_b$*-measurable;*
(ii) $\mathbb{E}\int_a^b f(t)dB_t = 0$;
(iii) $\mathbb{E}|\int_a^b f(t)dB_t|^2 = \mathbb{E}\int_a^b |f(t)|^2 dt$;
(vi) $\int_a^b [\alpha f(t) + \beta g(t)]dB_t = \alpha \int_a^b f(t)dB_t + \beta \int_a^b g(t)dB_t$.

The proof is left to the reader as an exercise. The next theorem improves the results (ii) and (iii) of Theorem 1.21.

Theorem 1.22 *Let* $f \in \mathcal{M}^2([a,b];\mathbb{R})$*. Then*

$$\mathbb{E}\Big(\int_a^b f(t)dB(t)|\mathcal{F}_a\Big) = 0, \tag{1.17}$$

$$\begin{aligned}\mathbb{E}\Big(\Big|\int_a^b f(t)dB(t)\Big|^2|\mathcal{F}_a\Big) &= \mathbb{E}\Big(\int_a^b |f(t)|^2 dt|\mathcal{F}_a\Big)\\ &= \int_a^b \mathbb{E}(|f(t)|^2|\mathcal{F}_a)dt.\end{aligned} \tag{1.18}$$

We need a simple lemma.

Lemma 1.6 *If* $f \in \mathcal{M}^2([a,b];\mathbb{R})$ *and* ξ *is a real-valued bounded* $\mathcal{F}_a$*-measurable random variable, then* $\xi f \in \mathcal{M}^2([a,b];\mathbb{R})$ *and*

$$\int_a^b \xi f(t)dB_t = \xi \int_a^b f(t)dB_t. \tag{1.19}$$

Proof. It is clear that $\xi f \in \mathcal{M}^2([a,b];\mathbb{R})$. If f is a simple processes, then (1.19) follows from the definition of the stochastic integral. For general $f \in \mathcal{M}^2([a,b];\mathbb{R})$, let $\{g_k\}$ be a sequence of simple processes satisfying (1.16). Applying (1.19) to each g_k and taking $k \to \infty$, the required assertion (1.19) follows. □

Proof of Theorem 1.22. By the definition of conditional expectation, (1.17) holds if and only if

$$\mathbb{E}\Big(I_A \int_a^b f(t)dB(t)\Big) = 0$$

for all sets $A \in \mathcal{F}_a$. But by Lemma 1.6 and Theorem 1.21,

$$\mathbb{E}\Big(I_A \int_a^b f(t)dB(t)\Big) = \mathbb{E}\int_a^b I_A f(t)dB(t) = 0$$

as required. The proof of (1.18) is similar. □

Let $T > 0$ and $f \in \mathcal{M}^2([0,T];\mathbb{R})$. Clearly, for any $0 \le a < b \le T$, $\{f(t)\}_{a \le t \le b} \in \mathcal{M}^2([a,b];\mathbb{R})$ so $\int_a^b f(t)dB_t$ is well defined. It is easy to show that

$$\int_a^b f(t)dB_t + \int_b^c f(t)dB_t = \int_a^c f(t)dB_t \tag{1.20}$$

if $0 \le a < b < c \le T$.

Definition 1.23 Let $f \in \mathcal{M}^2([0,T];\mathbb{R})$. Define

$$I(t) = \int_0^t f(s)dB_s \quad \text{for } 0 \le t \le T,$$

where, by definition, $I(0) = \int_0^0 f(s)dB_s = 0$. We call $I(t)$ the indefinite Itô integral of f.

Clearly, $\{I(t)\}$ is $\{\mathcal{F}_t\}$-adapted. We now show the very important martingale property of the indefinite Itô integral.

Theorem 1.24 *If $f \in \mathcal{M}^2([0,T];\mathbb{R})$, then the indefinite integral $\{I(t)\}_{0 \le t \le T}$ is a square-integrable martingale with respect to the filtration $\{\mathcal{F}_t\}$. In particular,*

$$\mathbb{E}\left[\sup_{0 \le t \le T}\left|\int_0^t f(s)dB_s\right|^2\right] \le 4\mathbb{E}\int_0^T |f(s)|^2 ds. \tag{1.21}$$

Proof. Clearly, $\{I(t)\}_{0 \le t \le T}$ is square-integrable. To show the martingale property, let $0 \le s < t \le T$. By (1.20) and Theorem 1.22

$$\mathbb{E}(I(t)|\mathcal{F}_s) = \mathbb{E}(I(s)|\mathcal{F}_s) + \mathbb{E}\Big(\int_s^t f(r)dB_r|\mathcal{F}_s\Big) = I(s)$$

as desired. The inequality (1.21) now follows from Doob's martingale inequality (i.e. Theorem 1.11). □

Theorem 1.25 *If $f \in \mathcal{M}^2([0,T];\mathbb{R})$, then the indefinite integral $\{I(t)\}_{0 \le t \le T}$ has a continuous version.*

Proof. Let $\{g_k\}$ be a sequence of simple processes such that

$$\lim_{k \to \infty} \mathbb{E}\int_0^T |f(s) - g_k(s)|^2 ds = 0. \tag{1.22}$$

Note from the definition of the stochastic integral and the continuity of the Brownian motion that the indefinite integrals

$$I_k(t) = \int_0^t g_k(s)dB_s, \quad 0 \le t \le T$$

are continuous. By Theorem 1.24, $\{I_k(t) - I_j(t)\}$ is a martingale, for each pair of integers k, j. Hence, by Doob's martingale inequality (namely, Theorem 1.11), for any $\varepsilon > 0$

$$\mathbb{P}\Big\{\sup_{0\le t\le T} |I_k(t) - I_j(t)| \ge \varepsilon\Big\} \le \frac{1}{\varepsilon^2}\mathbb{E}|I_k(T) - I_j(T)|^2$$
$$= \frac{1}{\varepsilon^2}\mathbb{E}\int_0^T |g_k(s) - g_j(s)|^2 ds \to 0 \quad \text{as } k, j \to \infty.$$

For each $i = 1, 2, \cdots$, taking $\varepsilon = i^{-2}$, it follows that for some k_i sufficiently large,

$$\mathbb{P}\Big\{\sup_{0\le t\le T} |I_{k_i}(t) - I_j(t)| \ge \frac{1}{i^2}\Big\} \le \frac{1}{i^2} \quad \text{if } j \ge k_i.$$

One can then choose the k_i in such a way that $k_i \uparrow \infty$ as $i \to \infty$ and

$$\mathbb{P}\Big\{\sup_{0\le t\le T} |I_{k_i}(t) - I_{k_{i+1}}(t)| \ge \frac{1}{i^2}\Big\} \le \frac{1}{i^2}, \quad i \ge 1.$$

Since $\sum i^{-2} < \infty$, the Borel–Cantelli lemma (i.e. Lemma 1.2) implies that there exists a set $\Omega_0 \in \mathcal{F}$ with $\mathbb{P}(\Omega_0) = 0$ and an integer-valued random variable i_0 such that for every $\omega \in \Omega_0$,

$$\sup_{0\le t\le T} |I_{k_i}(t,\omega) - I_{k_{i+1}}(t,\omega)| < \frac{1}{i^2} \quad \text{if } i \ge i_0(\omega).$$

In other words, with probability 1, $\{I_{k_i}(t)\}_{i\ge 1}$ is uniformly convergent in $t \in [0,T]$, and therefore the limit, denoted by $J(t)$, is continuous in $t \in [0,T]$ for almost all $\omega \in \Omega$. Since (1.22) implies

$$\lim_{i\to\infty} I_{k_i}(t) = \int_0^t f(s)dB_s \quad \text{in } L^2(\Omega, \mathbb{R}),$$

it follows that

$$J(t) = \int_0^t f(s)dB_s \quad a.s.$$

That is, the indefinite integral has a continuous version.

From now on, when we speak of the indefinite integral we always mean a continuous version of it.

Theorem 1.26 *Let $f \in \mathcal{M}^2([0,T];\mathbb{R})$. Then the indefinite integral $I = \{I(t)\}_{0\le t\le T}$ is a square-integrable continuous martingale and its quadratic variation is given by*

$$\langle I, I\rangle_t = \int_0^t |f(s)|^2 ds, \quad 0 \le t \le T. \tag{1.23}$$

Proof. Obviously we need only to show (1.23). By the definition of the quadratic variation we need to show that $\{I^2(t) - \langle I, I\rangle_t\}$ is a continuous martingale vanishing at $t = 0$. But, obviously $I^2(0) - \langle I, I\rangle_0 = 0$. Moreover, if $0 \le r < t \le T$, by Theorem 1.22,

$$\begin{aligned}
&\mathbb{E}(I^2(t) - \langle I, I\rangle_t|\mathcal{F}_r) \\
&= I^2(r) - \langle I, I\rangle_r + 2I(r)\mathbb{E}\Big(\int_r^t f(s)dB_s|\mathcal{F}_r\Big) \\
&+ \mathbb{E}\Big(\Big|\int_r^t f(s)dB_s\Big|^2|\mathcal{F}_r\Big) - \mathbb{E}\Big(\int_r^t |f(s)|^2 ds|\mathcal{F}_r\Big) \\
&= I^2(r) - \langle I, I\rangle_r
\end{aligned}$$

as desired. □

Let us now proceed to define the stochastic integrals with stopping time. We observe that if τ is an $\{\mathcal{F}_t\}$-stopping time, then $\{I_{[[0,\,\tau]]}(t)\}_{t\ge 0}$ is a bounded right continuous $\{\mathcal{F}_t\}$-adapted process. In fact, the boundedness and right continuity are obvious. Moreover, for each $t \ge 0$,

$$\{\omega : I_{[[0,\,\tau]]}(t,\omega) \le r\} = \begin{cases} \emptyset \in \mathcal{F}_t & \text{if } r < 0, \\ \{\omega : \tau(\omega) < t\} \in \mathcal{F}_t & \text{if } 0 \le r < 1, \\ \Omega \in \mathcal{F}_t & \text{if } r \ge 1, \end{cases}$$

that is, $I_{[[0,\,\tau]]}(t)$ is $\mathcal{F}_t$-measurable. Therefore, $\{I_{[[0,\,\tau]]}(t)\}_{t\ge 0}$ is also predictable.

Definition 1.27 Let $f \in \mathcal{M}^2([0,T];\mathbb{R})$, and let τ be an $\{\mathcal{F}_t\}$-stopping time such that $0 \le \tau \le T$. Then, $\{I_{[[0,\,\tau]]}(t)f(t)\}_{0\le t\le T} \in \mathcal{M}^2([0,T];\mathbb{R})$ clearly, and we define

$$\int_0^\tau f(s)dB_s = \int_0^T I_{[[0,\,\tau]]}(s)f(s)dB_s.$$

Furthermore, if ρ is another stopping time with $0 \le \rho \le \tau$, we define

$$\int_\rho^\tau f(s)dB_s = \int_0^\tau f(s)dB_s - \int_0^\rho f(s)dB_s.$$

It is easy to see that

$$\int_\rho^\tau f(s)dB_s = \int_0^T I_{]]\rho,\ \tau]]}(s)f(s)dB_s. \tag{1.24}$$

If applying Theorem 1.21 to this we immediately obtain:

Theorem 1.28 *Let $f \in \mathcal{M}^2([0,T];\mathbb{R})$, and let ρ, τ be two stopping times such that $0 \le \rho \le \tau \le T$. Then*

$$\mathbb{E}\int_\rho^\tau f(s)dB_s = 0,$$

$$\mathbb{E}\Big|\int_\rho^\tau f(s)dB_s\Big|^2 = \mathbb{E}\int_\rho^\tau |f(s)|^2 ds.$$

However, the next theorem improves these results and is also a generalisation of Theorem 1.22.

Theorem 1.29 *Let $f \in \mathcal{M}^2([0,T];\mathbb{R})$, and let ρ, τ be two stopping times such that $0 \le \rho \le \tau \le T$. Then*

$$\mathbb{E}\Big(\int_\rho^\tau f(s)dB_s|\mathcal{F}_\rho\Big) = 0, \tag{1.25}$$

$$\mathbb{E}\Big(\Big|\int_\rho^\tau f(s)dB_s\Big|^2|\mathcal{F}_\rho\Big) = \mathbb{E}\Big(\int_\rho^\tau |f(s)|^2 ds|\mathcal{F}_\rho\Big). \tag{1.26}$$

We need a useful lemma.

Lemma 1.7 *Let $f \in \mathcal{M}^2([0,T];\mathbb{R})$, and let τ be a stopping time such that $0 \le \tau \le T$. Then*

$$\int_0^\tau f(s)dB_s = I(\tau),$$

where $\{I(t)\}_{0\le t\le T}$ is the indefinite integral of f given by Definition 1.23.

We leave the proof of this lemma to the reader, but prove Theorem 1.29.

Proof of Theorem 1.29. By Theorem 1.26 and the Doob martingale stopping theorem (i.e. Theorem 1.5),

$$\mathbb{E}(I(\tau)|\mathcal{F}_\rho) = I(\rho) \tag{1.27}$$

and

$$\mathbb{E}(I^2(\tau) - \langle I, I\rangle_\tau | \mathcal{F}_r) = I^2(\rho) - \langle I, I\rangle_\rho, \tag{1.28}$$

where $\{\langle I, I\rangle_t\}$ is defined by (1.23). Applying Lemma 1.7 one then sees from (1.27) that

$$\mathbb{E}\Big(\int_\rho^\tau f(s)dB_s|\mathcal{F}_\rho\Big) = \mathbb{E}(I(\tau) - I(\rho)|\mathcal{F}_\rho) = 0$$

which is (1.25). Moreover, by (1.27) and (1.28),

$$\begin{aligned}&\mathbb{E}(|I(\tau) - I(\rho)|^2|\mathcal{F}_\rho) = \mathbb{E}(I^2(\tau)|\mathcal{F}_\rho) - 2I(\rho)\mathbb{E}(I(\tau)|\mathcal{F}_\rho) + I^2(\rho)\\ &= \mathbb{E}(I^2(\tau)|\mathcal{F}_\rho) - I^2(\rho) = \mathbb{E}(\langle I, I\rangle_\tau - \langle I, I\rangle_\rho|\mathcal{F}_\rho) = \mathbb{E}\Big(\int_\rho^\tau |f(s)|^2 ds|\mathcal{F}_\rho\Big)\end{aligned}$$

which, by Lemma 1.7, is the required (5.21). □

Corollary 1.30 *Let $f, g \in \mathcal{M}^2([0,T];\mathbb{R})$, and let ρ, τ be two stopping times such that $0 \le \rho \le \tau \le T$. Then*

$$\mathbb{E}\Big(\int_\rho^\tau f(s)dB_s \int_\rho^\tau g(s)dB_s|\mathcal{F}_\rho\Big) = \mathbb{E}\Big(\int_\rho^\tau f(s)g(s)ds|\mathcal{F}_\rho\Big).$$

Proof. By Theorem 1.29,

$$\begin{aligned}&4\mathbb{E}\Big(\int_\rho^\tau f(s)dB_s \int_\rho^\tau g(s)dB_s|\mathcal{F}_\rho\Big)\\ &= \mathbb{E}\Big(\Big|\int_\rho^\tau (f(s)+g(s))dB_s\Big|^2|\mathcal{F}_\rho\Big) - \mathbb{E}\Big(\Big|\int_\rho^\tau (f(s)-g(s))dB_s\Big|^2|\mathcal{F}_\rho\Big)\\ &= \mathbb{E}\Big(\int_\rho^\tau (f(s)+g(s))^2 ds|\mathcal{F}_\rho\Big) - \mathbb{E}\Big(\int_\rho^\tau (f(s)-g(s))^2 ds|\mathcal{F}_\rho\Big)\\ &= 4\mathbb{E}\Big(\int_\rho^\tau f(s)g(s)ds|\mathcal{F}_\rho\Big)\end{aligned}$$

as desired. □

Let us now begin to extend the Itô stochastic integral to the multi-dimensional case. Let $\{B_t = (B_t^1, \cdots, B_t^m)^T\}_{t\ge0}$ be an m-dimensional Brownian motion defined on the complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ adapted to the filtration $\{\mathcal{F}_t\}$. Let $\mathcal{M}^2([0,T];\mathbb{R}^{n\times m})$ denote the family of all $n \times m$-matrix-valued measurable $\{\mathcal{F}_t\}$-adapted processes $f =$

$\{(f_{ij}(t))_{n\times m}\}_{0\le t\le T}$ such that

$$\mathbb{E}\int_0^T |f(s)|^2 dt < \infty.$$

Here, and throughout this book, $|A|$ will denote the trace norm for matrix A, i.e. $|A| = \sqrt{\text{trace}(A^T A)}$.

Definition 1.31 Let $f \in \mathcal{M}^2([0,T];\mathbb{R}^{n\times m})$. Using matrix notation, we define the multi-dimensional indefinite Itô integral

$$\int_0^t f(s)dB_s = \int_0^t \begin{pmatrix} f_{11}(s) & \cdots & f_{1m}(s) \\ \vdots & & \vdots \\ f_{d1}(s) & \cdots & f_{dm}(s) \end{pmatrix} \begin{pmatrix} dB_s^1 \\ \vdots \\ dB_s^m \end{pmatrix}$$

to be the n-column-vector-valued process whose i'th component is the following sum of one-dimensional Itô integrals

$$\sum_{j=1}^m \int_0^t f_{ij}(s)dB_s^j.$$

Clearly, the Itô integral is an $\mathbb{R}^n$-valued continuous martingale with respect to $\{\mathcal{F}_t\}$. Besides, it has the following important properties.

Theorem 1.32 *Let $f \in \mathcal{M}^2([0,T];\mathbb{R}^{n\times m})$, and let ρ,τ be two stopping times such that $0 \le \rho \le \tau \le T$. Then*

$$\mathbb{E}\Big(\int_\rho^\tau f(s)dB_s|\mathcal{F}_\rho\Big) = 0, \tag{1.29}$$

$$\mathbb{E}\Big(\Big|\int_\rho^\tau f(s)dB_s\Big|^2|\mathcal{F}_\rho\Big) = \mathbb{E}\Big(\int_\rho^\tau |f(s)|^2 ds|\mathcal{F}_\rho\Big). \tag{1.30}$$

Assertion (1.29) follows from the definition of multi-dimensional Itô integral and Theorem 1.29, while (1.30) follows from Theorem 1.29 and the following lemma.

Lemma 1.8 *Let $\{B_t^1\}_{t\ge0}$ and $\{B_t^2\}_{t\ge0}$ be two independent one-dimensional Brownian motions. Let $f,g \in \mathcal{M}^2([0,T];\mathbb{R})$, and let ρ,τ be two stopping times such that $0 \le \rho \le \tau \le T$. Then*

$$\mathbb{E}\Big(\int_\rho^\tau f(s)dB_s^1 \int_\rho^\tau g(s)dB_s^2|\mathcal{F}_\rho\Big) = 0. \tag{1.31}$$

Proof. We first claim that if $\varphi, \phi \in \mathcal{M}^2([a,b];\mathbb{R})$. Then

$$\mathbb{E}\Big(\int_a^b \varphi(s)dB_s^1 \int_a^b \phi(s)dB_s^2\Big) = 0. \tag{1.32}$$

In fact, let φ, ϕ be simple processes with the forms

$$\varphi(t) = \xi_0 I_{[t_0,\, t_1]}(t) + \sum_{i=1}^{k-1} \xi_i I_{(t_i,\, t_{i+1}]}(t)$$

and

$$\phi(t) = \zeta_0 I_{[\bar{t}_0,\, \bar{t}_1]}(t) + \sum_{j=1}^{m-1} \zeta_i I_{(\bar{t}_j,\, \bar{t}_{j+1}]}(t).$$

Then

$$\mathbb{E}\Big(\int_a^b \varphi(s)dB_s^1 \int_a^b \phi(s)dB_s^2\Big) = \sum_{i=0}^{k-1}\sum_{j=0}^{m-1} \mathbb{E}\big[\xi_i\zeta_j(B^1_{t_{i+1}} - B^1_{t_i})(B^2_{\bar{t}_{j+1}} - B^2_{\bar{t}_j})\big].$$

But for every pair of i, j, if $t_i \le \bar{t}_j$, then $B^2_{\bar{t}_{j+1}} - B^2_{\bar{t}_j}$ is independent of $\xi_i\zeta_j(B^1_{t_{i+1}} - B^1_{t_i})$ and hence

$$\mathbb{E}\big[\xi_i\zeta_j(B^1_{t_{i+1}} - B^1_{t_i})(B^2_{\bar{t}_{j+1}} - B^2_{\bar{t}_j})\big] = 0.$$

Similarly, it still holds if $t_i > \bar{t}_j$. In other words, we have shown that (1.32) holds for simple processes φ, ϕ, but the general case follows by the approximation procedure.

We next observe that for any $0 \le r < t \le T$

$$\mathbb{E}\Big(\int_r^t f(s)dB_s^1 \int_r^t g(s)dB_s^2 | \mathcal{F}_r\Big) = 0, \tag{1.33}$$

since, by (1.32) and Lemma 1.6, for any $A \in \mathcal{F}_r$

$$\mathbb{E}\Big(I_A \int_r^t f(s)dB_s^1 \int_r^t g(s)dB_s^2\Big) = \mathbb{E}\Big(\int_r^t I_A f(s)dB_s^1 \int_r^t g(s)dB_s^2\Big) = 0.$$

Therefore

$$\begin{aligned}
&\mathbb{E}\Big(\int_0^t f(s)dB_s^1 \int_0^t g(s)dB_s^2|\mathcal{F}_r\Big)\\
&= \int_0^r f(s)dB_s^1 \int_0^r g(s)dB_s^2 + \int_0^r f(s)dB_s^1 \mathbb{E}\Big(\int_r^t g(s)dB_s^2|\mathcal{F}_r\Big)\\
&+ \int_0^r g(s)dB_s^2 \mathbb{E}\Big(\int_r^t f(s)dB_s^1|\mathcal{F}_r\Big) + \mathbb{E}\Big(\int_r^t f(s)dB_s^1 \int_r^t g(s)dB_s^2|\mathcal{F}_r\Big)\\
&= \int_0^r f(s)dB_s^1 \int_0^r g(s)dB_s^2.
\end{aligned}$$

That is, $\{\int_0^t f(s)dB_s^1 \int_0^t g(s)dB_s^2\}_{0\le t\le T}$ is a martingale with respect to $\{\mathcal{F}_t\}$. Hence, by the Doob martingale stopping theorem,

$$\mathbb{E}\Big(\int_0^\tau f(s)dB_s^1 \int_0^\tau g(s)dB_s^2|\mathcal{F}_\rho\Big) = \int_0^\rho f(s)dB_s^1 \int_0^\rho g(s)dB_s^2. \tag{1.34}$$

Now the required assertion (1.31) follows from (1.34) easily. The proof of the lemma, hence of Theorem 1.32 is now complete. □

We shall finally extend the stochastic integral to a larger class of stochastic processes. Let $\mathcal{L}^2(\mathbb{R}_+;\mathbb{R}^{n\times m})$ denote the family of all $n\times m$-matrix-valued measurable $\{\mathcal{F}_t\}$-adapted processes $f=\{f(t)\}_{t\ge 0}$ such that

$$\int_0^T |f(t)|^2 dt < \infty \quad \text{a.s. for every } T>0.$$

Let $\mathcal{M}^2(\mathbb{R}_+;\mathbb{R}^{n\times m})$ denote the family of all processes $f\in\mathcal{L}^2(\mathbb{R}_+;\mathbb{R}^{n\times m})$ such that

$$\mathbb{E}\int_0^T |f(t)|^2 dt < \infty \quad \text{for every } T>0.$$

Clearly, if $f\in\mathcal{M}^2(\mathbb{R}_+;\mathbb{R}^{n\times m})$, then $\{f(t)\}_{0\le t\le T}\in\mathcal{M}^2([0,T];\mathbb{R}^{n\times m})$ for every $T>0$. Hence, the indefinite integral $\int_0^t f(s)dB_s$, $t\ge 0$ is well defined, and it is an $\mathbb{R}^n$-valued continuous square-integrable martingale. However, we aim to define the integral for all processes in $\mathcal{L}^2(\mathbb{R}_+;\mathbb{R}^{n\times m})$. Let $f\in\mathcal{L}^2(\mathbb{R}_+;\mathbb{R}^{n\times m})$. For each integer $k\ge 1$, define the stopping time

$$\tau_k = k\wedge \inf\Big\{t\ge 0: \int_0^t |f(s)|^2 ds \ge k\Big\}.$$

Clearly, $\tau_k \uparrow \infty$ a.s. Moreover, $\{f(t)I_{[[0,\tau_k]]}(t)\}_{t\geq 0} \in \mathcal{M}^2(\mathbb{R}_+;\mathbb{R}^{n\times m})$ so the integral

$$I_k(t) = \int_0^t f(s)I_{[[0,\tau_k]]}(s)dB_s, \quad t \geq 0$$

is well defined. Note that for $1 \leq k \leq j$ and $t \geq 0$,

$$\begin{aligned} I_j(t \wedge \tau_k) &= \int_0^{t\wedge\tau_k} f(s)I_{[[0,\tau_j]]}(s)dB_s = \int_0^t f(s)I_{[[0,\tau_j]]}(s)I_{[[0,\tau_k]]}(s)dB_s \\ &= \int_0^t f(s)I_{[[0,\tau_k]]}(s)dB_s = I_k(t), \end{aligned}$$

which implies

$$I_j(t) = I_k(t), \quad 0 \leq t \leq \tau_k.$$

So we may define the indefinite stochastic integral $\{I(t)\}_{t\geq 0}$ as

$$I(t) = I_k(t) \quad \text{on } 0 \leq t \leq \tau_k. \tag{1.35}$$

Definition 1.33 Let $f = \{f(t)\}_{t\geq 0} \in \mathcal{L}^2(\mathbb{R}_+;\mathbb{R}^{n\times m})$. The indefinite Itô integral of f with respect to $\{B_t\}$ is the $\mathbb{R}^n$-valued process $\{I(t)\}_{t\geq 0}$ defined by (1.35). As before, we usually write $\int_0^t f(s)dB_s$ instead of $I(t)$.

It is clear that the Itô integral $\int_0^t f(s)dB_s$, $t \geq 0$ is an $\mathbb{R}^n$-valued continuous local martingale.

1.6 Itô's Formula

In the previous section we defined the Itô stochastic integrals. However the basic definition of the integrals is not very convenient in evaluating a given integral. This is similar to the situation for classical Lebesgue integrals, where we do not use the basic definition but rather the fundamental theorem of calculus plus the chain rule in the explicit calculations. For example, it is very easy to use the chain rule to calculate $\int_0^t \cos(s)ds = \sin(t)$ but not so if you use the basic definition. In this section we shall establish the stochastic version of the chain rule for the Itô integrals, which is known as Itô's formula. We shall see in this book that Itô's formula is not only useful in evaluating the Itô integrals but, more importantly, it plays a key role in stochastic analysis.

Let $B(t) = (B_1(t), \cdots, B_m(t))^T$, $t \geq 0$ be an m-dimensional Brownian motion defined on the complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$.

Definition 1.34 An n-dimensional Itô process is an $\mathbb{R}^n$-valued continuous adapted process $x(t) = (x_1(t), \cdots, x_n(t))^T$ on $t \geq 0$ of the form

$$x(t) = x(0) + \int_0^t f(s)ds + \int_0^t g(s)dB(s),$$

where $f = (f_1, \cdots, f_n)^T \in \mathcal{L}^1(\mathbb{R}_+; \mathbb{R}^n)$ and $g = (g_{ij})_{n\times m} \in \mathcal{L}^2(\mathbb{R}_+; \mathbb{R}^{n\times m})$. We shall say that $x(t)$ has a stochastic differential $dx(t)$ on $t \geq 0$ given by

$$dx(t) = f(t)dt + g(t)dB(t). \tag{1.36}$$

We shall sometimes speak of Itô process $x(t)$ and its stochastic differential $dx(t)$ on $t \in [a, b]$, and the meaning is apparent.

Let $C^{2,1}(\mathbb{R}^n \times \mathbb{R}_+; \mathbb{R})$ denote the family of all real-valued functions $V(x, t)$ defined on $\mathbb{R}^n \times \mathbb{R}_+$ such that they are continuously twice differentiable in x and once in t. If $V \in C^{2,1}(\mathbb{R}^n \times \mathbb{R}_+; \mathbb{R})$, we set

$$V_t = \frac{\partial V}{\partial t}, \quad V_x = \left(\frac{\partial V}{\partial x_1}, \cdots, \frac{\partial V}{\partial x_n}\right),$$

$$V_{xx} = \left(\frac{\partial^2 V}{\partial x_i \partial x_j}\right)_{n\times n} = \begin{pmatrix} \frac{\partial^2 V}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 V}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 V}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 V}{\partial x_n \partial x_n} \end{pmatrix}.$$

Clearly, when $V \in C^{2,1}(\mathbb{R} \times \mathbb{R}_+; \mathbb{R})$, we have $V_x = \frac{\partial V}{\partial x}$ and $V_{xx} = \frac{\partial^2 V}{\partial x^2}$.

We are now ready to state the well-known Itô formula.

Theorem 1.35 ***(Itô's formula)*** *Let $x(t)$ be an n-dimensional Itô process on $t \geq 0$ with the stochastic differential*

$$dx(t) = f(t)dt + g(t)dB(t),$$

where $f \in \mathcal{L}^1(\mathbb{R}_+; \mathbb{R}^n)$ and $g \in \mathcal{L}^2(\mathbb{R}_+; \mathbb{R}^{n\times m})$. Let $V \in C^{2,1}(\mathbb{R}^n \times \mathbb{R}_+; \mathbb{R})$. Then $V(x(t), t)$ is a real-valued Itô process with its stochastic differential given by

$$\begin{aligned} dV(x(t), t) = \Big[&V_t(x(t), t) + V_x(x(t), t)f(t) \\ &+\frac{1}{2}\text{trace}\big(g^T(t)V_{xx}(x(t), t)g(t)\big)\Big]dt + V_x(x(t), t)g(t)dB(t) \quad a.s. \end{aligned} \tag{1.37}$$

The proof can be found in many books e.g. [Mao (1997)] so is left to the reader as an exercise. Let us now introduce formally a multiplication table:

$$
\begin{aligned}
dtdt &= 0, \quad & dB_i dt &= 0, \\
dB_i dB_i &= dt, \quad & dB_i dB_j &= 0 \quad \text{if } i \neq j.
\end{aligned}
$$

Then, for example,

$$
dx_i(t)dx_j(t) = \sum_{k=1}^{m} g_{ik}(t)g_{jk}(t)dt. \tag{1.38}
$$

Moreover, the Itô formula can be written as

$$
\begin{aligned}
dV(x(t),t) &= V_t(x(t),t)dt + V_x(x(t),t)dx(t) \\
&\quad + \frac{1}{2}dx^T(t)V_{xx}(x(t),t)dx(t).
\end{aligned} \tag{1.39}
$$

Note that if $x(t)$ were continuously differentiable in t, then (by the classical calculus formula for total derivatives) the term $\frac{1}{2}dx^T(t)V_{xx}(x(t),t)dx(t)$ would not appear. For example, let $V(x,t) = x_1x_2$, then (1.38) and (1.39) yield

$$
\begin{aligned}
d[x_1(t)x_2(t)] &= x_1(t)dx_2(t) + x_2(t)dx_1(t) + dx_1dx_2 \qquad (1.40) \\
&= x_1(t)dx_2(t) + x_2(t)dx_1(t) + \sum_{k=1}^{m} g_{1k}(t)g_{2k}(t)dt,
\end{aligned}
$$

which is different from the classical formula of integration by parts $d(uv) = vdu + udv$ if both u, v are differentiable. More clearly, we have

$$
d[\sin^2(t)] = 2\sin(t)d\sin(t)
$$

but we don't have, if $B(t)$ is a scalar Brownian motion,

$$
d[B^2(t)] = 2B(t)dB(t),
$$

instead, we have

$$
d[B^2(t)] = 2B(t)dB(t) + dt.
$$

However we do have the stochastic version of integration by parts formula which is similar to the classical one.

Theorem 1.36 ***(Integration by parts formula)*** *Let $x(t)$, $t \geq 0$ be a one-dimensional Itô process with the stochastic differential*

$$dx(t) = f(t)dt + g(t)dB(t),$$

where $f \in \mathcal{L}^1(\mathbb{R}_+;\mathbb{R})$ and $g \in \mathcal{L}^2(\mathbb{R}_+;R^{1\times m})$. Let $y(t)$, $t \geq 0$ be a real-valued continuous adapted process of finite variation. Then

$$d[x(t)y(t)] = y(t)dx(t) + x(t)dy(t), \tag{1.41}$$

that is

$$x(t)y(t) - x(0)y(0) = \int_0^t y(s)[f(s)ds + g(s)dB(s)] + \int_0^t x(s)dy(s), \tag{1.42}$$

where the last integral is the Lebesgue–Stieltjes integral.

Given $V \in C^{2,1}(\mathbb{R}^n \times \mathbb{R}_+;\mathbb{R})$, define an operator $LV : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}$ by

$$LV(x,t) = V_t(x,t) + V_x(x,t)f(t) + \frac{1}{2}\text{trace}\big(g^T(t)V_{xx}(x,t)g(t)\big), \tag{1.43}$$

which is called the *diffusion operator* of the Itô process (1.36) associated with the $C^{2,1}$-function V. With this diffusion operator, the Itô formula (1.37) can be written as

$$dV(x(t),t) = LV(x(t),t)dt + V_x(x(t),t)g(t)dB(t) \quad a.s. \tag{1.44}$$

If $0 \leq \tau \leq \rho < \infty$ are two stopping times such that

$$\mathbb{E}\int_\tau^\rho |V_x(x(t),t)g(t)|^2 dt < \infty$$

and the integrations involved below are all finite, then

$$\mathbb{E}V(x(\rho),\rho) - \mathbb{E}V(x(\tau),\tau) = \mathbb{E}\int_\tau^\rho LV(x(t),t)dt. \tag{1.45}$$

Let us now give a number of examples to illustrate the use of Itô's formula in evaluating the stochastic integrals.

Example 1.37 Let $B(t)$ be a one-dimensional Brownian motion. To compute the stochastic integral

$$\int_0^t e^{-s/2+B(s)}dB(s),$$

we let $V(x,t) = e^{-t/2+x}$ and $x(t) = B(t)$, and then, by the Itô formula, we compute

$$\begin{aligned} d\Big[e^{-t/2+B(t)}\Big] &= -\frac{1}{2}e^{-t/2+B(t)}dt + e^{-t/2+B(t)}dB(t) + \frac{1}{2}e^{-t/2+B(t)}dt \\ &= e^{-t/2+B(t)}dB(t). \end{aligned}$$

That yields

$$\int_0^t e^{-s/2+B(s)}dB(s) = e^{-t/2+B(t)} - 1.$$

Rewrite this as

$$e^{-\frac{1}{2}p^2 t+pB(t)} = 1 + \int_0^t e^{-\frac{1}{2}p^2 s+pB(s)}dB(s).$$

It can be shown (see Exercise 1.10) that

$$\mathbb{E}\int_0^t |e^{-\frac{1}{2}p^2 s+pB(s)}|^2 ds = \frac{1}{p^2}[e^{p^2 t} - 1], \quad \forall t \geq 0.$$

So $\int_0^t e^{-\frac{1}{2}p^2 s+pB(s)}dB(s)$ is a martingale on $t \geq 0$ vanishing at $t = 0$. We hence obtain the following important result.

Theorem 1.38 *(The exponential martingale formula) Let $B(t)$ be a one-dimensional Brownian motion and $p > 0$. Then $e^{-\frac{1}{2}p^2 t+pB(t)}$ is a martingale on $t \geq 0$ with initial value 1 and hence for any bounded stopping time τ,*

$$\mathbb{E}\big[e^{-\frac{1}{2}p^2\tau+pB(\tau)}\big] = 1.$$

Example 1.39 Let $B(t)$ be a one-dimensional Brownian motion. What is the integration of the Brownian sample path over the time interval $[0,t]$, i.e. $\int_0^t B(s)ds$? The integration by parts formula yields

$$d[tB(t)] = B(t)dt + tdB(t).$$

Therefore

$$\int_0^t B(s)ds = tB(t) - \int_0^t sdB(s).$$

On the other hand, we may apply Itô's formula to $B^3(t)$ to obtain

$$dB^3(t) = 3B^2(t)dB(t) + 3B(t)dt,$$

which gives the alternative

$$\int_0^t B(s)ds = \frac{1}{3}B^3(t) - \int_0^t B^2(s)dB(s).$$

Example 1.40 Let $B(t)$ be an m-dimensional Brownian motion. Let $V : \mathbb{R}^m \to \mathbb{R}$ be C^2. Then Itô's formula implies

$$V(B(t)) = V(0) + \frac{1}{2}\int_0^t \Delta V(B(s))ds + \int_0^t V_x(B(s))dB(s),$$

where $\Delta = \sum_{i=1}^m \frac{\partial^2}{\partial x_i^2}$ is the Laplace operator. In particular, let V be a quadratic function, namely $V(x) = x^T Qx$, where Q is an $m \times m$ matrix. Then

$$B^T(t)QB(t) = \text{trace}(Q)t + \int_0^t B^T(s)(Q + Q^T)dB(s).$$

Example 1.41 Let $x(t)$ be an n-dimensional Itô process as given by Definition 1.34. Let Q be an $n \times n$ matrix. Then

$$\begin{aligned} &x^T(t)Qx(t) - x^T(0)Qx(0) \\ &= \int_0^t \Big(x^T(s)(Q + Q^T)f(s) + \frac{1}{2}\text{trace}[g^T(s)(Q + Q^T)g(s)]\Big)ds \\ &+ \int_0^t x^T(s)(Q + Q^T)g(s)dB(s). \end{aligned}$$

1.7 Markov Processes

In this section we will recall some basic facts about a Markov process. An n-dimensional $\mathcal{F}_t$-adapted process $X = \{X_t\}_{t\geq 0}$ is called a *Markov process* if the following *Markov property* is satisfied: for all $0 \leq s \leq t < \infty$ and $A \in \mathcal{B}(\mathbb{R}^n)$,

$$\mathbb{P}(X(t) \in A|\mathcal{F}_s) = \mathbb{P}(X(t) \in A|X(s)).$$

This is equivalent to the following one: for any bounded Borel measurable function $\varphi : \mathbb{R}^n \to \mathbb{R}$ and $0 \leq s \leq t < \infty$,

$$\mathbb{E}(\varphi(X(t))|\mathcal{F}_s) = \mathbb{E}(\varphi(X(t))|X(s)).$$

The *transition probability or function* of the Markov process is a function $P(s, x; t, A)$, defined on $0 \le s \le t < \infty$, $x \in \mathbb{R}^n$ and $A \in \mathcal{B}(\mathbb{R}^n)$, with the following properties:

(1) For every $0 \le s \le t < \infty$ and $A \in \mathcal{B}(\mathbb{R}^n)$,

$$P(s, X(s); t, A) = \mathbb{P}(X(t) \in A | X(s)).$$

(2) $P(s, x; t, \cdot)$ is a probability measure on $\mathcal{B}(\mathbb{R}^n)$ for every $0 \le s \le t < \infty$ and $x \in \mathbb{R}^n$.

(3) $P(s, \cdot; t, A)$ is Borel measurable for every $0 \le s \le t < \infty$ and $A \in \mathcal{B}(\mathbb{R}^n)$.

(4) The Kolmogorov–Chapman equation

$$P(s, x; t, A) = \int_{\mathbb{R}^n} P(u, y; t, A) P(s, x; u, dy)$$

holds for any $0 \le s \le u \le t < \infty, x \in \mathbb{R}^n$ and $A \in \mathcal{B}(\mathbb{R}^n)$.

Clearly, in terms of transition probability, the Markov property becomes

$$\mathbb{P}(X(t) \in A | \mathcal{F}_s) = P(s, X(s); t, A).$$

We shall use the notion

$$\mathbb{P}(X(t) \in A | X(s) = x) = P(s, x; t, A)$$

and

$$\mathbb{E}_{s,x} \varphi(X(t)) = \int_{\mathbb{R}^n} \varphi(y) P(s, x; t, dy).$$

A Markov process $X = \{X(t)\}_{t \ge 0}$ is said to be *homogeneous* if its transition probability $P(s, x; t, A)$ is stationary, namely

$$P(s + u, x; t + u, A) = P(s, x; t, A)$$

for all $0 \le s \le t < \infty, x \in \mathbb{R}^n, u \ge 0$ and $A \in \mathcal{B}(\mathbb{R}^n)$. In this case, the transition probability $P(s, x; t, A)$ depends only on $t-s$ and it can be simply written as $P(0, x; t, A) = P(t, x, A)$. Moreover, the Kolmogorov–Chapman equation becomes

$$P(t + s, x, A) = \int_{\mathbb{R}^n} P(s, y, A) P(t, x, dy).$$

Furthermore, with the notation

$$\mathbb{E}_x(\varphi(X(t))) = \int_{\mathbb{R}^n} \varphi(y)P(t, x, dy),$$

the Markov property becomes

$$\mathbb{E}(\varphi(X(t))|\mathcal{F}_s) = \mathbb{E}_{X(s)}\varphi(X(t-s)).$$

An n-dimensional process $\{X_t\}_{t\geq 0}$ is called a *strong Markov process* if the following *strong Markov property* is satisfied: for any bounded Borel measurable function $\varphi : \mathbb{R}^n \to \mathbb{R}$, any finite $\{\mathcal{F}_t\}$-stopping time τ and $t \geq 0$,

$$\mathbb{E}(\varphi(X(t+\tau))|\mathcal{F}_\tau) = \mathbb{E}(\varphi(X(t+\tau))|X(\tau)).$$

Especially, in the homogeneous case, this becomes

$$\mathbb{E}(\varphi(X(t+\tau))|\mathcal{F}_\tau) = \mathbb{E}_{X(\tau)}\varphi(X(t)).$$

A stochastic process $X = \{X(t)\}_{t\geq 0}$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with values in a countable set Ξ (to be called the *state space* of the process), is called a *continuous-time Markov chain* if for any finite set $0 \leq t_1 < t_2 < \cdots < t_n < t_{n+1}$ of "times", and corresponding set $i_1, i_2, \ldots, i_{n-1}, i, j$ of states in Ξ such that $\mathbb{P}\{X(t_n) = i, X(t_{n-1}) = i_{n-1}, \ldots, X(t_1) = i_1\} > 0$, we have

$$\begin{aligned}\mathbb{P}\{X(t_{n+1}) = j|X(t_n) = i, X(t_{n-1}) = i_{n-1}, \ldots, X(t_1) = i_1\}\\ = \mathbb{P}\{X(t_{n+1}) = j|X(t_n) = i\}.\end{aligned}$$

If for all s, t such that $0 \leq s \leq t < \infty$ and all $i, j \in \Xi$ the conditional probability $\mathbb{P}\{X(t) = j|X(s) = i\}$ depends only on $t - s$, we say that the process $X = \{X(t)\}_{t\geq 0}$ is *homogeneous*. In this case, then, $\mathbb{P}\{X(t) = j|X(s) = i\} = \mathbb{P}\{X(t-s) = j|X(0) = i\}$, and the function

$$P_{ij}(t) =: \mathbb{P}\{X(t) = j|X(s) = i\}, \quad i, j \in \Xi, t \geq 0,$$

is called the *transition function* or *transition probability* of the process. The function $P_{ij}(t)$ is called *standard* if $\lim_{t\to 0} P_{ii}(t) = 1$ for all $i \in \Xi$.

Theorem 1.42 **[Anderson (1991)]** *Let $P_{ij}(t)$ be a standard transition function, then $\gamma_i := \lim_{t\to 0}[1 - P_{ii}(t)]/t$ exists (but may be ∞) for all $i \in \Xi$.*

A state $i \in \Xi$ is said to be *stable* if $\gamma_i < \infty$.

Theorem 1.43 [**Anderson (1991)**] *Let $P_{ij}(t)$ be a standard transition function, and let j be a stable state. Then $\gamma_{ij} = P'_{ij}(0)$ exists and is finite for all $i \in \Xi$.*

Let $\gamma_{ii} = -\gamma_i$ and $\Gamma = (\gamma_{ij})_{i,j\in\Xi}$. Γ is called the *generator* of the Markov chain. If the state space is *finite* which we can take to be $\mathbb{S} = \{1, 2, \ldots, N\}$, then the process is called a continuous-time *finite* Markov chain. Throughout this book, we assume that all Markov chains are finite and all states are stable. For such a Markov chain, almost every sample path is a right continuous step function.

Theorem 1.44 [**Anderson (1991)**] *Let $P(t) = (P_{ij}(t))_{N\times N}$ be the transition probability matrix and $\Gamma = (\gamma_{ij})_{N\times N}$ be the generator of a finite Markov chain. Then*

$$P(t) = e^{t\Gamma}.$$

It is useful to emphasise that a continuous-time Markov chain $X(t)$ with generator $\Gamma = \{\gamma_{ij}\}_{N\times N}$ can be represented as a stochastic integral with respect to a Poisson random measure (see [Skorohod (1989)] and [Ghosh *et al.* (1997)]). Indeed, let Δ_{ij} be consecutive, left closed, right open intervals of the real line each having length γ_{ij} such that

$$\begin{aligned}
\Delta_{12} &= [0, \gamma_{12}), \\
\Delta_{13} &= [\gamma_{12}, \gamma_{12} + \gamma_{13}), \\
&\vdots \\
\Delta_{1N} &= \left[\sum_{j=2}^{N-1} \gamma_{1j}, \sum_{j=2}^{N} \gamma_{1j}\right), \\
\Delta_{21} &= \left[\sum_{j=2}^{N} \gamma_{1j}, \sum_{j=2}^{N} \gamma_{1j} + \gamma_{21}\right), \\
\Delta_{23} &= \left[\sum_{j=2}^{N} \gamma_{1j} + \gamma_{21}, \sum_{j=2}^{N} \gamma_{1j} + \gamma_{21} + \gamma_{23}\right), \\
&\vdots \\
\Delta_{2N} &= \left[\sum_{j=2}^{N} \gamma_{1j} + \sum_{j=1, j\neq 2}^{N-1} \gamma_{2j}, \sum_{j=2}^{N} \gamma_{1j} + \sum_{j=1, j\neq 2}^{N} \gamma_{2j}\right)
\end{aligned}$$

and so on. Define a function

$$h : \mathbb{S} \times \mathbb{R} \to \mathbb{R}$$

by

$$h(i,y) = \begin{cases} j-i & \text{if } y \in \Delta_{ij}, \\ 0 & \text{otherwise.} \end{cases} \tag{1.46}$$

Then

$$dX(t) = \int_{\mathbb{R}} h(X(t-), y)\nu(dt, dy),$$

with initial condition $X(0) = i_0$, where $\nu(dt, dy)$ is a Poisson random measure with intensity $dt \times \mu(dy)$, in which μ is the Lebesgue measure on $\mathbb{R}$.

1.8 Generalised Itô's Formula

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a complete probability space with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions (i.e. it is increasing and right continuous while $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets). Let $B(t) = (B_t^1, \ldots, B_t^m)^T$ be an m-dimensional Brownian motion defined on the probability space. Let $r(t), t \geq 0$, be a right-continuous Markov chain on the probability space taking values in a finite state space $\mathbb{S} = \{1, 2, \ldots, N\}$ with generator $\Gamma = (\gamma_{ij})_{N\times N}$ given by

$$\mathbb{P}\{r(t+\delta) = j | r(t) = i\} = \begin{cases} \gamma_{ij}\delta + o(\delta) & \text{if } i \neq j, \\ 1 + \gamma_{ii}\delta + o(\delta) & \text{if } i = j, \end{cases}$$

where $\delta > 0$. Here $\gamma_{ij} \geq 0$ is transition rate from i to j if $i \neq j$ while

$$\gamma_{ii} = -\sum_{j\neq i} \gamma_{ij}.$$

We assume that the Markov chain $r(\cdot)$ is independent of the Brownian motion $B(\cdot)$.

Let $x(t)$ be an n-dimensional Itô process on $t \geq 0$ with the stochastic differential

$$dx(t) = f(t)dt + g(t)dB(t),$$

where $f \in \mathcal{L}^1(\mathbb{R}_+; \mathbb{R}^n)$ and $g \in \mathcal{L}^2(\mathbb{R}_+; \mathbb{R}^{n\times m})$. The Itô formula established in Section 1.6 shows that a $C^{2,1}(\mathbb{R}^n \times \mathbb{R}_+; \mathbb{R}_+)$-function V maps the Itô process $x(t)$ into another Itô process $V(x(t), t)$. On the other hand, we will consider the paired process $(x(t), r(t))$ in this book and we need to know how a function $V : \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{S} \to \mathbb{R}$ will map $(x(t), r(t))$ into another process $V(x(t), t, r(t))$. For this purpose, let $C^{2,1}(\mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{S}; \mathbb{R})$ denote the family of all real-valued functions $V(x, t, i)$ on $\mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{S}$ which are continuously twice differentiable in x and once in t. If $V \in C^{2,1}(\mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{S}; \mathbb{R})$, define an operator LV from $\mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{S}$ to $\mathbb{R}$ by

$$\begin{aligned} LV(x,t,i) = & V_t(x,t,i) + V_x(x,t,i)f(t) \\ & + \frac{1}{2}\text{trace}[g^T(t)V_{xx}(x,t,i)g(t)], + \sum_{j=1}^{N} \gamma_{ij} V(x,t,j), \end{aligned} \tag{1.47}$$

where

$$V_t(x,t,i) = \frac{\partial V(x,t,i)}{\partial t}, \quad V_x(x,t,i) = \left(\frac{\partial V(x,t,i)}{\partial x_1}, \ldots, \frac{\partial V(x,t,i)}{\partial x_n} \right)$$

and

$$V_{xx}(x,t,i) = \left(\frac{\partial^2 V(x,t,i)}{\partial x_i \partial x_j} \right)_{n \times n}.$$

The following formula, known as the generalised Itô formula, reveals how V maps the paired process $(x(t), r(t))$ into a new process $V(x(t), t, r(t))$.

Theorem 1.45 *If $V \in C^{2,1}(\mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{S}; \mathbb{R})$, then for any $t \geq 0$*

$$\begin{aligned} & V(x(t), t, r(t)) \\ = & V(x(0), 0, r(0)) + \int_0^t LV(x(s), s, r(s))ds \\ & + \int_0^t V_x(x(s), s, r(s))g(x(s), s, r(s))dB(s) \\ & + \int_0^t \int_R (V(x(s), s, i_0 + h(r(s), l)) - V(x(s), s, r(s)))\mu(ds, dl), \end{aligned} \tag{1.48}$$

where the function h is defined by (1.46) and $\mu(ds, dl) = \nu(ds, dl) - \mu(dl)ds$ is a martingale measure while ν and μ have been defined in the end of Section 1.7.

The proof can be found in [Skorohod (1989)] on page 104 (namely, the proof of Lemma 3 there). In particular, taking the expectation on both sides of (1.48), we get the following useful lemma.

Lemma 1.9 *Let $V \in C^{2,1}(\mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{S}; \mathbb{R}_+)$ and τ_1, τ_2 be bounded stopping times such that $0 \le \tau_1 \le \tau_2$ a.s. If $V(x(t), t, r(t))$ and $LV(x(t), t, r(t))$ etc, are bounded on $t \in [\tau_1, \tau_2]$ with probability 1, then*

$$\mathbb{E}V(x(\tau_2), \tau_2, r(\tau_2)) = \mathbb{E}V(x(\tau_1), \tau_1, r(\tau_1)) + \mathbb{E}\int_{\tau_1}^{\tau_2} LV(x(s), s, r(s))ds. \tag{1.49}$$

1.9 Exercises

1.1 Let B_t, $0 \le t \le T$ be a scalar Brownian motion and let k be a positive integer. Set $\Delta = T/k$ and $t_j = j\Delta$ for $0 \le j \le k$. Show

$$\mathbb{E}(B_{t_j} - B_{t_{j-1}})^2 = \Delta \quad \text{and} \quad \mathbb{E}(B_{t_j} - B_{t_{j-1}})^4 = 3\Delta^2.$$

Deduce that

$$\mathbb{E}[(B_{t_j} - B_{t_{j-1}})(B_{t_i} - B_{t_{j-i}})] = 0 \quad \text{for } i \neq j.$$

Hence show that $\sum_{j=1}^k (B_{t_j} - B_{t_{j-1}})^2$ has mean T. Next, show that

$$\mathbb{E}\Big(\sum_{j=1}^{k}(B_{t_j} - B_{t_{j-1}})^2\Big)^2 = T^2 + 2T\Delta$$

and hence deduce that $\sum_{j=1}^k (B_{t_j} - B_{t_{j-1}})^2$ has variance of $O(\Delta)$.

1.2 Show that a scalar Brownian motion B_t has a scaling property: for any fixed $c \neq 0$,

$$X_t := \frac{B_{c^2 t}}{c}, \quad t \ge 0,$$

is a Brownian motion with respect to the filtration $\{\mathcal{F}_{c^2 t}\}$.

1.3 Prove properties (a)–(c) listed on page 21 for a Brownian motion.

1.4 Prove Lemma 1.4.

1.5 In Step 2 of the proof of Lemma 1.5, $\phi_k(t)$ is defined for the given process $\varphi(t)$. Show that for every ω, $\phi_k(\cdot, \omega)$ is continuous and and

$$\lim_{k \to \infty} \int_a^b |\varphi(t,\omega) - \phi_k(t,\omega)|^2 dt = 0.$$

1.6 Prove Theorem 1.21.

1.7 Prove Lemma 1.7.

1.8 Prove the Itô formula. (You may refer to [Mao (1997)].)

1.9 Let Q be an $m \times m$ matrix and define $V(x) = x^T Q x$ for $x \in \mathbb{R}^m$. Show that $V_x(x) = x^T(Q + Q^T)$ and $V_{xx} = Q + Q^T$ and hence show by the Itô formula that

$$B^T(t)QB(t) = \text{trace}(Q)t + \int_0^t B^T(s)(Q + Q^T)dB(s),$$

where $B(t)$ is an m-dimensional Brownian motion.

1.10 Let $\xi \sim N(0, \sigma^2)$ (i.e. a normal distribution with mean 0 and variance σ^2). Show

$$\mathbb{E}(e^{\xi}) = e^{\frac{1}{2}\sigma^2}.$$

Hence show that if $B(t)$ is a scalar Brownian motion and $p > 0$, then

$$\int_0^t \mathbb{E}|e^{-\frac{1}{2}p^2 s + pB(s)}|^2 ds = \frac{1}{p^2}[e^{p^2 t} - 1], \quad \forall t \geq 0.$$