Lecture Notes on Stochastic Analysis for Finance

Melih İşeri *

November 30, 2025

Abstract

These lecture notes cover the fundamentals of stochastic processes, starting from defining the Lebesgue integral and progressing through a variety of major theorems such as Itô's formula, the Martingale Representation Theorem, and Girsanov's Theorem, all proved under strong regularity assumptions.

Following the theoretical groundwork, the notes explore Stochastic Differential Equations (SDE)s, Forward-Backward SDEs, and the Feynman-Kac formula, discussing these topics without rigorously treating well-posedness.

The final part of the notes transitions to applications in quantitative finance, introducing the Black-Scholes model and discussing derivative pricing, risk-neutral valuation, and hedging strategies.

The content on measure theory largely follows Folland [3]. The material leading up to SDEs mainly follows Zhang [6], with additional details from Durrett [2], Karatzas, and Shreve [4], and Cohen [1]. Subsequently, the concepts pertaining to finance are informed by Björk [5].

*Department of Mathematics, University of Michigan, United States, iseri@umich.edu.

Contents

1 Fundamental Concepts

1.1 A Tour in Measure Theory

In this section, we will introduce the fundamentals of measure theory and provide an exposure to the core theorems.

The main objective of measure theory is to study functions that assign a real value to subsets of a given set. As a core structural requirement, we aim to understand functions that are additive under disjoint subsets.

In geometry, which is vital for intuition, this notion corresponds to length, area, volume etc. Importantly, measure theory allows the study of integration in a robust way, allowing powerful limit theorems to hold. In our context, probability theory is built upon measure theory.

It turns out that not all subsets of a given set can be considered. Vague intuition is that, a set has no constraints, it is just a collection. And without any requirements, one can construct pathological examples in mathematics. Banach and Tarski proved the following interesting result in 1924:

Let U, V be arbitrary bounded open sets in $\mathbb{R}^m$ for $m \geq 3$. There exists $k \in \mathbb{N}$ and subsets $E_1, \cdots, E_k$, $F_1, \cdots, F_k$ of $\mathbb{R}^m$ such that E_i's partition U, F_i's partition V and E_i is congurent (translation + rotation + reflection) to F_i.

Definition 1.1. A collection of subsets $\mathcal{F}$ of a set Ω is called a σ-algebra, if

- If $\{E_k\}_{k=1}^{\infty} \in \mathcal{F}$, then $\cup_k E_k \in \mathcal{F}$.
- If $E \in \mathcal{F}$, then $E^c \in \mathcal{F}$.

For any set Ω, all subsets 2^{Ω} and $\{\emptyset, \Omega\}$ (trivial) are σ-algebras. Furthermore, intersection of σ-algebras is itself a σ-algebra. Therefore, given any collection of subsets $\mathcal{A}$, we define $\sigma(\mathcal{A})$ as the σ-algebra generated by $\mathcal{A}$.

Exercise 1.2. Show that

(i) $\emptyset$ and Ω is contained in any σ-algebra,

(ii) σ-algebra is closed under countable intersections,

(iii) Intersection of arbitrary family of σ-algebras is a σ-algebra. Relying on this, define the σ-algebra generated by any collection of subsets $\mathcal{E}$, denoted as $\sigma(\mathcal{E})$.

So, which σ-algebra to take? For any topological space, the most important σ-algebra is the one generated by all the open sets. We call it the Borel σ-algebra. Roughly speaking, Borel σ-algebra contains all the countable intersections and unions of open sets. It is informative to keep in mind (i) geometric objects with partitions, (ii) intervals in (real) numbers, (iii) open balls around continuous functions. (i) serves well for an abstract case, and the discrete nature of it helps the intuition greatly. (ii) is crucial as numbers are, however, as in every area of mathematics, concepts becomes unimaginably powerful once applied to functions and (iii) will serve us as a basis in stochastic processes.

Next proposition is the underlying reason why cumulative distribution function characterizes a probability distribution. The proof (omitted) relies on the fact that every open set in $\mathbb{R}$ is a countable union of open intervals.

Proposition 1.3. *The Borel σ-algebra on $\mathbb{R}$ is generated by* intervals. *Any collection of intervals, such as (a, b), $[a, b]$, $(-\infty, a)$, etc. can be used as a generator of the Borel σ-algebra.*

Definition 1.4. A measure on $(\Omega, \mathcal{F})$ is a function $\mu : \mathcal{F} \to [0, \infty]$ such that

- $\mu(\emptyset) = 0$
- $\mu(\cup_{k=1}^{\infty} E_k) = \sum_{k=1}^{\infty} \mu(E_k)$ for any collection of disjoint subsets E_k in $\mathcal{F}$.

We call $(\Omega, \mathcal{F}, \mu)$ a measure space.

Example 1.5. Let $\mathcal{F} = 2^{\Omega}$[1] for any set Ω, and take any $\rho : \Omega \to [0, \infty]$. Then,

$$\mu(E) := \sum_{x \in E} \rho(x) := \sup \left\{ \sum_{x \in F} \rho(x) \ : \ F \subset E, F \text{ finite} \right\}$$

is a measure on $(\Omega, \mathcal{F})$. In general, ρ might be understood as a density. Two special cases are important:

(i) If $\rho(x) = 1$ for all x, it is called *counting measure*.

(ii) If $\rho(x_0) = 1$ for some $x_0 \in \Omega$ and 0 otherwise, it is called *Dirac measure* or *point mass*.

Next theorem clarifies why probability distributions are characterized by their cumulative distributions.[2]

Theorem 1.6. *Let $F : \mathbb{R} \to \mathbb{R}$ be any increasing, right continuous function. Then there exists a unique measure μ_F on $\mathbb{R}$ with $\mu_F((a, b)) = F(a) - F(b)$. If G is another such function, we have $\mu_F = \mu_G$ if and only if $F - G$ is constant.*

Note that we can generate a significant amount of measures on $\mathbb{R}$ by above theorem. Most important example is the so called Lebesgue measure $\mathbf{m}$ on $\mathbb{R}$, which is the measure associated with $F(x) = x$ where $\mathbf{m}((a, b)) = b - a$. Some basic properties are as follows,

Theorem 1.7. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.*

(Monotonicity) *If $E \subset F$, then $\mu(E) \leq \mu(F)$.*

(Subadditivity) *$\mu(\cup_k E_k) \leq \sum_k \mu(E_k)$.*

(Continuity) *If $E_1 \subset E_2 \subset \cdots$, then $\mu(\cup_k E_k) = \lim_k \mu(E_k)$.*
If $E_1 \supset E_2 \supset \cdots$ and $\mu(E_1) < \infty$ then $\mu(\cap_k E_k) = \lim_k \mu(E_k)$.

Proof. (Monotonicity) Since $\mu(F) = \mu(E \cup (F \setminus E)) = \mu(E) + \mu(F \setminus E)$, and $\mu(\cdot) \in [0, \infty]$, we are done.

(Subadditivity) Set $F_1 = E_1$ and $F_k = E_k \setminus (\bigcup_{j=1}^{k-1} E_j)$. Then, F_k's are disjoint and $\cup_{j=1}^{k} F_j = \cup_{j=1}^{k} E_j$ for all k. Then,

$$\mu(\cup_{k=1}^{\infty} E_k) = \mu(\cup_{k=1}^{\infty} F_k) = \sum_{k=1}^{\infty} \mu(F_k) \leq \sum_{k=1}^{\infty} \mu(E_k)$$

(Continuity from below)

$$\mu(\cup_{k=1}^{\infty} E_k) = \lim_{n \to \infty} \sum_{k=1}^{n} \mu(E_k \setminus E_{k-1}) = \lim_{n \to \infty} \mu(E_n)$$

[1] 2^{Ω} denotes all the subsets.

[2] Probability distribution means $\mu(\Omega) = 1$.

(Continuity from above) Reverse the sequence in the sense that $F_k := E_1 \setminus E_k$. Then $F_1 \subset F_2 \subset \cdots$. Also, $\mu(E_1) = \mu(F_k) + \mu(E_k)$ and $\bigcup_{k=1}^{\infty} F_k = E_1 \setminus (\bigcap_{k=1}^{\infty} E_k)$. Then, by continuity from below,

$$\mu(E_1) = \mu(\cap_{k=1}^{\infty} E_k) + \lim_{k\to\infty} \mu(F_k) = \mu(\cap_{k=1}^{\infty} E_k) + \lim_{k\to\infty} (\mu(E_1) - \mu(E_k))$$

and substract $\mu(E_1)$ from both sides to get the result. ■

Exercise 1.8. (i) Find a sequence E_k such that $\mu(E_1) = \infty$ and continuity from above fails.

(ii) Show that, if $\mu_1, \mu_2, \cdots$ are measures on $(\Omega, \mathcal{F})$, and $a_1, a_2, \cdots \in [0, \infty)$, then $\sum_{k=1}^{\infty} a_k \mu_k$ is a measure on $(\Omega, \mathcal{F})$.

(iii) $(\Omega, \mathcal{F}, \mu)$ is a measure space. Show that $\mu(E) + \mu(F) = \mu(E \cup F) + \mu(E \cap F), \forall E, F \in \mathcal{F}$.

(iv) $(\Omega, \mathcal{F}, \mu)$ is a measure space and fix $E \in \mathcal{F}$. Show that $\mu_E(A) := \mu(A \cap E)$ is a measure.

We say $E \in \mathcal{F}$ is a null set if $\mu(E) = 0$. If a statement is true for all $\omega \in \Omega$ excluding a null set, then we say it holds *almost surely*, or *almost everywhere*.

Next, we will discuss measurable functions. First, recall that any function $f : \Omega \to \Lambda$ induces a mapping $f^{-1} : 2^{\Omega} \to 2^{\Lambda}$ defined as $f^{-1}(E) = \{x \in X : f(x) \in E\}$ which preserves unions, intersection and complements. Therefore, if $\mathcal{G}$ is a σ-algebra for Λ, then $\{f^{-1}(E) : E \in \mathcal{B}\}$ is a σ-algebra for Ω.

Definition 1.9. Given two measurable spaces $(\Omega, \mathcal{F})$, $(\Lambda, \mathcal{G})$, a function $f : \Omega \to \Lambda$ is called measurable if $f^{-1}(E) \in \mathcal{F}$ for all $E \in \mathcal{G}$.

Proposition 1.10. *If X, Y are topological spaces, any continuous function is measurable when X, Y are equipped with Borel σ-algebras.*

In fact, measurable functions are closely related to continuous function but we will not explore this. As an informative example, note that the function $f(x) = 1$ for all $x \in [0, 1] \setminus \mathbb{Q}$ and 0 otherwise is a measurable function. Since $\mathbf{m}(\mathbb{Q}) = 0$, for arbitrary $\varepsilon > 0$, one can find a domain with measure $1 - \varepsilon$ for which f is continuous.

Introduce the *indicator function* or *characteristic function* as

$$\mathbf{1}_{\{E\}}(\omega) := \begin{cases} 1 & \text{if } \omega \in E \\ 0 & \text{if } \omega \notin E \end{cases}$$

which is measurable iff E is in the σ-algebra. Then we have the definition of functions that the integration is build on.

Definition 1.11. We say $\phi : \Omega \to \mathbb{R}$ is simple, if ϕ is measurable and the range is a finite subset of $\mathbb{R}$. The *standard representation* of ϕ is

$$\phi = \sum_{k=1}^{n} x_k \mathbf{1}_{\{E_k\}}, \text{ where } E_k = \phi^{-1}(x_k), \ \text{range}(\phi) = \{x_1, \cdots, x_n\}$$

We are ready to talk about the integration now. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. First, we define the integral of a simple function ϕ with the standard representation as

$$\int \phi d\mu := \sum_k x_k \mu(E_k), \quad \text{and} \quad \int_A \phi d\mu := \int \phi \mathbf{1}_{\{A\}} d\mu := \int \phi \mathbf{1}_{\{A\}}, \forall A \in \mathcal{F} \tag{1.1}$$

Define L^+ as the space of all measurable positive functions,

$$L^+ := \left\{ \text{all measurable } f : \Omega \to \mathbb{R}^+ \right\}$$

Proposition 1.12. *Let $\phi, \varphi \in L^+$ be simple functions. Then,*

(i) If $c > 0$, then $\int c\phi = c\int \phi$.

(ii) $\int(\phi + \varphi) = \int \phi + \int \varphi$

(iii) If $\phi \le \varphi$, then $\int \phi \le \int \varphi$.

(iv) The map $A \mapsto \int_A \phi d\mu$ is a measure.

Proof. (i) is obvious. For (ii) and (iii): Let $\sum_k x_k \mathbf{1}_{\{E_k\}}$ and $\sum_\ell y_\ell \mathbf{1}_{\{F_\ell\}}$ be the standard representations of ϕ and φ. Note that since $\cup_k E_k = \cup_\ell F_k = \Omega$ are disjoint decompositions of Ω, $E_k = \bigcup_\ell (E_k \cap F_\ell)$ and $F_\ell = \bigcup_k (F_\ell \cap E_k)$ are disjoint decompositions.
(iii): First, $\phi \le \varphi$ means $x_k \le y_\ell$ whenever $\mu(E_k \cap F_\ell) \ne 0$. Therefore

$$\int \phi = \sum_{k,\ell} x_k \mu(E_k \cap F_\ell) \le \sum_{k,\ell} y_\ell \mu(E_k \cap F_\ell) = \int \varphi$$

(ii): Next,

$$\int \phi + \int \varphi = \sum_{k,\ell} (x_k + y_\ell)\mu(E_k \cap F_\ell) = \int \phi + \varphi$$

(iv): Lastly, let $A_m \in \mathcal{F}$ be a collection of disjoint subsets. Then,

$$\int_{\cup_m A_m} \phi d\mu = \int \phi \mathbf{1}_{\{\cup_m A_m\}} = \sum_k x_k \mu(E_k \cap (\cup_m A_m)) = \sum_{k,m} x_k \mu(E_k \cap A_m) = \sum_m \int_{A_m} \phi$$

■

We now lift the definition of integral to any $f \in L^+$ as

$$\int f d\mu := \int f := \sup \left\{ \int \phi d\mu \ : \ 0 \le \phi \le f, \ \phi \text{ simple} \right\} \tag{1.2}$$

Since all the functions can be decomposed as $f = f^+ - f^-$[3] to negative and positive parts, we can define integrals if $\int |f| d\mu < \infty$, which we denote all such functions as L^1,

$$L^1 := \left\{ \text{all measurable } f : \Omega \to \mathbb{R} \text{ s.t. } \int |f| d\mu < \infty \right\}$$

Exercise 1.13. Show that,

(i) when f is a simple function, (1.2) agrees with (1.1).

(ii) $c\int f = \int cf$, and if $f \le g$ then $\int f \le \int g$.

Example 1.14 (Summation)**.** Let $\Omega = \mathbb{N}$, $\mathcal{F}$ all subsets of $\mathbb{N}$, and $\mu(E) = |E|$. Then

$$\int f d\mu = \sum_{n \ge 0} f(n)$$

[3] $f^+ = \max(0, f)$ and $f^- = \max(0, -f)$. Also, $\int f := \int f^+ - \int f^-$.

Example 1.15 (Riemann Integral)**.** Let $\Omega = [a, b]$, $\mathcal{B}$ be the Borel σ-algebra and $\mathbf{m}$ Lebesgue measure. Then $\int f d\mathbf{m} = \int_a^b f(x)dx$ if f has discontinuities only on a set of measure 0.

Recall the function f that is equal to 1 on $[0, 1]$ except $\mathbb{Q}$. We simply identify this function same as identically 1, and integral is well defined to be 1 too. Recall that we typically characterize Rimeann integral on continuous functions, whereas now we have a larger class of functions for which in particular allows us to 'ignore' zero measure events.

Example 1.16 (Probability)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. That is, $\mathbb{P}(\Omega) = 1$. Then,

$$\mathbb{E}X := \frac{1}{\mathbb{P}(\Omega)} \int_\Omega X d\mathbb{P}$$

for any random variable (i.e. measurable function) X.

We now list three basic convergence theorems, which forms the backbone of the theory. These convergence theorems with measurable functions allows one to carry out analysis, whereas working on continuous functions requires verifications case by case. We will omit the proofs for the sake of this course.

Theorem 1.17 (**Monotone Convergence Theorem**)**.** *If $f_k \in L^+$ and $f_k \le f_{k+1}$ for all $1 \le k$, then*

$$\lim_{k\to\infty} \int f_k d\mu = \int \lim_{k\to\infty} f_k d\mu$$

Theorem 1.18 (**Fatou's Lemma**)**.** *If $f_k \in L^+$ for all $1 \le k$,*

$$\int \underline{\lim}\, f_k d\mu \le \underline{\lim} \int f_k d\mu$$

Theorem 1.19 (**Dominated Convergence Theorem**)**.** *Suppose $f_k \in L^1$ and*

- $\lim_k f_k = f$ *almost everwhere*
- *there exists $g \in L^1$ such that $|f_k| \le g$ for all k,*

then

$$\lim_k \int f_k d\mu = \int f d\mu$$

Lastly, we will see two more important theorems, simplified considerably.

Theorem 1.20 (**Fubini-Tonelli**)**.** *If $f \in L^+(\Omega \times \Lambda)$ (Tonelli) or $f \in L^1(\Omega \times \Lambda)$ (Fubini), then*

$$\int_{\Omega\times\Lambda} f(x, y) d(\mu \times \nu)(x, y) = \int_\Omega \Big(\int_\Lambda f(x, y) d\nu(y)\Big) d\mu(x) = \int_\Lambda \Big(\int_\Omega f(x, y) d\mu(x)\Big) d\nu(y)$$

where $\mu \times \nu$ is the product measure on $\Omega \times \Lambda$.

Recall Example 1.14, and Fubini-Tonelli allows us to interchange summations.

Theorem 1.21 (**Radon-Nikodym**)**.** *Let μ, ν be σ-finite[4] measures on $(\Omega, \mathcal{F})$ where $\nu(E) = 0$ if $\mu(E) = 0$ (denoted as $\nu \ll \mu$). Then there exists a unique (almost everywhere) integrable function $f : \Omega \to \mathbb{R}$ such that*

$$d\nu = f d\mu, \quad \text{that is,} \quad \nu(E) = \int_E f d\mu$$

[4] That is, Ω is a countable union of sets with finite measures.

1.2 Basics of Probability Theory

In probability theory, measures (with total mass 1) are typically denoted by $\mathbb{P}$, and the integral is denoted by $\mathbb{E}$ or $\mathbb{E}^{\mathbb{P}}$. Measurable space is typically called the event space, and we typically do not model it except for the sake of introductory examples. Measurable functions $\Omega \to \mathbb{R}$ are called random variables (RVs), denoted as X, Y, Z etc. We always implicitly consider the Borel sigma algebra on $\mathbb{R}$. Moreover, we take the change of variable formula as granted:

$$\mathbb{E}[X] = \int_\Omega X d\mathbb{P} = \int_\mathbb{R} x d\mu_X(x)$$

where $\mu_X(A) = \mathbb{P}(X \in A)$ is the *law of* X.

Definition 1.22. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and X be a random variable. We denote the sigma algebra generated by X as $\sigma(X) := \{X^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\}$.

Example 1.23. Consider the event space as $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = 2^\Omega$ and let $X = 1$ on 6 and 0 otherwise. Then $\sigma(X) = \{\emptyset, 6, \{1, 2, 3, 4, 5\}, \Omega\}$.

As X is measurable, $\sigma(X) \subset \mathcal{F}$ but it might be strict as above. Also, if X is constant, $\sigma(X) = \{\emptyset, \Omega\}$. Roughly speaking, $\sigma(X)$ characterizes how much information X yields. Note that, if X takes finitely many values, then $\sigma(X)$ is generated by finitely many sets.

In this course, we will work with square integrable random variables (Hilbert space),

$$L^2 := \Big\{ \text{all RVs } X : \Omega \to \mathbb{R} \text{ s.t. } \mathbb{E}|X|^2 < \infty \Big\}$$

For $X, Y \in L^2$, introduce

$$\begin{aligned}
\operatorname{Var}(X) &:= \mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)^2\Big] = \mathbb{E}[|X|^2] - |\mathbb{E}[X]|^2 \\
\operatorname{Cov}(X, Y) &:= \mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)\big(Y - \mathbb{E}[Y]\big)\Big] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\
\rho(X, Y) &:= \rho_{X,Y} := \frac{\operatorname{Cov}(X, Y)}{\sigma_X \sigma_Y}, \quad \sigma_\cdot := \sqrt{\operatorname{Var}(\cdot)}
\end{aligned}$$

For random vectors $X = (X_1, \cdots, X_d)^\top : \Omega \to \mathbb{R}^d$, we similarly define $\mu_X(A) := \mathbb{P}(X \in A)$ for A in Borel σ-algebra of $\mathbb{R}^d$ and introduce the cumulative distribution function (cdf) as

$$F_X(x) := \mathbb{P}(X_1 \le x_1, \ldots, X_d \le x_d), \;\; x \in \mathbb{R}^d$$

We say random variables $X_1, \ldots, X_n$ are independent by the following equivalent definitions

(i)
$$F_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$$

(ii) $\mathbb{E}\big[\prod_{i=1}^n g_i(X_i)\big] = \prod_{i=1}^n \mathbb{E}\big[g_i(X_i)\big]$ for any bounded scalar Borel measurable functions $g_1, \ldots, g_n$.

Remark 1.24. Independent RVs $X_1, \ldots, X_n$ induces a product measure on Ω^n. Let $n = 2$ and recall Fubini-Tonelli Theorem 1.20 with $f(x, y) = xy$. We get that

$$\mathbb{E}[XY] = \int_{\mathbb{R}\times\mathbb{R}} xy(d\mu_X \times d\mu_Y) = \int_\mathbb{R} \Big(\int_\mathbb{R} xy d\mu_X \Big) d\mu_Y = \int_\mathbb{R} x d\mu_X \int_\mathbb{R} y d\mu_Y = \mathbb{E}[X]\mathbb{E}[Y]$$

Also, if $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, or equivalently, $\operatorname{Cov}(X, Y) = 0$, we say X and Y are uncorrelated.

Definition 1.25. Recall the Radon-Nikodym Theorem 1.21. Suppose the law of X satisfies $\mu_X \ll \mathbf{m}$. Then the Radon-Nikodym derivative f_X is called the density (pdf) of X. In particular,

$$\mu_X((a,b)) = \int_{(a,b)} f d\mathbf{m} = \int_a^b f(x)dx$$

Moreover, existence of density is equivalent to F_X being absolutely continuous. In this case, F_X is almost everywhere differentiable and $\partial_x F_X(x) = f(x)$.

Definition 1.26. Suppose $\{X^n\}_{n\geq 1}$ and X are random variables. We say $X^n \to X$

- almost surely if $\mathbb{P}(\lim_{n\to\infty} X^n = X) = 1$,
- in probability, if $\lim_{n\to\infty} \mathbb{P}(|X^n - X| > \varepsilon) = 0$,
- in distribution, if $\lim_{n\to\infty} F_{X^n}(x) = F_X(x)$ for all x where F_X is continuous at x.
- in L^p, if $\lim_{n\to\infty} \mathbb{E}\big[|X^n - X|^p\big] = 0$ for some $p \geq 1$,
- weakly in L^2, if we are considering $X^n, X \in L^2$, and $\lim_{n\to\infty} \mathbb{E}[X^n \eta] = \mathbb{E}[X\eta]$, $\forall \eta \in L^2$.

Note that $\lim_{n\to\infty} \int f d\mu_{X^n} = \int f d\mu_X$ for all bounded, continuous f is equivalent to convergence in distribution.

Remark 1.27. In this remark, we will try to clarify some differences between these convergences. There are a lot of connections to explore, which we are not aiming to do here. Suppose $\Omega = [0,1]$ with Lebesgue measure $\mathbf{m}$, and consider the following examples:
(i) $X^n(\omega) = n\mathbf{1}_{\{[0,1/n]\}}$,
(ii) $X^n(\omega) = \mathbf{1}_{\{[i/2^k,(i+1)/2^k]\}}$ where $n = 2^k + i$ with $0 \leq i < 2^k$,
(iii) $X^{2n}(\omega) = \omega$, $X^{2n-1}(\omega) = 1 - \omega$, and
(iv) X^n is i.i.d. sequence of uniform distributions with mean 0 and variance 1.

• Now, (i) converges to 0 almost surely (a.s.), however, does not converge in L^1. On the other hand, (ii) converges in L^1 whereas does not converge for any $x \in [0,1]$.

• (iii) obviously converges in distribution to the uniform distribution on $[0,1]$. However, it does not converge in probability.

• (iv) converges to 0 weakly in L^2, whereas it trivially converges to uniform measure in distribution. To see the weak convergence in L^2, note that $\{X^n\}$ is an orthonormal sequence. Given any $\eta \in L^2$, we can write

$$\eta = \sum_{n=1}^{\infty} a_n X^n + \eta^\top, \quad \text{where } a_n := \mathbb{E}[\eta X^n]$$

by Bessel's Inequality (1828), $\sum_{n=1}^{\infty} |a_n|^2 \leq \mathbb{E}[|\eta|^2] < \infty$. In particular, $a_n \to 0$ and this is exactly what we need to conclude X^n converges weakly in L^2.

Exercise 1.28. Prove that, if $X^n \to X$ in L^2, then $X^n \to X$ in probability.
[Hint: Chebyshev's inequality.]

We left reader to recall cdf and pdf of some common distributions:
(i) Bernoulli, (ii) Binomial, (iii) Geometric, (iv) Uniform, (v) Exponential

We say X have normal distribution, denoted as $X \sim \mathcal{N}(\mu, \sigma^2)$, if it has the pdf

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

Check that $\mathbb{E}[X] = \mu$ and $\mathrm{Var}[X] = \sigma^2$. Moreover, $aX + b \sim \mathcal{N}(\mu + b, a^2\sigma^2)$ and if $Y \sim \mathcal{N}(\nu, \rho^2)$ independent of X, then $X + Y \sim \mathcal{N}(\mu + \nu, \sigma^2 + \rho^2)$. We say Z have standard normal distribution if $Z \sim \mathcal{N}(0, 1)$.

We say $X = (X_1, \dots, X_n)^\top$ has a multivariate Gaussian distribution if any linear combination of $X_1, \dots, X_n$ has normal distribution. In particular, if $X_1, \dots, X_n$ are independent and have normal distribution, then X have multivariate Gaussian distribution. Also, if $X_1, \dots, X_n$ have Gaussian distribution, then they are independent if and only if they are pairwise uncorrelated. We say that $Z = (Z_1, \dots, Z_n)$ has a standard Gaussian distribution if $Z_1, \dots, Z_n$ has independent standard normal distribution. Equivalently, we may define $X = (X_1, \dots, X_n)^\top$ has multivariate Gaussian distribution if there exists $m \le n$, a vector $\mu = (\mu_1 \dots, \mu_n)$ and a $n \times m$ matrix A such that $X = \mu + AZ$ where $Z = (Z_1, \dots, Z_m)^\top$ has standard Gaussian distribution. We write $X \sim \mathcal{N}(\nu, \Sigma)$ where

$$\Sigma_{i,j} = \mathrm{Cov}(X_i, X_j) = \mathbb{E}\big[(X_i - \mu_i)(X_j - \mu_j)\big] = A_i^\top \mathbb{E}\big[ZZ^T\big]A_j = A_i^\top A_j$$

that is, $\Sigma = A^\top A$. If the covariance matrix Σ is invertible, then the density of $X \sim \mathcal{N}(\nu, \Sigma)$ is given by

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\Big(-\tfrac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu)\Big)$$

The crucial reason why the normal distribution is fundamental is given by the following theorem.

Theorem 1.29 (**Central Limit Theorem**). *Suppose $\{X_n\}_{n \ge 1}$ are independent and identically distributed (i.i.d.) with $\mathbb{E}[X_n] = \mu$ and $\mathrm{Var}(X_n) = \sigma^2$. Denote the sample mean $\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$ and $Z_n := \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma}$. Then, Z_n converges to $\mathcal{N}(0, 1)$ in distribution.*

Let us also quickly recall the strong Law of Large Numbers (SLLN).

Theorem 1.30 (**Strong Law of Large Numbers**). *Let $X_1, X_2, \cdots$ be pairwise independent identically distributed random variables, where $\mathbb{E}X_1$ exists. [5]. Then the sample mean $\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$ converges to $\mathbb{E}X_1$ almost surely.*

1.3 Conditional Expectation

We have an important result which characterizes the notion of measurability with respect to the σ-algebra of a measurable function:

Theorem 1.31 (**Doob-Dynkin**). *Let X, Y be random variables. Then Y is measurable with respect to $\sigma(X)$ if and only if $Y = h(X)$ for some (Borel) measurable $h : \mathbb{R} \to \mathbb{R}$.*

Definition 1.32 (Conditional Expectation)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $X \in L^1$. Consider a sub σ-algebra $\mathcal{H} \subset \mathcal{F}$. We call $\mathbb{E}[X|\mathcal{H}] \in L^1$ the conditional expectation of X given $\mathcal{H}$, satisfying

- $\mathbb{E}[X|\mathcal{H}]$ measurable with respect to $\mathcal{H}$, and
- $\int_H \mathbb{E}[X|\mathcal{H}] d\mathbb{P} = \int_H X d\mathbb{P}$ for all $H \in \mathcal{H}$.

Furthermore, if Y is an another random variable, we denote $\mathbb{E}[X|Y] := \mathbb{E}[X|\sigma(Y)]$.

[5]That is, $\mathbb{E}X_1^- < \infty$, where $X_1^- = -\min(0, X_1)$

Remark 1.33.

(i) By Doob-Dynkin lemma, $\mathbb{E}[X|Y] = h(Y)$ for some measurable h.

(ii) Equivalent condition to second condition is

$$\mathbb{E}[Y\mathbb{E}[X|\mathcal{H}]] = \mathbb{E}[YX]$$

for all Y measurable with respect to $\mathcal{H}$. If $\mathcal{H} = \sigma(Z)$ for some Z, then we can further write $Y = g(Z)$ and consider all Borel measurable functions g by Doob-Dynkin lemma.

(iii) Conditional expectation exists and unique by the Radon-Nikodym theorem. Namely, $\mu(H) := \int_H X d\mathbb{P}$ is a measure for $(\Omega, \mathcal{H})$, which satisfies $\mu \ll \mathbb{P}_{|\mathcal{H}}$.

(iv) X itself satisfies the second bullet. However, may not be $\mathcal{H}$ measurable.

Example 1.34. If $\mathcal{H} = \{\emptyset, \Omega\}$, then $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]$.

Example 1.35. If X is independent of $\mathcal{H}$, that is,

$$\mathbb{P}(\{X \in B\} \cap H) = \mathbb{P}(X \in B)\mathbb{P}(H) \text{ for all } B \in \mathcal{B}(\mathbb{R}), H \in \mathcal{H}$$

then $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]$. To see this, constanct functions are always measurable. Hence, take $H \in \mathcal{H}$, and then

$$\int_H X d\mathbb{P} = \mathbb{E}[X\mathbf{1}_{\{H\}}] = \mathbb{E}[X]\mathbb{E}[\mathbf{1}_{\{H\}}] = \int_H \mathbb{E}[X|\mathcal{H}] d\mathbb{P}$$

Example 1.36. If X is measurable with respect to $\mathcal{H}$, then $\mathbb{E}[X|\mathcal{H}] = X$.

Example 1.37. Suppose $\Omega_1, \Omega_2, \cdots$ is a partition of Ω, where $\mathbb{P}(\Omega_k) > 0$ for all $1 \le k$. Let $\mathcal{H} = \sigma(\Omega_1, \Omega_2, \cdots)$. Then we claim

$$\mathbb{E}[X|\mathcal{H}] = \frac{\mathbb{E}[X\mathbf{1}_{\{X \in \Omega_k\}}]}{\mathbb{P}(\Omega_k)} = \frac{1}{\mathbb{P}(\Omega_k)} \int_{\Omega_k} X d\mathbb{P} \quad \text{on } \Omega_k$$

To see this, note that $\mathbb{E}[X|\mathcal{H}]$ is constant on each Ω_k. Therefore, it is measurable with respect to $\mathcal{H}$. Then we need to check the second condition, and it suffices to check for $H = \Omega_k$, which is trivial.

Remark 1.38. As it follows from the example above, if Y is a random variable with discrete values, then

$$\mathbb{E}[X|Y = y] = \frac{\mathbb{E}[X\mathbf{1}_{\{Y=y\}}]}{\mathbb{P}(Y = y)}$$

Example 1.39. Consider two independent fair coin flips X_1, X_2. Then

$$\mathbb{E}[X_1|X_1 + X_2] = \begin{cases} \frac{\mathbb{E}[X_1\mathbf{1}_{\{X_1+X_2=0\}}]}{\mathbb{P}(X_1+X_2=0)} & \text{if } X_1 + X_2 = 0 \\ \frac{\mathbb{E}[X_1\mathbf{1}_{\{X_1+X_2=1\}}]}{\mathbb{P}(X_1+X_2=1)} & \text{if } X_1 + X_2 = 1 \\ \frac{\mathbb{E}[X_1\mathbf{1}_{\{X_1+X_2=2\}}]}{\mathbb{P}(X_1+X_2=2)} & \text{if } X_1 + X_2 = 2 \end{cases} = \frac{X_1 + X_2}{2}$$

Intuitively, if $X_1 + X_2$ is 0 or 1, we know the value of X_1. If the sum is 1, we have no information about X_1.

Example 1.40. Consider X, Y with joint density function $f(x, y)$. That is,

$$\mathbb{P}((X, Y) \in B) = \int_B f(x, y) dx dy \text{ for } B \in \mathcal{B}(\mathbb{R}^2)$$

If $\mathbb{E}[|g(X)|] < \infty$, then

$$\mathbb{E}[g(X)|Y] = h(Y), \quad \text{where } h(y) = \frac{1}{\int f(x,y)dx} \int g(x) f(x,y) dx$$

To verify this, since h itself is a measurable function, $h(Y)$ is measurable with respect to $\sigma(Y)$. Now, let $A = \{Y \in B\}$ for some $B \in \mathcal{B}(\mathbb{R})$. Then,

$$\int_A h(Y) d\mathbb{P} = \int h(Y) \mathbf{1}_{\{Y \in B\}} d\mathbb{P} = \int_{\mathbb{R}^2} h(y) \mathbf{1}_{\{y \in B\}} f(x,y) dx dy =$$
$$= \int_{\mathbb{R}} h(y) \mathbf{1}_{\{y \in B\}} \Big(\int_{\mathbb{R}} f(x,y) dx \Big) dy = \int_{\mathbb{R}^2} \mathbf{1}_{\{y \in B\}} g(x) f(x,y) dy = \int_A g(X) d\mathbb{P}$$

Proposition 1.41 (Properties of Conditional Expectation)**.**

(Linear) $\mathbb{E}[aX + Y|\mathcal{H}] = a\mathbb{E}[X|\mathcal{H}] + \mathbb{E}[Y|\mathcal{H}]$.

(Monotone) *If* $X \le Y$ *(almost surely), then* $\mathbb{E}[X|\mathcal{H}] \le \mathbb{E}[Y|\mathcal{H}]$.

(Jensen's inequality) *If* ϕ *is convex,* $\mathbb{E}|X|, \mathbb{E}[|\phi(X)|] < \infty$*, then* $\phi(\mathbb{E}[X|\mathcal{H}]) \le \mathbb{E}[\phi(X)|\mathcal{H}]$.

(Tower property) *If* $\mathcal{H} \subset \mathcal{G}$*, then* $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$.

- *If* $X \in \mathcal{H}$*,* $\mathbb{E}|Y| < \infty$*,* $\mathbb{E}|XY| < \infty$*, then* $\mathbb{E}[XY|\mathcal{H}] = X\mathbb{E}[Y|\mathcal{H}]$.

Remark 1.42. Let's note some particular cases. If $\phi(x) = x^2$, then $(\mathbb{E}[X|\mathcal{H}])^2 \le \mathbb{E}[X^2|\mathcal{H}]$. Since by taking $\mathcal{H} = \{\emptyset, \Omega\}$, we also conclude $(\mathbb{E}[X])^2 \le \mathbb{E}[X^2]$. By the same choice of $\mathcal{H}$, $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$.

Example 1.43 (Random walk)**.** Let ξ_k be i.i.d. random variables with mean μ. Define $Z_n := \sum_{k=1}^n \xi_k$. Then,

$$\mathbb{E}[Z_{n+1}|\xi_1, \ldots, \xi_n] = \mathbb{E}[Z_n + \xi_{n+1}|\xi_1, \cdots, \xi_n] = Z_n + \mu$$

1.4 Stochastic Processes

We now introduce the notion of filtrations, to accomodate stochastic processes.

Definition 1.44 (Filtration)**.** Let $\mathbb{I}$ be either $\mathbb{N}$ or $\mathbb{R}^+$. We say $\mathbb{F} = \{\mathcal{F}_n\}_{n \in \mathbb{I}}$ is a filtration if $\mathcal{F}_k \subset \mathcal{F}_n$ whenever $k \le n$.

Stochastic process is a collection of random variables, indexed by an ordered set $\mathbb{I}$. We will work in continuous time setting with $\mathbb{I} = [0, T]$, and say that stochastic process X is a mapping $[0, T] \times \Omega \to \mathbb{R}$. Instead of viewing a stochastic process as $\{X_t : 0 \le t \le T\}$, it is also typical to view it as family of paths $\{X_\cdot(\omega), \omega \in \Omega\}$.

Example 1.45. Typically, filtration is generated by a stochastic process. Let X be a stochastic process. Then

$$\mathbb{F}^X := \{\mathcal{F}_t^X\}_{t \in [0,T]}, \quad \mathcal{F}_t^X := \sigma(X_s : s \le t)$$

is the filtration generated by X.

Definition 1.46 (Adaptedness)**.** We say a stochastic process X is adapted to the filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t \ge 0}$ if X_t is $\mathcal{F}_t$ measurable.

Remark 1.47.

(i) X is always adapted to its own filtration $\mathbb{F}^X$. Recall the random walk $Z_n = \sum_{k=1}^n \xi_k$. Here, Z is adapted to the filtration formed by $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$.

(ii) We are simplifying the discussion here by considering adapted processes. In fact, one needs to consider progressively measurable processes. We call a process X progressively measurable, if restriction of X onto $[0,t]$ is $\mathcal{B}([0,t]) \times \mathcal{F}_t$-measurable for all t. Similarly, we also omit the discussion around what it means for two process to be equal.

Theorem 1.48 (**Kolmogorov's Extension**). *Let $\mu_{t_1,\dots,t_n}$ be a family of distributions on $\mathbb{R}^n$ satisfying*

$$\mu_{t_1,\dots,t_n}(A_1 \times \cdots \times A_{i-1} \times \mathbb{R} \times A_{i+1} \times \cdots \times A_n) = \mu_{t_1,\dots,t_{i-1},t_{i+1},\dots,t_n}(A_1 \times \cdots \times A_{i-1} \times A_{i+1} \times \cdots \times A_n)$$

for all i and A_i Borel measurable subsets of $\mathbb{R}$. Then, there exists $(\Omega, \mathcal{F}, \mathbb{P})$ and a stochastic process X where joint distribution of $(X_{t_1}, \dots, X_{t_n})$ is given by $\mu_{t_1,\dots,t_n}$.

Theorem 1.49 (**Kolmogorov's Continuity**). *Suppose X is a stochastic process where there exists $\alpha, \beta, C > 0$ such that*

$$\mathbb{E}\big[|X_{t,s}|^\alpha\big] \le C|t-s|^{1+\beta}, \quad \forall s,t \in [0,T] \quad \textit{where} \quad X_{t,s} := X_s - X_t$$

Then, for any $\gamma \in (0, \beta/\alpha)$, $X.(\omega)$ is γ-Hölder continuous almost surely.[6]

Definition 1.50 (Stopping time)**.** We say $\tau : \Omega \to [0,T]$ is a $\mathbb{F}$-stopping time if $\{\tau \le t\} \in \mathcal{F}_t$ for all $t \in [0,T]$. Moreover, we introduce the σ-field corresponding to the stopping time τ as

$$\mathcal{F}_\tau := \Big\{A \subset \Omega : A \cap \{\tau \le t\} \in \mathcal{F}_t, \ \forall 0 \le t \le T\Big\}$$

Intuitively, being measurable with respect to $\mathcal{F}_\tau$ implies that the function is determined by $(\tau, X_{[0,\tau]})$.

Lemma 1.51. *Suppose $A \subset \mathbb{R}^d$ is closed. Then, $\tau = \inf\{t > 0 : X_t \in A\}$ is a $\mathbb{F}^X$ stopping time.*

In case the filtration is generated by a stochastic process X, which is typically the case, τ is a stopping time means we can determine if 'τ ringed before time t' by knowing the path of X from 0 to t (denoted typically as $X_{[0,t]}$).

1.5 Martingales

Now, we are ready to give the definition of martingales. These are, in a rough sense, processes that do not drift deterministically.

Definition 1.52 (Martingale)**.** Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space. We say a stochastic process M_t is a $(\mathbb{F}, \mathbb{P})$-martingale if

- M is adapted to $\mathbb{F}$.
- $\mathbb{E}[|M_t|] < \infty$ for all t.
- $\mathbb{E}[M_t|\mathcal{F}_s] = M_s$ for all $s < t$.

Exercise 1.53. Show that if M_t is a martingale, $\mathbb{E}[M_t] = \mathbb{E}[M_0]$. In the discrete case, we assume $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$. Show that $\mathbb{E}[M_{n+k}|\mathcal{F}_n] = M_n$ for any $1 \le k$.

[6] To be more precise, one needs to say there exists a modification of X that is γ-Hölder continuous almost surely.

Next, we define the sub and super martingales. Roughly, increasing and decreasing processes, similar to the definition of martingale.

Definition 1.54. We say a stochastic process M_t is a $(\mathbb{F}, \mathbb{P})$-submartingale (supermartingale) if

- M is adapted to $\mathbb{F}$.
- $\mathbb{E}|M_t| < \infty$ for all t.
- $\mathbb{E}[M_t|\mathcal{F}_s] \geq (\leq) M_s$ for all $s < t$.

We can construct martingales from a given process. We will handle some particular cases, and for the sake of examples, we will consider discrete time. In the context of random walks, we reserve Z_n for $Z_n := \sum_k^n \xi_k$.

Example 1.55 (Asymmetric simple random walk)**.** Let ξ_i be 1 with probability p and -1 with probability $q = 1 - p$. Then,

$$\mathbb{E}[Z_{n+1}|\mathcal{F}_n^X] = Z_n + p - q,$$

hence if $p - q > 0$, Z_n is a submartingale, and else it is a supermartingale.

Exercise 1.56. Show that $M_n := Z_n - (p - q)n$ is a martingale.

Example 1.57. Define ξ_i as in Example 1.55. Then, $\hat{M}_n := (q/p)^{Z_n}$ is a martingale. It is obviously adapted to the filtration $\mathcal{F}_n^Z$. Next,

$$\mathbb{E}|\hat{M}_n| \leq (q/p)^n + (q/p)^{-n} < \infty$$

and

$$\mathbb{E}[\hat{M}_{n+1}|\mathcal{F}_n^Z] = \mathbb{E}[(q/p)^{Z_n}(q/p)^{\xi_{n+1}}|\mathcal{F}_n^Z] = \hat{M}_n\Big[(q/p)^1 p + (q/p)^{-1} q\Big] = \hat{M}_n$$

Exercise 1.58 (Random walk)**.** Suppose ξ_k's are i.i.d. with mean 0 and variance σ^2. Then Z_n and $Z_n^2 - n\sigma^2$ are both martingales.

Let us note further properties of martingales.

Theorem 1.59.

- *If M_n is a martingale, and ϕ a convex function where $\mathbb{E}|\phi(M_n)| < \infty$, then $\phi(M_n)$ is a submartingale. In particular, if $\mathbb{E}|M_n|^2 < \infty$, then M_n^2 is a submartingale.*
- *If M_n is a submartingale, and ϕ is non-decreasing convex function where $\mathbb{E}|\phi(M_n)| < \infty$, then $\phi(M_n)$ is a submartingale.*
- *If M_n is a martingale where $\mathbb{E}|M_n|^2 < \infty$, then for any $0 \leq \ell \leq k \leq m \leq n$,*

$$\mathbb{E}[(M_n - M_m)M_k] = 0 \text{ and } \mathbb{E}[(M_n - M_m)(M_k - M_\ell)] = 0$$

- *If M_n is a martingale where $\mathbb{E}|M_n|^2 < \infty$, then*

$$M_n^2 - \sum_{k=1}^n \mathbb{E}[(M_k - M_{k-1})^2|\mathcal{F}_{k-1}]$$

is a martingale.

Proof. **Exercise.** ■

1.6 Markov Processes

Suppose X is $\mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}$ adapted process. We say X is a Markov process if, for any $0 \leq s < t \leq T$ and bounded Borel measurable φ, it holds

$$\mathbb{E}\big[\varphi(X_t)|\mathcal{F}_s\big] = \mathbb{E}\big[\varphi(X_t)|X_s\big] \text{ a.s.}$$

Roughly, this means $\{X_t : t \leq s\}$ and $\{X_t : t \geq s\}$ are independent given X_s. By Doob-Dynkin's Lemma 1.31, $\mathbb{E}[\varphi(X_t)|X_s] = h(X_s)$ for some Borel measurable h.

Moreover, we say X is strong Markov process if, for any two stopping time $\tau, \tilde{\tau}$ satisfying $\tau \leq \tilde{\tau}$,

$$\mathbb{E}\big[\varphi(X_{\tilde{\tau}})|\mathcal{F}_\tau\big] = \mathbb{E}\big[\varphi(X_{\tilde{\tau}})|\sigma(\tau, X_\tau)\big] \text{ a.s.}$$

In this case, $\mathbb{E}\big[\varphi(X_{\tilde{\tau}})|\sigma(\tau, X_\tau)\big] = h(\tau, X_\tau)$.

To see the independence, let $B \in \mathcal{B}(\mathbb{R})$, $A \in \mathcal{F}_\tau$ and observe that,

$$\begin{aligned}
\mathbb{P}\big(\{X_{\tilde{\tau}} \in B\} \cap A|\sigma(\tau, X_\tau)\big) &=: \mathbb{E}\big[\mathbf{1}_{\{X_{\tilde{\tau}}\in B\}}\mathbf{1}_{\{A\}}|\sigma(\tau, X_\tau)\big] \\
&= \mathbb{E}\big[\mathbb{E}[\mathbf{1}_{\{X_{\tilde{\tau}}\in B\}}|\mathcal{F}_\tau]\mathbf{1}_{\{A\}}|\sigma(\tau, X_\tau)\big] \\
&= \mathbb{E}\big[\mathbf{1}_{\{X_{\tilde{\tau}}\in B\}}|\sigma(\tau, X_\tau)\big]\mathbb{E}\big[\mathbf{1}_{\{A\}}|\sigma(\tau, X_\tau)\big] \\
&:= \mathbb{P}\big(\mathbf{1}_{\{X_{\tilde{\tau}}\in B\}}|\sigma(\tau, X_\tau)\big)\mathbb{P}\big(\mathbf{1}_{\{A\}}|\sigma(\tau, X_\tau)\big)
\end{aligned}$$

Notable Results and Dates

- **Borel σ-algebra** (1898) – Émile Borel
- **Lebesgue Integration** (1902) – Henri Lebesgue
- **Dominated Convergence Theorem** (1904) – Henri Lebesgue
- **Monotone Convergence Theorem** (1905) – Henri Lebesgue
- **Fubini-Tonelli Theorem** (1907) – Guido Fubini, Leonida Tonelli
- **Fatou's Lemma** (1906) – Pierre Fatou
- **Central Limit Theorem** (1920s) – Aleksandr Lyapunov, Jarl Waldemar Lindeberg, Paul Lévy
- **Banach-Tarski Paradox** (1924) – Stefan Banach, Alfred Tarski
- **Strong Law of Large Numbers** (1929) – Andrey Kolmogorov
- **Radon-Nikodym Theorem** (1930) – Johann Radon, Otto Nikodym
- **Kolmogorov Extension Theorem** (1933) – Andrey Kolmogorov
- **Kolmogorov Continuity Theorem** (1933) – Andrey Kolmogorov
- **Doob-Dynkin Lemma** (1950s) – Joseph L. Doob, Eugene Dynkin
- **Doob's Martingale Convergence Theorem** (1953) – Joseph L. Doob

2 Brownian Motion

As it is the most studied object in mathematics, and plays a crucial role in every part of science, we now define and study some basic properties of the Brownian motion.

Definition 2.1 (**Brownian Motion**). We say B is a (standard) Brownian motion if

- $B_0 = 0$
- For any $t_1 < t_2 < \cdots < t_n$, $B_{t_1} - B_{t_0}, \cdots, B_{t_n} - B_{t_{n-1}}$ are independent.
- $B_t - B_s \sim \mathcal{N}(0, t-s)$, that is normally distributed zero mean, $t-s$ variance for all $s < t$.

By Kolmogorov Extension theorem 1.48, we know that Brownian motion exists. Moreover, by Kolmogorov's Continuity Theorem 1.49, we have the following result:

Proposition 2.2. *For any* $0 < \varepsilon < 1/2$*, standard Brownian motion* B *is* ε*-Hölder continuous almost surely. In particular,* B *is continuous.*

Proof. Since the increments are distributed normally as $B_t - B_s \sim \mathcal{N}(0, t-s)$, it is a standard result that

$$\mathbb{E}[|B_t - B_s|^p] = C_p |t-s|^{\frac{p}{2}}$$

and hence the Kolmogorov's continuity theorem implies B is $\frac{(p/2)-1}{p-1}$-Hölder continuous.[7] Let $p \to \infty$ to approach $1/2$. ■

Brownian motion is a Markov process [EXERCISE], and

$$\mathbb{E}[B_t|\mathcal{F}_s^B] = \mathbb{E}[B_t|\sigma(B_s)] = \mathbb{E}[B_t - B_s|\sigma(B_s)] + B_s = B_s$$

hence a martingale with respect to its own filtration.[8] Note that for any $0 = t_0 < t_1 < \cdots < t_n$,

$$(B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}})$$

are independent and distributed normally. Hence, it is a Gaussian distribution. As we can write

$$(B_{t_1}, \ldots, B_{t_n}) = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} B_{t_1} - B_{t_0} \\ \vdots \\ B_{t_n} - B_{t_{n-1}} \end{pmatrix}$$

which is then obvious that $(B_{t_1}, \ldots, B_{t_n})$ has Gaussian distribution. Such processes with finite distributions being Gaussian distribution are called Gaussian processes. We then have an equivalent definition of Brownian motion:

Definition 2.3. We say B_t is a (standard) Brownian motion if

- B_t is a Gaussian process,
- For any $t < s$, $\mathbb{E}[B_t] = 0$ and $\mathbb{E}[B_t B_s] = t$,

[7] by considering a modification if necessary, but I avoid that discussion completely.

[8] We do not need to use Markov property.

Proposition 2.4. *Suppose B is a Brownian motion. Then,*

(Translation Invariance) *for any $s > 0$, the process $B^s_t := B_{s+t} - B_s$,*

(Scale Invariance) *for any $c > 0$, $\hat{B}^c_t := \frac{1}{\sqrt{c}} B_{ct}$*

(Time Invariance) $\tilde{B}_t := t B_{1/t}$,

are all Brownian motions.

Exercise 2.5. Prove Proposition 2.4 using the definition 2.3.

Theorem 2.6 (Donsker's Invariance). *Let $\xi_1, \xi_2, \ldots$ be i.i.d. random variables with expected 0 and variance σ^2. Define*

$$\hat{Z}^n_t := \frac{1}{\sqrt{n\sigma^2}}\big(Z_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor)\xi_{\lfloor nt \rfloor + 1}\big), \quad Z_n := \sum_{i=1}^{n} \xi_i$$

Then, $\hat{Z}^n_t$ converges in distribution to the standard Brownian motion.

Note that when $t = 1$, this is essentially the CLT. Importantly, we now have a very easy method to construct approximate Brownian motions.

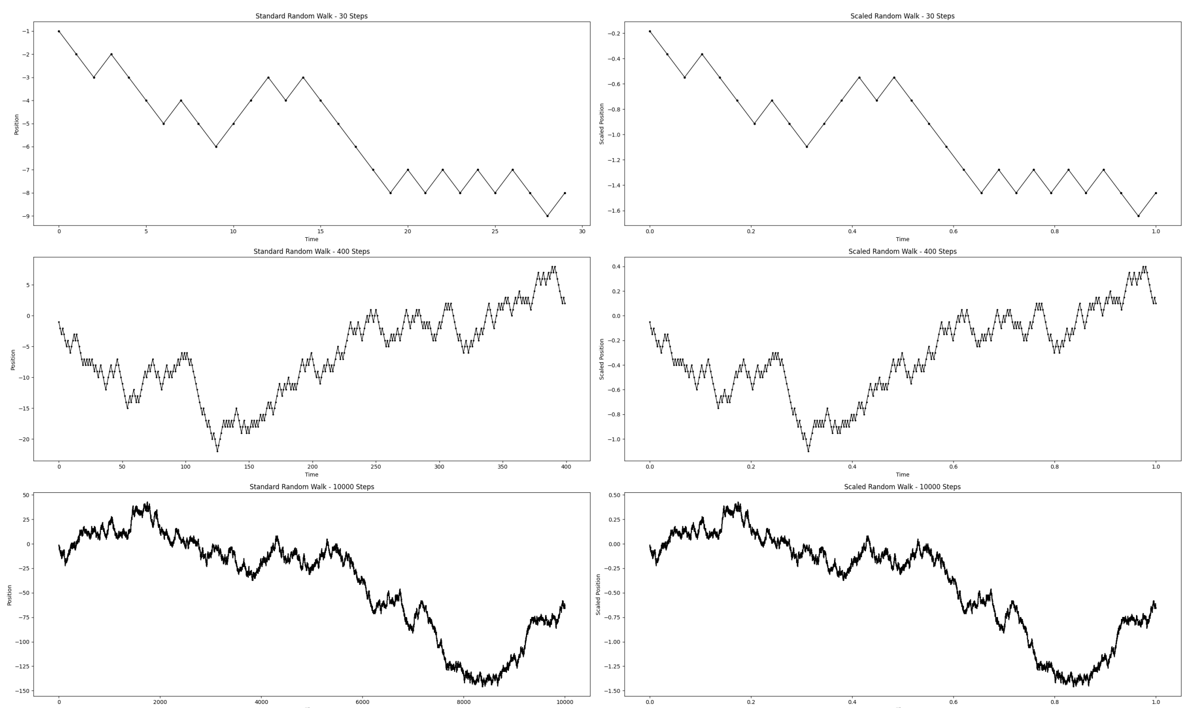


Figure 1: Visual convergence of random walk to Brownian motion. See the Functional CLT 2.6.

Let us discuss an important observation about the σ-algebra $\mathcal{F}^+_0 := \cap_{\varepsilon > 0} \mathcal{F}^B_\varepsilon$, called germ field, which represents limits of functions on the paths $\{B_s : 0 \le s \le \varepsilon\}$. Consider a φ which is measurable with respect to $\mathcal{F}^+_0$. In particular, it is measurable with respect to $\mathcal{F}^B_\varepsilon$ for some small $\varepsilon > 0$. Then, by Doob-Dynkin Lemma,

$$\varphi(\omega) = h(B_{[0,\varepsilon]}(\omega))$$

for some measurable h. Here $_{[0,\varepsilon]}$ represents the path. However, since it has to be measurable with respect to $s < \varepsilon$, it most hold that

$$h\big(B_{[0,s]}(\omega) +_s B_{(s,\varepsilon]}(\omega)\big) = h\big(B_{[0,s]}(\tilde{\omega}) +_s B_{(s,\varepsilon]}(\tilde{\omega})\big), \quad \text{whenever} \quad B_{[0,s]}(\omega) = B_{[0,s]}(\tilde{\omega})$$

Otherwise, it would contradict with $\varphi \in \mathcal{F}_s^B$.[9] That is, h is in fact independent of $B_{(s,\varepsilon]}$ and one can well define

$$\varphi = h(B_{[0,s]})$$

Because s is arbitrary, only $\lim_{s \to 0^+} B_s$ might matter, if it was not by continuity equal to 0. We deduce that

$$\varphi = h(0)$$

and hence a constant. To sum up, we have shown that, if $\varphi \in \mathcal{F}_0^+$, then it is a constant. This yields an important result:

Theorem 2.7 (**Blumenthal's 0-1 Law**). *Let $A \in \mathcal{F}_0^+$. Then either $\mathbb{P}(A) = 1$ or $\mathbb{P}(A) = 0$. In other words, $\mathcal{F}_0^+$ is almost the trivial σ-algebra.*

Proof. For a given $A \in \mathcal{F}_0^+$, consider the random variable $\mathbf{1}_{\{A\}}$. Since it is measurable, it has to be constant. Therefore, it is either 1 or 0 almost surely. ■

Altought at a first glance the result seems to carry only a little information, one notices that it yields important informations about the local behaviour of the process. We now explore some immediate consequences.

Theorem 2.8. *Let $\tau = \inf\{t \geq 0 : B_t > 0\}$, then $\mathbb{P}(\tau = 0) = 1$.*

Proof. For any t, $\mathbb{P}(\tau \leq t) \geq \mathbb{P}(B_t > 0) = 1/2$. By continuity of measures,

$$\mathbb{P}(\tau = 0) = \lim_{t \downarrow 0} \mathbb{P}(\tau \leq t) \geq 1/2$$

On the other hand, $\{\tau = 0\} \in \mathcal{F}_0^+$ and Blumenthal's 0-1 Law 2.7 concludes the argument. ■

As the Brownian motion is symmetric and continuous, immediate follow up result is:

Theorem 2.9. *If $\tau = \inf\{t > 0 : B_t = 0\}$, then $\mathbb{P}(\tau = 0) = 1$.*

In particular, B_t has infinitely many 0's in $[0, \varepsilon]$ for arbitrary $\varepsilon > 0$ almost surely. One can continue to study local behaviour, but we only state one important result as a remark and move on.

Remark 2.10. Law of Iterated Logarithms:

$$\overline{\lim_{\delta \to 0^+}} \frac{B_{t+\delta} - B_t}{\sqrt{2\delta \log\log(1/\delta)}} = 1, \quad \text{and} \quad \underline{\lim_{\delta \to 0^+}} \frac{B_{t+\delta} - B_t}{\sqrt{2\delta \log\log(1/\delta)}} = -1$$

In particular, B is not $\frac{1}{2}$-Hölder continuous and hence nowhere differentiable too.

[9] Here, I am hiding an important step for the sake of explanation. To claim that $\varphi = h(B_{[0,s]})$, we need the fact that Brownian motion has independent increments. More formally, let $\varphi = h(B_{[0,s]}, "B_{[s,\varepsilon]} - B_s")$. Then, since $\varphi \in \mathcal{F}_s$,

$$h(B_{[0,s]}, "B_{[s,\varepsilon]} - B_s") = \mathbb{E}[\varphi|\mathcal{F}_s] = \tilde{h}(B_{[0,s]})$$

a.s. for distribution on $B_{[0,s]}$, and independently almost surely for "$B_{[s,\varepsilon]} - B_s$". (Making it fully rigorous needs more notations)

Next, we will use the symmetry of Brownian motion to show the so called reflection principle. It is an important result to deduce facts about the supremum of the Brownian motion. In particular for finance, this is used for pricing Barrier options.

Theorem 2.11 (**Reflection Principle**).

$$\mathbb{P}\Big(\sup_{0\le s\le t} B_s \ge a\Big) = 2\mathbb{P}\Big(B_t \ge a\Big)$$

Proof. We will prove something slightly more general for $b \le a$:

$$\mathbb{P}\Big(\sup_{0\le s\le t} B_s \ge a, B_t \le b\Big) = \mathbb{P}\Big(B_t \ge 2a-b\Big)$$

given this, it is straightforward to finish the proof as

$$\mathbb{P}\Big(\sup_{0\le s\le t} B_s \ge a\Big) = \mathbb{P}\Big(\sup_{0\le s\le t} B_s \ge a, B_t \le a\Big) + \mathbb{P}\Big(\sup_{0\le s\le t} B_s \ge a, B_t \ge a\Big) = 2\mathbb{P}\Big(B_t \ge a\Big)$$

Now, we will see how useful the stopping times are. Recall lemma 1.51. Define

$$\tau_a = \inf\{t > 0 : B_t \ge a\}$$

and then by the Strong Markov Property, the process

$$\tilde{B}_s := B_{s+\tau_a} - B_{\tau_a}$$

is also a Brownian motion. Then,

$$\begin{aligned}
\mathbb{P}\Big(\sup_{0\le s\le t} B_s \ge a, B_t \le b\Big) &= \mathbb{P}\Big(\tau_a \le t, \tilde{B}_{t-\tau_a} \le b-a\Big) \\
(\text{Symmetry}) &= \mathbb{P}\Big(\tau_a \le t, -\tilde{B}_{t-\tau_a} \le b-a\Big) \\
&= \mathbb{P}\Big(\tau_a \le t, a+\tilde{B}_{t-\tau_a} \ge 2a-b\Big) \\
(2a-b\ge a) \ &= \mathbb{P}\Big(\tau_a \le t, B_t \ge 2a-b\Big) = \mathbb{P}\Big(B_t \ge 2a-b\Big)
\end{aligned}$$

■

Before continuing to define the integrals with respect to the Brownian motion, we first study the total variation and quadratic variation. Let $\pi := 0 = t_0 < t_1 < \cdots < t_n = T$ be a partition of $[0,T]$ and set $|\pi| := \max_i t_i - t_{i-1}$. Then, we define the total variation of a process X in $0 \le a < b \le T$ as

$$\bigvee_a^b(X) := \sup_\pi \sum_{i=1}^n |X_{t_i|_a^b} - X_{t_{i-1}|_a^b}|, \quad \text{where} \quad t|_a^b := a \vee t \wedge b$$

Furthermore, we define the quadratic variation of X, a process in $\mathbb{R}^{d\times d}$ as

$$\langle X\rangle_t := \lim_{|\pi|\to 0} \sum_{i=1}^n \big(X_{t_i|^t} - X_{t_{i-1}|^t}\big)\big(X_{t_i|^t} - X_{t_{i-1}|^t}\big)^T$$

if this limit exists in the sense of convergence in probability. Note that, we do not define in a pathwise manner, or say for each $\omega \in \Omega$. Next, we compute the quadratic variation of the Brownian motion:

Proposition 2.12. *Suppose $B, \tilde{B}$ are two independent standard Brownian motions. Then,*

$$\lim_{|\pi|\to 0} \mathbb{E}\Big[\Big(\sum_{i=1}^{n}\big(B_{t_i|^t} - B_{t_{i-1}|^t}\big)^2 - t\Big)^2\Big] = 0 \text{ and } \lim_{|\pi|\to 0} \mathbb{E}\Big[\Big(\sum_{i=1}^{n}\big(B_{t_i|^t} - B_{t_{i-1}|^t}\big)\big(\tilde{B}_{t_i|^t} - \tilde{B}_{t_{i-1}|^t}\big)^T\Big)^2\Big] = 0$$

Consequently, if B is a d-dimensional standard Brownian motion[10]*, then $\langle B\rangle_t = tI_d$ where I_d is the d-dimensional identity matrix.*

Proof. First of all, by noting the elements of $\langle B\rangle_t$, it follows $\langle B\rangle_t = tI_d$ since L^2 convergence implies the convergence in probability. We will only prove the first claim, as the second one follows similarly. Define

$$\Delta t_i := t_i - t_{t_{i-1}}, \quad \eta_i := |B_{t_i} - B_{t_{i-1}}|^2 - \Delta t_i$$

Note that η_i's are independent and $\mathbb{E}[\eta_i] = 0$. Then,

$$\text{Var}(\eta_i) = \text{Var}(|B_{t_i} - B_{t_{i-1}}|^2) = \mathbb{E}\big[|B_{t_i} - B_{t_{i-1}}|^4\big] - \Big(\mathbb{E}\big[|B_{t_i} - B_{t_{i-1}}|^2\big]\Big)^2 = 3\Delta t_i^2 - \Delta t_i^2$$

Therefore, we can conclude

$$\begin{aligned}\mathbb{E}\Big[\Big(\sum_{i=1}^{n}\big|B_{t_i|^t} - B_{t_{i-1}|^t}\big|^2 - t\Big)^2\Big] &= \mathbb{E}\Big[\Big(\sum_{i=1}^{n}\eta_i\Big)^2\Big] = \text{Var}\Big(\sum_{i=1}^{n}\eta_i\Big)\\ &= \sum_{i=1}^{n}\text{Var}\left(\eta_i\right) = 2\sum_{i=1}^{n}(\Delta t_i)^2 \le 2|\pi|\sum_{i=1}^{n}\Delta t_i = 2|\pi|T \to 0\end{aligned}$$

and hence we proved the result. ■

Corollary 2.13. $\bigvee_a^b(B) = \infty$ *almost surely.*

Proof. Observe that, for any partition π,

$$\sum_{i=1}^{n}\big|B_{t_i|_a^b} - B_{t_{i-1}|_a^b}\big|^2 \le \Big(\max_i |B_{t_i|_a^b} - B_{t_{i-1}|_a^b}|\Big)\sum_{i=1}^{n}\big|B_{t_i|_a^b} - B_{t_{i-1}|_a^b}\big| \le \Big(\max_i |B_{t_i|_a^b} - B_{t_{i-1}|_a^b}|\Big)\bigvee_a^b(B)$$

Since we know that Brownian motion is continuous, $\max_i |B_{t_i|_a^b} - B_{t_{i-1}|_a^b}| \to 0$ almost surely. On the other hand, we have seen that $\sum_{i=1}^n \big|B_{t_i|_a^b} - B_{t_{i-1}|_a^b}\big|^2 \to b - a$ which then clearly implies the result. By using monotonicity of total variation, one can in fact show that the result holds for all $0 \le a < b \le T$ almost surely. ■

Notable Results and Dates

- **Brownian Motion** (1827) – Robert Brown
- **Brownian Motion** (1900) – Louis Bachelier
- **Brownian Motion** (1905) – Albert Einstein, Marian Smoluchowski
- **Wiener Process** (1923) – Norbert Wiener
- **Levy Processes** (1930s) – Paul Lévy
- **Blumenthal's 0-1 Law** (1957) – Robert McCallum Blumenthal
- **Donsker's Invariance Principle** (1951) – Monroe D. Donsker

[10] That is, each component is an independent standard Brownian motion.

3 Stochastic Calculus

Fix a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ together with $\mathbb{F}$-Brownian motion.[11] In this section, we will formalize the meaning of $\int \sigma dB_t$ and introduce the Itô's formula. First, we need to endow our space of processes with L^2 norm, given as

$$\|X\|_2^2 := \mathbb{E}\Big[\int_0^T |X_t|^2 dt\Big]$$

which is then a complete metric space. [12] Recall that we defined the Lebesgue integral with simple functions, which were essentially piecewise constant functions. Similarly, we will first define the stochastic integral for simple processes which are $\mathbb{F}$-adapted processes that are piecewise constant in time. That is, σ is a simple process if $\sigma_t \in \mathcal{F}_t$ and there exists a partition $0 = t_0 < \cdots < t_n = T$ for which $\sigma_t = \sigma_{t_i}$ for all $t \in [t_i, t_{i+1})$. Now, define the stochastic integral as

$$\int_0^t \sigma_s dB_s := \sum_{i=1}^n \sigma_{t_{i-1}}(B_{t_i|^t} - B_{t_{i-1}|^t}), \quad \forall 0 \le t \le T$$

Lemma 3.1. *Suppose σ is a simple process with $\|\sigma\|_2 < \infty$. Denote $M_t := \int_0^t \sigma_s dB_s$. Then,*

(i) M is an $\mathbb{F}$-martingale. In particular $\mathbb{E}[M_t] = 0$.

(ii) $\|M\|_2 < \infty$ and $\hat{M}_t := M_t^2 - \int_0^t |\sigma_s|^2 ds$ is a martingale. In particular, we have the Itô isometry:

$$\mathbb{E}\big[|M_t|^2\big] = \mathbb{E}\Big[\int_0^t |\sigma_s|^2 ds\Big] \tag{3.1}$$

(iii) M is continuous almost surely.

Proof. (i): By the tower property of conditional expectations, it suffices to show that

$$\mathbb{E}[M_{t_i}|\mathcal{F}_t] = M_t, \quad t_{i-1} \le t \le t_i$$

because one can easily extend it to any $t \le t_i$ and $\mathbb{E}[M_s|\mathcal{F}_t] = \mathbb{E}[\mathbb{E}[M_{t_i}|\mathcal{F}_s]|\mathcal{F}_t]$ $t \le s \le t_i$. Now, above follows by a straightforward observation

$$\mathbb{E}[M_{t_i} - M_t|\mathcal{F}_t] = \mathbb{E}[\sigma_{t_{i-1}}(B_{t_i} - B_t)|\mathcal{F}_t] = \sigma_{t_{i-1}}\mathbb{E}[(B_{t_i} - B_t)|\mathcal{F}_t] = 0$$

(ii): It suffices to show that $\hat{M}$ is a martingale to argue L^2 integrability. First,

$$\hat{M}_{t_i} - \hat{M}_t = M_{t_i}^2 - M_t^2 - |\sigma_{t_{i-1}}|^2(t_i - t) = (M_{t_i} - M_t)^2 + 2M_t(M_{t_i} - M_t) - |\sigma_{t_{i-1}}|^2(t_i - t)$$

and now similar to (i),

$$\mathbb{E}\big[\hat{M}_{t_i} - \hat{M}_t|\mathcal{F}_t\big] = \mathbb{E}\big[(M_{t_i} - M_t)^2|\mathcal{F}_t\big] + 2M_t\mathbb{E}\big[(M_{t_i} - M_t)|\mathcal{F}_t\big] - |\sigma_{t_{i-1}}|^2(t_i - t) = 0$$

(iii) is obvious by the definition of M_t. ■

We will omit the proof of the next two lemmas for this course. First one is a crucial estimate that allows us to keep the stochastic integrals staying in L^2 space. Second one states that simple processes are dense in L^2 and hence allows us to extend the stochastic integral from simple processes to L^2 processes.

[11] We call B an $\mathbb{F}$-Brownian motion, if B is adapted to the filtration $\mathbb{F}$ and $(B_t - B_s)$ is independent of $\mathcal{F}_s$.

[12] In fact, it can be viewed as a closed subspace of $L^2([0, T] \times \Omega)$, but we will not discuss much. See Karatsaz&Shreve for a rigorous construction of stochastic integrals.

Lemma 3.2 (Doob's Maximum Inequality). *Suppose σ is a simple process with $\|\sigma\|_2 < \infty$ and let $M_t := \int_0^t \sigma_s dB_s$. Then,*

$$\mathbb{E}\big[|M_T|^2\big] \le \mathbb{E}\big[\sup_{0\le t\le T} |M_t|^2\big] \le 4\mathbb{E}\big[|M_T|^2\big]$$

Lemma 3.3. *For any process σ with $\|\sigma\|_2 < \infty$, there exists a sequence of simple processes σ^n with $\|\sigma^n\|_2 < \infty$ such that $\|\sigma - \sigma^n\|_2 \to 0$.*

Let us remark that, for σ continuous and bounded, it is easy to construct an approximating simple processes as

$$\sigma_t^n := \sum_{i=1}^n \sigma_{t_{i-1}} \mathbf{1}_{\{[t_{i-1},t_i)\}}, \quad \text{where } t_i := i\frac{T}{n},\ i = 0, \cdots, n$$

Finally, we collected all the results needed to introduce the stochastic integral for any L^2 process σ. Introduce M^n corresponding to σ^n, by Doob's maximum inequality and Itô isometry,

$$\mathbb{E}\big[\sup_{0\le t\le T} |M_t^n - M_t^m|^2\big] \le 4\mathbb{E}\Big[\Big|\int_0^T (\sigma_t^n - \sigma_t^m) dB_t\Big|^2\Big] = 4\mathbb{E}\Big[\int_0^T |\sigma_t^n - \sigma_t^m|^2 ds\Big] = 4\|\sigma^n - \sigma^m\|_2$$

and since σ^n approximates σ, $\|\sigma^n - \sigma^m\|_2 \to 0$ as $m, n \to \infty$. Since the space is complete, Cauchy sequence converges to some limit M. If $\tilde{\sigma}^n$ was another approximating sequence, by the similar observation,

$$\mathbb{E}\big[\sup_{0\le t\le T} |M_t^n - \tilde{M}_t^n|^2\big] \le 4\|\sigma^n - \tilde{\sigma}^n\|_2 \to 0$$

and hence the limit is independent of the approximating sequence. Therefore, we define

$$\int_0^t \sigma_s dB_s := \lim_{n\to\infty} \int_0^t \sigma_s^n dB_s \quad \text{where the convergence is in the sense:}$$

$$\lim_{n\to\infty} \mathbb{E}\Big[\sup_{0\le t\le T} \Big|\int_0^t \sigma_s dB_s - \int_0^t \sigma_s^n dB_s\Big|^2\Big] = 0$$

Remark 3.4. (i): Note that the convergence is not in a pathwise manner. That is, we cannot determine $\big(\int_0^t \sigma_s dB_s\big)(\omega)$.

(ii): One can further extend the integrals to locally L^2 processes σ, i.e. without requiring finite expectations, by using stopping times. In this case, $M_t := \int_0^t \sigma_s dB_s$ is a local martingale.

The following theorem holds due to the uniform convergence that we have in the definition of the stochastic integral.

Theorem 3.5. *Suppose σ is an L^2 process, i.e. $\|\sigma\|_2 < \infty$. Denote $M_t := \int_0^t \sigma dB_s$. Then all the results in Lemma 3.1 holds true.*

3.1 Itô's Formula

We will prove the Itô's formula, which can be viewed as the fundamental theorem of stochastic calculus. First, let us observe why the usual chain rule type of result does not work in the case of stochastic analysis.

Example 3.6. Let B be the standard Brownian motion. Then,

$$|B_t|^2 - t = 2\int_0^t B_s dB_s$$

Consequently, considering $f(x) = x^2$, it does not hold that $df(B_t) = f'(B_t)dB_t$.

Proof. We will argue for $t = T$ for notational simplicity. Let $\pi : 0 = t_0 < \cdots < t_n = T$ be a partition of $[0, T]$, and write

$$|B_T|^2 = \sum_{i=1}^{n} |B_{t_i}|^2 - |B_{t_{i-1}}|^2 = \sum_{i=1}^{n} |B_{t_i} - B_{t_{i-1}}|^2 - 2B_{t_{i-1}}(B_{t_i} - B_{t_{i-1}})$$

We know that as $|\pi| \to 0$, $\sum_{i=1}^{n} |B_{t_i} - B_{t_{i-1}}|^2 \to T$ in L^2. On the other hand, one can check that $\sum_i B_{t_{i-1}} \mathbf{1}_{\{t \in [t_{i-1}, t_i)\}}$ is a simple process approximating the Brownian motion. Then, by definition of stochastic integral we conclude the result. ■

Now, suppose b is L^1 and σ is L^2 processes adapted to the filtration [13]. Let

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dB_s, \quad \text{and} \quad \langle X \rangle_t := \int_0^t |\sigma_s|^2 ds$$

Theorem 3.7 (Itô Formula)**.** *Supose $f \in C^{1,2}([0, T] \times \mathbb{R}; \mathbb{R})$. Then*

$$\begin{aligned} df(t, X_t) &= \partial_t f(t, X_t)dt + \partial_x f(t, X_t)dX_t + \tfrac{1}{2}\partial_{xx} f(t, X_t)d\langle X \rangle_t \\ &= \big[\partial_t f + b_t \partial_x f + |\sigma_t|^2 \tfrac{1}{2}\partial_{xx} f\big](t, X_t)dt + \sigma_t \partial_x f(t, X_t)dB_t \end{aligned}$$

which means,

$$f(t, X_t) = f(0, X_0) + \int_0^t \big[\partial_t f + b_t \partial_x f + |\sigma_t|^2 \tfrac{1}{2}\partial_{xx} f\big](t, X_t)dt + \int_0^t \sigma_t \partial_x f(t, X_t)dB_t \qquad (3.2)$$

Proof. We will proceed the proof only for the simple case. We will assume that b, σ are simple processes and all the derivatives of f are bounded.

Because we assume that the derivatives of f are continuous, on a compact set, their supremum is finite. That is, they are locally integrable and therefore integrals in (3.2) are well defined. We will assume $t = T$ again. As a first step, assume b and σ are bounded and $\mathcal{F}_0$ measurable. Take a partition π of $[0, T]$ and write

$$f(T, X_T) - f(0, X_0) = \sum_{i=1}^{n} f(t_i, X_{t_i}) - f(t_{i-1}, X_{t_{i-1}})$$

denote $\Delta t_i := t_i - t_{i-1}$, $\Delta B_{t_i} := B_{t_i} - B_{t_{i-1}}$ and note that $\Delta X_{t_i} := X_{t_i} - X_{t_{i-1}} = b_0 \Delta t_i + \sigma_0 \Delta B_{t_i}$. By Taylor expansion,

$$\begin{aligned} &f(t_i, X_{t_i}) - f(t_{i-1}, X_{t_{i-1}}) = f(t_{i-1} + \Delta t_i, X_{t_{i-1}+\Delta t_i}) - f(t_{i-1}, X_{t_{i-1}}) \\ &= \partial_t f(t_{i-1}, X_{t_{i-1}})\Delta t_i + \partial_x f(t_{i-1}, X_{t_{i-1}})\Delta X_{t_i} \\ &\quad + \frac{1}{2}\partial_{tt} f(t_{i-1}, X_{t_{i-1}})|\Delta t_i|^2 + \partial_{tx} f(t_{i-1}, X_{t_{i-1}})\Delta t_i \Delta X_{t_i} \\ &\quad + \frac{1}{2}\partial_{xx} f(t_{i-1}, X_{t_{i-1}})|\Delta X_{t_i}|^2 + R_i^\pi \\ &= \left[\partial_t f + b_0 \partial_x f + \frac{1}{2}\partial_{xx} f |\sigma_0|^2\right](t_{i-1}, X_{t_{i-1}})\, \Delta t_i + \sigma_0 \partial_x f(t_{i-1}, X_{t_{i-1}})\, \Delta B_{t_i} \\ &\quad + \frac{1}{2}\partial_{xx} f(t_{i-1}, X_{t_{i-1}})|\sigma_0|^2 \big[|\Delta B_{t_i}|^2 - \Delta t_i\big] + I_i^\pi \end{aligned}$$

[13] We are ignoring the discussion around adapted vs progressively measurable processes. Also, in general locally integrable processes are sufficient to argue the Itô formula.

where

$$I_i^\pi := \frac{1}{2}\left[\partial_{tt} f + \partial_{tx} f b_0 + \partial_{xx} f |b_0|^2\right](t_{i-1}, X_{t_{i-1}})|\Delta t_i|^2$$
$$+ \left[\partial_{tx} f + b_0\sigma_0\partial_{xx} f\right](t_{i-1}, X_{t_{i-1}})\Delta t_i \Delta B_{t_i} + R_i^\pi,$$
$$\left|R_i^\pi\right| \le C\left[|\Delta t_i|^3 + |\Delta X_{t_i}|^3\right] \le C\left[|\Delta t_i|^3 + |\Delta B_{t_i}|^3\right]$$

Now, we let $|\pi| \to 0$. Send $|\pi| \to 0$. By the Dominated Convergence Theorem,

$$\sum_{i=1}^n \left[\partial_t f + b_0\partial_x f + \frac{1}{2}\partial_{xx} f |\sigma_0|^2\right](t_{i-1}, X_{t_{i-1}})\Delta t_i$$
$$\to \int_0^T \left[\partial_t f + b_0\partial_x f + \frac{1}{2}|\sigma_0|^2\partial_{xx} f\right](t, X_t)dt \quad \text{in } L^2$$

Again by Dominated Convergence Theorem,

$$\mathbb{E}\left[\sum_{i=1}^n \int_{t_{i-1}}^{t_i} |\partial_x f(t, X_t) - \partial_x f(t_{i-1}, X_{t_{i-1}})|^2 dt\right] \to 0$$

and thus, by Itô's Isometry,

$$\sum_{i=1}^n \sigma_0\partial_x f(t_{i-1}, X_{t_{i-1}})\Delta B_{t_i} \to \int_0^T \sigma_0\partial_x f(t, X_t)dB_t \quad \text{in } L^2.$$

This concludes all the terms we need. Now, we need to show that the rest of the terms converge to 0 in L^2 sense. By Example 3.6, we see that

$$|\Delta B_{t_i}|^2 - \Delta t_i = 2\int_{t_{i-1}}^{t_i} (B_t - B_{t_{i-1}})dB_t$$

Then

$$\sum_{i=1}^n \partial_{xx} f(t_{i-1}, X_{t_{i-1}})\left[|\Delta B_{t_i}|^2 - \Delta t_i\right] = 2\sum_{i=1}^n \partial_{xx} f(t_{i-1}, X_{t_{i-1}})\int_{t_{i-1}}^{t_i} (B_t - B_{t_{i-1}})dB_t$$
$$= 2\int_0^T \sum_{i=1}^n \partial_{xx} f(t_{i-1}, X_{t_{i-1}})(B_t - B_{t_{i-1}})\mathbf{1}_{\{[t_{i-1}, t_i)\}}dB_t$$

To show that this converges to 0 in L^2, by Itô Isomery,

$$\mathbb{E}\left[\int_0^T \left|\sum_{i=1}^n \partial_{xx} f(t_{i-1}, X_{t_{i-1}})(B_t - B_{t_{i-1}})\mathbf{1}_{\{[t_{i-1}, t_i)\}}\right|^2 dt\right]$$
$$= \mathbb{E}\left[\sum_{i=1}^n \int_{t_{i-1}}^{t_i} \left|\partial_{xx} f(t_{i-1}, X_{t_{i-1}})(B_t - B_{t_{i-1}})\right|^2 dt\right]$$
$$\le C\sum_{i=1}^n \int_{t_{i-1}}^{t_i} (t - t_{i-1})dt = C\sum_{i=1}^n |\Delta t_i|^2 \le C|\pi| \to 0$$

Lastly, we need to control I^{π}. Recall that $\mathbb{E}[|\Delta B_{t_i}|^p] = C_p|\Delta t_i|^{\frac{p}{2}}$. Then,

$$\mathbb{E}\left[\Big|\sum_{i=1}^{n} I_i^{\pi}\Big|^2\right] \le \mathbb{E}\left[n\sum_{i=1}^{n}|I_i^{\pi}|^2\right]$$
$$\le C\mathbb{E}\left[\frac{1}{|\pi|}\sum_{i=1}^{n}|\Delta t_i|^4 + |\Delta t_i|^2|\Delta B_{t_i}|^2 + |\Delta B_{t_i}|^6\right] \le C|\pi| \to 0$$

Note that, it is straightforward to extend to simple processes b and σ, because we can apply the above result on $[t_{i-1}, t_i)$ to get

$$f(t_i, X_{t_i}) - f(t_{i-1}, X_{t_{i-1}}) = \int_{t_{i-1}}^{t_i}\left[\partial_t f + b_t\partial_x f + |\sigma_t|^2\tfrac{1}{2}\partial_{xx} f\right](t, X_t)dt + \int_{t_{i-1}}^{t_i}\sigma_t\partial_x f(t, X_t)dB_t$$

and sum it up. ■

3.1.1 Multidimensional Itô Formula

Let $B = (B^1, \cdots, B^d)^\top$ be a d-dimensional $\mathbb{F}$-Brownian Motion, b^i's are L^1, σ^{ij}'s are L^2 processes adapted to the filtration. Set $b := (b^1, \cdots, b^{d_1})^\top$ and $\sigma := (\sigma^{ij})_{1\le i\le d_1, 1\le j\le d}$ which take values in $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_1\times d}$, respectively. Let $X = (X^1, \cdots, X^{d_1})^\top$ satisfy

$$dX_t^i := b_t^i dt + \sum_{j=1}^{d}\sigma_t^{ij}dB_t^j, \quad i = 1, \cdots, d_1; \quad \text{or equivalently,} \quad dX_t = b_t dt + \sigma_t dB_t$$

Theorem 3.8. *Assume $f \in C^{1,2}\big([0, T] \times \mathbb{R}^{d_1}; \mathbb{R}\big)$. Then*

$$\begin{aligned} &df(t, X_t)\\ &= \left[\partial_t f + (\partial_x f)^\top b_t + \frac{1}{2}\partial_{xx} f : (\sigma_t\sigma_t^\top)\right](t, X_t)dt + (\partial_x f(t, X_t))^\top\sigma_t dB_t\\ &= \left[\partial_t f + \sum_{i=1}^{d_1}\partial_{x_i} f b_t^i + \frac{1}{2}\sum_{i,j=1}^{d_1}\sum_{k=1}^{d}\partial_{x_i x_j} f\sigma_t^{ik}\sigma_t^{jk}\right](t, X_t)dt + \sum_{i=1}^{d_1}\partial_{x_i} f(t, X_t)\sum_{j=1}^{d}\sigma_t^{ij}dB_t^j. \end{aligned}$$

where $A : B := \mathrm{tr}\,(A^T B)$ is the trace operator.[14]

3.2 Martingale Representation Theorem

We know that given L^2 process σ, the stochastic integral $\int_0^t \sigma dB_s$ is a martingale. Martingale Representation Theorem deals with the converse.

Theorem 3.9. *Suppose ξ is measurable with respect to the σ-algebra of Brownian motion $\mathcal{F}_T^B := \sigma(B_s : 0 \le s \le T)$ and square integrable. Then, there exists unique process σ adapted to $\mathbb{F}_T^B$ with $\|\sigma\|_2 < \infty$ s.t.*

$$\xi = \mathbb{E}[\xi] + \int_0^T \sigma_t dB_t$$

Consequently, for any martingale M adapted to $\mathbb{F}^B$ satisfying $\mathbb{E}[|M_T|^2] < \infty$, there exists such a unique process σ such that

$$M_t = M_0 + \int_0^t \sigma_s dB_s$$

[14] Trace operator is essentially an inner product, generalizing the usual inner product of vectors.

Proof. Let us first show that the first claim implies the second one. Since $\mathbb{E}[|M_T|^2] < \infty$, there exists σ such that

$$M_T = \mathbb{E}[M_T] + \int_0^T \sigma_s dB_S$$

Define

$$\tilde{M}_t := \mathbb{E}[M_T] + \int_0^t \sigma_s dB_s$$

We know that $\tilde{M}$ is L^2 martingale and $\tilde{M}_T = M_T$. Therefore,

$$M_t = \mathbb{E}[M_T|\mathcal{F}_t^B] = \mathbb{E}[\tilde{M}_T|\mathcal{F}_t^B] = \tilde{M}_t, \quad \Big(\text{and in particular, } M_0 = \mathbb{E}[M_T]\Big)$$

Next, uniqueness is a direct consequence of Itô isometry. Suppose there exists an another such $\tilde{\sigma}$. Then,

$$\int_0^T (\sigma_s - \tilde{\sigma}_s) dB_s = 0$$

Consider the L^2 norm to obtain that

$$\mathbb{E}\Big[\int_0^T |\sigma_s - \tilde{\sigma}_s|^2 ds\Big] = 0, \quad \text{which implies } \sigma_s = \tilde{\sigma}_s \quad dt \times \mathbb{P} - \text{almost surely.}$$

The crucial part is the existence. Construction essentially relies on the connection between the heat equation and Brownian motion. We will only deal with the regular cases, and omit the dealing with the regularity extensions. Let us first assume that $\xi = g(B_T)$ for some bounded measurable g. In particular, ξ is $\sigma(B_T)$ measurable. Define

$$u(t,x) := \mathbb{E}\Big[g(x + B_T - B_t)\Big] = \int_{\mathbb{R}} g(y) p(T-t, y-x) dy, \quad \text{where } p(t,x) := \frac{1}{\sqrt{2\pi t}} e^{-\frac{-x^2}{2t}}$$

It is a straightforward computation to show that p satisfies the forward heat equation:

$$\partial_t p(t,x) - \frac{1}{2}\partial_{xx} p(t,x) = 0$$

Then it is clear that u satisfies the backward heat equation:

$$\partial_t u(t,x) + \frac{1}{2}\partial_{xx} u(t,x) = 0, \quad u(T,x) = g(x)$$

Define

$$M_t := u(t, B_t), \quad \text{and} \quad \sigma_t := \partial_x u(t, B_t)$$

Note that u is smooth as p is smooth, then apply the Itô formula to get

$$du(t, B_t) = \big[\partial_t u + \frac{1}{2}\partial_{xx} u\big](t, B_t) dt + \partial_x u(t, B_t) dB_t = \sigma_t dB_t$$

Therefore,

$$g(B_T) = u(T, B_T) = u(0,0) + \int_0^T \sigma_s dB_s = \mathbb{E}[g(B_T)] + \int_0^T \sigma_s dB_s$$

and since $\partial_x u$ is bounded (hence L^2), we conclude the result.

Next, we will show how to extend for the case $\xi = g(B_{t_1}, \dots, B_{t_n})$ where $0 < t_1 < \cdots < t_n \le T$ and for some measurable $g : \mathbb{R}^n \to \mathbb{R}$. Given $(B_{t_1}, \dots, B_{t_{n-1}})$, we can apply the first result on $[t_{n-1}, t_n]$ to get

$$g(B_{t_1}, \dots, B_{t_n}) = \mathbb{E}\big[g(B_{t_1}, \dots, B_{t_n})|\mathcal{F}^B_{t_{n-1}}\big] + \int_{t_{n-1}}^{t_n} \sigma^n_s dB_s$$

Now, $\mathbb{E}\big[g(B_{t_1}, \dots, B_{t_n})|\mathcal{F}^B_{t_{n-1}}\big]$ is measurable with respect to $\sigma(B_{t_1}, \dots, B_{t_{n-1}})$. In particular, since Brownian motion has independent increment, we can introduce

$$g^{n-1}(x_1, \dots, x_{n-1}) := \mathbb{E}[g(x_1, \dots, x_{n-1}, x_{n-1} + B_{t_n} - B_{t_{n-1}})], \text{ and hence}$$
$$g^{n-1}(B_{t_1}, \dots, B_{t_{n-1}}) = \mathbb{E}\big[g(B_{t_1}, \dots, B_{t_n})|\mathcal{F}^B_{t_{n-1}}\big]$$

Repeat the same argument to get

$$\begin{aligned} g^{n-1}(B_{t_1}, \dots, B_{t_{n-1}}) &= \mathbb{E}\big[g^{n-1}(B_{t_1}, \dots, B_{t_{n-1}})|\mathcal{F}^B_{t_{n-2}}\big] + \int_{t_{n-2}}^{t_{n-1}} \sigma^{n-1}_s dB_s \\ &= g^{n-2}(B_{t_1}, \dots, B_{t_{n-2}}) + \int_{t_{n-2}}^{t_{n-1}} \sigma^{n-1}_s dB_s \end{aligned}$$

where we defined g^{n-2} similar to g^{n-1}. Repeating the same logic yields

$$g(B_{t_1}, \dots, B_{t_n}) = g^0 + \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \sigma^i_s dB_s = \mathbb{E}[g(B_{t_1}, \dots, B_{t_n})] + \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \sigma^i_s dB_s$$

and this concludes the result in this case. ■

3.3 The Girsanov Theorem

In this section, we will discuss changing the measure $\mathbb{P}$ where B is a $\mathbb{P}$-Brownian motion. In essence, consider many paths realized by the Brownian motion. We will construct an another measure $\mathbb{P}^\alpha$ still defined on the same set of paths but shifted in a differentiable way assigning different probabilities and hence B will not be a Brownian motion for $\mathbb{P}^\alpha$, but $B^\alpha_t := B_t - \int_0^t \alpha_s ds$ will be.

To better make sense of the Girsanov's Theorem, let us first discuss the canonical setup. We take $\Omega = C([0, T]; \mathbb{R})$ as the space of continuous functions. Then, we set $B : \Omega \to \mathbb{R}$ as the canonical process, that is, $B_t(\omega) = \omega_t$ and let the filtration generated by the canonical process. Wiener formalized the fact that there exists a probability measure $\mathbb{P}$, called Wiener measure, for which B is a Brownian motion.

To show the underlying mechanism under this explanation, take a partition $\pi : 0 = t_0 < \cdots < t_n = T$ with $\Delta t_i = t_i - t_{i-1}$ and consider the distribution generated by $(\Delta B_1, \dots, \Delta B_n) = (B_{t_1} - B_{t_0}, \dots, B_{t_n} - B_{t_{n-1}})$. Then, take simple processes α, ΔB, and ΔB^α on this π as

$$\alpha(t, \omega) := \sum_{i=1}^n \alpha_i(\omega) \mathbf{1}_{\{[t_{i-1}, t_i)\}}(t),$$
$$\Delta B(t, \omega) := \sum_{i=1}^n \frac{\Delta B_i}{\Delta t_i} \mathbf{1}_{\{[t_{i-1}, t_i)\}}(t), \quad \Delta B^\alpha(t, \omega) := \sum_{i=1}^n \frac{\Delta B_i - \alpha_i \Delta t_i}{\Delta t_i} \mathbf{1}_{\{[t_{i-1}, t_i)\}}(t)$$

We remark that

$$\int_0^t \Delta B_t dt = \sum_{i=1}^n \Delta B_i \sim B_t, \quad \int_0^t \Delta B_t^\alpha dt = \sum_{i=1}^n \Delta B_i - \alpha_i \Delta t_i \sim B_t - \int_0^t \alpha_t dt, \quad \text{and}$$

$$\int_0^t |\Delta B_t|^2 dt = \sum_{i=1}^n \frac{|\Delta B_i|^2}{\Delta t_i} = \begin{pmatrix}\Delta B_1 & \cdots & \Delta B_n\end{pmatrix}^\top \left(\Sigma^\pi\right)^{-1} \begin{pmatrix}\Delta B_1 & \cdots & \Delta B_n\end{pmatrix}$$

where $\Sigma^\pi = \operatorname{diag}(\Delta t_1, \dots, \Delta t_n)$.

We denote the joint density of $(\Delta B_1, \dots, \Delta B_n)$ as $d\mathbb{P}(z_1, \dots, z_n)$ and compute it at $d\mathbb{P}(\Delta B_1, \dots, \Delta B_n)$. This is for notational convenience in the upcoming computation, and one can compute the conditional expectations to recover the density. Now, we are ready to introduce a new probability measure $\mathbb{P}^\alpha$ with the density given as;

$$\begin{aligned}
d\mathbb{P}^\alpha(\Delta B_1, \dots, \Delta B_n) &:= \exp\Big(\int_0^T \alpha_s dB_s - \frac{1}{2}\int_0^T |\alpha_s|^2 ds\Big) d\mathbb{P}(\Delta B_1, \dots, \Delta B_n) \\
&= \exp\Big(\int_0^T \alpha_s dB_s - \frac{1}{2}\int_0^T |\alpha_s|^2 ds\Big) \frac{1}{2^{n/2}|\Sigma^\pi|^{1/2}} \exp\Big(-\frac{1}{2}\int_0^T |\Delta B_s|^2 ds\Big) \\
&= \frac{1}{2^{n/2}|\Sigma^\pi|^{1/2}} \exp\Big(-\frac{1}{2}\int_0^T |\Delta B_s - \alpha_s|^2 ds\Big)
\end{aligned}$$

Note that this is the same joint density, but the mean is shifted by the process α. Let us note explicitly

$$\begin{aligned}
\int_0^T |\Delta B_s - \alpha_s|^2 ds = \sum_{i=1}^n \Big(\frac{\Delta B_i}{\Delta t_i} - \alpha_i\Big)^2 \Delta t_i &= \sum_{i=1}^n \Big(\Delta B_i - \alpha_i \Delta t_i\Big)^2 \frac{1}{\Delta t_i} \\
&= [\Delta B_1 - \alpha_1 \Delta t_1 \ \dots \ \Delta B_n - \alpha_n \Delta t_n]\, (\Sigma^\pi)^{-1}\, [\Delta B_1 - \alpha_1 \Delta t_1 \ \dots \ \Delta B_n - \alpha_n \Delta t_n]^\top
\end{aligned}$$

which in particular shows ΔB_i has expected $\alpha_i \Delta t_i$ under $\mathbb{P}^\alpha$, and variance is not changed. To sum up the main message, under $\mathbb{P}^\alpha$, $\Delta B_i - \alpha_i \Delta t_i$ has exactly the same distribution as ΔB_i under the original $\mathbb{P}$. In other words, ΔB has the same distribution under $\mathbb{P}$ as ΔB^α under $\mathbb{P}^\alpha$. To finish the argument, ignoring the small corrections due to partitions,

$$\int_0^t \Delta B ds \simeq B_t, \quad \text{and} \quad \int_0^t \Delta B^\alpha ds \simeq B_t - \int_0^t \alpha_s ds$$

are both Brownian motions under their corresponding measures. Altough it is not the rigorous proof, we will rely on this observation and only state the main theorem. See Karatzas&Shreve [4].

Theorem 3.10 (Girsanov's Theorem). *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered space where B is $\mathbb{F}$-Brownian motion under $\mathbb{P}$. Suppose α be L^2 process and*

$$M_t := \exp\Big(\int_0^t \alpha_s dB_s - \frac{1}{2}\int_0^t |\alpha_s|^2 ds\Big) \tag{3.3}$$

is a martingale. Then,

$$\mathbb{P}^\alpha(A) := \mathbb{E}\Big[M_T \mathbf{1}_{\{A\}}\Big] = \int_A M_T(\omega) d\mathbb{P}(\omega), \quad \forall A \in \mathcal{F}_T \quad (\textit{that is, } d\mathbb{P}^\alpha := M_T d\mathbb{P})$$

is a probability measure. Moreover, the process

$$B_t^\alpha := B_t - \int_0^t \alpha_s ds$$

is a Brownian motion under $\mathbb{P}^\alpha$.

Exercise 3.11. (i): Prove that $\mathbb{P}^\alpha$ is a probability measure.
[Hint: Find the result where we have already proved $\mathbb{P}^\alpha$ is a measure.]

(ii): Show that $M_t^\alpha = M_t$ defined in (3.3) satisfies

$$M_t = 1 + \int_0^t M_s^\alpha \alpha_s dB_s$$

[Hint: Itô's Formula.]

One sufficient condition for M_t to be a martingale is the following result:[15]

Lemma 3.12 (**Novikov's Condition**). *Suppose*

$$\mathbb{E}\Big[\exp\big(\tfrac{1}{2}\int_0^T |\alpha_t|^2 dt\big)\Big] < \infty$$

then M_t in (3.3) is a martingale.

Notable Results and Dates

- **Itô's Formula** (1944) – Kiyosi Itô
- **Martingale Representation Theorem** (1960s-1970s) – Paul-André Meyer, Claude Dellacherie
- **Girsanov's Theorem** (1960) – Igor Girsanov

[15] See Kazamaki's condition for a more general condition.

4 Stochastic Differential Equations

In this section, we will briefly explore the notion of Stochastic Differential Equations (SDEs), which can be seen as a stochastic version of ordinary differential equations.

Fix a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, where B is an d-dimensional Brownian motion. Our interest here is the solution to the following equation:

$$X_t = X_0 + \int_0^t \alpha_s(\omega, X_s)ds + \int_0^t \sigma_s(\omega, X_s)dB_s, \ \ 0 \le t \le T, \ \ \text{a.s.} \tag{4.1}$$

As usual, we will omit ω from notations. Here $X_0 \in \mathbb{R}^n$ is $\mathcal{F}_0$ measurable, and

$$\alpha : [0, T] \times \Omega \times \mathbb{R}^n \to \mathbb{R}^n, \qquad \sigma : [0, T] \times \Omega \times \times \mathbb{R}^n \to \mathbb{R}^{n \times d}$$

Moreover, $\alpha_t(X_t)$ and $\sigma_t(X_t)$ are (progressively) measurable with respect to the filtration $\mathbb{F}$.

Assumption 4.1. *α and σ are uniformly Lipschitz continuous in x with a Lipschitz constant L. That is,*

$$|\alpha_t(x_1) - \alpha_t(x_2)| + |\sigma_t(x_1) - \sigma_t(x_2)| \le L|x_1 - x_2|, \ \ \forall x_1, x_2 \in \mathbb{R}^n, \quad dt \times d\mathbb{P} - a.s.$$

Also, X_0 is in L^2 and

$$\mathbb{E}\Big[\Big(\int_0^T |\alpha_s(0)|ds\Big)^2\Big] < \infty, \quad \mathbb{E}\Big[\Big(\int_0^T |\sigma_s(0)|^2 ds\Big)\Big] < \infty$$

We won't be rigorously prove the well-posedness, but rather state the necessary steps. Uniqueness can be seen immediately from the following important estimate;

Theorem 4.2. *Suppose $(X_0^1, \alpha^1, \sigma^1)$ and $(X_0^2, \alpha^2, \sigma^2)$ satisfy the Assumption 4.1 and X^1, X^2 are the corresponding solutions to the SDE (4.1). Then,*

$$\mathbb{E}\Big[\sup_{0 \le t \le T} |\Delta X_t|^2\Big] \le C \ \mathbb{E}\Big[|\Delta X_0|^2 + \Big(\int_0^T |\Delta\alpha_t(X_t^1)|dt\Big)^2 + \int_0^T |\Delta\sigma_t(X_t^1)|^2 dt\Big]$$

where $\Delta\varphi := \varphi^1 - \varphi^2$ for $\varphi \in \{X_0, X, \alpha, \sigma\}$.

Theorem 4.3. *Suppose (X_0, α, σ) satisfy the Assumption 4.1. Then SDE (4.1) admits a unique solution.*

Proof. The idea is the same as Picard Iteration. We will only sketch it. Let $X_t^0 := X_0$ for all $0 \le t \le T$ and define

$$X_t^{n+1} := X_0 + \int_0^t \alpha_s(X_s^n)ds + \int_0^t \sigma_s(X_s^n)dB_s$$

Denote $\Delta X^n := X^n - X^{n-1}$ and then show the estimate

$$\mathbb{E}\Big[\int_0^T |\Delta X_s^{n+1}|^2 ds\Big] \le 2L^2(1 + T) \ T \ \mathbb{E}\Big[\int_0^T |\Delta X_s^n|^2 ds\Big]$$

Choose T small enough such that

$$\|\Delta X^{n+1}\|_2 \le \tfrac{1}{2}\|\Delta X^n\|_2$$

Then, since the space of such processes is complete, by Banach Fixed Point theorem, we have the limit satisfying the SDE. Now, to extend to arbitrary times, simply partition to small times and solve at each subinterval. Well-posedness in a small time is almost always much simpler to show. The key observation here is that we can obtain the small T depending only on the uniform Lipschitz constant L. ■

4.1 Forward-Backward SDEs & Feynman-Kac Formula

In this section, as in previous section, we will be less concerned about rigorous arguments but rather will go over some important concepts in a non-formal way. Altough the discussion can be carried out in higher dimensions, to keep the notations simple and typically assume scalar valued functions implicitly.

Similar to the well-posedness of SDEs, one can work out the well-posedness of Backward SDEs (BSDE)s, which is crucial to capture the structure of stochastic optimization in general:

$$Y_t = Y_T + \int_t^T f_s(Y_s, Z_s)ds - \int_t^T Z_s dB_s, \quad 0 \le t \le T, \quad \mathbb{P} - a.s. \tag{4.2}$$

where (Y, Z) pair is the L^2 solution with appropriate dimensions. Note that we are familiar to the concept from financial contracts, where the terminal value is given and we are interested in the Y_t above for pricing purposes. It is worth mentioning that BSDEs can be seen as a nonlinear version of the martingale representation theorem. Next result is in this direction:

Lemma 4.4. *Suppose $f : [0, T] \times \Omega \to \mathbb{R}^n$ is independent of (y, z) and define*

$$Y_t := \mathbb{E}\Big[Y_T + \int_t^T f_s ds \big| \mathcal{F}_t\Big]$$

Then, there exists a unique L^2 process Z such that (Y, Z) is the solution to (4.2).

Proof. We first observe that $Y_t + \int_0^t f_r dr$ is a martingale:

$$\mathbb{E}\Big[Y_t + \int_0^t f_r dr | \mathcal{F}_s\Big] = \mathbb{E}\Big[\mathbb{E}\big[Y_T + \int_s^T f_r dr | \mathcal{F}_s\big] + \int_0^s f_r dr | \mathcal{F}_s\Big] = Y_s + \int_0^s f_r dr$$

Note that we used adaptedness of f and properties of conditional expectation, which some are hidden. Then, by Martingale Representation Theorem, there exists a unique L^2 process Z such that

$$Y_t + \int_0^t f_r dr = \mathbb{E}\Big[Y_T + \int_0^T f_r dr\Big] + \int_0^t Z_r dB_r$$

and in particular

$$Y_T + \int_0^T f_r dr = \mathbb{E}\Big[Y_T + \int_0^T f_r dr\Big] + \int_0^T Z_r dB_r$$

Let us collect the terms as we need to observe (4.2):

$$Y_T + \int_t^T f_r dr - \int_t^T Z_r dB_r = \mathbb{E}\Big[Y_T + \int_0^T f_r dr\Big] - \int_0^t f_r dr + \int_0^t Z_r dB_r = Y_t$$

and hence we conclude the result. ∎

We now further include the forward dynamics into the discussion, typically models the price of underlying assets in finance, and consider the backward equation slightly less general. That is, we will be interested in the following decoupled Forward-Backward SDE (FBSDE) structure:

$$\begin{cases} X_s^{t,x} &= x + \int_t^s \alpha(r, X_r^{t,x})dr + \int_t^s \sigma(r, X_r^{t,x})dB_r & \text{(forward)} \\ Y_s^{t,x} &= \Phi(X_T^{t,x}) + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x})dr - \int_s^T Z_r^{t,x} dB_r & \text{(backward)} \end{cases} \tag{4.3}$$

[16]Notice that we included the initial time and state (t, x) into our notations, and if $t = 0$, we will drop it from superscripts. Note that all the functions are state dependent, and does not rely on paths of the processes. One can show that the solution triplet (X, Y, Z) is Markov, and consequently

$$Y_s^{t,X_t^{0,x}} = Y_s^{0,x}$$

On the right, we have the value process where forward dynamics are initiated at time 0. On the left, however, forward dynamics are initiated at time t, but starts from $X_t^{0,x}$.

Now, define

$$u(t, x) := Y_t^{t,x}$$

and we remark that $u(t, x) \in \mathcal{F}_t$ and again by Markov structure, is independent of $\mathcal{F}_t$. Therefore, it is a well-defined deterministic function. Moreover, it holds that

$$u(t, X_t) := u(t, X_t^{0,x}) = Y_t^{t,X_t^{0,x}} = Y_t^{0,x} =: Y_t$$

We are now ready to observe the Feynman-Kac formula directly. Suppose $u \in C^{1,2}$, and write down the Itô formula:

$$dY_t = du(t, X_t) = \big[\partial_t u + \alpha \partial_x u + \tfrac{1}{2}\mathrm{tr}\,(\sigma\sigma^\top \partial_{xx} u)\big](t, X_t)dt + [\sigma \partial_x u](t, X_t)dB_t$$

On the other hand,

$$dY_t = -f(t, X_t, Y_t, Z_t)dt + Z_t dB_t$$

Therefore, we observed

$$\big[\partial_t u + \alpha \partial_x u + \tfrac{1}{2}\mathrm{tr}\,(\sigma\sigma^\top \partial_{xx} u) + f(\cdot, Y_t, Z_t)\big](t, X_t) = 0 \quad \text{and} \quad Z_t = [\sigma \partial_x u](t, X_t)$$

We then write down the Feynman-Kac Formula, and adapt it to our needs in the pricing of contracts. First, we will assume f is independent of (y, z) and second, we will consider a discounting factor.

Theorem 4.5 (**Feynman-Kac Formula**). *Suppose some regularity conditions (such as Lipschitz continuity etc.) holds for the coefficients, and $u \in C^{1,2}([0, T] \times \mathbb{R}^d; \mathbb{R})$ is a classical solution to the following PDE:*

$$\big[\partial_t u + \alpha^\top \partial_x u + \tfrac{1}{2}\mathrm{tr}\,(\sigma\sigma^\top \partial_{xx} u) - ru + f\big](t, x) = 0, \quad u(T, x) = \Phi(x) \tag{4.4}$$

Then,

$$u(t, x) = \mathbb{E}\left[e^{-r(T-t)}\Phi(X_T^{t,x}) + \int_t^T e^{-r(s-t)} f(s, X_s^{t,x})ds\right] \tag{4.5}$$

Remark 4.6. (i): Note that the $-ru$ term is due to the discounting we typically encounter in finance. By setting $\tilde{u}(t, x) = e^{-r(T-t)}u(t, x)$, you can modify the above discussion in a straightforward manner.

(ii): Notice that this is a representation result for the solution of the PDE (4.4). The result follows directly from Lemma 4.4. Other terms are incorporating the discounting factor appropriately.

Example 4.7. The heat equation is given by

$$\partial_t u - \tfrac{1}{2}\partial_{xx} u = 0$$

with some initial condition $u(0, x) = u_0(x)$. Then, by taking $(\alpha, \sigma, r, f) = (0, 1, 0, 0)$, and changing the time from backward to forward, one can observe that the solution is given by

$$u(t, x) = \mathbb{E}[u_0(x + B_t)]$$

This shows how the Brownian motion can directly model the diffusion of heat.

[16]Do not confuse the integration variable r with the interest rate, which will also be denoted as r.

Notable Results and Dates

- **Picard Iteration Method** (1890) – Émile Picard
- **Lévy Processes** (1930s) – Paul Lévy
- **Feynman-Kac Formula** (1940s) – Mark Kac, Richard Feynman
- **Forward-Backward SDEs** (1990s) – Étienne Pardoux, Shige Peng, Jin Ma, Jiongmin Yong

5 Portfolio Structure & Market Dynamics

5.1 Portfolio Structure

In this section, we will review concepts related to portfolios. To reserve the notation B to the risk-free asset, we will switch to denoting the Brownian motion (or Wiener process) as W.

We assume there are N risky-assets $S = (S^1, \ldots, S^N)$, and a risk-free asset B[17]. To simplify the notations, as long as we don't need to explicitly mention, we will treat $S = (S^0, S^1, \ldots, S^N)$ with $S^0 = B$.

We characterize a portfolio by an adapted process $h_t = (h_t^0, h_t^1, \ldots, h_t^N)$ denoting the number of assets in the portfolio, and an another scalar adapted process c_t called consumption process. The portfolio value V^h is given by

$$V_t^h := h_t^\top S_t = h_t \cdot S_t = h_t^0 B_t + \sum_{i=1}^{N} h_t^i S_t^i$$

Assumption 5.1. *We only consider portfolios such that the portfolio value V^h is an L^2 process.*

Definition 5.2. We say (h, c) is a self-financing portfolio if

$$dV_t^h = h_t \cdot dS_t - c_t dt \tag{5.1}$$

Given h, we can create the relative-portfolio u as

$$u_t^i := \frac{h_t^i S_t^i}{V_t^h}, \qquad 0 \le i \le N$$

and note that $\sum_{i=0}^{N} u^i = 1$. We can write down the self-financing condition in terms of relative-portfolio:

Definition 5.3. We say (h, c) is a self-financing portfolio if

$$dV_t^h = V_t^h \sum_{i=0}^{N} u_t^i \frac{dS_t^i}{S_t^i} - c_t dt \tag{5.2}$$

Let us state an immediate observation as a lemma:

Lemma 5.4. *Suppose there are scalar valued processes c_t, Z_t, u_t^i, $0 \le i \le N$ where*

$$dZ_t = Z_t \sum_{0=1}^{N} u_t^i \frac{dS_t^i}{S_t^i} - c_t dt, \quad \textit{and} \quad \sum_{0=1}^{N} u_t^i = 1$$

Define a portfolio h_t by

$$h_t^i = \frac{u_t^i Z_t}{S_t^i} \tag{5.3}$$

Then the value process V^h is given by $V_t^h = Z_t$, the pair (h, c) is self-financing, and u is the corresponding relative portfolio to h.

[17] Do not confuse with Brownian motion.

Proof. By definition $V_t^h = h_t S_t$, and hence

$$V_t^h = \sum_{i=0}^{N} h_t^i S_t^i = \sum_{i=0}^{N} u_t^i Z_t = Z_t \sum_{i=0}^{N} u_t^i = Z_t$$

By definition of u_t^i, it is immediate that u is indeed the relative portfolio corresponding to h. By (5.2), it is also clear that (h, c) is a self-financing portfolio. ■

We present this familiar notion of relative portfolio, which is convenient in certain cases such as portfolio management. However, as we are mainly interested in pricing the contracts, we will typically encounter arbitrage portfolios with no value, which does not immediately allow the construction of relative portfolios.

Let us motivate the self-financing condition by the discrete case. Suppose we index times with $0, 1$. Then, initial portfolio is $V_0 = h_0^\top S_0$, and at time 1 it becomes $V_1 = h_1^\top S_1$.[18] If the portfolio is self financing,

$$h_1^\top S_1 + c_1 = h_0^\top S_1 \implies (h_1 - h_0)^\top S_1 + c_1 = 0$$

Therefore,

$$V_1 - V_0 = (h_1^\top - h_0^\top) S_1 + h_0^\top (S_1 - S_0) = h_0^\top (S_1 - S_0) - c_1$$

In words, change in portfolio value is due to the change in the price of the assets and the consumption only. (5.1) in the definition reflects the same idea.

If one wants to see the connection between discrete and continuous indexing more formally, with appropriate definitions and no consumption, it suffices to write that for a partition $0 = t_0 < t_1 < \cdots < t_n = t$,

$$\sum_{i=1}^{n} V_{t_i} - V_{t_{i-1}} = \sum_{i=1}^{n} h_{t_{i-1}}^\top (S_{t_i} - S_{t_{i-1}}) + (h_{t_i} - h_{t_{i-1}})^\top S_{t_i} = \sum_{i=1}^{n} h_{t_{i-1}}^\top (S_{t_i} - S_{t_{i-1}})$$

and then notice that as the partition size converges to 0, this is exactly (5.1).

Let us also present the core notion of finance that pricing of derivatives are build upon. Consumption is irrelevant in the pricing of derivatives, so we will assume $c = 0$ and don't mention it.

Definition 5.5. We say a self-financing portfolio h is an arbitrage portfolio if

$$V_0^h = 0, \qquad \mathbb{P}(V_T^h \geq 0) = 1, \qquad \mathbb{P}(V_T^h > 0) > 0$$

We will not be treating dividends, but to briefly mention, suppose stochastic processes $D_t = (D_t^1, \ldots, D_t^N)$ are given representing the cumulative dividends paid for each stock. We may assume that cumulative dividends paid for each stock will be given as $D_t^i = \int_0^t \delta_s^i ds$, where the process δ_t^i is called the dividend yield. Now, define the so called gain process $G_t = S_t + D_t$ and the self-financing condition becomes

$$dV_t^h = h_t^\top dG_t - c_t dt$$

Note that, altough the price of the underlying asset is S_t, if one considers a portfolio holding this asset, the portfolio value is following the gain process.

[18] To handle predictibility of the portfolio, in discrete time, typically one considers index as shifted. We won't rigorously discuss this point further.

5.2 Market Dynamics

First, let us start with generalized geometric Brownian motion, which we model our price dynamics of risky assets.

Example 5.6 (Generalized Geometric Brownian Motion)**.** We call the stochastic process S_t

$$S_t = S_0 \exp\Big(\int_0^t \sigma_s dW_s + \int_0^t \big[\alpha_s - \tfrac{1}{2}\sigma_s^2\big]ds\Big) \tag{5.4}$$

the generalized geometric Brownian motion, which satisfies the following SDE:

$$dS_t = \alpha_t S_t dt + \sigma_t S_t dW_t \tag{5.5}$$

Here, α, σ are bounded, adapted processes. We typically call α as drift and σ as volatility.

Exercise 5.7. (i): Show that (5.4) indeed satisfies the SDE (5.5).

(ii): Suppose α and σ are constants. Find the ODE that $\bar{S}(t) := \mathbb{E}[S(t)]$ satisfies and write down the explicit solution.

Definition 5.8. We say B is a risk-free asset if the dynamics are given as

$$dB_t = r_t B_t dt, \quad \text{that is,} \quad B_t = B_0 \exp\Big(\int_0^t r_s ds\Big) \tag{5.6}$$

for some bounded, adapted process r_t.

An important special case is the famous Black-Scholes model (BS). The dynamics of BS is modeled by one risk-free and one risky asset, as in (5.5) and (5.6), where (r, α, σ) are all constant.

When we aim to consider N-assets, similar to (5.5),we model it as the solution to the following SDE

$$dS_t = S_t \circ \alpha_t dt + S_t \circ \sigma_t dW_t \tag{5.7}$$

Here, $\alpha \in \mathbb{R}^N, \sigma \in \mathbb{R}^{N\times d}$ are bounded, adapted stochastic processes, W is a d-dimensional standard Brownian motion, and $\circ$ is the Hadamard product (element-wise product) between two vectors. To be able to make sense of

$$S_t \circ (\sigma_t dW_t) = (S_t \circ \sigma_t) dW_t \tag{5.8}$$

we extend the Hadamard product $(x \circ \sigma)$ between a vector in $x \in \mathbb{R}^N$ and a matrix in $\sigma \in \mathbb{R}^{N\times d}$ as ith element of the vector multiplying the ith row. If $x \in \mathbb{R}^d$, then we set right multiplication $(\sigma \circ x)$ as the ith element of the vector multiplying the whole ith column. Moreover, we denote x^{-1} as the elementwise inverse, that is $x^{-1} := [1/x^1, \ldots 1/x^d]$. Our choice here is tailored to be compatible with Hadamard product and if $N = d$ satisfy

$$(x^{-1} \circ \sigma^{-1})(\sigma \circ x) = \mathrm{Id}_{N\times N} = (x \circ \sigma)(\sigma^{-1} \circ x^{-1})$$

where σ^{-1} is the usual inverse of $N \times N$ square matrix.

To be more explicit, for any $i = 1, \ldots, N$,

$$S_t^i = S_0^i + \int_0^t S_s^i \alpha_s^i ds + \int_0^t S_s^i \sum_{j=1}^d \sigma_s^{i,j} dW_s^j$$

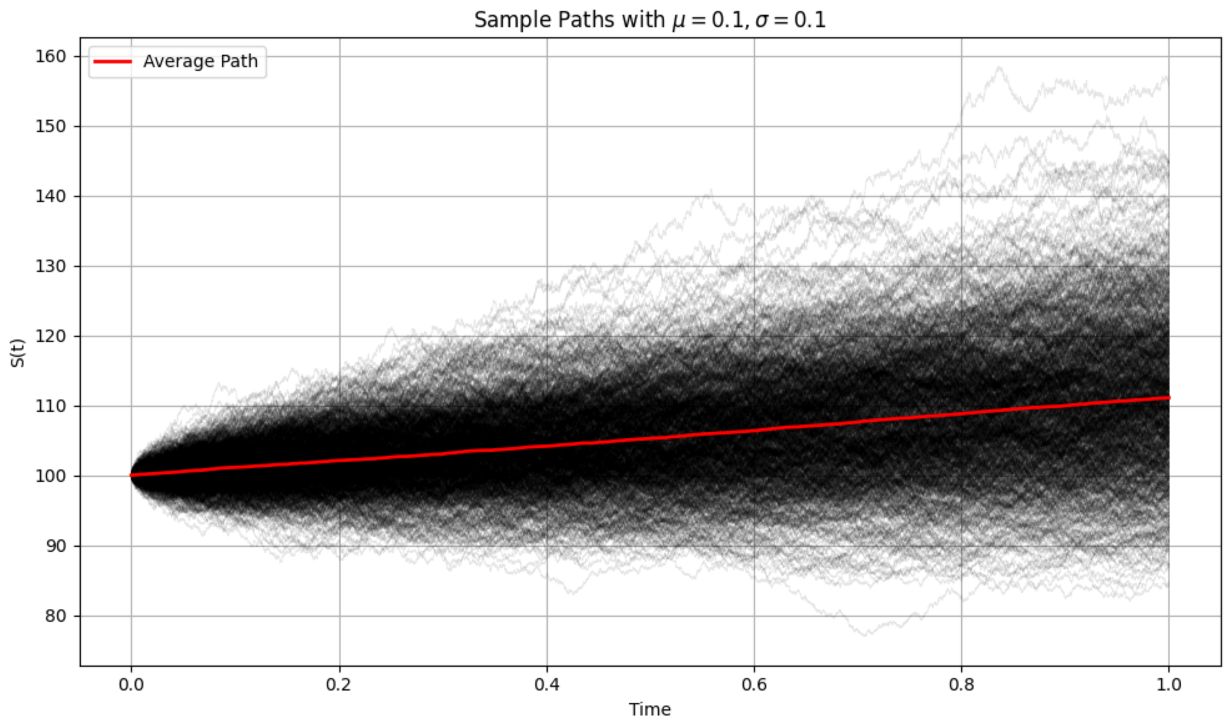


Figure 2: Sample paths of S_t in the Black-Scholes model.

5.3 Contingent Claims

One of the central interest in math finance is to understand how to price contracts, or so called contingent claims. Recall that we fix a time horizon T[19], a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ and typically $\mathcal{F}_t = \mathcal{F}_t^S = \sigma(S_s : 0 \leq s \leq t)$ given by the dynamics of risky asset S.

Definition 5.9. We say $\mathcal{X}$ is a contingent claim if $\mathcal{X} \in \mathcal{F}_T$. We say $\mathcal{X}$ is a simple claim if $\mathcal{X} \in \sigma(S_T)$. In this case, by Doob-Dynkin Theorem 1.31, $\mathcal{X} = \Phi(S_T)$ for some measurable deterministic function Φ called contract function.

Example 5.10. Europen call and put options with strike price K are simple claims where

$$\Phi^{\text{call}}(x) = (x - K)^+, \qquad \Phi^{\text{put}}(x) = (K - x)^+$$

Before moving on to the main focus of pricing these contingent claims, let us briefly review some examples, without delving into exact details, to illustrate their high relevance in the finance industry. Call and Put options, also known as vanilla options, are standardized contracts extensively traded for equities on public exchanges such as the New York Stock Exchange (NYSE) and NASDAQ. There are also non-standardized contracts known as exotic options, which are typically traded over-the-counter (OTC) where market makers are investment banks like Goldman Sachs, J.P. Morgan, and Barclays. For instance, Asian options provide exposure to the average price of an asset, which is particularly useful in the commodity market for producers, such as in the oil or agriculture industries. Another example is lookback options, which depend on the maximum or minimum price of the underlying asset and are commonly used in currency markets to hedge against unfavorable exchange rate movements by locking in the best historical rate.

[19] T is typically called maturity date, terminal time, time horizon etc. all refers to time when contract terminates.

Additionally, there are other major classes of derivatives. For interest rates, contracts called swaptions are commonly used by hedge funds, institutional investors, and banks to manage interest rate risk. For credit risk, credit default swaps (CDS) are used to hedge the default risk of corporate debt, protecting against the risk that a company will fail to meet its debt obligations. Furthermore, there are weather derivatives for temperature, rainfall, or other adverse weather events, which are typically issued by companies such as Swiss Re and Aon. Agricultural producers, energy companies, and insurance firms use these derivatives to hedge weather-related risks. Although the underlying dynamics of these contracts may require different models, the central idea of risk-neutral pricing will remain the same across all of them.

A standard notation for the price of a contingent claim is $\Pi(t;\mathcal{X})$ in general, and $\Pi(t;\Phi)$ in case $\mathcal{X}$ is a simple claim.[20] The price of the contract at the terminal time T has no ambiguity. That is, by definition, $\Pi(T;\mathcal{X}) = \mathcal{X}$. We are interested in the price of a contract for $0 \le t \le T$.

We have an important assumption which will always be in consideration.

Assumption 5.11. *All the assets, such as risk-free, risky and derivatives of risky assets, are all tradeable without any restrictions or costs.*

We have seen in discrete models that 'frictions' in the markets give rise to an interval of arbitrage free pricing, rather than a unique price. For the purposes of this course, we will not attempt to model such generality in continuous framework.

Proposition 5.12. *Suppose there exists a self-financing portfolio h, such that the value process V^h satisfies*

$$dV_t^h = k_t V_t^h dt, \quad V_0^h > 0$$

for some adapted process k_t. If $r_t = k_t$ $dt \times d\mathbb{P}$-a.s. does not hold, then there exists an arbitrage portfolio.

Proof. Let us directly construct a portfolio $\hat{h}$ investing in in two assets (B, V^h) as follows:

$$\hat{h}(t,\omega) := \big(m_t/B_t,\ s_t/V_t^h\big), \quad \text{where} \quad s_t := \text{sign}(k_t - r_t)$$

Note that, the value of portfolio $\hat{h}$ is

$$V_t^{\hat{h}} = m_t + s_t$$

Instead of determining m_t, we set $m_t := V_t^{\hat{h}} - s_t$, and determine what $V_t^{\hat{h}}$ must be from the self-financing condition:

$$dV_t^{\hat{h}} = \tfrac{m_t}{B_t} dB_t + \tfrac{s_t}{V_t} dV_t = \tfrac{V_t^{\hat{h}}}{B_t} dB_t - \tfrac{s_t}{B_t} dB_t + \tfrac{s_t}{V_t} dV_t = \tfrac{V_t^{\hat{h}}}{B_t} dB_t + s_t(k_t - r_t)dt$$

Noting that $dV_t^{\hat{h}} = d(B_t(V_t^{\hat{h}}/B_t)) = (V_t^{\hat{h}}/B_t)dB_t + B_t d(V_t^{\hat{h}}/B_t)$,

$$B_t d(V_t^{\hat{h}}/B_t) = s_t(k_t - r_t)dt = |k_t - r_t|dt$$

More explicitly, and since $V_0^{\hat{h}} = 0$,

$$V_t^{\hat{h}} = B_t \int_0^t \tfrac{|k_u - r_u|}{B_u} du$$

This not only well-defines m_t, but also concludes the result. Note that, by assumption $|k_u - r_u| > 0$ on a set of positive probability, and hence $V_T^{\hat{h}} > 0$ with positive probability. ∎

[20] We may omit $\mathcal{X}$ (or Φ) if it is clear.

Assumption 5.13. *We assume the market formed by financial assets* $(B, S, \mathcal{X})$ *is arbitrage free.*

Next, we will derive Black-Scholes equation, the PDE that the price of the contingent claim satisfies under some simplifying assumptions.

Theorem 5.14 (Black-Scholes Equation). *Consider a simple claim* $\mathcal{X} = \Phi(S_T)$ *under the Black-Scholes model. Suppose Assumptions 5.11 and 5.13 holds. Moreover, we assume that the price of the contract is determined by* $F \in C^{1,2}([0,T] \times \mathbb{R}; \mathbb{R})$ *where*

$$\Pi(t; \Phi) = F(t, S_t)$$

Then, F is the classical solution to the Black-Scholes PDE

$$\Big[\partial_t F + rx\partial_x F + \frac{1}{2}x^2\sigma^2\partial_{xx}F - rF\Big](t,x) = 0, \quad \textit{with terminal} \quad F(T,x) = \Phi(x)$$

Proof. Let us consider a self-financing portfolio V formed by m_t many risk-free asset, k_t many underlying asset S, 1 simple claim $\mathcal{X}$. Note that, we cannot omit m_t, otherwise there are no self-financing portfolios. It is intuitively easy to believe that for any k_t, we can find the m_t, which is simply accounting the price differences as reflected on risk-free asset. We will determine m_t at the end too.

Now, we can write dV by self-financing property and Itô's formula:

$$\begin{aligned}
&dV_t \\
&= dF(t,S_t) + k_t dS_t + m_t dB_t \quad \text{(self-financing)} \\
&= \Big[\partial_t F + \alpha S_t \partial_x F + \frac{1}{2}\sigma^2 S_t^2 \partial_{xx} F + k_t \alpha S_t + m_t r B_t\Big](t,S_t)dt + \Big[\sigma S_t \partial_x F + \sigma S_t k_t\Big](t,S_t)dW_t
\end{aligned}$$

Now, we set

$$k_t := -\partial_x F(t, S_t)$$

which is essentially the so called delta-hedging. Then,

$$dV_t = \Big[\partial_t F + \frac{1}{2}\sigma^2 S_t^2 \partial_{xx} F + m_t r B_t\Big](t,S_t)dt$$

It is important to note that at this step α dissapears from analysis. By Proposition 5.12,

$$\partial_t F + \frac{1}{2}\sigma^2 S_t^2 \partial_{xx} F + m_t r B_t = rV_t = r(F - \partial_x F S_t + m_t B_t)(t, S_t)$$

Rearrange the terms to get:

$$\Big[\partial_t F + rS_t\partial_x F + \frac{1}{2}\sigma^2 S_t^2 \partial_{xx} F - rF\Big](t,S_t) = 0$$

which holds $dt \times d\mathbb{P}$-almost surely. Since S_t has full range on $\mathbb{R}_+$, we conclude the result. It is not stated in the theorem, but the solution is satisfies for $t \in [0,T)$, and $x > 0$.

Lastly, let us also explicitly determine what m_t should be in order to have a self-financing portfolio. Note that, in general, $dV_t = dF(t,S_t) + d(k_t S_t) + d(m_t B_t)$. Therefore, since risk-free asset has no martingale term, it follows

$$\begin{aligned}
d(k_t S_t) + d(m_t B_t) = k_t dS_t + m_t dB_t &\implies d(k_t S_t) + m_t dB_t + B_t dm_t = k_t dS_t + m_t dB_t \\
&\implies m_t = m_0 + \int_0^t \frac{k_s}{B_s} dS_s - \int_0^t \frac{1}{B_s} d(k_s S_s)
\end{aligned}$$

where $k_s S_s$ is also a semi-martingale[21], hence the integration is well-defined. Let us also note that we need to choose m_0 to make sure $V_0 > 0$. Here, we are representing m_t in terms of a process k_t directly. If we were to follow the proof of 5.12, we would get

$$m_t := \frac{V_0}{B_0} + \int_0^t \frac{1}{B_u} dF(u, S_u) - \frac{1}{B_t} F(t, S_t) + \int_0^t \frac{k_u}{B_u} dS_u - \frac{1}{B_t}(k_t S_t) - \int_0^t (F(u, S_u) + k_u S_u) r du$$

■

Definition 5.15. We say that contingent claim $\mathcal{X}$ can be replicated (or reachable, or hedgeable) if there exists a self-financing portfolio h investing in risk-free and the underlying risky asset such that

$$\mathcal{X} = \Pi(T; \mathcal{X}) = V_T^h \quad \mathbb{P} - a.s.$$

Lemma 5.16. *Suppose $\mathcal{X}$ is hedgeable by h. If there is no arbitrage, then $\Pi(t, \mathcal{X}) = V_t^h$ for all t $\mathbb{P}$-almost surely. Moreover, if an another self-financing portfolio g is hedging $\mathcal{X}$ too, then $V_t^h = V_t^g$ for all t almost surely.*

Proof. Idea is quite simple. If this does not hold, sell one and buy the other one to create an arbitrage. ■

Definition 5.17. We say that a market is complete, if every contingent claim is replicable.

Theorem 5.18 (**Risk Neutral Valuation**). *Suppose Assumptions 5.11 and 5.13 holds. Consider the generalized Black-Scholes model:*

$$dB_t = r_t B_t dt, \qquad dS_t = \alpha_t S_t dt + \sigma_t S_t dW_t \tag{5.9}$$

where r_t, α_t and $\sigma_t \geq c > 0$ are adapted bounded processes.
(i): Generalized Black-Scholes model (5.9) is complete.
(ii): For any contingent claim $\mathcal{X} \in \mathcal{F}_T^S$ written on S, the price is given by

$$\Pi(t; \mathcal{X}) = B_t \mathbb{E}^\theta \Big[\big(\mathcal{X} / B_T \big) \big| \mathcal{F}_t^S \Big] \tag{5.10}$$

where $\mathbb{E}^\theta$ is taken under risk-neutral probability measure $\mathbb{P}^\theta$ for which

$$dS_t = r_t S_t dt + \sigma_t S_t dW_t^\theta, \quad \text{where } W^\theta \text{ is } \mathbb{P}^\theta\text{-Brownian motion}$$

(iii): If we further assume the structure in Theorem 5.14, representation of F is given by

$$F(t, x) := e^{-r(T-t)} \mathbb{E}\big[\Phi(X_T^{t,x}) \big] \tag{5.11}$$

where

$$dX_s^{t,x} = r X_s^{t,x} ds + \sigma X_s^{t,x} dW_s, \quad X_t^{t,x} = x \tag{5.12}$$

Proof. (iii) is exactly the Feynman-Kac Theorem 4.5 by setting

$$f = 0, \quad \alpha(t, x) = rx, \quad \sigma(t, x) = \sigma x, \quad \text{and } u = F$$

[21] Continuous function of a semi-martingale is a semi-martingale, and product of two semimartingales is also a semimartingale.

(i) and (ii) will follow by the construction of the replicating portfolio. Define

$$\theta_t = -\frac{\alpha_t - r_t}{\sigma_t}, \qquad M_t^\theta := \exp\left(\int_0^t \theta_s dW_s - \frac{1}{2}\int_0^t |\theta_s|^2 ds\right)$$

and $d\mathbb{P}^\theta := M_T d\mathbb{P}$, $W_t^\theta := W_t - \int_0^t \theta_s ds$. Let us note that $-\theta_t$ is called the Sharpe ratio. By Girsanov's Theorem 3.10, W_t^θ is a $\mathbb{P}^\theta$ Brownian motion. Notice that,

$$dS_t = r_t S_t dt + S_t \sigma_t dW_t^\theta$$

and in particular, the martingale term can be replicated by the underlying market:

$$S_t \sigma_t dW_t^\theta = dS_t - \frac{S_t}{B_t} dB_t$$

Now, define $V_t := B_t \mathbb{E}^\theta[\mathcal{X}/B_T|\mathcal{F}_t^S]$.[22] For the moment, assume that the process S generates the same filtration as W^θ. Then, by Martingale Representation Theorem 3.9, there exists unique L^2 stochastic process Z_t such that

$$V_t/B_t = \mathbb{E}^\theta[\mathcal{X}/B_T] + \int_0^t Z_s dW_s^\theta$$

Now, obviously if we can write V_t as a self-financing portfolio, it will replicate the contract, and this will also conclude (ii) by Lemma 5.16. To do so, we simply write the self-financing condition to notice the appropriate portfolio:

$$\begin{aligned} dV_t &= B_t d(V_t/B_t) - V_t/B_t dB_t \\ &= (V_t/B_t)dB_t + \frac{B_t Z_t}{S_t \sigma_t}(S_t \sigma_t dW_t^\theta) \\ &= (V_t/B_t)dB_t + \frac{B_t Z_t}{\sigma_t S_t}\left(dS_t - \frac{S_t}{B_t}dB_t\right) = \left(\frac{V_t}{B_t} - \frac{Z_t}{\sigma_t}\right)dB_t + \left(\frac{B_t Z_t}{\sigma_t S_t}\right)dS_t \end{aligned}$$

Then the portfolio has to be

$$h_t^0 := \frac{V_t}{B_t} - \frac{Z_t}{\sigma_t}, \qquad h_t^1 := \frac{B_t Z_t}{\sigma_t S_t}$$

Let us verify that $V_t^h = V_t - Z_t B_t/\sigma_t + B_t Z_t/\sigma_t = V_t$, and hence we conclude the results.

Lastly, let us discuss the filtrations. First, just to note, it is obvious that $\mathbb{F}^W = \mathbb{F}^{W^\theta}$. It is also clear that $\mathbb{F}^S \subset \mathbb{F}^{W^\theta}$. We need to write W^θ as a function of S:

$$dW_t^\theta = \frac{1}{\sigma_t}\frac{dS_t}{S_t} + \frac{r_t}{\sigma_t}dt$$

where we can define $\int dS_t$ exactly as $\int dW_t$. ■

Example 5.19 (Forward Contract)**.** Let us consider the forward contract, which is characterized by $\Phi(x) = x - F$, where F is the aggreement price of the contract to be determined. Then, (5.10) implies

$$F(t,x) := e^{-r(T-t)}\mathbb{E}\big[X_T^{t,x}\big] - Fe^{-r(T-t)}$$

where we know that $\mathbb{E}\big[X_T^{t,x}\big] = xe^{r(T-t)}$. Moreover, we design the forward contract such that $F(0, S_0) = 0$, hence

$$F(0, S_0) = 0 = S_0 - Fe^{-rT} \implies F = S_0 e^{rT}$$

[22] Here $\mathbb{E}^\theta$ denotes the expectation under the measure $\mathbb{P}^\theta$.

Example 5.20 (European Call)**.** Another well known example is the European call contract, for which $\Phi(x) = (x - K)^+$. Note that, by (5.4), we know the explicit solution and

$$\log(X_T^{t,x}/x) \sim \mathcal{N}\Big((\alpha - \tfrac{1}{2}\sigma^2)(T-t), \sigma\sqrt{T-t}\Big)$$

By Risk Neutral Valuation;

$$F(t,x) = e^{-r(T-t)} \int_{\mathbb{R}} \Phi(xe^z) f(z) dz$$

where f is the pdf of $\log(X_T^{t,x}/x)$. By explicit computations, one arrives at

$$C_K^E(t, S_t; T) := F(t, S_t) = S_t N(d_+) - Ke^{-r(T-t)} N(d_-)$$

where N is the cdf of standard normal distribution and

$$d_\pm := \frac{\log(S_t/K) + (r \pm \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$$

Example 5.21 (Asian & Lookback Options)**.** We will only introduce here to give an example of non-simple contingent claims. Fixed strike Asian options' terminal payoff is given by

$$\mathcal{X} = \Big(\frac{1}{T}\int_0^T S_t dt - K\Big)^+$$

and floating strike lookback options terminal payoff is given by

$$\mathcal{X} = S_T - \inf_{0 \le t \le T} S_t$$

Note that in both cases $\mathcal{X} \in \mathcal{F}_T^S$ but $\mathcal{X} \notin \sigma(S_T)$.[23]

5.4 Contracts are dense & Hedging the Greeks

Let us first note that the pricing formula (5.10) is linear in the contract:

Lemma 5.22. *Let $\mathcal{X}, \mathcal{Y}$ are two contingent claims and $a \in \mathbb{R}$. Then,*

$$\Pi(t; a\mathcal{X} + \mathcal{Y}) = a\Pi(t;\mathcal{X}) + \Pi(t;\mathcal{Y})$$

Now, our aim is to approximate any continuous terminal payoff with a constant (or buy-and-hold) portfolio. Linear combinations of bond and the risky-asset is not sufficient to replicate much, however, together with linear combinations of call options with all strikes is rich enough to approximate every continuous function. Let us now introduce the associated contracts

$$\Phi_S(x) := x \text{ (risky-asset)}, \quad \Phi_B(x) := 1 \text{ (bond)}, \quad \Phi_C^K(x) := (x-K)^+, \quad \Phi_P^K(x) := (K-x)^+$$

with pricing

$$\Pi(t;\Phi_S) = S(t), \;\; \Pi(t;\Phi_B) = e^{-r(T-t)}, \;\; \Pi(t;\Phi_C^K(x)) =: C_K^E(t,x), \;\; \Pi(t;\Phi_P^K(x)) =: P_K^E(t,x)$$

We may add (T, r, σ) to the notations of contracts when necessary. Recall the Put-Call parity, which follows from the fact:

$$\text{(Synthetic Future)} \quad \Phi_C^K - \Phi_P^K = \Phi_S - K\Phi_B$$

[23]This σ represents the σ-algebra, do not confuse with the volatility.

Lemma 5.23 (Put-Call Parity). *Price of the Europen call and put contracts satisfies*

$$C_K^E(t,S;T) - P_K^E(t,S;T) = S - Ke^{-r(T-t)}$$

Note that, in particular, put contract can be replicated by holding one call contract, K bonds and selling the underlying asset. Then, it is easy to argue that

Proposition 5.24. *The set*

$$\left\{ a_S\Phi_S + a_B\Phi_B + \sum_{i=1}^{n} \gamma^i \Phi_C^{K_i} \, : \, \forall n \in \mathbb{N}, a_S, a_B, \gamma^i, K_i \in \mathbb{R}, 1 \le i \le n \right\}$$

is dense in the set of continuous functions with compact domain under the supremum norm.

and hence we coclude that every simple contingent claim with a compact continuous terminal can be (approximately) replicated with a buy-and-hold strategy.

5.4.1 The Greeks

In this section, we will discuss the sensitivity of the contract to the underlying parameters, and discuss how to find portfolios that are insensitive to them.

Definition 5.25. Let $D = D(t,s,T,r,\sigma)$ be any pricing function. Then, we define "the greeks":

$$\Delta_D := \partial_s D, \quad \Gamma_D := \partial_{ss} D, \quad \rho_D := \partial_r D, \quad \Theta_D := \partial_t D, \quad \mathcal{V}_D := \partial_\sigma D$$

We note that D might represent the price of a single contract or a whole portfolio.

Proposition 5.26. *Under the Black-Scholes model, the greeks of European call price is given by*

$$\Delta = N(d_+), \quad \Gamma = \frac{f(d+)}{S\sigma\sqrt{T-t}}, \quad \rho = K(T-t)e^{-r(T-t)}N(d_-),$$
$$\Theta = \frac{-sf(d_+)\sigma}{2\sqrt{T-t}} - rKe^{-r(T-t)}N(d_-), \quad \mathcal{V} = sf(d_+)\sqrt{T-t}$$

where f, N are pdf and cdf of standard normal distribution.

5.4.2 Hedging the Greeks

We say a portfolio is neutral if the mentioned greek is 0. For example, we say portfolio V is Δ-neutral if $\Delta_V = 0$. Recall the proof of Black-Scholes Theorem 5.14, where we replicated the contract by matching the amount of the underlying asset exactly to the Δ of the contract. In the perspective of a market maker, they can sell any contract and in theory replicate it by the underlying asset. However, as we cannot adjust portfolios continuously without any frictions, it is important to be able to create portfolios that are insensitive to some parameters.

We will briefly discuss how we can create portfolios that are Δ and Γ-neutral. Before that, let us mention two more greeks

$$\text{Vanna}_D := \partial_\sigma(\Delta_D), \quad \text{Charm}_D := \partial_t(\Delta_D)$$

which is important for market makers. As we discussed why Δ is important, how Δ is changing is also important.

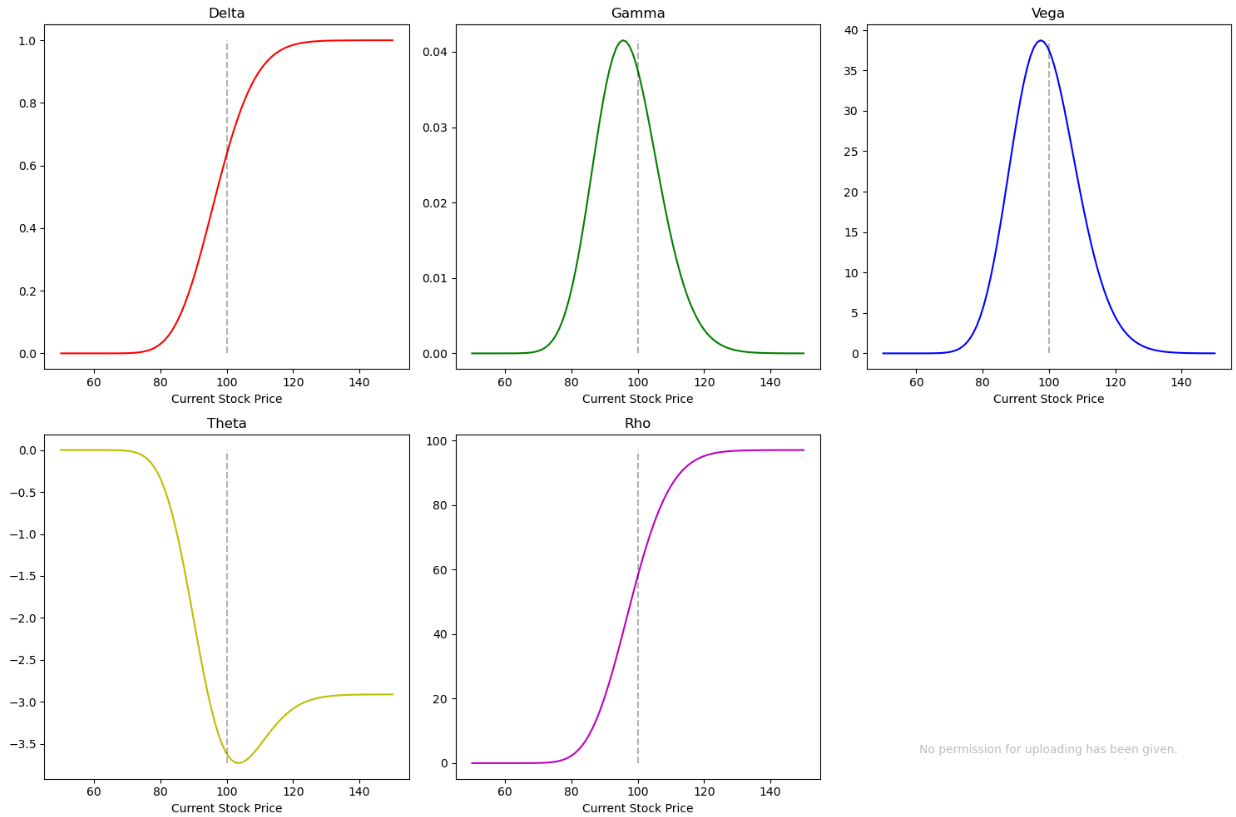


Figure 3: The greeks of a European Call option depending on S where $K = 100, T = 1, r = 3\%, \sigma = 10\%$.

Suppose we are given two contracts D and D', for example two call options with different strike or maturities. Let $(h_1, h_2, 1)$ be an absolute portfolio investing in (risky-asset, D, D'). Note that the price of this portfolio is

$$V^h = h_1 S + h_2 D + D'$$

Here V^h is treating the parameters as given variables, but not like S_t as usualy. For example, $\partial_t V^h$ is independent of the change in the current price. Then,

$$\Delta_{V^h} = h_1 + h_2 \Delta_D + \Delta_{D'}, \qquad \Gamma_{V^h} = h_2 \Gamma_D + \Gamma_{D'}$$

Set them equal to 0 to get

$$h_2 = -\frac{\Gamma_{D'}}{\Gamma_D}, \qquad h_1 = \frac{\Gamma_{D'}}{\Gamma_D} \Delta_D - \Delta_{D'}$$

That is, if you want to hold D' in your portfolio but does not want to be sensitive to the change in underlying assets price up to second order, you need to add (h_1, h_2) many (asset, D) in your portfolio.

5.5 Multiple asset case

Now, we will carry out a similar analysis when there are N assets and dynamics are given by the SDE (5.7).

Remark 5.27.

(i): The definitions 5.9, 5.15, 5.17, and assumptions 5.11, 5.13 are independent of the number of assets. Moreover, as we consider a scalar value process with a single risk-free asset, proposition 5.12 and lemma 5.16 are also independent of number of assets.

(ii): In section 3, not only Itô formula has multidimensional version, but Martingale representation theorem and Girsanov's theorem also extends to d-dimensional Brownian motion in a straightforward manner. Moreover, the discussion in section 4 carried out in one dimensions, but by using the multidimensional Itô formula, we can get the multidimensional Feynman-Kac formula.

We will observe that the only notable difference between the scalar case is due to the volatility. In the one-dimensional case, it was enough to assume the volatility is non-zero to define the Sharpe ratio. Here, we assume that there exists an adapted stochastic process θ such that

$$\sigma_t \theta_t = -(\alpha_t - r_t \mathbf{1}) \tag{5.13}$$

holds almost surely, where $\mathbf{1}$ is the N-dimensional vector with all the entries equal to 1. In other words, $-(\alpha_t - r_t \mathbf{1}) \in \mathrm{Im}(\sigma_t)$ almost surely. Let

$$M_t^\theta := \exp\Big(\int_0^t \theta_s \cdot dW_s - \frac{1}{2}\int_0^t |\theta_s|^2 ds \Big), \quad d\mathbb{P}^\theta := M_T^\theta d\mathbb{P}, \quad W_t^\theta := W_t - \int_0^t \theta_s ds$$

By the Girsanov's theorem, W^θ is a standard Brownian motion under $\mathbb{P}^\theta$. Then, the dynamics of the price processes are given by

$$dS_t = S_t \circ \alpha_t dt + S_t \circ \sigma_t (dW_t^\theta + \theta_t dt) = r_t S_t dt + S_t \circ \sigma_t dW_t^\theta$$

That is, discounted price processes are martingale under $\mathbb{P}^\theta$. We call such θ admissible, if it satisfies the conditions of Girsanov's theorem. That is, $\|\theta\|_2 < \infty$ and M^θ is a martingale. Furthermore, $\lambda := -\theta$ is called the market prices of risk. Note that the right hand side of (5.13) is the excess return of assets over the risk-free rate, and in particular for $i \in \{1, \ldots, N\}$,

$$\alpha_t^i - r_t = \sum_{j=1}^d \sigma_t^{i,j} \lambda_t^j$$

Here, important point is that the same λ determines the excess return of the asset, independent of i.

Altough we introduced the multi asset case with N many assets and d many Brownian motions, we will assume $N = d$ and then σ becomes a square matrix. Then, we will require σ to be nonsingular, or $\mathrm{IM}(\sigma_t) = \mathbb{R}^N$. In the next section, we will discuss the underlying reasons for such choice.

For notational purposes, recall the discussion around the Hadamard product (5.8). We again remark that, up to multidimensional notations and above discussion, Theorem 5.14 & 5.18 are exactly the same.

Theorem 5.28 (**Black-Scholes Equation**). *Consider a simple claim $\mathcal{X} = \Phi(S_T)$ under the Black-Scholes model.*[24] *Suppose Assumptions 5.11 and 5.13 holds. Moreover, we assume that the price of the contract is determined by $F \in C^{1,2}([0,T] \times \mathbb{R}^N; \mathbb{R})$ where*

$$\Pi(t; \Phi) = F(t, S_t)$$

Then, F is the classical solution to the Black-Scholes PDE

$$\Big[\partial_t F + r x^\top \nabla_x F + \frac{1}{2}\mathrm{tr}\,\Big((x \circ \sigma)(x \circ \sigma)^\top \nabla_{xx}^2 F\Big) - rF\Big](t,x) = 0 \quad \textit{with terminal} \quad F(T,x) = \Phi(x)$$

where $\nabla_x F = [\partial_{x_1} F, \ldots, \partial_{x_N} F]^\top$ and similarly $\nabla_{xx}^2 F$ is the Hessian of F.

[24] In the multidimensional case, the Black-Scholes model still refers to the case where (r, α, σ) are constants.

Proof. Let us consider a self-financing portfolio V formed by k_t many (stochastic process) underlying assets S and 1 simple claim $\mathcal{X}$. Recall that, we need to keep track of accounting for the risk-free asset to ensure the porfolio is self-financing, however, we won't keep track in the multivariate case.

Now, we can write dV by self-financing property and Itô's formula:

$$\begin{aligned} dV &= dF(t, S_t) + k_t dS_t \quad \text{(self-financing)} \\ &= \Big[\partial_t F + \nabla_x F^\top (S_t \circ \alpha) + \tfrac{1}{2}\mathrm{tr}\Big((S_t \circ \sigma)(S_t \circ \sigma)^\top \partial_{xx} F\Big) + k_t^\top (S_t \circ \alpha)\Big](t, S_t)dt \\ &\quad + \Big[\nabla_x F^\top (S_t \circ \sigma) + k_t^\top (S_t \circ \sigma)\Big](t, S_t) dW_t \end{aligned}$$

Now, we set

$$k_t := -\nabla_x F$$

which is essentially the so called delta-hedging. Then,

$$dV = \Big[\partial_t F + \tfrac{1}{2}\mathrm{tr}\Big((S_t \circ \sigma)(S_t \circ \sigma)^\top \partial_{xx} F\Big)\Big](t, S_t)dt$$

It is important to note that at this step α_t dissapears from analysis. By Proposition 5.12,

$$\partial_t F + \tfrac{1}{2}\mathrm{tr}\Big((S_t \circ \sigma)(S_t \circ \sigma)^\top \partial_{xx} F\Big) = rV(t, S_t) = r(F - \nabla_x F^\top S_t)(t, S_t)$$

Rearrange the terms to get:

$$\Big[\partial_t F + r\nabla_x F^\top S_t + \tfrac{1}{2}\mathrm{tr}\Big((S_t \circ \sigma)(S_t \circ \sigma)^\top \partial_{xx} F\Big) - rF\Big](t, S_t) = 0$$

which holds $dt \times d\mathbb{P}$-almost surely. Since S_t has full range, we conclude the result. ■

Theorem 5.29 (Risk Neutral Valuation). *Suppose Assumptions 5.11 and 5.13 holds. Consider the generalized Black-Scholes model:*

$$dB_t = r_t B_t dt, \qquad dS_t = S_t \circ \alpha_t dt + S_t \circ \sigma_t dW_t \tag{5.14}$$

where r_t, α_t and σ_t are any adapted processes, and $\sigma \in \mathbb{R}^{N \times N}$ is nonsingular $dt \times d\mathbb{P}$-almost surely.
(i): Generalized Black-Scholes model (5.14) is complete.
(ii): For any contingent claim $\mathcal{X} \in \mathcal{F}_T^S$ written on S, the price is given by

$$\Pi(t; \mathcal{X}) = B_t \mathbb{E}^\theta\Big[\big(\mathcal{X}/B_T\big)\big|\mathcal{F}_t^S\Big] \tag{5.15}$$

where $\mathbb{E}^\theta$ is taken under risk-neutral probability measure $\mathbb{P}^\theta$ for which

$$dS_t = r_t S_t dt + S_t \circ \sigma_t dW_t^\theta, \qquad \text{where } W^\theta \text{ is } \mathbb{P}^\theta\text{-Brownian motion}$$

(iii): If we further assume the structure in Theorem 5.28, representation of F is given by

$$F(t, x) := e^{-r(T-t)}\mathbb{E}\big[\Phi(X_T^{t,x})\big] \tag{5.16}$$

where

$$dX_s^{t,x} = rX_s^{t,x} ds + X_s^{t,x} \circ \sigma\, dW_s, \quad X_t^{t,x} = x \tag{5.17}$$

Proof. (iii) is exactly the Feynman-Kac Theorem 4.5 (after generalizing to multidimensions) by setting

$$f = 0, \quad \alpha(t,x) = rx, \quad \sigma(t,x) = x \circ \sigma, \quad \text{and } \; u = F$$

(i) and (ii) will follow by the construction of the replicating portfolio. As we already discussed, relying on nonsingular assumption, define

$$\theta_t = -\sigma_t^{-1}(\alpha_t - r_t \mathbf{1}), \qquad M_t^\theta := \exp\Big(\int_0^t \theta_s \cdot dW_s - \frac{1}{2}\int_0^t |\theta_s|^2 ds\Big)$$

and $d\mathbb{P}^\theta := M_T d\mathbb{P}$, $W_t^\theta := W_t - \int_0^t \theta_s ds$. By Girsanov's Theorem 3.10, W_t^θ is a $\mathbb{P}^\theta$ Brownian motion. Recall,

$$dS_t = r_t S_t dt + S_t \circ \sigma_t dW_t^\theta$$

Now, define $V_t := B_t \mathbb{E}^\theta[\mathcal{X}/B_T | \mathcal{F}_t^S]$.[25] For the moment, assume that the process S generates the same filtration as W^θ. Then, by Martingale Representation Theorem 3.9, there exists unique L^2 stochastic process $Z_t \in \mathbb{R}^{1\times N}$ such that

$$V_t/B_t = \mathbb{E}^\theta[\mathcal{X}/B_T] + \int_0^t Z_s dW_s^\theta$$

Now, obviously if we can write V_t as a self-financing portfolio, it will replicate the contract, and this will also conclude (ii) by Lemma 5.16. To do so, let us recall the notation $S_t^{-1} = [1/S_t^1, \ldots, 1/S_t^N]$ to carry out the computation, and we simply write the self-financing condition to notice the appropriate portfolio:

$$\begin{aligned} dV_t &= (V_t/B_t)dB_t + B_t Z_t\big((\sigma_t^{-1} \circ S_t^{-1})(S_t \circ \sigma_t)dW_t^\theta\big) \\ &= (V_t/B_t)dB_t + B_t Z_t(\sigma_t^{-1} \circ S_t^{-1})\Big(dS_t - \frac{S_t}{B_t}dB_t\Big) = \Big(\frac{V_t}{B_t} - Z_t\sigma_t^{-1}\mathbf{1}\Big)dB_t + B_t Z_t(\sigma_t^{-1} \circ S_t^{-1})\, dS_t \end{aligned}$$

Then the portfolio has to be

$$h_t^0 := \Big(\frac{V_t}{B_t} - Z_t\sigma_t^{-1}\mathbf{1}\Big) \in \mathbb{R}, \qquad (h_t^1)^\top := B_t Z_t(\sigma_t^{-1} \circ S_t^{-1}) \in \mathbb{R}^N$$

Let us verify that

$$\begin{aligned} V_t^h &= \big[V_t/B_t - Z_t\sigma_t^{-1}\mathbf{1}\big]B_t + \big[B_t Z_t(\sigma_t^{-1} \circ S_t^{-1})\big]S_t \\ &= V_t - B_t Z_t \sigma_t^{-1}\mathbf{1} + B_t Z_t \sigma_t^{-1}\mathbf{1} = V_t \end{aligned}$$

and hence we conclude the results.

Lastly, let us discuss the filtrations. First, just to note, it is obvious that $\mathbb{F}^W = \mathbb{F}^{W^\theta}$. It is also clear that $\mathbb{F}^S \subset \mathbb{F}^{W^\theta}$. We need to write W^θ as a function of S:

$$(\sigma_t^{-1} \circ S_t^{-1})dS_t - r_t\sigma_t^{-1}\mathbf{1}dt = dW_t^\theta$$

where we can define $\int dS_t$ exactly as $\int dW_t$. ■

[25] Here $\mathbb{E}^\theta$ denotes the expectation under the measure $\mathbb{P}^\theta$.

5.6 Fundamental Theorems of Asset Pricing

In this section, we will closely follow the discussion given in [5]. Our main interest is to understand the fundamental theorems characterizing the completeness and arbitrage opportunities in the market.

But first, let us briefly mention a common rule of thumb for market completeness and arbitrage opportunities in our model with multiple assets, sometimes referred to as a Meta Theorem. Recall that we assume $N = d$, that is, the number of assets equals the number of independent sources of noise. Under the assumptions we have discussed, this is a typical case where the market is both complete and arbitrage-free.

If there are more assets than sources of noise ($N > d$), the market may still remain complete, meaning that all contingent claims can be hedged.[26] However, there is a higher risk of encountering arbitrage opportunities. For example, in the extreme case where $d = 1$ and N is very large, either all assets share a single Sharpe ratio, or arbitrage opportunities exist. To see this in our proofs, recall the construction of $\mathbb{P}^\theta$ by the Girsanov's theorem. If $d = 1$, then (5.13) must be satisfied with a scalar θ.

On the other hand, if there are more sources of noise than assets ($N < d$), the market becomes incomplete, as it becomes difficult to hedge all possible contingent claims. However, it also becomes less likely to find arbitrage opportunities, since many sources of noise cannot be fully hedged. For example, in the extreme case where there is only a single asset driven by multiple Brownian motions, arbitrage opportunities are absent, but there may still be contingent claims that cannot be perfectly hedged by trading the single asset. In the proof of Risk Neutral Valuation, one will notice that $\mathbb{F}^S \subset \mathbb{F}^W$ is strict and thus we cannot invoke Martingale Representation Theorem.

Now, recall that while introducing the portfolio dynamics, we had $S = (S^0, S^1, \dots, S^N)$ where S^0 represents the risk-free asset. We assume that $S^0 > 0$ almost surely, and change our perspective to so called normalized economy:

Definition 5.30. The normalized economy (also referred as Z-economy) is defined by the price process Z given by

$$Z_t = \frac{S_t}{S_t^0} = \left(1, \frac{S_t^1}{S_t^0}, \dots, \frac{S_t^N}{S_t^0}\right)$$

Recall that, given a portfolio $h_t = (h_t^0, h_t^1, \dots, h_t^N)$, we defined the value $V_t^h = h_t^\top S_t$. To distinguish it from the normalized economy, we will call this value as S-value of h, and might denote it as V^S if h is clear and we need to distinguish it from the normalized economy. Then, we call $V_t^Z := h_t^\top Z_t$ as the Z-value of h, and say that h is Z-self-financing portfolio if $dV_t^Z = h_t dZ_t$. We then have an equivalence between these two economies:

Lemma 5.31 (Invariance Lemma)**.**

(i) h is a self-financing portfolio if and only if it is a self-financing portfolio in the normalized economy.

(ii) For any portfolio h, $S_t^0 V_t^Z = V_t^S$.

(iii) A contingent claim is $\mathcal{X}$ is replicable if and only if $\mathcal{X}/S_T^0$ is replicable in the normalized economy.

Proof. (ii) and (iii) are obvious. To argue (i), first note that, since the risk-free asset has a finite total variation,

$$dS_t = d(S_t^0 Z_t) = d(S_t^0)Z_t + S_t^0 dZ_t$$

[26] If there are arbitrage opportunities in the market, it is not that meaningful to discuss completeness.

Then, assuming that $dV_t^S = h_t dS_t$,

$$
\begin{aligned}
dV_t^Z &= d((S_t^0)^{-1}V_t^S) \\
&= d((S_t^0)^{-1})V_t^S + (S_t^0)^{-1}dV_t^S \\
&= -(S_t^0)^{-2}d(S_t^0)V_t^S + (S_t^0)^{-1}h_t dS_t \\
&= -(S_t^0)^{-1}d(S_t^0)V_t^Z + (S_t^0)^{-1}h_t d(S_t^0)Z_t + (S_t^0)^{-1}h_t S_t^0 dZ_t \\
&= -(S_t^0)^{-1}d(S_t^0)V_t^Z + (S_t^0)^{-1}d(S_t^0)V_t^Z + (S_t^0)^{-1}h_t S_t^0 dZ_t = h_t dZ_t
\end{aligned}
$$

and hence we conclude $dV_t^Z = h_t dZ_t$. Similarly, if we start by assuming h is self-financing in the normalized economy, then

$$dV_t^S = d(S_t^0 V_t^Z) = d(S_t^0)V_t^Z + S_t^0 dV_t^Z = d(S_t^0)V_t^Z + S_t^0 h_t dZ_t = d(S_t^0)V_t^Z + h_t(dS_t - d(S_t^0)Z_t)$$

which then concludes the result by noting $h_t Z_t = V_t^Z$. ■

Relying on this invariance, from now on, we will always assume $S_t^0 = 1$. So far, we have not really restricted our set of portfolios, but only assumed regularities and adaptedness, which only means it depends on the current information and not the future.[27] However, this class is far too large and not realistic. For example, we can consider a strategy similar to the classic doubling down by borrowing more and more money, betting higher and higher on a risky asset if we loose. Such strategies are guaranteed to win (with probability 1), but expected time to win and hence the amount needed to be borrowed is not finite. To avoid such portfolios, we have the following definition;

Definition 5.32. We say a portfolio h is admissible, if

$$\int_0^t h_s dS_t \geq -C^h > -\infty$$

almost surely for some constant C^h that depends on h.

Note that, when h is self-financing, admissibility means $V_t^h - V_0^h \geq -C^h$. That is, your potential loses are not infinite. To get familiar with the terms, let us also make the following definition

Definition 5.33. We say Q is an equivalent (local) martingale measure (martingale measure, EMM) if

(i) $Q \sim \mathbb{P}$, that is, Q is equivalent to $\mathbb{P}$, and

(ii) the price process S is a (local) martingale under Q.

In our case, we already constructed the EMM by Girsanov's theorem. Note that, in the normalized economy,

$$
\begin{aligned}
dZ_t &= d((S_t^0)^{-1}S_t) \\
&= -(S_t^0)^{-2}d(S_t^0)S_t + (S_t^0)^{-1}dS_t \\
&= -(S_t^0)^{-1}d(S_t^0)Z_t + (S_t^0)^{-1}(r_t S_t dt + S_t \circ \sigma_t dW_t^\theta) \\
&= -r_t Z_t dt + r_t Z_t dt + Z_t \circ \sigma_t dW_t^\theta
\end{aligned}
\qquad \left[dZ_t = Z_t \circ \sigma_t dW_t^\theta \right]
$$

That is, Z is a martingale. Also, recall that $\mathbb{P}^\theta(A) := \mathbb{E}[M_T \mathbf{1}_{\{A\}}]$. Hence, $\mathbb{P}(A) = 0 \iff \mathbb{P}^\theta(A) = 0$. This concludes $\mathbb{P}^\theta$ is an EMM for the normalized economy.

We are ready to start collecting the main results:

[27] When we integrate portfolio h with respect to semimartingale price processes, we require h to be predictable. Again, we avoid such discussions in this course.

Lemma 5.34. *Consider the N-asset market defined in section 5.2. Due to the existence of an equivalent martingale measure, there cannot be any arbitrage portfolio.*

Proof. To contradict, suppose there exists an arbitrage portfolio h. That is,

$$V_0^h = 0, \qquad \mathbb{P}(V_T^h \geq 0) = 1, \qquad \mathbb{P}(V_T^h > 0) > 0$$

By definition of arbitrage, h has to be an self-financing portfolio. Moreover, by the invariance lemma, we may assume price processes are already discounted. Then, it follows

$$dV_t^h = h_t \cdot dS_t = h_t^\top (S_t \circ \sigma_t) dW_t^\theta$$

We assumed the portfolio value $V_t^h = h_t^\top S_t$ is in L^2, and since σ is bounded, V_t^h is a martingale. We are done, because

$$0 < \mathbb{E}^{\mathbb{P}^\theta}[V_T^h] = \mathbb{E}^{\mathbb{P}^\theta}[V_0^h] = 0$$

which is a contradiction. ■

For our purposes, since we already have the existence of EMM in our framework, Lemma 5.34 is the First Fundamental Theorem of Asset Pricing. Under appropriate generalization of arbitrage, these conditions are equivalent in a considerably more general setting:

Theorem 5.35 (First Fundamental Theorem of Asset Pricing). *Let the price process S be a locally bounded real valued semi-martingale. Then, there exists an equivalent martingale measure if and only if (NFLVR) condition holds.*

Now, we will explore the definitions in this theorem. First, in the context of stochastic processes, "local" means that the corresponding property holds only up to some random time τ^n, but for which $\tau^n \uparrow T$ as $n \to \infty$. Therefore, locally bounded means $\hat{S}_t := S_{\tau^n \wedge t}$ are bounded, and it is quite general. For example, all continuous adapted processes are locally bounded, with $\tau^n := \inf\{t > 0 : |S_t| \geq n\}$. Processes that exhibits jumps are also locally bounded, as long as jump sizes are bounded. Thus, locally bounded price process is quite a general assumption.

Next, semi-martingale means that S is of the form

$$S_t = A_t + M_t$$

where M is a local martingale, and A is a cádlág[28] adapted process with locally bounded variation. This class is also quite general, and it is important to note that this is the largest class where Itô integrals are defined. Due to this generality, even if the market is not well-described by our framework in section 5.2, equivalence between existence of a martingale measure and no arbitrage condition still holds true in the markets.

To explain the more general arbitrage condition that the theorem requires, we first introduce some spaces of functions. First,

$$\mathcal{K} := \left\{ \int_0^T h_t dS_t \ : \ \text{for all admissible } h_t \right\}$$

is the space of contingent claims reachable by portfolios with 0 initial value. Next,

$$\mathcal{C} := \left\{ g \in L^\infty(\mathcal{F}_T^S)^{29} : \text{ there exists } f \in \mathcal{K} \text{ s.t. } g \leq f \right\}$$

[28] Cádlág process means left limits exists and right continuous, almost surely.

[29] $L^\infty(\mathcal{F}_T^S)$ is the space of almost surely bounded functions, measurable to $\mathcal{F}_T^S = \sigma(S_t : 0 \leq t \leq T)$.

is the set of bounded contingent claims dominated by reachable portfolios with 0 initial value. The reason to consider $\mathcal{C}$ is to generalize our market and include contingent claims (or contracts) into consideration. A contingent might not be exactly achievable by a replicating portfolio (only with underlying assets), but might be dominated by one. Then, a more general no-arbitrage condition including any bounded contingent claim is given by

$$\mathcal{C} \cap L^{\infty,+}(\mathcal{F}_T^S) = \{0\}$$

That is, there is no bounded contingent claim with only positive payoff, dominated by a reachable portfolio with no initial investment. Then, we have a slight generalization as follows:

Definition 5.36. We say that the market has no free lunch with vanishing risk (NFLVR) if

$$\bar{\mathcal{C}} \cap L^{\infty,+}(\mathcal{F}_T^S) = \{0\}$$

where $\bar{\mathcal{C}}$ is the closure $\mathcal{C}$ in $L^{\infty}(\mathcal{F}_T^S)$.[30]

To understand the condition, suppose it does not hold. Then there is a non-zero contingent claim with positive values. Then, for any nearby contingent claim in $L^{\infty}(\mathcal{F}_T^S)$, we can find a portfolio in $\mathcal{K}$ dominating this contingent claim. It does not require that these portfolios are positive, however, as their limit dominates a positive contingent claim, 'risk is vanishing'. And since the contingent claim is not exactly 0, limit of the portfolios becomes an arbitrage.

We are not interested in the technical details of the proof, but the overall scheme goes as follows: If there exists an equivalent martingale measure, and hence the price processes are martingales under this measure, as long as the portfolios generated are also martingales we are done as in lemma 5.34. In general settings, this might not be the case and one might need to argue supermartingale property by relying on admissibility of portfolios. Now, the difficult part is to start by assuming NFLVR condition. One needs to study convex analysis, and when convex sets can be seperated by hyperplanes. Hahn-Banach is the central one, but here one needs a more technical result by Kreps-Yan. By noting that $\mathcal{C}$ and $L^{\infty,+}$ are convex sets with only common function 0, there exists an integrable, strictly positive function ξ such that

$$\mathbb{E}[\xi X] \leq 0 \ \ \forall X \in \mathcal{C}, \ \text{ and } \ \mathbb{E}[\xi X] \geq 0 \ \ \forall X \in L^{\infty,+}$$

Here, ξ seperates these two convex sets of functions. Given that one can rigorously show this, the rest of the proof is easier to handle. Define $dQ := \xi d\mathbb{P}$, and note that this is a measure by 1.12. Since ξ is integrable, after a scaling, it is a probability measure. This is our candidate equivalent martingale measure. Equivalency follows by strict positivity of ξ. Now, take any $s < t$, $A \in \mathcal{F}_s^S$, and define the self-financing portfolio

$$h(r, \omega) := \mathbf{1}_{\{[s,t]\}}(r)\mathbf{1}_{\{A\}}(\omega)(-S_s^i(\omega), 0, \dots, 0, 1, 0, \dots, 0)$$

where 1 appears at the ith place. In words, we buy 1 asset S^i at time s if A occurs, and finance it by borrowing money (at zero interest by invariance lemma), and close those positions back at time t. Now, ignoring the admissibility of the portfolio h, $\pm V_t^h \in \mathcal{K}$. Also, note that as $\mathbb{E}[\xi X] \leq 0$ for all $X \in \mathcal{C}$, we have $\mathbb{E}[\xi X] \leq 0$ for all $X \in \mathcal{K}$. Since $\pm V_t^h \in \mathcal{K}$ ("linear subspace"), this implies $\mathbb{E}[\xi V_t^h] = 0$. Writing this explicitly,

$$\mathbb{E}\left[\xi \int_0^T h_t dS_t\right] = \mathbb{E}^Q\left[\int_0^T h_t dS_t\right] = \mathbb{E}^Q[(S_t^i - S_s^i)\mathbf{1}_{\{A\}}] = 0$$

Since $A \in \mathcal{F}_s$ is arbitrary, this is an equivalent way of defining martingale, and we conclude the price processes are martingales.

We now move to discuss the characterization of completeness (see definition 5.17) of the market.

[30] Closure of $\mathcal{C}$ is taken under the topology generated by the operator norm, where L^{∞} functions acts on L^1 functions.

Theorem 5.37 (**Second Fundamental Theorem of Asset Pricing**)**.** *Suppose there exists an equivalent local martingale measure Q. Then, the market is complete if and only if Q is the unique local martingale measure.*

Suppose the market is complete. Then for any $\mathcal{X} \in \mathcal{F}_T^S$, there exists a self-financing portfolio h with $V_T^h = \mathcal{X}$. By assuming sufficient regularities (to avoid discussions around 'local'), since $V_t^h - V_0^h = \int_0^t h_s dS_s$ is a stochastic integral with respect to a martingale, V_t^h is also a martingale. Therefore,

$$V_0^h = \mathbb{E}^Q[V_T^h] = \mathbb{E}^Q[\mathcal{X}]$$

Suppose there exists an another equivalent martingale measure $\tilde{Q}$. Consider any set $E \in \mathcal{F}_T$ and set $\mathcal{X} = \mathbf{1}_{\{E\}}$. By above, and since the market is complete,

$$Q(E) = \mathbb{E}^Q[\mathcal{X}] = V_0^h = \mathbb{E}^{\tilde{Q}}[\mathcal{X}] = \tilde{Q}(E)$$

That is, Q is unique.

Now, suppose Q is unique and take a contingent claim $\mathcal{X} \in \mathcal{F}_T^S$. Then, by a generalization of Martingale Representation Theorem that holds for any (local) martingale, there exists Z such that

$$\mathcal{X} = \mathbb{E}^Q[\mathcal{X}] + \int_0^T Z_t dS_t$$

and we are done as Z provides the replicating portfolio. Here, uniqueness is used to invoke the representation theorem.

6 A Tour in Stochastic Optimal Controls

In countless settings, whenever there exists an agent aiming to decide on various actions, it is crucial to understand how to behave optimally. In this section, we will explore situations where the underlying dynamics are characterized by SDEs. We will apply these concepts to Merton's problem, where a financial advisor aims to decide how much risk to take for a retiring customer. It is important to note that there is ongoing cutting-edge research in this field, where both the dynamics and objectives are evolving, such as in reinforcement learning.

Let us first introduce the dynamics given by the SDE $X^{t,x,\pi} = X^{\pi}$:

$$dX_t^{\pi} = \mu(t, X_t^{\pi}, \pi(t, X^{\pi}))dt + \sigma(t, X_t^{\pi}, \pi(t, X^{\pi}))dW_t \tag{6.1}$$

The core addition is that there exists a space called action space, denote it as $\mathbb{A}$, where the agent has choices to make. We take $\pi : [0,T] \times C([0,T]; \mathbb{R}^d) \to \mathbb{A}$ as the admissible control of the agent whenever (6.1) is well-posed, and denote $\mathcal{A}$ as the set of all such controls. For example, the agent might be constructing a portfolio and amounts of financial assets are the choices of the agent. Note that, we let the controls to depend on the whole process X, which should also satisfy adaptedness in the sense that $\pi(t,x)$ can only depend on $x_{[0,t]}$.

To find what is the optimal control, one needs to set the preferences. What we know is that, given π, we have a flow of distributions $\{\mathcal{L}_{X_t^{\pi}}\}_{0\le t\le T}$ for the evolution of the dynamics. Then, the agent should decide on which states are prefered, in order to determine which π is the best. To do so, the agent assigns two functions $F : [0,T] \times \mathbb{R}^d \times \mathbb{A} \to \mathbb{R}$ and $G : \mathbb{R}^d \to \mathbb{R}$ as running cost and terminal cost, which orders the states. Moreover, running cost incorporates the potential costs of taken actions. For a different example, imagine that you are controlling a robot and trying to direct it to a particular position. In this case, it might be costly to accelerate and agent might aim to minimize energy consumption by penalizing it through the running cost. Now, we assign the cost of the control π as

$$J(t,x,\pi) := \mathbb{E}\Big[G(X_T^{\pi}) + \int_t^T F(s, X_s^{\pi}, \pi(s, X^{\pi}))ds\Big], \qquad \text{where } X^{\pi} = X^{t,x,\pi}$$

Note that, minimizing or maximizing is essentially the same problem up to a scalar constant, and we will continue with maximization problem. Thus, the value[31] of the problem is

$$V(t,x) = \sup_{\pi\in\mathcal{A}} J(t,x,\pi) \tag{6.2}$$

An important property of V is that it is time-consistent. It is typically called Dynamic Programming Principle (DPP), but the origin of the name is not due to mathematical concern. Let us state it first and then discuss:

Theorem 6.1 (**Time Consistency**). *For any $t \le T_0 \le T$, it holds*

$$V(t,x) = \sup_{\pi} \mathbb{E}\Big[V(T_0, X_{T_0}^{\pi}) + \int_t^{T_0} F(s, X_s^{\pi}, \pi(s, X^{\pi}))ds\Big]$$

where $X^{\pi} = X^{t,x,\pi}$.

Note that the left hand side $V(t,x)$ is the original problem starting from t ending at T. On the right, we have the same optimization problem again starting from t, but ends at a middle time T_0 with the terminal

[31] This is typically called the value, as there is no conceptual difference between minimizing and maximizing; we use whichever term is more suitable.

cost taken as $V(T_0,\cdot)$. This is the crucial property that will allow us to obtain the PDE approach. Note that, one can solve the optimization problem $[T-\delta,T]$ first, and then use the value at $T-\delta$ as a terminal to solve $[T-2\delta,T-\delta]$ and so on.

Before the proof, let us mention the core idea of it. Easier direction says the optimal control for $[t,T]$ problem is also optimal for appropriately defined $[t,T_0]$ and $[T_0,T]$ problem. Direction that is harder to obtain is if we have optimal controls for $[t,T_0]$ and $[T_0,T]$, we can obtain an optimal control for $[t,T]$ problem.

Proof. We will sketch the main ideas only. First, we observe that

$$\begin{aligned} J(T_0, X^\pi_{T_0}, \pi) &= \mathbb{E}\Big[G(X_T^{T_0, X^\pi_{T_0}, \pi}) + \int_{T_0}^T F(\cdots)ds\Big] \\ &= \mathbb{E}\Big[G(X^\pi_T) + \int_{T_0}^T F(s, X^\pi_s, \pi(s, X^\pi))ds | \mathcal{F}^X_{T_0}\Big] \end{aligned} \tag{6.3}$$

To see this, we remark that

$$X_T^{T_0, X^\pi_{T_0}, \pi} = X_T^{t,x,\pi} =: X^\pi_T$$

and the first expectation in (6.3) is not taken over $X^\pi_{[0,T_0]}$, which is given as it appears in $J(T_0, X^\pi_{T_0}, \pi)$. To point out a subtle point, from generalization of Doob-Dynkin, we know that the right-hand side of (6.3) is a function of $X_{[0,T_0]}$. Similarly, we need to be careful about the left-hand side and note that the control π depends on the path. Thus, given a realization of $X_{[0,T_0]}$, π on the left-hand side fixes this path as given, and defined as a control on $[T_0,T]$, which is not clear from our notations.

Given this, one direction of the proof is much simpler. Note that, due to the tower-property of conditional expectations and (6.3),

$$\begin{aligned} J(t,x,\pi) &= \mathbb{E}\Big[J(T_0, X^\pi_{T_0}, \pi) + \int_t^{T_0} F(s, X^\pi_s, \pi(s, X^\pi))ds\Big] \\ &\leq \mathbb{E}\Big[V(T_0, X^\pi_{T_0}) + \int_t^{T_0} F(s, X^\pi_s, \pi(s, X^\pi))ds\Big] \end{aligned}$$

Take the supremum over all π to obtain one side of the result.

It requires a lot more technical details to complete the other direction. Instead, we will assume that there exists an optimal control π^x for the optimization problem in $[T_0,T]$, starting from any x. That is,

$$V(T_0,x) = J(T_0,x,\pi^x) \tag{6.4}$$

Then, given any control π, we introduce

$$\hat{\pi}(t,x) := \pi(t,x)\mathbf{1}_{\{t<T_0\}} + \pi^{x_{T_0}}(t,x)\mathbf{1}_{\{t\geq T_0\}},$$

Recall that $x \in C([0,T];\mathbb{R}^d)$ when it appears in π, and this construction of $\hat{\pi}$ is necessarily path dependent. Without any technical considerations, we assume that $\hat{\pi}(t,x)$ is admissible. It is not quite clear that we can form this, as there is no apriori regularity of which optimal control we are chosing for each x_{T_0}.[32] Now, by definition of the value V and (6.4),

$$\begin{aligned} V(t,x) \geq J(t,x,\hat{\pi}) &= \mathbb{E}\Big[J(T_0, X^\pi_{T_0}, \pi^{X^\pi_{T_0}}) + \int_t^{T_0} F(s, X^\pi_s, \pi(s, X_s))ds\Big] \\ &= \mathbb{E}\Big[V(T_0, X^\pi_{T_0}) + \int_t^{T_0} F(s, X^\pi_s, \pi(s, X_s))ds\Big] \end{aligned}$$

[32]See *measurable selection* as a keyword, or one can construct $\hat{\pi}$ with discretization of space if there are sufficient regularities.

Notice that, we are again seperating $J(t, x, \hat{\pi})$ into $[t, T_0]$ and $[T_0, T]$ where in the first region we are using arbitrary π and in the second region we are using the optimal control $\pi^{X^{\pi}_{T_0}}$. ■

Now, given this time-consistency, we will derive a PDE for the value function $V(t, x)$. It is called Hamilton-Jacobi-Bellman (HJB) equation, and is given as follows:

$$\partial_t V + \sup_{a \in \mathbb{A}} \left[\mu \partial_x V + \tfrac{1}{2}\sigma^2 \partial_{xx} V + F\right](t, x, a) = 0, \qquad V(T, x) = G(x) \tag{6.5}$$

Let us briefly discuss what is going on. First of all, terminal value has to be the terminal cost G by definition. Next, the equation consists of two parts: the first from the Itô formula, representing the change in value with respect to the state process, and the second, the running (immediate) cost F. It is in a sense obvious, but important fact that we are not finding the optimal action that only maximing the running (or immediate) cost F. Instead, we should consider how our actions would change our value V due to the change of the underlying dynamics, and choose the supremum accordingly.

We do care about the value function because if we can solve it by any method and obtain $V(t, x)$, we can reduce our optimization problem to a static one and construct optimal π^*. This is the important message of this section. For this purpose, we rewrite the HJB equation and separate the optimization part. Introduce the Hamiltonian $\mathcal{H}$ as follows:

$$h(t, x, z, \gamma, a) := \mu(t, x, a) z + \frac{1}{2}\sigma^2(t, x, a)\gamma + F(t, x, a), \qquad \mathcal{H}(t, x, z, \gamma) = \sup_{a \in \mathbb{A}} h(t, x, z, \gamma) \tag{6.6}$$

which isolates the optimization independent of $\partial_x V, \partial_{xx} V$, but for any given z, γ instead. Then, the HJB equation reads

$$\partial_t V(t, x) + \mathcal{H}(t, x, \partial_x V(t, x), \partial_{xx} V(t, x)) = 0$$

Theorem 6.2. *Let V be defined as in (6.2). Then,*

(i) If $V \in C^{1,2}$, then V satisfies the HJB equation (6.5).

(ii) If $U \in C^{1,2}$ solves the HJB equation, then $U = V$.

(Verification) Suppose $U \in C^{1,2}$ solves the HJB equation, and there exists a Borel measurable function I such that

$$h(t, x, z, \gamma, I(t, x, z, \gamma)) = \mathcal{H}(t, x, z, \gamma)$$

and the following SDE has a strong solution:

$$X^*_t = x_0 + \int_0^t \mu(s, X^*_s, I(s, X^*_s, \partial_x U(s, X^*_s), \partial_{xx} U(s, X^*_s)))ds + \int_0^t \sigma(\cdots) dW_s$$

Then, $\pi^*(t, x) := I(t, x, \partial_x V, \partial_{xx} V)$ is an optimal control for $V(0, x_0)$.

Proof. We will only sketch the proof, and omit (ii) but show it under the assumption (iii).

Now, suppose $V \in C^{1,2}$. Note that, DPP or time-consistency is

$$\begin{aligned} 0 &= \inf_{\pi} \mathbb{E}\left[V(t + \delta, X^{\pi}_{t+\delta}) - V(t, x) + \int_t^{t+\delta} F(s, X^{\pi}_s, \pi(s, X^{\pi}))ds \right] \\ &= \inf_{\pi} \mathbb{E}\left[\int_t^{t+\delta} \left[\partial_t V + \mu \partial_x V + \tfrac{1}{2}\sigma^2 \partial_{xx} V + F\right](s, X^{\pi}_s, \pi)ds \right] \end{aligned}$$

where the last line follows from Itô's formula, and the martingale term is ignored under the expectation.Then, divide above by δ, send $\delta \to 0$ and replace $\sup_\pi$ with $\sup_a$.

Next, assume (iii) holds. Then, by Itô formula, for any π,

$$U(T, X_T^\pi) - U(0, x_0) = \int_0^T \left[\partial_t U + \mu \partial_x U + \tfrac{1}{2}|\sigma|^2 \partial_{xx} U\right] ds + \int_0^T \sigma \partial_x U dW_t$$

Note that the terminal is given $U(T, X_T^\pi) = G(X_T^\pi)$. Then, add the running cost on both sides, use that U solves the HJB equation and take the expectation to get:

$$U(0, x_0) - J(0, x_0, \pi) = -\mathbb{E}\left[\int_0^T \left[\partial_t U + \mu \partial_x U + \tfrac{1}{2}|\sigma|^2 \partial_{xx} U + F\right] ds\right] \geq 0$$

That is, $U(0, x_0) \geq V(0, x_0)$. Now, if we plug in π^*, inequality becomes equality. Then,

$$U(0, x_0) \geq V(0, x_0) \geq J(0, x_0, \pi^*) = U(0, x_0)$$

Notice that, we have shown V is the unique solution under the assumption that there exists an optimal control. Moreover, π^* is clearly an optimal control: $V(0, x_0) = J(0, x_0, \pi^*)$, which is state dependent. ■

Let us reiterate the important fact that finding the optimal control is a matter of solving the static optimization problem in (6.6) for any (t, x, z, γ). Then, if we know the value function V, optimal control is simply given by π^* as in the verification part.

6.1 Merton's Portfolio Problem

The Merton Problem addresses the allocation of a portfolio between a risky asset, such as the S&P 500, and a riskless asset, like money market funds. We will now investigate this problem under standard Black-Scholes model with a CRRA utility function, which enables the explicit computation of the optimal allocation. This subsection closely follows the lecture notes of Lacker [7].

Suppose our control π determines the relative propotion of our portfolio in the risky asset, and hence $1-\pi$ is in risk-free asset. That is, dynamics of our state process X^π representing the wealth of the portfolio is given by

$$dX_t^\pi = \frac{\pi_t X_t^\pi}{S_t} dS_t + \frac{(1-\pi_t) X_t^\pi}{B_t} dB_t = \left[\alpha \pi_t X_t^\pi + r(1-\pi_t) X_t^\pi\right] dt + \left[\sigma \pi_t X_t^\pi\right] dW_t$$

Suppose we do not have a running cost $F = 0$, but the terminal cost is given by

$$G(x) := \frac{x^{1-\eta}}{1-\eta}, \quad \eta > 0, \eta \neq 1$$

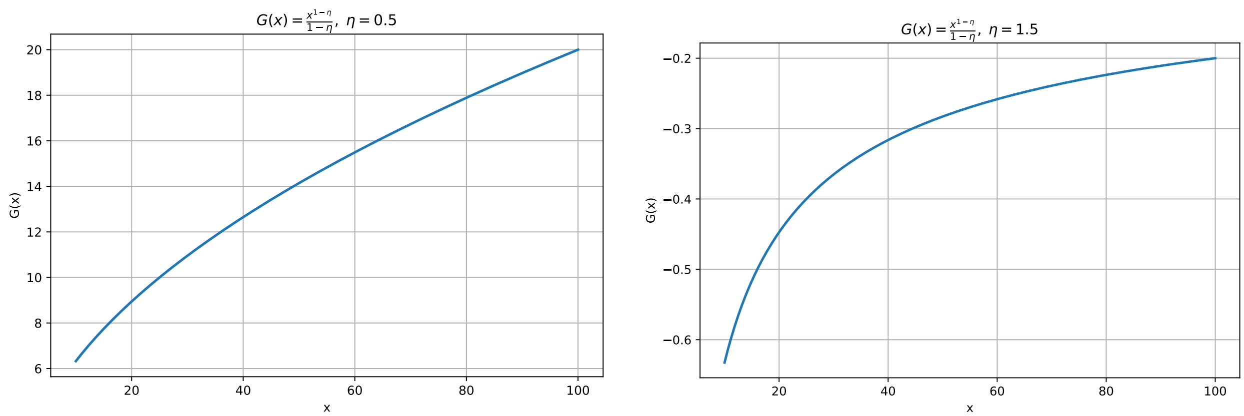
Note that $\partial_x G(x) = x^{-\eta}$, reflecting that as the wealth increases, change in utility decrease. Here, $\eta > 1$ represents risk-averse and $\eta < 1$ represents risk-seeking preferences, see Figure 4.

Let us write down the Hamiltonian (6.6):

$$\mathcal{H}(t, x, z, \gamma) = \sup_{a \in \mathbb{A}} \left[(\alpha\ a\ x + r(1-a)x)\ z + \tfrac{1}{2}|\sigma\ a\ x|^2 \gamma\right] = \sup_{a \in \mathbb{A}} \left[r\ x\ z + a(\alpha - r)xz + \tfrac{1}{2}(\sigma^2 x^2 \gamma)\ a^2\right]$$

We set $\mathbb{A} = \mathbb{R}$, allowing short selling and borrowing. Obviously if $\gamma > 0$, Hamiltonian becomes infinity. Hence, we assume $\gamma < 0$. Then, it is just a quadratic equation over a that we are trying to maximize. We can easily deduce

$$\mathcal{H}(t, x, z, \gamma) = rxz - \frac{(\alpha - r)^2 z^2}{2\sigma^2 \gamma}$$



(a) Risk-seeking

(b) Risk-averse

Figure 4: CRRA utility for different risk preferences.

where the maximizer is

$$I(t, x, z, \gamma) = -\frac{(\alpha - r)z}{\sigma^2 x \gamma}$$

Let us write down the (6.5):

$$\partial_t V(t, x) + rx\partial_x V(t, x) - \frac{1}{2}\left(\frac{\alpha - r}{\sigma}\right)^2 \frac{(\partial_x V)^2}{\partial_{xx} V} = 0 \tag{6.7}$$

To solve it, we make an ansatz as $V(t, x) = f(t)G(x)$. Then,

$$\partial_t V(t, x) = f'(t)G(x), \quad \partial_x V(t, x) = f(t)x^{-\eta}, \quad \partial_{xx} V(t, x) = -\eta f(t)x^{-\eta-1}$$

and plugging into (6.7) yields

$$x^{1-\eta}\left[\frac{1}{1-\eta} f'(t) + rf(t) + \frac{1}{2\eta}\left(\frac{\alpha - r}{\sigma}\right)^2 f(t)\right] = 0$$

The terminal condition for f is $f(T) = 1$. The solution is then

$$f(t) = e^{C(T-t)}, \quad \text{where } C = r(1-\eta) + \frac{1-\eta}{2\eta}\left(\frac{\alpha - r}{\sigma}\right)^2$$

Thus, we conclude

$$V(t, x) = e^{C(T-t)} G(x)$$

solves the HJB equation. Observe that $\partial_{xx} V < 0$, hence our $\gamma < 0$ assumption is satisfied.

Given the value function, together with the maximizer of the Hamiltonian 6.1, optimal control is given by the verification theorem:

$$\pi^*(t, x) = \frac{(\alpha - r)}{\sigma^2 \eta} = a^*$$

We conclude that optimal allocation is constant, which requires continuous (or roughly) adjustments to the portfolio. An another way of looking at it is, given your current allocation and estimates for (α, r, σ), you can find your own risk aversian parameter. This might keep you aligned within some reasonable range.

7 Appendix

Theorem 7.1 (Banach Fixed Point)**.** *Let* $(\mathcal{S}, d)$ *be a complete metric space and* $T : \mathcal{S} \to \mathcal{S}$ *be a contraction mapping with constant* $0 < \lambda < 1$*, i.e.*

$$d(T(x), T(y)) \leq \lambda d(x, y)$$

Then there exists a unique fixed point x^* *of* T*, i.e.*

$$T(x^*) = x^*$$

Proof. Choose arbitrary $x_0 \in \mathcal{S}$. Define the sequence $x_n := T(x_{n-1})$ for all $n \geq 1$. Observe that

$$d(x_{n+1}, x_n) = d(T(x_n), T(x_{n-1})) \leq \lambda d(x_n, x_{n-1}) \leq \lambda^n d(x_1, x_0)$$

We now show that x_n is a Cauchy sequence, and then since $\mathcal{S}$ is complete, by definition it converges to some x^*. By triangle inequality,

$$d(x_m, x_n) \leq \sum_{k=n+1}^{m} d(x_k, x_{k-1}) \leq d(x_1, x_0) \sum_{k=n+1}^{m} \lambda^k = d(x_1, x_0)\lambda^n \frac{1 - \lambda^{m-n}}{1 - \lambda}$$

which tends to 0 as $n \to 0$, hence x_n is a Cauchy sequence. Next, we claim the limit x^* is the fixed point.

$$\begin{aligned} d(T(x^*), x^*) &\leq d(T(x^*), x_n) + d(x_n, x^*) \\ &= d(T(x^*), T(x_{n-1})) + d(x_n, x^*) \\ &\leq \lambda d(x^*, x_{n-1}) + d(x_n, x^*) \to 0 \end{aligned}$$

Lastly, suppose there exists another fixed point x. Then,

$$d(x^*, x) = d(T(x^*), T(x)) \leq \lambda d(x^*, x)$$

which is a contradiction if $d(x^*, x) > 0$. ■

References

[1] A. Cohen, *Lecture Slides for Math 474: Introduction to Stochastic Analysis for Finance*, University of Michigan, Winter 2022.

[2] R. Durrett, *Probability: Theory and Examples*, 5th ed., Cambridge University Press, New York, 2019.

[3] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd ed., Wiley, New York, 1999.

[4] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed., Springer-Verlag, New York, 1991.

[5] T. Björk, *Arbitrage Theory in Continuous Time*, Oxford University Press, Oxford, 2009.

[6] J. Zhang, *Backward Stochastic Differential Equations*, Springer, New York, 2017.

[7] Daniel Lacker, *Mean Field Games and Interacting Particle Systems*, Lecture notes, Spring 2018, Columbia University IEOR.