STOCHASTIC DIFFERENTIAL EQUATIONS

Adair Antonio da Silva Neto

June 24, 2023

Contents

Chapter 1

Introduction

An equation that models an evolution process that contains 'noise', with a certain randomness, is an equation of the form:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = b(t, X_t) + \sigma(t, X_t) \cdot \text{`white noise'} \tag{1.1}$$

Problems like this appears naturally in biology (populational growth models), physics (charge in electrical circuits), engineering (filtering problems, like the Kalman's Filter exihibted in the Figure 1.1) and finance (optimal stopping, optimal portfolio and option pricing).

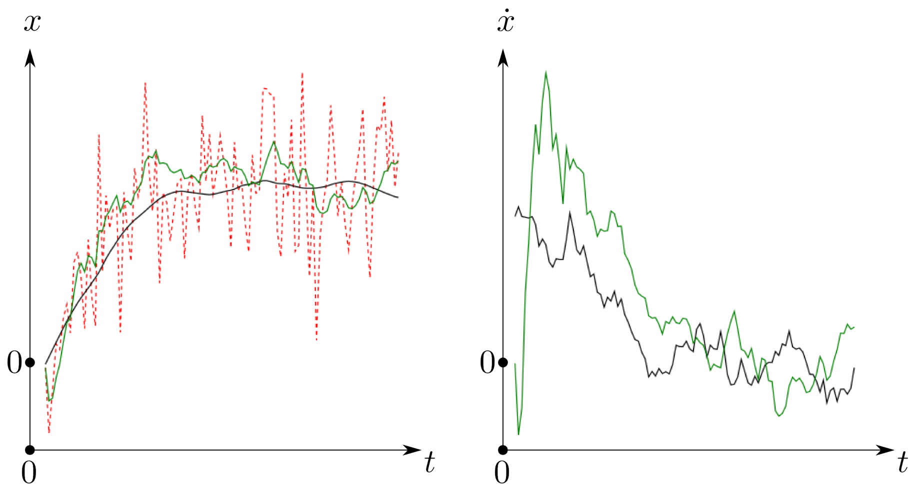

Figure 1.1: Kalman's Filter: Observed data; Filtered process; Real data [Wik22].

From the mathematical viewpoint, we need to make sense of this kind of equations, since with the usual tools of calculus, it is not possible to treat them. Our goal then, in this work, is to give the foundations that allow us to treat this equations. In particular, firstly we're going to define the Itô's Integral and then we'll be ready to tackle stochastic differential equations.

Chapter 2

Probability Theory

2.1 Probability Spaces

A probability space starts with a set Ω called the **sample space**, which is, intuitively, a list of all possible outcomes of an experiment. Each $\omega \in \Omega$ is a **sample point**, and each subset $A \subseteq \Omega$ is an **event**.

To filter 'well-behaved' subsets, where it will be possible to measure a probability, the following definition is necessary.

Definition 2.1.1 (σ-algebra)**.** If Ω is a set, then a σ**-algebra** $\mathfrak{F}$ on Ω is a family of subsets of Ω satisfying:

1. $\emptyset \in \mathfrak{F}$.
2. If $A \in \mathfrak{F}$, then $A^c \in \mathfrak{F}$.
3. If $A_1, A_2, \ldots \in \mathfrak{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathfrak{F}$.

The pair $(\Omega, \mathfrak{F})$ is said to be a **measurable space**, and any subset $B \subseteq \Omega$ that also belongs to $\mathfrak{F}$ is called a **measurable set**.

Intuitively, a σ-algebra represents the information available at the time.

Example 2.1.1 (Examples of σ-algebras)**.** The following are σ-algebras.

1. The family $\{\emptyset, \Omega\}$ is called the **trivial σ-algebra**, and is the smallest one possible.
2. The power set $\mathfrak{P}(\Omega)$ is called the **discrete σ-algebra**, and is the largest one possible.
3. The family $\{\emptyset, \Omega, A, A^c\}$ is the **σ-algebra generated by the set** A and is usually denoted by $\mathfrak{F}_A$.

But the most important σ-algebra for probability theory is the Borel σ-algebra, which will be denoted by $\mathfrak{B}$. Taking $\Omega = \mathbb{R}$, the Borel σ-algebra is generated by the intersection of all σ-algebras containing the real line intervals.

Notice that the Borel σ-algebra contains all open sets, closed sets, and all their countable operations with union $\cup$, intersection $\cap$, and their complements ^{c}. This is the smallest σ-algebra containing all open subsets.

More generally, if U is the collection of all open subsets of a topological space Ω, then $\mathfrak{B}$ is called the **Borel σ-algebra** on Ω. The elements $B \in \mathfrak{B}$ are called **Borel sets**.

Given a measurable space, it's possible to assign each outcome to a probability.

Definition 2.1.2 (Probability Measure). Let $(\Omega, \mathfrak{F})$ be a measurable space. A **probability measure** P is the function

$$P : \mathfrak{F} \longrightarrow [0,1]$$

satisfying

1. $P(\emptyset) = 0$ and $P(\Omega) = 1$.
2. (σ-additivity). If $A_1, A_2, \ldots \in \mathfrak{F}$ and $A_i \cap A_j = \emptyset$, for $i \neq j$ (i.e. are mutually exclusive), then

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

And the triple $(\Omega, \mathfrak{F}, P)$ is called a **probability space**.

In order to restrict the functions to sets in the σ-algebra, the following definition will be needed.

Definition 2.1.3 ($\mathfrak{F}$-measurable function). Given a probability space $(\Omega, \mathfrak{F}, P)$, a function Y from the sample space Ω to $\mathbf{R}^n$ is called $\mathfrak{F}$-measurable if

$$Y^{-1}(U) = \{\omega \in \Omega : Y(\omega) \in U\} \in \mathfrak{F}$$

for all open sets $U \in \mathbf{R}^n$ (i.e. for all Borel sets $U \subseteq \mathbf{R}^n$). In other words, the inverse image of U is in the σ-algebra.

And to attach numerical values to each $\omega \in \Omega$, define the following function.

Definition 2.1.4 (Random variable). An $\mathfrak{F}$-measurable function $X : \Omega \longrightarrow \mathbf{R}^n$ is called a **random variable** on a complete probability space $(\Omega, \mathfrak{F}, P)$.

This definition means that if we know which event U in the $\mathfrak{F}$ has occurred, then we know which value of X has occurred.

Consider the measurable space $(\Omega, \mathfrak{P}(\Omega))$, and let X be a random variable with values x_i, $i = 1, 2, \ldots, k$. The sets

$$A_i = \{\omega \in \Omega : X(\omega) = x_i\} \subseteq \Omega$$

form a partition of Ω, and the σ-algebra generated by this partition is called the **σ-algebra generated by** X.

Notice that this is the smallest σ-algebra that contains all the sets A_i. This σ-algebra is often denoted by $\mathfrak{F}_X$. Intuitively, this σ-algebra represents all information available about the sample point ω by observing X.

2.2 Distribution and Density of a Random Variable

However, given only the probability measure P, it is not always immediate how to compute the probability of a given interval, set or value. For that purpose, the following three definitions will be useful.

Definition 2.2.1 (Distribution). The **distribution** of a random variable X is a function $\mu_X : \Omega \longrightarrow \mathbf{R}^n$ defined as

$$\mu_X(B) = P[X^{-1}(B)] = P[X \in B] = P(\{\omega : X(\omega) \in B\})$$

Definition 2.2.2 (Distribution Function). The **cumulative distribution function** (CDF) of a random variable X is defined as

$$F_X(x) = P[X \leq x] = P(\{\omega : X(\omega) \leq x\})$$

where $x \in \mathbf{R}$. Notice that F is non-decreasing, and right-continuous, and approaches 0 at $-\infty$ and 1 at $+\infty$.

Definition 2.2.3 (Joint Distribution). If X and Y are random variables, then their **joint distribution function** is

$$F_{X,Y}(x,y) = P[X \leq x, Y \leq y]$$

where $x, y \in \mathbf{R}$.

The distribution function gives the probability that the random variable X is on the interval $(-\infty, x]$. This also allows to look for an interval $[a,b]$ using $F_X(b) - F_X(a)$.

While the distribution of X returns the probability of an event $\{\omega : X(\omega) \in B\}$.

Definition 2.2.4 (Density function). A random variable X has a **probability density function** (PDF) $f : \mathbf{R}^n \longrightarrow \mathbf{R}$ if f is a measurable function and

$$F_X(x) = P[X \leq x] = \int_{-\infty}^{x} f(y)\, dy$$

2.3 Expected Value and Variance

Another useful tool for any given random variable is to know its mean value and how much it varies. This motivates the following.

Definition 2.3.1 (Expectation). Let $(\Omega, \mathfrak{F}, P)$ be a probability space and X a random variable. If $\int_\Omega |X(\omega)|\, dP(\omega) < \infty$, then the **expected value** (or **mean value** of X with respect to P is

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, dP(\omega) = \int_{\mathbf{R}^n} x\, d\mu_X(x)$$

Notice that the expectation is linear. I.e., if X and Y are integrable and a and b are constants, then

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

If f is a Borel measurable function and $\int_\Omega |f(X)|\,dP(\omega) < \infty$, then

$$\mathbb{E}[f(X)] = \int_\Omega f(X(\omega))\,dP(\omega) = \int_{\mathbf{R}^n} f(X)\,d\mu_X(x)$$

Theorem 2.3.1 (Chebychev's inequality)**.**

$$P[\,|X| \geq \lambda\,] \leq \frac{1}{\lambda^p}\mathbb{E}[\,|X|^p\,]$$

for all $\lambda \geq 0$.

Definition 2.3.2 (Variance)**.** Let $\mu := \mathbb{E}[X]$, i.e., the expected value of a random variable X. Then the **variance** of X is given by

$$\mathrm{Var}[X] = \mathbb{E}[\|X-\mu\|^2] = \int_\Omega |X-\mu|^2\,dP(\omega)$$

Definition 2.3.3 (Covariance)**.** Let X and Y be integral random variables. Let $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$. If XY is integrable, then the **covariance** of X and Y is:

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

Notice that $\mathrm{Var}[X] = \mathrm{Cov}[X,X]$.

The variance of X can be computed more simply using the following theorem.

Theorem 2.3.2. Let X be a random variable. Then,

1. $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.
2. $\mathrm{Var}[aX+b] = a^2\mathrm{Var}[X]$.
3. $\mathrm{Cov}[X,Y] = \mathrm{Cov}[Y,X]$.
4. $\mathrm{Cov}[aX+bY,Z] = a\mathrm{Cov}[X,Z] + b\mathrm{Cov}[Y,Z]$.
5. $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X,Y]$.

Definition 2.3.4 (Independence)**.** Two events $A, B \in \mathfrak{F}$ are **independent** if

$$P[A \cap B] = P[A] \cdot P[B]$$

More generally, any collection of events A_i, $i = 1, 2, \ldots$, is called independent if for any $n \in \mathbf{N}$ and any choice of indices i_k, $k = 1, 2, \ldots$,

$$P\left(\bigcap_{k=1}^{n} A_{i_k}\right) = \prod_{k=1}^{n} P\left(A_{i_k}\right)$$

Definition 2.3.5 (Conditional Expectation)**.** Let X and Y be random variables. Then, given that $f_Y(y) > 0$, the **conditional distribution** of X given $Y = y$ is given by the conditional density

$$f(x \mid y) = \frac{f(x,y)}{f_Y(y)}$$

And the **conditional expectation** of X given $Y = y$ is

$$\mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x f(x \mid y)\, dx$$

The intuition behind this definition is to build an estimate of the random variable X given the information available in Y.

Some important and very useful properties of conditional expectation are listed below.

Theorem 2.3.3. Let $X, Y : \Omega \longrightarrow \mathbf{R}^n$ be random variables with $\mathbb{E}[\,|X|\,] < \infty$ and $\mathbb{E}[\,|Y|\,] < \infty$, $\mathfrak{H} \subseteq \mathfrak{F}$ a σ-algebra, and let $a, b \in \mathbf{R}$.

1. $\mathbb{E}[aX + bY \mid \mathfrak{H}] = a\mathbb{E}[X \mid \mathfrak{H}] + b\mathbb{E}[Y \mid \mathfrak{H}]$.
2. $\mathbb{E}[\mathbb{E}[X \mid \mathfrak{H}]] = \mathbb{E}[X]$.
3. $\mathbb{E}[X \mid \mathfrak{H}] = X$ if X is $\mathfrak{H}$-measurable.
4. $\mathbb{E}[X \mid \mathfrak{H}] = \mathbb{E}[X]$ if X is independent of $\mathfrak{H}$.
5. $\mathbb{E}[Y \cdot X \mid \mathfrak{H}] = Y \cdot \mathbb{E}[X \mid \mathfrak{H}]$ if Y is $\mathfrak{H}$-measurable.
6. If $\mathfrak{G} \subseteq \mathfrak{H}$ is a σ-algebra, then

$$\mathbb{E}[X \mid \mathfrak{G}] = \mathbb{E}[\mathbb{E}[X \mid \mathfrak{H}] \mid \mathfrak{G}]$$

7. If $\varphi : \mathbf{R} \longrightarrow \mathbf{R}$ is convex and $\mathbb{E}[\,|\varphi(X)|\,] < \infty$, then

$$\varphi(\mathbb{E}[X \mid \mathfrak{H}]) \leq \mathbb{E}[\varphi(X) \mid \mathfrak{H}]$$

Definition 2.3.6 (Infinitely often)**.** If $A_1, A_2, \ldots, A_n, \ldots$ are events in the probability space, then the event

$$\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \{\omega \in \Omega : \omega \text{ belongs to infinitely many of the } A_n\}$$

is called A_n **infinitely often**, or simply 'A_n **i.o.**'.

The next lemma helps us check if some sequence of events occurs infinitely often.

Lemma 2.3.4 (Borel-Cantelli)**.** If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then

$$P(A_n \text{ i.o.}) = P\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\right) = 0$$

The most important distribution in this text is the **normal** (or **gaussian**) distribution.

Definition 2.3.7 (Normal Distribution)**.** If random variable X has mean μ, variance σ^2 and a density function of the form

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

then X is said to have a **normal distribution**, which we denote by $X \sim N(\mu, \sigma^2)$.

Notice that X can be normalized (i.e. transformed into a distribution of the form $N(0,1)$) by taking $Z = \frac{X-\mu}{\sigma}$.

Definition 2.3.8 (Lipschitz and Hölder Conditions)**.** A function f is said to be **Hölder continuous** of order α, $0 < \alpha \leq 1$, on an interval $[a,b] \subseteq \mathbf{R}$ if there exists a constant $K > 0$, such that for all $x, y \in [a,b]$ we have

$$|f(x) - f(y)| \leq K|x-y|^{\alpha}$$

If $\alpha = 1$, then the function f is said to be **Lipschitz continuous**.

Corollary 2.3.5. Let $X \in L^1(\mathbb{P})$, $\{\mathfrak{N}_k\}_{k=1}^{\infty}$ be an increasing family of σ-algebras, $\mathfrak{N}_k \subset \mathfrak{F}$ and define $\mathfrak{N}_\infty$ to be the σ-lagebra generated by $\{\mathfrak{N}_k\}_{k=1}^{\infty}$.

Then,

$$\mathbb{E}[X \mid \mathfrak{N}_k] \longrightarrow \mathbb{E}[X \mid \mathfrak{N}_\infty] \text{ as } k \to \infty$$

a.e. $\mathbb{P}$ and in $L^1(\mathbb{P})$.

2.4 Stochastic Processes

Definition 2.4.1 (Stochastic Process)**.** A **stochastic process** is a collection of random variables $\{X(t)\}$ parametrized by time $t \in T$.

For each point $\omega \in \Omega$, the mapping $t \longrightarrow X_t(\omega)$ is the respective **sample path**, also called **realization** or **trajectory**.

Definition 2.4.2 (Filtration)**.** A filtration $\mathbf{F}$ is a collection of σ-algebras

$$\mathbf{F} = \{\mathfrak{F}_0, \mathfrak{F}_1, \ldots, \mathfrak{F}_t, \ldots, \mathfrak{F}_T\}$$

such that $\mathfrak{F}_t \subseteq \mathfrak{F}_{t+1}$.

The idea behind this definition is to model the flow of information. As the time t passes, the observer has more information, and does not lose any previous data. Notice that this implies finer partitions of the sample space Ω.

Given a stochastic process $\{X(t)\}$, let $\mathfrak{F}_t$ be the σ-algebra generated by the random variables X_s, $s = 0, \ldots, t$. Since $\mathfrak{F}_t \subseteq \mathfrak{F}_{t+1}$, these σ-algebras form a filtration called the **natural filtration** of the process $\{X(t)\}$, and contain all available information of the process up to the time t.

Definition 2.4.3 (Martingale)**.** A stochastic process $\{M_t\}$ on $(\Omega, \mathfrak{F}, P)$ is called a **martingale** with respect to a filtration $\mathfrak{M}_t$ if

1. M_t is $\mathfrak{M}_t$-measurable for all t.
2. $\mathbb{E}[\,|M_t|\,] < \infty$ for all t.
3. $\mathbb{E}[M_s \mid \mathfrak{M}_t] = M_t$ for all $s \geq t$.

Intuitively, a martingale is a stochastic process in which the future has no tendency to go up or down. In other words, the expected value of any time in the future is equal to the value of the process at the present time.

Theorem 2.4.1 (Doob's martingale inequality). If M_t is a martingale such that the mapping $t \longrightarrow M_t(\omega)$ is continuous a.s., then for all $p \geq 1$, $T \geq 0$ and $\lambda > 0$,

$$P\left[\sup_{0 \leq t \leq T} |M_t| \geq \lambda\right] \leq \frac{1}{\lambda^p}\mathbb{E}[\,|M_T|^p\,]$$

2.5 Brownian Motion

In 1828 the Scottish botanist Robert Brown described the irregular motion of pollen grains suspended in fluid, that is now known as the Brownian Motion. To describe this mathematically, we use a stochastic process $B_t(\omega)$, which is the position at time t of the particle ω.

This can also be understood as the model for the cumulative effect of 'noise'. And is also called Wiener process, after N. Wiener, who formalized mathematically the Brownian motion.

Definition 2.5.1 (Brownian Motion). A Brownian motion $\{B(t)\}$ is a stochastic process with the following properties:

1. (Independence of increments). $B(t) - B(s)$, for $t > s$, are independent. That means that the direction that the particle will go does not depend on the past.
2. (Normal increments). $B(t) - B(s)$ has normal distribution with mean 0 and variance $t - s$. Notice that by taking $s = 0$ we have $B(t) - B(0) \sim N(0, t)$.
3. (Continuity of paths). $B(t)$, $t \geq 0$, are continuous functions of t.

The existence and continuity of Brownian motion is proved using Kolmogorov's extension and continuity theorem [Oks13].

Theorem 2.5.1 (Quadratic variation). If $t \geq s$,

$$\mathbb{E}[(B_t - B_s)^2] = t - s$$

More generally (in $\mathbf{R}^n$):

$$\mathbb{E}[(B_t - B_s)^2] = n(t - s)$$

Theorem 2.5.2.

$$\mathbb{E}[f(B_t)] = \frac{1}{\sqrt{2\pi t}} \int_{\mathbf{R}} f(x) e^{-\frac{x^2}{2t}}\, dx$$

Theorem 2.5.3. The expected value of odd moments of B_t (other than one) are zero. And the even moments are given by

$$\mathbb{E}[B_t^{2k}] = \frac{(2k)!}{2^k k!} t^k$$

Last but not least, two important identities, useful for further calculations are:

$$B_j(B_{j+1} - B_j) = \frac{1}{2}[B_{j+1}^2 - B_j^2 - (B_{j+1} - B_j)^2]$$

$$B_j^2(B_{j+1} - B_j) = \frac{1}{3}(B_{j+1}^3 - B_j^3) - B_j(B_{j+1} - B_j)^2 - \frac{1}{3}(B_{j+1} - B_j)^3$$

2.6 Riemann-Stieltjes Integral

Definition 2.6.1 (Riemann-Stieltjes Integral)**.** If φ is a continuous function on $[a,b]$ and F is a distribution function, we define the **Riemann-Stieltjes Integral** of φ on $[a,b]$ in relation to F (or weighted by F) as

$$\lim_{\|\Delta\| \to 0} \sum_{i=1}^{n} \varphi(y_i)[F(x_{i+1}) - F(x_i))] = \int_a^b \varphi(x)\ dF(x)$$

where $a = x_1 < x_2 < \ldots < x_{n+1} - b$, y_i is an arbitrary point of $[x_i, x_{i+1}]$, and $\|\Delta\| = \max_{1 \le i \le n}(x_{i+1} - x_i)$.

If F is a discrete random variable, then

$$\int_a^b \varphi\ dF = \sum_{i: a < x_i \le b} \varphi(x_i) p(x_i)$$

And if F is a continuous random variable, then

$$\int_a^b \varphi\ dF = \int_a^b \varphi(x) f(x)\ dx$$

2.7 Modes of Convergence

Definition 2.7.1. A **space** $L^p(\Omega, \mathfrak{F}, P)$ is the space of measurable functions in which the p-th power of the absolute value is Lebesgue integrable. I.e., functions of the form

$$\|f\|_p := \left(\int_\Omega |f|^p\, d\mu \right)^{1/p} < \infty$$

We can also write this as

$$\|X\|_p := \sqrt[p]{\mathbb{E}[|X|^p]} < \infty$$

A sequence of random variable can converge in different ways. The following definitions are given in an increasing order of strength.

Definition 2.7.2 (Convergence in Distribution)**.** A sequence of random variables $\{X_n\}$ **converges in distribution** to X if their distribution functions $F_{X_n}(x)$ converge to the distribution function $F_X(x)$ at any point of continuity of F_X.

In other words, for any bounded function $f : \mathbf{R} \longrightarrow \mathbf{R}$ we have

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$$

I.e.,

$$\lim_{n\to\infty} \int_{\Omega_n} f(X_n)\,dP_n = \int_{\Omega} f(X)\,dP$$

Definition 2.7.3 (Convergence in Probability)**.** A sequence of random variables $\{X_n\}$ **converges in probability** to X if for any $\varepsilon > 0$, we have

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

Definition 2.7.4 (Convergence Almost Surely)**.** A sequence of random variables $\{X_n\}$ **converges almost surely (a.s.)** to X if for any ω outside a set of probability zero we have that

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

Alternatively, we can write

$$P[X_n(\omega) \to X(\omega)] = 1$$

Definition 2.7.5 (L^p-Convergence)**.** A sequence of random variables $\{X_n\}$ **converges in** L^p to X if $\{X_n\} \subseteq L^p$, $p \in [1, \infty)$, and

$$\lim_{n\to\infty} \|X_n - X\|_p = 0$$

I.e., $\mathbb{E}[\,|X_n|^p\,] < \infty$ and

$$\lim_{n\to\infty} \mathbb{E}\left[\,|X_n - X|^p\,\right] = 0$$

Chapter 3

Stochastic Integrals

3.1 Formalizing the 'noise'

To study the equation (1.1), the first step is to give a formalization for the 'noise' that we are going to consider. In our case, we'll consider the 'white noise', which is a generalized stochastic process W_t with gaussian probability distribution with mean zero, finite variance such that $\mathbb{E}[W_s W_t] = 0$ whenever $t \neq s$.

In other words, the 'noise' will be represented by some stochastic process W_t and our goal is to satisfy the following properties:

- If $t_1 \neq t_2$, then W_{t_1} and W_{t_2} are independent.
- The collection $\{W_t\}$ is stationary, i.e., the joint distribution of any collections of W_t does not depend on t.
- $\mathbb{E}[W_t] = 0$ for all t.

However, this process W_t does not have reasonable properties (does not have continuous paths, and it is not measurable for finite time). So our first goal is to replace W_t with some convenient stochastic process.

Let $0 = t_0 < t_1 < \ldots < t_m = t$ and consider the discrete version

$$X_{k+1} - X_k = b(t_k, X_k)\Delta t_k + \sigma(t_k, X_k) W_k \Delta t_k$$

Now, replace $W_k \Delta t_k = \Delta V_k = V_{k+1} - V_k$, where V_t is a stochastic process. Since we want V_t to have stationary independent increments with mean zero, the only suitable process with continuous paths is the Brownian motion, represented in the Figure 3.1.

Hence, we take $V_t = B_t$, thus obtaining

$$X_k = X_0 + \sum_{j=0}^{k-1} b(t_j, X_j)\Delta t_j + \sum_{j=0}^{k-1} \sigma(t_j, X_j)\Delta B_j$$

Taking the limit as Δt_j goes to zero, we have

$$X_k = X_0 + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s$$

Notice that

$$B_t - B_s = W_s(t-s)$$

for all $t \geq s$, i.e., the 'white noise' is the derivative, with respect to time, of the Brownian motion.

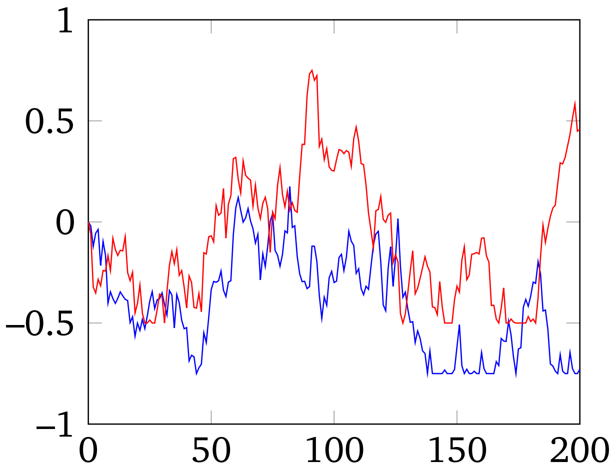

Figure 3.1: Two realizations of the Brownian motion [Jak]

3.2 Preparing the Terrain

Now we can start treating the equation. In order to do that, notice that in the determinist case, in which there isn't a 'noise', we have an equation of the form

$$\frac{dX_t}{dt} = b(t, X_t), \quad X_0 = x_0 \tag{3.1}$$

In this case, a solution is a function X_t such that

$$X_t - X_0 = \int_0^T b(s, X_s)\, ds \tag{3.2}$$

This function can be understood as a stochastic process over a space of probability given by a single point.

Using the formalism above, we propose as a solution to the equation (1.1) an identity of the form

$$X_t - X_0 = \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s \tag{3.3}$$

where the left integral is a Lebesgue-Stieltjes' integral, but the right integral is what still needs to be formalized.

As said earlier, our task is to understand the following integral

$$\int_S^T f(t, \omega)\, dB_t(\omega) \tag{3.4}$$

where $f : [0, \infty] \times \Omega \longrightarrow \mathbf{R}$.

The idea will be the following: we'll first define the integral for the most elementary functions and, after that, using approximations and convergence criteria, we'll define it for the other classes of functions.

Since it is natural to approximate a given function $f(t,\omega)$ by

$$\sum_j f(t_j^*,\omega)\chi_{[t_j,t_{j+1})}(t)$$

where $t_j^* \in [t_j, t_{j+1}]$, we could define the integral (3.4) as follows

$$\int_S^T f(t,\omega)\,dB_t(\omega) = \lim_{n\to\infty} f(t_j^*,\omega)[B_{t_{j+1}} - B_{t_j}](\omega)$$

However, the choice of points t_j^* make a difference here. If we take $t_j^* = t_j$, the left end point, then we obtain the **Itô integral**, denoted by

$$\int_S^T f(t,\omega)\,dB_t(\omega)$$

If, on the other hand, we take $t_j^* = (t_j+t_{j+1})/2$, the mid point, then we obtain the **Stratonovich integral**, denoted by

$$\int_S^T f(t,\omega)\circ dB_t(\omega)$$

One of the advantages of Itô integral is its application in Finance. Intuitively, by taking the leftmost point, one does not have to know the future, like whether a stock goes up or down.

Related to this intuition, and developing the Ito integral, we must restrict ourselves to functions f that only depends on the behaviour of $B_s(\omega)$ up to the time t_j.

Definition 3.2.1. Let $B_t(\omega)$ be a Brownian motion. We define $\mathfrak{F}_t$ as the σ-algebra generated by the random variables B_s, where $s \leq t$. I.e., $\mathfrak{F}_t$ is the smallest σ-algebra containing all sets

$$\{\omega : B_{t_1}(\omega) \in F_1, \dots, B_{t_k}(\omega)\} \in F_k$$

where $t_j \leq t$ and $F_j \subseteq \mathbf{R}^n$ are Borel sets, $j \leq k = 1, 2, \dots$.

The intuition behind this definition is that $\mathfrak{F}_t$ can be thought as the 'history of B_s up to the time t'. A function $h(\omega)$ is $\mathfrak{F}_t$-measurable if h depends only on the values from B_0 up to B_t. I.e., it does not depend on the 'future'.

Definition 3.2.2 ($\mathfrak{N}_t$-Adapted)**.** Let $\{\mathfrak{N}_t\}_{t\geq 0}$ be an increasing family of σ-algebras of subsets of Ω. A process $g(t,\omega) : [0,\infty) \times \Omega \longrightarrow \mathbf{R}^n$ is **$\mathfrak{N}_t$-adapted** if for each $t \geq 0$ the function $\omega \longrightarrow g(t,\omega)$ is $\mathfrak{N}_t$-measurable.

Example 3.2.1. The process $h_1(t,\omega) = B_{t/2}(\omega)$ is $\mathfrak{F}_t$-measurable and, hence, $\mathfrak{F}_t$-adapted. Notice that this process only depends on previous information.

However, the process $h_2(t,\omega) = B_{2t}(\omega)$ is not $\mathfrak{F}_t$-measurable and, hence, it is not $\mathfrak{F}_t$-adapted. That's because the process h_2 depends on future information.

3.3 Constructing the Itô Integral

The class of functions for which the Itô integral will be first defined is the following.

Definition 3.3.1. Let $\mathfrak{V}(S,T)$ be the class of functions

$$f(t,\omega) : [0,\infty) \times \Omega \longrightarrow \mathbf{R}$$

such that

1. The function $(t,\omega) \longrightarrow f(t,\omega)$ is $\mathfrak{B} \times \mathfrak{F}$-measurable, where $\mathfrak{B}$ denotes the Borel σ-algebra on $[0,\infty)$.
2. The function $f(t,\omega)$ is $\mathfrak{F}_t$-adapted.
3. $\mathbb{E}\left[\int_S^T f(t,\omega)^2\,dt\right] < \infty$.

Given functions $f \in \mathfrak{V}$, in order to define the Ito integral

$$I[f](\omega) = \int_S^T f(t,\omega)\,dB_t(\omega)$$

we are going to define $I[\varphi]$ for a simple class of functions φ and then show how each $f \in \mathfrak{V}$ can be approximated by φ.

Definition 3.3.2 (Elementary functions)**.** A function $\varphi \in \mathfrak{V}$ is **elementary** if it has the form

$$\varphi(t,\omega) = \sum_j e_j(\omega)\chi_{[t_j,t_{j+1})}(t)$$

Since $\varphi \in \mathfrak{V}$, each function e_j is $\mathfrak{F}_j$-measurable. And, as we did before, we define

$$\int_S^T \varphi(t,\omega)\,dB_t(\omega) = \sum_{j\geq 0} e_j(\omega)[B_{j+1} - B_j](\omega)$$

To further our development of the Ito integral, we need the following result.

Theorem 3.3.1 (Itô isometry)**.** If $\varphi(t,\omega)$ is bounded and elementary, then

$$\mathbb{E}\left[\left(\int_S^T \varphi(t,\omega)\,dB_t(\omega)\right)^2\right] = \mathbb{E}\left[\int_S^T \varphi(t,\omega)^2\,dt\right]$$

Using this isometry, it is possible to extended the previous definition to functions in $\mathfrak{V}$ by the following steps.

1. Let $g(\omega) \in \mathfrak{V}$ be bounded and continuous for each ω. Then there exists elementary func-

tions $\varphi_n \in \mathfrak{V}$ such that

$$\mathbb{E}\left[\int_S^T (g - \varphi_n)^2 \, dt\right] \longrightarrow 0$$

as $n \to \infty$.

2. Let $h \in \mathfrak{V}$ be bounded, then there exists bounded functions $g_n \in \mathfrak{V}$ such that $g_n(\omega)$ are continuous for all ω and n, and

$$\mathbb{E}\left[\int_S^T (h - g_n)^2 \, dt\right] \longrightarrow 0$$

3. Let $f \in \mathfrak{V}$. Then there exists a sequence of functions $(h_n) \subseteq \mathfrak{V}$ such that h_n is bounded for each n and

$$\mathbb{E}\left[\int_S^T (f - h_n)^2 \, dt\right] \longrightarrow 0$$

as $n \to \infty$.

Please note that we started from bounded and continuous functions, generalized into bounded functions and then generalized further into any function in our class $\mathfrak{V}$.

These steps can be summarized in the following result, which guarantees that the elementary functions are dense in $\mathfrak{V}(S,T)$.

Theorem 3.3.2. If $g \in \mathfrak{V}(S,T)$, then there exists a sequence of elementary functions $\varphi_n \in \mathfrak{V}(S,T)$ such that $\varphi_n \to g$ in $L^2(P)$

Finally, we can define the Itô integral as follows.

Definition 3.3.3 (The Itô integral)**.** Let $f \in \mathfrak{V}(S,T)$, then the **Itô integral** of f, from S to T is

$$\int_S^T f(t,\omega) \, dB_t(\omega) = \lim_{n\to\infty} \int_S^T \varphi_n(t,\omega) \, dB_t(\omega)$$

where the limit is in $L^2(P)$ and (φ_n) is a sequence of elementary functions such that

$$\mathbb{E}\left[\int_S^T (f(t,\omega) - \varphi_n(t,\omega))^2 \, dt\right] \longrightarrow 0$$

as $n \to \infty$.

This new definition induces a generalized form of the Ito isometry.

Theorem 3.3.3 (Ito isometry)**.** For all $f \in \mathfrak{V}(S,T)$,

$$\mathbb{E}\left[\left(\int_S^T f(t,\omega) \, dB_t(\omega)\right)^2\right] = \mathbb{E}\left[\int_S^T f(t,\omega)^2 \, dt\right]$$

Corollary 3.3.4. If $f(t,\omega) \in \mathfrak{V}(S,T)$ and $f_n(t,\omega) \in \mathfrak{V}(S,T)$ and

$$\mathbb{E}\left[\left(\int_S^T f_n(t,\omega) - f(t,\omega)\right)^2 dt\right] \longrightarrow 0$$

as $n \to \infty$, then

$$\int_S^T f_n(t,\omega)\, dB_t(\omega) \longrightarrow \int_S^T f(t,\omega)\, dB)t(\omega)$$

in $L^2(P)$ as $n \to \infty$.

Example 3.3.1. We wish to compute the integral

$$\int_0^T B(t)\, dB(t)$$

Step 1. Let

$$\varphi_n(t) = \sum_{i=0}^{n-1} B_n(t_i^n)[B_n(t_{i+1}^n) - B_n(t_i^n)]$$

be a sequence of elementary functions. Then,

$$\mathbb{E}\left[\int_0^T (\varphi_n - B_s)^2\, ds\right] = \mathbb{E}\left[\sum_j \int_{t_j}^{t_{j+1}} (B_j - B_s)^2\, ds\right] = \sum_j \int_{t_j}^{t_{j+1}} (s - t_j)^2\, ds$$
$$= \sum_j \frac{1}{2}\left(t_{j+1} - t_j\right)^2 \longrightarrow 0 \text{ when } \Delta t_j \to 0$$

I.e., $\varphi_n(t)$ converges to $B(t)$ almost surely as $\max_i(t_{i+1}^n - t_i^n) \to 0$ by the continuity of $B(t)$.

Step 2. Now notice that

$$\begin{aligned} B_n(t_i)[B_n(t_{i+1}) - B_n(t_i)] &= B_n(t_{i+1})B_n(t_i) - B_n^2(t_i) + B_n^2(t_{i+1}) - B_n^2(t_{i+1}) \\ &= \frac{1}{2}[B_n^2(t_{i+1}) - B_n^2(t_i) - (B_n(t_{i+1}) - B_n(t_i))^2] \end{aligned}$$

Step 3. With these two results, we can write the original integral as

$$\int_0^T X_n(t)\, dB(t) = \frac{1}{2}\sum_{i=0}^{n-1}\left[B_n^2(t_{i+1}) - B_n^2(t_i)\right] - \frac{1}{2}\sum_{i=0}^{n-1}[B_n(t_{i+1}) - B_n(t_i)]^2$$

Step 4. Since the first sum is a telescopic sum and the second one converges in probability to T by the quadratic variation of Brownian motion, we have

$$\int_0^T X_n(t)\, dB(t) = \frac{1}{2}B^2(t) - \frac{1}{2}B^2(0) - \frac{1}{2}T = \frac{1}{2}B^2(t) - \frac{1}{2}T$$

Hence, the integral converges in $L^2(P)$ to

$$\int_0^T B(t)\, dB(t) = \lim_{n\to\infty} \int_0^T X_n(t)\, dB(t) = \frac{1}{2}B^2(t) - \frac{1}{2}T$$

3.4 Properties

Before heading to more theoretical and important results, let us notice some natural facts of the Itô integral.

Theorem 3.4.1. Let $f, g \in \mathfrak{V}(0,T)$ and let $0 \leq S < U < T$. Then

1. $\int_S^T f\, dB_t = \int_S^U f\, dB_t + \int_U^T f\, dB_t$ for a.a. ω.
2. $\int_S^T (cf + g)\, dB_t = c\int_S^T f\, dB_t + \int_S^T g\, dB_t$ where $c \in \mathbf{R}$.
3. $\mathbb{E}\left[\int_S^T f\, dB_t\right] = 0$.
4. $\int_S^T f\, dB_t$ is $\mathfrak{F}_T$-measurable.

With the Doob's martingale inequality, it can be proved that the Itô integral can be chosen to depend continuously on t. A proof of this fact is on the third chapter of [Oks13].

Now an important result is that the Itô integral is a martingale.

Theorem 3.4.2. Let $f(t,\omega) \in \mathfrak{V}(0,T)$ for all T. Then

$$M_t(\omega) = \int_0^t f(s,\omega)\, dB_s$$

is a martingale w.r.t. $\mathfrak{F}_t$ and for $\lambda, T > 0$

$$P\left[\sup_{0\leq t\leq T} |M_t| \geq \lambda\right] \leq \frac{1}{\lambda^2}\mathbb{E}\left[\int_0^T f(s,\omega)^2\, ds\right]$$

Summarizing, the Itô integral is continuous, adapted, linear, a martingale, satisfies the Itô isometry and has Quadratic variation.

3.5 Extensions

Using the concept of martingales, we can generalize the Itô integral for a larger class of functions than $\mathfrak{V}$.

Considering the Definition 3.3.1, we can relax the condition 2. into

2.' There exists an increasing familiy of σ-algebras $\mathfrak{H}_t$ such that B_t is a martingale w.r.t. $\mathfrak{H}_t$ and f_t is $\mathfrak{H}_t$-adapted.

The idea here is that B_t must remain a martingale with respect to the history of f_s.

Another way of extending the Itô integral definition is by weaking the condition 3. into

3.' $P\left[\int_S^T f(s,\omega)^2\,ds < \infty\right] = 1$

Too understand why this works, let $\mathfrak{W}_{\mathfrak{H}}(S,T)$ denote the class of processes satisfying conditions 1, 2' and 3' above. Then, in the 1-dimensional Brownian motion, for all t there exists step functions $f_n \in \mathfrak{W}_{\mathfrak{H}}$ such that

$$\int_0^t |f_n - f|^2\,ds \longrightarrow 0$$

in probability.

Now, since $\int_0^t f_n(s,\omega)\,dB_s(\omega)$ converges in probability to a random variable and the limit depends only on f, we define

$$\int_0^t f(s,\omega)\,dB_s(\omega) = \lim_{n\to\infty} \int_0^t f_n(s,\omega)\,dB_s(\omega)$$

where the limit is in probability and $f \in \mathfrak{W}_{\mathfrak{H}}$.

3.6 Stratonovich integral

As we saw, the Itô integral is one possible interpretation of the integral

$$\int_S^T f(t,\omega)\,dB_t(\omega)$$

The Stratonovich integral is another possibility and, in general, leads to different results (except when the integrating function has a derivative). In some cases, the Stratonovich definition may be adequate.

In other cases, Itô's feature of 'not looking in the future' (as [Oks13] puts it), justifies its use in biology and finance, for example. Moreover, Itô integral is a martingale, and Stratonovich's is not. This gives an important computational advantage to Itô's definition.

And, to quote [S+04],

> However, it [Stratonovich's integral] is inappropriate for finance. In finance, the integrand represents a position in an asset and the integrator represents the price of that asset. We cannot decide at 1:00 p.m. which position we took at 9:00 a.m. We must decide the position at the beginning of each time interval, and the Ito integral is the limit of the gain achieved by that kind of trading as the time between trades approaches zero.

However, Stratonovich integral is used in manifolds and is of particular interest to physicists for obeying rules of classical calculus, such as the chain rule.

In this section, we presented a formalism for modeling evolution process with noise using the Brownian motion. We saw that the classical notion of solution doesn't work and we proposed a new interpretation. As a result, it was necessary to construct a new theory of integration and, with that, the Itô and Stratonovich integrals. We hence studied the properties of these integrals.

Given that, we can advance on more elaborated questions, for example: does the equation (1.1) has a solution? Under which hypothesis? How can we study the behavior of the solution?

Chapter 4

The Itô Formula and the Martingale Representation Theorem

4.1 1-Dimensional Itô Processes

Itô formula: Integral version of the Chain Rule

Idea: introduce **Itô processes (stochastic integrals)** as sums of a $\mathrm{d}B_s$– and a $\mathrm{d}s$-integral. This family of integrals is stable under smooth maps (what does this mean?)

Definition 4.1.1 (1-dimensional Itô processes)**.** B_t 1-dimensional Brownian motion on $(\Omega, \mathfrak{F}, \mathbb{P})$. A 1-dimensional Itô process (or stochastic integral) is a stochastic process X_t on $(\Omega, \mathfrak{F}, \mathbb{P})$ of the form

$$X_t = X_0 + \int_0^t u(s,\omega)\ \mathrm{d}s + \int_0^t v(s,\omega)\ \mathrm{d}B_s \tag{4.1}$$

where $v \in \mathfrak{W}_{\mathfrak{H}}$ and

$$\mathbb{P}\left[\int_0^t v(s,\omega)^2\ \mathrm{d}s < \infty \text{ for all } t \geq 0\right] = 1$$

and u is $\mathfrak{H}_t$-adapted (as in the condition 2') and

$$\mathbb{P}\left[\int_0^t |u(s,\omega)|\ \mathrm{d}s < \infty \text{ for all } t \geq 0\right] = 1$$

If X_t is an Itô process of this form, then the equation (4.1) is written in the differential form

$$\mathrm{d}X_t = u\mathrm{d}t + v\mathrm{d}B_t$$

Hence, we can write

$$\mathrm{d}\left(\frac{1}{2}B_t^2\right) = \frac{1}{2}\mathrm{d}t + B_t\mathrm{d}B_t$$

Theorem 4.1.1 (The 1-dimensional Itô formula)**.** X_t be an Itô process given by

$$\mathrm{d}X_t = u\mathrm{d}t + v\mathrm{d}B_t$$

and $g(t,x) \in \mathfrak{C}^2([0,\infty) \times \mathbf{R})$. Then

$$Y_t = g(t, X_t)$$

is an Itô process, and

$$dY_t = \frac{\partial g}{\partial t}(t,X_t)dt + \frac{\partial g}{\partial x}(t,X_t)dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t,X_t)(dX_t)^2 \tag{4.2}$$

where $(dX_t)^2$ is computed according to the rules

$$dt \cdot dt = dt \cdot dB_t = dB_t \cdot dt = 0,\ dB_t \cdot dB_t = dt$$

Proof. Notice that the claim is equivalent to the expression

$$g(t,X_t) = g(0,X_0) + \int_0^t \left(\frac{\partial g}{\partial s}(s,X_s) + u_s \frac{\partial g}{\partial x}(s,X_s) + \frac{1}{2}v_s^2 \frac{\partial^2 g}{\partial x^2}(s,X_s) \right) ds + \int_0^t v_s \frac{\partial g}{\partial x}(s,X_s)\ dB_s$$

where $u_s = u(s,\omega)$ and $v_s = v(s,\omega)$.

We'll assume that g and its derivatives above are bounded. In the general case, we can approximate g by $\mathfrak{C}^2$ functions that converge uniformly to g.

Then write the Taylor expansion with respect to the arguments (in this case, t and X_t), taking the first-order terms for all arguments with zero quadratic variation (here, t) and up to the second order for every argument with non zero quadratic variation (here, X_t).

Notice that the terms with crossed Δt_j and ΔB_j go to zero by Itô isometry.

Also notice that the terms with crossed $(\Delta B_i)^2 - \Delta t_i$ and $(\Delta B_j)^2 - \Delta t_j$, $i \neq j$, vanish because they are indepedent. For the similar terms with $i = j$, the sum converges to a dt integral, showing that $(dB_t)^2 = dt$.

Since the remainder goes to zero as $\Delta t_j \to 0$, we complete the proof. □

Example 4.1.1. Consider again the following integral

$$I = \int_0^t B_s\ dB_s$$

Choose $X_t = B_t$ and $g(t,x) = \frac{1}{2}x^2$, then

$$Y_t = g(t, B_t) = \frac{1}{2}B_t^2$$

Then, by Itô's formula,

$$\begin{aligned} dY_t &= \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial x}dB_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(dB_t)^2 \\ &= B_t dB_t + \frac{1}{2}(dB_t)^2 \\ &= B_t dB_t + \frac{1}{2}dt \end{aligned}$$

I.e.,

$$d\left(\frac{1}{2}B_t^2\right) = B_t dB_t + \frac{1}{2}dt$$

Put it another way,

$$\frac{1}{2}B_t^2 = \int_0^t B_s\ dB_s + \frac{1}{2}t$$

Example 4.1.2. Now consider

$$\int_0^t s\ dB_s$$

Put

$$g(t,x) = tx \text{ and } Y_t = g(t, B_t) = tB_t$$

By Itô's formula,

$$dY_t = \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial x}dB_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(dB_t)^2$$

i.e.

$$d(tB_t) = dY_t = B_t dt + t dB_t + \cdot 0$$

In the integral form,

$$tB_t = \int_0^t B_s\ ds + \int_0^t s\ dB_s$$

or

$$\int_0^t s\ dB_s = tB_t - \int_0^t B_s\ ds$$

This example is generalized in the next theorem.

Theorem 4.1.2 (Integration by parts)**.** Suppose that $f(s,\omega) = f(s)$ only depends on s and that f is continuous and of bounded variation in $[0,t]$. Then

$$\int_0^t f(s)\ dB_s = f(t)B_t - \int_0^t B_s\ df_s$$

4.2 The Multi-Dimensional Itô Formula

Definition 4.2.1 (n-dimensional Itô process)**.** Let $B(t,\omega) = (B_{t_1}(t,\omega), \ldots, B_m(t,\omega))$ denote m-dimensional Brownian motion. If each $u_i(t,\omega)$ and $v_{ij}(t,\omega)$, for $1 \le i \le n$ and $1 \le j \le m$, satisfies the conditions given in the definition of Itô process, then we can form n Itô processes

$$\begin{cases} dX_1 = u_1 dt + v_{11} dB_1 + \ldots + v_{1m} dB_m \\ \quad\vdots \\ dX_n = u_n dt + v_{n1} dB_1 + \ldots + v_{nm} dB_m \end{cases}$$

Or simply

$$dX(t) = udt + vdB(t)$$

where

$$X(t) = \begin{bmatrix} X_1(t) \\ \vdots \\ X_n(t) \end{bmatrix}, u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}, v = \begin{bmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nm} \end{bmatrix}, dB(t) = \begin{bmatrix} dB_1(t) \\ \vdots \\ dB_m(t) \end{bmatrix}$$

The process $X(t)$ is called an **n-dimensional Itô process**.

Theorem 4.2.1 (General Itô Formula)**.** Let

$$dX(t) = udt + vdB(t)$$

be an n-dimensional Itô process. And let $g(t,x) = (g_1(t,x), \ldots, g_p(t,x))$ be a $\mathfrak{C}^2$ map from $[0,\infty) \times \mathbf{R}^n$ into $\mathbf{R}^p$. Then the process

$$Y(t,\omega) = g(t, X(t))$$

is an Itô process, whose k-th component is given by

$$dY_k = \frac{\partial g_k}{\partial t}(t,X)dt + \sum_i \frac{\partial g_k}{\partial x_i}(t,X)dX_i + \frac{1}{2}\sum_{i,j} \frac{\partial^2 g_k}{\partial x_i \partial x_j}(t,X)dX_i dX_j$$

where $dB_i dB_j = \delta_{ij}dt$, $dB_i dt = dt dB_i = 0$.

4.3 The Martingale Representation Theorem

Let $B(t)$ be an n-dimensional Brownian motion. We saw that if $v \in \mathfrak{V}^n$, then the Itô integral

$$X_t = X_0 + \int_0^t v(s,\omega)\, dB(s),\ t \geq 0$$

is always a martingale w.r.t. filtration $\mathfrak{F}_t^n$ and the probability measure $\mathbb{P}$.

In this section, we show that the converse is also true: Any $\mathfrak{F}_t^n$-martingale (w.r.t to $\mathbb{P}$) can be represented as an Itô integral. This result is called the martingale representation theorem, and is important in many applications, such as mathematical finance.

Lemma 4.3.1. Fix $T > 0$. The set of random variables

$$\{\varphi(B_{t_1}, \ldots, B_{t_n})\}$$

where $t_i \in [0,T]$, $\varphi \in C_0^\infty(\mathbf{R}^n)$, and $n = 1, 2, \ldots$ is dense in $L^2(\mathfrak{F}_T, \mathbb{P})$.

Lemma 4.3.2. The linear span of random variables of the type

$$\exp\left\{\int_0^T h(t)\ dB_t(\omega) - \frac{1}{2}\int_0^T h^2(t)\ dt\right\}$$

where $h \in L^2[0,T]$ is dense in $L^2(\mathfrak{F}_T, \mathbb{P})$.

Suppose that $B(t)$ is n-dimensional. If $v(s,\omega) \in \mathfrak{V}^n(0,T)$, then the random variable

$$V(\omega) := \int_0^T v(t,\omega)\ dB(t)$$

is $\mathfrak{F}_t^n$-measurable and by the Itô isometry

$$\mathbb{E}[V^2] = \int_0^T \mathbb{E}[v^2(t,\cdot)]\ dt < \infty$$

so $V \in L^2(\mathfrak{F}_T, \mathbb{P})$.

Theorem 4.3.3 (Itô Representation Theorem)**.** Let $F \in L^2(\mathfrak{F}_T, \mathbb{P})$. Then there exists a unique stochastic process $f(t,\omega) \in \mathfrak{V}^n(0,T)$ such that

$$F(\omega) = \mathbb{E}[F] + \int_0^T f(t,\omega)\ dB(t)$$

Theorem 4.3.4 (The Martingale Representation Theorem)**.** Let $B(t)$ be an n-dimensional Brownian motion. Suppose M_t is an $\mathfrak{F}_t^{(n)}$-martingale w.r.t. $\mathbb{P}$ and that $M_t \in L^2(\mathbb{P})$ for all $t \geq 0$.

Then there exists a unique stochastic process $g(s,\omega)$ such that $g \in \mathfrak{V}^{(n)}(0,t)$ for all $t \geq 0$ and

$$M_t(\omega) = \mathbb{E}[M_0] + \int_0^t g(s,\omega)\ dB(s)$$

almost surely, for all $t \geq 0$.

Chapter 5

Stochastic Differential Equations

Stochastic differential equations model evolution processes with continuous random noise, and are of the form
$$\frac{dX_t}{dt} = b_t(X_t) + \sigma_t(X_t)W_t \tag{5.1}$$
where $b_t(x), \sigma_t(x) \in \mathbf{R}$ and W_t is the one-dimensional 'white noise'.

Inspired by the deterministic case, we propose as a solution a stochastic process X_t which satisfies the following identity
$$X_t = X_0 + \int_0^t b_s(X_s)\, ds + \int_0^t \sigma_s(X_s)\, dB_s \tag{5.2}$$

I.e.,
$$dX_t = b_t(X_t)\, dt + \sigma_t(X_t)\, dB_t \tag{5.3}$$

Notice that to get from (5.1) to (5.3) we replace W_t by $\frac{dB_t}{dt}$ and multiply by dt.

In this chapter, we address the following questions. When does a solution exists, and when is it unique? What are its properties? How to solve such an equation?

5.1 Solution Methods

We start this section by the following example, which shows an 'Itô correction' to the exponential.

Example 5.1.1 (Stochastic Exponential). Consider
$$dU_t = U_t dX_t, \quad U_0 = 1 \iff U_t = 1 + \int_0^t U_s\, dX_s \tag{5.4}$$

U is called the **stochastic exponential of** X. If X_t have finite variation, then the solution of the equation (5.4) is given by $U_t = e^{X_t}$. The proof follows imediatelly from the Itô's formula (see [Kle12]).

For the Brownian motion B_t, the stochastic exponential is given by

$$U_t = e^{B_t - \frac{1}{2}t} \tag{5.5}$$

Let $Y_t = B_t - \frac{1}{2}t$ and $g(t,x) = x - \frac{1}{2}t$. Then

$$dY_t = -\frac{1}{2}dt + dB_t$$

Noting that $U_t = e^{Y_t}$, define $g(t,x) = e^x$. Then $U_t = g(t, Y_t)$ and

$$\begin{aligned} dU_t &= e^{Y_t}dY_t + \frac{1}{2}e^{Y_t}(dY_t)^2 = U_t\left(-\frac{1}{2}dt + dB_t\right) + \frac{1}{2}U_t dt \\ &= U_t dB_t \end{aligned}$$

A linear SDE in one dimension is of the form

$$dX_t = (\alpha_t + \beta_t X_t)dt + (\gamma_t + \delta_t X_t)dB_t \tag{5.6}$$

where $\alpha, \beta, \gamma, \delta$ are adapted processes and continuous functions of t.

Example 5.1.2 (Stochastic Exponential SDE). If $\alpha_t = 0 = \gamma_t$, we have

$$dU_t = \beta_t U_t dt + \delta_t U_t dB_t$$

The solution for this equation is

$$U_t = U_0 \exp\left(\int_0^t \left(\beta_s - \frac{1}{2}\delta_s^2\right) ds + \int_0^t \delta_s \, dB_s\right)$$

To verify this, suppose for simplicity that $U_0 = 1$, and let

$$g(t,x) = \exp\left(\int_0^t \left(\beta_s - \frac{1}{2}\delta_s^2\right) ds + \int_0^t \delta_s \, dx\right)$$

Note that $U_t = g(t, B_t)$ and compute

$$\frac{\partial g}{\partial t} = \left(\beta_t - \frac{1}{2}\delta_t^2\right)U_t, \quad \frac{\partial g}{\partial x} = \delta_t U_t, \quad \frac{\partial^2 g}{\partial x^2} = \delta_t^2 U_t$$

By Itô's formula,

$$\begin{aligned} dU_t &= \left(\beta_t - \frac{1}{2}\delta_t^2\right)U_t dt + \delta_t U_t dB_t + \frac{1}{2}\delta_t^2 U_t dt \\ &= \beta_t U_t dt + \delta_t U_t dB_t \end{aligned}$$

Another way to look at this equation is by remarking that it is the stochastic exponential of Y_t, i.e.,

$$dU_t = U_t dY_t \tag{5.7}$$

in which Y_t is an Itô process defined by

$$dY_t = \beta_t dt + \delta_t dB_t$$

Therefore,

$$\begin{aligned} U_t &= U_0 \exp\left(Y_t - Y_0 - \frac{1}{2}[Y,Y](t)\right) \\ &= U_0 \exp\left(\int_0^t \beta_s \, ds + \int_0^t \delta_s \, dB_s - \frac{1}{2}\int_0^t \delta_s^2 \, ds\right) \\ &= U_0 \exp\left(\int_0^t \left(\beta_s - \frac{1}{2}\delta_s^2\right) ds + \int_0^t \delta_s \, dB_s\right) \end{aligned}$$

Example 5.1.3 (Stochastic Logarithm)**.** How can we compute the integral $\int_0^t \frac{dX_s}{X_s}$? Applying Itô formula for $g(t,x) = \ln x$, where $x > 0$,

$$\begin{aligned} d(\ln X_t) &= \frac{1}{X_t} dX_t + \frac{1}{2}\left(-\frac{1}{X_t^2}\right)(dX_t)^2 \\ &= \frac{dX_t}{X_t} - \frac{1}{2X_t^2}(dX_t)^2 \end{aligned}$$

Hence,

$$\frac{dX_t}{X_t} = d(\ln X_t) + \frac{1}{2X_t^2}(dX_t)^2 \tag{5.8}$$

Example 5.1.4 (Population Growth)**.** Consider the following Stochastic Differential Equation (SDE):

$$\frac{dN_t}{dt} = a_t N_t, \quad N_0 \in \mathbf{R}$$

where $a_t = r_t + \alpha W_t$.

If r_t is a constant r, by our previous discussion we have

$$\frac{dN_t}{N_t} = r dt + \alpha dB_t \iff dN_t = r N_t dt + \alpha N_t dB_t$$

In the integral form,

$$\int_0^t \frac{dN_s}{N_s} = rt + \alpha B_t$$

Using (5.8)

$$\frac{dN_t}{N_t} = d(\ln N_t) + \frac{1}{2N_t^2}\left(\alpha^2 N_t^2 dt\right) = d(\ln N_t) + \frac{1}{2}\alpha^2 dt$$

Hence,

$$\ln\left(\frac{N_t}{N_0}\right) = \left(r - \frac{1}{2}\alpha^2\right)t + \alpha B_t \iff N_t = N_0 \exp\left(\left(r - \frac{1}{2}\alpha^2\right)t + \alpha B_t\right)$$

Remark that if we used the Stratonovich interpretation, then the solution would be

$$N_t = N_0 \exp(rt + \alpha B_t)$$

Example 5.1.5 (Stock Prices). Let $P(t)$ denote the price of a stock at time t. We may assume that the relative change of price evolves according to the SDE

$$\frac{dP}{P} = \mu dt + \sigma dB_t \iff dP = \mu P dt + \sigma P dB_t$$

where $\mu > 0$ is the **drift**, and σ is the **volatility** of the stock.

By Itô's formula,

$$\begin{aligned} d(\ln P) &= \frac{dP}{P} - \frac{1}{2}\frac{\sigma^2 P^2}{P^2} dt \\ &= \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dB_t \end{aligned}$$

Hence,

$$P(t) = p_0 \exp\left(\sigma B_t + \left(\mu - \frac{\sigma^2}{2}\right) t\right)$$

Since the integral form of the SDE is

$$P(t) = p_0 + \int_0^t \mu P \, ds + \int_0^t \sigma P \, dB_t$$

and

$$\mathbb{E}\left(\int_0^t \sigma P \, dB_t\right) = 0$$

we have that

$$\mathbb{E}(P(t)) = p_0 + \int_0^t \mu \mathbb{E}(P(s)) \, ds$$

Hence,

$$\mathbb{E}(P(t)) = p_0 e^{\mu t} \quad \text{for } t \geq 0$$

The above example is a generalization of Example 5.1.4, where B_t is independent of N_0, i.e., there's no noise in a_t.

In this case, we obtain

$$\mathbb{E}[N_t] = \mathbb{E}[N_0] e^{rt}$$

For the Stratonovich solution,

$$\mathbb{E}[N_t] = \mathbb{E}[N_0] e^{\left(r + \frac{1}{2}\alpha^2\right)t}$$

Notice that for Itô solution we have

1. If $r > \frac{1}{2}\alpha^2$, then $N_t \to \infty$ as $t \to \infty$ a.s.
2. If $r < \frac{1}{2}\alpha^2$, then $N_t \to 0$ as $t \to \infty$ a.s.

3. If $r = \frac{1}{2}\alpha^2$, then N_t will fluctuate between arbitrary large and small values as $t \to \infty$ a.s.

Definition 5.1.1 (Geometric Brownian Motion). Processes of the type

$$X_t = X_0 \exp(\mu t + \alpha B_t)$$

where μ, α are constants, are called **geometric Brownian motions**.

Example 5.1.6 (Electric Circuit). Let $Q(t)$ be the charge at time t at a fixed point in a electric circuit. Then the charge satisfies the following differential equation

$$LQ''(t) + RQ'(t) + \frac{1}{C}Q(t) = F(t), \quad Q(0) = Q_0,\ Q'(0) = I_0$$

In which L is inductance, R is resistance, C is capacitance and $F(t)$ the potential source.

It may be the case that some coefficient, say F_t, are not deterministic, but of the form

$$F(t) = G(t) + \alpha W(t)$$

We reduce this second order equation to a system of first order equations by defining

$$X(t,\omega) = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} Q_t \\ Q'_t \end{bmatrix}$$

and then obtaining

$$\begin{cases} X'_1 = X_2 \\ LX'_2 = -RX_2 - \frac{1}{2}X_1 + G_t + \alpha W_t \end{cases} \tag{5.9}$$

Letting

$$dX = \begin{bmatrix} dX_1 \\ dX_2 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 1 \\ -\frac{1}{CL} & -\frac{R}{L} \end{bmatrix}, \quad H_t = \begin{bmatrix} 0 \\ \frac{1}{L}G_t \end{bmatrix}, \quad K = \begin{bmatrix} 0 \\ \frac{\alpha}{L} \end{bmatrix}$$

we can write the system (5.9) in a matricial form

$$dX_t = AX_t dt + H_t dt + K dB_t \tag{5.10}$$

Now, rewrite (5.10) as

$$\exp(-At)dX_t - \exp(-At)AX_t dt = \exp(-At)[H_t dt + K dB_t] \tag{5.11}$$

Applying Itô's formula to $d(\exp(-At)X_t)$,

$$d(\exp(-At)X_t) = (-A)\exp(-At)X_t dt + \exp(-At)dX_t$$

and replacing it into (5.11) and using integration by parts,

$$\exp(-At)X_t - X_0 = \int_0^t \exp(-As)H_s\, ds + \int_0^t \exp(-As)K\, dB_s$$

i.e.,

$$X_t = \exp(At)\left(X_0 + \exp(-At)KB_t + \int_0^t \exp(-As)[H_s + AKB_s]\, ds\right)$$

For more solution methods, see [Eva12], [Gar88], and [Kle12].

5.2 Existence and Uniqueness

Theorem 5.2.1. Let $T > 0$ and $b : [0,T] \times \mathbf{R}^n \longrightarrow \mathbf{R}^n$, $\sigma : [0,T] \times \mathbf{R}^n \longrightarrow \mathbf{R}^{n\times m}$ are measurable functions.

If the following conditions are satisfied

1. Coefficients satisfy the linear growth condition

$$|b(t,x)| + |\sigma(t,x)| \leq C(1 + |x|), \quad x \in \mathbf{R}^n,\ t \in [0,T]$$

for some constant C;

2. Coefficients are locally Lipschitz in x uniformly in t

$$|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq D|x-y|, \quad x,y \in \mathbf{R}^n,\ t \in [0,T]$$

for some constant D;

3. Z is a random variable which is independent of the σ-algebra $\mathfrak{F}_\infty^{(m)}$ generated by B_s $(s \geq 0)$ and such that $\mathbb{E}[|Z|^2] < \infty$.

Then the stochastic differential equation

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad 0 \leq t \leq T,\ X_0 = Z \tag{5.12}$$

has a unique t-continuous solution $X_t(\omega)$ with the property that $X_t(\omega)$ is adapted to the filtration $\mathfrak{F}_t^Z$ generated by Z and B_s $(s \leq t)$, and

$$\mathbb{E}\left[\int_0^T |X_t|^2\, dt\right] < \infty$$

Proof. Existence: successive approximations (Picard iterations). By Gronwall's lemma, the Lipschitz condition and Itô isometry imply uniqueness. □

5.3 Weak and Strong Solutions

The solution that we found is called a **strong solution**, in which the version of Brownian motion is given, and the solution is $\mathfrak{F}_t^Z$-adapted.

However, what if we're to find a pair of processes $((\bar{X}_t, \bar{B}_t), \mathfrak{H}_t)$ on a probability space such that the differential form (5.12) holds? In this case, the solution $(\bar{X}_t, \bar{B}_t)$ is called a **weak solution**. Put another way, 'weak solutions are solutions in distribution' [Kle12, 137].

The uniqueness of the Theorem 5.2.1 is **strong** or **pathwise** uniqueness. **Weak uniqueness** means that any two solutions are identical in law (have the same finite-dimensional distributions).

Notice that any strong solution is a weak solution (a result from Yamada and Watanabe) and any solution is weakly unique. The next example shows that it is possible to have no strong solutions, but a weakly unique weak solution.

Example 5.3.1 (The Tanaka Equation)**.** Consider

$$dX_t = \text{sign}(X_t)dB_t, \quad X_0 = 0 \tag{5.13}$$

Since $\sigma(t,x) = \text{sign}(x)$ is not continuous, it is not Lipschitz and hence the Theorem 5.2.1 does not apply. In fact, (5.13) has no strong solution.

Let X_t be any Brownian motion. Then

$$Y_t = \int_0^t \text{sign}(X_s)\, dX_s \iff dY_t = \text{sign}(X_t)dX_t$$

Hence,

$$dX_t = \text{sign}(X_t)dY_t$$

We'll see in Levy's theorem that Y_t is a Brownian motion (more than that, any weak solution is a Brownian motion) and X_t is a weakly unique weak solution.

Chapter 6

Diffusion Processes

This chapter studies the properties of solutions of stochastic differential equations. Since these solutions can be interpreted as in the physical Brownian motion, such stochastic processes are called **(Itô) diffusions**.

6.1 Basic Definitions

Throughout this chapter, consider a stochastic differential equation

$$\mathrm{d}X_t = \mu(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}B_t \tag{6.1}$$

with $X_t \in \mathbf{R}^n$, $\mu(t,x) \in \mathbf{R}^n$, $\sigma(t,x) \in \mathbf{R}^{n\times m}$ and B_t is m-dimensional Brownian motion. We call μ the **drift coefficient** and σ (or $\frac{1}{2}\sigma\sigma^{\mathrm{T}}$) the **diffusion coefficient**.

Definition 6.1.1 (Itô Diffusion)**.** A (time-homogeneous) **Itô diffusion** is a stochastic process

$$X_t(\omega) : [0, \infty] \times \Omega \longrightarrow \mathbf{R}^n$$

satisfying a stochastic differential equation of the form

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t, \quad t \geq s, \quad X_s = x$$

satisfying the conditions of the Theorem 5.2.1, which in this case simplify to:

$$|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \leq D|x - y|, \quad x, y \in \mathbf{R}^n$$

with $|\sigma|^2 = \sum |\sigma_{ij}|^2$.

We denote the solution as $X_t = X_t^{s,x}$ and, for $s = 0$, we write $X_t^x = X_t^{0,x}$.

Notice that the process $X_t(\omega)$ has the property of being **time-homogeneous**, i.e.

$$\begin{aligned} X_{s+h}^{s,x} &= x + \int_s^{s+h} \mu(X_u^{s,x})\ \mathrm{d}u + \int_s^{s+h} \mu(X_u^{s,x})\ \mathrm{d}B_u \\ &= x + \int_0^h \mu(X_{s+v}^{s,x})\ \mathrm{d}v + \int_0^h \mu(X_{s+v}^{s,x})\ \mathrm{d}\bar{B}_v \end{aligned}$$

with $u = s + v$, $\overline{B}_v = B_{s+v} - B_s$, $v \geq 0$.

Taking $s = 0$,

$$X_h^{0,x} = x + \int_0^h \mu(X_v^{0,x})\, dv + \int_0^h \mu(X_v^{0,x})\, dB_v$$

Since the processes $\{\overline{B}_v\}$ and $\{B_v\}$ have the same P^0-distribution, by weak uniqueness of the solution, we have that $\{X_{s+h}^{s,x}\}_{h\geq 0}$ and $\{X_h^{0,x}\}_{h\geq 0}$ have the same P^0-distributions. Thus, $\{X_t\}_{t\geq 0}$ is time-homogeneous.

Let us now introduce new probability laws Q^x that give the distribution of $\{X_t\}_{t\geq 0}$ assuming that $X_0 = x$. Let $\mathfrak{M}_\infty$ be the σ-algebra generated by the random variables $\omega \longrightarrow X_t(\omega) = X_t^y(\omega)$, with $t \geq 0$ and $y \in \mathbf{R}^n$. We define Q^x on $\mathfrak{M}$ as

$$Q^x[X_{t_1} \in E_1, \ldots, X_{t_k} \in E_k] = P^0[X_{t_1}^x \in E_1, \ldots, X_{t_k}^x \in E_k]$$

where $E_1 \subset \mathbf{R}^n$ are Borel sets and $1 \leq i \leq k$.

Also let $\mathfrak{F}_t^{(m)}$ be the σ-algebra generated by $\{B_r : r \leq t\}$ and $\mathfrak{M}_t$ be the σ-algebra generated by $\{X_r : r \leq t\}$. Since X_t is measurable w.r.t. $\mathfrak{F}_t^{(m)}$, it follows that $\mathfrak{M}_t \subseteq \mathfrak{F}_t^{(m)}$.

6.2 The Markov Property

Another feature of the solution X_t is the Markov property, which says that, given the present state, the future is independent of the past. Put another way, given what happened up to time t, the future of the process is the same as the behavior obtained by starting the process at X_t. More precisely,

Definition 6.2.1 (Markov Property)**.** For any $0 \leq s < t$,

$$P[X_t \leq y \mid \mathfrak{F}_s] = P[X_t \leq y \mid X_s] \quad \text{a.s.}$$

The next results states that a solution to an SDE X_t is a Markov process w.r.t. $\{\mathfrak{F}_t^{(m)}\}_{t\geq 0}$.

Theorem 6.2.1 (Markov Property for Itô Diffusions)**.** Let $f : \mathbf{R}^n \longrightarrow \mathbf{R}$ be a bounded Borel function. For $t, h \geq 0$,

$$\mathbb{E}^x[f(X_{t+h}) \mid \mathfrak{F}_t^{(m)}]_{(\omega)} = \mathbb{E}^{X_t(\omega)}[f(X_h)] \tag{6.2}$$

where $\mathbb{E}^x$ denotes the expected value w.r.t. the probability measure Q^x, i.e., $\mathbb{E}^x[f(X_h)] = [f(X_h^x)]$ w.r.t. the probability measure P^0.

Proof. By uniqueness (of SDE solutions), $X_r(\omega) = X_r^{t,X_t}(\omega)$.

Define

$$F(x, t, r, \omega) = X_r^{t,x}(\omega), \quad r \geq t$$

and use this to write $X_r(\omega) = F(X_t, t, r, \omega)$.

Using that $\omega \longrightarrow F(x, t, r, \omega)$ is $\mathfrak{F}_t^{(m)}$-independent and the time-homogeneity, we have that

$$\mathbb{E}[f(F(X_t, t, t+h, \omega)) \mid \mathfrak{F}_t^{(m)}] = \mathbb{E}[f(F(X_t, 0, h, \omega))]$$

Let $g(t,\omega) = f \circ F(x,t,t+h,\omega)$ and notice that it is measurable. Thus, we can approximate g pointwise boundedly by

$$\sum_{k=1}^{m} \varphi_k(x)\psi_k(\omega)$$

Compute

$$\mathbb{E}[g(X_t,\omega) \mid \mathfrak{F}_t^{(m)}] = \mathbb{E}[g(X_t,\omega)]$$

Using that $\{X_r\}$ is time-homogeneous, the result follows. □

Remark. Since $\mathfrak{M}_t \subseteq \mathfrak{F}_t^{(m)}$, we have that X_t is also a Markov process w.r.t. $\{\mathfrak{M}_t\}_{t\geq 0}$. With the fact $\mathbb{E}^{X_t}[f(X_h)]$ is $\mathfrak{M}_t$-measurable and the Theorem 2.3.3 items 3 and 6, it follows that

$$\begin{aligned}
\mathbb{E}^x[f(X_{t+h}) \mid \mathfrak{M}_t] &= \mathbb{E}^x[\mathbb{E}^x[f(X_{t+h}) \mid \mathfrak{F}_t^{(m)}] \mid \mathfrak{M}_t] \\
&= \mathbb{E}^x[\mathbb{E}^{X_t}[f(X_h)] \mid \mathfrak{M}_t] \\
&= \mathbb{E}^{X_t}[f(X_h)]
\end{aligned}$$

What happens if we replace the time t in (6.2) by a random time $\tau(\omega)$, i.e., a stopping time? The Strong Markov Property states that the equation (6.2) continues to hold, but before proving it, we need to make precise some terms.

To define a stopping time, the intuition is that we should be able to verify whether or not $\tau \leq t$ has occurred using the information contained in an increasing family of σ-algebras.

Definition 6.2.2 (Stopping time)**.** Let $\{\mathfrak{N}_t\}$ be an increasing family of σ-algebras of subsets of Ω. A (strict) **stopping time** (also called a **Markov time**) w.r.t. $\{\mathfrak{N}_t\}$ is a function $\tau : \Omega \longrightarrow [0,\infty]$ such that $\{\omega : \tau(\omega) \leq t\} \in \mathfrak{N}_t$ for all $t \geq 0$.

Definition 6.2.3 (First Exit Time)**.** Given $H \subset \mathbf{R}^n$, we define the **first exist time** from H as

$$\tau_H = \inf\{t > 0 : X_t \notin H\}$$

Remark. The family $\{\mathfrak{M}_t\}$ is right-continuous and τ_H is a stopping time for any Borel set H (assuming that the sets of measure zero are included in $\mathfrak{M}_t$).

Definition 6.2.4. Let τ be a stopping time w.r.t. $\{\mathfrak{N}_t\}$, and let $\mathfrak{N}_\infty$ be the smallest σ-algebra containing $\mathfrak{N}_t$ for all $t \geq 0$. Then the σ-algebra $\mathfrak{N}_t$ consists of all sets $N \in \mathfrak{N}_\infty$ such that

$$N \bigcap \{\tau \leq t\} \in \mathfrak{N}_t, \quad \forall\, t \geq 0$$

Remark. If $\mathfrak{N}_t = \mathfrak{M}_t$, we can describe $\mathfrak{M}_t$ as the σ-algebra generated by $\{X_{\min(s,\tau)} : s \geq 0\}$.

If $\mathfrak{N}_t = \mathfrak{F}_t^{(m)}$, we can describe $\mathfrak{F}_t^{(m)}$ as the σ-algebra generated by $\{B_{s\wedge\tau} : s \geq 0\}$.

Theorem 6.2.2 (Strong Markov Property for Itô Diffusions)**.** Let f be a bounded Borel function defined on $\mathbf{R}^n$, τ a stopping time w.r.t. $\mathfrak{F}_t^{(m)}$, $\tau < \infty$ a.s. then

$$\mathbb{E}^x[f(X_{\tau+h}) \mid \mathfrak{F}_\tau^{(m)}]_{(\omega)} = \mathbb{E}^{X_\tau(\omega)}[f(X_h)], \quad \forall\, h \geq 0 \tag{6.3}$$

Proof. The idea here is to imitate the proof of the Theorem 6.2.1. ☐

The equation (6.3) can be extended as follows.

Corollary 6.2.3. If $f_1, \dots, f_k$ are bounded Borel functions on $\mathbf{R}^n$, τ a stopping time w.r.t. $\mathfrak{F}_t^{(m)}$, $\tau < \infty$ a.s. then

$$\mathbb{E}^x[f_1(X_{\tau+h_1})f_2(X_{\tau+h_2})\cdots f_k(X_{\tau+h_k}) \mid \mathfrak{F}_t^{(m)}] = \mathbb{E}^{X_\tau}[f_1(X_{\tau+h_1})f_2(X_{\tau+h_2})\cdots f_k(X_{\tau+h_k})] \tag{6.4}$$

Proof. By induction on k. ☐

6.2.1 Hitting distribution and Harmonic measure

We end this section by applying the concept of the shift operator to the expected hitting time to a boundary and the probabilistic representation of Harmonic functions.

Definition 6.2.5 (Shift Operator)**.** Let $\mathfrak{H}$ be the set of all real $\mathfrak{M}_\infty$-measurable functions, and $t \geq 0$. The **shift operator**

$$\theta_t : \mathfrak{H} \longrightarrow \mathfrak{H}$$

as follows.

If $\eta = g_1(X_{t_1})\cdots g_k(X_{t_k})$, with each g_i Boreal measurable and $t_i \geq 0$, we define

$$\theta_t \eta = g_1(X_{t_1+t})\cdots g_k(X_{t_k+t})$$

And we extend to all functions in $\mathfrak{H}$ in a natural way: by taking limits of sums of such functions.

It follows from (6.4) that

$$\mathbb{E}^x[\theta_\tau \eta \mid \mathfrak{F}_\tau^{(m)}] = \mathbb{E}^{X_\tau}[\eta] \tag{6.5}$$

for all stopping times τ and all bounded $\eta \in \mathfrak{H}$, with

$$(\theta_\tau \eta)(\omega) = (\theta_t \eta)(\omega), \quad \text{if } \tau(\omega) = t$$

Theorem 6.2.4. Let $H \subset \mathbf{R}^n$ be measurable, τ_H be the first exit time from H for an Itô diffusion X_t, α another stopping time, g a bounded continuous function on $\mathbf{R}^n$ and define

$$\eta = g(X_{\tau_H})\chi_{\{\tau_H < \infty\}}, \quad \tau_H^\alpha = \inf\{t > \alpha : X_t \notin H\}$$

Then

$$\theta_\alpha \eta \chi_{\{\alpha<\infty\}} = g(X_{\tau_H^\alpha})\chi_{\{\tau_H^\alpha < \infty\}}$$

Proof. First approximate η by functions $\eta^{(k)}$.

Then compute $\theta_t \chi_{[t_J, t_{j+1})}(\tau_H) = \chi_{[t_j+t, t_{j+1}+t]}(\tau_H^t)$.

Finally, notice that $\theta_t \eta = g(X_{\tau_H^t})\chi_{\{\tau_H^t < \infty\}}$. ☐

In particular, if $\alpha = \tau_G$, with $G \subset\subset H$ (i.e. $\overline{G}$ is compact and $\overline{G} \subset H$) measurable, and $< \infty$ a.s. w.r.t. Q^x, then $\tau_H^\alpha = \tau_H$ and

$$\theta_{\tau_G} g(X_{\tau_H}) = g(X_{\tau_H}) \tag{6.6}$$

Thus, if f is a bounded measurable function, from (6.5) and (6.6) we have that

$$\mathbb{E}^x[f(X_{\tau_H})] = \mathbb{E}^x[\mathbb{E}^{X_{\tau_G}}[f(X_{\tau_H})]] = \int_{\partial G} \mathbb{E}^y[f(X_{\tau_H})] Q^x[X_{\tau_G} \in dy]$$

for $x \in G$.

Definition 6.2.6 (Harmonic Measure). The **harmonic measure** of X on ∂G, denoted by μ_G^x, is defined as

$$\mu_G^x(F) = Q^x[X_{\tau_G} \in F], \qquad \text{for } F \subset \partial G,\ x \in G$$

The function

$$\varphi(x) = \mathbb{E}^x[f(X_{\tau_H})]$$

satisfies the **mean value property**:

$$\varphi(x) = \int_{\partial G} \varphi(y)\, d\mu_G^x(y), \quad \forall\, x \in G \tag{6.7}$$

for all Borel sets $G \subset\subset H$.

6.3 The Generator of an Itô Diffusion

In this section, we associate a second order partial differential operator to an Itô diffusion.

Definition 6.3.1 (Generator). Let $\{X_t\}$ be a time-homogeneous Itô diffusion in $\mathbf{R}^n$. The (infinitesimal) **generator** A of X_t is defined by

$$A f(x) = \lim_{t \downarrow 0} \frac{\mathbb{E}^x[f(X_t)] - f(x)}{t}, \quad x \in \mathbf{R}^n \tag{6.8}$$

The set of functions $f : \mathbf{R}^n \longrightarrow \mathbf{R}$ such that the limit exists at x is denoted by $\mathfrak{D}_A(x)$, and $\mathfrak{D}_A$ denotes the set of functions for which the limit exists for all $x \in \mathbf{R}^n$.

Now consider the differential operator associated with the SDE (6.1)

$$L f(x) = \sum_i \mu_i(x) \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j} (\sigma \sigma^T)_{i,j}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}$$

But how are the generator and the coefficients of the stochastic differential equation related? We'll prove that A and L coincide on $\mathfrak{C}_0^2(\mathbf{R}^n)$. To show this, we need the following lemma.

Lemma 6.3.1. Let $Y_t = Y_t^x$ be an Itô process in $\mathbf{R}^n$ of the form

$$Y_t^x(\omega) = x + \int_0^t u(s,\omega)\, ds + \int_0^t v(s,\omega)\, dB_s(\omega)$$

where B is m-dimensional.

And let $f \in \mathfrak{C}_0^2(\mathbf{R}^n)$ (notice that this means that f has compact support), τ be a stopping time w.r.t. $\{\mathfrak{F}_t^{(m)}\}$ and assume that $\mathbb{E}^x[\tau] < \infty$. Also assume that $u(t,\omega)$ and $v(t,\omega)$ are bounded on the set of (t,ω) such that $Y(t,\omega)$ belongs to the support of f. Then

$$\mathbb{E}^x[f(Y_\tau)] = f(x) + \mathbb{E}^x\left[\int_0^\tau \left(\sum_i u_i(s,\omega)\frac{\partial f}{\partial x_i}(Y_s) + \frac{1}{2}\sum_{i,j}(vv^T)_{i,j}(s,\omega)\frac{\partial^2 f}{\partial x_i \partial x_j}(Y_s)\right) ds\right]$$

where $\mathbb{E}^x$ is the expectation w.r.t. the probability law R^x for Y_t starting at x:

$$R^x[Y_{t_1} \in F_1, \dots, Y_{t_k} \in F_k] = P^0[Y_{t_1}^x \in F_1, \dots, Y_{t_k}^x \in F_k], \quad F_i \text{ are Borel sets}$$

Proof. First, apply Itô's formula to $f(Y)$.

Notice that $(vdB)_i(vdB)_j = (vv^T)_{ij}dt$ and use this to write an expression for $f(Y_t)$.

Compute the expected value of $f(Y_\tau)$.

Using that g is a bounded Borel function, notice that the expected value of the dB part goes to zero (separate the measurable functions and write the integral as the difference of the measurable and non-measurable parts). □

Theorem 6.3.2 (Formula for the generator)**.** Let X_t be the Itô diffusion

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t$$

If $f \in \mathfrak{C}_0^2(\mathbf{R}^n)$, then $f \in \mathfrak{D}_A$ and

$$Af(x) = \sum_i \mu_i(x)\frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i,j}(\sigma\sigma^T)_{i,j}(x)\frac{\partial^2 f}{\partial x_i \partial x_j} \tag{6.9}$$

Proof. Follows from the Lemma 6.3.1. □

Example 6.3.1 (Standard Brownian Motion)**.** The standard n-dimensional Brownian motion B_t, which satisfies the SDE $dX_t = dB_t$, i.e., with $\mu = 0$ and $\sigma = I_n$, has the generator

$$Af = \frac{1}{2}\sum \frac{\partial^2 f}{\partial x_i^2} = \frac{1}{2}\Delta, \quad f = f(x_1,\dots,x_n) \in \mathfrak{C}_0^2(\mathbf{R}^n)$$

Example 6.3.2 (Graph of Brownian Motion)**.** Consider the two-dimensional process X_t satisfying

$$dX_t = \begin{pmatrix} dt \\ dB_t \end{pmatrix}$$

i.e., with $\mu = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\sigma = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

X_t may be interpreted as the graph of Brownian motion. Its generator is given by

$$Af = \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}, \quad f = f(t,x) \in \mathfrak{C}_0^2(\mathbf{R}^n)$$

6.4 Dynkin's Formula

Using the theory we already have, we can easily draw a useful formula.

Theorem 6.4.1 (Dynkin's Formula)**.** Let $f \in \mathfrak{C}_0^2(\mathbf{R}^n)$ and suppose that τ is a stopping time with $\mathbb{E}^x[\tau] < \infty$. Then

$$\mathbb{E}^x[f(X_\tau)] = f(x) + \mathbb{E}^x\left[\int_0^\tau Af(X_s)\, ds\right] \tag{6.10}$$

Proof. Follows from combining Lemma 6.3.1 with (6.9). □

Notice that, if τ is the first exit time of a bounded set, then the result holds for any function $f \in \mathfrak{C}^2$.

Example 6.4.1. Consider an n-dimensional Brownian motion $B = (B_1, \dots, B_n)$ starting at $a = (a_1, \dots, a_n)$ with $|a| < R$. Let us compute the expected value of the first exit time τ_K of B from the ball

$$K = K_R = \{x \in \mathbf{R}^n : |x| < R\}$$

Choose an integer k and apply Dynkin's formula with $X = B$, $\tau = \sigma_k = \min(k, \tau_K)$, and $f \in \mathfrak{C}_0^2$ such that $f(x) = |x|^2$ for $|x| \le R$:

$$\begin{aligned}\mathbb{E}^a[f(B_{\sigma_k})] &= f(a) + \mathbb{E}^a\left[\int_0^{\sigma_k} \frac{1}{2}\Delta f(B_s)\, ds\right] \\ &= |a|^2 + \mathbb{E}^a\left[\int_0^{\sigma_k} n\, ds\right] \\ &= |a|^2 + n\mathbb{E}^a[\sigma_k]\end{aligned}$$

Since $\mathbb{E}^a[f(B_{\sigma_k})] \le R^2$,

$$\mathbb{E}^a[\sigma_k] \le \frac{1}{n}(R^2 - |a|^2), \quad \forall\ k$$

Taking $k \to \infty$, we have that $\tau_k = \lim \sigma_k < \infty$ a.s. and

$$\mathbb{E}^a[\tau_K] = \frac{1}{n}(R^2 - |a|^2)$$

Now assume that $n \ge 2$ and $|b| > R$. We'll compute the probability that B starting at b ever hits the ball K.

Let α_k be the first exit time from the annulus

$$A_k = \{x : R < |x| < 2^k R\}, \quad k = 1, 2, \dots$$

and put

$$T_K = \inf\{t > 0 : B_t \in K\}$$

We define $f = f_{n,k} \in \mathfrak{C}^2$ with compact support such that

$$\mathrm{R} \le |x| \le 2^k \mathrm{R} \implies f(x) = \begin{cases} -\log|x|, & \text{when } n = 2 \\ |x|^{2-n}, & \text{when } n > 2 \end{cases}$$

Since $\Delta f = 0$ in A_k, by Dynkin's formula,

$$\mathbb{E}^b[f(\mathrm{B}_{\alpha_k})] = f(b), \quad \forall\, k \tag{6.11}$$

Define

$$p_k = \mathrm{P}^b[|\mathrm{B}_{\alpha_k}| = \mathrm{R}] \quad \text{and} \quad q_k = \mathrm{P}^b[|\mathrm{B}_{\alpha_k}| = 2^k\mathrm{R}]$$

In the case $n = 2$, by (6.11) and the definition of f,

$$-\log \mathrm{R} \cdot p_k - (\log \mathrm{R} + k \log 2) \cdot q_k = -\log|b|, \quad \forall\, k$$

Therefore, $\lim_{k\to\infty} q_k = 0$, and thus

$$\mathrm{P}^b[\mathrm{T_K} < \infty] = 1$$

i.e., Brownian motion is **recurrent** in $\mathbf{R}^2$.

In the case $n > 2$,

$$p_k \mathrm{R}^{2-n} + q_k (2^k\mathrm{R})^{2-n} = |b|^{2-n}$$

Since $q_k \in [0,1]$,

$$\lim_{k\to\infty} p_k = \mathrm{P}^b[\mathrm{T_K} < \infty] = \left(\frac{|b|}{\mathrm{R}}\right)^{2-n}$$

i.e., Brownian motion is **transient** in $\mathbf{R}^n$ for $n > 2$.

6.5 The Characteristic Operator

We now introduce an operator closely related to the generator, which is often more suitable, such as in the Dirichlet problem.

Definition 6.5.1 (Characteristic Operator). Let $\{\mathrm{X}_t\}$ be an Itô diffusion. The **characteristic operator** $\mathfrak{A} = \mathfrak{A}_\mathrm{X}$ of $\{\mathrm{X}_t\}$ is defined by

$$\mathfrak{A} f(x) = \lim_{\mathrm{U}\downarrow x} \frac{\mathbb{E}^x[f(\mathrm{X}_{\tau_\mathrm{U}})] - f(x)}{\mathbb{E}^x[\tau_\mathrm{U}]}, \quad x \in \mathbf{R}^n \tag{6.12}$$

where each U is an open set U_k decreasing to the point x in the sense that $\mathrm{U}_{k+1} \subset \mathrm{U}_k$ and $\bigcap_k \mathrm{U}_k = \{x\}$, and $\tau_\mathrm{U} = \inf\{t > 0 : \mathrm{X}_t \notin \mathrm{U}\}$ is the first exit time from U for X_t.

The set of functions $f : \mathbf{R}^n \longrightarrow \mathbf{R}$ such that the limit (6.12) exists for all $x \in \mathbf{R}^n$ and all $\{\mathrm{U}_k\}$ is denoted by $\mathfrak{D}_\mathfrak{A}$.

If $\mathbb{E}^x[\tau_\mathrm{U}] = \infty$ for all open sets $\mathrm{U} \ni x$, we define $\mathfrak{A} f(x) = 0$.

In fact, $\mathfrak{D}_{\mathrm{A}} \subseteq \mathfrak{D}_{\mathfrak{A}}$ and

$$\mathrm{A} f = \mathfrak{A} f \quad \forall f \in \mathfrak{D}_{\mathrm{A}}$$

i.e., the characteristic operator is a generalization of the generator.

We will prove that $\mathfrak{A}_{\mathrm{X}}$ and L_{X} coincide on $\mathfrak{C}^2$. Before that, we need a property of exit times.

Definition 6.5.2 (Trap)**.** A point $x \in \mathbf{R}^n$ is called a **trap** for $\{\mathrm{X}_t\}$ if

$$\mathrm{Q}^x[\{\mathrm{X}_t = t \text{ for all } t\}] = 1$$

Put another way, x is a trap iff. $\tau_{\{x\}} = \infty$ a.s. w.r.t. Q^x.

For example, if $\mu(x_0) = \sigma(x_0) = 0$, then x_0 is a trap for X_t by strong uniqueness of X_t. The intuition here is that the process never leaves the point.

Lemma 6.5.1. If x is not a trap for X_t, then there exists an open set U containing x such that $\mathbb{E}^x[\tau_{\mathrm{U}}] < \infty$.

In words, if the point is not a trap, there exists an open set containing it such that the process starting at it leaves it in finite time.

Theorem 6.5.2. If $f \in \mathfrak{C}^2$, then $f \in \mathfrak{D}_{\mathrm{A}}$ and $\mathfrak{A} f = \mathrm{L} f$.

Proof. First, we need to show that if x is a trap for $\{\mathrm{X}_t\}$, then $\mathfrak{A} f(x) = 0$. To do that, choose a bounded open set containing x. Modifying f to f_0 outside V such that $f_0 \in \mathfrak{C}_0^2(\mathbf{R}^n)$, this claim follows.

Now, if x is not a trap, choose a bounded open set U containing x such that $\mathbb{E}^x[\tau_{\mathrm{U}}] < \infty$. Applying Dynkin's formula and using that $\mathrm{L} f$ is continuous, we have the result. □

Remark. From the discussions in this chapter, we obtained that an Itô diffusion is

1. Continuous;
2. A strong Markov process;
3. Such that the domain of definition of its characteristic operator includes $\mathfrak{C}^2$.

This means that an Itô diffusion is a diffusion in the sense of Dynkin.

6.6 The Feynman-Kac Formula

Relates stochastic differential equations and partial differential equations. Special case: the relationship between geometric Brownian motion and Black-Scholes-Merton model.

Theorem 6.6.1 (Feynman-Kac)**.** Consider the stochastic differential equation

$$\mathrm{dX}_t = \mu(t, \mathrm{X}_t)\mathrm{d}t + \sigma(t, \mathrm{X}_t)\mathrm{dB}_t \tag{6.13}$$

Let $h(y)$ be a Borel-measurable function and define

$$g(t,x) = \mathbb{E}^{t,x}[h(X_T)] \tag{6.14}$$

which we suppose to satisfy $\mathbb{E}^{t,x}[\,|h(X_T)|\,] < \infty$ for all t and x.

Then $g(t,x)$ satisfies the partial differential equation

$$\frac{\partial g}{\partial t}(t,x) + \mu(t,x)\frac{\partial g}{\partial x}(t,x) + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 g}{\partial x^2}(t,x) = 0$$

and the terminal condition

$$g(T,x) = h(x), \quad \forall\, x$$

Lemma 6.6.2. Let X_u be a solution to the stochastic differential equation (6.13) with initial condition given at $t=0$. And let $h(y)$ be a Borel-measurable function, $T>0$, and $g(t,x)$ given by (6.14). Then the stochastic process $g(t,X_t)$, $0 \le t \le T$, is a martingale.

Proof. Let $0 \le s \le t \le T$. By the Markov Property for Itô Diffusions,

$$\mathbb{E}[h(X_T) \mid \mathfrak{F}_s] = g(s,X_s), \quad \mathbb{E}[h(X_T) \mid \mathfrak{F}_t] = g(t,X_t)$$

Taking conditional expectation of the second equation and using iterated expectation,

$$\begin{aligned}\mathbb{E}[g(t,X_t) \mid \mathfrak{F}_s] &= \mathbb{E}[\mathbb{E}[h(X_T) \mid \mathfrak{F}_t] \mid \mathfrak{F}_s] \\ &= \mathbb{E}[h(X_T) \mid \mathfrak{F}_s] = g(s,X_s)\end{aligned}$$

□

Now, we're ready to prove Feynman-Kac.

Proof. **First step: Find the martingale.**

Let X_t be the solution to the stochastic differential equation (6.13) starting at zero.

Since $g(t,X_t)$ is a martingale (by the lemma 6.6.2), the term dt must be zero.

Second step: take the differential.

Thus (omitting the argument (t,X_t)),

$$\begin{aligned}dg &= g_t dt + g_x dX + \frac{1}{2} g_{xx} dX dX \\ &= g_t dt + \mu g_x dt + \sigma g_x dB_t + \frac{1}{2}\sigma^2 g_{xx} dt \\ &= \left(g_t + \mu g_x + \frac{1}{2}\sigma^2 g_{xx}\right) dt + \sigma g_x dB_t\end{aligned}$$

Third step: set the dt term equal to zero.

Setting dt to zero,

$$g_t + \mu g_x + \frac{1}{2}\sigma^2 g_{xx} = 0$$

along every path of X. □

Theorem 6.6.3 (Discounted Feynman-Kac)**.** Consider the stochastic differential equation

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t$$

Let $h(y)$ be a Borel-measurable function and r be a constant. Fix $T > 0$ and let $t \in [0, T]$. Define

$$g(t,x) = \mathbb{E}^{t,x}[e^{-r(T-t)}h(X_T)]$$

which we suppose to satisfy $\mathbb{E}^{t,x}[\,|h(X_T)|\,] < \infty$ for all t and x.

Then $g(t,x)$ satisfies the partial differential equation

$$\frac{\partial g}{\partial t}(t,x) + \mu(t,x)\frac{\partial g}{\partial x}(t,x) + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 g}{\partial x^2}(t,x) = rg(t,x)$$

and the terminal condition

$$g(T,x) = h(x), \quad \forall\ x$$

Proof. As in the previous proof, we let X_t be the solution of the stochastic differential equation starting at zero.

The difference here is that $g(t, X_t)$ is not a martingale. For $0 \le s \le t \le T$, we have

$$\begin{aligned}\mathbb{E}[g(t,X_t) \mid \mathfrak{F}_s] &= \mathbb{E}[\mathbb{E}[e^{-r(T-t)}h(X_T) \mid \mathfrak{F}_t] \mid \mathfrak{F}_s] \\ &= \mathbb{E}[e^{-r(T-t)}h(X_T) \mid \mathfrak{F}_s]\end{aligned}$$

However,

$$g(s, X_s) = \mathbb{E}[e^{-r(T-s)}h(X_T) \mid \mathfrak{F}_s]$$

The idea now is to 'complete the discounting', since the random variable should not depend on t. Notice that

$$e^{-rt}g(t,X_t) = \mathbb{E}[e^{-rT}h(X_T) \mid \mathfrak{F}_t]$$

We apply iterated conditioning to show that $e^{-rt}g(t, X_t)$ is a martingale:

$$\begin{aligned}\mathbb{E}[e^{-rt}g(t,X_t) \mid \mathfrak{F}_s] &= \mathbb{E}[\mathbb{E}[e^{-rT}h(X_T) \mid \mathfrak{F}_t] \mid \mathfrak{F}_s] \\ &= \mathbb{E}[e^{-rT}h(X_T) \mid \mathfrak{F}_s] \\ &= e^{-rs}g(s, X_s)\end{aligned}$$

Computing the differential,

$$\begin{aligned}d(e^{-rt}g(t,X_t)) &= e^{-rt}\left[-rg\,dt + \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial x}dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(dX_t)^2\right] \\ &= e^{-rt}\left[-rg + \frac{\partial g}{\partial t} + \mu\frac{\partial g}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 g}{\partial x^2}\right]dt + e^{-rt}\sigma\frac{\partial g}{\partial x}dB_t\end{aligned}$$

Set dt to zero and we have our result. □

Chapter 7

Application to Option Pricing

In this chapter, we present an application of the stochastic differential equations to the problem of option pricing.

7.1 Introduction

7.1.1 Risk Neutral Valuation: Discrete Model Intuition

The problem we're interested in is how to price derivatives whilst minimizing the risk. Derivatives are contingent claims 'that promise some payment or delivery in the future contingent on an underlying stock's behavior.' [BRR96, p. 3]

The simplest derivative is a **forward contract**, which pays to the holder $S-K$ at the time T, where S_T is the **stock price** at T and K is the **exercise price** or **strike price**. The time T is called **maturity time**.

Most commonly in this chapter, we'll be interested in a kind of derivative called **options**. The main difference here is that the buyer has the right, but not the obligation, to exercise the contract.

1. A **call option** pays $\max\{S-K, 0\}$ at time T. Intuitively, it's a bet that the price will go up, giving the buyer the option to receive the stock for the strike price K.

2. A **put option** pays $\max\{K-S, 0\}$ at time T. It's a bet that the price will go down and gives the buyer the option to sell the stock for the strike price K.

The options defined above are called **European**. Another kind is the **American option**, which can be executed before maturity.

Given this language, how do you price an option at the time $t=0$? And how to hedge it?

If the economy includes a stock S and a **riskless asset** B (a bond, for example) with **interest rate** r and derivative claim f, we may assume only two outcomes at time Δt either it goes up with probability p or down, with probability $(1-p)$ (see Figure 7.1).

We may think that the strike price K should be defined by the real-world probability p. However, this leads to arbitrage problems. For simplicity, we assume a **viable market**, i.e., a market without arbitrage opportunity.

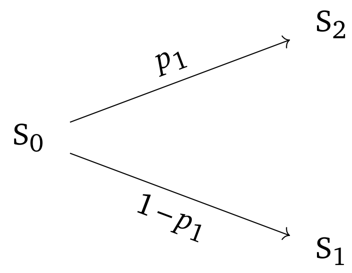

Figure 7.1: Binomial branch

Consider instead the following strategy. Borrow S_0 to buy the stock and enter a forward contract with strike price K_0. In time Δt, deliver stock in exchange for K_0 and repay the loan, which now costs $S_0 e^{r\Delta t}$.

Then, the outcome is

- If $K_0 > S_0 e^{r\Delta t}$, then we made riskless profit.
- If $K_0 < S_0 e^{r\Delta t}$, then we lost money.

Hence, the enforced strike price must be $K_0 = S_0 e^{r\Delta t}$, i.e., 'the current price of a derivative claim is determined by current price of a portfolio which exactly replicates the payoff of the derivative at the maturity'.

We want to find a, b such that

$$f_2 = aS_2 + bB_0 e^{r\Delta t}$$
$$f_1 = aS_1 + bB_0 e^{r\Delta t}$$

where f_2 is the price of the derivative if the price goes up and f_1, if it goes down.

Solving for a and b,

$$a = \frac{f_2 - f_1}{S_2 - S_1}, \quad b = \frac{S_2 f_1 - S_1 f_2}{(S_2 - S_1)B_0 e^{r\Delta t}}$$

Substituting the values above into

$$f_0 = aS_0 + bB_0$$

we obtain

$$f_0 = e^{-r\Delta t}\left(S_0 e^{r\Delta t}\frac{f_2 - f_1}{S_2 - S_1} + \frac{S_2 f_1 - S_1 f_2}{S_2 - S_1}\right)$$

Thus, it's possible to rewrite

$$f_0 = e^{-r\Delta t}(f_2 q + f_1(1-q))$$

where

$$q = \frac{S_0 e^{r\Delta t} - S_1}{S_2 - S_1}, \quad 0 < q < 1$$

is the **risk-neutral probability**, or **martingale probability**, i.e., the **arbitrage probability** that the stock price goes up.

Hence,

$$S_2 q + S_1(1-q) = e^{r\Delta t} S_0$$

Theorem 7.1.1. Summarizing,

1. Arbitrage probability: $q = \frac{S_0 e^{r\Delta t} - S_1}{S_2 - S_1}$;
2. Claim value: $f_0 = e^{r\Delta t}(qf_2 + (1-q)f_1)$;
3. Claim value at time 0: $V = \mathbb{E}_Q(B_T^{-1}X)$, where X is the claim pay-off;
4. Stock holding strategy: $\varphi = \frac{f_2 - f_1}{S_2 - S_1}$;
5. Bond holding strategy: $\psi = B_0^{-1}(f_0 - \varphi S_0)$.

Notice that $u = e^{\sigma \Delta t}$, where σ is the underlying volatility.

Given S_0, K, T, r as defined, let N be the number of time steps, u the up-factor and let $d = 1/u$ to ensure that the tree is recombinant. Then, we have the following algorithm.

Algorithm 1: Cox-Ross-Rubinstein Binomial Model

Input: S_0, K, T, r, N, u, d
Output: C_0

```
// Initialize constants
```
$dt \leftarrow T/N$;
$q \leftarrow (e^{r\cdot dt} - d)/(u-d)$;
$disc \leftarrow exp^{-r\cdot dt}$;

```
// Compute asset prices at maturity
```
$S[0] = S_0 \cdot d^N$;
for $j \in \{1, \ldots, N+1\}$ **do**
  $S[j] \leftarrow S[j-1] \cdot u/d$;

```
// Compute option values at maturity
```
for $j \in \{0, 1, \ldots, N+1\}$ **do**
  $C[j] \leftarrow \max\{0, S[j] - K\}$;

```
// Step backwards through the tree
```
for $i \in \{N, N-1, \ldots, 0\}$ **do**
  for $j \in \{0, \ldots, i\}$ **do**
    $C[j] = disc \cdot \big(q \cdot C[j+1] + (1-q) \cdot C[j]\big)$;

return $C[0]$

7.1.2 Continuous Model Intuition

Since the claim in discrete model has binomial distribution, by the central limit theorem the distribution of the claim converges to a normal. Then, as the intertick time Δt gets smaller and the number of time steps N gets larger, the distribution of S_t becomes log-normal. Thus, we can write

$$dS = \mu S dt + \sigma S dW$$

where W is a Brownian motion.

Our goal is to find a replicating portfolio such that

$$df = adS + bdB$$

Applying Itô's formula to df and substituting dS, df, $dB = rBdt$ and $(dS)^2 = \sigma^2 S^2 dt$,

$$df = (a\mu S + brB)dt + a\sigma S dW$$

Comparing terms,

$$a = \frac{\partial f}{\partial S}, \quad brB = \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2$$

Since $bB = f - aS$ is deterministic and $dB = rBdt$,

$$d(f - aS) = r(f - aS)dt$$

and we obtain the Black-Scholes equation

$$\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2 + \frac{\partial f}{\partial S} rS - rf = 0$$

Remark.
1. Notice that the Black-Scholes model assumes that the volatility is constant.
2. Any tradable derivative satisfies the equation.
3. Since there's no dependence on the drift μ, this gives a hedging strategy by replicating portfolio.

Changing variables, we have the heat equation

$$\frac{\partial u}{\partial \tau} = \frac{\partial^2 u}{\partial x^2}$$

where the pay-off of the derivative gives boundary and final conditions.

For an European Call, we have:

$$C(S, T) = \max\{S - K, 0\}$$

$$C(0, t) = 0, \quad C(\infty, t) \approx S$$

For an European Put:

$$P(S, T) = \max\{K - S, 0\}$$

$$P(0, t) = Ke^{-r(T-t)}, \quad P(\infty, t) = 0$$

For European call/put, we have the following analytical solution:

$$C_t = e^{-r(T-t)}\left(e^{r(T-t)} S N(d_1) - K N(d_2)\right)$$

$$P_t = e^{-r(T-t)}\left(K N(-d_2) - e^{r(T-t)} S N(-d_1)\right)$$

where

$$d_1 = \frac{\ln(S/K) + (r + \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}}$$

$$d_2 = \frac{\ln(S/K) + (r - \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}}$$

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du$$

With Q being the risk neutral measure under which $ds = rSdt + \sigma Sdw$, we have

$$f_t = e^{-r(T-t)}\mathbb{E}_Q[f_T]$$

In more complicated options, it is necessary to use numerical methods, like finite difference, tree methods, and Monte Carlo.

7.1.3 Option Price and Probability Duality

In this section, we'll interpret option prices as probability distributions. Let us fix the following notation:

- Call option with payout at T: $\max(S_T - K, 0)$.
- ZCB (zero coupon bond): 1.
- Digital option: $\chi_{\{S_T > T\}}$.
- T is maturity.
- K is the strike.
- Price of a call at $t \leq 1$: $C_K(t, T)$.
- Price of the ZCB at $t \leq 1$: $Z(t, T)$.
- Price of the digital at $t \leq 1$: $D_K(t, T)$.

Consider the portfolio: λ calls with strike K, $-\lambda$ calls with strike $K + 1/\lambda$. The price at t is given by $\lambda C_K(t, T) - \lambda C_{K+1/\lambda}(t, T)$. The payout at T (call spread) is 0 between 0 and K and 1 after $K + 1/\lambda$ with a straight line between them.

Let $\lambda \to \infty$: $-\frac{\partial C}{\partial K}$. Then the payout function becomes 'instantaneously one', it is the payout of the digital:

$$-\frac{\partial C}{\partial K}(t, T) = D_K(t, T)$$

By the fundamental theorem of asset pricing, prices today (at t) are the expected payout at T suitably discounted. Thus, we can write

$$D_K(t, T) = Z(t, T)\mathbb{E}^*[\chi_{\{S_T > K\}} \mid S_t] = Z(t, T)\mathbb{P}^*[S_T > K \mid S_t]$$

under the risk-neutral distribution.

In fact,

$$\frac{D_K(t, T)}{Z(t, T)}$$

is a martingale.

By equating two prices:

$$-\frac{\partial C}{\partial K}(t,T) = Z(t,T)\mathbb{P}^*[S_T > K \mid S_t] \iff \mathbb{P}^*[S_T < K \mid S_t] = 1 + \frac{\partial C}{\partial K}(t,T)\frac{1}{Z(t,T)}$$

Differentiating, we obtain the density function

$$f_{S_T \mid S_t}(x) = \frac{1}{Z(t,T)} \frac{\partial^2 C}{\partial K^2}\bigg|_x$$

Thus, given the set of call prices, we can determine the density of the underlying asset.

Now consider the **call butterfly** portfolio, given as follows: λ calls with strike $K - 1/\lambda$, -2λ calls with strike K and λ calls with strike $K + 1/\lambda$, $\lambda > 0$.

This approximates the density function. Its price is

$$\lambda(C_{K-1/\lambda}(t,T) - C_K(t,T)) - \lambda(C_K(t,T) - C_{K+1/\lambda}(t,T)) =: B_{K,\lambda}(t,T)$$

Take

$$\lambda B_{K,\lambda}(t,T) \approx \frac{\partial^2 C}{\partial K^2}(t,T)$$

For large λ,

$$B_{K,\lambda}(t,T) \approx \frac{1}{\lambda} f(x)\bigg|_K$$

Note that none of this depends on the option price. Putting Black-Scholes at C and taking the second derivative with respect to K, we end up with a log-normal distribution.

The fundamental theorem of asset pricing states that for a derivative with payout $D(T,T) = g(S_T)$ at T and price $D(t,T)$ at $t \leq T$, we have

$$\frac{D(t,T)}{Z(t,T)} = \mathbb{E}^*\left[\frac{D(T,T)}{Z(T,T)} \,\middle|\, S_t\right]$$

In other words, $\frac{D(t,T)}{Z(t,T)}$ is a martingale with respect to the stock price under the risk-neutral distribution.

Then

$$D(t,T) = Z(t,T)\mathbb{E}^*[D(T,T) \mid S_t] = Z(t,T)\int g(x) f_{S_T \mid S_t}(x)\, dx$$

The theorem means we can go from $f(x)$, the probability density, to $D(t,T)$, the derivative price.

Remark. Call prices do span all derivative prices. Piecewise linear g replicating portfolio.

Write the Taylor expansion:

$$g(S_T) = g(0) + S_T g'(0) + \int_0^\infty (S_T - K)^+ g''(K)\, dK$$

Take the discounted expected value:

$$D(t,T) = Z(t,T)g(0) + g'(0)S_t + \int_0^\infty C_K(t,T) g''(K)\, dK$$

which gives the number of bonds and stocks, and the portfolio of calls.

7.2 Discrete Time Models

7.2.1 European Options

More generally, letting N be the horizon of our investment, i.e., the maturity of our options, we have the following definitions

Definition 7.2.1 (Market).

1. Our market consists of $d+1$ assets with prices $S_n^0, S_n^1, \dots, S_n^d$ at time n. These prices are positive random variables which are $\mathfrak{F}_n$-measurable.
2. The asset indexed by zero is the **riskless asset**, which we set $S_0^0 = 1$ with return r over one period. I.e., $S_n^0 = (1+r)^n$.
3. The assets indexed by $i = 1, 2, \dots, d$ are the **risky assets**.
4. The factor $\beta_n = 1/S_n^0$ is the **discount factor**.

Definition 7.2.2 (Trading Strategy). A **trading strategy** is the stochastic process

$$\varphi = ((\varphi_n^0, \varphi_n^1, \dots, \varphi_n^d))_{0 \le n \le N}$$

where each φ_n^i is the number of shares of the asset i at time n.

We assume that this process is **predictable**, i.e., each φ_0^i is $\mathfrak{F}_0$-measurable and, for $n \ge 1$, φ_n^i is $\mathfrak{F}_{n-1}$-measurable.

Definition 7.2.3 (Value of the portfolio). We define

1. The **value of the portfolio** at time n:

$$V_n(\varphi) = \varphi_n S_n = \sum_{i=0}^{d} \varphi_n^i S_n^i$$

2. The **discounted prices**:

$$\tilde{S}_n = \beta_n S_n = (1, \beta_n S_n^1, \dots, \beta_n S_n^d)$$

3. The **discounted value of the portfolio**:

$$\tilde{V}_n(\varphi) = \beta_n(\varphi_n S_n) = \varphi_n \tilde{S}_n$$

Definition 7.2.4 (Self-financing strategy). A trading strategy is called **self-financing** if it readjusts positions without bringing or consuming wealth, i.e.,

$$\varphi_n S_n = \varphi_{n+1} S_n, \quad \forall\ 0 \le n \le N-1$$

Remark. The identity $\varphi_n S_n = \varphi_{n+1} S_n$ is equivalent to

$$\varphi_{n+1}(S_{n+1} - S_n) = \varphi_{n+1} S_{n+1} - \varphi_{n+1} S_n = \varphi_{n+1} S_{n+1} - \varphi_n S_n$$

and

$$V_{n+1}(\varphi) - V_n(\varphi) = \varphi_{n+1}(S_{n+1} - S_n)$$

Now we present a characterization of self-financing strategies.

Proposition 7.2.1. The following are equivalent:

1. The strategy φ is self-financing.
2. For any $n \in \{1, \dots, N\}$,
$$V_n(\varphi) = V_0(\varphi) + \sum_{i=1}^{n} \varphi_i \Delta S_i$$
3. For any $n \in \{1, \dots, N\}$,
$$\tilde{V}_n(\varphi) = V_0(\varphi) + \sum_{i=1}^{n} \varphi_i \Delta \tilde{S}_i$$

Proof. To show that 1. and 2. are equivalent, notice that

$$V_n(\varphi) - V_0(\varphi) = \sum_{i=1}^{n} [V_i(\varphi) - V_{i-1}(\varphi)] = \sum_{i=1}^{n} \varphi_i \Delta S_i$$

by the previous remark.

Since $\varphi_n S_n = \varphi_{n+1} S_n \iff \varphi \tilde{S}_n = \varphi_{n+1} \tilde{S}_n$, the equivalence 1. and 3. holds. □

Proposition 7.2.2. Let $((\varphi_n^1, \dots, \varphi_n^d))_{0 \leq n \leq N}$ be a predictable process and V_0 a $\mathfrak{F}_0$-measurable random variable. There exists a unique predictable process $(\varphi_n^0)_{0 \leq n \leq N}$ such that $(\varphi^0, \varphi^1, \dots, \varphi^d)$ is self-financing and its initial value is V_0.

Proof. By the self-financing condition, we can write

$$\begin{aligned}
\tilde{V}_n(\varphi) &= \varphi_0^n + \varphi_n^1 \tilde{S}_n^1 + \dots + \varphi_n^d \tilde{S}_n^d \\
&= V_0 + \sum_{i=1}^{n} \varphi_i \Delta \tilde{S}_i
\end{aligned}$$

Isolating φ_n^0 we obtain the process and, by canceling the nth term, predictability follows. □

Definition 7.2.5 (Admissible strategy)**.** A strategy is called **admissible** if it is self-financing and $V_n(\varphi) \geq 0$ for all n.

Definition 7.2.6 (Arbitrage)**.** 'An **arbitrage** strategy is an admissible strategy with zero initial value and non-zero final value.'

To exclude arbitrage, we use martingales. In the context of financial assets, saying that (S_n^i)

is a martingale means that the best estimate for S^i_{n+1} is S_i.

Proposition 7.2.3. Let (M_n) be a martingale and (H_n) a predictable sequence w.r.t. the filtration $(\mathfrak{F}_n)$. The sequence (X_n) given by

$$X_n = \begin{cases} H_0M_0, & \text{if } n = 0 \\ H_0M_0 + \sum_{i=1}^n H_i\Delta M_i, & \text{if } n \geq 1 \end{cases}$$

is a martingale w.r.t. $(\mathfrak{F}_n)$.

Proof. Using that (H_n) is predictable and (M_n) is a martingale, compute

$$\mathbb{E}[X_{n+1} - X_n \mid \mathfrak{F}_n] = \mathbb{E}[H_{n+1}\Delta M_{n+1} \mid \mathfrak{F}_n] = H_{n+1}\mathbb{E}[M_{n+1} - M_n \mid \mathfrak{F}_n] = 0$$

Thus, (X_n) is a martingale. □

This proposition implies that, if the vector of discounted prices is a martingale, then the expected value of the wealth generated by a self-financing strategy equals the initial wealth.

The following is a characterization of Martingales.

Proposition 7.2.4. An adapted sequence of real-valued random variables (M_n) is a martingale if and only if for any predictable sequence (H_n), we have

$$\mathbb{E}\left(\sum_{n=1}^N H_n\Delta M_n\right) = 0$$

Proof. $(\Rightarrow)$ Let

$$X_n = \begin{cases} 0, & \text{if } n = 0 \\ \sum_{i=1}^n H_i\Delta M_i, & \text{if } n \geq 1 \end{cases}$$

By the previous proposition, (X_n) is a martingale. Thus, $\mathbb{E}[X_n] = \mathbb{E}[X_0] = 0$.

$(\Leftarrow)$ Fix $1 \leq j \leq N$ and an $\mathfrak{F}_j$-measurable set A.

Consider

$$H_n = \begin{cases} 0, & \text{if } n \neq j+1 \\ \chi_A, & \text{if } n = j+1 \end{cases}$$

Then H_n is predictable and

$$0 = \mathbb{E}\left[\sum_{n=1}^N H_n\Delta M_n\right] = \mathbb{E}[\chi_A(M_{j+1} - M_j)] = \mathbb{E}[M_{j+1} - M_j \mid \mathfrak{F}_j]$$

Hence, (M_n) is a martingale. □

Definition 7.2.7 (Viable Markets)**.** A market is called **viable** if there's no arbitrage opportunity.

In the next result, known as the Fundamental Theorem of Asset Pricing, we'll show how to use martingales to exclude arbitrage. Before that, we need two definitions and a lemma.

Definition 7.2.8 (Equivalent probability measures)**.** Two probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ are **equivalent** if, for any event A, we have that $\mathbb{P}_1(A) = 0 \iff \mathbb{P}_2(A) = 0$.

Intuitively, equivalent probabilities agree on what is possible.

Notice that this definition means that, for any $\omega \in \Omega$, $\mathbb{P}^*[\{\omega\}] > 0$.

Definition 7.2.9 (Cumulative Discounted Gain)**.** The cumulative discounted gain realized by following the self-financing strategy $\varphi_n^1, \dots, \varphi_n^d$ is

$$\tilde{G}_n(\varphi) = \sum_{j=1}^{n} \left(\varphi_j^1 \Delta \tilde{S}_j^1 + \cdots + \varphi_j^d \Delta \tilde{S}_j^d\right)$$

Lemma 7.2.5. Let Γ be the set of all non-negative random variables X such that $\mathbb{P}[X > 0] > 0$, $(\varphi_n^1, \dots, \varphi_n^d)$ be an admissible process, and $\tilde{G}_n(\varphi)$ be the cumulative discounted gain.

If the market if viable, then any predictable process $(\varphi^1, \dots, \varphi^d)$ satisfies $\tilde{G}_N(\varphi) \notin \Gamma$.

Proof. We proceed by contrapositive. Suppose that $\tilde{G}_N(\varphi) \in \Gamma$. If all $\tilde{G}_n(\varphi) \geq 0$, the market is not viable.

Suppose that not all $\tilde{G}_n(\varphi)$ are non-negative. Let

$$n = \sup\{k : \mathbb{P}[\tilde{G}_k(\varphi) < 0] > 0\}$$

Then we can construct a new process

$$\psi_j(\omega) = \begin{cases} 0 & \text{if } j \leq n \\ \chi_A(\omega)\varphi_j(\omega) & \text{if } j > n \end{cases}$$

where A is the event $\{\tilde{G}_n(\varphi) < 0\}$.

Thus, $\tilde{G}_j(\omega) \geq 0$ for all $j \in \{0, 1, \dots, N\}$ and the market is not viable. □

Theorem 7.2.6 (Fundamental Theorem of Asset Pricing)**.** The market is viable if and only if there exists a probability $\mathbb{P}^*$, equivalent to $\mathbb{P}$, such that the discounted prices are $\mathbb{P}^*$-martingales.

Proof. ($\Leftarrow$) Suppose that $\mathbb{P}^*$ and $\mathbb{P}$ are equivalent and that the discounted prices $(\tilde{S}_n(\varphi))$ are $\mathbb{P}^*$-martingales.

By the Proposition 7.2.1,

$$\tilde{V}_n(\varphi) = V_0(\varphi) + \sum_{j=1}^{n} \varphi_j \Delta \tilde{S}_j$$

And by the Proposition 7.2.3, $(\tilde{V}_n(\varphi))$ is a $\mathbb{P}^*$-martingale. Thus,

$$\mathbb{E}^*[\tilde{V}_N(\varphi)] = \mathbb{E}^*[\tilde{V}_0(\varphi)]$$

To prove that there's no arbitrage opportunity, suppose that the strategy is admissible and its initial value is zero. Then,

$$\tilde{V}_N(\varphi) \geq 0 \quad \text{and} \quad \mathbb{E}^*[\tilde{V}_N(\varphi)] = 0$$

Since the probability measures are equivalent, $\mathbb{P}^*[\{\omega\}] > 0$. Which implies that $\tilde{V}_N(\varphi) = 0$.

($\Rightarrow$) Let Γ be as in the Lemma 7.2.5 and $\tilde{G}_n(\varphi)$ be the cumulative discounted gain. Notice that Γ is a convex cone in the vector space of real-valued random variables (Hahn–Banach separation theorem: the separation in $\mathbb{R}^n$ is a hyperplane).

Since we're supposing that the market is viable, we have that

$$V_0(\varphi) = 0 \implies \tilde{V}_N(\varphi) \notin \Gamma$$

By the Proposition 7.2.2, there exists a unique (φ_n^0) such that the strategy $((\varphi_n^0, \ldots, \varphi_n^d))$ is self-financing and has zero initial value.

By the hypothesis that the market is viable and $\tilde{G}_n(\varphi) \geq 0$, it follows that $\tilde{G}_N(\varphi) = 0$. By the Lemma 7.2.5, a stronger fact holds: even if not all $\tilde{G}_n(\varphi)$ are non-negative, we still have that $\tilde{G}_N(\varphi) = 0$.

Now we can construct our risk-neutral measure. Notice that

$$\mathfrak{V} = \{\tilde{G}_N(\varphi : \varphi \in \mathbb{R}^d \text{ is predictable})\}$$

is a vector subspace of $\mathbb{R}^\Omega$.

By the Lemma 7.2.5, $\mathfrak{V} \cap \Gamma = \emptyset$. Thus, $\mathfrak{V}$ doesn't intersect

$$K = \{X \in \Gamma : \sum_\omega X(\omega = 1)\} \subseteq \Gamma$$

which is a convex compact set.

Using the convex sets separation theorem, there exists $(\lambda(\omega))_\omega$ such that

1. For all $X \in K$,
$$\sum_\omega \lambda(\omega) X(\omega) > 0$$

2. If φ is predictable,
$$\sum_\omega \lambda(\omega) \tilde{G}_N(\varphi)(\omega) = 0$$

By the first property, $\lambda(\omega) > 0$ for all $\omega \in \Omega$. Thus, the probability $\mathbb{P}^*$ defined by

$$\mathbb{P}^*(\{\omega\}) = \frac{\lambda(\omega)}{\sum_{\omega' \in \Omega} \lambda(\omega')}$$

is equivalent to $\mathbb{P}$.

Using the second property,

$$\mathbb{E}^*\left(\sum_{j=1}^{N}\varphi_j\Delta\tilde{S}_j\right)=0 \implies \mathbb{E}^*\left(\sum_{j=1}^{n}\varphi_j^i\Delta\tilde{S}_j^i\right)=0$$

Hence, by the Proposition 7.2.4, the discounted prices $(\tilde{S}_n^1),\dots,(\tilde{S}_n^d)$ are $\mathbb{P}^*$-martingales. □

A European option is characterized by its payoff h, which is a non-negative $\mathfrak{F}_n$-measurable random variable.

1. For a **call** on the asset i with strike price K, we have $h=(S_N^i-K)$.
2. For a **put** on the same asset, $h=(K-S_N^i)$.

Definition 7.2.10 (Attainable claim)**.** A contingent claim defined by h is **attainable** if there exists an admissible strategy worth h at maturity N.

Remark. If the market is viable, then it is sufficient to find a self-financing strategy worth h at maturity to say that h is attainable.

In fact, if φ is self-financing and $\mathbb{P}^*$ is a risk-neutral measure, then $(\tilde{V}_n(\varphi))$ is a $\mathbb{P}^*$-martingale and $\tilde{V}_n(\varphi)=\mathbb{E}^*[\tilde{V}_N(\varphi)\mid\mathfrak{F}_n]$.

Thus, if $\tilde{V}_N(\varphi)\geq 0$, the strategy is admissible.

Definition 7.2.11 (Complete market)**.** The market is **complete** if every contingent claim is attainable.

Altough a complete market is a restrictive assumption, it allows us to deduce an important theory of option pricing and hedging.

Theorem 7.2.7. A viable market is complete if and only if there exists unique measure $\mathbb{P}^*$ equivalent to $\mathbb{P}$ under which discounted prices are martingales.

Proof. ($\Rightarrow$) If the market is viable and complete, any non-negative $\mathfrak{F}_N$-measurable random variable h can be written as $h=V_N(\varphi)$, in which φ is an admissible strategy that replicates the claim h.

Since φ is self-financing,

$$\frac{h}{S_N^0}=\tilde{V}_N(\varphi)=V_0(\varphi)+\sum_{j=1}^{N}\varphi_j\Delta\tilde{S}_j$$

If there are two probability measures under which the discounted prices are martingales, by computing the expected value of $\tilde{V}_N(\varphi)$ under each measure, we see that the measures agree on the whole σ-algebra $\mathfrak{F}_N$, i.e., the probability measure is unique.

($\Leftarrow$) Notice that the Fundamental Theorem of Asset Pricing implies that the market is viable.

Suppose that the market is viable and incomplete. Then there exists a random variable

$h \geq 0$ that is not attainable. Define $\tilde{\mathfrak{V}}$ the set of random variables of the form

$$U_0 + \sum_{n=1}^{N} \varphi_n \Delta \tilde{S}_n$$

in which U_0 is $\mathfrak{F}_0$-measurable and $((\varphi_n^1, \ldots, \varphi_n^d))$ is predictable.

By the Proposition 7.2.2 and the previous remark, h/S_N^0 does not belong to $\tilde{\mathfrak{V}}$, i.e., $\tilde{\mathfrak{V}}$ is a strict subset of all random variables on $(\Omega, \mathfrak{F})$.

Define the following scalar product on the set of random variables

$$(X, Y) \longmapsto \mathbb{E}^*[XY]$$

and notice that there exists a non-zero random variable X orthogonal to $\tilde{\mathfrak{V}}$ (via orthogonal projection).

Now define

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

Since $\mathbb{E}^*[X] = 0$, we obtain a new probability measure equivalent to $\mathbb{P}$ and different from $\mathbb{P}^*$ satisfying

$$\mathbb{E}^{**}\left(\sum_{n=1}^{N} \varphi_n \Delta \tilde{S}_n\right) = \int_\Omega \sum_{n=1}^{N} \varphi_n \Delta \tilde{S}_n \, d\mathbb{P}^* + \int_\Omega \frac{1}{2\|X\|_\infty} \left(\sum_{n=1}^{N} \varphi_n \Delta \tilde{S}_n\right) \cdot X \, d\mathbb{P}^* = 0$$

where the first integral is zero, since the process is a martingale, and the second one is zero because it is the definition of inner product and X is orthogonal do $\tilde{\mathfrak{V}}$.

Thus, from the Proposition 7.2.4, $(\tilde{S}_n)$ is a $\mathbb{P}^{**}$-martingale, contradicting the uniqueness hypothesis. □

Example 7.2.1. How to price and hedge contingent claims in a viable and complete market?

Suppose $h = V_N(\varphi)$, i.e., we have an admissible strategy replicating the claim.

Since $(\tilde{V}_n)$ are $\mathbb{P}^*$-martingales, we have

$$V_0(\varphi) = \mathbb{E}^*[\tilde{V}_N(\varphi)] = \mathbb{E}^*[h/S_N^0]$$

Hence,

$$V_n(\varphi) = S_n^0 \mathbb{E}^*\left[\frac{h}{S_N^0} \middle| \mathfrak{F}_n\right]$$

is completely determined by h.

Thus, we may call $V_n(\varphi)$ the **value of the option** at n.

Note that the investor is **perfectly hedged** if, at time zero, he sells the option for $\mathbb{E}^*[h/S_N^0]$, which is called the **fair price** of the option. The computation of the option price only requires the knowledge of the risk-neutral probability and doesn't depend on the true probability.

7.2.2 American Options

An American option may be exercised at any time before maturity. So we define as a non-negative sequence $(Z_n)_{0\leq n\leq N}$ adapted to $\mathfrak{F}_n$ in which Z_n is the profit made by exercising the option at time n.

Naturally, for a call option we have $Z_n = (S_n - K)$ and for a put option, $Z_n = (K - S_n)$. But how do we price American options?

At maturity, the value of the option U_N is equal to Z_N. At $N-1$, it is the maximum between Z_{N-1} and the value at $N-1$ of an strategy paying Z_N at N, i.e.,

$$U_{N-1} = \max\{Z_{N-1},\ S^0_{N-1}\mathbb{E}^*[\tilde{Z}_N \mid \mathfrak{F}_{N-1}]\}$$

where $\tilde{Z}_N = Z_N/S^0_N$.

By induction,

$$U_{n-1} = \max\left\{Z_{n-1},\ S^0_{n-1}\mathbb{E}^*\left[\frac{U_n}{S^0_n}\,\middle|\,\mathfrak{F}_{n-1}\right]\right\}$$

If $S^0_n = (1+r)^n$,

$$U_{n-1} = \max\left\{Z_{n-1},\ \frac{1}{1+r}\mathbb{E}^*[U_n \mid \mathfrak{F}_{n-1}]\right\}$$

Proposition 7.2.8. Let $\tilde{U}_n = U_n/S^0_n$ be the discounted price of the American option. Then $(\tilde{U}_n)$ is a $\mathbb{P}^*$-supermartingale and is the smallest $\mathbb{P}^*$-supermartingale that dominates $(\tilde{Z}_n)$.

Proof. By definition of $\tilde{U}_n$,

$$\tilde{U}_{n-1} = \max\{\tilde{Z}_{n-1},\ \mathbb{E}^*[\tilde{U}_n \mid \mathfrak{F}_{n-1}]\}$$

Thus, $(\tilde{U}_n)$ is a supermartingale that dominates $(\tilde{Z}_n)$.

To show that it is the smallest, consider a supermartingale $(\tilde{T}_n)$ that dominates $(\tilde{Z}_n)$. By induction, it dominates $(\tilde{U}_n)$.

Notice that $\tilde{T}_N \geq \tilde{U}_N$. If $\tilde{T}_n \geq \tilde{U}_n$, then

$$\tilde{T}_{n-1} \geq \mathbb{E}^*[\tilde{T}_n \mid \mathfrak{F}_{n-1}] \geq \mathbb{E}^*[\tilde{U}_n \mid \mathfrak{F}_{n-1}]$$

Thus,

$$\tilde{T}_{n-1} \geq \max\{\tilde{Z}_{n-1},\ \mathbb{E}^*[\tilde{U}_n \mid \mathfrak{F}_{n-1}]\} = \tilde{U}_{n-1}$$

□

7.2.3 Optimal Stopping

In this subsection, we apply the concept of stopping time to find the optimal exercise date of an American option. We start with a basic definition and a result.

Definition 7.2.12. Let (X_n) be an adapted sequence and ν a stopping time. The **sequence stopped at a stopping time** ν is defined as

$$X_n^{\nu}(\omega) = X_{\nu(\omega)\wedge n(\omega)}$$

Proposition 7.2.9. Let (X_n) be an adapted sequence and ν a stopping time.

1. The stopped sequence is adapted.

2. If (X_n) is a martingale (supermartingale), then (X_n^{ν}) is a martingale(supermartingale).

Proof. For $n \geq 1$,

$$X_{\eta\wedge n} = X_0 + \sum_{j=1}^{n} \varphi_j \Delta X_j$$

in which $\varphi_j = \chi_{\{j\leq\nu\}}$.

Since $\{j \leq \nu\}$ is the complement of $\{\nu < j\} = \{\nu \leq j-1\}$, (φ_n) is predictable. Thus, $(X_{\nu\wedge n})$ is $\mathfrak{F}_n$-adapted.

Given that $(X_{\nu\wedge n})$ is a martingale transform of (X_n), it is also a martingale.

For supermartingales or submartingales, the procedure is analogous. □

Now we turn to the concept of the Snell envelope, which is a fundamental concept to solve the problem at hand.

Definition 7.2.13 (Snell envelope). Consider (Z_n) and adapted sequence and

$$U_n = \begin{cases} U_N = Z_N \\ U_n = \max\{Z_n,\ \mathbb{E}[U_{n+1} \mid \mathfrak{F}_n]\}, \quad n = 0,\dots,N-1 \end{cases}$$

The sequence (U_n), which is the smallest supermartingale that dominates (Z_n), is called the **Snell envelope** of the sequence (Z_n).

The idea is that by stopping at an appropriate time, it is possible to obtain a martingale.

Proposition 7.2.10. Let

$$\nu_0 = \inf\{n \geq 0 : U_n = Z_n\}$$

The stopped sequence $(U_{n\wedge\nu_0})$ is a martingale.

Proof. 1. Write

$$U_n^{\nu_0} = U_{n\wedge\nu_0} = U_0 + \sum_{j=1}^{n} \varphi_j \Delta U_j$$

with $\varphi_j = \chi_{\{j\leq\nu_0\}}$.

2. Compute the differences

$$\begin{aligned} U_{n+1}^{\nu_0} - U_n^{\nu_0} &= \varphi_{n+1}(U_{n+1} - U_n) \\ &= \chi_{\{n+1\leq\nu_0\}}(U_{n+1} - U_n) \end{aligned}$$

3. Use the definition of U_n and the fact that $U_n > Z_n$ for $n+1 \le \nu_0$ to conclude that
$$U_n = \mathbb{E}[U_{n+1} \mid \mathfrak{F}_n]$$

4. Replace this in the previous equation,
$$U_{n+1}^{\nu_0} - U_n^{\nu_0} = \chi_{\{n+1\le\nu_0\}}(U_{n+1} - \mathbb{E}[U_{n+1} \mid \mathfrak{F}_n])$$

5. Take the conditional expectation
$$\mathbb{E}[U_{n+1}^{\nu_0} - U_n^{\nu_0} \mid \mathfrak{F}_n] = \chi_{\{n+1\le\nu_0\}}\mathbb{E}[(U_{n+1} - \mathbb{E}[U_{n+1} \mid \mathfrak{F}_n] \mid \mathfrak{F}_n] = 0$$

□

Let $\mathfrak{I}_{n,N}$ denote the set of stopping times taking values in $\{n, n+1, \ldots, N\}$. The next corollary relates the Snell envelope to the optimal stopping problem.

Corollary 7.2.11. The stopping time ν_0 satisfies
$$U_0 = \mathbb{E}[Z_{\nu_0} \mid \mathfrak{F}_0] = \sup_{\nu\in\mathfrak{I}_{0,N}} \mathbb{E}[Z_\nu \mid \mathfrak{F}_0]$$

Proof. 1. Use that U^{ν_0} is a martingale
$$U_0 = U_0^{\nu_0} = \mathbb{E}[U_N^{\nu_0} \mid \mathfrak{F}_0] = \mathbb{E}[U_{\nu_0} \mid \mathfrak{F}_0] = \mathbb{E}[Z_{\nu_0} \mid \mathfrak{F}_0]$$

2. Suppose that $\nu \in \mathfrak{I}_{0,N}$. Then the stopped sequence U^ν is a supermatingale. Thus,
$$\begin{aligned} U_0 &\ge \mathbb{E}[U_N^\nu \mid \mathfrak{F}_0] = \mathbb{E}[U_\nu \mid \mathfrak{F}_0] \\ &\ge \mathbb{E}[Z_\nu \mid \mathfrak{F}_0] \end{aligned}$$

□

Hence, if we interpret Z_n as the total winnings of a gambler after n games, then stopping at ν_0 maximizes the expected gain given $\mathfrak{F}_0$.

More generally,
$$U_n = \sup_{\nu\in\mathfrak{I}_{n,N}} \mathbb{E}[Z_\nu \mid \mathfrak{F}_n] = \mathbb{E}[Z_{\nu_n} \mid \mathfrak{F}_n] \tag{7.1}$$
where $\nu_n = \inf\{j \ge n : U_j = Z_j\}$.

Definition 7.2.14 (Optimal Stopping Time). A stopping time ν^* is called **optimal** for the sequence (Z_n) if
$$\mathbb{E}[Z_{\nu^*} \mid \mathfrak{F}_0] = \sup_{\nu\in\mathfrak{I}_{0,N}} \mathbb{E}[Z_\nu \mid \mathfrak{F}_0]$$

ν_0 is optimal. The following theorem shows that ν_0 is the smallest optimal time.

Theorem 7.2.12. A stopping time ν is optimal if, and only if, $Z_\nu = U_\nu$ and $(U_{\nu\wedge n})$ is a martingale.

Proof. ($\Leftarrow$) If U^ν is a martingale, $U_0 = \mathbb{E}[U_\nu \mid \mathfrak{F}_0]$.

By hypothesis, $U_0 = \mathbb{E}[Z_\nu \mid \mathfrak{F}_0]$.

By the Corollary 7.2.11, ν is optimal.

($\Rightarrow$) Supposing that ν is optimal, we have

$$U_0 = \mathbb{E}[Z_\nu \mid \mathfrak{F}_0] \leq \mathbb{E}[U_\nu \mid \mathfrak{F}_0]$$

Since U^ν is a supermartingale, $\mathbb{E}[U_\nu \mid \mathfrak{F}_0] \leq U_0$. Thus,

$$\mathbb{E}[U_\nu \mid \mathfrak{F}_0] = \mathbb{E}[Z_\nu \mid \mathfrak{F}_0]$$

Using that $U_\nu \geq Z_\nu$, it follows that $U_\nu = Z_\nu$.

Now use that $\mathbb{E}[U_\nu \mid \mathfrak{F}_0] = U_0$ and, by the supermartingale property of (U_n^ν), that

$$U_0 \geq \mathbb{E}[U_{\nu \wedge n} \mid \mathfrak{F}_0] \geq \mathbb{E}[U_\nu \mid \mathfrak{F}_0]$$

to obtain

$$\mathbb{E}[U_{\nu \wedge n} \mid \mathfrak{F}_0] = \mathbb{E}[U_\nu \mid \mathfrak{F}_0] = \mathbb{E}[\mathbb{E}[U_\nu \mid \mathfrak{F}_n] \mid \mathfrak{F}_0]$$

Since $U_{\nu \wedge n} \geq \mathbb{E}[U_\nu \mid \mathfrak{F}_n]$, it follows that $U_{\nu \wedge n} = \mathbb{E}[U_\nu \mid \mathfrak{F}_n]$, and thus (U_n^ν) is a martingale. □

Now that we know what is an optimal stopping time, we'll use a decomposition of supermartingales in viable complete markets to associate any supermartingale with a trading strategy in which consumption is allowed.

Proposition 7.2.13 (Doob Decomposition)**.** Every supermartingale (U_n) has the unique following decomposition:

$$U_n = M_n - A_n$$

where (M_n) is a martingale and (A_n) is a non-decreasing, predictable process and null at zero.

Proof. 1. For $n = 0$, $M_0 = U_0$ and $A_0 = 0$.

2. Write the difference

$$U_{n+1} - U_n = M_{n+1} - M_n - (A_{n+1} - A_n)$$

3. Conditioning w.r.t. $\mathfrak{F}_n$:

$$\mathbb{E}[U_{n+1} \mid \mathfrak{F}_n] - U_n = -(A_{n+1} - A_n)$$

and

$$M_{n+1} - M_n = U_{n+1} - \mathbb{E}[U_{n+1} \mid \mathfrak{F}_n]$$

4. These equations entirely determine (M_n) and (A_n), (M_n) is a martingale and, since (U_n) is a supermartingale, (A_n) is a predictable and non-decreassing process. □

The next result shows how to characterize the largest optimal stopping time for (Z_n) using the process (A_n) of the Doob decomposition.

Proposition 7.2.14. The largest optimal stopping time for (Z_n) is given by

$$\nu_{\max} = \begin{cases} N, & A_n = 0 \\ \inf\{n,\ A_{n+1} \neq 0\} & A_n \neq 0 \end{cases}$$

Proof. 1. $\nu_{\max}$ is a stopping time: follows from the fact that (A_n) is predictable.

2. Optimality of $\nu_{\max}$:

- Using that $U_n = M_n - A_n$ and $A_j = 0$ for $j \leq \nu_{\max}$, it follows that $U^{\nu_{\max}} = M^{\nu}_{\max}$ and, thus, $U^{\nu_{\max}}$ is a martingale.
- Thus, it is sufficient to show that $U_{\nu_{\max}} = Z_{\nu_{\max}}$.
- To do that, notice

$$\begin{aligned} U_{\nu_{\max}} &= \sum_{j=0}^{N-1} \chi_{\nu_{\max}=j} U_j + \chi_{\nu_{\max}=N} U_n \\ &= \sum_{j=0}^{N-1} \chi_{\nu_{\max}=j} \max\{Z_j,\ \mathbb{E}[U_{j+1} \mid \mathfrak{F}_j]\} + \chi_{\nu_{\max}=N} Z_n \end{aligned}$$

- We have $\mathbb{E}[U_{j+1} \mid \mathfrak{F}_j] = M_j - A_{j+1}$, and that $A_j = 0$, $A_{j+1} > 0$ on $\{\nu_{\max} = j\}$.
- Thus, $U_j = M_j$ and $\mathbb{E}[U_{j+1} \mid \mathfrak{F}_j] < U_j$.
- Then, $U_j = \max\{Z_j,\ \mathbb{E}[U_{j+1} \mid \mathfrak{F}_j]\} = Z_j$ and $U_{\nu_{\max}} = Z_{\nu_{\max}}$.

3. It is the greatest optimal stopping time.

- Let ν be a stopping time such that $\nu \geq \nu_{\max}$ with positive probability.
- Then
$$\mathbb{E}[U_\nu] = \mathbb{E}[M_\nu] - \mathbb{E}[A_\nu] = \mathbb{E}[U_0] - \mathbb{E}[A_\nu] < \mathbb{E}[U_0]$$
- Which implies that U^ν is not a martingale.

□

Before applying these concepts to the pricing of American options, it is necessary to know how to compute Snell envelopes in a Markovian setting.

Recall that a sequence (X_n) of random variables taking their values in a finite set E is called a **Markov chain** if, for any integer $n \geq 1$ and $x_0, x_1, \ldots, x_{n-1}, x, y \in E$, we have that

$$\mathbb{P}[X_{n+1} = y \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = x] = \mathbb{P}[X_{n+1} = y \mid X_n = x]$$

The chain is **homogeneous** if the value $P(x,y) = \mathbb{P}[X_{n+1} = y \mid X_n = x]$ does not depend on n.

And the matrix $P = (P(x,y))_{(x,y)\in E\times E}$ is called the **transition matrix** of the chain. It has non-negative entries and satisfies $\sum_{y\in E} P(x,y) = 1$ for all $x \in E$ and is also called **stochastic matrix**.

Alternatively, a sequence (X_n) of random variables taking values in E is a **homogeneous**

Markov chain with respect to the filtration $(\mathfrak{F}_n)$ and with transition matrix P if (X_n) is adapted and for any real valued function f on E we have

$$\mathbb{E}[f(X_{n+1}) \mid \mathfrak{F}_n] = Pf(X_n)$$

where Pf is the function that maps $x \in E$ to

$$Pf(x) = \sum_{y \in E} P(x,y)f(y)$$

Note that the previous definition is a Markov chain with respect to its natural filtration.

The next result is used to compute the prices of American options in discrete models.

Proposition 7.2.15. Let (Z_n) be the adapted sequence defined by $Z_n = \psi(n, X_n)$ in which (X_n) is a homogeneous Markov chain with transition matrix P and taking values in E, and $\psi : \mathbb{N} \times E \longrightarrow \mathbb{R}$.

The Snell envelope (U_n) of (Z_n) is given by $U_n = u(n, X_n)$ in which

$$u(n,x) = \begin{cases} \psi(n,x), & n = N \\ \max\{\psi(n,x),\ Pu(n+1,x)\}, & n \leq N-1 \end{cases}$$

Proof. A consequence of the definition of Markov chain and Snell envelope. □

Now we are ready to apply these concepts to American options.

Example 7.2.2. Suppose that we are in a viable complete market. Let $\mathbb{P}^*$ be the unique probability under which the discounted asset prices are martingales.

1. Pricing the option

The sequence $(\tilde{U}_n)$ of discounted prices of the option is the Snell envelope, under $\mathbb{P}^*$ of $(\tilde{Z}_n)$.

By the equation (7.1),

$$\tilde{U}_n = \sup_{\nu \in \mathfrak{I}_{n,N}} \mathbb{E}^*[\tilde{Z}_\nu \mid \mathfrak{F}_n]$$

and thus

$$U_n = S_n^0 \sup_{\nu \in \mathfrak{I}_{n,N}} \mathbb{E}^*\left[\frac{Z_\nu}{S_\nu^0} \,\middle|\, \mathfrak{F}_n\right]$$

Using the Doob Decomposition,

$$\tilde{U}_n = \tilde{M}_n - \tilde{A}_n$$

in which $\tilde{M}_n$ is a $\mathbb{P}^*$-martingale and $(\tilde{A}_n)$ is an increasing predictable process null at zero.

Since the market is complete, there exists a self-financing strategy φ such that

$$V_N(\varphi) = S_N^0 \tilde{M}_N \implies \tilde{V}_N(\varphi) = \tilde{M}_N$$

And given that $(\tilde{V}_n(\varphi))$ is a $\mathbb{P}^*$-martingale,

$$\tilde{V}_n(\varphi) = \mathbb{E}^*[\tilde{V}_N(\varphi) \mid \mathfrak{F}_n] = \mathbb{E}^*[\tilde{M}_N \mid \mathfrak{F}_n] = \tilde{M}_n$$

Hence,

$$\tilde{U}_n = \tilde{V}_n(\varphi) - \tilde{A}_n \implies U_n = V_n(\varphi) - A_n$$

where $A_n = S_n^0 \tilde{A}_n$.

Notice that a perfect hedging is available: receive the premium $U_0 = V_0(\varphi)$, generate wealth equal to $V_n(\varphi)$ at n which is bigger than $U_n \geq Z_n$.

2. Optimal exercise date

If $U_n > Z_n$, we would be trading an asset worth U_n for an amount (Z_n). Therefore, it must be at a stopping time τ such that $U_\tau = Z_\tau$.

Also, it should not be after $\nu_{\max} = \inf\{j,\ A_{j+1} \neq 0\}$ (see Proposition 7.2.14). Selling gives $U_{\nu_{\max}} = V_{\nu_{\max}}(\varphi)$. (What happens after that?)

Thus, we set $\tau \leq \nu_{\max}$ and then U^τ is a martingale. The optimal dates of exercise are optimal stopping times for $(\tilde{Z}_n)$ under $\mathbb{P}^*$.

If the writer hedges himself with the strategy φ defined earlier and the buyer exercises at a non-optimal time τ, then either $U_\tau > Z_\tau$ or $A_\tau > 0$. In both cases, the writer makes a profit

$$V_\tau(\varphi) - Z_\tau = U_\tau + A_\tau - Z_\tau > 0$$

The following proposition relates American and European options.

Proposition 7.2.16. Let C_N be the value, at time n, of an American option described by an adapted sequence (Z_n), and (c_n) be the value, at time n, of the Europen option defined by the $\mathfrak{F}_N$-measurable random variable $h = Z_n$.

Then,

1. $C_n \geq c_n$;
2. If $c_n \geq Z_n$ for any n, then $c_n = C_n$ for all $n \in \{0, 1, \dots, N\}$.

Proof. $(\tilde{C}_n)$ is a supermartingale under $\mathbb{P}^*$:

$$(\tilde{C}_n) \geq \mathbb{E}^*[\tilde{C}_N \mid \mathfrak{F}_n] = \mathbb{E}^*[\tilde{c}_N \mid \mathfrak{F}_n] = \tilde{c}_n \implies C_n \geq c_n$$

If $c_n \geq Z_n$ for any n, then $(\tilde{c}_n)$ is a martingale under $\mathbb{P}^*$ is an upper bound for the $(\tilde{Z}_n)$ and thus $\tilde{C}_n \leq \tilde{c}_n$ for all $n \in \{0, 1, \dots, N\}$. □

Remark that if the conditions didn't hold, there would be arbitrage opportunities. The property doesn't hold for puts or calls on currencies or dividend-paying stocks.

7.3 Continuous Time: The Black-Scholes model

In this section, we present the Black-Scholes model. The intuition gained from discrete time will prove itself useful.

7.3.1 The behavior of prices

Suppose that we have one risk asset S_t and a riskless asset S_t^0 such that

$$dS_t^0 = rS_t^0 dt$$

where $r \geq 0$ is the instantaneous interest rate.

Setting $S_0^0 = 1$, we have $S_t^0 = e^{rt}$. We also assume that the following stochastic differential equation determines the behavior of the stock price

$$dS_t = \mu S_t dt + \sigma S_t dB_t$$

From the Example 5.1.5, we have that its solution is

$$S_t = S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t\right) \tag{7.2}$$

Notice that the law of S_t is lognormal and that the hypotheses for this model are the same as the Brownian motion.

7.3.2 Self-financing Strategies

Before pricing options, we need to fix our terminology for the continuous case.

Definition 7.3.1 (Strategy). We define a **strategy** as a process $\varphi = (\varphi_t) = (H_t^0, H_t)$ adapted to the natural filtration $(\mathfrak{F}_t)$ of Brownian motion, where H_t^0 and H_t are the quantities of the riskless and risky asset at t, respectively.

Definition 7.3.2 (Value of the portfolio). The **value of the portfolio** at t is given by:

$$V_t(\varphi) = H_t^0 S_t^0 + H_t S_t$$

To characterize self-financing strategies, we have the following identity

$$dV_t(\varphi) = H_t^0 dS_t^0 + H_t dS_t$$

In order to make this equality meaningful, we set the following conditions

$$\int_0^T |H_t^0|\ dt < +\infty \text{ a.s.} \quad \text{and} \quad \int_0^T H_t^2 < +\infty \text{ a.s.}$$

Thus,

$$\int_0^T H_t^0\ dS_t^0 = \int_0^T H_t^0 r e^{rt}\ dt$$

and

$$\int_0^T H_t^0\ dS_t^0 = \int_0^T (H_t S_t \mu)\ dt + \int_0^T \sigma H_t S_t\ dB_t$$

are well-defined.

Definition 7.3.3 (Self-Financing Strategy). A **self-financing strategy** is a pair of adapted processes (H_t^0) and (H_t) satisfying

$$\int_0^T |H_t^0| \, dt + \int_0^T H_t^2 \, dt < +\infty \text{ a.s.}$$

and

$$H_t^0 S_t^0 + H_t S_t = H_0^0 S_0^0 + H_0 S_0 + \int_0^t H_u^0 \, dS_u^0 + \int_0^t H_u \, dS_u \text{ a.s. }, \quad \forall\, t \in [0,T]$$

We call $\tilde{S}_t = e^{-rt} S_t$ the **discounted price** of risky asset.

Proposition 7.3.1. Suppose that $\varphi = (H_t^0, H_t)$ is an adapted process, valued in $\mathbb{R}^2$, satisfying

$$\int_0^T |H_t^0| \, dt + \int_0^T H_t^2 \, dt < +\infty \text{ a.s.}$$

and let $V_t(\varphi) = H_t^0 S_t^0 + H_t S_t$ and $\tilde{V}_t(\varphi) = e^{-rt} V_t(\varphi)$.

Then φ defines a self-financing strategy if, and only if,

$$\tilde{V}_t(\varphi) = V_0(\varphi) + \int_0^t H_u \, d\tilde{S}_u \text{ a.s.}, \quad \forall\, t \in [0,T]$$

Proof. If φ is self-financing, differentiating the product of (e^{rt}) and $(V_t(\varphi))$ yields

$$d\tilde{V}_t(\varphi) = -r\tilde{V}_t(\varphi)dt + e^{-rt} dV_t(\varphi)$$

Rewriting, we have that $d\tilde{V}_t(\varphi) = H_t d\tilde{S}_t$. The converse follows similarly. □

Notice that no condition of predictability is needed.

7.3.3 Change of Measure

We need the following definition and result to define equivalent probabilities in the continuous context.

Definition 7.3.4 (Absolutely Continuous Measure). A probability $\mathbb{Q}$ is **absolutely continuous** with respect to $\mathbb{P}$ if

$$\mathbb{P}(A) = 0 \implies \mathbb{Q}(A) = 0, \quad \forall\, A \in \mathfrak{F}$$

Theorem 7.3.2. A probability measure $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$ if, and only if, there exists a non-negative random variable Z such that

$$\mathbb{Q}(A) = \int_A Z \, d\mathbb{P}, \quad \forall\, A \in \mathfrak{F}$$

Z is called the **density** of $\mathbb{Q}$ with respect to $\mathbb{P}$ and is denoted by $\frac{d\mathbb{Q}}{d\mathbb{P}}$.

Definition 7.3.5 (Equivalent Probabilities)**.** The probabilities $\mathbb{P}$ and $\mathbb{Q}$ are **equivalent** if each of them is absolutely continuous with respect to the other.

Notice that $\mathbb{P}$ and $\mathbb{Q}$ are equivalent if, and only if, $\mathbb{P}[Z > 0] = 1$.

Theorem 7.3.3 (Girsanov)**.** Let (θ_t) be an adapted process satisfying

$$\int_0^T \theta_s^2 \, ds < \infty \text{ a.s.}$$

and such that the process (L_t) given by

$$L_t = \exp\left(-\int_0^t \theta_s \, dB_s - \frac{1}{2} \int_0^t \theta_s^2 \, ds \right)$$

is a martingale.

Then, under the probability $\mathbb{P}^L$ with density L_T with respect to $\mathbb{P}$, the process (W_t) defined by

$$W_t = B_t + \int_0^t \theta_s \, ds$$

is an $(\mathfrak{F}_t)$-Brownian motion.

Proof. Refer to [S⁺04, Theorem 5.2.3]. □

Remark (Novikov's criterion)**.** If

$$\mathbb{E}\left[\exp\left(\frac{1}{2} \int_0^T \theta_t^2 \, dt \right) \right] < \infty$$

then (L_t) is a martingale.

7.3.4 Pricing and Hedging Options in the Black-Scholes Model

To price and hedge options, we first show that there exists a probability equivalent to $\mathbb{P}$ under which the discounted prices are martingales.

Using that $\tilde{S}_t = e^{-rt} S_t$ and applying Itô's formula to

$$dS_t = \mu S_t dt + \sigma S_t dB_t$$

we obtain

$$d\tilde{S}_t = -re^{-rt} S_t dt + e^{-rt} dS_t = (\mu - r)\tilde{S}_t dt + \sigma \tilde{S}_t dB_t$$

Setting $W_t = B_t + (\mu - r)t/\sigma$,

$$d\tilde{S}_t = \tilde{S}_t \sigma dW_t \tag{7.3}$$

By the Girsanov's theorem applied to $\theta_t = (\mu - r)/\sigma$, there exists a probability $\mathbb{P}^*$ equivalent to $\mathbb{P}$ under which (W_t) is a standard Brownian motion.

Using that the stochastic integral is invariant by change of equivalent probability, it follows that, under $\mathbb{P}^*$, $(\tilde{S}_t)$ is a martingale and

$$\tilde{S}_t = \tilde{S}_0 \exp\left(\sigma W_t - \frac{\sigma^2 t}{2}\right)$$

With that, we may price the option following a procedure analogous to the discrete case. We define a European option by a non-negative, $\mathfrak{F}_T$-measurable random variable $h = f(S_T)$, with $f(x) = \max\{(x-K), 0\}$ for a call, or $f(x) = \max\{(K-x), 0\}$ for a put.

Definition 7.3.6 (Admissible Strategy)**.** A strategy $\varphi = (H_t^0, H_t)$ is **admissible** if

1. It is self-financing;
2. The discounted value
$$\tilde{V}_t(\varphi) = H_t^0 + H_t \tilde{S}_t$$
of the portfolio is non-negative for all $0 \le t \le T$;
3. $\sup_{t\in[0,T]} \tilde{V}_t \in L^2(\mathbb{P}^*)$.

An option is **replicable** if its payoff at maturity is equal to the final value of an admissible strategy.

The next result shows that, under some conditions, any option is replicable and gives the value of any replicating portfolio.

Theorem 7.3.4. Any option defined by a non-negative, $\mathfrak{F}_T$-measurable random variable h in $L^2(\mathbb{P}^*)$ is replicable (in the Black-Scholes model).

In fact, the value at time t of any replicating portfolio is given by

$$V_t = \mathbb{E}^*\left[e^{-r(T-t)} h \,\middle|\, \mathfrak{F}_t\right]$$

Hence, option value at t can be naturally defined by $\mathbb{E}^*\left[e^{-r(T-t)} h \,\middle|\, \mathfrak{F}_t\right]$.

Proof. Suppose that we already found an admissible strategy (H^0, H) replicating the option. The value of the portfolio at time t is given by

$$V_t = H_t^0 S_t^0 + H_t S^t$$

and the discounted value is

$$\tilde{V}_t = H_t^0 + H_t \tilde{S}_t$$

Since the strategy is self-financing, by Proposition 7.3.1,

$$\tilde{V}_t = V_0 + \int_0^t H_u \, d\tilde{S}_u$$

and, by the equation (7.3),

$$\tilde{V}_t = V_0 + \int_0^t H_u \sigma \tilde{S}_u \, dW_u$$

Using that $\sup_{t\in[0,T]} \tilde{V}_t \in L^2(\mathbb{P}^*)$, the process $(\tilde{V}_t)$ is a stochastic integral with respect to (W_t) it follows that $(\tilde{V}_t)$ is a square-integrable martingale under $\mathbb{P}^*$.

Hence,

$$\tilde{V}_t = \mathbb{E}^*[\tilde{V}_T \mid \mathfrak{F}_t] \implies V_t = \mathbb{E}^*[e^{-r(T-t)}h \mid \mathfrak{F}_t]$$

proves the second part of our theorem.

Now, we need to prove that the option is replicable. That means to find $(H_t)^0$ and (H_t) defining an admissible strategy satisfying

$$H_t^0 S_t^0 + H_t S_t = \mathbb{E}^*[e^{-r(T-t)}h \mid \mathfrak{F}_t]$$

The process $M_t = \mathbb{E}^*[e^{-r(T-t)}h \mid \mathfrak{F}_t]$ is a square-integrable $\mathbb{P}^*$-martingale. Thus, from the Martingale Representation Theorem, there exists an adapted process (K_t) such that

$$\mathbb{E}^*\left[\int_0^T K_s^2 \, ds\right] < \infty, \quad M_t = M_0 + \int_0^t K_s \, dW_s \text{ a.s. } \forall\, t \in [0,T]$$

Applying the Proposition 7.3.1 and the equation (7.3) again, the strategy $\varphi = (H^0, H)$ with $H_t = K_t/(\sigma\tilde{S}_t)$ and $H_t^0 = M_t - H_t\tilde{S}_t$ is a self-finacing strategy with value at t given by

$$V_t(\varphi) = e^{rt}M_t = \mathbb{E}^*[e^{-r(T-t)}h \mid \mathfrak{F}_t]$$

Thus, φ is the admissible strategy replicating h. □

The next result expresses the option value only as a function of t and S_t. Put another way, it is an explicit solution of the Black-Scholes formula. Before that, we need the following.

Lemma 7.3.5. Let X and Y be two random variables with values in $(E, \mathfrak{E})$ and $(F, \mathfrak{F})$ respectively. Suppose that X is $\mathfrak{B}$-measurable and that Y is independent of $\mathfrak{B}$. Then, for any non-negative (or bounded) Borel function Ψ on $(E \times F, \mathfrak{E} \otimes \mathfrak{F})$, the function ψ defined by

$$\psi(x) = \mathbb{E}[\Psi(x, Y)], \quad x \in E$$

is a Borel function on $(E, \mathfrak{E})$.

And we have

$$\mathbb{E}[\Psi(X, Y) \mid \mathfrak{B}] = \psi(X) \text{ a.s.}$$

Proof. Refer to Proposition A.2.5. [LL11, p. 240]. □

Proposition 7.3.6. The option value V_t can be expressed as $V_t = F(t, S_t)$ in which

$$F(t, x) = x\Phi(d_1) - Ke^{-r\theta}\Phi(d_2)$$

for a call and

$$F(t, x) = Ke^{-r\theta}\Phi(-d_2) - x\Phi(-d_1)$$

for a put, where $\Phi(x), d_1$ and d_2 are given by

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-u^2/2}\, du$$

and

$$d_1 = \frac{\ln(x/K) + (r + \sigma^2/2)\theta}{\sigma\sqrt{\theta}}, \quad d_2 = \frac{\ln(x/K) + (r - \sigma^2/2)\theta}{\sigma\sqrt{\theta}}$$

Proof. Using (7.2),

$$\begin{aligned} V_t(\varphi) &= \mathbb{E}^*[e^{-r(T-t)} f(S_T) \mid \mathfrak{F}_t] \\ &= \mathbb{E}^*\left[e^{-r(T-t)} f\left(S_t e^{r(T-t)} e^{\sigma(W_T - W_t)} e^{-(\sigma^2/2)(T-t)} \right) \,\middle|\, \mathfrak{F}_t \right] \end{aligned}$$

S_t is $\mathfrak{F}_t$-measurable and, under $\mathbb{P}^*$, $W_T - W_t$ is independent of $\mathfrak{F}_t$, by the previous lemma we have that $V_t = F(t, S_t)$, where

$$F(t,x) = \mathbb{E}^*\left[e^{-r(T-t)} f\left(x e^{r(T-t)} e^{\sigma(W_T - W_t)} e^{-(\sigma^2/2)(T-t)} \right) \right]$$

Since, under $\mathbb{P}^*$, $(W_T - W_t) \sim N(0,\ T-t)$,

$$F(t,x) = e^{-r(T-t)} \int_{-\infty}^{+\infty} f\left(x e^{(r-\sigma^2/2)(T-t) + \sigma y \sqrt{T-t}} \right) \frac{e^{-y^2/2}}{\sqrt{2\pi}}\, dy$$

For a call option, $f(x) = \max\{(x-K), 0\}$. Hence,

$$\begin{aligned} F(t,x) &= \mathbb{E}^*\left[e^{-r(T-t)} \max\left(x e^{(r-\sigma^2/2)(T-t) + \sigma y (W_T - W_t)} - K,\ 0 \right) \right] \\ &= \mathbb{E}\left[\max\left(x e^{\sigma\sqrt{\theta} g - \sigma^2\theta/2} - K e^{-r\theta},\ 0 \right) \right] \end{aligned}$$

where g is a standard normal random variable and $\theta = T - t$.

Using d_1 and d_2:

$$\begin{aligned} F(t,x) &= \mathbb{E}\left[\left(x e^{\sigma\sqrt{\theta} g - \sigma^2\theta/2} - K e^{-r\theta} \right) \chi_{\{g + d_2 \geq 0\}} \right] \\ &= \int_{-d_2}^{+\infty} \left(x e^{\sigma\sqrt{\theta} y - \sigma^2\theta/2} - K e^{-r\theta} \right) \frac{e^{-y^2/2}}{\sqrt{2\pi}}\, dy \\ &= \int_{-\infty}^{d_2} \left(x e^{\sigma\sqrt{\theta} y - \sigma^2\theta/2} - K e^{-r\theta} \right) \frac{e^{-y^2/2}}{\sqrt{2\pi}}\, dy \end{aligned}$$

Writing as the difference of two integrals and using the change of variables $z = y + \sigma\sqrt{\theta}$ in the first one, we obtain the result. □

With these results, how can we build a replicating portfolio to hedge an option? It must have, at time t, a discounted value $\tilde{V}_t = e^{-rt} F(t, S_t)$.

It can be proved that, under some hypothesis on f, the function F is of class $\mathfrak{C}^\infty$ on $[0,T) \times \mathbb{R}$. Therefore, we can set $\tilde{F}(t,x) = e^{-rt} F(t, x e^{rt})$ and, by applying Itô's formula and using that $\tilde{F}(t, \tilde{S}_t)$

is a $\mathbb{P}^*$-martingale, we can deduce that

$$\tilde{F}(t,\tilde{S}_t) = \tilde{F}(0,\tilde{S}_0) + \int_0^t \frac{\partial \tilde{F}}{\partial x}(u,\tilde{S}_u)\, d\tilde{S}_u$$

Thus, the hedging process H_t is given by

$$H_t = \frac{\partial \tilde{F}}{\partial x}(t,\tilde{S}_t) = \frac{\partial F}{\partial x}(t,S_t)$$

Notice that, in the case of a call:

$$\frac{\partial F}{\partial x}(t,S_t) = \Phi(d_1)$$

and for a put:

$$\frac{\partial F}{\partial x}(t,S_t) = -\Phi(-d_1)$$

Definition 7.3.7 ('Greeks')**.** The quantity $\frac{\partial F}{\partial x}(t,S_t)$ above is called the **delta** of the option. It measures the sensitivity of the portfolio with respect to the variations of the asset price at time t.

The **gamma** is the second derivative, **theta** is the time derivative, and **vega** is the derivative with respect to the volatility.

7.3.5 American Options

Definition 7.3.8 (Trading strategy with consumption)**.** A **trading strategy with consumption** is an adapted process $\varphi = (H_t^0, H_t)$ with values in $\mathbb{R}^2$ satisfying

1. $\int_0^T |H_t^0|\, dt + \int_0^T H_t^2\, dt < +\infty$ a.s.
2. $H_t^0 S_t^0 + H_t S_t = H_0^0 S_0^0 + H_0 S_0 + \int_0^t H_u^0\, dS_u^0 + \int_0^t H_u\, dS_u - C_t$ for all $t \in [0,T]$, where (C_t) is an adapted, continuous, non-decreasing process null at $t = 0$, corresponding to the cumulative consumption up to time t.

An American option is an adapted non-negative process (h_t). We'll consider only payoffs of the form $h_t = \psi(S_t)$, where $\psi : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ is a continuous function satisfying $\psi(x) \leq A + Bx$ for some non-negative constants A and B.

Denote by Φ^ψ the set of all trading strategies with consumption hedging the American option of the form $h_t = \psi(S_t)$.

Notice that φ hedges h_t if, setting $V_t(\varphi) = H_t^0 S_t^0 + H_t S_t$, we have $V_t(\varphi) \geq \psi(S_t)$ for all $t \in [0,T]$ almost surely.

Theorem 7.3.7. Let $u : [0,T] \times \mathbb{R}^+ \longrightarrow \mathbb{R}$ be given by

$$u(t,x) = \sup_{t \in \mathfrak{I}_{t,T}} \mathbb{E}^*[e^{-r(\tau-t)}\psi(x\exp((r-(\sigma^2/2))(\tau-t)+\sigma(W_\tau - W_t)))]$$

where $\mathfrak{I}_{t,T}$ is the set of all stopping times with values in $[t,T]$.

Then there exists a strategy $\overline{\varphi} \in \Phi^{\psi}$ such that $V_t(\overline{\varphi}) = u(t, S_t)$ for all $t \in [0, T]$.

Also, for any strategy $\varphi \in \Phi^{\psi}$, $V_t(\varphi) \geq u(t, S_t)$ for all $t \in [0, T]$.

Proof. 1. Show that $(e^{-rt}u(t, S_t))$ is the Snell envelope of $(e^{-rt}\psi(S_t))$.

2. Since the discounted value of a trading strategy with consumption is a $\mathbb{P}^*$-martingale, $V_t(\varphi) \geq u(t, S_t)$ for all $\varphi \in \Phi^{\psi}$.

3. Use the decomposition of martingales and the representation theorem to show the existence of $\overline{\varphi}$ such that $V_t(\overline{\varphi}) = u(t, S_t)$.

□

This result inspires us to interpret $u(t, S_t)$ as the price for the American option at time t.

Remark. The minimal wealth that hedges all possible exercises is given by

$$u(0, S_0) = \sup_{\tau \in \mathfrak{I}_{0,T}} \mathbb{E}^*[e^{-r\tau}\psi(S_\tau)]$$

Notice that the American call price, on a non-dividend paying stock, equals the European call price. However, this is not the case for puts, which need numerical methods.

7.3.6 Perpetual Puts and Critical Price

We study qualitative properties of $u(t, x)$.

Taking T to infinity,

$$\begin{aligned} u(0, x) &= \sup_{\tau \in \mathfrak{I}_{0,T}} \mathbb{E}^*[\max\{e^{-r\tau} - x\exp(\sigma W_\tau - \sigma^2\tau/2),\ 0\}] \\ &\leq \sup_{\tau \in \mathfrak{I}_{0,T}} \mathbb{E}[\max\{e^{-r\tau} - x\exp(\sigma B_\tau - \sigma^2\tau/2),\ 0\}\chi_{\tau<\infty}] =: u^{\infty}(x) \end{aligned}$$

where $u^{\infty}(x)$ is called the **value of perpetual put**.

Proposition 7.3.8.

$$u^{\infty}(x) = \begin{cases} K - x, & x \leq x^* \\ (K - x^*)\left(\frac{x}{x^*}\right)^{-\gamma} & x > x^* \end{cases} \tag{7.4}$$

with $x^* = K\gamma/(1+\gamma)$ and $\gamma = 2r/\sigma^2$.

Proof. 1. Notice that u^{∞} is convex, decreasing on $[0, \infty)$ and

$$u^{\infty}(x) \geq \max\{(K - x), 0\}$$

2. For all $T > 0$, $u^{\infty}(x) \geq \mathbb{E}[\max\{e^{-rT} - x\exp(\sigma B_T - \sigma^2 T/2),\ 0\}]$ implies that $u^{\infty}(x) > 0$ for all $x \geq 0$.

3. Define $x^* = \sup\{x \geq 0 : u^{\infty}(x) = K - x\}$.

4. Conclude that

$$\forall\, x \leq x^*, \quad u^{\infty}(x) = K - x \quad \text{and} \quad \forall\, x \geq x^*, \quad u^{\infty}(x) > \max\{(K - x), 0\} \tag{7.5}$$

5. Define the stopping time

$$\tau_x = \inf\{t \geq 0 : e^{rt} u^{\infty}(X_t^x) = e^{-rt} \max\{K - X_t^x, 0\}\}$$

6. Using the continuous time version of Snell envelope,

$$u^{\infty}(x) = \mathbb{E}[\max\{e^{-r\tau_x} - x\exp(\sigma B_{\tau_x} - \sigma^2 \tau_x/2), 0\}\chi_{\tau_x < \infty}]$$

7. Notice that τ_x is an optimal stopping time.

8. From (7.5), it follows that

$$\tau_x = \inf\{t \geq 0 : X_t^x \leq x^*\} = \inf\{t \geq 0 : (r - \sigma^2/w)t + \sigma B_t \leq \log(x^*/x)\}$$

9. Define a new stopping time $\tau_{x,z} = \inf\{t \geq 0 : X_t^x \leq z\}$ and notice that the optimal stopping time $\tau_x = \tau_{x,x^*}$.

10. Now define

$$\varphi(z) = \mathbb{E}\left[e^{-r\tau_{x,z}} \chi_{\tau_{x,z} < \infty} \max\{K - X_{\tau_{x,z}}^x, 0\}\right]$$

11. Since τ_{x,x^*} is optimal, $\varphi(z)$ attains its maximum at $z = x^*$.

12. Compute φ explicitely and maximize it to find x^* and $u^{\infty}(x) = \varphi(x^*)$.

□

Remark that for an American put with finite maturity T, we may apply the same argument above. For $t \in [0, t)$, there exists $s(t) \in [0, K]$ such that

$$\forall\, x \leq s(t), \quad u(t, x) = K - x$$

and

$$\forall\, x > s(t), \quad u(t, x) > \max\{K - x, 0\}$$

Using (7.4), we have that $s(t) \geq x^*$ for all $t \in [0, t)$.

The number $s(t)$ is interpreted as the **critical price** at time t. If the price of the underlying asset at t is less than or equal to $s(t)$, the buyer of the option should exercise the option.

7.3.7 Implied and local volatilities

Notice that the Black-Scholes model only depends on the volatility σ. The natural question, then, is how to evaluate it? We start by listing the two most used methods.

1. **Historical method**:
    - $\sigma^2 T$ variance of $\log(S_T)$;
    - $\log(S_T/S_0), \ldots, \log(S_{NT}, S_{(n-1)T})$ are i.i.d.
    - σ is estimated by past observations.
2. **Implied volatility**:
    - The price of the options is an increasing function of σ.

- By inversion of Black-Scholes, associate an 'implied' volatility.
- Used for hedging more than pricing.

A more consistent approach, however, is to replace the constant σ by a stochastic process (σ_t). With this, our model becomes

$$dS_t = \mu S_t dt + \sigma_t S_t dB_t$$

If σ_t is adapted and bounded, we can extend the approach developed previously.

In the **local volatility** model, $\sigma_t = \sigma(t, S_t)$ is a deterministic function of time and current price. Using market prices of calls, we can build a local volatility model that provides the same prices as the market. If $C(T,K)$ is the market price, the consistent local volatility is given by the **Dupire's formula**:

$$\frac{\partial C}{\partial T}(T,K) = \frac{\sigma^2(T,K)K^2}{2}\frac{\partial^2 C}{\partial K^2}(T,K) - rK\frac{\partial C}{\partial T}(T,K)$$

However, this is not easy to implement and has unstability issues. Thus, it is prefered, by practicioners, to use stochastic volatility models with jumps.

In a **stochastic volatility** model,

- (σ_t) satisfies an stochastic differential equation following another Brownian motion, which may not be correlated with (B_t).
- No longer adapted to the natural filtration of Brownian motion.
- Are incomplete, replication may not be possible.

7.3.8 Call/Put Symmetry

Notice that our previous model assumes that the underlying stock does not distribute dividends. In this subsection, we model options with dividends by assuming that the holder receives $\delta S_t dt$ in an infinitesimal time interval.

In this context, the self-financing condition yields

$$dV_t = H_t^0 dS_t^0 + H_t dS_t + \delta H_t S_t dt$$

and

$$d\tilde{V}_t = H_t d\tilde{S}_t + \delta H_t \tilde{S}_t dt = \sigma H_t \tilde{S}_t dW_t^\delta$$

where

$$W_t^\delta = B_t + (\mu - \delta - r)t/\sigma$$

and $\mathbb{P}^\delta$ is the measure under which (W_t^δ) is a standard Brownian motion.

Remark that, under the probability $\mathbb{P}^\delta$, $(e^{(\delta-r)t}S_t)$ is a martingale.

We denote by $C_e(t,x,K,r,\delta)$ and $P_e(t,x,K,r,\delta)$ the price, at time t of an European call (respectively put), with current stock price x, strike price K, interest rate r and dividend yield δ. Analogously, we define C_a and P_a for American call and put.

Then,

$$C_e(t,x,K,r,\delta) = \mathbb{E}^\delta\left[e^{-r(T-t)}\max\left\{xe^{(r-\delta-(\sigma^2/2)(T-t))+\sigma(W_T^\delta - W_t^\delta)} - K,\ 0\right\}\right]$$

Proposition 7.3.9 (Call/Put Symmetry)**.**

$$\mathrm{C}_e(t,x,\mathrm{K},r,\delta) = \mathrm{P}_e(t,\mathrm{K},x,r,\delta) \quad \text{and} \quad \mathrm{C}_a(t,x,\mathrm{K},r,\delta) = \mathrm{P}_a(t,\mathrm{K},x,r,\delta)$$

Proof. We prove for American options.

1. Assume $t=0$ and write

$$\mathrm{C}_a(0,x,\mathrm{K},r,\delta) = \sup_{\tau\in\mathfrak{I}_{0,\mathrm{T}}} \mathbb{E}^\delta\left[e^{-r\tau}\max\left\{xe^{(r-\delta-(\sigma^2/2)\tau)+\sigma \mathrm{W}^\delta_\tau} - \mathrm{K},\ 0\right\}\right]$$

2. Let $\hat{\mathrm{W}}^\delta_t = \mathrm{W}^\delta_t - \sigma t$ and $\hat{\mathbb{P}}^\delta$ be the probability measure with density

$$\frac{\mathrm{d}\hat{\mathbb{P}}^\delta}{\mathrm{d}\mathbb{P}^\delta} = e^{\sigma \mathrm{W}^\delta_\mathrm{T} - (\sigma^2/2)\mathrm{T}}$$

3. Thus, using that $(\hat{\mathrm{W}}^\delta_t)$ is a $\hat{\mathbb{P}}^\delta$-martingale,

$$\begin{aligned}
&\mathbb{E}^\delta\left[e^{-r\tau}\max\left\{xe^{(r-\delta-(\sigma^2/2)\tau)+\sigma \mathrm{W}^\delta_\tau} - \mathrm{K},\ 0\right\}\right] \\
&\quad = \mathbb{E}^\delta\left[e^{-\delta\tau}e^{\sigma \mathrm{W}^\delta_\tau - (\sigma^2/2)\tau}\max\left\{x - \mathrm{K}e^{(\delta-r-(\sigma^2/2)\tau)+\sigma \mathrm{W}^\delta_\tau},\ 0\right\}\right] \\
&\quad = \mathbb{E}^\delta\left[e^{-\delta\tau}e^{\sigma \mathrm{W}^\delta_\mathrm{T} - (\sigma^2/2)\mathrm{T}}\max\left\{x - \mathrm{K}e^{(\delta-r-(\sigma^2/2)\tau)+\sigma \hat{\mathrm{W}}^\delta_\tau},\ 0\right\}\right]
\end{aligned}$$

4. Hence,

$$\begin{aligned}
&\mathbb{E}^\delta\left[e^{-r\tau}\max\left\{xe^{(r-\delta-(\sigma^2/2)\tau)+\sigma \mathrm{W}^\delta_\tau} - \mathrm{K},\ 0\right\}\right] \\
&\quad = \hat{\mathbb{E}}^\delta\left[e^{-\delta\tau}\max\left\{x - \mathrm{K}e^{(\delta-r-(\sigma^2/2)\tau)+\sigma \hat{\mathrm{W}}^\delta_\tau},\ 0\right\}\right]
\end{aligned}$$

$\square$

Appendix A

Measure Theory

A.1 Measurable Spaces

Given a set S, on what collection $\mathfrak{S}$ of subsets of S are suitable to be a domain of measure?

Definition A.1.1. Let S be a set and $\mathfrak{S}$ be a family of subsets of S. Then $\mathfrak{S}$ is called a **σ-algebra** on S if

1. $\emptyset, S \in \mathfrak{S}$.
2. If $A \in \mathfrak{S}$, then $A^c \in \mathfrak{S}$.
3. If $A_1, A_2, \ldots \in \mathfrak{S}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathfrak{S}$.

The pair $(S, \mathfrak{S})$ is said to be a **measurable space**, and any subset $A \subseteq \mathfrak{S}$ is called an **$\mathfrak{S}$-measurable set**.

For any set S and any collection $\mathfrak{A}$ of subsets of S there is at least one σ-algebra containing $\mathfrak{A}$: the family of all subsets of S. Taking the intersection of all the σ-algebras containing $\mathfrak{A}$, we obtain the smallest σ-algebra containing $\mathfrak{A}$, which is called the **σ-algebra generated by** $\mathfrak{A}$.

Particularlly, the smallest σ-algebra containing all of the open sets of $\mathbf{R}$ is called the **Boreal algebra** for $\mathbf{R}$ and is denoted by $\mathfrak{B}$. Any set in $\mathfrak{B}$ is called a **Borel set**.

A.2 Measures

Now, how can we assign a size (or a probability) to all the sets in $\mathfrak{S}$?

Definition A.2.1. Let $(S, \mathfrak{S})$ be a measurable space. A **measure** is an extended real-valued function $\mu : \mathfrak{S} \longrightarrow \overline{\mathbf{R}}$ such that

1. $\mu(\emptyset) = 0$;
2. $\mu(A) \geq 0$, for all $A \in \mathfrak{S}$;

3. If $(A_n)_{n=1}^{\infty}$ is a countable, disjoint sequence of subsets in $\mathfrak{S}$, then $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$.

The triple $(S, \mathfrak{S}, \mu)$ is called a **measure space**.

A measure is nonnegative, assigns zero to the null set, and is **countably additive**. If $\mu(S) < \infty$, then μ is finite.

We say that a proposition holds **almost everywhere (a.e.)** if there exists a set $A \in \mathfrak{S}$ with $\mu(A) = 0$, such that the proposition holds on the complement of A. Intuitively, the proposition holds everywhere except on sets of measure zero.

For example, a sequence of functions (f_n) on S converges a.e. to a function f if there exists $A \in \mathfrak{S}$ with $\mu(A) = 0$ such that $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in A^c$.

If $\mu(S) = 1$, then μ is a **probability measure** and $(S, \mathfrak{S}, \mu)$ is called a **probability space**. Any measurable set $A \in \mathfrak{S}$ is called and **event** and $\mu(A)$ is the **probability of the event** A. In a probability space, the phrase **almost surely (a.s.)** is used interchangeably with 'almost everywhere'.

Notice that if $A, B \in \mathfrak{S}$ and $A \subseteq B$, then $\mu(A) \leq \mu(B)$. If $\mu(A) < \infty$, then $\mu(B \backslash A) = \mu(B) - \mu(A)$.

Theorem A.2.1. Let $(S, \mathfrak{S}, \mu)$ be a measure space.

1. If $(A_n)_{n=1}^{\infty}$ is an increasing sequence in $\mathfrak{S}$ (i.e. if $A_n \subseteq A_{n+1}$ for all n), then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n)$$

2. If $(B_n)_{n=1}^{\infty}$ is a decreasing sequence in $\mathfrak{S}$ (i.e. if $B_{n+1} \subseteq B_n$ for all n) and if $\mu(B_m) < \infty$ for some m, then

$$\mu\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mu(B_n)$$

First, we define a measure on a small family of sets and then present an extension theorem.

Definition A.2.2. Let S be a set and $\mathfrak{A}$ be a family of subsets of S. Then $\mathfrak{A}$ is called an **algebra** if

1. $\emptyset, S \in \mathfrak{A}$;
2. $A \in \mathfrak{A}$ implies $A^c \in \mathfrak{A}$;
3. $A_1, A_2, \ldots, A_n \in \mathfrak{A}$ implies $\bigcup_{i=1}^{n} A_i \in \mathfrak{A}$.

I.e., an algebra is closed under complementation and finite union. On an algebra, the idea of measure is very much similar to the definition on a σ-algebra.

Definition A.2.3. Let S be a set, and let $\mathfrak{A}$ be an algebra of subsets of S. A **measure** is a real-valued function $\mu : \mathfrak{A} \longrightarrow \overline{\mathbf{R}}$ such that

1. $\mu(\emptyset) = 0$;
2. $\mu(A) \geq 0$, for all $A \in \mathfrak{A}$;

3. If $(A_n)_{n=1}^{\infty}$ is a disjoint sequence of sets in $\mathfrak{A}$ with $\bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}$, then $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$.

Notice that last item is different from the definition on a σ-algebra. Here, we require that the union is contained in the algebra.

Although it is easier to define measures on algebras, it is more convenient to work with σ-algebras.

Theorem A.2.2 (Caratheodory Extension Theorem)**.** Let S be a set, $\mathfrak{A}$ an algebra of its subsets, and μ a measure on $\mathfrak{A}$. Let $\mathfrak{S}$ be the smallest σ-algebra containing $\mathfrak{A}$. Then there exists a measure μ^* on $\mathfrak{S}$ such that $\mu(A) = \mu^*(A)$, for all $A \in \mathfrak{A}$.

To rule out the possibility of more than one extension of μ to $\mathfrak{S}$, we'll need the following definition

Definition A.2.4. Let S be a set, $\mathfrak{A}$ an algebra of its subsets, and μ a measure on $\mathfrak{A}$. If there is a countable sequence of sets $(A_i)_{i=1}^{\infty} \in \mathfrak{A}$ with $\mu(A_i) < \infty$ for all i, and $S = \bigcup_{i=1}^{\infty} A_i$, then μ is called **σ-finite**.

By definition, any probability measure is σ-finite. And the next theorem shows that the extension of a σ-finite measure is unique.

Theorem A.2.3 (Hahn Extension Theorem)**.** Let S, $\mathfrak{A}$, μ and $\mathfrak{S}$ be as specified in Caratheodory Extension Theorem. If μ is σ-finite, then the extension μ^* to $\mathfrak{S}$ is unique.

Definition A.2.5. Let $(S, \mathfrak{S}, \mu)$ be a measure space, $A \in \mathfrak{S}$ be any set with measure zero, and let $C \subseteq A$. Denoting by $\mathfrak{C}$ the family of such sets C, i.e.,

$$\mathfrak{C} = \{C \subset S : C \subseteq A \text{ for some } A \in \mathfrak{S} \text{ with } \mu(A) = 0\}$$

The **completion** of $\mathfrak{S}$ is the family $\mathfrak{S}'$ constructed by starting with any set $B \in \mathfrak{S}$, and then adding and subtracting from it sets in $\mathfrak{C}$. That is

$$\mathfrak{S}' = \{B' \subseteq S : B' = (B \cup C_1) \setminus C_2,\ B \in \mathfrak{S},\ C_1, C_2 \in \mathfrak{C}\}$$

Intuitively, $\mathfrak{S}'$ consists of all the subsetes of S that differ from a set in $\mathfrak{S}$ by a setset of a set of μ-measure zero. Using this definition, a measure μ on $(S, \mathfrak{S})$ can be extended to $(S, \mathfrak{S}')$.

A.3 Measurable Functions

Definition A.3.1. Given a measurable space $(S, \mathfrak{S})$, a real-valued function $f : S \longrightarrow \mathbf{R}$ is **measurable with respect to** $\mathfrak{S}$ (or **$\mathfrak{S}$-measurable**) if

$$\{s \in S : f(s) \leq a\} \in \mathfrak{S},\ \forall a \in \mathbf{R}$$

If the space is a probability space, then f is called a **random variable**.

Some cases in which measurability can be easily verified:

- Any monotone or continuous function $f : \mathbf{R} \longrightarrow \mathbf{R}$ is measurable with respect to Borel sets.

- If S is a countable set and $\mathfrak{S}$ is a complete σ-algebra for S (i.e. contains all subsets of S), then all functions $f : S \longrightarrow \mathbf{R}$ are $\mathfrak{S}$-measurable.

In more general cases, the measurability of a function is established by showing that it is the limit of a sequence of 'simpler' functions. We'll do this progressively.

First, let $(S, \mathfrak{S})$ be a measurable space and consider the indicator function $\chi_A : S \longrightarrow \mathbf{R}$ of the from

$$\chi_A(s) = \begin{cases} 1, & \text{if } s \in A \\ 0, & \text{if } s \notin A \end{cases}$$

Clearly, χ_A is $\mathfrak{S}$-measurable iff. $A \in \mathfrak{S}$.

Now, consider the finite weighted sums of indicator functions:

$$\varphi(s) = \sum_{i=1}^{n} a_i \chi_{A_i}(s)$$

where $(A_i)_{i=1}^n$ is a sequence of subsets of S, and $(a_i)_{i=1}^n$ is a sequence of real numbers. Functions like this are called **simple functions**.

If the sets (A_i) form a partition of S and if all of the values a_i are distinct, then the previous equation is the **standard representation** of the function. If this is the case, then φ is measurable iff. each $A_i \in \mathfrak{S}$.

As we'll see in the next two theorems, the set of all measurable functions consists of those that are pointwise limits of measurable simple functions.

First, any function f that is pointwise limit of a sequence (f_n) of measurable function is itself measurable.

Theorem A.3.1 (Pointwise convergence preserves measurability)**.** Let $(S, \mathfrak{S})$ be a measurable space, and let (f_n) be a sequence of $\mathfrak{S}$-measurable functions converging pointwise to f, i.e.,

$$\lim_{n\to\infty} f_n(s) = f(s),\ \forall s \in S$$

Then f is also $\mathfrak{S}$-measurable.

Second, any measurable function f can be expressed as the pointwise limit of a sequence (φ_n) of measurable simple functions. If f is nonnegative, the sequence can be chosen to be strictly increasing. And if f is bounded, the sequence can be chosen to converge uniformly.

The idea is that we can construct a class of measurable functions by taking the class of measurable simple function and then closing this under pointwise convergence. The next result shows that this set contains all the measurable functions.

Theorem A.3.2 (Approximation of measurable functions by simple functions)**.** Let $(S, \mathfrak{S})$ be a measurable space. If $f : S \longrightarrow \mathbf{R}$ is $\mathfrak{S}$-measurable, then there exists a sequence of measurable simple functions (φ_n) such that $\varphi_n \longrightarrow f$ pointwise. If $0 \le f$, then the sequence can be chosen so that

$$0 \le \varphi_n \le \varphi_{n+1} \le f$$

for all n. If f is bounded, then the sequence can be chosen so that $\varphi_n \longrightarrow f$ uniformly.

With these two theorems, we see that a function is measurable iff. it is the pointwise limit of a sequence of measurable simple functions. The standard way to prove that a function is measurable is to find such an approximation.

Some important properties follow. Suppose that f and g are $\mathfrak{S}$-measurable functions on S and $c \in \mathbf{R}$. Then,

- The functions $f+g, f \cdot g, |f|$ and $c \cdot f$ are $\mathfrak{S}$-measurable.
- If (f_n) is a sequence of $\mathfrak{S}$-measurable functions, then $\inf f_n$, $\sup f_n$, $\liminf f_n$ and $\limsup f_n$ are all $\mathfrak{S}$-measurable.
- All continuous functions on $\mathbf{R}^l$ are $\mathfrak{B}^l$-measurable.
- The composition of Borel measurable functions are Borel measurable. However, this is not true for Lebesgue measurable functions.

Definition A.3.2. Let $(\mathrm{S}, \mathfrak{S})$ and $(\mathrm{T}, \mathfrak{T})$ be measurable spaces. Then the function $f : \mathrm{S} \longrightarrow \mathrm{T}$ is **measurable** if the inverse image of every measurable set is measurable, i.e.,

$$\{s \in \mathrm{S} : f(s) \in \mathrm{A}\} \in \mathfrak{S}, \ \forall \mathrm{A} \in \mathfrak{T}$$

Hence, if $(\mathrm{S}, \mathfrak{S})$, $(\mathrm{T}, \mathfrak{T})$, and $(\mathrm{U}, \mathfrak{U})$ are measurable spaces, and $f : \mathrm{S} \longrightarrow \mathrm{T}$ and $g : \mathrm{T} \longrightarrow \mathrm{U}$ are measurable functions, then $h : \mathrm{S} \longrightarrow \mathrm{U}$ defined by $h(s) = g \circ f(s)$ is a measurable function.

Definition A.3.3 (Measurable selection). Let $(\mathrm{S}, \mathfrak{S})$ and $(\mathrm{T}, \mathfrak{T})$ be measurable spaces, and let Γ be a correspondence of S into T. Then the function $h : \mathrm{S} \longrightarrow \mathrm{T}$ is a **measurable selection from** Γ if h is measurable and $h(s) \in \Gamma(s)$, for all $s \in \mathrm{S}$.

Theorem A.3.3 (Measurable Selection). Let $\mathrm{S} \subseteq \mathbf{R}^l$ and $\mathrm{T} \subseteq \mathbf{R}^m$ be Borel sets, with their Borel subsets $\mathfrak{S}$ and $\mathfrak{T}$. Let $\Gamma : \mathrm{S} \longrightarrow \mathrm{T}$ be a nonempty compact-valued and upper hemi-continuous (UHC) correspondence. Then there exists a measurable selection from Γ.

A.4 Integration

Let $(\mathrm{S}, \mathfrak{S}, \mu)$ be a fixed measure space and $\mathrm{M}(\mathrm{S}, \mathfrak{S})$ be the space of measurable, extended real-valued functions on S.

For a nonnegative, measurable and simple function, we have the following definition.

Definition A.4.1. Let $\varphi \in \mathrm{M}^+(\mathrm{S}, \mathfrak{S})$ be a measurable simple function, with the standard representation $\varphi(s) = \sum_{i=1}^n a_i \chi_{\mathrm{A}_i}(s)$. Then the **integral of** φ **with respect to** μ is

$$\int_{\mathrm{S}} \varphi(s)\, \mu(\mathrm{d}s) = \sum_{i=1}^n a_i \mu(\mathrm{A}_i)$$

To extended from simple functions to all $\mathrm{M}^+(\mathrm{S}, \mathfrak{S})$, we define

Definition A.4.2. For $f \in \mathrm{M}^+(\mathrm{S}, \mathfrak{S})$, the **integral of** f **with respect to** μ is

$$\int_{\mathrm{A}} f(s)\, \mu(\mathrm{d}s) = \int_{\mathrm{S}} f(s) \chi_{\mathrm{A}}(s) \mu(\mathrm{d}s)$$

We also denote the integrals above as $\int f\,\mathrm{d}\mu$ and $\int_A f\,\mathrm{d}\mu$.

Since any function $f \in M^+(S, \mathfrak{S})$ can be expressed as the limit of an increasing sequence (φ_n) of simple functions in $M^+(S, \mathfrak{S})$. However, we cannot define the integral as the limit of such a sequence because it would leave open the possibility that the limit depends on the particular sequence chosen. The Monotone Convergence Theorem shows that the limit value is unique.

Before that, the following result shows that a simple function on a measure space can be used to define a new measure on the space.

Lemma A.4.1. If $\varphi \in M^+(S, \mathfrak{S})$ is a simple function and $\lambda : \mathfrak{S} \longrightarrow \mathbf{R}$ is defined by

$$\lambda(A) = \int_A \varphi\,\mathrm{d}\mu$$

for all $A \in \mathfrak{S}$, then λ is a measure on $\mathfrak{S}$.

Theorem A.4.2 (Monotone Convergence Theorem)**.** If (f_n) is a monotone increasing sequence of functions in $M^+(S, \mathfrak{S})$ converging pointwise to f then

$$\int f\,\mathrm{d}\mu = \lim_{n\to\infty} \int f_n\,\mathrm{d}\mu$$

Some important properties:

1. If $f, g \in M^+(S, \mathfrak{S})$ and $c \geq 0$, then

$$\int (f+g)\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\mu$$

   and

$$\int c \cdot f\,\mathrm{d}\mu = c \int f\,\mathrm{d}\mu$$

2. If $f, g \in M^+(S, \mathfrak{S})$ and $f \leq g$, then

$$\int f\,\mathrm{d}\mu \leq \int g\,\mathrm{d}\mu$$

3. If $f \in M^+(S, \mathfrak{S})$ and $A, B \in \mathfrak{S}$ with $A \subseteq B$, then

$$\int_A f\,\mathrm{d}\mu \leq \int_B f\,\mathrm{d}\mu$$

4. If $f \in M^+(S, \mathfrak{S})$ and $A \in \mathfrak{S}$ with $\mu(A) = 0$, then

$$\int_A f\,\mathrm{d}\mu = 0$$

5. If $f \in M^+(S, \mathfrak{S})$, $\int_S f\,\mathrm{d}\mu < \infty$, and $A = \{s \in S : f(s) = +\infty\}$, then $\mu(A) = 0$.

6. If $f \in M^+(S, \mathfrak{S})$ and $\lambda : \mathfrak{S} \longrightarrow \mathbf{R}_+$, defined as $\lambda(A) = \int_A f\,\mathrm{d}\mu$ for all $A \in \mathfrak{S}$, then λ is a measure on $(S, \mathfrak{S})$.

7. Let (g_i) be a sequence of functions in $\mathrm{M}^+(\mathrm{S}, \mathfrak{S})$. Then

$$\int \left(\sum_{i=1}^{n} g_i \right) \mathrm{d}\mu = \sum_{i=1}^{n} \int g_i \, \mathrm{d}\mu$$

Lemma A.4.3 (Fatou's Lemma)**.** If (f_n) is a sequence of functions in $\mathrm{M}^+(\mathrm{S}, \mathfrak{S})$, then

$$\int (\liminf f_n) \, \mathrm{d}\mu \leq \liminf \int f_n \, \mathrm{d}\mu$$

To generalize to functions that take on both positive and negative values, we begin by defining the positive and negative parts of a function f^+ and f^-.

$$f^+(s) = \begin{cases} f(s), & \text{if } f(s) \geq 0 \\ 0, & \text{if } f(s) < 0 \end{cases}$$

and

$$f^-(s) = \begin{cases} -f(s), & \text{if } f(s) \leq 0 \\ 0, & \text{if } f(s) > 0 \end{cases}$$

Hence, if f is measurable, f^+ and f^- are both in $\mathrm{M}^+(\mathrm{S}, \mathfrak{S})$ and $f = f^+ - f^-$. This motivates the following generalization.

Definition A.4.3. Let $(\mathrm{S}, \mathfrak{S}, \mu)$ be a measure space, and let f be a measurable real-valued function on S. If f^+ and f^- both have finite integrals with respect to μ, then f is integrable and the **integral of f with respect to μ** is

$$\int f \, \mathrm{d}\mu = \int f^+ \, \mathrm{d}\mu - \int f^- \, \mathrm{d}\mu$$

In probability, if $(\mathrm{S}, \mathfrak{S}, \mu)$ is a probability space and f is a measurable function, then f is a random variable. And if f is integrable, the integral of f with respect to μ is called the **expected value of f** and is denoted by

$$\mathbb{E}[f] = \int f \, \mathrm{d}\mu$$

Now let $\mathrm{L}(\mathrm{S}, \mathfrak{S}, \mu)$ denote the set of all $\mathfrak{S}$-measurable, real-valued functions on S that are integrable with respect to μ. How can we determine whether a function is integrable?

1. If f is bounded, measurable, real-valued function on S, and $\mu(\mathrm{S}) < \infty$, then f is μ-integrable.
2. f is μ-integrable iff. $|f|$ is μ-integrable. In that case, $|\int f \, \mathrm{d}\mu| \leq \int |f| \, \mathrm{d}\mu$.
3. If f is $\mathfrak{S}$-measurable, g is μ-integrable, and $|f| \leq |g|$, then f is μ-integrable and $\int |f| \, \mathrm{d}\mu \leq \int |g| \, \mathrm{d}\mu$.
4. If f and g are μ-integrable and $\alpha \in \mathbf{R}$, then αf and $f + g$ are μ-integrable, satisfying the standard rules of calculus.

Theorem A.4.4 (Lebesgue Dominated Convergence Theorem)**.** Let $(\mathrm{S}, \mathfrak{S}, \mu)$ be a measure space, and let (f_n) be a sequence of integrable functions that converges almost everywhere to a measurable function f.If there exists an integrable function g such that $|f_n| \leq g$ for all n, then f is integrable and

$$\int f \,\mathrm{d}\mu = \lim \int f_n \,\mathrm{d}\mu$$

Definition A.4.4. Let λ and μ be finite measures on $(\mathrm{S}, \mathfrak{S})$. If

$$\mu(\mathrm{A}) = 0 \implies \lambda(\mathrm{A}) = 0 \,\forall \mathrm{A} \in \mathfrak{S}$$

then λ is **absolutely continuous** with respect to μ, written $\lambda << \mu$.

If there is $\mathrm{A} \in \mathfrak{S}$ such that $\lambda(\mathrm{B}) = \lambda(\mathrm{A} \cap \mathrm{B})$, for all $\mathrm{B} \in \mathfrak{S}$, then λ is **concentrated on** A.

If there are disjoint sets $\mathrm{A}, \mathrm{B} \in \mathfrak{S}$ such that λ is concentrated on A and μ is concentrated on B, then λ and μ are **mutually singular**, written $\lambda \perp \mu$.

Theorem A.4.5 (Radon-Nikodym Theorem)**.** Let λ and μ be σ-finite positive measures on $(\mathrm{S}, \mathfrak{S})$ with $\lambda << \mu$. Then there exists an integrable function h such that

$$\lambda(\mathrm{A}) = \int_{\mathrm{A}} h(s)\, \mu(\mathrm{d}s), \;\text{ all } \mathrm{A} \in \mathfrak{S}$$

And the function h is called the **Radon-Nikodym derivative** of λ with respect to μ.

Lemma A.4.6. Let λ_1 and λ_2 be finite measures on $(\mathrm{S}, \mathfrak{S})$. Then there exists a triple of measures $\gamma, \alpha_1, \alpha_2$ such that

$$\lambda_i = \gamma + \alpha_1$$

where $i = 1, 2$, and $\alpha_1 \perp \alpha_2$.

A.5 Product Spaces

Let $(\mathrm{X}, \mathfrak{X})$ and $(\mathrm{Y}, \mathfrak{Y})$ be measurable spaces, and let $\mathrm{Z} = \mathrm{X} \times \mathrm{Y}$.

Our first task is to define a σ-algebra of subsets of Z that is a natural product of $\mathfrak{X}$ and $\mathfrak{Y}$.

Definition A.5.1 (Measurable rectangle)**.** A set $\mathrm{C} = \mathrm{A} \times \mathrm{B} \subseteq \mathrm{Z}$ is a **measurable rectangle** if $\mathrm{A} \in \mathfrak{X}$ and $\mathrm{B} \in \mathfrak{Y}$.

We denote by $\mathfrak{C}$ the set of all measurable rectangles and $\mathfrak{E}$ the set of all finite unions of measurable rectangles. Notice that $\mathfrak{E}$ is an algebra and that every set in $\mathfrak{E}$ can be written as the finite union of disjoint measurable rectangles.

Definition A.5.2 (Product Space)**.** Let $\mathfrak{Z} = \mathfrak{X} \times \mathfrak{Y}$ be the σ-algebra generated by $\mathfrak{E}$. The measurable space $(\mathrm{Z}, \mathfrak{Z})$ is called the **product space**.

Theorem A.5.1. Let $(\mathrm{X}, \mathfrak{X})$, $(\mathrm{Y}, \mathfrak{Y})$, $\mathfrak{C}$, and $\mathfrak{E}$ be as specified above. Let $\mu : \mathfrak{C} \longrightarrow \mathbf{R}_+$ have the following properties:

- $\mu(\emptyset) = 0$;
- If $(C_i) = ((A_i \times B_i))_{i=1}^{\infty}$ is a sequence of disjoint sets in $\mathfrak{C}$ and $\bigcup_{i=1}^{\infty} C_i$ is in $\mathfrak{C}$, then $\mu\left(\bigcup_{i=1}^{\infty} C_i\right) = \sum_{i=1}^{\infty} \mu(C_i)$.

Then there is a measure on $\mathfrak{E}$ that coincides with μ on $\mathfrak{C}$.

This theorem can be naturally extended to any space that is the product of a finite number of measurable spaces.

Definition A.5.3. Let $(X, \mathfrak{X})$ and $(Y, \mathfrak{Y})$ be measurable spaces, and let $(Z, \mathfrak{Z})$ be the product space. And let $E \subseteq Z$ and $x \in X$. Then the ***x*-section of** E is the set $E_x = y \in Y : (x,y) \in E$. And the ***y*-section of** E is the set $E_y = x \in X : (x,y) \in E$.

Let $f : Z \longrightarrow \mathbf{R}$ and let $x \in X$. The ***x*-section of** f is the function $f_x : Y \longrightarrow \mathbf{R}$ defined by $f_x(y) = f(x,y)$. And the ***y*-section of** f is the function $f_y : X \longrightarrow \mathbf{R}$ defined by $f_y(x) = f(x,y)$.

Intuitively, the x-section of a function is found by fixing x and viewing f only as a function of y.

Theorem A.5.2. Let $(X, \mathfrak{X})$ and $(Y, \mathfrak{Y})$ be measurable spaces, and let $(Z, \mathfrak{Z})$ be the product space.

1. If the set E in Z is $\mathfrak{Z}$-measurable, then every section of E is measurable.
2. If the function $f : Z \longrightarrow \mathbf{R}$ is $\mathfrak{Z}$-measurable, then every section of f is measurable.

A.6 The Monotone Class Lemma

The main question that this section aims to answer is whether a propery P holds at a given σ-algebra.

Definition A.6.1 (Monotone Class)**.** A **monotone class** is a nonempty collection $\mathfrak{M}$ of sets such that it contains

1. The union of every nested increasing sequence $A_1 \subseteq A_2 \subseteq \ldots$ of sets in $\mathfrak{M}$.
2. The intersection of every nested decreasing sequence $A_1 \supseteq A_2 \supseteq \ldots$ of sets in $\mathfrak{M}$.

Some important facts:

1. Every σ-algebra is a monotone class.
2. Given a nonempty collection of subsets of S, denoted by $\mathfrak{A}$, there exists a smallest monotone class containing $\mathfrak{A}$, which is called the monotone class generated by $\mathfrak{A}$.
3. The σ-algebra generated by $\mathfrak{A}$ contains the monotone class generated by $\mathfrak{A}$.
4. If a monotone class is an algebra, then it is a σ-algebra.

Lemma A.6.1 (Monotone class)**.** Leet S be a set and let $\mathfrak{A}$ be an algebra of subsets of S. Then the monotone class $\mathfrak{M}$ generated by $\mathfrak{A}$ is the same as the σ-algebra $\mathfrak{S}$ generated by $\mathfrak{A}$.

Using this lemma, our original task in this section is solved. To establish that P holds for all sets in a product σ-algebra, it suffices to show that

1. P holds for all finite unions of disjoint measurable rectangles.
2. The family of sets $\mathfrak{C}$ for which P holds is a monotone class.

A.7 Modes of Convergence

To the traditional ideas of convergence **pointwise** and **uniformly**, learned in an undergraduate course in Analysis, we'll add new forms of convergence, as the functions f_n acquire an infinite and inequivalent number of ways to approach their limit f.

Let $(X, \mathfrak{B}, \mu)$ be a measure space, and the functions f_n (and their limit f) measurable with respect to this space. Then, we say that f_n converges to f

1. **pointwise almost everywhere** if, for (μ–)almost everywhere $x \in X$, $f_n(x)$ converges to $f(x)$.
2. **uniformly almost everywhere, essentially uniformly** , or **in** L^∞ **norm** if, for every $\varepsilon > 0$, there exists N such that for every $n \geq N$,
$$|f_n(x) - f(x)| \leq \varepsilon$$
for μ-almost every $x \in X$.
3. **almost uniformly** if, for every $\varepsilon > 0$, there exists a set $E \in \mathfrak{B}$ of measure $\mu(E) \leq \varepsilon$ such that f_n converges uniformly to f on the complement of E.
4. **in** L^1 **norm** if the quantity
$$\|f_n - f\|_{L^1(\mu)} = \int_x |f_n(x) - f(x)|\, d\mu$$
converges to 0 as $n \to \infty$.
5. **in measure** if, for every $\varepsilon > 0$, the measures
$$\mu(\{x \in X : |f_n(x) - f(x)| \geq \varepsilon\})$$
converge to zero as $n \to \infty$.

In probability theory, if f_n and f are random variables, convergence in L^1 is usually referred to as **convergence in mean**, pointwise convergence almost everywhere is referred to as **almost sure convergence**, and convergence in measure is often referred to as **convergence in probability**.

Recall that a property $P(x)$ is said to hold **almost everywhere** if the set of x for which $P(x)$ fails has Lebesgue measure zero.

A.8 Conditional Expectation

Let $(\Omega, \mathfrak{F}, \mu)$ be a probability space.

Definition A.8.1 (Conditional Probability). For a measurable set A satisfying $\mu(A) > 0$ and for any measurable set B we define the **conditional probability** of B given A as

$$\mu_A(B) := \frac{\mu(B \cap A)}{\mu(A)}$$

We'll also use $P(B \mid A)$ to denote this probability.

Now notice that $\mu_A : \mathfrak{F} \longrightarrow [0,1]$, i.e., it is also a probability measure.

Definition A.8.2 (Conditional Expectation). The **conditional expectation** of f given A is defined as

$$\mathbb{E}[f \mid A] = \int f \, d\mu_A$$

Definition A.8.3 (Measurable Partition). A family ${A_\eta}_{\eta \in H}$ of subsets of Ω is a **measurable partition** of Ω if the following conditions hold:

1. $A_\eta \in \mathfrak{F}$, $\forall \eta \in H$.
2. $\bigcup_{\eta \in H} A_\eta = H$.
3. $A_\eta \cap A_{\eta'} = 0$, $\forall \eta \neq \eta'$.

If the index set H is countable, then we call the measurable partition **countable**.

To formalize it better, extending the concept to σ-algebras not generated by a countable partition, we'll introduce the following.

Definition A.8.4 (Conditional Expectation). Let $A \subset \mathfrak{F}$ be a σ-algebra and let $f : \Omega \longrightarrow \mathbf{R}$ be an integrable function. Then the **conditional expectation of f relative to** A is an $\mathfrak{A}$-measurable function $\mathbb{E}[f \mid \mathfrak{A}] : \Omega \longrightarrow \mathbf{R}$ such that

$$\int_C \mathbb{E}[f \mid \mathfrak{A}](\omega)\mu\,(d\omega) = \int_C f(\omega)\mu\,(d\omega),\ \forall C \in \mathfrak{A}$$

It can be proved that this function exists and is unique in the sense that if g also satisties the condition above, then $\mathbb{E}[f \mid \mathfrak{A}] = g$.

Notice that to compute conditional probability $P(B \mid A_\eta)$ we may take f to be the indicator function χ_B. Then

$$\mathbb{E}[\chi_B \mid \mathfrak{A}](\hat{\omega})\mu(A_\eta) = \int_{A_\eta} \chi_b \, d\mu = \mu(B \cap A_\eta)$$

for all $\hat{\omega} \in A_\eta$, all $\eta \in H$.

If $\mu(A_\eta) > 0$, then

$$\mathbb{E}[\chi_B \mid \mathfrak{A}](\hat{\omega}) = \frac{\mu(B \cap A_\eta)}{\mu(A_\eta)}$$

Summarizing this fact,

$$P(B \mid A_\eta) = \mathbb{E}[\chi_B \mid \mathfrak{A}](\hat{\omega})$$

for all $\hat{\omega} \in A_\eta$, all $\eta \in H$.

We end this section with an important property of conditional expectations. Let $\mathfrak{A}_1 \subseteq \mathfrak{A}_2 \subseteq \mathfrak{F}$. Then we have

$$\begin{aligned}\int_C \mathbb{E}[\mathbb{E}(f \mid \mathfrak{A}_2) \mid \mathfrak{A}_1]\, d\mu &= \int_c \mathbb{E}(f \mid \mathfrak{A}_2)(\omega)\, d\mu \\ &= \int_C f(\omega)\, d\mu \\ &= \int_C \mathbb{E}[f \mid \mathfrak{A}_1](\omega)\, d\mu,\ \forall C \in \mathfrak{A}_1\end{aligned}$$

This is known as the **law of the iterated expectation**. To state it in another way,

$$\int_C \mathbb{E}[f - \mathbb{E}(f \mid \mathfrak{A}_2) \mid \mathfrak{A}_1]\, d\mu = 0,\ \forall C \in \mathfrak{A}_1$$

Bibliography

[BRR96] Martin Baxter, Andrew Rennie, and Andrew JO Rennie. *Financial calculus: an introduction to derivative pricing*. Cambridge university press, 1996.

[Dur19] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[Eva12] Lawrence C Evans. *An introduction to stochastic differential equations*, volume 82. American Mathematical Soc., 2012.

[Gar88] Thomas C Gard. *Introduction to Stochastic Differential Equations*. Dekker, Inc, 1988.

[Jak] Jake. How to draw brownian motions in tikz/pgf. https://tex.stackexchange.com/a/59934. [Acesso em: 07/07/2022].

[Kle12] Fima C Klebaner. *Introduction to stochastic calculus with applications*. World Scientific Publishing Company, 2012.

[LL11] Damien Lamberton and Bernard Lapeyre. *Introduction to stochastic calculus applied to finance*. CRC press, 2011.

[Oks13] Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

[S+04] Steven E Shreve et al. *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer, 2004.

[Wik22] Wikipedia. Kalman filter — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Kalman%20filter&oldid=1091166052, 2022. [Online; accessed 06-July-2022].