Chun Yip Yau

Stochastic Calculus with Financial Applications

RMSC 4005 Lecture Note

February 9, 2018

Springer

Contents

Chapter 1
Introduction

1.1 Fundamental Concepts in Finance

Consider a hypothetical situation that the financial market is formed by n basic components called *financial assets/instruments*, denoted by $S_1, \ldots, S_n$. One basic mathematical model can be defined as follows:

Definition 1.1. (Discrete time financial model).

i) At **initial** time 0, the values of the n assets are $\mathbf{S}^0 = (S_1^0, \ldots, S_n^0)'$.
ii) At **maturity** time T. There are m sets of possible values, or **scenarios**. For scenario i $(1 \le i \le m)$, the value of asset S_j $(1 \le j \le n)$ is denoted as S_j^i. The information of the assets at T can be represented by the matrix

$$\mathbf{S} = \begin{pmatrix} S_1^1 & S_2^1 & \cdots & S_n^1 \\ S_1^2 & S_2^2 & \cdots & S_n^2 \\ S_1^3 & S_2^3 & \cdots & S_n^3 \\ \vdots & \vdots & & \vdots \\ S_1^m & S_2^m & \cdots & S_n^m \end{pmatrix}. \tag{1.1}$$

We assume that $S_j^i \ge 0$ for all i and j, since an asset has no liability. Bear in mind that superscript $(^i)$ indicates the possible *scenario* and subscript $(_j)$ indicates the *index* of an asset. □

We call $\{\mathbf{S}^0, \mathbf{S}\}$ the asset structure of the market. In this book, we consider the pricing problem:

Pricing problem

Suppose that there is a financial instrument F taking value F^i, $i = 1, \ldots, m$, at each of the m scenarios in (1.1), what is the price/value of F at time 0, say F^0?

Before we answer this question, we first define some terms.

Definition 1.2. (Terms in Finance)

- **Portfolio.** A **portfolio** is a collection of assets. A portfolio can be written as $\sum_{j=1}^{n} w_j S_j$, where $\mathbf{w} = (w_1, \ldots, w_n)'$ is a vector called **weight**, which indicates a holding of w_j units of asset S_j in the portfolio.
- **Payoff Vector.** The vector that represents the values of a portfolio/financial instrument at different scenarios at maturity is called the **payoff vector**, or simply **payoff**. For example, $\mathbf{F} = (F^1, \ldots, F^m)'$ is the payoff vector of the financial instrument F, and $\mathbf{Sw}$ is the payoff vector of the portfolio $\sum_{j=1}^{n} w_j S_j$.
- **Complete.** A market is said to be **complete** if every payoff vector is attainable by one or more portfolios, i.e., for all $\mathbf{F} \in \mathbb{R}^m$, there exists $\mathbf{w} \in \mathbb{R}^n$ such that $\mathbf{F} = \mathbf{Sw}$.
- **Hedging.** A hedge is an investment position intended to offset potential losses or gains. For example, if a person is holding one unit of F and he thinks it is too risky, he can short a portfolio of $0.8\mathbf{w}$, where $\mathbf{w}$ satisfies $\mathbf{F} = \mathbf{Sw}$, so as to offset 80% of the potential losses/gains. (such a $\mathbf{w}$ can be found if the market is complete)
- **Arbitrage.** An arbitrage is a portfolio $\mathbf{w}$ that has non-positive initial value ($\mathbf{w}'\mathbf{S}^0$, a scalar) and non-negative value at maturity ($\mathbf{Sw}$, an m-dimensional vector), excluding the case $\mathbf{w}'\mathbf{S}^0 = 0$ and $S\mathbf{w} = \mathbf{0}$. Mathematically, we have

$$\underbrace{\mathbf{w}'\mathbf{S}^0}_{\text{Price}} \leq 0 \ \text{ and } \ \underbrace{\mathbf{Sw}}_{\text{Payoff}} > \mathbf{0} \quad \text{or} \quad \underbrace{\mathbf{w}'\mathbf{S}^0}_{\text{Price}} < 0 \ \text{ and } \ \underbrace{\mathbf{Sw}}_{\text{Payoff}} \geq \mathbf{0}, \tag{1.2}$$

  where

  - $\mathbf{Sw} \geq \mathbf{0}$ means all entries of $\mathbf{Sw}$ are non-negative.
  - $\mathbf{Sw} > \mathbf{0}$ means all entries of $\mathbf{Sw}$ are non-negative with at least one element being positive.
  - $\mathbf{Sw} \gg \mathbf{0}$ means all entries of $\mathbf{Sw}$ are positive. (used in Definition 1.6)

  In words, arbitrage (for the first case) means that initially we can hold a portfolio without any cost, while the portfolio at maturity has non-negative values in all scenarios, and has positive values in some scenarios. Roughly speaking, arbitrage is a **free lunch**. In finance we assume that arbitrage does not exist. Whenever arbitrage exists, demand and supply forces will quickly drive the prices back to a position without arbitrage. In short, "These chances don't belong to me".

□

Remark 1.1. **(No Arbitrage).** We say that an arbitrage opportunity exists if there exists $\mathbf{w} \in \mathbb{R}^n$ such that (1.2) holds. Thus, to check whether an arbitrage opportunity exists, we can verify the negation of (1.2), i.e., for all $\mathbf{w} \in \mathbb{R}^n$,

$$\underbrace{\mathbf{Sw}}_{\text{Payoff}} > \mathbf{0} \Rightarrow \underbrace{\mathbf{w}'S^0}_{\text{Price}} > 0 \quad \text{and} \quad \underbrace{\mathbf{Sw}}_{\text{Payoff}} \geq \mathbf{0} \Rightarrow \underbrace{\mathbf{w}'S^0}_{\text{Price}} \geq 0 \tag{1.3}$$

This negation follows from the fact that for any two statements A and B, the two statements

$$\sim (A \text{ and } B), \quad \text{and} \quad B \Rightarrow \sim A,$$

are logically equivalent. Here the notation $\sim A$ denotes the negation of statement A. For simplicity, in this book we will verify the former case of no-arbitrage condition (1.3), and the latter case can be handled similarly. □

Example 1.1. Suppose that markets A, B and C have asset structure $\mathbf{S}_A^0 = \mathbf{S}_B^0 = \mathbf{S}_C^0 = (1,1)'$, and

$$\mathbf{S}_A = \begin{pmatrix} 2 & 0 \\ 2 & 0 \\ 0 & 3 \end{pmatrix}, \mathbf{S}_B = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \mathbf{S}_C = \begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix}. \tag{1.4}$$

First we consider completeness of the market. Market A is not complete as payoffs of the form $\mathbf{F} = (a,b,c)'$ with $a \neq b$ cannot be replicated. Similarly, market C is not complete as payoffs of the form $\mathbf{F} = (a,b)'$ with $2a \neq b$ cannot be replicated. On the other hand, market B is complete because any payoff $\mathbf{F}$ can be replicated by $\mathbf{w} = \mathbf{S}_B^{-1}\mathbf{F}$, since it implies $\mathbf{S}_B\mathbf{w} = \mathbf{S}_B\mathbf{S}_B^{-1}\mathbf{F} = \mathbf{F}$. Indeed, a market $(\mathbf{S}^0, \mathbf{S})$ is complete if the inverse of $\mathbf{S}$ exists.

Next we consider arbitrage by verifying (1.2) or (1.3). To have payoff $\mathbf{S}_A\mathbf{w} = (2w_1, 2w_1, 3w_2)' > \mathbf{0}$, we need both w_1 and w_2 non-negative with at least one strictly positive. Thus, the initial price $\mathbf{w}'\mathbf{S}^0 = w_1 + w_2 > 0$. This implies no arbitrage by (1.3). The existence of arbitrage for market B will be considered in Example 1.3. For market C, it is clear that Asset 1 has the same initial price as Asset 2 but a better payoff in all scenarios. Therefore, the portfolio of holding Asset 1 and shorting Asset 2 ($\mathbf{w} = (1,-1)'$) gives a payoff $\mathbf{S}\mathbf{w} = (1,2)' > \mathbf{0}$ and price $\mathbf{w}'\mathbf{S}^0 = 1 - 1 = 0$. This is an arbitrage by (1.2).

Finally, suppose that we are holding in market A a derivative F with payoff $\mathbf{F} = (4,5,-6)'$. Note from the above discussion that we cannot completely replicate $\mathbf{F}$ using the assets $\mathbf{S}_A$. However, if we think the loss (-6) in Scenario 3 is too risky, we can hold one unit of Asset 2 to partially offset (hedge) the loss. The resulting payoff will be $(4,5,-6)' + (0,0,3)' = (4,5,-3)'$. The price of the reduced risk at maturity is the additional cost of purchasing Asset 2 in the beginning. □

1.2 Principle of No Arbitrage Pricing

In finance, the **Principle of No Arbitrage Pricing** is the key to pricing financial instruments. The key idea is: To price any financial instrument F, we find a portfolio $\sum_{j=1}^n w_j S_j$ such that its payoff matches F at every scenario at T. Having the same payoff structure at T, F and $\sum_{j=1}^n w_j S_j$ should have the same initial price. Otherwise, one can make a profit immediately by buying the cheaper one and selling the more expensive one (see Exercise 1.12). Therefore, the price F^0 is the initial value of the replicating portfolio, $\mathbf{w}'\mathbf{S}^0$. Mathematically, we do the following steps:

Principle of no arbitrage pricing:

Step 1. **(Replication).** *Given a target payoff vector* $\mathbf{F}$, *find a portfolio* $\mathbf{w} = (w_1, w_2, \dots, w_n)'$ *such that*

$$\mathbf{F} \triangleq \begin{pmatrix} F^1 \\ F^2 \\ F^3 \\ \vdots \\ F^m \end{pmatrix} = \begin{pmatrix} S_1^1 & S_2^1 & \dots & S_n^1 \\ S_1^2 & S_2^2 & \dots & S_n^2 \\ S_1^3 & S_2^3 & \dots & S_n^3 \\ \vdots & \vdots & & \vdots \\ S_1^m & S_2^m & \dots & S_n^m \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_n \end{pmatrix} \triangleq \mathbf{S}\mathbf{w}. \tag{1.5}$$

That is, we replicate the payoff of F *in every scenario by holding* $(S_1, S_2, \dots, S_n)'$ *with weight* $(w_1, \dots, w_n)'$. *The portfolio* $\mathbf{w}$ *is known as the* **replicating portfolio**.

Step 2: **(Pricing).** *The price of the instrument,* F^0, *is given by*

$$F^0 = \mathbf{w}'\mathbf{S}^0 = \sum_{j=1}^{n} w_j S_j^0. \tag{1.6}$$

The weight $\mathbf{w}$ of the replicating portfolio is found by solving the system of equations (1.5) with m equations and n unknowns. When does a solution exist? Is the solution unique? This reduces to a linear algebra problem. Observe from (1.5) that for any $\mathbf{F}$, the weight $\mathbf{w}$ exists if $\mathbf{F}$ is in the column space of $\mathbf{S}$, i.e., $\mathbf{F}$ can be represented by linear combinations of columns of $\mathbf{S}$. Thus the column space of $\mathbf{S}$ is the key to the pricing problem.

We impose the following assumptions to obtain an effective set of assets.

Assumption 1.1 (Linearly Independent Columns of S). *In the market defined in Definition 1.1,* $\mathbf{S}$ *has linearly independent columns.*

If the columns of $\mathbf{S}$ are linearly dependent, then at least one of the columns can be written as a linear combination of other columns. Physically, it means that some assets can be constructed by other assets (their final payoffs are the same in all scenarios). If this happens, then the asset is redundant and may be ignored. That is, we can remove columns of $\mathbf{S}$ until it has linearly independent columns.

For simplicity, we also impose the following assumption. This assumption will be relaxed in Section 1.10.

Assumption 1.2 *In the market defined in Definition 1.1, the number of scenarios is the same as the number of assets, i.e.,* $m = n$.

The assumption $m = n$ is imposed to ensure that the inverse $\mathbf{S}^{-1}$ of S exists, which facilitates the derivation of the following replicating portfolio weights and pricing formulas .

Theorem 1.3. (Replicating Portfolio)
Under Assumptions 1.1 and 1.2, for any given payoff vector $\mathbf{F}$, *the weight of replicating portfolio is given by*

$$\mathbf{w} = \mathbf{S}^{-1}\mathbf{F}\,. \tag{1.7}$$

Proof. Since every square matrix ($m = n$) with linearly independent columns is invertible, Assumptions 1.1 and 1.2 together imply the existence of $\mathbf{S}^{-1}$. Since a replicating portfolio $\mathbf{w}$ satisfies $\mathbf{F} = \mathbf{S}\mathbf{w}$, multiplying $\mathbf{S}^{-1}$ on both sides gives (1.7). □

Note that Theorem 1.3 implies that the **market is complete** because all payoff vectors can be replicated. Using the weight in Theorem 1.3, we obtain the following simple pricing formula:

Corollary 1.4 (Pricing) *Under Assumptions 1.1 and 1.2, for any payoff* $\mathbf{F}$ *and asset structure* $\{\mathbf{S}^0, \mathbf{S}\}$*, the price is given by*

$$F^0 \triangleq (\mathbf{S}^0)'\mathbf{S}^{-1}\mathbf{F}\,. \tag{1.8}$$

Proof. Recall from (1.6) that given the replicating portfolio $\mathbf{w}$, the price is given by $F^0 = \mathbf{w}'\mathbf{S}^0$. Since F^0 is a scalar, we can write $F^0 = (F^0)' = (\mathbf{S}^0)'\mathbf{w}$. Substituting the weight $\mathbf{w} = \mathbf{S}^{-1}\mathbf{F}$ from (1.7) into $F^0 = (\mathbf{S}^0)'\mathbf{w}$, the result (1.8) follows. □

Example 1.2. For the markets A, B and C in Example 1.1, it is clear that only Market B satisfies Assumptions 1.1 and 1.2. Given a payoff $\mathbf{F} = (3,3)'$ in Market B, the replicating portfolio is $\mathbf{w} = \mathbf{S}_B^{-1}\mathbf{F} = (1,1)'$. Thus, the price of F is given by $F^0 = (\mathbf{S}^0)'\mathbf{w} = (1,1)\cdot(1,1)' = 2$. □

1.3 State price vector and risk neutral probabilities

In this section we introduce a new concept called state price vector, which gives a probabilistic interpretation of the pricing procedure. To motivate, recall the pricing formula $F^0 \triangleq (\mathbf{S}^0)'\mathbf{S}^{-1}\mathbf{F}$ in (1.8). If we define the **state price vector** as

$$\psi' = (\mathbf{S}^0)'\mathbf{S}^{-1}\,, \tag{1.9}$$

(an m-dimensional vector), then (1.8) can be written as

$$F^0 = \psi'\mathbf{F} = \sum_{j=1}^{m} \psi_j F^j\,. \tag{1.10}$$

From (1.10), ψ can be regarded as a "probability distribution" that gives weights to the scenarios $(F^1,\dots,F^m)'$. The price F^0 is regarded as the expected value of $\mathbf{F}$ under the discrete "probability measure" $(\psi_1,\dots,\psi_m)'$. Now the question is "Is ψ really a probability distribution?" Recall that any probability must be non-negative and the total probability is 1. Thus ψ has to satisfy

$$\text{(i) } \psi > \mathbf{0} \quad \text{and} \quad \text{(ii) } \sum_{j=1}^{m} \psi_j = 1\,, \tag{1.11}$$

in order to be a probability distribution.

Proposition 1.5 *Under Assumptions 1.1 and 1.2, there is no arbitrage opportunity in the market if and only if the state price vector given in* (1.9) *satisfies* $\psi \gg \mathbf{0}$.

Proof. Suppose that $\psi \gg 0$. If the payoff $\mathbf{F} = \mathbf{Sw} > \mathbf{0}$, then from (1.10) the price is $F^0 = \psi'\mathbf{F} = \psi'\mathbf{Sw} > 0$ because every entry in ψ is positive and every entry of $\mathbf{Sw}$ is non-negative. This verifies the "no arbitrage condition" (1.3).

Conversely, suppose that there is no arbitrage opportunity in the market. Define $\mathbf{1}_{k,m} = (0,\ldots,0,1,0,\ldots,0)'$ to be an m-dimensional vector with all entries equal 0 except the k-th entry being 1. The "no arbitrage condition" (1.3) implies that if the payoff structure is $\mathbf{F} = \mathbf{1}_{k,m}$ (gain \$ 1 when scenario k happens), then the price $F^0 = \psi'\mathbf{F} = \psi_k$ must be positive, i.e., $\psi_k > 0$. Repeating the same argument for $k = 1,\ldots,m$, we have $\psi \gg \mathbf{0}$. □

Remark 1.2. **(Interpretation of State Price Vector.)** As a by-product of the above proof, it can be seen that ψ_k is the **price** of a product with final payoff of \$ 1 at scenario (**state**) k. This explains the name "**State Price**". □

Example 1.3. Consider Market B in Example 1.1. The state price vector is given by $\psi' = (\mathbf{S}^0)'\mathbf{S}^{-1} = (1/3,1/3)$. Since $\psi \gg \mathbf{0}$, Proposition 1.5 implies that there is no arbitrage in the market. Note that the price of a derivative with payoff $\mathbf{F} = (3,3)'$ can be computed by $F^0 = \psi'\mathbf{F} = 2$, which is consistent with the results in Example 1.2. □

We have just seen that (i) $\psi > \mathbf{0}$ of (1.11) holds, since $\psi \gg \mathbf{0}$ implies $\psi > \mathbf{0}$. However, (ii) $\sum_{j=1}^{m} \psi_j = 1$ does not hold in general. Consider the counter-example:

Example 1.4. Consider a risk free asset (e.g. bond) with $F^0 = 1$ and $F^i = 1 + r$ for $i = 1,\ldots,m$. Using the pricing formula (1.10), we have

$$1 = F^0 = \sum_{j=1}^{m} \psi_j F^j = (1+r)\sum_{j=1}^{m} \psi_j\,,$$

giving

$$\overline{\psi} \triangleq \sum_{i=1}^{m} \psi_i = \frac{1}{1+r}\,,$$

which is the <u>discount factor</u> of the risk free asset. Here, $\overline{\psi} = \sum_{i=1}^{m} \psi_i \neq 1$ unless $r = 0$. □

Remark 1.3. **(Discount factor.)** Writing $\overline{\psi} = \sum_{i=1}^{m} \psi_i = \psi'\mathbf{1}_m$ and using (1.10), $\overline{\psi}$ is the price for a product with *riskless payoff* $\mathbf{1}_m$. Thus $\overline{\psi}$ is naturally interpreted as a discount factor. □

Example 1.5. Consider Market B in Example 1.1. Although the two assets in $\mathbf{S}_B$ are not risk free, a risk free asset with payoff $\mathbf{F} = (1,1)'$ can be found by Theorem 1.3 as $\mathbf{w} = \mathbf{S}_B^{-1}\mathbf{F} = (1/3,1/3)'$. The price, i.e., the discount factor $\overline{\psi}$, is $F^0 = (\mathbf{S}^0)'\mathbf{w} = 2/3$.

Alternatively, from Example 1.3 we have the state price vector $\psi = (1/3, 1/3)'$. The discount factor can be obtained as $\overline{\psi} = \sum_{i=1}^{2} \psi_i = 2/3$. □

Since (ii) in (1.11) does not hold, ψ is in general not a probability distribution. However, it is interesting to see from the above example that, $\overline{\psi} = \sum_{i=1}^{m} \psi_i$ is the discount factor of the risk free asset. If we rescale ψ and rewrite (1.10) as

$$F^0 = \psi'\mathbf{F} = \overline{\psi}\left(\frac{\psi'}{\overline{\psi}}\right)\mathbf{F} = \overline{\psi}\left(\frac{\psi_1}{\overline{\psi}}, \ldots, \frac{\psi_m}{\overline{\psi}}\right)\mathbf{F}, \tag{1.12}$$

where $\overline{\psi} = \sum_{i=1}^{m} \psi_i$, then $\widetilde{\psi} \triangleq \psi/\overline{\psi}$ can be regarded as a probability distribution. Note that $\overline{\psi}$ is a scalar and both ψ and $\widetilde{\psi}$ are m-dimensional vectors.

State Price Vector (ψ) and Risk Neutral Probabilities ($\widetilde{\psi}$):

- Multiply State Price Vector ψ to the payoff $\mathbf{F}$, i.e. $\psi'\mathbf{F}$, gives the fair price.
- $\overline{\psi} \triangleq \sum_{i=1}^{m} \psi_i$ is the discount factor from time 0 to T.
- $\widetilde{\psi} \triangleq \frac{1}{\overline{\psi}}\psi \gg 0$ is a probability distribution known as the risk neutral probability distribution or measure.

Remark 1.4. **(Interpretation of Risk Neutral Probability)** Rewrite (1.9) as

$$(\mathbf{S}^0)' = \psi'\mathbf{S} = \overline{\psi}\widetilde{\psi}'\mathbf{S}. \tag{1.13}$$

The j-th entry ($j = 1, \ldots, n$) of (1.13) is $S_j^0 = \overline{\psi}\sum_{i=1}^{m} \widetilde{\psi}_i S_j^i$. Rewrite as

$$\sum_{i=1}^{m} \widetilde{\psi}_i \frac{S_j^i}{S_j^0} = \frac{1}{\overline{\psi}}, \qquad j = 1, \ldots, n. \tag{1.14}$$

Since S_j/S_j^0 is the gain of the j-th asset, the left hand side of (1.14) is the expected gain of the j-th asset under the risk neutral measure $\widetilde{\psi}$. Thus, (1.14) implies that the expected gain is equal to the same $\overline{\psi}^{-1}$ for all assets. This explains the name "risk neutral". □

We conclude this section by summarizing the pricing formula in terms of expectation under the risk neutral probability measure.

Theorem 1.6. (Pricing)
Consider a market with asset structure $\{\mathbf{S}^0, \mathbf{S}\}$. Under Assumptions 1.1 and 1.2, if a payoff $\mathbf{F}$ is replicable, then the price is given by

$$F^0 = \psi'\mathbf{F} \triangleq \overline{\psi} E_{\widetilde{\psi}}(\mathbf{F}),$$

where $\psi' = (\mathbf{S}^0)'\mathbf{S}^{-1}$ is the **state price vector**, *$\overline{\psi} = \sum_{i=1}^{m} \psi_i$ is the* **discount factor**, *$\widetilde{\psi} \triangleq \psi/\overline{\psi}$ is the* **risk neutral probability distribution**, *$E_{\widetilde{\psi}}$ is the expectation under the risk neutral probability distribution $\widetilde{\psi}$. For notational simplicity, sometimes we* *write $F^0 = E_\psi(\mathbf{F})$. It is a slight abuse of notation because ψ is not a probability distribution.*

Example 1.6. Consider Market B in Example 1.1. Since the state price vector is $\psi = (1/3, 1/3)'$ (Example 1.3) and the discount factor is $\overline{\psi} = 2/3$ (Example 1.5), the risk neutral probability measure is given by $\widetilde{\psi} = \psi/\overline{\psi} = (1/2, 1/2)'$.

As in Example 1.3, the price of the derivative F with payoff $\mathbf{F} = (3,3)'$ can be computed by $F^0 = \psi'\mathbf{F} = 2$, or equivalently,

$$F^0 = \overline{\psi}\mathrm{E}_{\widetilde{\psi}}(\mathbf{F}) = \frac{2}{3} \times \left(\frac{1}{2}3 + \frac{1}{2}3\right) = 1 + 1 = 2\,.$$

□

1.4 Physical Probability v.s. Risk Neutral Probability

Note that in the previous discussion the "physical" probability about the price evolution has not been taken into account. This is because in the principle of No Arbitrage Pricing, the asset structure already controls the price of all replicable financial instruments. To be precise, consider the following example.

Example 1.7. Suppose that the market has two assets S_1, S_2 satisfying

$$\mathbf{S}^0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \ \mathbf{S} = \begin{pmatrix} 1.2 & 10 \\ 1.2 & 1 \end{pmatrix}, \ \mathrm{P}(\text{Scenario 1}) = 0.9\,, \ \mathrm{P}(\text{Scenario 2}) = 0.1\,.$$

Note that the first asset is a bond with 20% interest rate. Recall that a Call option is a financial instrument that offers a chance to purchase the underlying asset for price K (strike price) at time T. Consider a Call option F on the second asset S_2 with a strike price $K = 2$. F is said to be a financial derivative of S_2. In the first scenario, we can purchase S_2 with \$2 and sell it immediately for \$10 to earn \$8. In the second scenario, we do nothing as the strike price is higher than the market price. Thus, the payoff is given by

$$\mathbf{F} = \begin{pmatrix} 8 \\ 0 \end{pmatrix}.$$

Simple probability calculation suggests that the expected value of F at time T equals to $8(0.9) + 0(0.1) = 7.2$. Then, discounting by the risk free rate gives the physical price $F^0_{pp} = \$7.2/1.2 = \$\,6$. However, using (1.8), the risk neutral price can be calculated to be $F^0_{rnp} = (\mathbf{S}^0)'\mathbf{S}^{-1}\mathbf{F} = \$\,0.148 < 6$.

It is tempting to think that F is worth buying if the price is less than \$ 6. If F is now selling at \$ 5, say, should we buy it?

The answer is negative. Formally, we can argue that if $F^0 = 5$, there exists an arbitrage opportunity (Exercise 1.15). To get more intuition, note that although it seems to be a good deal to buy F at \$ 5, we should not take that because S_2 **itself**

is a much better deal! To see this, just notice that if you buy F, you pay \$ 5 at the beginning and receive \$ 8 or \$ 0. But if you buy S_2, you pay \$ 1 at the beginning and receive \$ 10 or \$ 1 (pay less and get more).

In fact, in a complete market, every payoff structure can be replicated by trading of assets $\{S_i\}_{i=1,\ldots,n}$. In other words, the structure $\{\mathbf{S}^0, \mathbf{S}\}$ controls the risk neutral price of financial instruments with any payoff structure. If probability measure other than the risk neutral measure (e.g. physical measure) is used for pricing, arbitrage opportunities may occur.

In conclusion, this example points out that pricing is governed by the predetermined structure of the assets, and physically what will be happening is irrelevant. □

Example 1.8. In the previous example, S_2 is a much better deal than F if $F^0 = 5$. In fact, one can check that the *physical price* of S_2 should be \$ $(10(0.9) + 1(0.1))/1.2 =$ \$ 7.58. How come it is selling at \$ 1?

One may think in this way. In the computation of the physical price, it just weighs the payoff by the probability of occurrence of each scenario but **ignores the importance of different scenarios**. For example, if Scenario 2 is a state of bad health condition and one needs \$ 1.2 for the medical expenses, then S_1, which guarantees a payoff of \$ 1.2, will be worth more than S_2.

On the other hand, the state price vector is found to be $\psi' = (\mathbf{S}^0)'\mathbf{S}^{-1} = (0.019, 0.815)$, thus the risk-neutral probability distribution is $\tilde{\psi} = (0.022, 0.978)'$. Since more weight is assigned to Scenario 2, the risk-neutral probabilities suggest that Scenario 2 is much more important than Scenario 1. In other words, the risk-neutral probability captures the importance of different scenarios from the market structure.

After all, the setting of asset price structure $\{\mathbf{S}^0, \mathbf{S}\}$ boils down to **how we measure risk**, which is a matter of taste/utility, and hence is not questionable unless arbitrage opportunities exist. □

1.5 Binomial trees

One well known example of the discrete time financial model (1.1) is the binomial model, or binomial tree:

Definition 1.3. (Binomial Model)
Binomial model is a discrete time financial model with $n = m = 2$. It consists of

- two assets, bond (B) and stock (S) in the market
- two scenarios, upward (u) and downward (d) movements of the stock $(u > d)$.

Specifically,

$$\mathbf{S}^0 = (1, s)' \text{ and } \mathbf{S} = \begin{pmatrix} e^{rT} & su \\ e^{rT} & sd \end{pmatrix}. \tag{1.15}$$

In (1.15), s is the initial price of the stock and r is the continuously compounded risk free interest rate for the bond, thus the accumulation factor is e^{rT}.

The properties of Binomial model are summarized in the following Theorem.

Theorem 1.7. *In the binomial tree model,*

a) **(Replicating Portfolio)**. *For any payoff structure* $\mathbf{F} = (F^1, F^2)' \triangleq (F_u, F_d)'$, *the replicating portfolio is given by*

$$\mathbf{w} = \mathbf{S}^{-1}\mathbf{F} = \left(\frac{F_d u - F_u d}{e^{rT}(u-d)}, \frac{F_u - F_d}{s(u-d)} \right)',$$

i.e., a holding of $(F_d u - F_u d)/e^{rT}(u-d)$ *units of bonds and* $(F_u - F_d)/s(u-d)$ *units of stocks.*

b) **(Risk Neutral Probabilities)**. *The state price vector is unique and is given by*

$$\psi = \left((\mathbf{S}^0)'\mathbf{S}^{-1}\right)' = \frac{1}{e^{rT}(u-d)} \left(e^{rT} - d,\, u - e^{rT} \right)'. \tag{1.16}$$

In particular, the discount factor is $\overline{\psi} = \sum_{j=1}^{2} \psi_j = e^{-rT}$ *and the risk neutral probability vector is given by*

$$\widetilde{\psi} = \left(\frac{e^{rT} - d}{u-d}, \frac{u - e^{rT}}{u-d} \right)'.$$

c) **(Pricing)**. *The price of any instrument with payoff* $\mathbf{F} = (F_u, F_d)'$ *is given by*

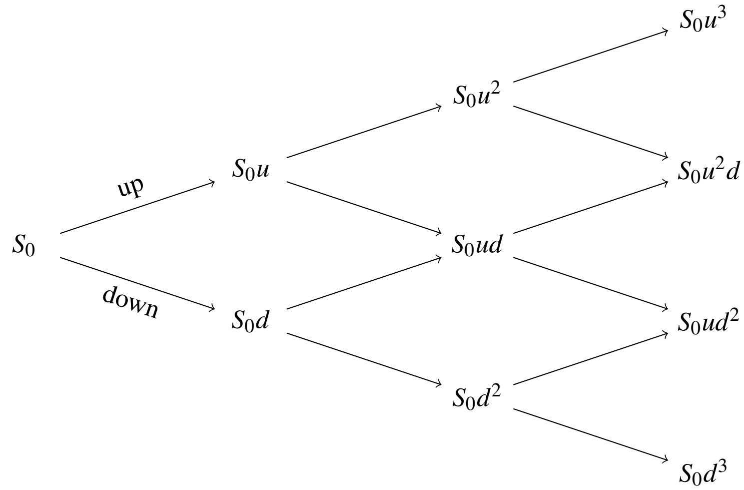
$$F^0 = \psi'\mathbf{F} = \frac{F_u(e^{rT} - d) + F_d(u - e^{rT})}{e^{rT}(u-d)}.$$

d) **(Arbitrage)** *Arbitrage opportunities exist if* $u > d > e^{rT}$ *or* $d < u < e^{rT}$.

Proof. Note that now $m = n = 2$ and $u > d$ imply that Assumption 1.1 holds. Thus, the explicit formulas in the previous sections can be used. First, a) follows from (1.7). Next, it is easy to see that b) and c) follow from (1.9) and (1.10) respectively. For d), suppose first that $u > d > e^{rT}$. Note that the first entry of the state price vector ψ is negative. By Remark 1.2, the payoff vector $\mathbf{F} = (1,0)'$ creates an arbitrage opportunity (check). Intuitively, the stock now outperforms the bond in every scenario, so we can short the bond and long the stock to obtain arbitrage. In fact, from (1.7), the precise arbitrage portfolio is found to be $\mathbf{w} = \mathbf{S}^{-1}(1,0)'$. The case $d < u < e^{rT}$ can be handled similarly. □

1.6 Multi-period Binomial models

Obviously, a discrete time financial model is inadequate to describe the evolution of an asset price, no matter how many scenarios are chosen, since the dynamics of the price is not modeled. Therefore, to construct a more sophisticated market model, one natural way is to bring together copies of binomial models to form a multi-period Binomial model, or Binomial tree, as shown below.



Definition 1.4. (Multi-period Binomial Model).
In an N-period binomial model, the market is observable at time $0 = t_0 < t_1 < t_2 < \cdots < t_N = T$, where $t_i = i\delta$. Over each time interval $[t_i, t_{i+1}]$, the assets follow the binomial model with a bond and a stock. In particular, the values of the bond and stock at time t_i are given by (B_i, S_i), where $(B_0, S_0) = (1, s)$ at time 0 and

$$B_i = e^{rt_i}, \qquad S_i = \begin{cases} S_{i-1}u & \text{if market goes up} \\ S_{i-1}d & \text{if market goes down} \end{cases}.$$

Since at each time point (each branch of the tree) the market can go up or down, the N-period multinomial model can handle 2^N possible scenarios without specifying 2^N assets. Since the same u and d are used in all periods, some of the 2^N scenarios collapse to the same one (e.g., $udd = dud = ddu$). As a result, there are only $i+1$ distinct scenarios for (B_i, S_i) at time t_i (see the above figure).

Remark 1.5. **(State Price Vector at each branch).** Suppose that at time t_i a branch of the tree takes value (e^{rt_i}, S_i). The binomial model at this branch can be expressed in matrix form (1.15) by

$$\mathbf{S}^0 = (e^{rt_i}, S_i)' \text{ and } \mathbf{S} = \begin{pmatrix} e^{rt_{i+1}} & S_i u \\ e^{rt_{i+1}} & S_i d \end{pmatrix}. \tag{1.17}$$

Note that $t_{i+1} - t_i = \delta$. From (1.9), the state price vector is given by

$$\begin{aligned} \psi' &= (\mathbf{S}^0)'\mathbf{S}^{-1} = \frac{1}{e^{r\delta}(u-d)} \left(e^{r\delta} - d,\, u - e^{r\delta} \right) \\ &= (\psi_u, \psi_d), \end{aligned} \tag{1.18}$$

say. By construction, the state price vector is common for all branches. □

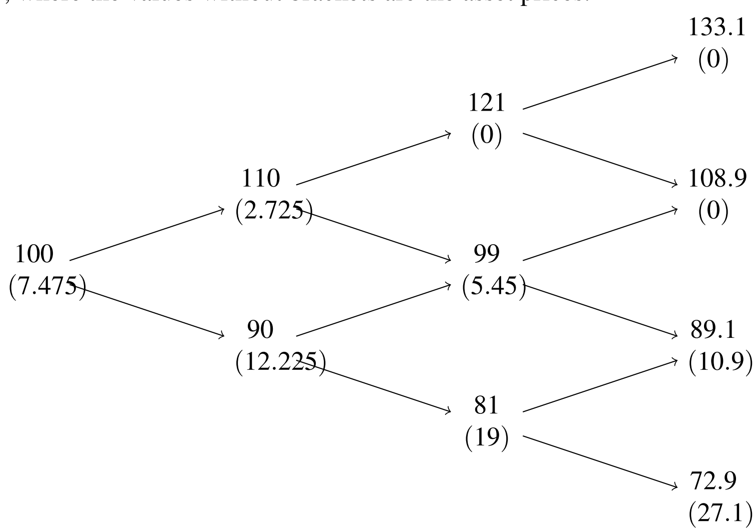
From Remark 1.5, the pricing procedure can be performed by repeating the same idea in the single period model at each branch of the tree. The procedure is known as **Backward Induction** on tree.

Backward Induction on Binomial tree

Suppose that we want to price a product F where only the values of F_N (the price of F at the N-th period, i.e. time T) at each of the 2^N scenarios are known. The pricing procedure proceeds as follows:

i) Set $j = N$. At each of the 2^{j-1} branches of the tree in the $(j-1)$-th period, use a single period model to compute F_{j-1}.
ii) Repeat step i) for $j = N-1, N-2, \ldots, 1$.
iii) The price is given by F_0.

Example 1.9. Suppose that the asset prices evolve according to the tree shown below, where the values without brackets are the asset prices.



To illustrate the method, suppose that the risk-free interest rate is zero. What is the price of a three-month American put option with strike price $K = 100$?

Solution: First, note that this is a particular example of (1.17) where $r = 0$, $S_0 = 100$, $u = 1.1$ and $d = 0.9$. From (1.18), it can be checked that the state price vector for all branches is $\psi = (0.5, 0.5)'$. Using the backward induction of trees, we have:

1) The values of the option at time 3, reading from top to bottom, are 0, 0, 10.9 and 27.1, respectively.
2) At time 2, we must consider two possibilities: the value if we exercise the option immediately, and the value if we hold the option until the next period. For the top node it is easy since the values are zero in both cases. For the second node,

if we exercise the option, then the value is 1. On the other hand, if we hold the option, then from the analysis of the single step binomial model, the value of the option is the expected value under the risk-neutral probabilities of the claim at time 3, i.e. $0 \times 0.5 + 10.9 \times 0.5 = 5.45$. As $5.45 > 1$, we should hold the option and the value of the option should be 5.45. For the bottom node, using the same reasoning, we arrive at the value $\max(19, 0.5(10.9 + 27.1)) = 19$.

3) Now consider the two nodes at time 1. The value at the top node is $\max(0, 0.5(0 + 5.45)) = 2.725$. For the bottom node, the option is worth $\max(10, 0.5(5.45 + 19)) = 12.225$.
4) Finally, at time 0, the price is given by $\max(0, 0.5(2.725 + 12.225)) = 7.475$.

Thus, the option price at time 0 is 7.475. The price at each node is shown in the brackets of the figure above.

Next we discuss the general theory of pricing European options by the multi-period binomial model. We begin with the notion of *path*.

Definition 1.5. (Path) For an N-period binomial model and $t = 1, \ldots, N$, the path $\omega_t = (Z_1, \ldots, Z_t)'$ is a t-dimensional vector specifying the evolution of the market up to time t, where for $i = 1, \ldots, t$,

$$Z_i = \begin{cases} 1 & \text{if market goes up,} \\ -1 & \text{if market goes down.} \end{cases}$$

Noticing that once ω_N is given, we know ω_t for $t \leq N$.

Example 1.10. Using the dynamics of Example 1.9, if the stock price goes up in the first two periods, then $\omega_2 = (1, 1)'$. In this case, we denote $S_1(\omega_2) = 110$, $S_2(\omega_2) = 121$; $F_1(\omega_2) = 2.725$, $F_2(\omega_2) = 0$. However, $S_3(\omega_2)$ and $F_3(\omega_2)$ are unknown.

The concept of risk neutral probability or state price vector can be generalized from single period model to multiple period model by the following theorem.

Theorem 1.8. (State Price Vector and Pricing in Multi-period Binomial model) *Let F^j, ω_N^j, $j = 1, \ldots, 2^N$ be the 2^N possible payoffs of F at time N and their corresponding paths, respectively. Define the* **State Price Vector** $\Psi = (\Psi^1, \ldots, \Psi^{2^N})'$, *where*

$$\Psi^j = \prod_{t=1}^{N} \left(\psi_u \mathbb{I}_{\{Z_t^j = 1\}} + \psi_d \mathbb{I}_{\{Z_t^j = -1\}} \right)$$

is the "state price" associated with the path $\omega_N^j = (Z_1^j, \ldots, Z_N^j)'$, $\mathbb{I}_B$ is the indicator function of the set B, and (ψ_u, ψ_d) is defined in (1.18). *Given the payoff vector $\mathbf{F} = (F^1, \ldots, F^{2^N})'$, the price of F at $t = 0$ is*

$$F^0 = \sum_{j=1}^{2^N} \Psi^j F^j \triangleq E_\Psi(\mathbf{F}), \tag{1.19}$$

where E_Ψ is defined as the expectation under the discrete distribution Ψ.

Proof. From the operation of the Backward Induction algorithm, the payoff F^j is multiplied by the discounted risk neutral probabilities ψ_u or ψ_d at each period, corresponding to the upward and downward movement of the market. The product of all the N branches along the path ω_N^j is exactly Ψ^j. Summing over all the contributions of the 2^N payoffs gives (1.19). □

Example 1.11. **(Risk Neutral Probabilities)**
Summing over all the Ψ^js, we recover the N-period discount factor

$$\overline{\Psi} \triangleq \sum_{j=1}^{2^N} \Psi^j = \sum_{k=0}^{N} C_k^N \psi_u^k \psi_d^{N-k} = (\psi_u + \psi_d)^N \overset{(1.18)}{=} \left(e^{-r\delta}\right)^N = e^{-rT},$$

thus the Risk Neutral Probability vector in the N-period Binomial model is simply

$$\widetilde{\Psi} = \overline{\Psi}^{-1} \Psi. \tag{1.20}$$

Intuitively, the vector $\widetilde{\Psi}$ describes the (risk neutral) probabilities of occurrence for each of the 2^N possible scenarios for the final stock price S_N and the state price vector Ψ is the "discounted" version of $\widetilde{\Psi}$ so that multiplying to $\mathbf{F}$ in (1.19) yields the price F^0. □

Lastly, recall that hedging is an action to reduce risk by taking an investment position intended to offset potential losses/gains. In a multi-period Binomial model, the hedging of F can be performed by adjusting the replicating portfolio at the beginning of each period.

Example 1.12. **(Replicating Portfolio in Multi-period Binomial model)**
Given that at period k the asset values are $(B_k, S_k) = (e^{rk\delta}, s)$ and the value of an instrument is F_k. Suppose that the two possible values of F at period $k+1$ are found to be $\mathbf{F}_{k+1} = (F_{k+1}^u, F_{k+1}^d)'$. Then, using Theorem 1.7, the weights of the replicating portfolio are given by

$$\mathbf{w}_{k+1} = \left(\frac{F_{k+1}^d u - F_{k+1}^u d}{e^{r(k+1)\delta}(u-d)}, \frac{F_{k+1}^u - F_{k+1}^d}{s(u-d)} \right)'. \tag{1.21}$$

Note that s, F_{k+1}^d and F_{k+1}^u are known given the path ω_k of the market. In other words, to hedge for time $k+1$, the weight $\mathbf{w}_{k+1}$ has to be specified at time k using the information ω_k. One can verify the correctness of the hedging by showing that the hedging portfolio takes exactly the same values as $\mathbf{F}_{k+1}$ at time $k+1$. □

1.7 Extensions to Continuous Model

It is clear that a continuous model is more appropriate to describe the real situations. One natural way to achieve an approximately continuous model is to use a discrete model with very small time periods. We will use the following particular multi-period binomial model:

1. The time period of each interval is of length δ.
2. The cash bond takes the form $B_k = e^{r\delta k}$.
3. The initial stock price is S_0.
4. The stock price process is modeled by a binomial tree. If the stock price at the end of the k-th period is $S_{k\delta}$, then over the next time period it moves to new values

$$S_{(k+1)\delta} = S_{k\delta} \exp\left\{\mu\delta + \sigma\sqrt{\delta} Z_{k+1}\right\}, \quad \text{where} \quad Z_{k+1} = \begin{cases} 1 & \text{if up,} \\ -1 & \text{if down.} \end{cases} \tag{1.22}$$

Remark 1.6. Note that the above model is a particular case in Definition 1.4, with

$$u = \exp(\mu\delta + \sigma\sqrt{\delta}), \text{ and } d = \exp(\mu\delta - \sigma\sqrt{\delta}). \tag{1.23}$$

Roughly speaking, the stock is expected to grow with the rate $\exp(\mu\delta)$ with volatility (fluctuation) of size $\exp(\pm\sigma\sqrt{\delta})$. □

Without loss of generality, we assume $T = 1$. Similar to the previous section, we partition the time interval $[0,T] = [0,1]$ into n intervals of length δ, but later we let $\delta \to 0$, $n \to \infty$ such that $n\delta = 1$. From (1.22), it can be easily checked that at time $t = m\delta$ (the end of the m-th period), $1 \le m \le n$,

$$S_t = S_0 \exp\left\{\mu m\delta + \sigma\sqrt{\delta}\sum_{k=1}^{m} Z_k\right\}. \tag{1.24}$$

Note that all the randomness of S_t is from $\sum_{k=1}^{m} Z_k$. Note also that the physical probability for the stock to go up or down is not important. Substitute (1.23) into (1.18), the risk neutral probabilities (before discounting) of going up ($Z_k = 1$) and down ($Z_k = -1$) are respectively

$$\begin{aligned} \psi_u e^{r\delta} &\triangleq P(Z_k = 1) \triangleq p = \frac{e^{r\delta} - e^{\mu\delta - \sigma\sqrt{\delta}}}{e^{\mu\delta + \sigma\sqrt{\delta}} - e^{\mu\delta - \sigma\sqrt{\delta}}}, \\ \psi_d e^{r\delta} &\triangleq P(Z_k = -1) = 1 - p. \end{aligned} \tag{1.25}$$

Recall from the basic calculation of Bernoulli random variable that for each k,

$$\begin{aligned} \mathrm{E}(Z_k) &= p - (1-p) = 2p - 1, \\ \mathrm{Var}(Z_k) &= p + (1-p) - (2p-1)^2 = 4p(1-p). \end{aligned}$$

Normalizing $\sum_{k=1}^{m} Z_k$ by the mean and standard deviation, we can rewrite (1.24) as

$$\begin{aligned}S_t &= S_0 \exp\left\{\mu m\delta + \sigma\sqrt{\delta}\left(\sum_{k=1}^{m}\frac{Z_k-(2p-1)}{\sqrt{4p(1-p)}}\sqrt{4p(1-p)}+m(2p-1)\right)\right\}\\ &= S_0 \exp\left\{\mu m\delta + \sigma m\delta\frac{2p-1}{\sqrt{\delta}} + \sigma\sqrt{4\delta p(1-p)}\sum_{k=1}^{m}Y_k\right\}, \qquad (1.26)\end{aligned}$$

where $Y_k \triangleq \frac{Z_k-(2p-1)}{\sqrt{4p(1-p)}}$ is a random variable with mean 0 and variance 1.

Next we investigate the limit of (1.26) as $\delta \to 0$. In (1.25), using a second order Taylor's expansion around 0 on each of the four exponential functions in p (i.e. $e^x = 1+x+x^2/2+\cdots$), we have the approximation

$$p \approx \frac{\delta(r-\mu-\frac{1}{2}\sigma^2)+\sigma\sqrt{\delta}}{2\sigma\sqrt{\delta}}; \qquad (1.27)$$

see Exercise 1.19. Therefore, if we approximate the continuous time situation by taking $\delta \to 0$, we have

$$p \to \frac{1}{2}, \quad \frac{2p-1}{\sqrt{\delta}} \to -\frac{1}{\sigma}\left(\mu+\frac{1}{2}\sigma^2-r\right). \qquad (1.28)$$

For the limit of the random part $\sum_{k=1}^{m} Y_k$, we borrow two powerful theorems from probability theory:

Functional Central Limit Theorem (FCLT)

Suppose that $\{Y_k\}_{k=1,2,\ldots}$ is a sequence of independent random variables with mean 0 and variance 1. Define the **partial sum process** $S_Y(t) \triangleq \sum_{k=1}^{[nt]} Y_k$. Here, $S_Y(t)$ is a **random** function from $t \in (0,\infty)$ to $\mathbb{R}$. As $n \to \infty$, the rescaled partial sum satisfies

$$\frac{S_Y(t)}{\sqrt{n}} = \frac{\sum_{k=1}^{[nt]} Y_k}{\sqrt{n}} \xrightarrow{D} W_t,$$

where W_t is the Brownian Motion (Wiener process) with index t, and $\xrightarrow{D}$ denotes weak convergence of stochastic processes.

Continuous Mapping Theorem

Suppose that $X_n \xrightarrow{D} X$ as $n \to \infty$ and $f(\cdot)$ is a continuous function on the space of X. Then

$$f(X_n) \xrightarrow{D} f(X).$$

as $n \to \infty$.

The precise definition of Brownian Motion W_t will be given in chapter 6. One useful property is that $W_t \sim N(0,t)$ for each $t \geq 0$.

Recall that by construction, $n\delta = 1$ and $m\delta = t$. Thus, we can write $\delta = \frac{1}{n}$ and $m = nt$. Combining with (1.28), FCLT and the Continuous Mapping Theorem, we can compute the limit of (1.26) as

$$\begin{aligned} S_t &= S_0 \exp\left\{ \mu t + \sigma t \frac{2p-1}{\sqrt{\delta}} + \sigma\sqrt{4p(1-p)} \frac{\sum_{k=1}^{nt} Y_k}{\sqrt{n}} \right\} \\ &\xrightarrow{D} S_0 \exp\left\{ \mu t - \sigma t \frac{1}{\sigma}\left(\mu + \frac{1}{2}\sigma^2 - r\right) + \sigma W_t \right\} \\ &= S_0 \exp\left\{ \left(r - \frac{1}{2}\sigma^2\right) t + \sigma W_t \right\}, \end{aligned} \tag{1.29}$$

which we call the <u>**Continuous Time Pricing Model**</u>, or the <u>**Black-Scholes Model**, for asset S_t</u>.

Remark 1.7. i) In continuous time model, S_t takes infinitely many possible values from $[0,\infty)$. Thus, we cannot write down the risk neutral probability vector $\widetilde{\Psi}$ as in (1.20). Instead, the **probability density function** of the random variable $S_0 \exp\left\{\left(r - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right\}$ in (1.29) describes the **risk neutral probability distribution** of S_t, and thus can be regarded as the *continuous time extension* of $\widetilde{\Psi}$.

ii) The name <u>pricing model</u> comes from the fact that the dynamics of S_t follows the risk neutral probability distribution for pricing purposes. Again, the physical probability is never taken into account.

iii) The continuous compounding $B_t = e^{rt}$ is already a continuous time pricing model <u>for the bond</u>.

□

Using the Continuous Time Pricing Model (1.29), we can readily price one particular kind of financial derivative: the derivative with payoff taking the form $f(S_T)$, i.e., a function of the underlying price at maturity.

Theorem 1.9. (Pricing under Continuous Time Model) *Consider a financial instrument F under the Continuous Time Pricing Model. If the payoff of F at T is $F_T = f(S_T)$, then the price F^0 is given by*

$$F^0 = E_{\psi}(f(S_T)) = e^{-rT} \int f\left(S_0 \exp\left\{\left(r - \frac{1}{2}\sigma^2\right) T + \sigma x\right\}\right) \phi_T(x)\, dx,$$

where $\phi_T(x)$ is the density of the Normal distribution $N(0,T)$.

Proof. Recall that the price (before discounting) of an instrument F with payoff $F_T = f(S_T)$ is the expected of payoff F_T under the risk neutral probability distribution of S_t. Thus, from (1.29), the price is

$$\begin{aligned} \mathrm{E}_{\tilde{\psi}}(f(S_T)) &= \mathrm{E}_{\tilde{\psi}}(f(S_0 \exp\left\{\left(r - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right\})) \\ &= \int f\left(S_0 \exp\left\{\left(r - \frac{1}{2}\sigma^2\right) T + \sigma x\right\}\right) \phi_T(x)\, dx, \end{aligned}$$

where $\phi_T(x)$ is the density of the Normal distribution $N(0,T)$, and $\mathrm{E}_{\tilde{\psi}}$ is the expectation under the risk neutral probability distribution of S_t. The price is thus obtained from discounting $\mathrm{E}_{\tilde{\psi}}(f(S_T))$ with the discount factor e^{-rT}, and the proof is complete. $\square$

Example 1.13. **(Black-Scholes' Call Option Price).**
Recall that the European Call option with pre-specified strike price K has a payoff structure $f(S_T) = (S_T - K)^+$ at maturity time T. Thus Theorem 1.9 is applicable, and the price of the Call option is given by (after some straightforward but tedious algebra)

$$\begin{aligned} F^0 &= e^{-rT} \int \left(S_0 \exp\left\{ \left(r - \frac{1}{2}\sigma^2 \right) T + \sigma x \right\} - K \right)^+ \phi_T(x)\, dx \\ &= N(d_1)S_0 - N(d_2)Ke^{-rT}\,, \end{aligned}$$

where $N(\cdot)$ is the distribution function of a standard Normal random variable, $d_1 = \left[\ln\left(\frac{S_0}{K}\right) + (r+\sigma^2/2)T\right]/\sigma\sqrt{T}$ and $d_2 = d_1 - \sigma\sqrt{T}$. This pricing formula is first developed by Black and Scholes in 1973.

This section determines the risk neutral probability distribution for S_t for pricing purpose, which is analogous to evaluating the state price vector in the discrete time models. More advanced results will require more machinery in probability theory and will be discussed in the latter chapters.

1.8 Geometric Brownian Motion

A well-known model for the dynamics of financial assets is to describe the asset return as a **Geometric Brownian Motion**:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t\,, \tag{1.30}$$

where

- S_t is the value (price) of a financial asset at time t.
- dS_t/S_t is the return (the % change in price).
- μ is the *drift*.
- σ is the *volatility*.
- W_t is a *Brownian Motion*.

In fact, it can be shown that (1.30) is equivalent to the continuous time model (1.29). Therefore, (1.30) can be regarded as another formulation for deriving the pricing procedure in continuous time. With such a formulation, it is easier to price more complicated products and construct replicating portfolios.

However, there is a considerable amount of probability theory behind (1.30). We will try to fully understand these theories before putting the equation into practical use. For example,

- What is a Brownian Motion? How do we understand Brownian Motion as a random function? What properties does it have?
- What is the relationship between W_t and $W_{t+\delta}$?
- The Brownian motion satisfies $W_t \sim N(0,t)$. How do we justify its existence?
- What does it mean by differentiating a Brownian Motion in (1.30)?
- What is a stochastic process?
- What is weak convegence $\stackrel{D}{\to}$ of stochastic process?

This course aims to provide a solid mathematical treatment to mathematical finance based on modern probability theories. Starting from the fundamental issue of precisely defining a random variable, we build on concepts including sigma field, probability measure, Lesbegue integration, conditional probabilities and expectations, martingales, stopping times, Ito Calculus and stochastic integrals. After acquiring all the relevant machinery, we can establish a unified theory for mathematical finance.

1.9 Exercises

Exercise 1.10 *Let* $(S_1^1, S_2^1, S_1^2, S_2^2) = (1,2,3,6)'$.

i) Write down the matrix $\mathbf{S}$.
ii) For the portfolio with weight $\mathbf{w} = (3,4)'$*, find the payoff vector.*
iii) Find a weight for the portfolio such that the payoff is $\mathbf{F} = (3,9)'$.
iv) Can you find a weight for the portfolio with payoff $\mathbf{F} = (4,9)'$*? Is the market complete?*
v) If the initial price is $\mathbf{S}^0 = (1,4)'$*. Is there any arbitrage opportunity?*
vi) If the initial price is $\mathbf{S}^0 = (1,3)'$*. Is there any arbitrage opportunity?*

Exercise 1.11 *consider*

$$\mathbf{S}_a^0 = (1,1)', \ \ \mathbf{S}_a = \begin{pmatrix} 8 & 3 \\ 3 & 10 \\ 7 & 9 \end{pmatrix}, \quad \mathbf{S}_b^0 = (1,1,1)', \ \ \mathbf{S}_b = \begin{pmatrix} 8 & 3 & 3.001 \\ 3 & 10 & 9.999 \\ 7 & 9 & 9 \end{pmatrix}.$$

Is there any arbitrage opportunity for the market asset structure $\{\mathbf{S}_a^0, \mathbf{S}_a\}$*? Is there any arbitrage opportunity for the market asset structure* $\{\mathbf{S}_b^0, \mathbf{S}_b\}$*?*

Exercise 1.12 *Given that*

$$\mathbf{S} = \begin{pmatrix} 2 & 3 \\ 1 & 4 \\ 5 & 0 \end{pmatrix}, \mathbf{F} = (5,5,5)', \mathbf{S}^0 = (2,1)'.$$

i) *How many assets are there in this market?*
ii) *How many scenarios are there in this market?*
iii) *What is the value of the 2nd asset under the 1st scenario?*
iv) *What is the value of the third asset under the 1st scenario?*
v) *Is the market complete? Why?*
vi) *What is the price of F at time 0?*
vii) *If $F^0 = 3.5$, does arbitrage opportunity exist? If yes, explain how could one take the arbitrage opportunity.*
viii) *What is the replicating portfolio?*

Exercise 1.13 *Suppose that S_1, S_2 have payoffs $(10, -10, 3)'$ and $(9, -9, 2)'$ in 3 possible scenarios at T.*

i) *If $\mathbf{S}^0 = (1,1)'$. Is there any arbitrage opportunity?*
ii) *If $\mathbf{S}^0 = (1,0.8)'$. Is there any arbitrage opportunity?*

Assume that $\mathbf{S}^0 = (1,0.8)'$. For any asset A with initial value $A^0 > 0$ and payoff $(A^1, \ldots, A^m)'$, one may quantify risk by the criterion

$$Risk = \max_{j=1\ldots,m} \frac{-A^j}{A^0},$$

which measures the maximum of negative gain per unit initial value.

iii) *Find the Risk for portfolio $V_a = S_1$.*
iv) *Find the Risk for the portfolio $V_b = S_1 - S_2$.*
v) *Can you construct a portfolio using linear combinations of S_1 and S_2 with the minimum risk?*

Exercise 1.14 *Calculate the state price vector for the asset structure in Example 1.7. How is it connect to the importance of the states discussed in Example 1.8.*

Exercise 1.15 *Find the replicating portfolio for F in Example 1.7. If $F^0 = 5$, state how you take the arbitrage opportunity.*

Exercise 1.16 *Consider the eight combinations of the following three pairs of disjoint events:*

i) $\{m > n\}$, $\{m < n\}$.
ii) *{ the market is complete },{ the market is not complete }.*
iii) *{ no arbitrage opportunity exists }, { arbitrage opportunity does exist }.*

For each of the eight cases, give an example of the market in terms of $\mathbf{S}$, $\mathbf{F}$ *and* $\mathbf{S}^0$, *with columns of* $\mathbf{S}$ *being linearly independent. (Hint: not all combinations are possible.)*

Exercise 1.17 *Consider a future contract with strike price $K = 105$ and maturity $T = 1$. Suppose the initial price of the underlying stock is $S_0 = 100$ and the one-period risk-free rate is $r = 30\%$. The factor u and d are 1.2 and 0.9 respectively. If we model the dynamics by the one-period binomial tree model, then prove that arbitrage exists by*

i) constructing a portfolio explicitly;
ii) using Theorem 1.22;
iii) using Theorem 1.7.

Exercise 1.18 *Note that the choice of $u = exp(\sigma\sqrt{\delta}z)$, where $z = \pm 1$, is not arbitrary. Repeat the same derivation for* (1.29) *using $u = exp(\sigma(\delta)^q z)$ for $q < 1/2$ and $q > 1/2$. What is the resulting limit?*

Exercise 1.19 *Show equation* (1.27).

Exercise 1.20 *Give the detailed calculations for Example 1.13.*

1.10 Appendix: Pricing in the incomplete market, $m \neq n$

In Sections 1.2 and 1.3, we assumed $m = n$ which implies a complete market and facilitates the derivation of explicit formulas (see Theorem 1.3). Specifically, the portfolio weight $\mathbf{w} = \mathbf{S}^{-1}\mathbf{F}$ in (1.7) and the state price vector $\psi' = (\mathbf{S}^0)'\mathbf{S}^{-1}$ in (1.9) can only be defined when $\mathbf{S}^{-1}$ exists ($m = n$). For general m, n, the inverse $\mathbf{S}^{-1}$ does not exist and hence a more complicated treatment is required.

Although we relax Assumption 1.2 to consider general m, n, we keep Assumption 1.1 of linearly independent columns to avoid redundant assets. Note that each column of $\mathbf{S}$ is an m-dimensional vector summarizing the value of an asset at m different scenarios. Thus, the requirement of linearly independent columns implies that the number of columns (number of assets) n is not greater than m. Therefore, we focus on the case $m \geq n$.

When $m > n$, the n asset vectors are not enough to span the space $\mathbb{R}^m$ of all possible scenarios. Thus, not all payoff can be replicated. i.e., the **market is not complete**. The following theorem suggests a simple method to determine whether a given payoff $\mathbf{F}$ is replicable, and obtain the weights of the replicating portfolio upon existence.

Theorem 1.21. *Given a market with asset structure $\{\mathbf{S}^0, \mathbf{S}\}$ and a payoff $\mathbf{F}$. Then, the payoff $\mathbf{F}$ is replicable if and only if*

$$\mathbf{F} = \mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\mathbf{F}. \tag{1.31}$$

Moreover, if (1.31) *holds, then the weight of the replicating portfolio is given by*

$$\mathbf{w} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\mathbf{F}. \tag{1.32}$$

Proof. If $\mathbf{F}$ is replicable, then $\mathbf{F} = \mathbf{Sw}$ for some $\mathbf{w}$. Substituting to the right side of (1.31) gives

$$\mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\mathbf{F} = \mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'(\mathbf{Sw}) = \mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}(\mathbf{S}'\mathbf{S})\mathbf{w} = \mathbf{Sw} = \mathbf{F},$$

verfiying (1.31). Conversely, if (1.31) holds, then substituting (1.32) to (1.31) gives $\mathbf{F} = \mathbf{S}\mathbf{w}$, which is exactly the definition that $\mathbf{F}$ is replicable by $\mathbf{w}$. □

Remark 1.8. Readers who are familiar with statistics may recognize that $\mathbf{w}$ is the regression coefficient of regressing $\mathbf{F}$ against $\mathbf{S}$. Indeed, the financial concept of "replicating the payoff $\mathbf{F}$ by the assets $\mathbf{S}$" is equivalent to the statistical concept of "explaining the response $\mathbf{F}$ by the predictors $\mathbf{S}$".

From regression theory, if (1.31) holds, then $\mathbf{F}$ is perfectly explained by $\mathbf{S}$, i.e., $\mathbf{F}$ is replicable by $\mathbf{S}$. Even if (1.31) does not hold, the weight $\mathbf{w} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\mathbf{F}$ represents a portfolio $\mathbf{S}\mathbf{w}$ that is *closest* to the payoff $\mathbf{F}$ in the sense that the *least squares criterion* $(\mathbf{F}-\mathbf{S}\mathbf{w})'(\mathbf{F}-\mathbf{S}\mathbf{w})$ is minimized. □

Next we study the properties of the state price vector ψ under the incomplete market $m > n$ where $\mathbf{S}^{-1}$ does not exist. Although ψ cannot be defined as $(\mathbf{S}^0)'\mathbf{S}^{-1}$ as in (1.9), the idea of "weighting payoffs in different scenarios to obtain the initial price" in (1.10) motivates the following definition.

Definition 1.6. (**State Price Vector.**) A state price vector is an m-dimensional vector $\psi \gg 0$ satisfying

$$(\mathbf{S}^0)' = \psi'\mathbf{S}\,. \tag{1.33}$$

The notation $\gg$ means all the entries of ψ are strictly positive. □

In (1.9), the state price vector ψ is constructed explicitly. However, in the case $m > n$, the existence of ψ satisfying (1.33) requires justifications. Moreover, in Proposition 1.5, it is shown that $\psi \gg 0$ if there is no arbitrage opportunity in the market. The proof relies heavily on market completeness because it assumes that $\mathbf{1}_{k,m}$ is replicable for all $k = 1,\dots,m$. Although this argument does not work in the case $m > n$, the result remains valid.

Theorem 1.22. (No Arbitrage and State Price Vector). *Under Assumption 1.1, there is no arbitrage opportunity in the market if and only if a state price vector $\psi \gg 0$ in Definition 1.6 exists.*

Proof. The proof in Proposition 1.5 that $\psi \gg 0$ implies no arbitrage opportunity remains valid here because the market completeness is not needed for this direction.

It remains to prove the converse: no arbitrage opportunity implies the existence of $\psi \gg 0$ satisfying (1.33). Equivalently, we now show that if there does not exist $\psi \gg 0$ satisfying (1.33), then there exists arbitrage opportunities.

First we introduce some geometric interpretation of (1.33). Let $\mathbf{s}_1,\dots,\mathbf{s}_m$ be the row vectors of the $m \times n$ matrix $\mathbf{S}$ and $\mathbb{C} = \{\sum_{i=1}^m \alpha_i \mathbf{s}_i | \, \alpha_i > 0, i = 1,\dots,m\}$ be the **convex cone** of $\{\mathbf{s}_1,\dots,\mathbf{s}_m\}$. The name *cone* is motivated by the shape of the set $\mathbb{C}$, see Figure 1.1. Note that (1.33) can be expressed as $(\mathbf{S}^0)' = \sum_{i=1}^m \psi_i \mathbf{s}_i$; $\psi_i > 0$, i.e., (1.33) is equivalent to $(\mathbf{S}^0)' \in \mathbb{C}$.

Now, if there does not exist $\psi \gg 0$ satisfying (1.33), i.e., $(\mathbf{S}^0)' \notin \mathbb{C}$, then it is clear that we can find a hyperplane $P = \{\mathbf{x} \in \mathbb{R}^m | \mathbf{w}'\mathbf{x} = 0\}$ that *separates* $(\mathbf{S}^0)'$ and

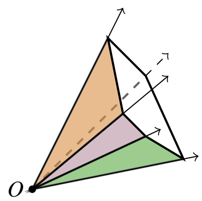


Fig. 1.1 Example of $\mathbb{C}$.

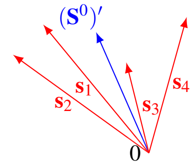


Fig. 1.2 $(\mathbf{S}^0)' \in \mathbb{C}$.

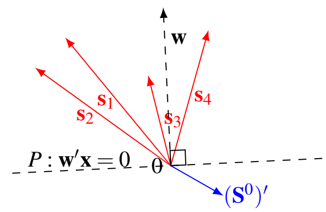


Fig. 1.3 $(\mathbf{S}^0)' \notin \mathbb{C}$.

$\{\mathbf{s}_1, \dots, \mathbf{s}_m\}$ (dash line in Figures 1.3). By definition, $\mathbf{w}'\mathbf{x} = 0$ for all $\mathbf{x} \in P$, this is, $\mathbf{w}$ is orthogonal to every element $\mathbf{x}$ in P. Thus, $\mathbf{w}$ is known as the *normal vector* of the plane P. Since the hyperplane $P = \{\mathbf{x} \in \mathbb{R}^m | \mathbf{w}'\mathbf{x} = 0\}$ is unchanged if we replace $\mathbf{w}$ by $-\mathbf{w}$, $\mathbf{w}$ can be chosen on the same side of the $\mathbf{s}_i$s and opposite to $(\mathbf{S}^0)'$. By the definition of dot product ($\mathbf{a}'\mathbf{b} \triangleq \mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos \theta$, where θ is the angle between $\mathbf{a}$ and $\mathbf{b}$), we have

$$(\mathbf{S}^0)'\mathbf{w} = \mathbf{w}'\mathbf{S}^0 \leq 0\,, \qquad \text{and} \qquad \mathbf{S}\mathbf{w} > \begin{pmatrix} \mathbf{s}_1'\mathbf{w} \\ \vdots \\ \mathbf{s}_m'\mathbf{w} \end{pmatrix} \geq 0\,,$$

which, by (1.2), implies that $\mathbf{w}$ defines a portfolio enjoying arbitrage opportunities. This completes the proof. □

Theorem 1.22 is seemingly striking, however, it is not useful in practice because it does not discuss how to check whether a $\psi \gg 0$ exists for a given market asset structure $\{\mathbf{S}^0, \mathbf{S}\}$. Also, when arbitrage exists (not exists), we do not get the state price vector ψ (arbitrage portfolio $\mathbf{w}$) explicitly. Nevertheless, we will see in the following that ψ and $\mathbf{w}$ satisfy some linear equations with inequality constraints. Thus **linear programming** can be employed to solve for ψ or $\mathbf{w}$.

Remark 1.9. (**Negative and Non-unique State Price Vector.**) Following the idea in deriving (1.6), since (1.32) gives a replicating portfolio $\mathbf{w}$, the price of any replicable payoff $\mathbf{F}$ can be computed by

$$F^0 = (\mathbf{S}^0)'\mathbf{w} = (\mathbf{S}^0)'(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\mathbf{F} = \psi'\mathbf{F}\,, \tag{1.34}$$

where $\psi' = (\mathbf{S}^0)'(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$. However, such ψ may not be used as the state price vector because it may not have strictly positive entries. See Exercise 1.24.

Since $\mathbf{S}$ is of size $m \times n$, $m > n$, there exists infinitely many non-zero vectors b satisfying $\mathbf{b}'\mathbf{S} = \mathbf{0}$. For such b, we can define $\psi_b' = (\mathbf{S}^0)'(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}' + b$ which satisfies (1.33). Therefore, if no arbitrage opportunity exists, then there exists infinitely many state price vectors of the form ψ_b. □

In the following we develop the **Check for Arbitrage (CFA)** algorithm, which determines the existence of arbitrage opportunity and constructs state price vector or arbitrage portfolio using linear programing techniques. First we review the **standard form** of linear programing problem.

Standard form

For fixed $n \times m$ matrix A, $\mathbf{c} \in \mathbb{R}^m$, and $\mathbf{b} \in \mathbb{R}^n$,

$$\max_{\mathbf{x}\in\mathbb{R}^m} \mathbf{c}'\mathbf{x} \qquad \text{subject to} \qquad A\mathbf{x} \le \mathbf{b}\,, \mathbf{x} \ge 0\,. \tag{1.35}$$

The dual of (1.35) is

$$\min_{\mathbf{y}\in\mathbb{R}^n} \mathbf{y}'\mathbf{b} \qquad \text{subject to} \qquad A'\mathbf{y} \ge \mathbf{c}\,, \mathbf{y} \ge 0\,. \tag{1.36}$$

If (1.35) has a solution $\mathbf{x}^*$ and (1.36) has a solution $\mathbf{y}^*$, then

$$\mathbf{c}'\mathbf{x}^* = (\mathbf{y}^*)'\mathbf{b}\,.$$

The standard form is readily solvable in many computing softwares, such as the command `simplex` in the `boot` package of `R`.

Next, we see how the standard form of linear programing is closely connected to portfolio weight and state price vector. Consider a market with asset structure $\{\mathbf{S}^0, \mathbf{S}\}$ and a derivative of F with payoff $\mathbf{F}$ and price F^0. Putting $(A, \mathbf{b}, \mathbf{c}, \mathbf{x}, \mathbf{y}) = (\mathbf{S}', \mathbf{S}^0, \mathbf{F}, \psi, \mathbf{w})$ in (1.35) and (1.36), we arrive at the following familiar equations:

$$\max_{\psi\in\mathbb{R}^m} \psi'\mathbf{F} \qquad \text{subject to} \qquad \psi'\mathbf{S} \le (\mathbf{S}^0)'\,, \psi \ge 0 \tag{1.37}$$

$$\min_{\mathbf{w}\in\mathbb{R}^n} \mathbf{w}'\mathbf{S}^0 \qquad \text{subject to} \qquad \mathbf{S}\mathbf{w} \ge \mathbf{F}\,, \mathbf{w} \ge 0\,. \tag{1.38}$$

Intuitively, $\mathbf{w}'\mathbf{S}^0$ is the price of a portfolio (weight $\mathbf{w}$) with payoff $\mathbf{S}\mathbf{w}$ not less than $\mathbf{F}$ at the maturity. If $\mathbf{w}'\mathbf{S}^0 < F^0$, then arbitrage opportunity exists. On the other hand, if the equality sign holds for the first constraint of (1.37), then ψ is the state price vector and $\psi'\mathbf{F}$ is the price of F.

However, if we are only given an asset structure $\{\mathbf{S}^0, \mathbf{S}\}$, we do not have $\{F^0, \mathbf{F}\}$. How can we apply (1.37) and (1.38)? The following lemma suggests a particular asset structure such that (1.37) and (1.38) are applicable. More importantly, this lemma shows that the existence of arbitrage opportunities is closely related to the solution of a linear programming of the form (1.38).

Lemma 1.1. *Suppose that $m > n$ and the asset structure $\{\mathbf{S}^0, \mathbf{S}\}$ is of the form*

$$\mathbf{S}^0 = \mathbf{1}_n\,, \qquad \mathbf{S} = \begin{pmatrix} D_{n-1} & \mathbf{0} \\ \widetilde{\mathbf{S}} & \mathbf{F} \end{pmatrix}, \tag{1.39}$$

where $\mathbf{1}_n = (1, 1, \ldots, 1) \in \mathbb{R}^n$, D_{n-1} is a diagonal matrix of size $n-1$ with strictly positive entries, $\widetilde{\mathbf{S}}$ is an $(m-n+1) \times (n-1)$ matrix, $\mathbf{F} \in \mathbb{R}^{m-n+1}$. Then arbitrage opportunities exist in the market if and only if $(\mathbf{w}_*)'\mathbf{1}_{n-1} < 1$, where $\mathbf{w}_*$ is the solution to the linear programming problem

$$\min_{\widetilde{\mathbf{w}} \in \mathbb{R}^{n-1}} \widetilde{\mathbf{w}}' \mathbf{1}_{n-1} \qquad \textit{subject to} \qquad \widetilde{\mathbf{S}}\widetilde{\mathbf{w}} \geq \mathbf{F}, \widetilde{\mathbf{w}} \geq 0\,. \tag{1.40}$$

Proof. Denote $\mathbf{w} = (\mathbf{w}_s, w_f)$, where $\mathbf{w}_s \in \mathbb{R}^{n-1}$, $w_f \in \mathbb{R}$. Recall from (1.2) that existence of arbitrage opportunities is equivalent to the existence of $\mathbf{w}$ such that

$$\begin{aligned}
&\mathbf{w}' S^0 < 0 \quad \text{and} \quad \mathbf{S}\mathbf{w} \geq \mathbf{0} \\
\Leftrightarrow\ & \mathbf{w}_s' \mathbf{1}_{n-1} < -w_f\,,\ \widetilde{\mathbf{S}}\mathbf{w}_s + w_f \mathbf{F} \geq \mathbf{0}\,,\ \mathbf{w}_s \geq \mathbf{0} && (\text{since } D_{n-1} \gg 0 \text{ is diagonal}) \\
\Leftrightarrow\ & \left(\frac{\mathbf{w}_s'}{-w_f}\right)\mathbf{1}_{n-1} < 1\,,\ \widetilde{\mathbf{S}}\left(\frac{\mathbf{w}_s}{-w_f}\right) \geq \mathbf{F}\,,\ \left(\frac{\mathbf{w}_s}{-w_f}\right) \geq \mathbf{0} && (\text{since } -w_f > \mathbf{w}_s'\mathbf{1}_{n-1} \geq 0) \\
\Leftrightarrow\ & \widetilde{\mathbf{w}}' \mathbf{1}_{n-1} < 1\,,\ \widetilde{\mathbf{S}}\widetilde{\mathbf{w}} \geq \mathbf{F}\,,\ \widetilde{\mathbf{w}} \geq \mathbf{0}\,, && \left(\text{replace } \tfrac{\mathbf{w}_s}{-w_f} \text{ by } \widetilde{\mathbf{w}}\right)
\end{aligned}$$

which is equivalent to $\mathbf{w}_*' \mathbf{1}_{n-1} < 1$, where $\mathbf{w}_*$ is the solution to (1.40). □

Remark 1.10. Lemma 1.1 suggests that the problem of not having a derivative F can be tackled by regarding the last asset as F. Moreover, note that upper part of $\mathbf{S}$ consists of the entry-wise positive diagonal matrix D_{n-1} and the zero matrix $\mathbf{0}$. This indicates that the *i*-th asset is the only valuable asset in the *i*-th scenario. We say that the ***i*-th asset dominates the *i*-th scenario**. The consequence is that, when arbitrage opportunity exists, the condition $\mathbf{S}\mathbf{w} > 0$ in (1.3) forces the weight $\mathbf{w}_s$ to be non-negative. Hence, we arrive at (1.40), which achieves the standard form (1.38).

We look also at a specific example where we can easily deduce no arbitrage opportunity without using linear programming.

Lemma 1.2. *Suppose that $m > n$ and the asset structure $\{\mathbf{S}^0, \mathbf{S}\}$ is of the form*

$$\mathbf{S}^0 = \mathbf{1}_n\,, \qquad \mathbf{S} = \begin{pmatrix} D_n \\ \widetilde{\mathbf{S}} \end{pmatrix}, \tag{1.41}$$

where $\mathbf{1}_n = (1, 1, \ldots, 1) \in \mathbb{R}^n$, D_n is a diagonal matrix of size n with strictly positive entries, $\widetilde{\mathbf{S}}$ is an $(m-n) \times n$ matrix. Then there is no arbitrage opportunities in the market. A state price vector is given by $\psi \in \mathbb{R}^m_+$, where

$$\psi_i = \begin{cases} \frac{1 - c\sum_{k=n+1}^m \widetilde{S}_i^k}{d_i}\,, & i = 1, \ldots, n\,, \\ c\,, & i = n+1, \ldots, m\,, \end{cases} \tag{1.42}$$

where

$$c = \frac{0.9}{\max_{j=1,\ldots,n} \left| \sum_{k=n+1}^m \widetilde{S}_j^k \right|}\,, \tag{1.43}$$

and $d_i > 0$ is the i-th diagonal entry of D_n.

Proof. Recall from (1.2) that existence of arbitrage opportunities is equivalent to the existence of $\mathbf{w} \in \mathbb{R}^m$ such that

$$
\begin{aligned}
& \mathbf{w}' S^0 < 0 \quad \text{and} \quad \mathbf{S}\mathbf{w} \geq \mathbf{0} \\
\Longleftrightarrow & \sum_{j=1}^{n} w_j < 0\,,\ D_n \mathbf{w} \geq \mathbf{0} \quad \text{and} \quad \widetilde{S}\mathbf{w} \geq \mathbf{0} \qquad \text{(definition of } \mathbf{S}^0 \text{ and } \mathbf{S}) \\
\Longleftrightarrow & \sum_{j=1}^{n} w_j < 0\,,\ \mathbf{w} \geq \mathbf{0} \quad \text{and} \quad \widetilde{S}\mathbf{w} \geq \mathbf{0}\,. \qquad \text{(since } D_n \gg 0)
\end{aligned}
$$

Since $\sum_{j=1}^{n} w_j < 0$ and $\mathbf{w} \geq \mathbf{0}$ cannot happen together, there is no arbitrage opportunity. From the definitions of ψ in (1.42) and $\mathbf{S}^0$, $\mathbf{S}$ in (1.41), direct calculations yield that ψ has strictly positive entries and satisfies $\psi'\mathbf{S} = (\mathbf{S}^0)'$. Thus it satisfies the definition of state price vector in (1.33). $\square$

The main idea of the CFA algorithm is to transform the asset structure $\{\mathbf{S}^0, \mathbf{S}\}$ to $\{\widetilde{\mathbf{S}}^0, \widetilde{\mathbf{S}}\}$ which takes the form in (1.39) or (1.41). The transformations are given by $(\widetilde{\mathbf{S}}^0)' = (\mathbf{S}^0)'\mathbf{M}$ and $\widetilde{\mathbf{S}} = \mathbf{S}\mathbf{M}$, where $\mathbf{M}$ is an invertible matrix of size $n \times n$. Multiplying $(\mathbf{S}^0)'$ and $\mathbf{S}$ by $\mathbf{M}$ on the right can be interpreted as forming a new set of assets $\{\widetilde{S}_1, \ldots, \widetilde{S}_n\}$, where the j-th new asset is

$$
\widetilde{S}_j = M_{1,j} S_1 + M_{2,j} S_2 + \cdots + M_{n,j} S_n\,,
$$

and $M_{k,j}$ is the (k, j) entry of $\mathbf{M}$. Let ψ and $\mathbf{w}$ be the state price vector and portfolio weight corresponding to $\{\mathbf{S}^0, \mathbf{S}\}$. By the definitions of ψ and $\widetilde{\mathbf{S}}$, we have

$$
\psi'\widetilde{\mathbf{S}} = \psi'\mathbf{S}\mathbf{M} = (\mathbf{S}^0)'\mathbf{M} = (\tilde{S}^0)'\,,\ \text{and}\ \widetilde{\mathbf{S}}(\mathbf{M}^{-1}\mathbf{w}) = \mathbf{S}\mathbf{M}\mathbf{M}^{-1}\mathbf{w} = \mathbf{S}\mathbf{w}\,, \quad (1.44)
$$

i.e., the state price vector and portfolio weight corresponding to $\{\tilde{S}^0, \widetilde{\mathbf{S}}\}$ are ψ and $\mathbf{M}^{-1}\mathbf{w}$ respectively. The following lemma summarizes four typical transformations.

Lemma 1.3. *In each of* T1 *to* T4 *below, the asset structure* $\{\widetilde{\mathbf{S}}^0, \widetilde{\mathbf{S}}\}$ *is equivalent to* $\{\mathbf{S}^0, \mathbf{S}\}$ *in the sense that they can be transformed between each others by* $(\widetilde{\mathbf{S}}^0)' = (\mathbf{S}^0)'\mathbf{M}$, $\widetilde{\mathbf{S}} = \mathbf{S}\mathbf{M}$, *and* $(\mathbf{S}^0)' = (\widetilde{\mathbf{S}}^0)'\mathbf{M}^{-1}$, $\mathbf{S} = \widetilde{\mathbf{S}}\mathbf{M}^{-1}$.

[T1]**(Replacing asset S_{j_2} by $aS_{j_1} + bS_{j_2}$)** *For any $a,b \in \mathbb{R}$,*

$$
\widetilde{\mathbf{S}}^0 = \begin{pmatrix} S_1^0 \\ \vdots \\ aS_{j_1}^0 + bS_{j_2}^0 \\ \vdots \\ S_n^0 \end{pmatrix},
\widetilde{\mathbf{S}} = \begin{pmatrix} S_1^1 & \cdots & aS_{j_1}^1 + bS_{j_2}^1 & \cdots & S_n^1 \\ S_1^2 & \cdots & aS_{j_1}^2 + bS_{j_2}^2 & \cdots & S_n^2 \\ \vdots & & \vdots & & \vdots \\ S_1^m & \cdots & aS_{j_1}^m + bS_{j_2}^m & \cdots & S_n^m \end{pmatrix},
\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \cdots & a & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots & b & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix},
$$

where the a and b in M are located at the (j_1, j_2) and (j_2, j_2) entry, respectively.
[T2]**(Exchanging S_{j_1} and S_{j_2})**

$$\widetilde{\mathbf{S}}^0 = \begin{pmatrix} S_1^0 \\ \vdots \\ S_{j_2}^0 \\ \vdots \\ S_{j_1}^0 \\ \vdots \\ S_n^0 \end{pmatrix}, \widetilde{\mathbf{S}} = \begin{pmatrix} S_1^1 & \cdots & S_{j_2}^1 & \cdots & S_{j_1}^1 & \cdots & S_n^1 \\ S_1^2 & \cdots & S_{j_2}^2 & \cdots & S_{j_1}^2 & \cdots & S_n^2 \\ \vdots & & \vdots & & \vdots & & \vdots \\ S_1^m & \cdots & S_{j_2}^m & \cdots & S_{j_1}^m & \cdots & S_n^m \end{pmatrix}, \mathbf{M} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix},$$

where the two off-diagonal 1*'s in M are located at the* (j_1, j_2) *and* (j_2, j_1) *entries.*
[T3] **(Making asset S_j dominate scenario i)** *For* $S_j^i \neq 0$*, let* $c_k = S_k^i / S_j^i$*,*

$$\widetilde{\mathbf{S}}^0 = \begin{pmatrix} S_1^0 - S_j^0 c_1 \\ S_2^0 - S_j^0 c_2 \\ \vdots \\ S_j^0 \\ \vdots \\ S_n^0 - S_j^0 c_n \end{pmatrix}, \widetilde{\mathbf{S}} = \begin{pmatrix} S_1^1 - S_j^1 c_1 & \cdots & S_j^1 & \cdots & S_n^1 - S_j^1 c_n \\ S_1^2 - S_j^2 c_1 & \cdots & S_j^2 & \cdots & S_n^2 - S_j^2 c_n \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & S_j^i & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ S_1^m - S_j^m c_1 & \cdots & S_j^m & \cdots & S_n^m - S_j^m c_n \end{pmatrix}, \mathbf{M} = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ -c_1 & -c_2 & \cdots & 1 & \cdots & -c_n \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix},$$

where $(-c_1, -c_2, \ldots, 1, \ldots, -c_n)$ *is the* j*-th row of* $\mathbf{M}$.
[T4] **(Standardizing the initial price)** *Suppose that* $S_i^0 \neq 0$ *for all* $i = 1, \ldots, n$. *Let* $R_j^i = S_j^i / S_j^0$.

$$\widetilde{\mathbf{S}}^0 = \mathbf{1}_n, \widetilde{\mathbf{S}} = \begin{pmatrix} R_1^1 & R_2^1 & \cdots & R_n^1 \\ R_1^2 & R_2^0 & \cdots & R_n^2 \\ \vdots & \vdots & & \vdots \\ R_1^m & R_2^m & \cdots & R_n^m \end{pmatrix}, \mathbf{M} = \begin{pmatrix} 1/S_1^0 & 0 & \cdots & 0 \\ 0 & 1/S_2^0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/S_n^0 \end{pmatrix}.$$

Proof. In each of T1 to T4, verify directly that $(\widetilde{\mathbf{S}}^0)' = (\mathbf{S}^0)'\mathbf{M}$ and $\widetilde{\mathbf{S}} = \mathbf{SM}$. $\square$

Check for Arbitrage (CFA) Algorithm for asset structure $\{\mathbf{S}^0, \mathbf{S}\}$:

1. (**Setting non-zero initial prices**) Let $j_1 \in \{1, \ldots, n\}$ be the smallest index such that $S^0_{j_1} \neq 0$. For all $j_2 \neq j_1$ such that $S^0_{j_2} = 0$, apply T1 with $(a,b) = (1,1)$.
2. (**Standardizing initial prices**) Apply T4. Set counting index $k = 1$ and the row index $R = \{1, \ldots, m\}$ for Steps 3 to 5.
3. (**Determining dominating scenario**) Let $(i^*, j^*) = \arg\max_{i \geq k, j \geq k} S^i_j$ and $j_\dagger = \arg\min_{j \geq k} S^{i^*}_j$ (pick the smallest index if not unique).
   a. If $S^{i^*}_{j^*} \leq 0$, then arbitrage opportunity exists: For each $j \geq k$, $\tilde{S}^0_j = 1$ in the beginning and $S^i_j \leq S^{i^*}_{j^*} \leq 0$ at maturity indicates that a short position of $\tilde{S}_j$ creates arbitrage profit.
   b. If $S^{i^*}_{j^*} > 0$ and $S^{i^*}_{j^*} > S^{i^*}_{j_\dagger}$, then proceed to 4.
   c. If $S^{i^*}_{j^*} > 0$ and $S^{i^*}_{j^*} = S^{i^*}_{j_\dagger}$, take the smallest $i \geq k$ such that $\max_{j \geq k} S^i_j > \min_{j \geq k} S^i_j$. Rename i and the maximizer and minimizer of S^i_j as $(i^*, j^*, j_\dagger)$.
      i. If $S^{i^*}_{j^*} > 0$, then proceed to 4.
      ii. If $S^{i^*}_{j^*} \leq 0$, then apply T1 with $(a,b,j_1,j_2) = \left(-\frac{S^{i^*}_{j^*} + S^{i^*}_{j_\dagger}}{2S^{i^*}_{j_\dagger}}, 1, j_\dagger, j^*\right)$.
      Then apply T4 and proceed to Step 4.
4. **(Setting dominating scenario)** Define $J = \{j \neq j^* | S^{i^*}_j \geq S^{i^*}_{j^*}\}$. For each $j \in J$, apply T1 with $(a,b,j_1,j_2) = \left(\frac{S^{i^*}_j - S^{i^*}_{j^*}}{S^{i^*}_{j^*} - S^{i^*}_{j_\dagger}} + 0.1, 1, j_\dagger, j\right)$. Next apply T3 to make asset S_{j^*} dominate scenario i^*. Then apply T4.
5. **(Rearranging columns)** Apply T2 to exchange the k-th and the j^*-th asset. Exchange the k-th and the i^*-th row of $\mathbf{S}$. Exchange the k-th and the i^*-th element of R. Set $k = k+1$.
6. Repeat Steps 3 to 5 up to $k = n-1$.
7. **(Final column)** For $k = n$, the current asset structure $\{\mathbf{S}^0, \mathbf{S}\}$ is of the form

$$\mathbf{S}^0 = \mathbf{1}_n, \qquad \mathbf{S} = \begin{pmatrix} D_{n-1} & \mathbf{0} \\ \widetilde{\mathbf{S}} & \mathbf{F} \end{pmatrix},$$

   in Lemma 1.1.
   a. If $\mathbf{F} \leq 0$, then arbitrage opportunity exists (a short position on S_n).
   b. If $F_i > 0$ and $F_i > \max_{j<n} S^i_j$ for some $i \in \{n, \ldots, m\}$, where S^i_j is the (i,j) entry of $\mathbf{S}$, then do Step 4 and 5. The resulting asset structure is of the form (1.41) in Lemma 1.2. Hence, no arbitrage opportunity exists and a state price vector can be constructed by (1.42).
   c. If the conditions in Step 7b do not hold, then solve the linear programming program (1.40) and use Lemma 1.1 to determine the existence of arbitrage opportunity and search for the arbitrage portfolio or state price vector.

Remark 1.11. **(Interpretations)** The intuition of CFA algorithm is discussed below:

a) Step 1 ensures that all initial values are non-zero so that standardization T4 in Step 2 can be performed.
b) In Steps 3 and 4 we search for a unique maximum value $S^{i^*}_{j^*}$ to conduct T3. The fact that $S^{i^*}_{j^*}$ is the maximum positive value ensures that the transformed initial price $\tilde{S}^0$ remains positive. Hence, the sign of the entries of $\mathbf{S}$ does not change when we standardize again using T4.
c) After the standardization in Step 4, we have $\mathbf{S}^0 = \mathbf{1}_n$ and the $S^{i^*}_{j^*} > 0$ is the only non-zero entry in the i^*-th row of $\mathbf{S}$. Step 5 then moves the (i^*, j^*)-entry to the (k,k)-entry of $\mathbf{S}$ to achieve the form in (1.39). The exchange of columns is handled by T2. The exchange of row is merely a renaming of scenarios and does not change the structure of the problem. Moreover, the row index R stores the correspondence between the original and the swapped scenarios. Thus, we can rename the scenarios in the end of the algorithm.
d) Once Steps 3 to 5 are completed for a fixed k, the j-th assets (of the current asset structure) dominates the j-th scenario, $j = 1, 2, \ldots, k$. That is, only the j-th asset can produce profit in the j-th scenario. To proceed to search for dominating assets in other scenarios, the first k assets cannot be used again since another T3 will destroy the original dominance. This explains why the search spaces have to be restricted to $i \geq k$ and $j \geq k$.
e) In Step 3c, $S^{i^*}_{j^*} = S^{i^*}_{j_\dagger}$ means that $S^{i^*}_k = S^{i^*}_{k+1} = \cdots = S^{i^*}_n$. In this case the maximum is not unique and T3 will produce initial prices of zero-value. To avoid this, we give up the i^* row and try another row i which satisfies $\max_{j \geq k} S^i_j > \min_{j \geq k} S^i_j$. Note that the situation $S^i_k = S^i_{k+1} = \cdots = S^i_n$ could not happen for all rows. Otherwise, the k-th to the n-th columns of $\mathbf{S}$ are identical, which violates the linear independence assumption.
f) In Step 3(c)ii, if $S^{i^*}_{j^*} \leq 0$, then $0 < S^{i^*}_{j^*}/S^{i^*}_{j_\dagger} < 1$. Apply T1 with $(a, b, j_1, j_2) = \left(-\frac{S^{i^*}_{j^*} + S^{i^*}_{j_\dagger}}{2S^{i^*}_{j_\dagger}}, 1, j_\dagger, j^*\right)$ yields $\tilde{S}^{i^*}_{j^*} = \frac{1}{2}\left(S^{i^*}_{j^*} - S^{i^*}_{j_\dagger}\right) > 0$ and also $\tilde{S}^0_{j^*} = a + b = \frac{1}{2}\left(1 - \frac{S^{i^*}_{j^*}}{S^{i^*}_{j_\dagger}}\right) > 0$, recovering a unique maximum positive value at (i^*, j^*) entry.
g) In Step 4, the $S^{i^*}_{j^*}$ may not be the unique maximum value in the i^*-th row. Also, there may be bigger $S^i_{j^*}$ values at $i < k$. To tackle this, the transformation T1 uses the smallest entry $S^{i^*}_{j_\dagger}$ to "dampen" the bigger entries so that $S^{i^*}_{j^*}$ becomes the unique maximum. To see this, if $S^{i^*}_j > S^{i^*}_{j^*}$, then in order that the transformations T1 and T4 in Step 4 give $\tilde{S}^{i^*}_{j^*} = S^{i^*}_{j^*} > \tilde{S}^{i^*}_j$, we must have

$$\frac{aS^{i^*}_{j_\dagger} + bS^{i^*}_j}{a+b} < S^{i^*}_{j^*} \iff \frac{a}{b} > \frac{S^{i^*}_j - S^{i^*}_{j^*}}{S^{i^*}_{j^*} - S^{i^*}_{j_\dagger}},$$

which motivates the choice of (a, b). Of course, the 0.1 in a can be replaced by any positive number. □

Remark 1.12. **(Implementations)**

1. The linear programming problem in Step 7 can be solved by classical algorithm such as the **Simplex method**. The command `simplex` in the `boot` package of `R` can be used.
2. Step 1 does not apply to the case $S_j^0 = 0$ for all j. In this case, we can pick one scenario of $\mathbf{S}$ and treat it as $-\mathbf{S}^0$ so that the previous method can be used. Specifically, the initial price of any portfolio is $\sum_{j=1}^n \alpha_j S_j^0 = 0$. Hence, arbitrage exists if we can find $\mathbf{w} \in \mathbb{R}^m$ such that $\mathbf{S}\mathbf{w} > \mathbf{0}$ (see (1.2)). Without loss of generality, assume that $\mathbf{S} = \begin{pmatrix} -(\widetilde{\mathbf{S}}^0)' \\ \widetilde{\mathbf{S}} \end{pmatrix}$, where $\widetilde{\mathbf{S}}^0 \neq \mathbf{0}$ is in $\mathbb{R}^n$ and $\widetilde{\mathbf{S}}$ is an $(m-1) \times n$ matrix. Then $\mathbf{S}\mathbf{w} > \mathbf{0}$ is equivalent to either "$(\widetilde{\mathbf{S}}^0)'\mathbf{w} \leq 0$ and $\widetilde{\mathbf{S}}\mathbf{w} > \mathbf{0}$" or "$(\widetilde{\mathbf{S}}^0)'\mathbf{w} < 0$ and $\widetilde{\mathbf{S}}\mathbf{w} \geq \mathbf{0}$", which is the definition (1.2) of arbitrage for the asset structure $(\widetilde{\mathbf{S}}^0, \widetilde{\mathbf{S}})$. Since $\widetilde{\mathbf{S}}^0 \neq \mathbf{0}$, the CFA algorithm can be employed on $(\widetilde{\mathbf{S}}^0, \widetilde{\mathbf{S}})$.

□

We conclude this section by the the following pricing theorem under general situations, i.e., without assuming $m = n$ or completeness of the market.

Theorem 1.23. (Pricing)
Suppose that there exists a state price vector in the market. If a payoff $\mathbf{F}$ *is replicable, then the price is given by*

$$F^0 = \psi'\mathbf{F} \triangleq \overline{\psi} E_{\widetilde{\psi}}(\mathbf{F}),$$

where $E_{\widetilde{\psi}}$ *is the expectation under the risk neutral probability distribution* $\widetilde{\psi}$*. For notational simplicity, sometimes we write* $F^0 = E_{\psi}(\mathbf{F})$*. It is a slight abuse of notation because* ψ *is not a probability distribution.*

Proof. If $\mathbf{F}$ is replicable, then $\mathbf{F} = \mathbf{S}\mathbf{w}$ for some $\mathbf{w}$. By the principle of no arbitrage pricing and (1.33), the price is given by $F^0 = (S^0)'\mathbf{w} = \psi'\mathbf{S}\mathbf{w} = \psi'\mathbf{F}$. □

1.11 Exercises

Exercise 1.24 *Give a numerical example of asset structure such that there is no arbitrary opportunity, but the state price vector constructed in Remark 1.9 has negative entries.*

Exercise 1.25 *consider* $\mathbf{F}_a = (6,8,9)'$, $\mathbf{F}_b = (9,4,1,3)'$,

$$\mathbf{S}_a^0 = (1,1)', \quad \mathbf{S}_a = \begin{pmatrix} 8 & 3 \\ 3 & 10 \\ 7 & 9 \end{pmatrix}, \quad \mathbf{S}_b^0 = (1,2,2)', \quad \mathbf{S}_b = \begin{pmatrix} 8 & 3 & 3 \\ 3 & 10 & 5 \\ 7 & 9 & 7 \\ 2 & -3 & 6 \end{pmatrix}.$$

i) Is $\mathbf{F}_1$ *replicable in market* $\{\mathbf{S}_a^0, \mathbf{S}_a\}$*?*

ii) Is $\mathbf{F}_2$ *replicable in market* $\{\mathbf{S}_b^0, \mathbf{S}_b\}$*?*

iii) Use CFA algorithm to check if arbitrage opportunity exists for the market $\{\mathbf{S}_a^0, \mathbf{S}_a\}$ *and* $\{\mathbf{S}_b^0, \mathbf{S}_b\}$*. If arbitrage opportunity exists, find an arbitrage portfolio. Otherwise, find a state price vector.*

Chapter 2
Measure Theory and Probability

In **probability theory**, the formal definition of random variable (r.v.) involves **measure theory**, which is a discipline on generalizing the concepts of length. In particular, one works with the **Probability Space** $(\Omega, \mathscr{F}, \mathbb{P})$, where

- Ω is the **sample space**,
- $\mathscr{F}$ is the **σ-algebra** of Ω, and
- $\mathbb{P}$ is the **probability measure**.

In this chapter we motivate the necessities of using measure theory on probability. First, in Sections 2.1 and 2.2 we discuss the problem of measuring the size of a bizarre subset of the real line. Since assigning probability to events is closely connected to measuring the size of subsets of the real line, this problem implies that we are not able to assign probability to all possible events. Hence, we need to focus on some *nice* subsets ($\mathscr{F}$) that we can assign probabilities. This motivates the need of defining the probability space, which will be elaborated in the subsequent sections.

2.1 Countability of set

First, we review some basic definitions in mathematics about the size of a set: **finite, countable** and **uncountable**.

Definition 2.1. For any positive integer n, let $J_n = \{1, 2, \cdots, n\}$ and $J = \mathbb{Z}^+$, the set of all positive integers. For two sets X and Y, denote $X \sim Y$ if there exists a bijective function between X and Y, or equivalently, elements in X and Y can be put into one-one correspondence. Then, for any set A, we say:

i) A is *finite* if $A \sim J_n$ for some n (the empty set is also considered to be finite).
ii) A is *infinite* if A is not finite.
iii) A is *countable* if $A \sim J$ or $A \sim J_n$ for some n.
iv) A is *uncountable* if A is neither finite nor countable.

Example 2.1. If $A = \{9,5,2,7\}$, then we can define a function $f : A \to J_4$ such that $f(9) = 1$, $f(5) = 2$, $f(2) = 3$, $f(7) = 4$. Since f is bijective, A is finite. □

Example 2.2. If A is the set of even positive integers $\{2,4,\cdots\}$, then we can define a bijective function $f : A \to J$ such as $f(x) = x/2$. (It is straightforward to check that f is a bijective function; see Exercise 2.4). Thus, A is countable. However, A is not finite since the number of elements in A is greater than n for any integer n. Therefore, no bijective function $f : A \to J_n$ can be defined for any fixed n. □

Example 2.3. The set of integers $A = \{\cdots,-3,-2,-1,0,1,2,3,\cdots\}$ is countable, although it seems to be twice as large as J. In this case, a bijective function $f : A \to J$ can be defined, e.g., $f(x) = (2x+1)1_{\{x\geq 0\}} + (-2x)1_{\{x<0\}}$. Note that f need not be unique. (Exercise 2.4). □

Example 2.4. A sequence $\{x_i\}_{i=1,2,3,\ldots}$ is clearly countable (e.g. $f(x_i) = i$). A **double array** $\{x_{ij}\}_{i=1,2,3\cdots,\ j=1,2,3\cdots}$ is also countable. The way of counting is achieved by the following **diagonal argument**:

$$
\begin{tikzcd}
x_{11} & x_{12} & x_{13} & x_{14} & \cdots \\
x_{21} \arrow[ur] & x_{22} \arrow[ur] & x_{23} \arrow[ur] & x_{24} & \cdots \\
x_{31} \arrow[ur] & x_{32} \arrow[ur] & x_{33} & x_{34} & \cdots \\
x_{41} \arrow[ur] & x_{42} & x_{43} & x_{44} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \cdots
\end{tikzcd}
$$

□

Example 2.5. Rational number $\mathbb{Q} = \{\frac{p}{q} : p,q \in \mathbb{Z}, q \neq 0\}$ is countable, since $\mathbb{Q}$ can be regarded as $\{x_{pq}\}_{p,q\in J}$. Thus the result of Example 2.4 can be applied directly.

Example 2.6. Let A be the set of infinite sequence of binary numbers, i.e., A is a set with elements taking the form such as $0.101100\cdots$ and $0.011001100\cdots$. The set A is uncountable. To see this, suppose on the contrary that A is countable. Then, there exists a bijective function $f : A \to J$ such that one can enumerate the elements in A as $x_1, x_2, \ldots, x_n, \ldots$. For example,

$$
\begin{aligned}
x_1 &= 0.\underline{1}011000\cdots \\
x_2 &= 0.0\underline{0}10011\cdots \\
x_3 &= 0.10\underline{1}0011\cdots \\
x_4 &= 0.100\underline{0}011\cdots \\
\vdots &= \qquad \vdots
\end{aligned}
\tag{2.1}
$$

Now we point out a contradiction by obtaining a binary number taking the form $y = 0.a_1a_2a_3\cdots$, where $a_i = 0$ or 1, but y does not belong to A. The construction of y is as follows: $a_i = 1-$ the i-th digit of x_i. That is, change the i-th digit of x_i from 0 to 1 or from 1 to 0. Continuing the example in (2.1), $y = 0.0101\cdots$. By construction, for any integer n, $y \neq x_n$ since their n-th digits are different. Thus $y \notin A$, contradicting that A contains all the infinite sequences of binary numbers.

The same argument can be generalized to show that real numbers are uncountable (Exercise 2.7). □

2.2 Measurablility of Sets

Intuitively, we can measure the size of any subset of the real line using intervals of the form $I = [a,b]$, $a,b \in \mathbb{R}$. Define $l(E)$ as the measure of the size of a set E. Obviously, $l([a,b]) = b-a$ and l is **countably additive** in the sense that

$$l(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} l(E_i)\,, \tag{2.2}$$

if E_is are disjoint. In this section we first consider four sets, E_1,E_2,E_3,E_4, that can be measured using intervals. Finally, we will introduce a bizarre set E_5 which cannot be measured using intervals.

2.2.1 Singleton

The set of a single element such as

$$E_1 := \left\{\frac{1}{3}\right\},$$

is known as a **singleton**. Obviously, the length, or the measure, of E_1 is 0. Formally, we use the following argument: For any ε, $l(E_1) \leq l\left([\frac{1}{3}-\frac{\varepsilon}{2}, \frac{1}{3}+\frac{\varepsilon}{2}]\right) = \varepsilon$. Since ε is arbitrary, $l(E_1) = 0$.

In general, a set with “zero length” is known as *null set*.

Definition 2.2. (Null set) A null set $A \subseteq \mathbb{R}$ is a set that can be covered by a sequence of intervals of arbitrarily small total length, i.e. given any $\varepsilon > 0$ we can find a sequence $\{I_n : n \geqslant 1\}$ of intervals such that $A \subseteq \bigcup_{n=1}^{\infty} I_n$ and $\sum_{n=1}^{\infty} l(I_n) < \varepsilon$. In particular, $l(A) = 0$. □

Remark 2.1. All empty sets are null, but not all null sets are empty (e.g. E_1).

2.2.2 Open Intervals

By definition we measure sets by *closed* intervals $[a,b]$, $a,b \in \mathbb{R}$. However, whether the boundary is closed or open is irrelevant. To find the measure of the open interval

$$E_2 := \left(0, \frac{1}{3}\right),$$

we use the countably additivity property of measure (2.2),

$$l\left(\left[0, \frac{1}{3}\right]\right) = l(\{0\}) + l(E_2) + l\left(\left\{\frac{1}{3}\right\}\right).$$

As l is well defined on closed sets and singletons, we have $l(E_2) = \frac{1}{3} - 0 - 0 = \frac{1}{3}$.

2.2.3 Rational Numbers

We have seen that singleton has measure 0. What if we have infinitely many singletons, such as

$$E_3 := [0,1] \cap \mathbb{Q},$$

the set of all rational numbers on the interval $[0,1]$?

Surprisingly, the set E_3 is a null set although it contains infinitely many elements. To show that $l(E_3) = 0$, from Definition 2.2, we need to construct a *cover* $\bigcup_{i=1}^{\infty} I_n$ for E_3 with arbitrary small measure: Since E_3 is a countable set, it can be arranged in the form of $E_3 = \{x_1, x_2, \cdots\}$. Fix $\varepsilon > 0$, a cover of E_3 can be constructed by the following sequence of intervals:

$$\begin{aligned}
I_1 &= (x_1 - \frac{\varepsilon}{8}, x_1 + \frac{\varepsilon}{8}), \\
I_2 &= (x_2 - \frac{\varepsilon}{16}, x_2 + \frac{\varepsilon}{16}), \\
I_3 &= (x_3 - \frac{\varepsilon}{32}, x_3 + \frac{\varepsilon}{32}), \\
&\vdots \\
I_n &= (x_n - \frac{\varepsilon}{2^{n+2}}, x_n + \frac{\varepsilon}{2^{n+2}}),, \\
&\vdots
\end{aligned}$$

Observe that $l(I_n) = \frac{\varepsilon}{2} \cdot \frac{1}{2^n}$. Since $\sum_{n=1}^{\infty} \frac{1}{2^n} = 1$, we have $E_3 \subseteq \bigcup_{n=1}^{\infty} I_n$ and $\sum_{n=1}^{\infty} l(I_n) = \frac{\varepsilon}{2} < \varepsilon$. Hence, E_3 is a null set.

Although rational numbers can be found nearly everywhere in $[0,1]$, from the fact that $l([0,1]) = 1$ and $l(E_3) = 0$, there are far more irrational numbers in $[0,1]$ than rational numbers.

2.2.4 Cantor Set

Not only countable set but also uncountable sets can be null. The *Cantor set* is a typical example of uncountable null set.

Define the interval

$$C_0 = [0,\, 1].$$

Remove the segment $(\frac{1}{3},\, \frac{2}{3})$, and let the union of the remaining segments as

$$C_1 = \left[0,\, \frac{1}{3}\right] \cup \left[\frac{2}{3},\, 1\right].$$

Remove the middle thirds of these intervals, and let C_2 be the union of the intervals

$$C_2 = \left[0,\, \frac{1}{9}\right] \cup \left[\frac{2}{9},\, \frac{3}{9}\right] \cup \left[\frac{6}{9},\, \frac{7}{9}\right] \cup \left[\frac{8}{9},\, 1\right].$$

Continuing in this way, we obtain a sequence of C_n such that $C_1 \supset C_2 \supset C_3 \supset \cdots$ and C_n is the union of 2^n intervals, each of length 3^{-n}. The limit

$$E_4 = \bigcap_{n=1}^{\infty} C_n\,, \tag{2.3}$$

is defined as the Cantor set.

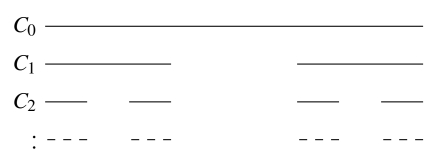


Fig. 2.1 Cantor set $E_4 = \bigcap_{n=1}^{\infty} C_n$.

The Cantor set is closely related to ternary expansion: To locate a point in $[0,1]$, we may use $0,1,2$ to indicate whether the point is in the first, middle or last thirds of $[0,1]$. For example, a 2 indicates the point is located in the last 1/3 of $[0,1]$, i.e., $[\frac{2}{3},1]$. Then, we can further use $0,1,2$ to indicate whether the point is in the first, middle or last thirds of $[\frac{2}{3},1]$. This process goes on and on and finally the point can be accurately located. Consequently, every point in $[0,1]$ can be expressed in the form $x = 0.a_1a_2\cdots$ with $a_k = 0,1,2$, which is called **ternary expansion**. For example, the point $x_1 \in [\frac{6}{9}, \frac{7}{9}]$ has a ternary expansion $x_2 = 0.20\cdots$, and the point $x_2 \in [\frac{1}{9}, \frac{2}{9}]$ has a ternary expansion $x_1 = 0.01\cdots$.

By construction, the Cantor set E_4 always removes the middle thirds of an interval. It follows that the digit "1" never appears in the ternary expansion of any $x \in E_4$. In other words, $x \in E_4$ if and only if $a_k = 0$ or 2 for all k in the ternary expansion $x = 0.a_1a_2\cdots$. Using the same argument as in Example 2.6, we see that the **Cantor set** E_4 **has uncountably many elements**.

Next we show that E_4 is a null set. Given any $\varepsilon > 0$, choose a sufficiently large n such that $\left(\frac{2}{3}\right)^n < \varepsilon$. Since $E_4 \subseteq C_n$ and C_n is a union of 2^n intervals of length 3^{-n}, we have $l(E_4) \leq l(C_n) = \left(\frac{2}{3}\right)^n < \varepsilon$. Since ε is arbitrary, $l(E_4) = 0$.

2.2.5 Non-measurable Set

The previous examples involve sets that can be measured by operations of intervals. However, the variety of subsets in $[0,1]$ can be far more complicated. In this section we define a subset of $[0,1]$, E_5, which is not measurable by operations of intervals.

First we introduce the **axiom of choice**, which is a fundamental set-theoretical axiom for the construction of E_5.

Axiom 2.2.1 (The axiom of choice) *Suppose that $\mathscr{A} = \{A_\alpha : \alpha \in \Gamma\}$ is a non-empty collection, indexed by a set Γ, of non-empty disjoint subsets of Ω. Then there exists a set $E \subset \Omega$ which contains precisely one element from each of the sets A_α.* □

Example 2.7. Let $\Gamma = \{1,2,\ldots,n\}$ and $A_\alpha = [\alpha - 0.1, \alpha + 0.1]$ for $\alpha \in \Gamma$, then it is easy to see that we can take $\alpha - 0.05 \in A_\alpha$ for $\alpha = 1,\ldots,n$ to form the set $E = \{0.95, 1.95, \ldots, n - 0.05\}$. However, if Γ is an uncountable set, say $\mathbb{R}$, then we cannot enumerate the elements in Γ, and hence it is hard to imagine how to choose precisely one element from each A_α. This motivates the need of Axiom 2.2.1. □

The construction of E_5 involves two steps:

<u>Step 1:</u> Define a **grouping rule** of real numbers as follows:

If $y - x \in \mathbb{Q}$, then we say $x \sim y$, or x and y are in the same group.

For example, $\sqrt{2} \sim \sqrt{2} + 0.1$, but $\sqrt{2} \nsim \sqrt{3} + 0.1$ and $e \nsim \pi$. Using this rule, we can partition $[0,1]$ into disjoint *equivalence classes* $\{A_\alpha, \alpha \in \Gamma\}$, for some index set Γ, such that $[0,1] = \bigcup_{\alpha \in \Gamma} A_\alpha$. One can imagine

$$A_{\bar{\alpha}} = \left(\bigcup_{q\in\mathbb{Q}} \{\sqrt{2}+q\}\right) \cap [0,1], \ A_{\tilde{\alpha}} = \left(\bigcup_{q\in\mathbb{Q}} \{\sqrt{3}+q\}\right) \cap [0,1], \ A_{\check{\alpha}} = \left(\bigcup_{q\in\mathbb{Q}} \{\pi+q\}\right) \cap [0,1], \tag{2.4}$$

etc…, for $\bar{\alpha}$, $\tilde{\alpha}, \check{\alpha} \in \Gamma$. By construction, any two elements $x, y \in A_\alpha$ differ by a rational number. For example, if $x, y \in A_{\bar{\alpha}}$, then $x - y = \sqrt{2} + q_x - (\sqrt{2} + q_y) =$

$q_x - q_y$ is rational for some $q_x, q_y \in \mathbb{Q}$. On the other hand, elements of different classes always differ by an irrational number. For instance, if $x \in A_{\bar{\alpha}}$ and $y \in A_{\tilde{\alpha}}$, then $x - y = \sqrt{2} - \sqrt{3} + q_x - q_y$ is irrational.
Note also that each A_α is countable (since $\mathbb{Q}$ is countable). However, since $[0,1] = \bigcup_{\alpha \in \Gamma} A_\alpha$ and $[0,1]$ is uncountable, there must be uncountably many different classes, i.e., Γ is uncountable (see Exercise 2.5). This is why we use the notations $\bar{\alpha}$, $\tilde{\alpha}$, $\check{\alpha}$ instead of $\alpha_1, \alpha_2, \alpha_3$ in (2.4).

Step 2: Use the **axiom of choice** to construct $E_5 \subset [0,1]$ by picking exactly one member a_α from each of the class A_α. That is,

$$E_5 = \bigcup_{\alpha \in \Gamma} \{a_\alpha\} .$$

Since Γ is uncountable, E_5 is an uncountable set. Also, since elements of different classes always differ by an irrational number, we have the property

$$\text{If } x, y \in E_5 \text{ and } x \neq y, \text{ then } x - y \text{ is an irrational number.} \tag{2.5}$$

This set is developed by an Italian mathematician Giuseppe Vitali in 1905 and is known as the **Vitali set**.

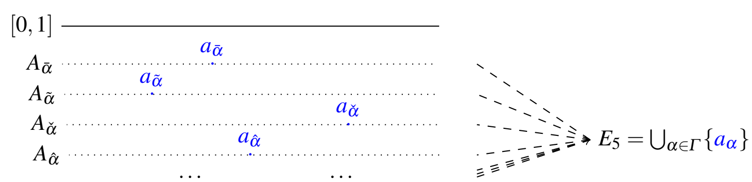


Fig. 2.2 Vitali set E_5. $A_\alpha = \left(\bigcup_{q \in \mathbb{Q}} \{a_\alpha + q\}\right) \cap [0,1]$.

To show that E_5 is not measurable, we define a sequence of *translates* of E_5 by

$$S_n = E_5 + q_n , \tag{2.6}$$

where $\{q_n\}$ is the enumeration of the rationals in $[-1,1]$ (we can enumerate as rationals are countable). Note the following properties from the construction:

1) If E_5 is measurable, then so is each S_n and $l(S_n) = l(E_5)$ for all n.
2) The sets in $\{S_n\}$ are disjoint:

   Proof: Suppose on the contrary that there exists $z \in S_m \cap S_n$ for some $m \neq n$. Then we can write $z = a_\alpha + q_m = a_\beta + q_n$ for some $a_\alpha, a_\beta \in E_5$, $a_\alpha \neq a_\beta$, and $q_m, q_n \in \mathbb{Q}$. Rearranging gives $a_\alpha - a_\beta = q_n - q_m \in \mathbb{Q}$. This contradicts Property (2.5) and thus such a z does not exist.

3) $[0,1] \subseteq \bigcup_{n=1}^{\infty} S_n \subseteq [-1,2]$.

<u>Proof</u>: By definition, E_5 contains exactly one element a_α from each A_α. From the grouping rule, each element of A_α is of the form $a_\alpha + q_n$ (for some n), which belongs to $S_n = E_5 + q_n$. As a result, $A_\alpha \subset \bigcup_{n=1}^\infty S_n$ for all $\alpha \in \Gamma$, and thus $[0,1] = \bigcup_{\alpha\in\Gamma} A_\alpha \subseteq \bigcup_{n=1}^\infty S_n$. Since $E_5 \subset [0,1]$ and $q_n \in [-1,1]$ in (2.6), $S_n = E_5 + q_n \subset [-1,2]$ and thus $\bigcup_{n=1}^\infty S_n \subseteq [-1,2]$.

By countable additivity and monotonicity of l, the above three properties imply:

$$1 = l([0,1]) \overset{(3)}{\leqslant} l(\cup_{n=1}^\infty S_n) \overset{(2)}{=} \sum_{n=1}^\infty l(S_n) \overset{(1)}{=} \sum_{n=1}^\infty l(E_5) \overset{(3)}{\leqslant} l([-1,2]) = 3\,,$$

where the second to fifth (in)equalities follow from properties 3,2,1,3, respectively. The conclusion $\sum_{n=1}^\infty l(E_5) \in [1,3]$ is clearly impossible since an infinite sum of the same positive constant can either be 0 or ∞. Hence, we have proved by contradiction that E_5 is not measurable.

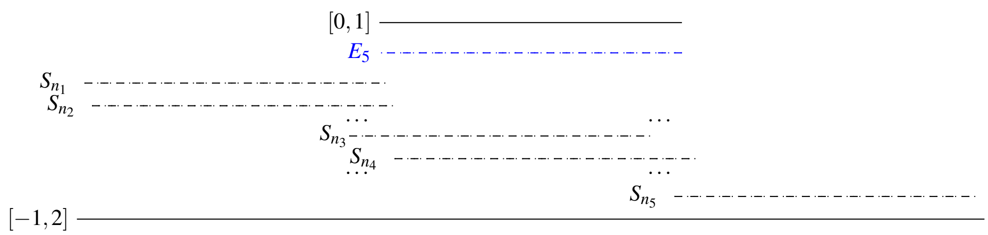


Fig. 2.3 Vitali set E_5 is not measurable. $S_n = E_5 + q_n$. $l(E_5) = l(S_n)$ for all n. $[0,1] \subseteq \bigcup_{n=1}^\infty S_n \subseteq [-1,2]$.

2.3 Probability Space

The Vitali set in Section 2.2.5 indicates that there are complicated sets that cannot be measured intuitively by using intervals. Therefore, it is necessary to restrict our consideration on a collection of some “nice” enough sets for practical purposes. To achieve this, we define the **Probability Space** as follows.

Definition 2.3. *(Probability space)* A probability space is a triple $(\Omega, \mathscr{F}, \mathbb{P})$, where

- Ω, **the Sample Space**, is a set describing all possible events.
- $\mathscr{F}$, the **σ-field**, is a class of subsets of Ω. These subsets are the “measurable sets”, where we can assign values about their probabilities.
- $\mathbb{P}$, the **Probability Measure**, is a function $\mathbb{P} : \mathscr{F} \to [0,1]$ satisfying $\mathbb{P}(\Omega) = 1$ and $\mathbb{P}(\cup_{i=1}^\infty A_i) = \sum_{i=1}^\infty \mathbb{P}(A_i)$ for disjoint $A_i \in \mathscr{F}$.

□

Note that the domain of the function $\mathbb{P}$ is the σ-field $\mathscr{F}$, which contains **subsets of elements** of Ω. Precisely speaking, if $\omega \in \Omega$, then $\mathbb{P}(\{\omega\})$ is well defined if $\{\omega\} \in \mathscr{F}$, but $\mathbb{P}(\omega)$ is undefined (ω is an element of Ω but not an element of $\mathscr{F}$).

In the following subsections we elaborate the elements in the triple $(\Omega, \mathscr{F}, \mathbb{P})$ one by one.

2.3.1 *Sample Space*

The **Sample Space** Ω is an arbitrary set which may be finite, countable or uncountable. Its elements are usually denoted by ω. Generally, Ω consists of all possible outcomes ω of an experiment or observation. An element ω of Ω is called a **sample point** and a subset of Ω (e.g. $\{\omega\}, \{\omega_1, \ldots, \omega_n\}$) is called an **event**.

Example 2.8. a) When you randomly place your finger on a ruler of length 1, then

$$\Omega_a = \{\text{The finger points to } x |\ x \in [0,1]\}.$$

b) When a die is thrown once, the sample space is

$$\Omega_b = \{\text{The face of the die is } i |\ i = 1,2,3,4,5,6\}.$$

c) To describe the results of the n tosses of a coin, the sample space is

$$\Omega_c = \{(\omega_1, \omega_2, \cdots, \omega_n) : \omega_i = H \text{ or } T\},$$

which is a set of size 2^n. Each element in Ω is a vector of length n.

d) To describe the position of a particle, the space is

$$\Omega_d = \{\text{The particle is at } x |\ x \in \mathbb{R}^3\},$$

where $\mathbb{R}^3$ is the three-dimensional Euclidean space.

e) For the price of a stock, the sample space Ω_e is an abstract space describing the whole operation in the company, which is difficult to quantify.

□

2.3.2 *σ-Field*

Intuitively, σ-field is a class of subsets of Ω (i.e. class of events) that we want to assign probabilities to. When Ω is finite, the class of all possible subsets of Ω is the power set 2^Ω, which is also finite. (The notation 2^Ω is motivated by the fact that any possible subset $A \subseteq \Omega$ can be formed by providing each element in Ω with two choices: to be included or not included in A). If the probability $\mathbb{P}(\{\omega\})$ is defined for each sample point $\omega \in \Omega$, then the probabilities of all events (all elements in

2^{Ω}) can be defined according to the **finite additivity rule** $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ for disjoint events A and B.

However, if Ω is infinite or uncountable, then it becomes problematic to assign probabilities to events (e.g. Vitali set in Section 2.2.5). Therefore, mathematicians resort to identifying a class of sets which is simple but useful enough for practical purposes. It is natural to think of a class of sets that is closed under the fundamental operations in set theory including union, intersection and complement. This class of sets is known as σ-*field* or σ- *algebra*.

Definition 2.4. **(σ-Field)** A class $\mathscr{F}$ of subsets of Ω is called a σ-*field* if

1. $\emptyset, \Omega \in \mathscr{F}$;
2. $A \in \mathscr{F} \Rightarrow A^c \in \mathscr{F}$; and
3. $A_1, A_2, \cdots \in \mathscr{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$.

□

Remark 2.2. Note that the definition of σ-field implies the closure of countable intersections, i.e., $A_1, A_2, \cdots \in \mathscr{F} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathscr{F}$ (Exercise 2.8).

Given a class of subset $\mathscr{F}$ of Ω, we can determine whether $\mathscr{F}$ is a σ-field by directly verifying the three conditions in Definition 2.4.

Example 2.9. Let $\mathscr{F} = \{(-\infty, x) | x \in \mathbb{R} \cup \{\pm\infty\}\}$. Then $\mathscr{F}$ is closed under countable intersection or unions: If $A_1 = (-\infty, a_1), \ldots, A_n = (-\infty, a_n), \ldots \in \mathscr{F}$, then $\bigcup_{i=1}^{n} A_i = (-\infty, \sup_{i \geq 1} a_i) \in \mathscr{F}$, and $\bigcap_{i=1}^{n} A_i = (-\infty, \inf_{i \geq 1} a_i) \in \mathscr{F}$. However, $\mathscr{F}$ is not closed under complement: $(-\infty, 1)^c = [1, \infty) \notin \mathscr{F}$. Hence, $\mathscr{F}$ is not a σ-field. □

Example 2.10. Suppose that $\mathscr{F}$ consists of the countable and the co-countable sets (A is **co-countable** if A^c is countable) in Ω. First we verify that $\mathscr{F}$ is a σ-field:

1. $\emptyset \in \mathscr{F}$ as empty set is countable. On the other hand, Ω is co-countable because $\emptyset = \Omega^c$ is countable. Thus, $\Omega \in \mathscr{F}$.
2. If $A \in \mathscr{F}$, then A is either countable or co-countable. Thus by definition, A^c is either co-countable or countable, implying that $A^c \in \mathscr{F}$.
3. If $A_1, A_2, \cdots \in \mathscr{F}$, then A_is are either countable or co-countable. If all A_is are countable, then the diagonal argument in Example 2.4 yields that $\bigcup_{i=1}^{\infty} A_i$ is countable. If some A_k is co-countable, then A_k^c is countable. Thus $\left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c$ is countable because it is a subset of the countable set A_k^c. In both cases, $\bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$.

If $\Omega = [0, 1]$, then $A = [0, 0.5]$ has the property that A and A^c are both uncountable. Thus $A \notin \mathscr{F}$, which shows that a σ-field may not contain all the subsets of Ω.

Furthermore, we may write $A = \bigcup_{x \in [0, 0.5]} \{x\}$ as the (uncountable) union of each $x \in A$. Since $\{x\} \in \mathscr{F}$ and $A \notin \mathscr{F}$, we see that a σ-field may not be closed under the formation of uncountable unions. □

Example 2.11. The largest σ-field in Ω is the *power set* 2^{Ω}, consisting of all possible subsets of Ω; the smallest σ-field is $\{\emptyset, \Omega\}$. The verifications are straightforward and are deferred to Exercise 2.9. □

When the sample space Ω is finite, σ-fields can be explicitly expressed by enumeration work.

Example 2.12. Let $\Omega = \{1,2,3,4,5,6\}$. Then it is straightforward to verify from definition that $\mathscr{F}_1 = \{\emptyset, \{1,2,3\}, \{4,5,6\}, \Omega\}$ and $\mathscr{F}_2 = \{\emptyset, \{1,3,5\}, \{2,4,6\}, \Omega\}$ are σ-fields. Note that if one is interested in "small/large" or "odd/even", then $\mathscr{F}_1$ and $\mathscr{F}_2$, respectively, are enough for assigning probabilities to possible events. However, $\mathscr{F}_3 = \{\emptyset, \{1,2\}, \{4,6\}, \Omega\}$ is not a σ-field since the set $\{1,2,4,6\} = \{1,2\} \cup \{4,6\}$ does not belong to $\mathscr{F}_3$. □

Sometimes we may have a class of subsets of Ω, say $\mathscr{A}$, to which we want to assign probabilities. To define a probability space, we need a σ-field that contains $\mathscr{A}$. This motivates the following definition.

Definition 2.5. (σ-field generated by a class of sets)
Let $\mathscr{A}$ be a class of subsets of Ω. The σ-field *generated* by $\mathscr{A}$, denoted by $\sigma(\mathscr{A})$, is the smallest σ-field that contains $\mathscr{A}$. □

Example 2.13. Let $\Omega = \{1,2,3,4,5,6\}$ and $\mathscr{A} = \{\{1\}, \{1,3\}\}$. Then it is straightforward to verify from definition that

$$\sigma(\mathscr{A}) = \{\emptyset, \{1\}, \{1,3\}, \{3\}, \{2,4,5,6\}, \{2,3,4,5,6\}, \{1,2,4,5,6\}, \Omega\}.$$ □

When Ω is finite, one can always construct the σ-field explicitly by union and complement operations. However, when Ω and $\mathscr{A}$ are not finite, then one does not have explicit expressions for the elements of $\sigma(\mathscr{A})$ in general. Nevertheless, the following theorem guarantees the existence of $\sigma(\mathscr{A})$.

Theorem 2.1. (σ-field generated by a class of sets)
Let $\mathscr{A}$ be a class of subsets of Ω. The smallest σ-field generated *by $\mathscr{A}$ is uniquely defined by*

$$\sigma(\mathscr{A}) = \bigcap_{\substack{\mathscr{F} \text{ is a } \sigma\text{-field} \\ \mathscr{A} \subseteq \mathscr{F}}} \mathscr{F}. \tag{2.7}$$

Proof. The existence of $\sigma(\mathscr{A})$ is guaranteed if we can find a σ-field $\mathscr{F}$ such that $\mathscr{A} \subseteq \mathscr{F}$. Obviously, the power set 2^Ω is one example (see Example 2.11).

Next we verify that $\sigma(\mathscr{A})$ is a σ-field containing $\mathscr{A}$. To simplify notation we denote the σ-fields inside the intersection of (2.7) by $\mathscr{F}_\alpha$s, where α belongs to a possibly uncountable index set Λ. Note that $\sigma(\mathscr{A})$ contains $\mathscr{A}$ as every $\mathscr{F}_\alpha$ contains $\mathscr{A}$. Following the standard arguments, we verify that $\sigma(\mathscr{A})$ is a σ-field:

1. $\emptyset, \Omega \in \mathscr{F}$ since $\emptyset, \Omega \in \mathscr{F}_\alpha$ for all $\alpha \in \Lambda$.
2. If $A \in \mathscr{F}$, then $A \in \mathscr{F}_\alpha$ for all $\alpha \in \Lambda$. As $\mathscr{F}_\alpha$s are σ-fields, $A^c \in \mathscr{F}_\alpha$ for all $\alpha \in \Lambda$, and thus $A^c \in \mathscr{F}$.
3. If $A_1, A_2, \cdots \in \mathscr{F}$, then $A_i \in \mathscr{F}_\alpha$ for all $\alpha \in \Lambda$. As $\mathscr{F}_\alpha$s are σ-fields, $\bigcup_{i=1}^\infty A_i \in \mathscr{F}_\alpha$ for all $\alpha \in \Lambda$, and thus $\bigcup_{i=1}^\infty A_i \in \mathscr{F}$.

Finally, the minimality and uniqueness of $\sigma(\mathscr{A})$ can be argued as follows: For any σ-field $\mathscr{F}' \supseteq \mathscr{A}$, $\sigma(\mathscr{A}) \subseteq \mathscr{F}'$ because $\mathscr{F}'$ is one component among the intersection in (2.7). □

Consider the uncountable sample space $\Omega = \mathbb{R}$. One obvious σ-field for $\mathbb{R}$ is the power set $2^{\mathbb{R}}$. However, we are not able to measure the size of all elements in $2^{\mathbb{R}}$ (e.g., Vitali set $E_5 \in 2^{\mathbb{R}}$). Since our intuition of measuring sets is based on intervals, it is fundamental to require a σ-field that contains all possible union/intersection/complement of intervals. Yet, each element in the σ-field is measurable. Definition 2.5 suggests a way to define such a σ-field.

Example 2.14. **(Borel σ-field)**
Let $\Omega = \mathbb{R}$ and let $\mathscr{A}$ be the collection of all finite open intervals, i.e., $\mathscr{A} = \{(l,u) | -\infty < l < u < \infty\}$. The *Borel σ-field* is defined as

$$\mathscr{B} := \mathscr{B}_{\mathbb{R}} = \sigma(\mathscr{A}),$$

which is the smallest σ-field generated by finite open intervals in $\mathscr{A}$. The elements of the Borel σ-field are called *Borel sets*. Particularly, the Borel σ-field on $[a,b]$ is denoted by $\mathscr{B}_{[a,b]} := \{[a,b] \cap B : B \in \mathscr{B}\}$. See Exercises 2.10 and 2.11. □

In Section 2.3.3 we will show that we can formally define a measure λ on $\mathscr{B}$, i.e., for each element $B \in \mathscr{B}$, the value $\lambda(B)$ is well-defined. Hence, by focusing on $\mathscr{B}$, we have a mathematically justified framework to discuss the size of commonly encountered sets, namely the sets generated by union/intersection/complement of intervals.

Example 2.15. **(Borel Product σ-field on $\mathbb{R}^n$)**
The idea in Example 2.14 can be easily generalized from measuring length to measuring area and volume . Consider $\Omega = \mathbb{R}^3$. We can extend the idea to define a σ-field $\mathscr{F}$ for $\mathbb{R}^3$ that contains all open *cubes*:

$$\mathscr{F} = \sigma(I_1 \times I_2 \times I_3 : I_i = (l_i, u_i) \subset \mathbb{R}, \text{ for } i = 1,2,3).$$

Motivated by the name **product space** for $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$, $\mathscr{F}$ is known as **Product σ-field**, and may be denoted as $\mathscr{B}^3 = \sigma(\mathscr{B} \times \mathscr{B} \times \mathscr{B})$. For higher dimensions, the Borel-measure on $\mathbb{R}^n$, $\mathscr{B}^n = \sigma(\mathscr{B} \times \mathscr{B} \times \cdots \times \mathscr{B})$ can be analogously defined. □

We end this subsection with examples of σ-fields associated with the sample spaces in Example 2.8.

Example 2.16. In Example 2.8b),c), since the sample spaces $\Omega_b = \{\text{The face of the die is } i |\, i = 1,2,3,4,5,6\}$ and $\Omega_c = \{(\omega_1, \omega_2, \cdots, \omega_n) : \omega_i = H, T\}$ are finite, it is simple to associate them with the σ-field 2^{Ω}.

In Example 2.8a),d),e), the sample spaces $\Omega_a = \{\text{The finger points to } x |\; x \in [0,1]\}$, $\Omega_d = \{\text{The particle is at } x |\; x \in \mathbb{R}^3\}$ and $\Omega_e = \{\text{operations of the company}\}$ are neither countable nor quantitative. It is difficult to associate meaningful and manipulatable σ-fields with these spaces. However, as will be discussed in Section 2.4, when we have a random variable X defined on Ω, we can generate an "induced" σ-field $\mathscr{F}$ for Ω based on X. □

Remark 2.3. It may be tempting to think that $\mathscr{B}$ and $\sigma(\mathscr{B}\times\mathscr{B}\times\mathscr{B})$ are the σ-fields for the sample spaces in Example 2.8a) and d), respectively. However, precisely speaking, this is not correct because $\mathscr{B}$ and $\sigma(\mathscr{B}\times\mathscr{B}\times\mathscr{B})$ are only the σ-fields for $\mathbb{R}$ and $\mathbb{R}^3$, respectively. Note that $\Omega_a \neq \mathbb{R}$ and $\Omega_d \neq \mathbb{R}^3$ since Ω_a and Ω_d are events in the physical world but $\mathbb{R}$ and $\mathbb{R}^3$ are numbers. □

2.3.3 Probability Measure

Given a sample space Ω and the associated σ-field $\mathscr{F}$, we call the duple $(\Omega,\mathscr{F})$ a *measurable space*. Our goal is to define a probability measure for the events in $\mathscr{F}$. Hence, a *measure* should be a function $\nu : \mathscr{F} \to \mathbb{R}$ which maps subsets of Ω to real numbers.

Definition 2.6. *(Measure)* Let $(\Omega,\mathscr{F})$ be a measurable space. A function $\nu : \mathscr{F} \to \mathbb{R}$ is called a *measure* if it has the following properties:

1. $0 \leqslant \nu(A) \leqslant \infty$ for any $A \in \mathscr{F}$;
2. $\nu(\emptyset) = 0$; and
3. (**Countable additivity**) If $A_i \in \mathscr{F}$ for $i = 1,2,\cdots$, and A_i's are **disjoint**, i.e. $A_i \cap A_j = \emptyset$ for any $i \neq j$, then

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

□

If a measure ν satisfies all the three conditions in Definition 2.6 and $\nu(\Omega) = 1$, then ν is called a *probability measure*. We usually denote a probability measure by $\mathbb{P}$ instead of ν. In this case, $(\Omega,\mathscr{F},\mathbb{P})$ is called a *probability space*. The *support* of $\mathbb{P}$ is defined to be the minimal set A in $\mathscr{F}$ such that $\mathbb{P}(A) = 1$.

Recall that the "measure" l in (2.2) is not well-defined since we only assumed $l([a,b]) = b-a$, i.e., l only assigns values to intervals in $\mathbb{R}$. It takes efforts to extend l to assigning values to arbitrary elements in a σ-field on $\mathbb{R}$. The following construction of a measure on $\mathbb{R}$ that satisfies Definition 2.6 is due to Lebesgue. Some discussions will be followed by Remark 2.4.

Construction of Lebesgue Measure

1) Define the *Lebesgue outer measure* for any set $A \subset \mathbb{R}$ by

$$m^*(A) = \inf \left\{ \sum_{n=1}^{\infty} l(I_n) : A \subseteq \bigcup_{n=1}^{\infty} I_n, \;\; I_n \text{ are intervals} \right\}, \tag{2.8}$$

which is the infimum of all possible covers of A. However, m^* is not really a measure as it is not countably additive (see Remark 2.4ii)).
2) Define the class of Lebesgue-Measurable set $\mathscr{M}$ as follows: The set E is in $\mathscr{M}$ if for every set $A \subseteq \mathbb{R}$, we have

$$m^*(A) = m^*(A \cap E) + m^*(A \cap E^c). \tag{2.9}$$

In this case we say that E is Lebesgue-Measurable, or $E \in \mathscr{M}$.
3) Verify that the class of sets $\mathscr{M}$ is a σ-field.
4) Verify that if we restrict $m^*(\cdot)$ on sets in $\mathscr{M}$, then $m^*(\cdot)$ satisfies the definition of a measure (Definition 2.6).
5) Lebesgue Measure λ is then defined as the outer measure $m^*(\cdot)$ on the σ-field $\mathscr{M}$. The measure space is denoted as $(\mathbb{R}, \mathscr{M}, \lambda)$

Remark 2.4. i) Since $m^*(A)$ is well-defined on any subset $A \subset \mathbb{R}$, m^* is well-defined on the σ-field $2^{\mathbb{R}}$.

ii) **(m^* is not countably additive on $2^{\mathbb{R}}$):** Consider the Vitali set E_5 and $S_n = E_5 + q_n$, $q_n \in \mathbb{Q} \cap [-1,1]$. In Section 2.2.5 we have shown that $[0,1] \subseteq \bigcup_{n=1}^{\infty} S_n \subseteq [-1,2]$. Thus $m^*\left(\bigcup_{n=1}^{\infty} S_n\right) \in [1,3]$. But $\sum_{n=1}^{\infty} m^*(S_n)$ can either be 0 or ∞. Thus, $m^*\left(\bigcup_{n=1}^{\infty} S_n\right) \neq \sum_{n=1}^{\infty} m^*(S_n)$ and m^* is not countably additive.

iii) **(m^* is countably sub-additive):** From (2.8), for every $\varepsilon > 0$, we can find intervals $\{I_n^k\}_{n \geq 1}$ such that $S_k \subseteq \bigcup_{n=1}^{\infty} I_n^k$ and $\sum_{n=1}^{\infty} l(I_n^k) - \frac{\varepsilon}{2^k} \leq m^*(S_k) \leq \sum_{n=1}^{\infty} l(I_n^k)$ for $k = 1, \ldots$. As $\bigcup_{k=1}^{\infty} S_k \subseteq \bigcup_{k=1}^{\infty} \bigcup_{n=1}^{\infty} I_n^k$, we have

$$m^*\left(\bigcup_{k=1}^{\infty} S_k\right) \leq \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} l(I_n^k) \leq \sum_{k=1}^{\infty} \left(m^*(S_k) + \frac{\varepsilon}{2^k}\right) = \sum_{k=1}^{\infty} m^*(S_k) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we have $m^*\left(\bigcup_{n=1}^{\infty} S_k\right) \leq \sum_{k=1}^{\infty} m^*(S_k)$, which is known as the countable sub-additivity property.

iv) Step 2) is the key step in the whole construction of Lebesgue measure, which cleverly restricts m^* on a class of sets, $\mathscr{M}$, which is nice (is a σ-field and contains $\mathscr{B}$; see Theorem 2.2) and useful enough (m^* is countably additive, thus is a measure, on $\mathscr{M}$). □

Theorem 2.2. (Lebesgue and Borel σ-field) *Lebesgue σ-field contains Borel σ-field. That is, $\mathscr{B} \subset \mathscr{M}$.*

Proof. First we show that $\mathscr{B} \subseteq \mathscr{M}$. From the definition in Example 2.14, $\mathscr{B}$ is the smallest σ-algebra containing all open intervals of the form (a,b). Since $\mathscr{M}$ is also a σ-field, it suffices to show $\mathscr{M}$ contains open intervals of the form (a,b). In particular, we need to verify

$$m^*(A) = m^*(A\cap(a,b)) + m^*(A\cap C)\,. \tag{2.10}$$

where $C = (a,b)^c = (-\infty,a]\cup[b,\infty)$. By the definition of outer measure (2.8), for any $\varepsilon > 0$, we can find intervals $\{I_n\}$ such that

$$\text{i) } A \subseteq \bigcup_{n=1}^{\infty} I_n \;\text{ and }\; \text{ii) } \sum_{n=1}^{\infty} l(I_n) < m^*(A) + \varepsilon\,. \tag{2.11}$$

From i) in (2.11), $A\cap(a,b) \subseteq \bigcup_{n=1}^{\infty} I_n\cap(a,b)$ and $A\cap C \subseteq \bigcup_{n=1}^{\infty} I_n\cap C$. As each of $I_n\cap(a,b)$ and $I_n\cap C$ is an interval or a union of intervals, the definition of outer measure implies that

$$m^*(A\cap(a,b)) \le \sum_{n=1}^{\infty} l(I_n\cap(a,b)) \quad\text{and}\quad m^*(A\cap C) \le \sum_{n=1}^{\infty} l(I_n\cap C)\,. \tag{2.12}$$

Combining (2.11)ii), (2.12), and the fact $l(I_n\cap(a,b)) + l(I_n\cap C) = l(I_n)$, we have

$$m^*(A\cap(a,b)) + m^*(A\cap C) \overset{(2.12)}{\le} \sum_{n=1}^{\infty} l(I_n\cap(a,b)) + \sum_{n=1}^{\infty} l(I_n\cap C) = \sum_{n=1}^{\infty} l(I_n) \overset{(2.11)ii)}{\le} m^*(A) + \varepsilon\,.$$

As $\varepsilon > 0$ can be set arbitrarily small, we have $m^*(A) \ge m^*(A\cap(a,b)) + m^*(A\cap C)$. The opposite direction $m^*(A) \le m^*(A\cap(a,b)) + m^*(A\cap C)$ follows from the countably sub-additivity of outer measure; see Remark 2.4iii). Therefore, (2.10) follows. (The intuition behind the countably sub-additivity of outer measure is that, putting together the covers for all A_is must cover $\bigcup_{i=1}^{n} A_i$.)

Next we outline a counter example to show that $\mathscr{B}$ is strictly smaller than $\mathscr{M}$.

i) Define **Cantor function** $c : [0,1] \to [0,1]$ by

$$c(x) = \begin{cases} \frac{1}{2}\sum_{n=1}^{\infty} \frac{a_n}{2^n} & x \in \mathscr{C}\,, \\ c(\sup\{z\in\mathscr{C} : z < x\}) & x \in [0,1]\backslash\mathscr{C} \end{cases}$$

where $\mathscr{C} \triangleq E_4$ is the Cantor set and $x = \sum_{n=1}^{\infty} a_n$ is the ternary expansion of x. Note that c is non-decreasing and is constant outside $\mathscr{C}$.

ii) Define $f : [0,1] \to [0,2]$ by $f(x) = c(x) + x$. Check the following properties:

1) $f(x)$ is continuous and monotonic increasing. Thus it is bijective and maps intervals to intervals, and hence Borel sets to Borel sets.
2) On the removed intervals of middle thirds, i.e., $[0,1]\backslash\mathscr{C}$, c is locally a constant function. Thus, the increase in $f(x)$ is solely due to the "x" part but not the "$c(x)$" part. As a result,

$$\lambda(f([0,1]\backslash\mathscr{C})) = 1\,. \tag{2.13}$$

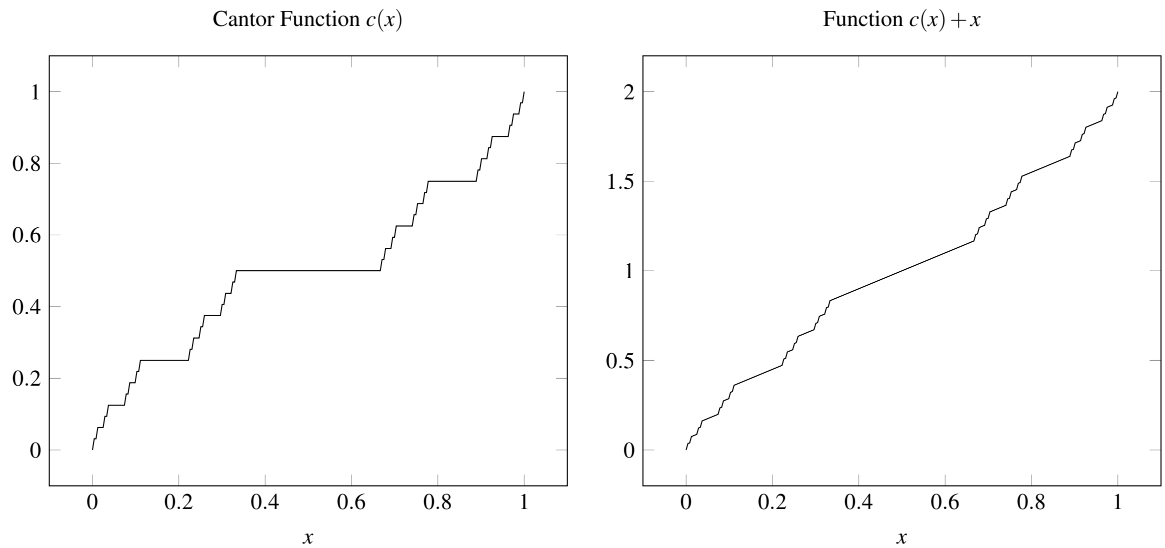


Fig. 2.4 Left: Approximation to Cantor function $c(x)$. Right: Approximation to $f(x) = c(x) + x$.

From (2.13) and $\lambda(f([0,1])) = \lambda([0,2]) = 2$, it follows that

$$\lambda(f(\mathscr{C})) = \lambda(f([0,1])) - \lambda(f([0,1] \backslash \mathscr{C})) = 2 - 1 = 1\,. \qquad (2.14)$$

iii) Modify the construction of Vitali set E_5 to show the existence of some non-measurable set $N \subseteq f(\mathscr{C})$.
iv) Define $E_6 = f^{-1}(N)$.

- Note that $E_6 \subseteq f^{-1}(f(\mathscr{C})) = \mathscr{C}$ so that $\lambda(E_6) \leq \lambda(\mathscr{C}) = 0$. Thus E_6 is Lebesgue measurable (measure equals 0).
- Check that E_6 is not Borel measurable: Otherwise, Property ii)1) implies that $N = f(E_6)$ is Borel measurable, contradicting iii) that N is non-measurable.

Thus, we have arrived at a set N that satisfies $N \notin \mathscr{B}$ but $N \in \mathscr{M}$.

□

Example 2.17. **(Lebesgue Measure and Uniform distribution)**
Formally, *Lebesgue measure* is the measure λ on $(\mathbb{R}, \mathscr{M})$ satisfying (2.8) and (2.9). Since $\mathscr{B} \subset \mathscr{M}$, λ is also defined on $(\mathbb{R}, \mathscr{B})$. In particular, it satisfies $\lambda([a,b]) = b - a$ for every finite interval $[a,b]$, where $-\infty < a < b < \infty$. In this special case, $\lambda(\cdot)$ is also known as length function $l(\cdot)$.

If we restrict λ to the measurable space $([0,1], \mathscr{M}_{[0,1]})$, where $\mathscr{M}_{[0,1]} = \{A \cap [0,1] | A \in \mathscr{M}\}$, then λ is a probability measure. Another name for the probability space $([0,1], \mathscr{M}_{[0,1]}, \lambda)$ is the **uniform distribution on [0,1]**. □

Example 2.18. **(Lebesgue-Stieltjes Measure)**. Let $F : \mathbb{R} \to \mathbb{R}$ be a right-continuous, non-decreasing function. The ***Lebesgue-Stieltjes measure on*** $\mathbb{R}$ ***associated with*** F is the measure on $(\mathbb{R}, \mathscr{M})$, denoted by λ_F, such that for $-\infty < a < b < \infty$,

$$\lambda_F((a,b]) = F(b) - F(a)\,. \tag{2.15}$$

The measure λ_F can be constructed in the same way as the Lebesgue measure λ, but with l in (2.8) replaced by l_F, where $l_F((a,b]) = F(b) - F(a)$.

If F satisfies $F(-\infty) = 0$ and $F(\infty) = 1$, then F is called a **cumulative distribution function**. The Lebesgue-Stieltjes measure $\lambda_F(\mathbb{R}) = \lambda_F((-\infty,\infty)) = F(\infty) - F(-\infty) = 1$ defines a probability measure. In particular, $\lambda_F((a,b]) = F(b) - F(a)$ is the probability of the event $(a,b]$.

The requirement of right-continuity of F arises from the half open interval $(-a,b]$ used in (2.15). In particular, if λ_F has a point mass at x, say $\lambda_F(\{x\}) = c > 0$, then

$$\lim_{n\to\infty} \lambda_F\left(\left(-\infty, x + \frac{1}{n}\right]\right) = \lambda_F((-\infty,x]) = \lim_{n\to\infty} \lambda_F\left(\left(-\infty, x - \frac{1}{n}\right]\right) + c\,,$$

which can be expressed using as (using (2.15))

$$\lim_{n\to\infty} F\left(x + \frac{1}{n}\right) = F(x) = \lim_{n\to\infty} F\left(x - \frac{1}{n}\right) + c\,. \tag{2.16}$$

The first equality in (2.16) yields the definition of right-continuity of F. Indeed, if there is no point mass at x, i.e., $c = 0$, then the second equailty in (2.16) suggests that F is also left-continuous at x. □

Example 2.19. **(Lebesgue Product Measure on** $\mathbb{R}^n$**)**. Let $\{a_i, b_i\}_{i=1}^n$ be real numbers satisfying $a_i < b_i$ for $i = 1, \ldots, n$. Given a rectangle $R = [a_1,b_1] \times [a_2,b_2]$ and a cube $C = [a_1,b_1] \times [a_2,b_2] \times [a_3,b_3]$, we know that the area of R and volume of C are given respectively by $|b_1 - a_1||b_2 - a_2|$ and $|b_1 - a_1||b_2 - a_2||b_3 - a_3|$. Therefore, for a **rectangular block in** $\mathbb{R}^n$, $B = [a_1,b_1] \times [a_2,b_2] \times \cdots \times [a_n,b_n]$, we can define a **generalized length function**

$$l(B) = \prod_{j=1}^{n} |b_j - a_j|\,. \tag{2.17}$$

Using the same construction method, we obtain the Lebesgue measure λ_n on $(\mathbb{R}^n, \mathscr{M}^n)$, where $\mathscr{M}^n$ is a σ-field containing all rectangular blocks of dimension n. In particular, $\mathscr{M}^n$ contains $\mathscr{B}^n$, the Borel product σ-field for $\mathbb{R}^n$ defined in Example 2.15.

Of course, Lebesgue-Stieltjes Product Measure λ_F on $\mathbb{R}^n$ can be defined similarly as in Example 2.18, where $F(\cdot,\cdot,\cdot)$ is a non-decreasing function in each dimension satisfying $F(\infty,\infty,\infty) = 1$ and $F(-\infty,-\infty,-\infty) = 0$. However, in contrast to (2.15) and (2.17), it could be tedious to write down the $\lambda_F(B)$ explicitly. For example, in $\mathbb{R}^2$,

$$\lambda_F([l_1,u_1]\times[l_2,u_2]) = F(u_1,u_2) - F(l_1,u_2) - F(u_1,l_2) + F(l_1,l_2)\,;$$

while in $\mathbb{R}^3$,

$$\begin{aligned}
&\lambda_F([l_1,u_1]\times[l_2,u_2]\times[l_3,u_3]) \\
= &F(u_1,u_2,u_3) - F(l_1,u_2,u_3) - F(u_1,l_2,u_3) - F(u_1,u_2,l_3) \\
&+F(l_1,l_2,u_3) + F(l_1,u_2,l_3) + F(u_1,l_2,l_3) - F(l_1,l_2,l_3)\,.
\end{aligned}$$

□

Example 2.20. In this example we consider probability measures associated with the sample spaces and σ-fields in Example 2.8 and 2.16. Since the sample spaces $\Omega_b = \{\text{The face of the die is } i|\ i = 1,2,3,4,5,6\}$ and $\Omega_c = \{(\omega_1,\omega_2,\cdots,\omega_n) : \omega_i = H,T\}$ are finite, for $\Omega = \Omega_b$ or Ω_c, it is simple to define probability $\mathbb{P}(\{\omega\})$ on each element $\omega \in \Omega$ such that $\sum_{\omega\in\Omega}\mathbb{P}(\{\omega\}) = 1$. Definition 2.6 of probability measure can be straightforwardly verified.

On the other hand, it is difficult to associate probability measures with the infinite spaces $\Omega_a = \{\text{The finger points to } x|\ x \in [0,1]\}$, $\Omega_d = \{\text{The particle is at } x|\ x \in \mathbb{R}^3\}$ and $\Omega_e = \{\text{operations of the company}\}$, since the σ-fields are difficult to define. However, as will be discussed in Section 2.4, when we have a random variable X defined on Ω, it is sufficient in practice to consider the σ-field and probability measure induced by X. □

2.4 Random Variables

In practice, we frequently use real-valued variables to quantify random events in the real world. For example, we may use $0,1$ to represent head, tail of a coin toss; we use stock price or other indexes to summarize the performance or the value of a company. These variables are known as **random variables**. To be precise, a **random variable** is a function that maps from the sample space Ω to the real line $\mathbb{R}$.

Definition 2.7. (Random Variable) On a measurable space $(\Omega,\mathscr{F})$, a function $X : \Omega \to \mathbb{R}$ is said to be an **$\mathscr{F}$-measurable random variable** if the inverse image $X^{-1}(B) \in \mathscr{F}$ for every Borel set $B \in \mathscr{B}$.

In Definition 2.7, the requirement $X^{-1}(B) \in \mathscr{F}$ means that for any *reasonable* values that X can take (i.e., $X \in B$), the corresponding event (i.e., $\{\omega : X(\omega) \in B\} \subseteq \Omega$), is $\mathscr{F}$ measurable. This ensures that we can assign probability to any *reasonable* possibility of X once a probability measure $\mathbb{P}$ is defined on $\mathscr{F}$. For example, we may assign probability to the event $X \in (0,1)$, but we have no hope to assign probability to $X \in E_5$, the Vitali set.

Remark 2.5. For notational simplicity, we use $\{X \in B\}$ to represent the set $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$. If $B = (-\infty,a]$, then $X^{-1}(B)$ is written as $\{X \le a\}$. We write $X^{-1}(\mathscr{B}) \subseteq \mathscr{F}$ to denote $X^{-1}(B) \in \mathscr{F}$ for every Borel set $B \in \mathscr{B}$, i.e., X is $\mathscr{F}$ measurable.

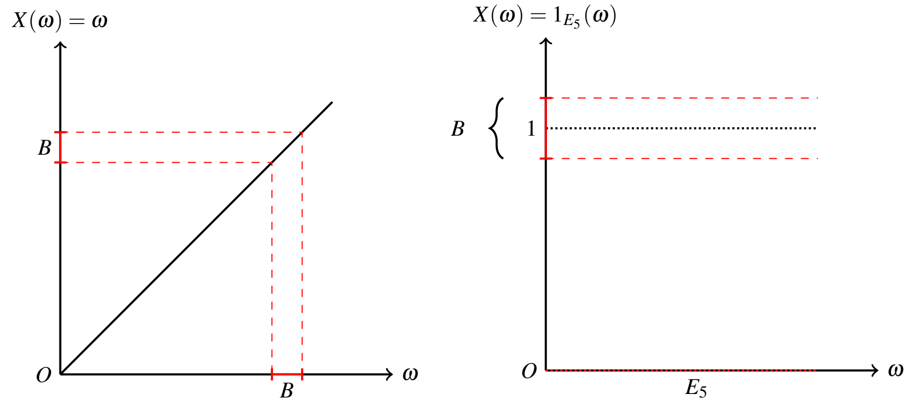


Fig. 2.5 *a*): $X^{-1}(B) \in \mathscr{B}$; X is $\mathscr{B}$ measurable. *b*): $X^{-1}(B) = E_5 \notin \mathscr{B}$; X is not $\mathscr{B}$ measurable.

Example 2.21. Let $\Omega = \{1,2,\ldots,6\}$, $\mathscr{F}_1 = 2^{\Omega}$, $\mathscr{F}_2 = \{\emptyset,\{1,2,3\},\{4,5,6\},\Omega\}$, X_1 and X_2 be functions from Ω to $\mathbb{R}$ such that $X_1(\omega) = 1_{\{\omega\geq 4\}}$ and $X_2(\omega) = \omega$. Note that

$$X_1^{-1}(B) = \begin{cases} \Omega & \text{if } \{0,1\} \subseteq B \\ \{4,5,6\} & \text{if } 0 \notin B \text{ and } 1 \in B \\ \{1,2,3\} & \text{if } 0 \in B \text{ and } 1 \notin B \\ \emptyset & \text{if } 0 \notin B \text{ and } 1 \notin B\,. \end{cases}$$

As the four sets on the right are measurable by $\mathscr{F}_1$ and $\mathscr{F}_2$, X_1 is a r.v. on both $(\Omega,\mathscr{F}_1)$ and $(\Omega,\mathscr{F}_2)$. However, X_2 is a r.v. on $(\Omega,\mathscr{F}_1)$ but not $(\Omega,\mathscr{F}_2)$, since $X_2^{-1}(\{4\}) = \{4\} \notin \mathscr{F}_2$. □

To check whether a function $X : \Omega \to \mathbb{R}$ is measurable, there is no need to exhaustively verify $X^{-1}(B) \in \mathscr{F}$ for all $B \in \mathscr{B}$:

Theorem 2.3. *Given a measurable space $(\Omega,\mathscr{F})$, a function $X : \Omega \to \mathbb{R}$ is $\mathscr{F}$-measurable if $X^{-1}((a,\infty))$ is measurable for all $a \in \mathbb{R}$.*

Proof. The proof is divided into two parts. 1) $X^{-1}((-\infty,a)) \in \mathscr{F}$ and $X^{-1}((a,b)) \in \mathscr{F}$ for all real numbers $a < b$. 2) Show that $X^{-1}(\mathscr{B}) \subseteq \mathscr{F}$.

1) First we show that for any function $X : \Omega \to \mathbb{R}$ and any subsets A, $A_n \subseteq \mathbb{R}$, $n = 1,2,\ldots$

$$\text{i) } X^{-1}(A^c) = \left(X^{-1}(A)\right)^c, \qquad \text{ii) } X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) = \bigcup_{n=1}^{\infty}\left(X^{-1}(A_n)\right). \tag{2.18}$$

Using the following standard arguments, we establish i):

$$\omega \in X^{-1}(A^c) \Leftrightarrow X(\omega) \in A^c \Leftrightarrow X(\omega) \notin A \Leftrightarrow \omega \notin X^{-1}(A) \Leftrightarrow \omega \in \left(X^{-1}(A)\right)^c. \tag{2.19}$$

Similar arguments give ii). Putting $A = (a, \infty)$ in i) we get

$$X^{-1}((-\infty, a]) = \left(X^{-1}((a, \infty))\right)^c \in \mathscr{F}$$

since $X^{-1}((a, \infty)) \in \mathscr{F}$ and $\mathscr{F}$ is a σ-field. Next, combining i) and ii) gives

$$\begin{aligned} X^{-1}((a,b]) &= X^{-1}\left([(-\infty,a] \cup (b,\infty)]^c\right) \overset{i)}{=} \left[X^{-1}\left((-\infty,a] \cup (b,\infty)\right)\right]^c \\ &\overset{ii)}{=} \left[X^{-1}((-\infty,a]) \cup X^{-1}((b,\infty))\right]^c \in \mathscr{F} \end{aligned} \tag{2.20}$$

since both $X^{-1}((-\infty, a])$ and $X^{-1}((b, \infty))$ are in $\mathscr{F}$. Finally, using similar arguments in (2.19) we have

$$(a,b) = \bigcup_{n=1}^{\infty} A_n\,, \text{ where } \qquad A_n = \left(a, b - \frac{1}{n}\right].$$

Since (2.20) holds for all a, b, we have $X^{-1}(A_n) = X^{-1}\left(\left(a, b - \frac{1}{n}\right]\right) \in \mathscr{F}$. Therefore, $X^{-1}((a,b)) = X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) = \bigcup_{n=1}^{\infty}\left(X^{-1}(A_n)\right) \in \mathscr{F}$ by ii). Similar arguments also yield $X^{-1}((-\infty, a)) \in \mathscr{F}$.

2) Define $\mathscr{G} = \{B \in \mathscr{B} : X^{-1}(B) \in \mathscr{F}\} \subseteq \mathscr{B}$ to be subsets of $\mathscr{B}$ such that $X^{-1}(B) \in \mathscr{F}$. Note that $X^{-1}(\mathscr{B}) \subseteq \mathscr{F}$ is equivalent to $\mathscr{G} = \mathscr{B}$. Therefore, to show 2), we only need to show that $\mathscr{G} = \mathscr{B}$. As $\mathscr{G} \subseteq \mathscr{B}$ by definition, it remains to show $\mathscr{B} \subseteq \mathscr{G}$. From 1), we have that $\mathscr{A} \subseteq \mathscr{G}$ where $\mathscr{A}$ is the collection of all finite open intervals in $\mathbb{R}$. Recall from the definition of Borel σ-field in Example 2.14 we have

$$\mathscr{B} = \sigma(\mathscr{A}) \subseteq \sigma(\mathscr{G}). \tag{2.21}$$

On the other hand, using (2.18), we can check that $\mathscr{G}$ is a σ-field. As $\sigma(\mathscr{G})$ is the smallest σ-field containing $\mathscr{G}$, we must have $\sigma(\mathscr{G}) = \mathscr{G}$. Therefore, (2.21) reduces to $\mathscr{B} \subseteq \mathscr{G}$, which completes the proof. □

Using Theorem 2.3, we can see that most of the commonly encountered functions are measurable.

Example 2.22. **(Measurable functions)**

1. Constant function $X(\omega) \equiv c$ is measurable, since

$$X^{-1}((a,\infty)) = \begin{cases} \Omega & \text{if } a < c \\ \emptyset & \text{otherwise} \end{cases},$$

and $\emptyset, \Omega \in \mathscr{F}$.

2. For a measurable set $A \in \mathscr{F}$, the indicator function $X(\omega) = 1_{\{\omega \in A\}}$ is measurable, since

$$X^{-1}((a,\infty)) = \begin{cases} \Omega & \text{if } a < 0 \\ A & \text{if } 0 \le a < 1 \\ \emptyset & \text{if } a \ge 1\,, \end{cases}$$

and $\emptyset, A, \Omega \in \mathscr{F}$. However, if $A = E_5$ in Section 2.2.5, then X is not a measurable random variable.

3. If $(\Omega, \mathscr{F}) = (\mathbb{R}, \mathscr{B})$, then any continuous function $X : \Omega \to \mathbb{R}$ is measurable. We divide the proof in three parts:

   i) If $\omega \in X^{-1}((a,\infty))$, then some interval $(\omega - \delta, \omega + \delta) \subseteq X^{-1}((a,\infty))$:
   For $\omega \in X^{-1}((a,\infty))$, we have $X(\omega) \in (a,\infty)$, and there exists some ε such that $(X(\omega) - \varepsilon, X(\omega) + \varepsilon) \subset (a,\infty)$. Continuity of X implies that there exists some $\delta > 0$ such that

   $$X(\omega - \delta, \omega + \delta) \subseteq (X(\omega) - \varepsilon, X(\omega) + \varepsilon) \subset (a,\infty)\,.$$

   In other words, $(\omega - \delta, \omega + \delta) \subset X^{-1}((a,\infty))$.

   ii) $X^{-1}((a,\infty))$ can be written as a *countable* union of intervals $\bigcup_{n=1}^{\infty}(l_n, u_n)$:
   For each $\omega \in X^{-1}((a,\infty))$, we find the largest open interval I_ω containing ω such that $I_\omega \subseteq X^{-1}((a,\infty))$. We claim that, if $\omega_1, \omega_2 \in X^{-1}((a,\infty))$ and $\omega_1 \ne \omega_2$, then either I_{ω_1} and I_{ω_2} are disjoint or identical. To see this, suppose on the contrary that $I_{\omega_1} \ne I_{\omega_2}$ have a non-empty intersection. Then their union $I_{\omega_1} \cup I_{\omega_2}$ is an open interval containing both ω_1 and ω_2, which contradicts to the maximality of I_{ω_1} and I_{ω_2}. Thus the claim is verified.
   As $I_\omega \subseteq X^{-1}((a,\infty))$, we can write $X^{-1}((a,\infty)) = \bigcup_{\omega \in X^{-1}((a,\infty))} I_\omega$. Note that each interval I_ω contains a rational number, say $r(\omega)$. Since rational numbers are countable and I_ωs are disjoint, the union (not necessarily countable) $\bigcup_{\omega \in X^{-1}((a,\infty))} I_\omega$ can be regarded as a *countable* union over intervals I_ω labeled by distinct rationals $r(\omega)$, and the proof of ii) is complete.

   iii) $X^{-1}((a,\infty)) \in \mathscr{F}$ for all $a \in \mathbb{R}$, thus X is measurable:
   For all $a \in \mathbb{R}$, from ii) we have that $X^{-1}((a,\infty)) = \bigcup_{n=1}^{\infty}(l_n, u_n)$. Since all intervals $(l_n, u_n) \in \mathscr{B}$ by the definition of Borel σ-field $\mathscr{B}$, and $\mathscr{F} = \mathscr{B}$, we have $X^{-1}((a,\infty)) = \bigcup_{n=1}^{\infty}(l_n, u_n) \in \mathscr{B} = \mathscr{F}$. □

In practice, we are often more interested in the values of the random variable than the corresponding events in the physical world. For example, we may be more interested in the stock price than the operational details of a company. Therefore, we shall only require that $X \in B$ is measurable for Borel set $B \in \mathscr{B}$. Nevertheless, to have a complete set-up, we can define a σ-field on the original sample space based on the requirement that $\{X \in B\} = \{\omega : X(\omega) \in B\}$ is measurable for all $B \in \mathscr{B}$.

Definition 2.8. (σ-field generated by a random variable.) The σ-field generated by a random variable $X : \Omega \to \mathbb{R}$ is defined as

$$\sigma(X) = \sigma\{X^{-1}(B) | B \in \mathscr{B}\}\,,$$

which is the smallest σ-field such that X is Borel measurable. Equivalently, using Theorem 2.3, we have

$$\sigma(X) = \sigma\{X^{-1}((a,\infty))|a \in \mathbb{R}\}.$$

Example 2.23. Consider $\Omega = \{1,2,\ldots,6\}$, $X_1(\omega) = 1_{\{\omega \geq 4\}}$, $X_2(\omega) = \omega$ and $X_3(\omega) = 1_{\{\omega \text{ is an even integer}\}}$. The σ-field generated by X_1, X_2 and X_3 are respectively $\mathscr{F}_1 = \{\emptyset, \{1,2,3\}, \{4,5,6\}, \Omega\}$, $\mathscr{F}_2 = 2^\Omega$ and $\mathscr{F}_3 = \{\emptyset, \{1,3,5\}, \{2,4,6\}, \Omega\}$.

On the other hand, if $\Omega = \mathbb{R}$, $X_1(\omega) = 1_{\{\omega \geq 4\}}$, $X_2(\omega) = \omega$ and $X_3(\omega) = 1_{\{\omega \text{ is an even integer}\}}$. The σ-field generated by X_1, X_2 and X_3 are respectively $\mathscr{F}_1 = \{\emptyset, (-\infty,4), [4,\infty), \mathbb{R}\}$, $\mathscr{F}_2 = \mathscr{B}$ and $\mathscr{F}_3 = \{\emptyset, \mathbb{R}\backslash\{2k : k \in \mathbb{Z}\}, \{2k : k \in \mathbb{Z}\}, \mathbb{R}\}$. □

Example 2.24. Consider $\Omega = \mathbb{R}$, $\mathscr{F} = \mathscr{B}$ and $X(\omega) = \omega^2$. Observe that $X^{-1}((a,\infty)) = (-\infty, -\sqrt{a}) \cup (\sqrt{a}, \infty)$ if $a > 0$ and $X^{-1}((a,\infty)) = \mathbb{R}$ if $a \leq 0$. Thus,

$$\sigma(X) = \sigma\{(-\infty,-c) \cup (c,\infty)|c \geq 0\} = \sigma\{[-c,c]|c \geq 0\}.$$

Note that $\sigma(X)$ is strictly smaller than $\mathscr{B}$ since the sets in $\sigma(X)$ are symmetric around 0.

Following Example 2.8, we can say that $(\mathbb{R}, \mathscr{B})$ is a measurable space **induced** from the original measurable space $(\Omega, \sigma(X))$. To complete the definition, we have to assign probability measures to the two measurable spaces.

Definition 2.9. (Induced Probability Measure) Every random variable $X : \Omega \to \mathbb{R}$ induces a probability measure $\mathbb{P}_X$ on $(\mathbb{R}, \mathscr{B})$ defined by

$$\mathbb{P}_X(B) = \mathbb{P}(\{\omega : X(\omega) \in B\}) \tag{2.22}$$

for each Borel set B of $\mathbb{R}$. The function $F_X : \mathbb{R} \to [0,1]$ defined by $F_X(x) = \mathbb{P}(X \leq x)$ is called the cumulative distribution function (c.d.f.) of X.

Remark 2.6. Although $\mathbb{P}(\{\omega : X(\omega) \in B\})$ and $\mathbb{P}_X(B)$ always take the same value, the former $\mathbb{P}$ measures events from the original sample space Ω and the latter $\mathbb{P}_X$ measures events from the induced sample space $\mathbb{R}$. Also, $\mathbb{P}(\{\omega : X(\omega) \in B\})$ is sometimes abbreviated as $\mathbb{P}(X \in B)$. Moreover, as will be seen in Example 2.25, we usually first design probability measures $\mathbb{P}_X$ on $(\mathbb{R}, \mathscr{B})$, and then define $\mathbb{P}$ on $(\Omega, \sigma(X))$ by (2.22) for completeness. □

Example 2.25. Some examples of random variables, generated σ-fields and induced probability measures in Example 2.8 are as follows:

a) For $\Omega_a = \{\text{The finger points to } x|\ x \in [0,1]\}$. The r.v. "*the nearest integer*" is given by $X_1 : \Omega_a \to \{0,1\}$ with $X_1(\omega_x) = [x]$. The r.v. "*the value shown on the ruler*" is given by $X_2 : \Omega_a \to [0,1]$ with $X_2(\omega_x) = x$. Since X_1 is equal to 0 or 1, the sample space induced by X_1 is $\Omega_X = \{0,1\}$, which is discrete. Thus σ-algebra and probability measure can be defined easily. On the other hand,

X_2 takes values in $\mathbb{R}$. We can define an induced measure $\mathbb{P}_X$ on $(\mathbb{R}, \mathscr{B})$ using Lebesgue measure. Using X_1 and X_2, we can then define σ-fields back on Ω by $\sigma(X_1)$ and $\sigma(X_2)$. Also, we can define probability measures on $(\Omega_a, \sigma(X_i))$ by $\mathbb{P}_i(\{\omega : X(\omega) \in B\}) = \mathbb{P}_{X_i}(B)$ for $i = 1, 2$, respectively. Although we cannot quantify σ_{X_i} and $\mathbb{P}_i$ explicitly, we do not care in practice since we work on the induced space $(\mathbb{R}, \mathscr{B}, \mathbb{P}_{X_i})$.

b) When a die is thrown once, $\Omega_b = \{\text{The face of the die is } i |\ i = 1,2,3,4,5,6\}$. If the elements in Ω_b are denoted by $\omega_i = \{\text{The face of the die is } i\}$, then the r.v. "*It is Big*" is given by $X_1 : \Omega_b \to \{0,1\}$ with $X_1(\omega_i) = 1_{\{i \geq 4\}}$. The r.v. "*value on the die*" is $X_2 : \Omega_b \to \{1, 2, \ldots, 6\}$ with $X_2(\omega_i) = i$. Define two sets $A_1 = \bigcup_{i=1}^{3} \{\omega_i\}$ and $A_2 = \bigcup_{i=4}^{6} \{\omega_i\}$. We have

$$\sigma(X_1) = \sigma(\{A_1, A_2\}), \quad \sigma(X_2) = \sigma(\{\{\omega_i\}, i = 1, \ldots, 6\}),$$
$$\mathbb{P}_{X_1}(B) = \frac{1}{2} 1_{\{1 \in B\}} + \frac{1}{2} 1_{\{0 \in B\}}, \quad \mathbb{P}_{X_2}(B) = \sum_{i=1}^{6} \frac{1}{6} 1_{\{i \in B\}},$$

where $B \in \mathscr{B}$, if the die is fair.

c) For n tosses of a coin, $\Omega_c = \{(\omega_1, \omega_2, \cdots, \omega_n) :\ \omega_i = H, T\}$. The r.v. "*number of heads*" is a function $X : \Omega_c \to \{0, 1, \ldots, n\}$, $X(\omega) = \sum_{i=1}^{n} 1_{\{\omega_i = H\}}$ for $\omega = (\omega_1, \ldots, \omega_n) \in \Omega_c$. Since X takes finite number of values, $\sigma(X)$ can be expressed explicitly by direct enumeration, albeit tedious. If the coin is fair, then the induced probability measure may be given by $\mathbb{P}_X(B) = \frac{1}{2^n} \sum_{k=0}^{n} C_k^n 1_{\{k \in B\}}$.

d) For the position of a particle, $\Omega_d = \{\text{The particle is at } x |\ x \in \mathbb{R}^3\}$, we may define the r.v. "*the coordinates of a particle*" by $X : \Omega_d \to \mathbb{R}^3$. Clearly, X is an identity function. However, we should think in a way that Ω_d describing the physical situation that the particle is lying somewhere, and X is how we use our coordinate system to state its location. To be precise, X is called a random vector (rather than variable) as it takes value in $\mathbb{R}^3$ instead of $\mathbb{R}$.

e) If Ω_e is an abstract space describing the operations of the company, the r.v. "*the stock price of the company*" is given by $X : \Omega_e \to \mathbb{R}^+$. It is impossible to write down a formula for X. However, in practice we are only interested in the value in that X takes. Thus, it is not important to be explicit about how X maps from Ω_e to $\{0, 1, \ldots, n\}$. We only need to focus on the induced probability space $(\mathbb{R}, \mathscr{B}, \lambda_F)$ if we have an idea about the c.d.f. F for X.

$\square$

2.5 Lebesgue Integral

A measure describes the length of a set. An integral describes the area under a curve over a set, which is a generalization of measure. In probability theory, measure is related to probability and integral is related to expectation. In this section, we discuss

Original Probability Space $(\Omega,\mathscr{F},\mathbb{P})$			Random Variable: X	Induced Probability Space $(\Omega_X,\mathscr{F}_X,\mathbb{P}_X)$		
Ω	$\mathscr{F}$	$\mathbb{P}$	X	Ω_X	$\mathscr{F}_X$	$\mathbb{P}_X$
$\Omega_a=\{\text{The finger points to } x \mid x\in[0,1]\}$ $\omega_x=\{\text{finger points to } x\}$	$\sigma(X_1)$	$\mathbb{P}(\{X=0\})=\mathbb{P}(\{X=1\})=\frac{1}{2}$	the Nearest Integer $X_1(\omega_x)=[x]$	$\{0,1\}$	$2^{\{0,1\}}$	$\mathbb{P}_{X_1}(\{0\})=\mathbb{P}_{X_1}(\{1\})=\frac{1}{2}$
	$\sigma(X_2)$	$\mathbb{P}(X\in[0,x])=x$	Value on ruler $X_2(\omega_x)=x$	$\mathbb{R}$	$\mathscr{B}_{[0,1]}$	$\mathbb{P}_{X_2}([0,x])=x$
$\Omega_b=\{\text{The face of die is } i \mid i=1,\dots,6\}$ $\omega_i=\{\text{The face of die is } i\}$	$\sigma(X_1)$ $=\sigma(\{\bigcup_{i=1}^3\{\omega_i\},\bigcup_{i=4}^6\{\omega_i\}\})$	$\mathbb{P}(\bigcup_{i=1}^3\{\omega_i\})=\mathbb{P}(\bigcup_{i=4}^6\{\omega_i\})=\frac{1}{2}$	Big $X_1(\omega_i)=1_{\{i\geq 4\}}$	$\{0,1\}$	$2^{\{0,1\}}$	$\mathbb{P}_{X_1}(\{0\})=\mathbb{P}_{X_1}(\{1\})=\frac{1}{2}$
	$\sigma(X_2)=2^{\Omega_b}$ $=\sigma(\{\{\omega_i\},i=1,\dots,6\})$	$\mathbb{P}(\omega_i)=\frac{1}{6}$	Value on the die $X_2(\omega_i)=i$	$\{1,2,\dots,6\}$	$2^{\{1,2,\dots,6\}}$	$\mathbb{P}_{X_2}(\{i\})=\frac{1}{6}$
$\Omega_c=\{(\omega_1,\omega_2,\cdots,\omega_n):\omega_i=H \text{ or } T\}$ $\omega=(\omega_1,\omega_2,\cdots,\omega_n),\omega_i=H \text{ or } T$	$\sigma(X)$	$\mathbb{P}(\omega)=\frac{1}{2^n}$	Number of heads $X(\omega)=\sum_{i=1}^n 1_{\{\omega_i=H\}}$	$\{0,1,2,\dots,n\}$	$2^{\{0,1,2,\dots,n\}}$	$\mathbb{P}_X(\{k\})=\frac{1}{2^n}C_k^n$
$\Omega_d=\{\text{The particle is at } x \mid x\in\mathbb{R}^3\}$ $\omega_x=\{\text{The particle is at } x\}$	$\sigma(X)$	$\mathbb{P}(\{X\in B\})=\mathbb{P}_X(B)=\lambda_F(B)$ $B\subseteq\mathbb{R}^3,B\in\mathscr{B}^3$	Coordinate of a particle $X(\omega_x)=x$	$\mathbb{R}^3$	$\mathscr{B}^3$	Lebesgue-Stieltjes measure λ_F on $(\mathbb{R}^3,\mathscr{M}^3)$ F is some joint cdf
$\Omega_e=\{\text{operations of the company}\}$ ω is hard to write down	$\sigma(X)$	$\mathbb{P}(\{X\in B\})=\mathbb{P}_X(B)=\lambda_F(B)$ $B\subseteq\mathbb{R},B\in\mathscr{B}$	Stock price of company $X:\Omega_e\to\mathbb{R}^+$ No explicit formula for X	$\mathbb{R}$	$\mathscr{B}$	Lebesgue-Stieltjes measure λ_F on $(\mathbb{R},\mathscr{M})$ F is some cdf

Table 2.1 Summary of examples of probability spaces.

Fig. 2.6 Measure

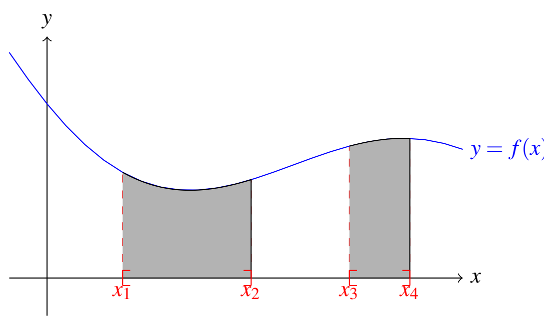


Fig. 2.7 Integral

how the Lebesgue measure can be used to define integrals that generalize Riemann integrals. First we recall the definition of Riemann integrals.

Definition 2.10. (Step Function) A function $f:[a,b]\to\mathbb{R}$ is a step function if $f(x)=a_i$ if $x_{i-1}\leq x<x_i$ where $a=x_0<x_1<\cdots<x_n=b$.

The Riemann integrals use step functions to approximate the area under a curve. For a step function, the Riemann integral is given by $\int_a^b f(x)dx=\sum_{i=1}^n a_i\Delta x_i$ where $\Delta x_i=x_i-x_{i-1}$ is the length of the subinterval $[x_{i-1},x_i]$. For an arbitrary function f, the Riemann integral of f is defined as follows.

Definition 2.11. (Riemann Integral) A finite set $P=\{a_0,a_1,\dots,a_n\}$ satisfying $a=a_0<a_1<a_2<\cdots<a_n=b$ is said to be a **partition** of $[a,b]$. Let $\Delta a_i=a_i-a_{i-1}$,

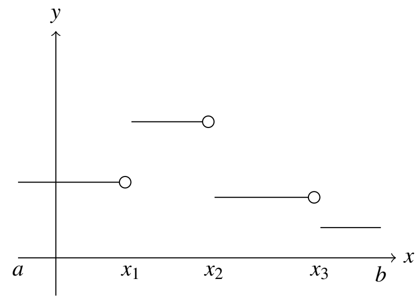


Fig. 2.8 A step function

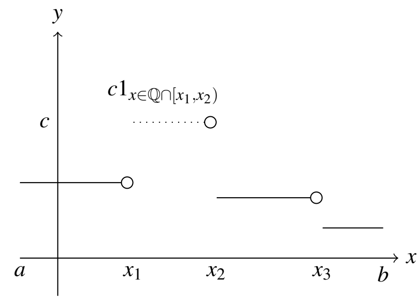


Fig. 2.9 A simple function that is not a step function

$$M_i = \sup_{a_{i-1}\le x\le a_i} f(x) \quad \text{and} \quad m_i = \inf_{a_{i-1}\le x\le a_i} f(x)\,.$$

Then the *upper* and *lower Riemann sums* corresponding to the partition P are defined by

$$U_R(P,f) = \sum_{i=1}^{n} M_i \Delta a_i, \quad \text{and} \quad L_R(P,f) = \sum_{i=1}^{n} m_i \Delta a_i\,.$$

If the supremum of $L_R(P,f)$ and the infimum of $U_R(P,f)$ (taking over all possible partitions P) are equal, then the common value is called the **Riemann Integral** of the function f, denoted by $\int_a^b f(x)\,dx$. □

However, there exist some functions that are not Riemann integrable.

Example 2.26. **(Non-Riemann Integrable function)** Let $f(x) = 1_{\{x\in\mathbb{Q}_{[0,1]}\}}$, where $\mathbb{Q}_{[0,1]}$ is the set of rational numbers in $[0,1]$. Note that any interval (a_{i-1},a_i) contains both rational and irrational numbers. It follows that $M_i = 1$ and $m_i = 0$ for any i and any partition P (see Definition 2.11). Thus, $\inf_P U_R(P,f) = 1 \neq 0 = \sup_P L_R(P,f)$ and Riemann Integral of f does not exist. □

The problem of Riemann integration is that it decomposes the integration domain (x-axis) into small parts, but the function (y-axis) may vary a lot as x changes. Lebesgue integral looks at integration in the opposite way: it decomposes the integration range (y-axis) into small parts.

Definition 2.12. (Simple function) A non-negative function $f : [a,b] \to [0,\infty)$ is a simple function if the range of f is a finite set of distinct non-negative real numbers $\{a_1,\ldots,a_n\}$. In this case, $i = 1,\ldots,n$,

$$f(x) = \sum_{i=1}^{n} a_i 1_{A_i}(x)\,,$$

where $1_{A_i}(x)$ is the indicator function of A_i, A_is are pairwise disjoint and $\bigcup_{i=1}^n A_i = [a,b]$. Sometimes we denote $A_i = f^{-1}(\{a_i\}) = \{x : f(x) = a_i\}$. □

While Riemann integration approximates the area of a function by step functions, Lebesgue integration performs the approximation by simple functions.

Definition 2.13. (Lebesgue Integral) Consider a partition $P = \{a_0, a_1, \ldots, a_n\}$ of the *range* of f, where $-\infty < a_0 < a_1 < a_2 < \cdots < a_n < \infty$. Let $A_i = [a_{i-1}, a_i)$. Then the partition P gives the *upper* and *lower Lebesgue sums*

$$U_\lambda(P, f) = \sum_{i=1}^{n} a_i \lambda(f^{-1}(A_i)), \text{ and } L_\lambda(P, f) = \sum_{i=1}^{n} a_{i-1} \lambda(f^{-1}(A_i)).$$

If the supremum of $L_\lambda(P, f)$ and the infimum of $U_\lambda(P, f)$ (taking over all possible partitions P) are equal, then the common value is called the **Lebesgue Integral** of the function f, denoted by $\int f \, d\lambda$.

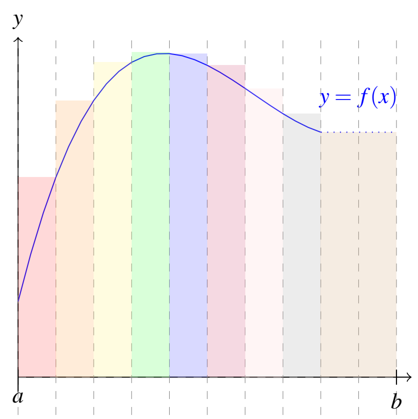


Fig. 2.10 Riemann integral: upper sum

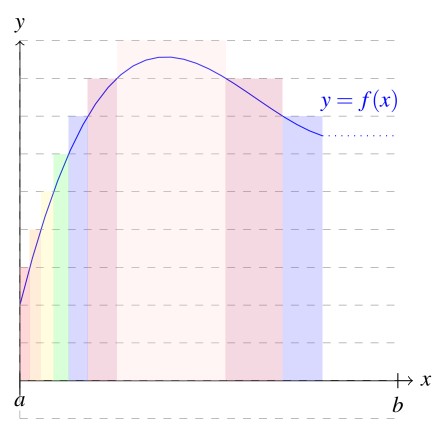


Fig. 2.11 Lebesgue integral: upper sum

Remark 2.7. When the partition P becomes finer, the a_i and a_{i-1} in Lebesgue integral converge to each other while M_i and m_i in Riemann Integral may not. Thus Lebesgue integration is a better device in defining an integral. In fact, it can be shown that if f is Riemann integrable, then f is Lebesgue Integral and the two integrals are the same (Exercise 2.25). The opposite is not true, as the following example shows. □

Example 2.27. Recall that the function $f(x) = 1_{\{x \in \mathbb{Q}_{[0,1]}\}}$ in Example 2.26 is not Riemann integrable. From the definition of Lebesgue integral,

$$\int f \, d\lambda = 1 \times \lambda(\mathbb{Q}_{[0,1]}) + 0 \times \lambda(\mathbb{R} \backslash \mathbb{Q}_{[0,1]}) = 0,$$

since $\lambda(\mathbb{Q}_{[0,1]}) = 0$. Thus f is Lebesgue integrable. □

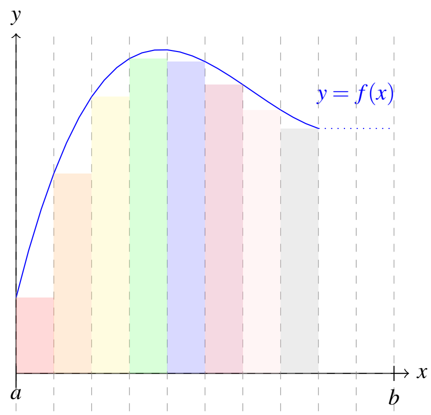


Fig. 2.12 Riemann integral: lower sum

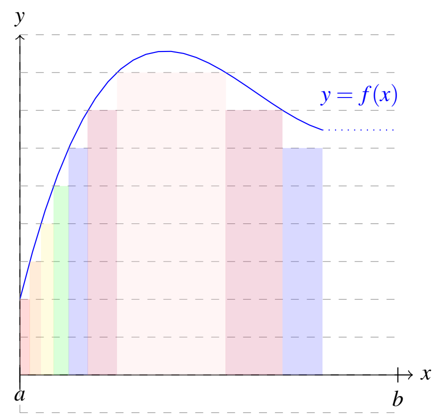


Fig. 2.13 Lebesgue integral: lower sum

Remark 2.8. A real function $f : [a,b] \to \mathbb{R}$ is said to be **Lebesgue measurable** if

$$f^{-1}(B) \in \mathscr{M} \text{ for each set } B \subset \mathscr{M}\,. \tag{2.23}$$

From Definition 2.13, to define the Lebesgue integral $\int f\,d\lambda$, f has to be a Lebesgue measurable function so that $\lambda(f^{-1}(A_i))$ exists for the interval A_i. For example, for the non-measurable Vitali set E_5 defined in Section 2.2.5, $f(x) = 1_{\{x \in E_5\}}$ is not Lebesgue measurable. Thus, $f(x) = 1_{\{x \in E_5\}}$ is not Lebesgue integrable. □

2.6 Lebesgue integration and Expectation

In the previous section we assumed that the integrands in the Lebesgue integral are Lebesgue measurable real functions (functions that satisfy (2.23)) . Although a r.v. $X : \Omega \to \mathbb{R}$, defined on $(\Omega, \mathscr{F}, \mathbb{P})$, is not mapping from $[a,b]$ to $\mathbb{R}$, the property that $X^{-1}(B) \in \mathscr{F}$ for every $B \in \mathscr{B}$ is analogous to (2.23). Thus, Lebesgue Integral can be extended to define the expected value of a random variable.

Consider first a discrete random variable X taking values from a countable set $\{x_1, x_2, \ldots\}$. The **integral** of X under probability measure $\mathbb{P}$ is defined as

$$\int_\Omega X\,d\mathbb{P} = \sum_{i=1}^{\infty} x_i \mathbb{P}(\{X = x_i\})$$

where $\mathbb{P}$ is the probability measure on Ω. Note that we partition Ω into disjoint subsets (i.e. $\{X = x_i\} = \{\omega : X(\omega) = x_i\}$, $i = 1, \ldots, \infty$) and then on each subset we multiply the value of the function by the size of that subset (i.e. the product $x_i \mathbb{P}(\{X = x_i\})$). The integral is then the sum of all such products.

For an arbitrary random variable X, we can approximate X by two discrete random variables $\overline{X}_P(\omega) = \sum_{i=1}^n x_i 1_{\{\omega \in A_i\}}$ and $\underline{X}_P(\omega) = \sum_{i=1}^n x_{i-1} 1_{\{\omega \in A_i\}}$, where $A_i = X^{-1}([x_{i-1}, x_i))$ and $P = \{x_0, x_1, \ldots, x_n\}$ is a partition on the range of X. The **integral** of X, denoted by

$$\int_\Omega X(\omega)\mathbb{P}(d\omega) \quad \text{or simply} \quad \int_\Omega X d\mathbb{P},$$

is then obtained by the same way in Definition 2.13. To be specific, let

$$U_{\mathbb{P}}(P,X) = \sum_{i=1}^n x_i \mathbb{P}(A_i) \;\text{ and }\; L_{\mathbb{P}}(P,X) = \sum_{i=1}^n x_{i-1}\mathbb{P}(A_i)$$

be the upper and lower $\mathbb{P}$-sums of X. Then $\int_\Omega X d\mathbb{P}$ is defined as $\sup_P L_{\mathbb{P}}(P,X)$ or $\inf_P U_{\mathbb{P}}(P,X)$, if the latter two quantities exist and are equal.

In particular, for any $F \in \mathscr{F}$, the integral of X over F, is defined by

$$\int_F X d\mathbb{P} = \int_\Omega X 1_F d\mathbb{P} = \int_\Omega X(\omega) 1_{\{\omega \in F\}} \mathbb{P}(d\omega).$$

We have defined the **integral** of a r.v. with respect to a probability measure $\mathbb{P}$. If the integral is finite, we call it **expectation of** X **under** $\mathbb{P}$.

Definition 2.14. (Random Variables in $\mathscr{L}^1$ and Expectation) A random variable $X : \Omega \to \mathbb{R}$ is said to be **integrable** (or in $\mathscr{L}^1$) if

$$\int_\Omega |X| d\mathbb{P} < \infty.$$

If $X \in \mathscr{L}^1$, then $E(X) = \int_\Omega X d\mathbb{P}$ exists and is called the **expectation** of X. In general, if $h : \mathbb{R} \to \mathbb{R}$ is a Borel measurable function and $\int_\Omega |h(X)| d\mathbb{P} < \infty$, then

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

is called the expectation of $h(X)$. □

Remark 2.9. Recall from Definition 2.9 that a random variable X induces a probability measure $\mathbb{P}_X$ on $(\mathbb{R}, \mathscr{B})$. Thus the expectation of $h(X)$ can be equivalently expressed as

$$E(h(X)) = \int_\Omega h(X) d\mathbb{P} = \int_\Omega h(X(\omega))\mathbb{P}(d\omega) = \int_{\mathbb{R}} h(x) d\mathbb{P}_X(x).$$

Moreover, for any $B \in \mathscr{B}$, we have

$$E(h(X)1_{\{X \in B\}}) = \int_{X^{-1}(B)} h(X) d\mathbb{P} = \int_{X^{-1}(B)} h(X(\omega))\mathbb{P}(d\omega) = \int_B h(x) d\mathbb{P}_X(x).$$

Definition 2.15. (Squared Integrable Function, $\mathscr{L}^2$ and Variance) A random variable $X : \Omega \to \mathbb{R}$ is called **squared integrable** (or in $\mathscr{L}^2$) if

$$\int_\Omega |X|^2 d\mathbb{P} < \infty .$$

Then the **variance** of X can be defined as

$$\mathrm{Var}(X) = E((X - E(X))^2) = \int_\Omega (X - E(X))^2 d\mathbb{P} .$$

Example 2.28. On $(\Omega, \mathscr{F}, \mathbb{P}) = ([0,1], \mathscr{B}_{[0,1]}, \lambda)$, let $X : [0,1] \to \mathbb{R}$ be defined as $X(\omega) = \min\{\omega, 1-\omega\}$. The expectation and variance are given by

$$\begin{aligned}
E(X) &= \int_{[0,1]} X\, d\lambda = \int_0^1 X(\omega)\, d\omega = \int_0^{0.5} \omega\, d\omega + \int_{0.5}^1 1 - \omega\, d\omega = 0.25 , \\
\mathrm{Var}(X) &= \int_{[0,1]} (X - E(X))^2\, d\lambda = \int_0^1 (X(\omega) - 0.25)^2\, d\omega \\
&= \int_0^{0.5} (\omega - 0.25)^2\, d\omega + \int_{0.5}^1 (0.75 - \omega)^2\, d\omega = 0.25^2/3 .
\end{aligned}$$

Note that the induced probability measure can be computed as

$$\mathbb{P}_X(X \le x) = \mathbb{P}(\{\omega : X(\omega) \le x\}) = \begin{cases} \mathbb{P}(\{\omega \le x\} \cup \{1 - \omega \le x\}) , & 0 \le x \le 0.5 \\ 1 & x > 0.5 \end{cases}$$

$$= \begin{cases} \mathbb{P}(\{\omega \le x\}) + \mathbb{P}(\{\omega \ge 1 - x\}) = x + [1 - (1 - x)] = 2x , & 0 \le x \le 0.5 \\ 1 & x > 0.5 . \end{cases}$$

The induced probability space can be taken as $([0,0.5], \mathscr{B}_{[0,0.5]}, \mathbb{P}_X)$ or $(\mathbb{R}, \mathscr{B}, \mathbb{P}_X)$. The expectation and variance can be found as

$$\begin{aligned}
E(X) &= \int X\, d\mathbb{P}_X = \int_0^{0.5} x\, d(2x) = \int_0^{0.5} 2x\, dx = 0.25 , \\
\mathrm{Var}(X) &= \int (X - E(X))^2\, d\mathbb{P}_X = \int_0^{0.5} (x - 0.25)^2\, d(2x) = 0.25^2/3 .
\end{aligned}$$

□

2.7 Exercises

Exercise 2.4 *Show that the function f in Examples 2.2 and 2.3 are bijective. In each case, construct another example of bijective function f.*

Exercise 2.5 *Consider the partition* $[0,1] = \bigcup_{\alpha \in \Gamma} A_\alpha$ *defined in Section 2.2.5. Using the facts that* A_α *is countable for each* $\alpha \in \Gamma$*, show by contradiction that* Γ *is uncountable.*

Exercise 2.6 *Show that the set of complex number with rational real and imaginary parts,* $\{a + bi : a, b \in \mathbb{Q}\}$*, is countable.*

Exercise 2.7 *Using similar argument as in Example 2.6, show that the set of real number is uncountable.*

Exercise 2.8 *Verify that a* σ*-field is closed under countable intersections.*

Exercise 2.9 *Verify that* 2^Ω *and* $\{\emptyset, \Omega\}$ *are* σ*-fields on the sample space* Ω*.*

Exercise 2.10 *Let* $-\infty < a < b < \infty$*. Show that the following sets belong to the Borel* σ*-field* $\mathscr{B}$*:*

1. singleton*:* $\{x\}$*, where* $x \in \mathbb{R}$.
2. half closed interval*:* $[a,b)$ *or* $(a,b]$
3. closed interval*:* $[a,b]$.
4. $(-\infty, b)$ *and* (a, ∞).

Exercise 2.11 *Show that* $\mathscr{B}_{[a,b]}$ *is a* σ*-field for any* $-\infty < a < b < \infty$*.*

Exercise 2.12 *Explain whether the following* $\mathscr{F}$*s are* σ*-fields or not.*

a *Suppose* $\Omega = \{1,2,3,4,5,6\}$ *and* $\mathscr{F} = \{\{1,2\},\{3,4,5\},\{6\}\}$.
b *Suppose* Ω *and* $\mathscr{F}$ *are defined as in part a). Let* $\mathscr{F}^* = \{A : A$ *is the union of some subset in* $\mathscr{F}\}$ *(e.g.,* $A = \{1,2\} \cup \{3,4,5\} = \{1,2,3,4,5\} \in \mathscr{F}^*$*). Is* $\mathscr{F}^*$ *a* σ*-field?*

Exercise 2.13 *Consider Example 2.25b). Let* $\mathscr{F} = \{\emptyset, \{1,2,3\}, \{4,5,6\}, \Omega\}$*. Show that* X_1 *is a r.v. in the measurable space* $(\Omega, \mathscr{F})$ *but* X_2 *is not.*

Exercise 2.14 *What is the difference between a singleton and an element?*

Exercise 2.15 *Suppose that the sample space* Ω *is a finite set. Suppose we want to define a* σ*-field and probability measure associate with* Ω*. Explain why the notation* $\sigma(\Omega)$ *is meaningless. Show that the smallest* σ*-field containing all singletons of* Ω *is* 2^Ω*. Compare this set between* $\sigma(\{\Omega\})$

Exercise 2.16 *Show that the* σ*-field generated by a class of set* $\mathscr{A}$ *can be expressed as*

$$\sigma(\mathscr{A}) = \cap_\alpha \mathscr{F}_\alpha \,,$$

where $\{\mathscr{F}_\alpha\}$ *are all the* σ*-fields (possibly uncountable) that contain* $\mathscr{A}$*.*

Exercise 2.17 *The function* $f : \mathbb{R} \to \mathbb{R}$ *is called a Borel measurable function if* $f^{-1}(B)$ *of any Borel set B is also a Borel set. Show that any continuous function* $f : \mathbb{R} \to \mathbb{R}$ *is a Borel measurable function.*

Exercise 2.18 *Recall the proof of Theorem 2.3. Show that*

1. $(a,b) = \bigcup_{n=1}^{\infty} \left[a + \frac{1}{n}, b - \frac{1}{n}\right]$.
2. $[a,b] = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b + \frac{1}{n}\right)$.
3. $\mathscr{G} = \{B \in \mathscr{B} : X^{-1}(B) \in \mathscr{F}\}$ *is a* σ*-field.*

Exercise 2.19 *Show that every step function* $X : \mathbb{R} \to \mathbb{R}$ *is* $\mathscr{B}$*-measurable.*

Exercise 2.20 *Let* $\{f_i\}_{i=1,2,\ldots}$, $f_i : \Omega \to \mathbb{R}$ *be a sequence of measurable random variables. Let* $a \in \mathbb{R}$ *and* $k \in \mathbb{N}$.

1. *Show that* $\{\omega : f(\omega) + g(\omega) < a\} = \{f + g < a\}$ *is equal to* $\bigcup_{n=1}^{\infty} [\{f < q_n\} \cap \{g < a - q_n\}]$, *where* $\{q_n\}$ *is a sequence representing all rational numbers.*
2. *Hence, show that* $f_1 + f_2$ *is measurable.*
3. *Show that* $\sum_{i=1}^{\infty} f_i$ *is measurable.*
4. *Show that* $\{\sup_{i\geq 1} f_i > a\} = \bigcup_{i=1}^{\infty} \{f_i > a\}$ *and* $\{\min_{i\leq k} f_i > a\} = \bigcap_{i=1}^{k} \{f_i > a\}$.
5. *Show that* $\sup_{i\geq 1} f_i$ *and* $\min_{i\leq k} f_i$ *are measurable.*

Exercise 2.21 *Verify that* $\sigma(X) = \sigma\{X^{-1}(B) | B \in \mathscr{B}\}$ *defined in Definition 2.8 can be written as* $\{X^{-1}(B) | B \in \mathscr{B}\}$. *That is, show that* $\{X^{-1}(B) | B \in \mathscr{B}\}$ *satisfies the definition of a* σ*-field.*

Exercise 2.22 *Suppose that* $\mathscr{F}$ *is a* σ*-field generated by a finite partition of* Ω, $\mathscr{A} = \{A_1, A_2, \ldots, A_n\}$, *i.e.,* $\Omega = \bigcup_{i=1}^{n} A_i$ *and* $A_i \cap A_j = \varnothing$ *if* $i \neq j$. *Show that every* $F \in \mathscr{F}$ *is a union of* A_i*s.*

Exercise 2.23 *Show that* $\{1,2,3\}$ *is a Lebesgue-null set using Definition 2.2.*

Exercise 2.24 *Give two examples of simple functions. Give two examples of step function. Is step function always a simple function? Is simple function always a step function?*

Exercise 2.25 *Using the fact that every step function is a simple function, show that if a function* f *is Riemann integrable, then it is Lebesgue integrable, and the values of the two integrals are the same.*

Exercise 2.26 *The* **(Doob-Dynkin)** *theorems states that*

Doob-Dynkin Theorem
Let X *be a random variable. Then each* $\sigma(X)$*-measurable random variable* Y *can be written as* $Y = f(X)$ *for some Borel function* $f : \mathbb{R} \to \mathbb{R}$.

Instead of showing the theorem we consider one illustrative example: Let $\Omega = \{1,2,3,4\}$ *and* $X : \Omega \to \mathbb{R}$ *be a random variable such that* $X = i^2$ *for* $i = 1,2,3,4$. *Clearly* $\sigma(X)$ *is just the power set* 2^{Ω} *of* Ω *(i.e. the set of all subsets of* Ω*). Let the r.v.* $Y : \Omega \to \mathbb{R}$ *be defined as*

$$Y(i) = \begin{cases} 0 \text{ if } & i = 1,2 \\ 1 \text{ if } & i = 3,4. \end{cases}$$

1) Find $\sigma(Y)$.
2) Show that Y is $\sigma(X)$ measurable.
3) Construct a Borel function $f : \mathbb{R} \to \mathbb{R}$ such that $Y = f(X)$.

Exercise 2.27 *Let $A_n = \{q_1, \dots, q_n\}$ and $f_n(x) = 1_{\{x \in A_n\}}$, where $q_i \in \mathbb{R}$. Find the Riemann integral and Lebesgue integral of f_n on $\mathbb{R}$.*

Exercise 2.28 *1) Show that if $X \in \mathscr{L}^2$ then $X \in \mathscr{L}^1$.*
2) If $\eta : \Omega \to [0,\infty)$ is in $\mathscr{L}^2$ and nonnegative, then

$$E(\eta^2) = 2\int_0^\infty t\mathbb{P}(\eta > t)dt\,.$$

Exercise 2.29 *Consider $(\Omega, \mathscr{F}, \mathbb{P}) = ([0,1], \mathscr{B}_{[0,1]}, \lambda_{[0,1]})$. Mark claims that, since $\{x\}$ and $\{y\}$ are disjoint sets if $x \neq y$, by the additivity of probability measure, one has*

$$\mathbb{P}([0,1]) = \mathbb{P}\left(\bigcup_{x\in[0,1]} \{x\}\right) = \sum_{x\in[0,1]} \mathbb{P}(\{x\})\,.$$

Explain whether this claim is correct.

Exercise 2.30 *Consider $(\Omega, \mathscr{F}, \mathbb{P}) = ([0,2], \mathscr{B}_{[0,2]}, \frac{1}{2}\lambda_{[0,2]})$, and*

$$X(\omega) = \begin{cases} 1 & \textit{if } \omega \in \mathbb{Q}\backslash\{1\}\,, \\ 3 & \textit{if } \omega \in [0,1)\backslash\mathbb{Q}\,, \\ 5 & \textit{if } \omega \in (1,2]\backslash\mathbb{Q}\,, \\ 7 & \textit{if } \omega = 1\,. \end{cases}$$

Find $E(X)$ and $\mathrm{Var}(X)$.

Chapter 3
Limit Theorems

In probability and statistics, we frequently encounter approximation of a quantity by an infinite sequence. For example,

i) the Cantor set $E_4 = \bigcap_{i=1}^{\infty} C_i$ in (2.3) is the limit of the intersection of the sequence of sets $\{C_1, C_2, \ldots,\}$, i.e. $\lim_{n\to\infty} \bigcap_{i=1}^{n} C_i$.
ii) in defining the Lebesgue integral $\int f\,d\lambda$ for a measurable function f, we approximate f by an increasing sequence of measurable functions $f_n(x) = \sum_{i=1}^{n} a_i 1_{A_i}(x)$, for which the integral $\int f_n\,d\lambda = \sum_{i=1}^{n} a_i \lambda(A_i)$ is well defined.
iii) in statistics, we estimate an unknown parameter θ by an estimator $\hat{\theta}_n$, which is a statistic computed from a sample of n observations.

In Example i), in order for the Cantor set to be well defined, we require that $\bigcap_{i=1}^{n} C_i$ converges to some set. In Example ii), we require that the sequence of functions f_n converges to f as n goes to infinity. In Example iii), we wish that $\hat{\theta}_n$ converges to θ as n goes to infinity. That is, when more data comes, the estimator $\hat{\theta}_n$ estimates the unknown parameter θ more accurately. However, since C_is are sets, f_ns are functions and $\hat{\theta}_n$s are random variables, in contrast to real numbers, there are many ways to measure the distance between two functions or two variables (see Section 3.1). Thus, we need to pay special attention to the meaning of "*convergence*" as we take the limit $n \to \infty$.

3.1 Measure of distance

When we are given two **real numbers** x and y, the distance between them is the absolute value

$$|x - y|\,.$$

If $x = (x_1, \ldots, x_m)$ and $y = (y_1, \ldots, y_m)$ are m-**dimensional vectors**, then the distance between them, $d(x,y)$, can be measured in various different ways. For example,

- j-**th component distance**: $d_j(x,y) = |x_j - y_j|\,.$

- **maximal distance**: $d_{\max}(x,y) = \max_{i=1,\dots,m} |x_i - y_i|$.
- **absolute distance**: $d_1(x,y) = \sum_{i=1}^m |x_i - y_i|$.
- **squared distance**: $d_2(x,y) = \sum_{i=1}^m (x_i - y_i)^2$.
- **Euclidean distance**: $d(x,y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2}$.

A measure of distance between two points is formally called a *metric* if it satisfies the following conditions.

Definition 3.1. (Metric) Let x, y and z be points on some space (real number, sequence, function, etc...). A real-valued distance function $d(x,y)$ is called a **metric** if it satisfies

i) $d(x,y) \geq 0$ and $d(x,y) = 0$ if and only if $x = y$;
ii) $d(x,y) = d(y,x)$;
iii) $d(x,y) \leq d(x,z) + d(z,y)$ for all z *(Triangle Inequality)* .

Note that not all measures of distance above are metrics. For example, if $x = (1,0,1)$ and $y = (1,2,3)$, then $x \neq y$ but the 1-st component distance between x and y is 0, violating (i). Indeed, it is straightforward to check that all distances above are metrics except the j-th component and squared distances (see Exercise 3.8).

Generally, there is no right or wrong in choosing which metric to use. The choice usually depends on the specific properties of the metrics. For example, absolute distance is less popular since difficulty arises from the absolute values when we consider differentiation. In this sense, squared distance is more preferable. However, the unit of the squared distance is not the same as the unit of the original x and y. Hence, the Euclidean distance, which preserves unit by its square root, naturally becomes the standard choice of metric for vectors in $\mathbb{R}^n$.

For two **sets** X and Y in some measurable space $(\Omega, \mathscr{F})$, distance can be defined as

- **μ-distance**: $d_\mu(X,Y) = \mu(X \triangle Y) = \mu(X \backslash Y) + \mu(Y \backslash X)$,

where μ is a measure satisfying Definition 2.6, and $X \triangle Y = (X \backslash Y) \cup (Y \backslash X)$. Since some non-empty sets may have zero μ-measure, $d_\mu(X,Y) = 0$ does not imply $X = Y$ in general.

Example 3.1. i) If $\mu = \lambda$, the Lebesgue measure, then $X = \mathbb{Q}$ and $Y = \emptyset$ are not equal but their distance is $d_\lambda(X,Y) = \lambda(\mathbb{Q}) = 0$.
ii) For the measure

$$m(A) = \begin{cases} 1 & \text{if } A \neq \emptyset \\ 0 & \text{if } A = \emptyset \end{cases}, \tag{3.1}$$

we have $d_m(X,Y) = 0$ if and only if $X = Y$ (see Exercise 3.9).

For two **measurable functions** $x = x(u)$ and $y = y(u)$ on $u \in \mathbb{R}$, some popular measures of distance include

- **pointwise distance at** t: $d_t(x,y) = |x(t) - y(t)|$ for a fixed $t \in \mathbb{R}$.
- **supremum distance**: $d_{\text{sup}}(x,y) = \sup_{u\in\mathbb{R}} |x(u) - y(u)|$.
- **total variation distance**: $d_1(x,y) = \int_{\mathbb{R}} |x-y|\, d\lambda$.
- **quadratic variation distance**: $d_2(x,y) = \int_{\mathbb{R}} (x-y)^2\, d\lambda$.
- $\mathscr{L}^p$ **distance**: $d(x,y) = (\int_{\mathbb{R}} |x-y|^p\, d\lambda)^{\frac{1}{p}}$.

Note that the pointwise, supremum, total variation, quadratic variation and $\mathscr{L}^2$ ($\mathscr{L}^p$ at $p=2$) distances of functions extend the j-th component, maximal, absolute, squared and Euclidean distance, respectively. However, all of these distances are not metrics except the supremum distance since they violate (i) in Definition 3.1.

Example 3.2. If $x(u) = 1_{\{u\in\mathbb{Q}_{[0,1]}\}}$ and $y(u) = 0$, then obviously $x \neq y$. However, as $\lambda(\mathbb{Q}) = 0$, we have $\int_{\mathbb{R}} |x-y|\, d\lambda = \int_{\mathbb{R}} (x-y)^2\, d\lambda = (\int_{\mathbb{R}} |x-y|^p\, d\lambda)^{\frac{1}{p}} = 0$. On the other hand, the supremum distance $\sup_{u\in\mathbb{R}} |x(u) - y(u)| = 1$ is able to spot out the difference between x and y.

Indeed, if $\sup_{u\in\mathbb{R}} |x(u) - y(u)| = 0$, then we must have $x(u) = y(u)$ for all $u \in \mathbb{R}$. Thus, the supremum distance satisfies (i) in Definition 3.1. □

Finally, we consider two **random variables** $X = X(\omega)$ and $Y = Y(\omega)$ on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Notice that random variables are functions on the domain Ω. Thus, the metrics are expected to be similar to the metrics of measurable functions. Nevertheless, the concepts of probability and distribution function introduce more possibilities in defining distances for random variables. Examples include

- **pointwise distance at** ω: $d_\omega(X,Y) = |X(\omega) - Y(\omega)|$ for a fixed $\omega \in \Omega$.
- **supremum distance**: $d_{\text{sup}}(X,Y) = \sup_{\omega\in\Omega} |X(\omega) - Y(\omega)|$.
- **total variation distance**: $d_1(X,Y) = \int_\Omega |X-Y|\, d\mathbb{P}$.
- **quadratic variation distance**: $d_2(X,Y) = \int_\Omega (X-Y)^2\, d\mathbb{P}$.
- $\mathscr{L}^p$ **distance**: $d_p(X,Y) = (\int_\Omega |X-Y|^p\, d\mathbb{P})^{\frac{1}{p}}$.
- **probability of ε-separation**: $d_{\text{pr}}(X,Y) = \mathbb{P}(|X-Y| > \varepsilon)$ for a fixed $\varepsilon > 0$.
- **distribution function distance at** c: $d_{\text{df}}(X,Y) = |\mathbb{P}(X \le c) - \mathbb{P}(Y \le c)|$ for a fixed c.

Note that $\mathbb{P}(X \le c)$ represents $\mathbb{P}(\{\omega : X(\omega) \le c\})$. Similar to measurable functions, it can be verified that only the supremum distance is a metric.

3.2 Modes of Convergence

Given a sequence $\{x_n\}_{n=1}^\infty = \{x_1, x_2, \ldots, x_n, \ldots\}$, it is natural to define its convergence by the distance between x_n and some *limit* x.

Definition 3.2. (Convergence) Given a distance function d, we say that the sequence $x_1, x_2, \ldots, x_n, \ldots$ converges to x in d if

$$d(x_n, x) \to 0, \quad \text{as } n \to \infty.$$

More precisely, $d(x_n,x) \to 0$ as $n \to \infty$ means that given any $\varepsilon > 0$, there exists an integer $N > 0$ such that $d(x_n,x) < \varepsilon$ for all $n > N$.

Example 3.3. If $\{x_n\}_{n=1}^{\infty}$ is a sequence of vectors, then convergence in maximal distance $d_{\max}$ is equivalent to convergence in j-component distance d_j for all $j = 1,\ldots,m$. To see this, first note that if $\{x_n\}_{n=1}^{\infty}$ converges in $d_{\max}$, then for $j = 1,\ldots,m$,

$$d_j(x_n,x) = |x_{n,j} - x_j| < \max_{j=1,\ldots,m} |x_{n,j} - x_j| = d_{\max}(x_n,x) \to 0\,, \tag{3.2}$$

as $n \to \infty$. Thus, $\{x_n\}_{n=1}^{\infty}$ converges in d_j for all j. Conversely, if $\{x_n\}_{n=1}^{\infty}$ converges in d_j for all j, then by Definition 3.2, given any $\varepsilon > 0$, there exists an integer $N_j > 0$ such that $d_j(x_n,x) < \varepsilon$ for all $n > N_j$. Taking $N = \max_{j=1,\ldots,m} N_j$, then for $n > N$,

$$d_{\max}(x_n,x) = \max_{j=1,\ldots,m} |x_{n,j} - x_j| = \max_{j=1,\ldots,m} d_j(x_n,x) < \varepsilon\,,$$

implying convergence in $d_{\max}$. □

Using arguments similar to Example 3.3, it can be shown that $d_{\max}$, d_1, d_2, and d are all equivalent for the convergence of *finite-dimensional vectors*. However, for a sequence of functions, we lose the equivalence in general since the domain of a real function is infinite dimensional.

Example 3.4. If $\{x_n\}_{n=1}^{\infty}$ is a sequence of real-valued functions on some domain $\mathbb{D}$, then convergence in $d_t(x,y)$ and $d_{\sup}$ correspond to **pointwise** and **uniform** convergence, respectively. Specifically, $x_n \to x$ **pointwisely** if $x_n(t) \to x(t)$ for all $t \in \mathbb{D}$ as $n \to \infty$. Also, $x_n \to x$ **uniformly** if $\sup_{t\in\mathbb{D}} |x_n(t) - x(t)| \to 0$ as $n \to \infty$.

In contrast to vectors, pointwise and uniform convergences of functions are not equivalent: Similar to (3.2), we have uniform convergence implies pointwise convergence. However, the opposite direction does not hold. For a counter example, consider $x_n(t) = t^n$ on $t \in \mathbb{D} \triangleq (0,1)$. Note that for each $t \in (0,1)$, $x_n(t) = t^n \to 0$. Hence, $\{x_n\}_{n=1}^{\infty}$ pointwisely converges to $x(t) \equiv 0$. However, for each $\varepsilon \in (0,1)$ and $n \in \mathbb{Z}^+$, we have

$$\begin{aligned}
\sup_{t\in(0,1)} |x_n(t) - x(t)| &= \sup_{t\in(0,1)} |t^n| && (\text{definition of } x_n \text{ and } x)\\
&\geq |(\varepsilon^{1/n})^n| = \varepsilon && (\text{substitute } t = \varepsilon^{1/n} \in (0,1))\,.
\end{aligned}$$

Thus, x_n does not converge to x uniformly. □

3.2.1 Convergence of Random Variables

For sequences of random variables, four common modes of convergence are listed as follows:

Definition 3.3. Consider a sequence of random variables $\{X_n\}_{n=1}^{\infty}$ on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$. The sequence $\{X_n\}_{n=1}^{\infty}$ is said to be convergent to a random variable X on $(\Omega, \mathscr{F}, \mathbb{P})$

i) **almost surely (a.s.)**, denoted as $X_n \xrightarrow{a.s.} X$, if for some $A \in \mathscr{F}$ with $\mathbb{P}(A) = 1$, we have

$$X_n(\omega) \to X(\omega)\,, \tag{3.3}$$

as $n \to \infty$ for all $\omega \in A$. Equivalently, we can write $\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$.

ii) **in probability**, denoted as $X_n \xrightarrow{p} X$, if for all $\varepsilon > 0$, we have

$$\mathbb{P}(|X_n - X| > \varepsilon) \to 0\,. \tag{3.4}$$

as $n \to \infty$.

iii) **in distribution**, denoted as $X_n \xrightarrow{d} X$, if for all $c \in \mathbb{R}$ such that $\mathbb{P}(X \le x)$ is continuous at $x = c$, we have

$$|\mathbb{P}(X_n \le c) - \mathbb{P}(X \le c)| \to 0\,, \tag{3.5}$$

as $n \to \infty$.

iv) **in** $\mathscr{L}^2$, denoted as $X_n \xrightarrow{\mathscr{L}^2} X$, if

$$\mathbb{E}(X_n - X)^2 = \int_{\Omega} (X_n - X)^2\, d\mathbb{P} \to 0\,, \tag{3.6}$$

as $n \to \infty$.

Clearly, convergence almost surely, in probability, in distribution and in $\mathscr{L}^2$ correspond to convergences in pointwise distance, probability of ε-separation, distribution function and quadratic variation distances, respectively. The following theorem describes some relationships between different modes of convergence.

Theorem 3.1. (Relationships between modes of convergence)
Consider a random variable X and a sequence $\{X_n\}_{n=1}^{\infty}$ on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

i) If $X_n \xrightarrow{a.s.} X$, then $X_n \xrightarrow{p} X$.
ii) If $X_n \xrightarrow{\mathscr{L}^2} X$, then $X_n \xrightarrow{p} X$.
iii) If $X_n \xrightarrow{p} X$, then $X_n \xrightarrow{d} X$.

Proof. i) From the definition of a.s. convergence and limit, we have

$$
\begin{aligned}
& X_n \xrightarrow{a.s.} X \\
\Leftrightarrow\ & \mathbb{P}\left(\lim_{n\to\infty} |X_n - X| = 0\right) = 1 \\
\Leftrightarrow\ & \mathbb{P}(\{\forall\varepsilon > 0\ \exists N \in \mathbb{Z}^+ : \forall n \geq N, |X_n - X| < \varepsilon\}) = 1 && \text{(definition of limit)} \\
\Leftrightarrow\ & \mathbb{P}(\{\exists\varepsilon > 0\ \forall N \in \mathbb{Z}^+ : \exists n \geq N, |X_n - X| \geq \varepsilon\}) = 0 && \text{(take complement)} \\
\Leftrightarrow\ & \mathbb{P}\left(\bigcup_{\varepsilon>0} \bigcap_{N\in\mathbb{Z}^+} \bigcup_{n=N}^{\infty} \{|X_n - X| \geq \varepsilon\}\right) = 0 \\
\Leftrightarrow\ & \forall\varepsilon > 0, \quad \mathbb{P}\left(\bigcap_{N\in\mathbb{Z}^+} \bigcup_{n=N}^{\infty} \{|X_n - X| \geq \varepsilon\}\right) = 0 \\
\overset{(3.15)}{\Leftrightarrow}\ & \forall\varepsilon > 0, \quad \lim_{N\to\infty} \mathbb{P}\left(\bigcup_{n=N}^{\infty} \{|X_n - X| \geq \varepsilon\}\right) = 0 \\
\Rightarrow\ & \forall\varepsilon > 0, \quad \lim_{N\to\infty} \mathbb{P}\left(\{|X_N - X| \geq \varepsilon\}\right) = 0 \\
\Leftrightarrow\ & X_n \xrightarrow{p} X\,.
\end{aligned}
$$

Note that the second last implication is only one-sided.

ii) If $X_n \xrightarrow{\mathscr{L}^2} X$, then $E\left(|X_n - X|^2\right) \to 0$. Thus, for all $\varepsilon > 0$,

$$
\begin{aligned}
\mathbb{P}(|X_n - X| > \varepsilon) &= \mathbb{P}(|X_n - X|^2 > \varepsilon^2) = E\left(1_{\{|X_n - X|^2 > \varepsilon^2\}}\right) \\
&\leq E\left(\frac{|X_n - X|^2}{\varepsilon^2} 1_{\{|X_n - X|^2 > \varepsilon^2\}}\right) && \text{(on } \{|X_n - X|^2 > \varepsilon^2\}\text{)} \\
&\leq \frac{1}{\varepsilon^2} E\left(|X_n - X|^2\right) && \text{(get rid of indicator)} \\
&\to 0\,,
\end{aligned}
$$

implying $X_n \xrightarrow{p} X$.

iii) First, we have

$$
\begin{aligned}
\mathbb{P}(X_n \leq c) &= \mathbb{P}(X_n \leq c, X \leq c + \varepsilon) + \mathbb{P}(X_n \leq c, X > c + \varepsilon) \\
&\leq \mathbb{P}(X \leq c + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)\,. \qquad (3.7)
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
\mathbb{P}(X \leq c - \varepsilon) &= \mathbb{P}(X_n \leq c, X \leq c - \varepsilon) + \mathbb{P}(X_n > c, X \leq c - \varepsilon) \\
&\leq \mathbb{P}(X_n \leq c) + \mathbb{P}(|X_n - X| > \varepsilon)\,. \qquad (3.8)
\end{aligned}
$$

Combining (3.7) and (3.8), we have

$$
\mathbb{P}(X \leq c - \varepsilon) - \mathbb{P}(|X_n - X| > \varepsilon) \leq \mathbb{P}(X_n \leq c) \leq \mathbb{P}(X \leq c + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)\,.
$$

Taking $n \to \infty$ and then $\varepsilon \to 0$, the continuity of $\mathbb{P}(X \le c)$ at c allows us to conclude that $\mathbb{P}(X_n \le c) \to \mathbb{P}(X \le c)$. $\square$

Example 3.5. We list some counter examples for the converse of each statement in Theorem 3.1.

i) Let $\{X_n\}_{n\in\mathbb{Z}^+}$ be random variables in $(\Omega,\mathscr{F},\mathbb{P}) = ([0,1],\mathscr{B}_{[0,1]},\lambda)$, with

$$X_1 = 1_{[0,\frac{1}{2}]}(\omega),\ X_2 = 1_{[\frac{1}{2},1]}(\omega),\ X_3 = 1_{[0,\frac{1}{4}]}(\omega),\ X_4 = 1_{[\frac{1}{4},\frac{1}{2}]}(\omega),\ X_5 = 1_{[\frac{1}{2},\frac{3}{4}]}(\omega),$$
$$X_6 = 1_{[\frac{3}{4},1]}(\omega),\ X_7 = 1_{[0,\frac{1}{8}]}(\omega),\ \ X_8 = 1_{[\frac{1}{8},\frac{1}{4}]}(\omega),$$

and so on .... Note that as n increases, the interval where $X_n \ne 0$ decreases in size, and moves over the space $[0,1]$. Thus, it is clear that for all $\varepsilon \in (0,1)$, $\mathbb{P}(X_n > \varepsilon) = \mathbb{P}(X_n = 1) \to 0$ as $n \to \infty$. Thus $X_n \xrightarrow{p} X$ with $X = 0$. However, for each $\omega \in [0,1]$, there are infinitely many n such that $X_n(\omega) = 1$. Thus $X_n(\omega) \nrightarrow 0$ for all $\omega \in [0,1]$. Hence $\mathbb{P}(\omega : \lim_{n\to\infty} X_n(\omega) = X) = 0 \ne 1$ and $X_n \overset{a.s.}{\nrightarrow} X$.

ii) Consider $X_n(\omega) = n \times 1_{\{0\le\omega\le 1/n\}}$ on $(\Omega,\mathscr{F},\mathbb{P}) = ([0,1],\mathscr{B}_{[0,1]},\lambda)$. For all $\varepsilon > 0$, $\mathbb{P}(X_n > \varepsilon) = \mathbb{P}(X_n = n) = \frac{1}{n} \to 0$ as $n \to \infty$. Thus $X_n \xrightarrow{p} X$ with $X = 0$. However, $E(X_n - X)^2 = \int n^2 1_{\{0\le\omega\le 1/n\}}\, d\lambda = n \to \infty$. Thus $X_n \overset{\mathscr{L}^2}{\nrightarrow} X$.

iii) If $X, X_1, X_2, \ldots$ are independent and identically distributed $N(0,1)$ random variables, then it is obvious that $X_n \xrightarrow{d} X$ and $X_n \overset{p}{\nrightarrow} X$.

$\square$

Example 3.6. **(a.s. and $\mathscr{L}^2$ convergence)**

i) **(a.s. does not imply $\mathscr{L}^2$ convergence)** Consider $X_n(\omega) = n^2\omega \times 1_{\{0\le\omega\le 2/n\}}$ and $X(\omega) = 0$ on $(\Omega,\mathscr{F},\mathbb{P}) = ([0,1],\mathscr{B}_{[0,1]},\lambda)$. Clearly, for $\omega = 0$, $X_n(\omega) = 0$ for all n. Also, for any $\omega > 0$, $X_n(\omega) = 0$ for $n > 2/\omega$. In other words, for each $\omega \in (0,1]$, $X_n(\omega) \to X(\omega)$. As $\mathbb{P}((0,1]) = 1$, we have $X_n \xrightarrow{a.s.} X$. However, $E(X_n - X)^2 = E(X_n^2) = n^4 \int \omega^2 \times 1_{\{0\le\omega\le 2/n\}}\, d\omega = \frac{8n}{3} \nrightarrow 0$. Thus $X_n \overset{\mathscr{L}^2}{\nrightarrow} X$.

ii) **($\mathscr{L}^2$ does not imply a.s. convergence)** Consider $X = 0$ and X_n in Example 3.5i) in which we showed that $X_n \overset{a.s.}{\nrightarrow} X$. However, using the fact that X_n is an indicator function, we have $E(X_n - X)^2 = E(X_n^2) = \mathbb{P}(X_n = 1) \to 0$. Hence, $X_n \xrightarrow{\mathscr{L}^2} X$. $\square$

Example 3.7. **(Importance of continuity point for convergence in distribution)** Recall the convergence in distribution in Definition 3.3iii), we only require $|\mathbb{P}(X_n \le c) - \mathbb{P}(X \le c)| \to 0$ at the **continuity points** of the distribution function of X, i.e, $\{c \in \mathbb{R} : \mathbb{P}(X \le x)$ is continuous at $x = c\}$. This exclusion of the discontinuity points is to avoid the effect of **point masses** of X, i.e., $\{c \in \mathbb{R} : \mathbb{P}(X = c) > 0\}$.

Consider $X_n = \left(1 + \frac{1}{n}\right) 1_{\{\omega\ge 0.5\}}$ and $X = 1_{\{\omega\ge 0.5\}}$ on $(\Omega,\mathscr{F},\mathbb{P}) = ([0,1],\mathscr{B}_{[0,1]},\lambda)$. Note that $X_n(\omega) \to X(\omega)$ for all $\omega \in [0,1]$. That is, $X_n \xrightarrow{a.s.} X$. Hence Theorem 3.1 i) and iii) ensures that $X_n \xrightarrow{d} X$. However,

$$\mathbb{P}(X_n \le x) = \begin{cases} 0 & \text{if } x < 0 \\ \mathbb{P}(\{\omega < 0.5\}) = \frac{1}{2} & \text{if } x \in [0, 1+\frac{1}{n}) \\ 1 & \text{if } x \ge 1+\frac{1}{n} \end{cases}, \quad \mathbb{P}(X \le x) = \begin{cases} 0 & \text{if } x < 0 \\ \mathbb{P}(\{\omega < 0.5\}) = \frac{1}{2} & \text{if } x \in [0,1) \\ 1 & \text{if } x \ge 1 \end{cases}.$$

Note that $\mathbb{P}(X_n \le x) \to \mathbb{P}(X \le x)$ for all $x \in \mathbb{R}\backslash\{1\}$, but $\mathbb{P}(X_n \le 1) = \frac{1}{2} \nrightarrow 1 = \mathbb{P}(X \le 1)$ for all $n \in \mathbb{Z}^+$.

In this example, X_n has a point mass at $1+\frac{1}{n}$ and X has a point mass at 1. Although there is no difference asymptotically, the distribution function of X_n does not converge to that of X at the point 1. Hence, to avoid this artifact, we focus on the continuity points of the distribution function. □

3.3 Limit, limsup and liminf

We have seen that a sequence $\{x_n\}_{n=1}^{\infty}$ may converge to a limit x in different distance measures. On the other hand, even if the limit of $\{x_n\}_{n=1}^{\infty}$ does not exist, we may still define the **limit supremum (limsup)** or **limit infimum (liminf)** to describe the behavior of x_n for large n. Intuitively, limsup (liminf) is the greatest (smallest) possible value in the limit.

Definition 3.4. (limsup and liminf) For a sequence of real numbers $\{x_n\}_{n=1}^{\infty}$,

$$\begin{aligned} \limsup_{n\to\infty} x_n &= \inf_{n\ge 1} \sup_{k\ge n} x_k\,, \\ \liminf_{n\to\infty} x_n &= \sup_{n\ge 1} \inf_{k\ge n} x_k\,. \end{aligned}$$

For a sequence of m-dimensional real vectors $\{\mathbf{x}_n\}_{n=1}^{\infty}$, where $\mathbf{x}_n = (x_{n,1}, \ldots, x_{n,m})$, the limsup and liminf are component-wisely defined; e.g., the j-th entry of $\limsup_{n\to\infty} \mathbf{x}_n$ is equal to $\limsup_{n\to\infty} x_{n,j}$.

For a sequence of real-valued functions $\{f_n\}_{n=1}^{\infty}$, the limsup and liminf are pointwisely defined; i.e., for each $u \in \mathbb{R}$,

$$\begin{aligned} \limsup_{n\to\infty} f_n(u) &= \inf_{n\ge 1} \sup_{k\ge n} f_k(u)\,, \\ \liminf_{n\to\infty} f_n(u) &= \sup_{n\ge 1} \inf_{k\ge n} f_k(u)\,. \end{aligned}$$

For a sequence of random variables $\{X_n\}_{n=1}^{\infty}$ on $(\Omega, \mathscr{F}, \mathbb{P})$, similar to real-valued functions, we define for each $\omega \in \Omega$,

$$\begin{aligned} \limsup_{n\to\infty} X_n(\omega) &= \inf_{n\ge 1} \sup_{k\ge n} X_k(\omega)\,, \\ \liminf_{n\to\infty} X_n(\omega) &= \sup_{n\ge 1} \inf_{k\ge n} X_k(\omega)\,. \end{aligned}$$

On the other hand, for a sequence of sets $\{A_n\}_{n=1}^{\infty}$,

$$\limsup_{n\to\infty} A_n = \bigcap_{n\geq 1}\bigcup_{k\geq n} A_k = \{\omega\in\Omega : \omega\in A_k \text{ for infinitely many } k\},$$
$$\liminf_{n\to\infty} A_n = \bigcup_{n\geq 1}\bigcap_{k\geq n} A_k = \{\omega\in\Omega : \omega\in A_k \text{ for all but finitely many } k\}.$$

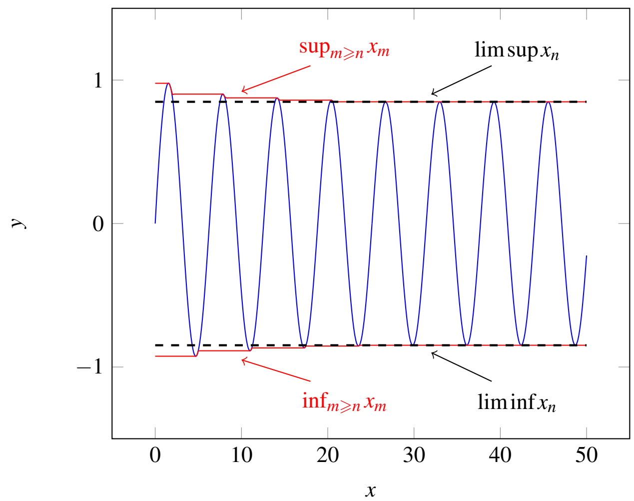


Fig. 3.1 limsup and liminf of a sequence.

Example 3.8. i) If $x_n = \frac{1}{n}$, then $\limsup_{n\to\infty} x_n = \liminf_{n\to\infty} x_n = 0$.
ii) If $x_n = (-1)^n$, then $\limsup_{n\to\infty} x_n = 1 \neq -1 = \liminf_{n\to\infty} x_n$.
iii) If $\mathbf{x}_n = ((-1)^{n+1}, (-1)^n)$, then $\limsup_{n\to\infty} \mathbf{x}_n = (1,1) \neq (-1,-1) = \liminf_{n\to\infty} x_n$.
iv) If $f_n(u) = (-u)^n$ on $u\in[0,1]$, then

$$\limsup_{n\to\infty} f_n(u) = \begin{cases} 1 & \text{for } u=1 \\ 0 & \text{for } u<1 \end{cases},$$
$$\liminf_{n\to\infty} f_n(u) = \begin{cases} -1 & \text{for } u=1 \\ 0 & \text{for } u<1 \end{cases}.$$

v) For a sequence of intervals $\{A_n\}$ with $A_n = (\sin n, 2)$, we have $\limsup A_n = (-1,2)$ and $\liminf A_n = (1,2)$.

Remark 3.1. For any sequence $\{x_n\}_{n\in\mathbb{Z}^+}$, if $\limsup_{n\to\infty} x_n = \liminf_{n\to\infty} x_n$, then we say the limit, denoted as $\lim_{n\to\infty} x_n$, exists. Note also that, while from Section 3.2 we have to specify the mode of convergence (i.e., distance function) when we define the limit, here the limsup and liminf are always the pointwise limits. □

3.4 Limit Theorems on Integrals

Given random variables $X_n \overset{a.s.}{\to} X$, it may seem obvious that $E(X_n) \to E(X)$. Surprisingly, this is not true in general, as the following example shows.

Example 3.9. Let $X_n(\omega) = n \times 1_{\{\omega<1/n\}}$ on $([0,1], \mathscr{B}_{[0,1]}, \lambda_{[0,1]})$. For each fixed ω in $(0,1]$, $X_n(\omega) = 0$ for all $n \geq 1/\omega$. This implies $\{\omega : \lim_{n\to\infty} X_n(\omega) = 0\} \supseteq (0,1]$, which by definition means $X_n \overset{a.s.}{\to} X \triangleq 0$ because

$$\mathbb{P}(\{\omega : \lim_{n\to\infty} X_n(\omega) = 0\}) \geq \mathbb{P}((0,1]) = 1\,.$$

Hence, $E(X) = \int \lim_{n\to\infty} X_n \, d\mathbb{P} = \int 0 \, d\lambda_{[0,1]} = 0$. However, for every integer n, $E(X_n) = \int X_n \, d\mathbb{P} = n\lambda_{[0,1]}\left(\left[0, \frac{1}{n}\right)\right) = 1$. As a result, we have

$$0 = E(X) < \lim_{n\to\infty} E(X_n) = 1\,.$$

The basic intuitions behind this example is that the a.s. limit ignores the magnitude of X_n on the set which becomes null in the limit. However, the magnitude (in $n \times 1_{\{\omega<1/n\}}$) still contributes significantly on the integral. □

Expressing in the integral form, the convergence $E(X_n) \to E(X)$ becomes

$$\lim_{n\to\infty} \int X_n \, d\mathbb{P} = \int \lim_{n\to\infty} X_n \, d\mathbb{P}\,, \tag{3.9}$$

which involves an **exchange of an integral and a limit**. Exchanging limits and integrals arises frequently in probability and statistics, and thus it is important to develop conditions that allow such exchanges. As these conditions apply to random variables, we first discuss the way to describe a statement about random variables.

A statement about a r.v. X is said to hold **almost surely (a.s.)** if $X(\omega)$ satisfies the statement for all $\omega \in A$ where A is some event of probability 1 (i.e., $\mathbb{P}(A) = 1$). For example, let $X : [0,1] \to \mathbb{R}$ be a r.v. on $([0,1], \mathscr{B}_{[0,1]}, \lambda)$ such that $X(\omega) = 1_{\mathbb{Q}}(\omega)$. We can say that $X = 0$ a.s., since $X(\omega) = 0$ for $\omega \in [0,1]\backslash\mathbb{Q}$ and $\lambda([0,1]\backslash\mathbb{Q}) = 1$. Note that A need not be Ω, but it can be $\Omega\backslash N$ for some null set N. This motivates the term "almost surely".

The following are three classical results about the exchange of limit and integral.

Theorem 3.2. (Monotone Convergence Theorem (MCT)) *If for all $n \geq 0$, $X_n \geq 0$ and $X_n \uparrow X$ a.s., (i.e., $X_n \leq X_{n+1}$ and $X_n \to X$ a.s.), then*

$$\lim_{n\to\infty} \int X_n \, d\mathbb{P} = \int X \, d\mathbb{P} = \int \lim_{n\to\infty} X_n \, d\mathbb{P}\,.$$

Proof. Since $X_n \leq X$ for all $n \geq 0$, we have $\int X_n \, d\mathbb{P} \leq \int X \, d\mathbb{P}$. Taking limit gives $\lim_{n\to\infty} \int X_n \, d\mathbb{P} \leq \int X \, d\mathbb{P}$.

To show the opposite direction, let Y be an arbitrary simple function satisfying $0 \leq Y \leq X$ and $\theta \in (0,1)$. Let $A_n = \{\theta Y \leq X_n\}$. By the definition of θ and Y, we have $\mathbb{P}(A_n) \to 1$ as $n \to \infty$. Note that

$$\theta \int_{A_n} Y\, d\mathbb{P} = \int_{A_n} \theta Y\, d\mathbb{P} \le \int_{A_n} X_n\, d\mathbb{P} \le \lim_{n\to\infty} \int X_n\, d\mathbb{P}\,. \tag{3.10}$$

Since $\mathbb{P}(A_n) \to 1$, θ and Y can be chosen arbitrarily close to 1 and X, respectively, the left hand side of (3.10) can be taken arbitrarily close to $\int X\, d\mathbb{P}$. This completes the proof. □

Theorem 3.3. (Fatou's Lemma) *If $X_n \ge 0$ a.s., then*

$$\int \liminf_{n\to\infty} X_n\, d\mathbb{P} \le \liminf_{n\to\infty} \int X_n\, d\mathbb{P}\,.$$

Proof. Let $Y_n = \inf_{k\ge n} X_k$. Note that Y_n is a monotone increasing function satisfying $Y_n \le X_n$ and $\lim_{n\to\infty} Y_n = \liminf_{n\to\infty} X_n$. Hence, we have

$$\int \liminf_{n\to\infty} X_n\, d\mathbb{P} = \int \lim_{n\to\infty} Y_n\, d\mathbb{P} \overset{\text{MCT}}{=} \lim_{n\to\infty} \int Y_n\, d\mathbb{P} = \liminf_{n\to\infty} \int Y_n\, d\mathbb{P} \le \liminf_{n\to\infty} \int X_n\, d\mathbb{P}\,.$$

□

Remark 3.2. **(Straight inequality of Fatou's Lemma)** Recall Example 3.9, $X_n(\omega) = n \times 1_{\{\omega<1/n\}}$ on $([0,1], \mathscr{B}_{[0,1]}, \lambda_{[0,1]})$. We have $X_n \overset{a.s.}{\to} 0$ and thus $\int \liminf_{n\to\infty} X_n\, d\mathbb{P} = \int 0\, d\lambda_{[0,1]} = 0$. However, for every integer n, $\int X_n\, d\mathbb{P} = n\lambda_{[0,1]}\left([0, \frac{1}{n})\right) = 1$. As a result, we have

$$0 = \int \liminf_{n\to\infty} X_n\, d\mathbb{P} < \liminf_{n\to\infty} \int X_n\, d\mathbb{P} = 1\,,$$

which is a straight inequality of Fatou's Lemma. □

Theorem 3.4. (Dominated Convergence Theorem (DCT)) *If $\lim_{n\to\infty} X_n = X$ a.s. and $|X_n| < Y$ where $\int Y\, d\mathbb{P} < \infty$, then*

$$\lim_{n\to\infty} \int X_n\, d\mathbb{P} = \int X\, d\mathbb{P} = \int \lim_{n\to\infty} X_n\, d\mathbb{P}\,.$$

Proof. Writing $\int X_n\, d\mathbb{P} = \int X_n^+\, d\mathbb{P} - \int X_n^-\, d\mathbb{P}$, where $X_n^+ = \max(X_n, 0) \ge 0$ and $X_n^- = -\min(X_n, 0) \ge 0$, we can assume without loss of generality that $X_n \ge 0$. By Fatou's Lemma, we have

$$\int \liminf_{n\to\infty} X_n\, d\mathbb{P} \le \liminf_{n\to\infty} \int X_n\, d\mathbb{P}\,. \tag{3.11}$$

Applying Fatou's Lemma again to $Y - X_n$ (≥ 0), we have

$$\begin{aligned}
& \int \liminf_{n\to\infty} Y - X_n\, d\mathbb{P} \le \liminf_{n\to\infty} \int Y - X_n\, d\mathbb{P} \\
\Rightarrow \quad & \int Y\, d\mathbb{P} + \int \liminf_{n\to\infty} (-X_n)\, d\mathbb{P} \le \int Y\, d\mathbb{P} + \liminf_{n\to\infty} \left(- \int X_n\, d\mathbb{P} \right) \\
\Rightarrow \quad & \limsup_{n\to\infty} \int X_n\, d\mathbb{P} \le \int \limsup_{n\to\infty} X_n\, d\mathbb{P}\,,
\end{aligned} \tag{3.12}$$

where the last inequality follows from $\liminf_{n\to\infty}(-X_n) = -\limsup_{n\to\infty} X_n$. Combining (3.11) and (3.12), we have

$$\liminf_{n\to\infty} \int X_n \, d\mathbb{P} \geq \int \liminf_{n\to\infty} X_n \, d\mathbb{P} = \int X \, d\mathbb{P} = \int \limsup_{n\to\infty} X_n \, d\mathbb{P} \geq \limsup_{n\to\infty} \int X_n \, d\mathbb{P}. \tag{3.13}$$

Since $\limsup_{n\to\infty} \int X_n \, d\mathbb{P} \geq \liminf_{n\to\infty} \int X_n \, d\mathbb{P}$, all terms in (3.13) are actually equal. Hence, $\lim_{n\to\infty} \int X_n \, d\mathbb{P} = \int X \, d\mathbb{P}$. □

Example 3.10. For any non-negative r.v. X on $(\Omega, \mathscr{F})$, there exists a sequence of discrete random variables X_n on $(\Omega, \mathscr{F})$ such that $X_n \uparrow X$ a.s.. For example, we can take

$$X_n = X_n(\omega) = \sum_{k=0}^{2^{2n}} \frac{k}{2^n} \cdot 1_{X^{-1}([\frac{k}{2^n}, \frac{k+1}{2^n}))}(\omega).$$

From the definition of X_n, it can be seen that for any $\omega \in \Omega$, $X_n(\omega) \uparrow X(\omega)$. Also, from the MCT, we have $E(X_n) = \int X_n \, d\mathbb{P} \to \int X \, d\mathbb{P} = E(X)$.

Recall that in Section 2.6, the expectation is defined by the limits of the upper and lower $\mathbb{P}$-sums over all possible partitions P. This example gives explicitly one sequence of partitions $\{P_n\}_{n\geq 1}$, where $P_n = \{0, \frac{1}{2^n}, \ldots, 2^n - \frac{1}{2^n}, 2^n\}$, that achieves the limit of the lower $\mathbb{P}$-sum as $n \to \infty$. Note that the small partition size of $\frac{1}{2^n}$ is necessarily for P_n to be nested, which ensures that X_n is strictly increasing.

□

Example 3.11. Given $f_n(x) = \frac{\sqrt{x}}{1+nx^3}$, DCT can be used to find

$$\lim_{n\to\infty} \int_1^\infty f_n(x) \, dx.$$

To see this, set $X_n : [1, \infty) \to \mathbb{R}^+$ on $([1, \infty), \mathscr{B}_{[1,\infty)}, \lambda)$ such that $X_n(\omega) = f_n(\omega)$. By construction we have $E(X_n) = \int_1^\infty X_n(\omega) \, \lambda(d\omega) = \int_1^\infty f_n(x) \, dx$ and $X_n \overset{a.s.}{\to} 0$.

Next we find the bound for X_n to apply DCT. Set $M : [1, \infty) \to \mathbb{R}^+$ such that $M(\omega) = \frac{1}{\omega^{2.5}}$. Note that $X_n(\omega) < \frac{\sqrt{\omega}}{\omega^3} = M(\omega)$ for all $n \geq 1$. Also, $M(\omega)$ satisfies $E(M) \triangleq \int_1^\infty M(\omega) \, d\lambda(\omega) = \frac{1}{1.5} < \infty$. Therefore, applying DCT gives

$$\lim_{n\to\infty} \int_1^\infty f_n(x) \, dx = \lim_{n\to\infty} E(X_n) = E\left(\lim_{n\to\infty} X_n\right) = E(0) = 0.$$

□

3.5 Convergence of sets

In this section we collect two important results related to convergence of sets.

Theorem 3.5. *Consider the probability space* $(\Omega, \mathscr{F}, \mathbb{P})$.

i) If $\{A_n\}$ *is an increasing sequence of sets (i.e.,* $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \subseteq \cdots$*) in* $\mathscr{F}$*, then*

$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \lim_{n\to\infty} \mathbb{P}(A_n). \tag{3.14}$$

ii) If $\{A_n\}$ *is a decreasing sequence of sets (i.e.,* $A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n \supseteq \cdots$*) in* $\mathscr{F}$*, then*

$$\mathbb{P}\left(\bigcap_{j=1}^{\infty} A_j\right) = \lim_{n\to\infty} \mathbb{P}(A_n). \tag{3.15}$$

Remark 3.3. For a fixed n, $\bigcup_{j=1}^{n} A_j = A_n$ since $\{A_n\}$ is increasing. Thus, it is clear that $\mathbb{P}\left(\bigcup_{j=1}^{n} A_j\right) = \mathbb{P}(A_n)$. Therefore, (3.14) can be restated as: $\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{j=1}^{n} A_j\right)$. Thus, Theorem 3.5 ensures the exchange between measure and limit *when the sequence of sets is monotonic increasing or decreasing*.

Proof. 1. Denote $A_0 = \emptyset$. Using the fact that $\{A_j \backslash A_{j-1}\}_{j=1}^{\infty}$ is a sequence of disjoint sets and $\bigcup_{j=1}^{\infty} A_j = \bigcup_{j=1}^{\infty} (A_j \backslash A_{j-1})$, we can obtain from countable additivity (Definition 2.6 (3)) that

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) &= \mathbb{P}\left(\bigcup_{j=1}^{\infty} (A_j \backslash A_{j-1})\right) \\
&= \sum_{j=1}^{\infty} \mathbb{P}\left(A_j \backslash A_{j-1}\right) && \text{(Definition 2.6 (3))} \\
&= \lim_{n\to\infty} \sum_{j=1}^{n} \mathbb{P}\left(A_j \backslash A_{j-1}\right) && \text{(Definition of limit)} \\
&= \lim_{n\to\infty} \sum_{j=1}^{n} \left(\mathbb{P}(A_j) - \mathbb{P}(A_{j-1})\right) \\
&= \lim_{n\to\infty} \mathbb{P}(A_n), && \text{(Telescopic sum)}
\end{aligned}$$

which completes the proof.

2. If $\{A_n\}$ is a decreasing sequence of sets, then $\{A_n^c\}$ is an increasing sequence of sets. Using (3.14) on the increasing $\{A_n^c\}$, we have

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{j=1}^{\infty} A_j\right) &= 1-\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j^c\right) \\
&= 1-\lim_{n\to\infty}\mathbb{P}\left(A_n^c\right) \qquad\qquad \text{(Using (3.14))} \\
&= 1-\lim_{n\to\infty}\left(1-\mathbb{P}(A_n)\right) \\
&= \lim_{n\to\infty}\mathbb{P}(A_n)\,,
\end{aligned}$$

which completes the proof.

□

Example 3.12. Consider a sequence of sets $\{B_n\}_{n=1}^{\infty}$. Since $A_n \triangleq \bigcup_{k=1}^{n} B_k$ is increasing in n, Theorem 3.5(i) implies

$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} B_j\right) = \lim_{n\to\infty}\mathbb{P}\left(\bigcup_{j=1}^{n} B_j\right).$$

Similarly, $\tilde{A}_n \triangleq \bigcap_{k=1}^{n} B_k$ is decreasing in n, Theorem 3.5(ii) implies

$$\mathbb{P}\left(\bigcap_{j=1}^{\infty} B_j\right) = \lim_{n\to\infty}\mathbb{P}\left(\bigcap_{j=1}^{n} B_j\right). \tag{3.16}$$

□

Theorem 3.6. (Borel Cantelli Lemma) *Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of events.*

i) If $\sum_{n=1}^{\infty}\mathbb{P}(A_n) < \infty$, then $\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0$.
ii) If $\{A_n\}_{n=1}^{\infty}$ are independent $\left(\text{ i.e., } \mathbb{P}\left(\bigcap_{j\in S} A_j\right) = \prod_{j\in S}\mathbb{P}(A_j) \text{ and } \mathbb{P}\left(\bigcap_{j\in S} A_j^c\right) = \prod_{j\in S}\mathbb{P}(A_j^c) \text{ for all } S\subset\mathbb{Z}^+ \right)$ and $\sum_{n=1}^{\infty}\mathbb{P}(A_n) = \infty$, then $\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 1$.

Proof. i) First note that for each k,

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = \mathbb{P}\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} A_n\right) \le \mathbb{P}\left(\bigcup_{n=k}^{\infty} A_n\right) \le \sum_{n=k}^{\infty}\mathbb{P}\left(A_n\right).$$

Since k is arbitrary and $\sum_{n=1}^{\infty}\mathbb{P}\left(A_n\right) < \infty$ implies that $\sum_{n=k}^{\infty}\mathbb{P}\left(A_n\right) \to 0$ as $k\to\infty$, the result follows.

ii) Using $1-t \le e^{-t}$, for each $k\in\mathbb{Z}^+$,

$$\mathbb{P}\left(\bigcap_{n=k}^{K} A_n^c\right) = \prod_{n=k}^{K}\left(1-\mathbb{P}(A_n)\right) \le \prod_{n=k}^{K} e^{-\mathbb{P}(A_n)} = e^{-\sum_{n=k}^{K}\mathbb{P}(A_n)} \to 0\,,$$

as $K\to\infty$. Thus, (3.16) implies $\mathbb{P}\left(\bigcap_{n=k}^{\infty} A_n^c\right) = \lim_{K\to\infty}\mathbb{P}\left(\bigcap_{n=k}^{K} A_n^c\right) = 0$. By countable sub-additivity of probability measure, we have

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}A_n^c\right) \le \sum_{k=1}^{\infty}\mathbb{P}\left(\bigcap_{n=k}^{\infty}A_n^c\right) = \sum_{k=1}^{\infty}0 = 0\,.$$

Taking complement, we obtain $\mathbb{P}(\limsup_{n\to\infty}A_n) = \mathbb{P}\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}A_n\right) = 1$.

Example 3.13. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of Bernoulli random variables satisfying $\mathbb{P}(X_n = 1) = \frac{1}{n^2}$ and $\mathbb{P}(X_n = 0) = 1 - \frac{1}{n^2}$. Denote $A_n = \{X_n = 1\}$, we have $\sum_{n=1}^{\infty}\mathbb{P}(A_n) = \sum_{n=1}^{\infty}\frac{1}{n^2} < \infty$. Then $\{X_n = 1 \text{ infinitely often}\} = \bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\{X_n = 1\}$ has probability 0 by Borel Cantelli Lemma (i).

Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of **independent** Bernoulli random variables satisfying $\mathbb{P}(X_n = 1) = \frac{1}{n}$ and $\mathbb{P}(X_n = 0) = 1 - \frac{1}{n}$. Denote $A_n = \{X_n = 1\}$, we have $\sum_{n=1}^{\infty}\mathbb{P}(A_n) = \sum_{n=1}^{\infty}\frac{1}{n} = \infty$. Then $\{X_n = 1 \text{ infinitely often}\} = \bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\{X_n = 1\}$ has probability 1 by Borel Cantelli Lemma (ii). Note that the independence assumption is important to apply Borel Cantelli Lemma (ii).

3.6 Cauchy Criterion

Suppose that we wish to check whether a sequence $\{x_n\}_{n=1}^{\infty}$ converges. In some simple cases (e.g. $x_n = \frac{1}{n}$, $x_n = \sqrt{n}$) we are able to guess the limits of the sequences. If it is not easy to guess the limit, then one standard procedure is to check for monotonicity. If a sequence is monotone and bounded, then it converges. A well known example is $x_n = \left(1 + \frac{1}{n}\right)^n$, which can be shown to be monotonic increasing and is bounded above by 3 (Indeed $x_n \to e$). However, this method is restricted to *real-valued* monotonic sequences. In practice, we often encounter non-monotonic sequence of functions or random variables. For example, in Chapter 7 we approximate a stochastic integral $\int_0^1 f(t)dW_t$ with a sequence of random variables, $\{I(f_n)\}_{n=1}^{\infty}$, where $I(f_n) = \sum_{k=1}^{n} f_n\left(\frac{k}{n}\right)\left(W_{\frac{k+1}{n}} - W_{\frac{k}{n}}\right)$, f_n is some function and W_t is a Brownian motion. In this case, the **Cauchy Criterion** provides an alternative way to check for convergence.

Definition 3.5. (Cauchy sequence) A sequence $\{x_n\}_{n=1}^{\infty}$ is called a *Cauchy sequence* if for any $\varepsilon > 0$, there exists an $N \in \mathbb{Z}^+$ such that for all $n,m > N$, the distance between x_n and x_m is bounded by ε, i.e.,

$$d(x_n, x_m) < \varepsilon\,,$$

where d is a metric between two points.

Theorem 3.7. (Cauchy Criterion)

i) Every convergent sequence is a Cauchy sequence.
ii) Every Cauchy sequence converges.

Proof. 1. If $\{x_n\}_{n=1}^{\infty}$ converges, then there exist an α and $N > 0$ such that for all $\varepsilon > 0$, $d(x_n, \alpha) \le \frac{\varepsilon}{2}$ for $n > N$. Therefore, for all $m,n > N$, by the triangle inequality,

$$d(x_m, x_n) \le d(x_m, \alpha) + d(\alpha, x_n) \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\,.$$

Thus, $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence.

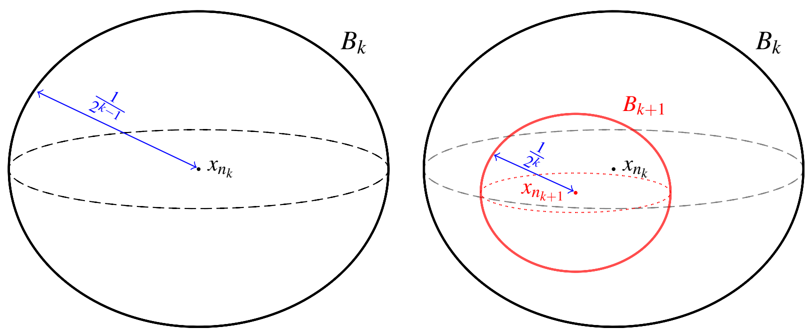
2. If $\{x_n\}_{n=1}^{\infty}$ is Cauchy, then we can find a subsequence $\{n_k\}_{k=1}^{\infty}$ such that

$$d(x_n, x_{n_k}) \le \frac{1}{2^k}\,. \tag{3.17}$$

for all $n > n_k$. Define $B_k = \left\{x : d(x, x_{n_k}) \le \frac{1}{2^{k-1}}\right\}$. Note from (3.17) and $\frac{1}{2^k} < \frac{1}{2^{k-1}}$ that $x_n \in B_k$ for all $n > n_k$. Also, for any $x \in B_{k+1}$, we have $d(x, x_{n_{k+1}}) \le \frac{1}{2^k}$. Combining with (3.17) yields

$$d(x, x_{n_k}) \le d(x, x_{n_{k+1}}) + d(x_{n_{k+1}}, x_{n_k}) \le \frac{1}{2^k} + \frac{1}{2^k} = \frac{1}{2^{k-1}}\,,$$

which implies that $x \in B_k$. Thus, we have shown that $B_{k+1} \subseteq B_k$. Since the argument holds for all $k \ge 1$, we have obtained a nested sequences of closed balls $B_1 \supseteq B_2 \supseteq \cdots \supseteq B_k \supseteq \cdots$. Thus, we can pick an element ℓ which is in B_k for all $k \ge 1$. As the radius of B_k goes to zero as $k \to \infty$ and B_k contains all x_n with $n > n_k$, we have $d(x_n, \ell) < \frac{1}{2^{k-1}}$ for all $n > n_k$. In other words, x_n converges to ℓ as $n \to \infty$. □



Example 3.14. Although we have shown that a Cauchy sequence x_n converges, it may not converge to an element in the space of x_n. For example, let $\{x_n\}_{n=1}^{\infty}$ be a sequence of rational numbers defined by rounding $\sqrt{2}$ to the nearest n-th decimal place. That is,

$$x_1 = 1.4\,,\; x_2 = 1.41\,,\; x_3 = 1.414\,,\; x_4 = 1.4142\,,\; \ldots\,.$$

Clearly, $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence since $|x_n - x_m| \le 10^{-k}$ for $n, m \ge k$. However $x_n \in \mathbb{Q}$ for all n but $x_n \to \sqrt{2} \notin \mathbb{Q}$.

In fact, whether a Cauchy sequence converges to an element in the space of x_n depends heavily on the nature of the space. It can be shown (see books in functional analysis) that $\mathbb{R}$, $\mathbb{C}$, $C[a,b]$ (continuous function on a bounded interval $[a,b]$), $\mathscr{L}^p$ ($p \geq 1$) have this desirable property. They are called **complete** spaces. □

Example 3.15. Consider the sequence $\{X_n\}_{n=1}^\infty$ in Example 3.5. As the support of X_n, i.e., $\{\omega : X_n(\omega) > 0\}$, is "moving around" as n grows, for each $\omega \in [0,1]$, there exist infinitely many ns such that $X_n(\omega) = 0$ and $X_{n+1}(\omega) = 1$. Thus $\{X_n\}_{n=1}^\infty$ is not Cauchy in d_ω and d_{sup} distance.

On the other hand, let α_n be the length of the support of X_n. Observe that $E(X_n) = E(X_n^2) = \mathbb{P}(X_n = 1) = \alpha_n$ and $\alpha_n \downarrow 0$ as $n \to \infty$. Thus, for each fixed $\varepsilon > 0$, we can find $2\alpha_N < \varepsilon$ such that for all $n,m > N$, $E(|X_n - X_m|) \leq E(|X_n|) + E(|X_m|) = \alpha_n + \alpha_m \leq 2\alpha_N < \varepsilon$. Thus $\{X_n\}_{n=1}^\infty$ is Cauchy in d_1 distance. Similar calculations yield that $\{X_n\}_{n=1}^\infty$ is Cauchy in d_p ($p \geq 1$), d_{pr} and d_{df} distances. □

3.7 Exercises

Exercise 3.8 *Among the jth-component, maximal, absolute, squared and Euclidean distances of real numbers, which one(s) is (are) metric(s)?*

Exercise 3.9 *For two subsets X,Y of an abstract space Ω, show that $X = Y$ if and only if $X \triangle Y = \emptyset$.*

Exercise 3.10 *Investigate the convergence of*

$$\int_a^\infty \frac{n^2 x e^{-n^2x^2}}{1+x^2}\,dx,$$

as $n \to \infty$, for $a > 0$ and $a = 0$.

Exercise 3.11 *i) Consider $X_n(\omega) = n \times 1_{\{0 \leq \omega \leq 1/n\}}$ on $(\Omega, \mathscr{F}, \mathbb{P}) = ([0,1], \mathscr{B}_{[0,1]}, \lambda)$. Find the c.d.f. of X_n.*

ii) Consider $X_n(\omega) = \sqrt{n} \times 1_{\{0 \leq \omega \leq 1/n\}}$ on $(\Omega, \mathscr{F}, \mathbb{P}) = ([0,1], \mathscr{B}_{[0,1]}, \lambda)$. Find the c.d.f. of X_n.

iii) Consider $X_n(\omega) = n \times 1_{\{0 \leq \omega \leq 1/n^2\}}$ on $(\Omega, \mathscr{F}, \mathbb{P}) = ([0,1], \mathscr{B}_{[0,1]}, \lambda)$. Find the c.d.f. of X_n.

For each of $i), ii)$ and $iii)$, investigate different modes of convergence of $\{X_n\}$.

Exercise 3.12 *Show that $E|X| < \infty$ implies $x\mathbb{P}(|X| > x) \to 0$ as $x \to \infty$.*

Exercise 3.13 *Consider $X_n = 1_{\{\omega \geq 0.5+1/n\}}$ and $X = 1_{\{\omega \geq 0.5\}}$ on $(\Omega, \mathscr{F}, \mathbb{P}) = ([0,1], \mathscr{B}_{[0,1]}, \lambda)$. Investigate the four modes of convergence for $\{X_n\}$. Does the distribution function of X_n converges to that of X at the discontinuity points?*

Exercise 3.14 *If $\{X_n\}$ is a sequence of random variables, then there exist $a_n \to \infty$ such that $\frac{X_n}{a_n} \xrightarrow{a.s.} 0$.*

Exercise 3.15 *Let $\{X_n\}$ be i.i.d. exponential distributed, i.e., $\mathbb{P}(X_i > x) = e^{-x}$. Let $M_n = \max_{1 \leq m \leq n} X_m$. Show that $\limsup_{n\to\infty} \frac{X_n}{\log n} = 1$ and $\frac{M_n}{\log n} = 1$ a.s.*

Exercise 3.16 *Consider $X_n(\omega) = n \times 1_{\{0 \le \omega \le 1/n\}}$ on $(\Omega, \mathscr{F}, \mathbb{P}) = ([0,1], \mathscr{B}_{[0,1]}, \lambda)$. Check whether $\{X_n\}_{n=1}^{\infty}$ is Cauchy in each measure of distance of random variables.*

Exercise 3.17 *i) For the space $\mathbb{R}$, find a sequence of open balls $\{B_n\}$ such that $B_1 \supseteq B_2 \supseteq \cdots \supseteq B_k \supseteq \cdots$ with $\bigcap_{n=1}^{\infty} B_n = \emptyset$.*

ii) Find a sequence of closed balls $\{B_n\}$ in $\mathbb{R}$ such that $B_1 \supseteq B_2 \supseteq \cdots \supseteq B_k \supseteq \cdots$ and $\bigcap_{n=1}^{\infty} B_n$ does not contain any rational number.

Chapter 4
Conditional Expectations

When we work on a particular branch of multi-period Binomial model, conditional probability is involved: conditioning on a node, we consider the probability of going up or down. Given a continuous time price process $\{S_t\}_{t\geq 0}$, if at time u we want to find a product's price at maturity T, we need to deal with the conditional distribution of S_T given S_u. In this section we provide a rigorous introduction to conditional probabilities and conditional expectations.

4.1 Conditional Probability and Independence

Definition 4.1. (Conditional Probability) For any events $A, B \in \mathscr{F}$ such that $\mathbb{P}(B) \neq 0$, the **conditional probability** of A given B is defined by

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}\,. \tag{4.1}$$

Example 4.1. **(Total Probability Formula)** For any event $A \in \mathscr{F}$ and any partition $B_1, B_2, \ldots$ of Ω such that $\mathbb{P}(B_i) \neq 0$ and $B_i \in \mathscr{F}$ for any i,

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{\infty} \mathbb{P}(A|B_i)\mathbb{P}(B_i)\,,$$

where the first and second equalities follow from the countable additivity of probability measure and (4.1) respectively.

Definition 4.2. (Independence of events) Two events $A, B \in \mathscr{F}$ are called **independent** if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Any n events $A_1, A_2, \ldots, A_n \in \mathscr{F}$ are **independent** if

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_k})$$

for any indices $1 \leq i_1 < i_2 < \cdots < i_k \leq n$. In general, an infinite number of events are **independent** if any finite subset of them are independent.

Definition 4.3. (Independence of random variables) A set of n random variables $X_1,\dots,X_n$ are **independent** if for any Borel sets $B_1,\dots,B_n \in \mathscr{B}$, the events

$$\{X_1 \in B_1\},\dots,\{X_n \in B_n\}$$

are independent. In general, an *infinite* number of random variables are **independent** if any finite subset of them are independent.

Definition 4.4. (Independence of σ-fields) A set of n σ-fields $\mathscr{G}_1,\dots,\mathscr{G}_n$ contained in $\mathscr{F}$ are **independent** if any n events $A_1 \in \mathscr{G}_1, A_2 \in \mathscr{G}_2,\dots,A_n \in \mathscr{G}_n$ are independent. In general, an infinite number of σ-fields is **independent** if any finite subset of them are independent.

Remark 4.1. i) The σ fields $\mathscr{G}_i$s must be contained in $\mathscr{F}$, i.e., $\mathscr{G}_i \subset \mathscr{F}$ for all i. Otherwise, the measure $\mathbb{P}$ is not able to assign probabilities to some $A_i \in \mathscr{G}_i$.

ii) For $B \in \mathscr{B}$, since $\{X \in B\}$ is the element of the σ-field generated by the random variable X, independence of random variables is equivalent to independence of σ-fields generated by the random variables.

Example 4.2. Consider the dice example $(\Omega,\mathscr{F},\mathbb{P})$ where $\Omega = \{1,2,3,4,5,6\}$ and $\mathscr{F} = 2^{\Omega}$. Let $\mathbb{P}(\{\omega\}) = 1/6$ for $\omega = 1,\dots,6$ and

$$A_1 = \{1,2,3\},\ A_2 = \{2,5\},\ X_i = 1_{A_i},\ \mathscr{F}_i = \sigma(\{A_i\}) \text{ for } i = 1,2\,.$$

First consider independence of events: $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2) = \frac{1}{6}$, thus A_1 and A_2 are independent.

Next consider independence of r.v.s. From Remark 4.1 we need to show that the σ-field generated by X_1 and X_2 are independent. With some elementary calculations, the σ-fields generated by X_1 and X_2 are found to be $\{\varnothing, A_1, \{4,5,6\}, \Omega\}$ and $\{\varnothing, A_2, \{1,3,4,6\}, \Omega\}$, respectively. We can conclude that X_1 and X_2 are independent if each pair of these two groups of events are independent (4×4=16 combinations). By elementary but tedious calculations, these 16 pairs of events can be shown to be independent. Thus, X_1 and X_2 are independent.

Finally, to show that $\mathscr{F}_1$ and $\mathscr{F}_2$ are independent, we need to show that each combination of events from $\mathscr{F}_1$ and $\mathscr{F}_2$ are independent. Notice that $\mathscr{F}_i$ coincides with the σ-field generated by X_i for $i = 1,2$, respectively. Thus, the calculations in the preceding paragraph show that $\mathscr{F}_1$ and $\mathscr{F}_2$ are indeed independent. □

Example 4.3. If two σ-fields $\mathscr{F}_a$ and $\mathscr{F}_b$ are independent, then any $\mathscr{F}_a$ measurable r.v. is independent of any $\mathscr{F}_b$ measurable r.v..

To see this, by Definition 4.3, we have to show that for any $\mathscr{F}_a$ measurable r.v. X_a and any $\mathscr{F}_b$ measurable r.v. X_b, the events $\{X_a \in B_a\}$ and $\{X_b \in B_b\}$ are independent, where $B_a, B_b \in \mathscr{B}$. Since $\{X_a \in B_a\} = X_a^{-1}(B_a) \in \mathscr{F}_a$ and $\{X_b \in B_b\} = X_b^{-1}(B_b) \in \mathscr{F}_b$ by the definition of r.v., putting $A_1 = X_a^{-1}(B_a)$ and $A_2 = X_b^{-1}(B_b)$ in Definition 4.4 completes the proof. □

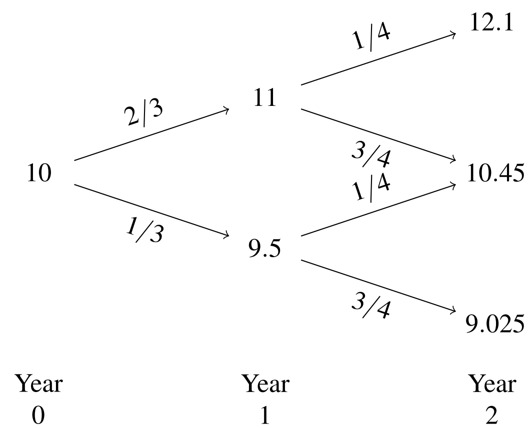
4.2 Conditional Expectation

4.2.1 Conditioning on an Event

Definition 4.5. (Conditional Expectation given an Event) For any random variable $X \in \mathscr{L}^1$ on $(\Omega, \mathscr{F}, \mathbb{P})$ and any event $B \in \mathscr{F}$ such that $\mathbb{P}(B) \neq 0$, the **conditional expectation** of X given B is defined by

$$E(X|B) = \frac{1}{\mathbb{P}(B)} \int_B X d\mathbb{P}. \tag{4.2}$$

Example 4.4. An investor plans to invest \$10 in an asset. According to his estimate, the asset value in the coming two years can be represented by the following diagram



The number along an arrow represents the probability that the event indicated by the arrow would occur. Let X be the value of the asset at the end of the second year. Recall the definition of **path** in Definition 1.5, which is $\omega_t = (z_1, \ldots, z_t)$, $z_i = \pm 1$ for $i = 1, \ldots, t$, indicates the evolution (up=1, down=-1) of the market. We can take the probability space to be $(\Omega_X, \mathscr{F}_X, \mathbb{P}_X)$ where $\Omega_X = \{\omega_2 = (a,b) | a = \pm 1, b = \pm 1\}$, $\mathscr{F}_X = 2^{\Omega_X}$ and $\mathbb{P}_X$ is the discrete probability measure satisfying $\mathbb{P}_X(\{(1,1)\}) = 1/6, \ldots, \mathbb{P}_X(\{(-1,-1)\}) = 1/4$.

Consider the event $B = \{\omega_1 = (1)\}$. Clearly, $\mathbb{P}(B) = 2/3$. Also, note that

$$X((1,1)) = 12.1\ , \quad X((1,-1)) = 10.45\,,$$

and

$$\mathbb{P}(\{(1,1)\}) = \frac{1}{6}\ ,\ \mathbb{P}(\{(1,-1)\}) = \frac{1}{2}\,.$$

Therefore

$$E(X|B) = \frac{1}{\mathbb{P}(B)} \int_B X d\mathbb{P} = \frac{1}{\mathbb{P}(B)} \left[(X((1,1))\mathbb{P}(\{(1,1)\}) + X((1,-1))\mathbb{P}(\{(1,-1)\}) \right]$$
$$= \frac{1}{2/3} \left(12.1 \times \frac{1}{6} + 10.45 \times \frac{1}{2} \right) = 10.8625.$$

□

4.2.2 Conditioning on a discrete random variable

Given an arbitrary random variable X in $\mathscr{L}^1$ (so expectation exists) and a **discrete random variable** $Z : \Omega \to \{z_1, \dots\}$, the conditional expectation of a r.v. X given Z, i.e. $E(X|Z)$, must depend solely on the random variable Z. Thus $E(X|Z)$ is itself a random variable. Since Z takes only m possible values, so does $E(X|Z)$. Therefore, we can characterize $E(X|Z)$ by m conditional expectations on events $\{Z = z_i\}$, $E(X|\{Z = z_i\})$, $i = 1, \dots, m$, using Definition 4.5.

Definition 4.6. (Conditional Expectation given a discrete r.v.) Let $X \in \mathscr{L}^1$ and Z be a discrete r.v. taking values on $\{z_i\}_{i=1,\dots}$. The **conditional expectation of** X **given** Z is defined as a discrete random variable $E(X|Z) : \Omega \to \mathbb{R}$ such that

$$E(X|Z)(\omega) = E(X|\{Z = z_i\}) \quad \text{on} \quad \{\omega : Z(\omega) = z_i\}, \tag{4.3}$$

for any $i = 1, 2, \dots$ □

Example 4.5. Three coins, 10, 20 and 50 cents, are tossed. The values of those coins with heads shown up are added to a total amount X. Let Z be the total amount of the 10- and 20-cent coins with head shown up. What is $E(X|Z)$?

First, the probability space can be taken as $\Omega = \{HHH, HHT, \dots, TTT\}$, where H and T stand for Head and Tail respectively. The associated σ-field can be taken as $\mathscr{F} = 2^{\Omega}$. The probability measure $\mathbb{P}$ can be defined on $\mathscr{F}$ with $\mathbb{P}(\{HHH\}) = \cdots = \mathbb{P}(\{TTT\}) = \frac{1}{8}$. The discrete r.v. Z on $(\Omega, \mathscr{F}, \mathbb{P})$ is given by $Z : \Omega \to \{0, 10, 20, 30\}$. Thus, finding $E(X|Z)$ is equivalent to finding $E(X|\{Z = 0\}), \dots, E(X|\{Z = 30\})$. For example,

$$E(X|\{Z = 0\}) = \frac{1}{\mathbb{P}(\{Z = 0\})} \int_{\{Z=0\}} X \, d\mathbb{P} = \frac{1}{1/4} \left((0 + 50)\frac{1}{8} + (0 + 0)\frac{1}{8} \right) = 25\,.$$

Using similar arguments for the other cases, we have

$$E(X|Z)(\omega) = \begin{cases} 25 & \text{if } Z(\omega) = 0\,, \\ 35 & \text{if } Z(\omega) = 10\,, \\ 45 & \text{if } Z(\omega) = 20\,, \\ 55 & \text{if } Z(\omega) = 30\,. \end{cases} \tag{4.4}$$

As we consider X conditioned on Z, the probability measure on Z is irrelevant. In particular, the above results still hold even if the 10- and 20-cent coins are biased. However, the results will be different if the 50-cent coin is assumed to be biased. $\square$

The following theorem gives some important properties of conditional expectation.

Theorem 4.1. *If $X \in \mathscr{L}^1$ and Z is a discrete random variable, then*

i) $E(X|Z)$ is $\sigma(Z)$ measurable ;
ii) For any $A \in \sigma(Z)$,

$$\int_A E(X|Z)\, d\mathbb{P} = \int_A X\, d\mathbb{P}. \tag{4.5}$$

Proof. i) To show that $E(X|Z)$ is $\sigma(Z)$ measurable, we have to show that for any $a \in \mathbb{R}$, $\{\omega : E(X|Z)(\omega) > a\} \in \sigma(Z)$ (see Theorem 2.3). Since Z is discrete, we can assume that Z takes values $\{z_i\}_{i=1,\ldots}$. From (4.3), $E(X|Z)$ also takes discrete values $\{y_i\}_{i=1,\ldots}$, where $y_i = E(X|\{Z = z_i\})$. Note that $\{\omega : E(X|Z)(\omega) = y_i\} = \{Z = z_i\}$. It follows that for any $a \in \mathbb{R}$, $\{\omega : E(X|Z)(\omega) > a\}$ is a disjoint union of some $\{Z = z_i\}$s, which belongs to $\sigma(Z)$. This completes the proof of i).

ii) Since $E(X|Z)(\omega)$ is constant y_i on $\{\omega : Z(\omega) = z_i\}$, we have

$$\begin{aligned}
\int_{\{Z=z_i\}} E(X|Z)d\mathbb{P} &= \int_{\{Z=z_i\}} E(X|\{Z = z_i\})d\mathbb{P} \\
&= E(X|\{Z = z_i\}) \int_{\{Z=z_i\}} d\mathbb{P} \\
&\overset{(4.2)}{=} \frac{\int_{\{Z=z_i\}} X d\mathbb{P}}{\int_{\{Z=z_i\}} d\mathbb{P}} \int_{\{Z=z_i\}} d\mathbb{P} \\
&= \int_{\{Z=z_i\}} X d\mathbb{P}.
\end{aligned}$$

In general, any $A \in \sigma(Z)$ is a countable union of sets of the form $\{Z = z_i\}$, which are pairwise disjoint. Thus (4.5) follows from the countable additivity of Lebesgue integral.

Example 4.6. Recall in Example 4.5 that the induced sample spaces for the r.v. X and Z are $\Omega_X = \{0, 10, 20, 30, 50, 60, 70, 80\}$ and $\Omega_Z = \{0, 10, 20, 30\}$, respectively. Let $A_1 = \{Z = 20\}$ and $A_2 = \{Z = 10 \text{ or } 30\}$. Note that $A_1, A_2 \in \mathscr{F}_Z \triangleq 2^{\Omega_Z}$. It is easy to verify that (4.5) holds for $A = A_1$ and A_2. In particular,

$$\begin{aligned}
\int_{A_1} X\,d\mathbb{P} &= 20\mathbb{P}(\{THT\})+70\mathbb{P}(\{THH\}) = 20\frac{1}{8}+70\frac{1}{8} = 11.25\,,\\
\int_{A_1} E(X|Z)\,d\mathbb{P} &\overset{(4.4)}{=} 45\int_{A_1} d\mathbb{P} = 45(\mathbb{P}(\{THT\})+\mathbb{P}(\{THH\})) = 45\frac{1}{4} = 11.25\,,\\
\int_{A_2} X\,d\mathbb{P} &= 10\mathbb{P}(\{HTT\})+\cdots+80\mathbb{P}(\{HHH\}) = (10+60+30+80)\frac{1}{8} = 22.5\,,\\
\int_{A_2} E(X|Z)\,d\mathbb{P} &= \int_{\{Z=10\}} E(X|Z)\,d\mathbb{P} + \int_{\{Z=30\}} E(X|Z)\,d\mathbb{P} \overset{(4.4)}{=} (35+55)\frac{1}{4} = 22.5\,.
\end{aligned}$$

□

4.2.3 Conditioning on an arbitrary random variable

If Z is a discrete r.v., then we see from (4.3) that $E(X|Z)$ is also a discrete random variable. When Z is an arbitrary r.v. (has discrete and continuous components), it is in general not easy to obtain an explicit formula for $E(X|Z)$. (The elementary way of writing $E(X|Z) = \int x f_{X,Z}(x,z)/f_Z(z)\,dx$ cannot be used if the p.d.f.s $f_{X,Z}$ or $f_Z(z)$ does not exist.) However, Theorem 4.1 suggests a way for us to define conditional expectation given an arbitrary r.v..

Definition 4.7. (Conditional Expectation given an arbitrary r.v.) Let $X \in \mathscr{L}^1$ and Z be an arbitrary r.v.. Then the **conditional expectation** of X given Z is defined to be a random variable $E(X|Z)$ such that

i) $E(X|Z)$ is a $\sigma(Z)$ measurable r.v. ;
ii) For any $A \in \sigma(Z)$,

$$\int_A E(X|Z)\,d\mathbb{P} = \int_A X\,d\mathbb{P}. \tag{4.6}$$

Since conditional expectation is defined implicitly by (4.6) rather than an explicit formula, we need to justify such a definition by showing that the r.v. $E(X|Z)$ is unique, i.e., $X = X'$ a.s. implies $E(X|Z) = E(X'|Z)$ a.s. Since $E(X|Z)$ is a random variable, the uniqueness depends on the probability measure $\mathbb{P}$ through the notion of **almost sure equivalence**.

Definition 4.8. (Almost Sure equivalence)

1. Two events $A, B \in \mathscr{F}$ are equal almost surely (a.s.) if $\mathbb{P}((A\backslash B)\cup(B\backslash A)) = 0$, i.e., the uncommon region of A and B has probability 0.
2. Two random variables X and Y on $(\Omega, \mathscr{F}, \mathbb{P})$ are equal almost surely (a.s.) if $\mathbb{P}(X = Y) \triangleq \mathbb{P}(\{\omega \in \Omega : X(\omega) = Y(\omega)\}) = 1$.

Example 4.7. Note that the a.s. equivalence of two events A and B does not imply that the two events are the same. It only means that the events $A\backslash B$ and

$B \backslash A$ are of $\mathbb{P}$-measure 0. For example, consider $A = (0, 0.5)$ and $B = (0, 0.5]$ on $([0,1], \mathscr{B}_{[0,1]}, \lambda_{[0,1]})$. It is clear that $A = B$ a.s. but $A \neq B$.

For the case of random variables, let $\Omega = \{\omega_1, \omega_2\}$, $\mathscr{F} = 2^{\Omega}$, and $\mathbb{P}$ be defined on $\mathscr{F}$ with $\mathbb{P}(\{\omega_1\}) = 1$. Suppose that $X_i : \Omega \to \mathbb{R}$ are r.v.s with

$$X_1(\omega) = \mathbb{I}_{\{\omega=\omega_1\}} + 2 \times \mathbb{I}_{\{\omega=\omega_2\}} \quad \text{and} \quad X_2(\omega) = \mathbb{I}_{\{\omega=\omega_1\}} + 3 \times \mathbb{I}_{\{\omega=\omega_2\}} .$$

Then $X_1 = X_2$ a.s. but $X_1 \neq X_2$. □

Theorem 4.2. (Uniqueness of Conditional Expectation) *$E(X|Z)$ is unique in the sense that if $X = X'$ a.s., then $E(X|Z) = E(X'|Z)$ a.s.*

The proof of Theorem 4.2 relies on the following lemma.

Lemma 4.3 *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\sigma(Z) \subseteq \mathscr{F}$. If Y is a $\sigma(Z)$ measurable r.v. and*

$$\int_A Y \, d\mathbb{P} = 0 ,$$

for any $A \in \sigma(Z)$, then $Y = 0$ a.s. That is, $\mathbb{P}(Y = 0) = 1$.

Proof. Observe that $P\left(Y \geq \frac{1}{n}\right) = 0$ for any positive integer n because

$$0 \leq \frac{1}{n}\mathbb{P}\left(Y \geq \frac{1}{n}\right) = \int_{\{Y \geq \frac{1}{n}\}} \frac{1}{n} d\mathbb{P} \leq \int_{\{Y \geq \frac{1}{n}\}} Y d\mathbb{P} = 0$$

The last equality follows from the fact that $\{Y \geq \frac{1}{n}\} = \{Y \in [\frac{1}{n}, \infty)\} \in \sigma(Z)$. Similarly, $P\{Y \leq -\frac{1}{n}\} = 0$ for any $n > 0$. Setting $A_n = \{-\frac{1}{n} < Y < \frac{1}{n}\}$, we have

$$P(A_n) = 1$$

for any positive integer n. Since $\{A_n\}$ forms a decreasing sequence of events and

$$\{Y = 0\} = \bigcap_{n=1}^{\infty} A_n ,$$

it follows that

$$\mathbb{P}\{Y = 0\} = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{m\to\infty} \mathbb{P}\left(\bigcap_{n=1}^{m} A_n\right) = \lim_{m\to\infty} \mathbb{P}(A_m) = 1 , \tag{4.7}$$

as required. The third equality follows from Theorem 3.5 for a decreasing sequence of events.

Proof. **(Theorem 4.2.)** Suppose that $X = X'$ a.s., i.e., $\mathbb{P}(X = X') = 1$. We want to show that $E(X|Z) = E(X'|Z)$ a.s.. Let $Y = E(X|Z) - E(X'|Z)$. Note that by the definition of conditional expectation, $E(X|Z)$ and $E(X'|Z)$ are $\sigma(Z)$ measurable, so is Y. For any $A \in \sigma(Z)$, we have

$$
\begin{aligned}
\int_A Y\,d\mathbb{P} &= \int_A \big(E(X|Z) - E(X'|Z)\big)\,d\mathbb{P} \\
&= \int_A (X - X')\,d\mathbb{P} && \text{(by (4.6))} \\
&= 0 && \text{(Since } X = X' \text{ a.s..)}
\end{aligned}
$$

From Lemma 4.3, we have $Y = 0$ a.s., i.e., $E(X|Z) = E(X'|Z)$ a.s., completing the proof of Theorem 4.2. □

Example 4.8. Consider the probability space $([0,1], \mathscr{B}_{[0,1]}, \lambda_{[0,1]})$. Let

$$X(\omega) = 2\omega^2, \ \ Z(\omega) = \begin{cases} 2 & \text{if } \ \omega \in [0, \frac{1}{2}) \\ \omega & \text{if } \ \omega \in [\frac{1}{2}, 1] \end{cases} .$$

The following heuristic can be used to guess the form of $E(X|Z)$: First note that there is a *discrete mass* at $Z = 2$. Following the arguments used in conditioning on discrete r.v., we obtain that, on $\omega \in [0, \frac{1}{2})$,

$$
\begin{aligned}
E(X|Z)(\omega) = E(X|\{Z=2\}) &= \frac{1}{\mathbb{P}(\omega \in [0, \frac{1}{2}))} \int_{[0,\frac{1}{2})} X(\omega)\,d\mathbb{P} \\
&= \frac{1}{1/2} \int_0^{\frac{1}{2}} 2\omega^2\,d\omega = \frac{1}{6}\,. \qquad (4.8)
\end{aligned}
$$

For the continuous part $Z(\omega) = \omega$ on $\omega \in [1/2, 1]$, note that $\{Z = z\} \in \sigma(Z)$ and $\{Z = z\} = \{\omega : Z(\omega) = z\} = \{\omega = z\}$ for $z \in [1/2, 1]$. Thus, we can heuristically *solve* for $E(X|Z)$ from the definition by

$$
\begin{aligned}
& \textstyle\int_{\{Z=z\}} E(X|Z)\,d\mathbb{P} = \int_{\{Z=z\}} X\,d\mathbb{P} \\
\Rightarrow\ & E(X|\{Z=z\}) \textstyle\int_{\{Z=z\}} d\mathbb{P} = \int_{\{\omega : Z(\omega)=z\}} X(\omega)\,\mathbb{P}(d\omega) \\
\Rightarrow\ & E(X|\{Z=z\}) = X(z) \frac{\int_{\{\omega=z\}} \mathbb{P}(d\omega)}{\int_{\{\omega=z\}} \mathbb{P}(d\omega)} = X(z) = 2z^2\,.
\end{aligned} \qquad (4.9)
$$

Expressing $E(X|Z)$ as a function of ω (sample space) instead of a function of Z (induced sample space), we have the guess

$$E(X|Z)(\omega) = \begin{cases} \frac{1}{6} & \text{if } \ \omega \in [0, \frac{1}{2})\,, \\ 2\omega^2 & \text{if } \ \omega \in [\frac{1}{2}, 1]\,. \end{cases} \qquad (4.10)$$

To verify that $E(X|Z)$ is really a conditional expectation, we need to verify Conditions i) and ii) in Definition 4.7. First we show i): $E(X|Z)$ is $\sigma(Z)$ measurable. Recall that $\sigma(Z) = \{Z^{-1}(B), B \in \mathscr{B}\}$. For any Borel set $B \in \mathscr{B}$, from the definition of Z, we have

$$Z^{-1}(B) = \begin{cases} B \cap \left[\frac{1}{2}, 1\right] & \text{if } 2 \notin B\,, \\ \left(B \cap \left[\frac{1}{2}, 1\right]\right) \cup \left[0, \frac{1}{2}\right) & \text{if } 2 \in B\,. \end{cases} \qquad (4.11)$$

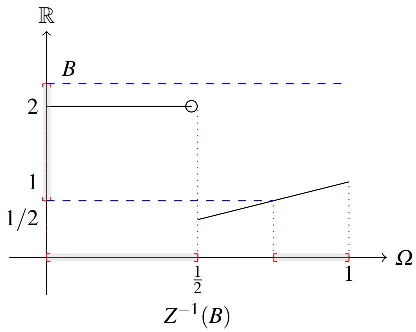


Fig. 4.1 Z

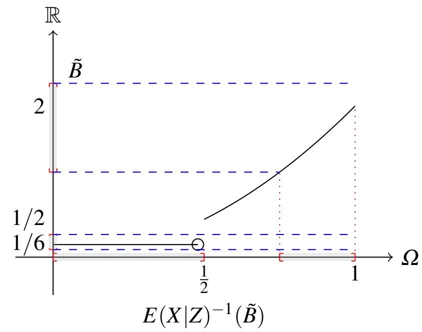


Fig. 4.2 $E(X|Z)(\omega)$

On the other hand, it can be checked that the inverse mapping of $\{\omega : E(X|Z)(\omega) \in B\}$, where $B \in \mathscr{B}$, also takes the same form as in (4.11). Hence, $\{\omega : E(X|Z)(\omega) \in B\} \subseteq \{Z^{-1}(B), B \in \mathscr{B}\} = \sigma(Z)$, i.e., $E(X|Z)$ is $\sigma(Z)$ measurable.

Finally, we verify ii) of Definition 4.7. From (4.11), we know that any $A \in \sigma(Z)$ is of the form $B \cap [\frac{1}{2}, 1]$ or $\big(B \cap [\frac{1}{2}, 1]\big) \cup [0, \frac{1}{2}]$, where $B \in \mathscr{B}$. Suppose first that $A = B \cap [\frac{1}{2}, 1]$. Then, by definition we have $\int_A X \, d\mathbb{P} = \int_{B \cap [\frac{1}{2}, 1]} 2\omega^2 \, d\omega$. On the other hand, (4.10) implies that

$$\int_A E(X|Z) \, d\mathbb{P} = \int_{B \cap [\frac{1}{2}, 1]} E(X|Z) \, d\mathbb{P} = \int_{B \cap [\frac{1}{2}, 1]} 2\omega^2 \, d\omega = \int_A X \, d\mathbb{P}. \tag{4.12}$$

Similarly, (4.12) holds for the case $A = \big(B \cap [\frac{1}{2}, 1]\big) \cup [0, \frac{1}{2}]$. Thus ii) of Definition 4.7 holds for all $A \in \sigma(Z)$. Hence, $E(X|Z)$ satisfies Definition 4.7, i.e., is a valid conditional expectation.

Example 4.9. **(Elementary definition of Conditional Expectation)** Recall that Definition 4.7 of conditional expectation $E(X|Z)$ does not require the existence of joint p.d.f. of X and Z. In this example we show that $E(X|Z)$ reduces to the elementary definition of conditional expectation when the joint p.d.f. of X and Z exists.

Suppose that the joint p.d.f. of random variables X and Z exists and is given by $f_{X,Z}(x,z)$. We can define the conditional p.d.f. by

$$f_{X|Z}(x|z) = f_{X,Z}(x,z)/f_Z(z) \quad \text{if} \quad f_Z(z) \neq 0,$$

and 0 otherwise, where $f_Z(z) = \int f_{X,Z}(x,z)\,dx$. Note that if the joint p.d.f. exists, then $d\mathbb{P} = f_{X,Z}(x,z)dxdz$. Following the same heuristic in Example 4.8, we *solve* for $E(X|Z)$ from the definition by

$$\begin{aligned} & \int_{\{Z=\mathfrak{z}\}} E(X|Z)\,d\mathbb{P} = \int_{\{Z=\mathfrak{z}\}} X\,d\mathbb{P} \\ \Rightarrow\ & E(X|\{Z=\mathfrak{z}\}) \iint 1_{\{z=\mathfrak{z}\}} f_{X,Z}(x,z)\,dxdz = \iint 1_{\{z=\mathfrak{z}\}} x f_{X,Z}(x,z)\,dxdz \\ \Rightarrow\ & E(X|\{Z=\mathfrak{z}\}) = \frac{\int x f_{X,Z}(x,\mathfrak{z})\,dx}{f_Z(\mathfrak{z})} = \int x f_{X|Z}(x|\mathfrak{z})\,dx =: g(\mathfrak{z}), \end{aligned} \tag{4.13}$$

say. Then $E(X|Z) \triangleq g(Z)$ is the conditional expectation of X given Z.

Next we justify Conditions i) and ii) in Definition 4.7. First, i) is trivial since $g(\mathfrak{z})$ is a function involving only $\mathfrak{z}$. Lastly we verify ii): for any $B \in \mathscr{B}$ and $A \triangleq \{\omega : Z(\omega) \in B\} \in \sigma(Z)$, we have

$$\int_A X \, d\mathbb{P} = \int \int \mathbb{I}_A(z) x f_{X,Z}(x,z) \, dxdz \, ,$$

and

$$\begin{aligned}\int_A g(Z) \, d\mathbb{P} &= \int \mathbb{I}_A(z) g(z) f_Z(z) \, dz = \int \mathbb{I}_A(z) \left(\int x f_{X|Z}(x|z) \, dx \right) f_Z(z) dz \\ &= \int \int \mathbb{I}_A(z) x f_{X,Z}(x,z) \, dxdz = \int_A X \, d\mathbb{P} \, .\end{aligned}$$

Thus the conditions in Definition 4.7 are fulfilled. □

4.3 Conditioning on a σ-field

In Definition 4.7 of conditional expectation given a r.v. Z, we only consider sets of the σ-field $\sigma(Z)$ rather than the actual values of Z. This idea helps to generalize the definition of conditional expectation to a given σ-field.

Definition 4.9. Consider $X \in \mathscr{L}^1$ on $(\Omega, \mathscr{F}, \mathbb{P})$, and let $\mathscr{G} \subset \mathscr{F}$ be a σ-field. Then the *conditional expectation* of X given $\mathscr{G}$ is defined as a random variable $E(X|\mathscr{G})$ such that

i) $E(X|\mathscr{G})$ is a $\mathscr{G}$ measurable r.v. ;
ii) For any $A \in \mathscr{G}$,

$$\int_A E(X|\mathscr{G}) \, d\mathbb{P} = \int_A X \, d\mathbb{P} \, . \tag{4.14}$$

Remark 4.2. Combining Definition 4.7 and 4.9, it can be seen that

$$E(X|\sigma(Z)) = E(X|Z) \, .$$

Note that different r.v.s can generate the same σ-field. For example, on the sample space $\Omega = \{H, T\}$, $Z_1 = 1_{\{H\}}$ and $Z_2 = 1_{\{T\}}$ generate the same σ-field $\mathscr{F} = \{\varnothing, \{H\}, \{T\}, \Omega\}$. Thus, conditional expectation given a σ-field is more general than conditional expectation given a r.v.. □

Example 4.10. If $\mathscr{G} = \{\phi, \Omega\}$, then $E(X|\mathscr{G}) = E(X)$ a.s. To see this, note that if $\mathscr{G} = \{\varnothing, \Omega\}$, then any constant, including $E(X)$, can be regarded as a $\mathscr{G}$ measurable random variable. Note also that

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

and

$$\int_{\varnothing} X \, d\mathbb{P} = 0 = \int_{\varnothing} E(X) \, d\mathbb{P} \, .$$

Therefore, $E(X)$ satisfies i) and ii) of Definition 4.9. Thus, $E(X|\mathscr{G}) = E(X)$, a.s.. □

Example 4.11. If X is $\mathscr{G}$ measurable, then $E(X|\mathscr{G}) = X$ a.s. because X trivially satisfies i) and ii) of Definition 4.9. □

Example 4.12 (Tower Property). For $B \in \mathscr{G}$, we have $E(E(X|\mathscr{G})|B) = E(X|B)$. To see this, note that from the definition of conditional expectation, for $B \in \mathscr{G}$ we have

$$\int_B E(X|\mathscr{G}) \, d\mathbb{P} = \int_B X \, d\mathbb{P} \, . \tag{4.15}$$

Thus Definition 4.5 implies that

$$E(E(X|\mathscr{G})|B) \overset{(4.2)}{=} \frac{1}{\mathbb{P}(B)} \int_B E(X|\mathscr{G}) \, d\mathbb{P} \overset{(4.15)}{=} \frac{1}{\mathbb{P}(B)} \int_B X \, d\mathbb{P} \overset{(4.2)}{=} E(X|B) \, .$$

□

Example 4.13. Consider the fair dice example $\Omega = \{1, \ldots, 6\}$ and $\mathbb{P}(\{\omega\}) = \frac{1}{6}$. Let $\mathscr{F}_1 = \{\varnothing, \Omega\}$, $\mathscr{F}_2 = \{\varnothing, \{1,2,3\}, \{4,5,6\}, \Omega\}$, $\mathscr{F}_3 = 2^{\Omega}$ and $X : \Omega \to \mathbb{R}$ satisfying $X(\omega) = (\omega - 4)^+$. Note that

$$\sigma(X) = \{\varnothing, \{5\}, \{6\}, \{5,6\}, \{1,2,3,4\}, \{1,2,3,4,5\}, \{1,2,3,4,6\}, \Omega\} \, .$$

Thus, $X \in \mathscr{F}_3$ but $X \notin \mathscr{F}_1$ and $X \notin \mathscr{F}_2$. From Example 4.10, $E(X|\mathscr{F}_1) = E(X) = (5-4)\frac{1}{6} + (6-4)\frac{1}{6} = 0.5$. Next, $E(X|\mathscr{F}_2)$ is a $\mathscr{F}_2$ measurable r.v.. Thus, it takes constant values on each of the sets $\{1,2,3\}$ and $\{4,5,6\}$. Simple calculations yield

$$E(X|\mathscr{F}_2)(\omega) = \begin{cases} \frac{\int_{\{1,2,3\}} X \, d\mathbb{P}}{\mathbb{P}(\{1,2,3\})} = 0 & \text{if } \omega \in \{1,2,3\} \, , \\ \frac{\int_{\{4,5,6\}} X \, d\mathbb{P}}{\mathbb{P}(\{4,5,6\})} = \frac{1}{1/2} \left((5-4)\frac{1}{6} + (6-4)\frac{1}{6} \right) = 1 & \text{if } \omega \in \{4,5,6\} \, . \end{cases}$$

Lastly, since $X \in \mathscr{F}_3$, we have $E(X|\mathscr{F}_3) = X$ a.s. from Example 4.11. □

4.4 Radon-Nikodym derivative

In Definition 4.7 and 4.9, it is implicitly assumed that the random variable $E(X|Z)$ or $E(X|\mathscr{G})$ exists. The existence of conditional expectation is non-trivial, but is guaranteed by the *Radon-Nikodym* Theorem, which is an important tool in probability theory that makes connections between measures.

Definition 4.10. (Absolute Continuous measures) Suppose that $\mathbb{P}$ and $\mathbb{Q}$ are probability measures on $(\Omega, \mathscr{F})$ and $\mathbb{P}(F) = 0$ implies $\mathbb{Q}(F) = 0$ when $F \in \mathscr{F}$. Then $\mathbb{Q}$ is said to be **absolutely continuous** with respect to $\mathbb{P}$ and is denoted by $\mathbb{Q} \ll \mathbb{P}$.

Example 4.14. Consider the measurable space $(\Omega, \mathscr{F})$ where $\Omega = \{1,2,3,4,5,6\}$ and $\mathscr{F} = 2^{\Omega}$. For $\omega \in \Omega$, let $\mathbb{P}_1(\{\omega\}) = 1/6$, $\mathbb{P}_2(\{\omega\}) = |\omega - 3.5|/9$, $\mathbb{P}_3(\{\omega\}) = \frac{1}{3}1_{\{\omega \geq 4\}}$ and $\mathbb{P}_4(\{\omega\}) = \frac{1}{3}1_{\{\omega \leq 3\}}$. Then, the only absolute continuous pairs are $\mathbb{P}_1 \ll \mathbb{P}_2$, $\mathbb{P}_2 \ll \mathbb{P}_1$, $\mathbb{P}_3 \ll \mathbb{P}_i$ and $\mathbb{P}_4 \ll \mathbb{P}_i$ for $i = 1,2$. Also, if $\mathbb{P}_5(\{\omega\}) = \frac{1}{2}1_{\{\omega \leq 2\}}$, then $\mathbb{P}_5 \ll \mathbb{P}_4$ but $\mathbb{P}_4 \not\ll \mathbb{P}_5$.

For the measurable space $(\mathbb{R}, \mathscr{B})$, let $\mathbb{P}_N$ and $\mathbb{P}_E$ satisfy $\mathbb{P}_N((a,b]) = F_N(b) - F_N(a)$ and $\mathbb{P}_E((a,b]) = F_E(b) - F_E(a)$, where F_N and F_E are the c.d.f.s of the standard normal and standard exponential distribution family, respectively. Then we have $\mathbb{P}_E \ll \mathbb{P}_N$, but $\mathbb{P}_N \not\ll \mathbb{P}_E$. □

Theorem 4.4. (Radon-Nikodym Theorem) *Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on the measurable space $(\Omega, \mathscr{F})$. If $\mathbb{Q} \ll \mathbb{P}$, then there exists an $\mathscr{F}$ measurable r.v. $Y \in \mathscr{L}^1$ such that*

$$\mathbb{Q}(A) = \int_A Y \, d\mathbb{P}, \tag{4.16}$$

for any $A \in \mathscr{F}$. The r.v. Y is called the **Radon-Nikodym derivative** *of $\mathbb{Q}$ w.r.t. $\mathbb{P}$ and is denoted by $Y = \frac{d\mathbb{Q}}{d\mathbb{P}}$. In particular, the expectation of a random variable X under $\mathbb{Q}$ is*

$$\mathbb{E}(X) = \int X \, d\mathbb{Q} = \int XY \, d\mathbb{P}.$$

Proof. The proof consists of four steps:

i) Assume $\mathbb{Q}(A) \leq \mathbb{P}(A)$ for all $A \in \mathscr{F}$. Construct a Y_n on a *finite* σ-field $\mathscr{F}_n \subset \mathscr{F}$.
ii) Take limit to obtain the required r.v. $Y = \lim_{n\to\infty} Y_n$ a.s.
iii) Show that the limit Y satisfies (4.16) for all $A \in \mathscr{F}$.
iv) Relax the assumption $\mathbb{Q}(A) \leq \mathbb{P}(A)$.

i) [*Construction of a finite σ-field*] Assume that $\mathbb{Q}(A) \leq \mathbb{P}(A)$ for all $A \in \mathscr{F}$. First we define the r.v. Y_n on a finite sub-σ field $\mathscr{F}_n \subset \mathscr{F}$ which is generated by a finite partition of Ω. That is, for some integer r_n, the sample space is partitioned as $\Omega = \bigcup_{j=1}^{r_n} F_j$, where $F_i \cap F_j = \varnothing$ if $i \neq j$. Then the finite sub-σ field is defined by $\mathscr{F}_n = \sigma(F_1, \ldots, F_{r_n})$. Define $Y_n : \Omega \to \mathbb{R}$ by

$$Y_n(\omega) = \begin{cases} \frac{\mathbb{Q}(F_j)}{\mathbb{P}(F_j)} & \text{if } \omega \in F_j \text{ and } \mathbb{P}(F_j) > 0, \\ 0 & \text{if } \mathbb{P}(F_j) = 0. \end{cases} \tag{4.17}$$

Simple algebra show that $\mathbb{Q}(F_j) = \int_{F_j} Y_n \, d\mathbb{P}$. Since every $F \in \mathscr{F}_n$ is a union of F_js, (4.16) holds for all $A \in \mathscr{F}_n$. Also, note that $Y_n \in \mathscr{L}^1$ since

$$E(|Y_n|) \overset{(Y_n \geq 0)}{=} \int_\Omega Y_n \, d\mathbb{P} = \sum_{j=1}^{r_n} \int_{F_j} Y_n \, d\mathbb{P} = \sum_{j=1}^{r_n} \mathbb{Q}(F_j) = \mathbb{Q}(\Omega) = 1 < \infty.$$

Moreover, $Y_n(\omega)$ is a constant $(\mathbb{Q}(F_j)/\mathbb{P}(F_j))$ for all $\omega \in F_j$, so Y_n is $\mathscr{F}_n$ measurable.

Next consider a refinement $\mathscr{F}_{n+1} = \sigma(\tilde{F}_1, \ldots, \tilde{F}_{r_{n+1}})$ of $\mathscr{F}_n$ such that $\{\tilde{F}_j\}_{j=1,\ldots,r_{n+1}}$ are disjoint and each $F_i \in \mathscr{F}_n$ is a union of some $\tilde{F}_j$s. Define $Y_{n+1} : \Omega \to \mathbb{R}$ by

$$Y_{n+1}(\omega) = \begin{cases} \frac{\mathbb{Q}(\tilde{F}_j)}{\mathbb{P}(\tilde{F}_j)} & \text{if } \omega \in \tilde{F}_j \text{ and } \mathbb{P}(\tilde{F}_j) > 0\,, \\ 0 & \text{if } \mathbb{P}(\tilde{F}_j) = 0\,. \end{cases}$$

Similarly, it can be shown that $Y_{n+1} \in \mathscr{L}^1$, Y_{n+1} is $\mathscr{F}_{n+1}$ measurable, and $\mathbb{Q}(A) = \int_A Y_{n+1}\, d\mathbb{P}$ for all $A \in \mathscr{F}_{n+1}$. Moreover,

$$\begin{aligned} \int_\Omega Y_{n+1}^2\, d\mathbb{P} &= \int_\Omega (Y_{n+1} - Y_n)^2\, d\mathbb{P} + \int_\Omega Y_n^2\, d\mathbb{P} + 2\int_\Omega Y_n(Y_{n+1} - Y_n)\, d\mathbb{P} \\ &= \int_\Omega (Y_{n+1} - Y_n)^2\, d\mathbb{P} + \int_\Omega Y_n^2\, d\mathbb{P} \\ &\geq \int_\Omega Y_n^2\, d\mathbb{P}\,, \end{aligned} \tag{4.18}$$

where the second equality follows from

$$\begin{aligned} \int_\Omega Y_n(Y_{n+1} - Y_n)\, d\mathbb{P} &= \sum_{j=1}^{r_n} \int_{F_j} Y_n(Y_{n+1} - Y_n)\, d\mathbb{P} \\ &= \sum_{j=1}^{r_n} \frac{\mathbb{Q}(F_j)}{\mathbb{P}(F_j)} \left(\int_{F_j} Y_{n+1}\, d\mathbb{P} - \int_{F_j} \frac{\mathbb{Q}(F_j)}{\mathbb{P}(F_j)}\, d\mathbb{P} \right) \\ &= \sum_{j=1}^{r_n} \frac{\mathbb{Q}(F_j)}{\mathbb{P}(F_j)} \left(\mathbb{Q}(F_j) - \mathbb{Q}(F_j) \right) = 0\,. \end{aligned}$$

ii) [*Passing to the limit*] From (4.18), it can be observed that if the partition of Ω gets finer and finer (e.g., a sequence of σ-field $\mathscr{F}_1 \subset \ldots \subset \mathscr{F}_k \subset \ldots$), then the corresponding Radon-Nikodym derivatives $\{Y_k\}$ are non-decreasing in the sense that $\int_\Omega Y_{k+1}^2\, d\mathbb{P} \geq \int_\Omega Y_k^2\, d\mathbb{P}$. On the other hand, $\int_\Omega Y_k^2\, d\mathbb{P}$ is bounded because $\mathbb{P}(\Omega) = 1$ and $Y_k \leq 1$ by (4.17) and the assumption $\mathbb{Q}(A) \leq \mathbb{P}(A)$ for all $A \in \mathscr{F}$. Thus, the limit $l := \lim_{k\to\infty} \int_\Omega Y_k^2\, d\mathbb{P}$ exists. Some limit arguments show that the a.s. limit $Y = \lim_{k\to\infty} Y_k$ also exists (see Exercise 4.33).

iii) [*Verify* (4.16)] Since $Y_n \in \mathscr{F}_n \subset \mathscr{F}$ and Y is the limit of Y_n, Y is $\mathscr{F}$ measurable. For any $A \in \mathscr{F}$, let $\mathscr{R}_n$ be the common refinement of $\mathscr{F}_n$ and $\{A, A^c\}$. It can be verified directly that $\mathbb{Q}(A) = \int_A Y_{\mathscr{R}_n}\, d\mathbb{P}$ where $Y_{\mathscr{R}_n}$ is defined similarly as Y_n but is on $\mathscr{R}_n$ instead of $\mathscr{F}_n$. Similar limit arguments as in ii) show that $\int_A Y_{\mathscr{R}_{n_k}}\, d\mathbb{P} \to \int_A Y\, d\mathbb{P}$ for some subsequence $\{n_k\}_{k=1,2,\ldots}$, yielding (4.16) (see Exercise 4.33).

iv) [*Relax the assumption* $\mathbb{Q}(A) \leq \mathbb{P}(A)$] In general, let $\mathbb{S} = \mathbb{P} + \mathbb{Q}$, so $\mathbb{P}(A) \leq \mathbb{S}(A)$ and $\mathbb{Q}(A) \leq \mathbb{S}(A)$ for all $A \in \mathscr{F}$. The previous results imply that there exist $Y_\mathbb{Q}$ and $Y_\mathbb{P}$ such that $\mathbb{Q}(A) = \int_A Y_\mathbb{Q}\, d\mathbb{S}$ and $\mathbb{P}(A) = \int_A Y_\mathbb{P}\, d\mathbb{S}$ for all $A \in \mathscr{F}$. Now, let $Y = \frac{Y_\mathbb{Q}}{Y_\mathbb{P}} 1_{\{Y_\mathbb{P} > 0\}}$. Then,

$$\begin{aligned}\mathbb{Q}(A) &= \int_A Y_{\mathbb{Q}}\, d\mathbb{S} \\ &= \int_{A\cap\{Y_{\mathbb{P}}>0\}} \frac{Y_{\mathbb{Q}}}{Y_{\mathbb{P}}} Y_{\mathbb{P}}\, d\mathbb{S} + \int_{A\cap\{Y_{\mathbb{P}}=0\}} Y_{\mathbb{Q}}\, d\mathbb{S} \\ &= \int_A Y Y_{\mathbb{P}}\, d\mathbb{S} + 0 \\ &= \int_A Y\, d\mathbb{P},\end{aligned}$$

where the third equality holds because on $B := \{Y_{\mathbb{P}} = 0\}$, $\mathbb{P}(B) = \int_B Y_{\mathbb{P}}\, d\mathbb{S} = 0$, thus $\mathbb{Q}(B) = 0$ as $\mathbb{Q} \ll \mathbb{P}$, which implies $\mathbb{S}(B) = 0$. The last equality follows from Exercise 4.33. Thus the proof is completed. □

Using the Radon-Nikodym Theorem, the existence of conditional expectation can be justified:

Theorem 4.5. (Existence of Conditional Expectation) *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $\mathscr{G}$ be a σ-field contained in $\mathscr{F}$. Then for any r.v. $X \in \mathscr{L}^1$ there exists a $\mathscr{G}$ measurable r.v. $X_{\mathscr{G}}$ such that for every $A \in \mathscr{G}$,*

$$\int_A X\, d\mathbb{P} = \int_A X_{\mathscr{G}}\, d\mathbb{P}.$$

Here $X_{\mathscr{G}}$ can be regarded as the conditional expectation $E(X|\mathscr{G})$. □

Proof. Given a **positive** r.v. X, let $\mathbb{Q}_X : \mathscr{F} \to \mathbb{R}$ be a set function satisfying

$$\mathbb{Q}_X(A) = \int_A X\, d\mathbb{P}, \tag{4.19}$$

for $A \in \mathscr{F}$. Without loss of generality, we can assume $E(X) = 1$. Otherwise we can consider $X/E(X)$ instead of X in the following arguments. It can be checked that $\mathbb{Q}_X$ is a probability measure and $\mathbb{Q}_X \ll \mathbb{P}$ on $\mathscr{G}$. Thus, applying Radon-Nikodym Theorem on the measure $\mathbb{Q}_X(\cdot)$ yields a $\mathscr{G}$ measurable r.v., say Y, such that $\mathbb{Q}_X(A) = \int_A Y\, d\mathbb{P}$ for $A \in \mathscr{G}$. Together with (4.19), Y satisfies i) and ii) of Definition 4.9. Thus $X_{\mathscr{G}} = Y$ is the desired conditional expectation of X given $\mathscr{G}$.

For a general r.v. X, we can write $X = X^+ - X^-$, where $X^+ = X1_{\{X>0\}}$ and $X^- = -X1_{\{X<0\}}$ are positive r.v.s. Then the preceding result can be applied to each of X^+ and X^-, and the difference of the two conditional expectations is the desired $X_{\mathscr{G}}$. □

Example 4.15. Refer to Example 4.14, we have $\frac{d\mathbb{P}_1}{d\mathbb{P}_2}(\omega) = 3/2|\omega - 3.5|$, $\frac{d\mathbb{P}_2}{d\mathbb{P}_1}(\omega) = 2|\omega - 3.5|/3$, $\frac{d\mathbb{P}_3}{d\mathbb{P}_1}(\omega) = 2\cdot 1_{\{\omega\geq 4\}}$ and $\frac{d\mathbb{P}_4}{d\mathbb{P}_1}(\omega) = 2\cdot 1_{\{\omega\leq 3\}}$. What are $\frac{d\mathbb{P}_3}{d\mathbb{P}_2}$, $\frac{d\mathbb{P}_4}{d\mathbb{P}_2}$ and $\frac{d\mathbb{P}_5}{d\mathbb{P}_4}$?

On the other hand, note that $\mathbb{P}_N((-\infty, x)) = F_N(x) = \int_{-\infty}^x f_N(x)\, dx$ and $\mathbb{P}_E((-\infty, x)) = F_E(x) = \int_0^x f_E(x)\, dx$ where $f_N(x)$ and $f_E(x)$ are the p.d.f.s of standard normal and standard exponential distribution, respectively. Note that $\frac{f_E(x)}{f_N(x)}$ is measurable and satisfies

$$\int_A \frac{f_E(x)}{f_N(x)}\, d\mathbb{P}_N = \int_A \frac{f_E(x)}{f_N(x)} f_N(x)\, dx = \int_A f_E(x)\, dx = \mathbb{P}_E(A)\,.$$

Thus, $\frac{d\mathbb{P}_E}{d\mathbb{P}_N} = \frac{f_E(x)}{f_N(x)}$. However, $\frac{d\mathbb{P}_N}{d\mathbb{P}_E}$ does not exists as $\mathbb{P}_N$ is not absolutely continuous w.r.t. $\mathbb{P}_E$ (because $f_N(x)/f_E(x)$ does not exist for $x < 0$). □

4.5 General Properties of Conditional Expectations

Theorem 4.6. *Let X, $Y \in \mathscr{L}^1$ be random variables defined on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$, and $\mathscr{G}$, $\mathscr{H}$ are sub-σ fields of $\mathscr{F}$. Conditional expectation of X given $\mathscr{G}$ has the following properties:*

1) $E(aX + bY|\mathscr{G}) = aE(X|\mathscr{G}) + bE(Y|\mathscr{G})$ (linearity).
2) $E(E(X|\mathscr{G})) = E(X)$.
3) $E(XY|\mathscr{G}) = XE(Y|\mathscr{G})$ if X is $\mathscr{G}$ measurable.
4) $E(X|\mathscr{G}) = E(X)$ if X is independent of $\mathscr{G}$.
5) $E(E(X|\mathscr{G})|\mathscr{H}) = E(X|\mathscr{H})$ if $\mathscr{H} \subset \mathscr{G}$ (tower property).
6) If $X \geq 0$, then $E(X|\mathscr{G}) \geq 0$ (positivity).

Here a,b are real numbers. In (3), we also assume that the product $XY \in \mathscr{L}^1$. All equalities and inequalities hold almost surely.

Proof. 1. For any $B \in \mathscr{G}$, by the linearity of Lebesgue integral,

$$\begin{aligned}\int_B (aE(X|\mathscr{G}) + bE(Y|\mathscr{G}))d\mathbb{P} &= a\int_B E(X|\mathscr{G})d\mathbb{P} + b\int_B E(Y|\mathscr{G})d\mathbb{P} \\ &= \int_B (aX + bY)d\mathbb{P}.\end{aligned}$$

The result thus follows from the definition of conditional expectation (4.6).

2. Putting $B = \Omega$ in Example 4.12 completes the proof. Also, it is a special case of (5) when $\mathscr{H} = \{\varnothing, \Omega\}$.
3. We first verify the result for $X = \mathbb{I}_A$ (indicator function) where $A \in \mathscr{G}$. In this case, for any $B \in \mathscr{G}$,

$$\begin{aligned}\int_B \mathbb{I}_A E(Y|\mathscr{G})d\mathbb{P} &= \int_{A\cap B} E(Y|\mathscr{G})d\mathbb{P} \\ &= \int_{A\cap B} Y d\mathbb{P} \\ &= \int_B \mathbb{I}_A Y d\mathbb{P}\,.\end{aligned}$$

Then, it follows from the definition of conditional expectation (4.6) that

$$\mathbb{I}_A E(Y|\mathscr{G}) = E(\mathbb{I}_A Y|\mathscr{G})\,.$$

Using 1), we can extend the preceding result to a simple function $X = \sum_{j=1}^{m} a_j \mathbb{I}_{A_j}$, i.e.,

$$XE(Y|\mathscr{G}) = \sum_{j=1}^{m} a_j \mathbb{I}_{A_j} E(Y|\mathscr{G}) = E\left(\sum_{j=1}^{m} a_j \mathbb{I}_{A_j} Y|\mathscr{G}\right) = E(XY|\mathscr{G}).$$

where $A_j \in \mathscr{G}$ for $j = 1, 2, \ldots, m$. Finally, the general result follows by approximating X with an increasing sequence of simple functions and use MCT.

4. Since X is independent of $\mathscr{G}$, for any $B \in \mathscr{G}$, the random variables X and $\mathbb{I}_B$ are independent. We first verify the result for $X = \mathbb{I}_A$ where $A \in \mathscr{F}$ and $\mathscr{F}$ is the σ-field where X is defined on. In particular,

$$\int_B E(\mathbb{I}_A) d\mathbb{P} = E(\mathbb{I}_A) \int_B d\mathbb{P} = \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A \cap B) = \int_{A \cap B} d\mathbb{P} = \int_B \mathbb{I}_A \, d\mathbb{P},$$

where the independence of $\mathbb{I}_A$ and $\mathbb{I}_B$ (thus A and B) is used in the third equality. Now we have obtained that $\int_B E(X) d\mathbb{P} = \int_B X d\mathbb{P}$ with $X = \mathbb{I}_A$ for all $A \in \mathscr{F}$ and $B \in \mathscr{G}$. Similar to 3), we can extend the results to simple function $X = \sum_{j=1}^{m} a_j \mathbb{I}_{A_j}$ and use MCT to generalize to any $\mathscr{F}$ measurable X. Thus we obtain $\int_B E(X) d\mathbb{P} = \int_B X d\mathbb{P}$ for all $B \in \mathscr{G}$, i.e., $E(X|\mathscr{G}) = E(X)$.

5. By Definition,

$$\int_B E(X|\mathscr{G}) d\mathbb{P} = \int_B X d\mathbb{P}$$

for every $B \in \mathscr{G}$ and

$$\int_B E(X|\mathscr{H}) d\mathbb{P} = \int_B X d\mathbb{P}$$

for every $B \in \mathscr{H}$. Since $\mathscr{H} \subset \mathscr{G}$,

$$\int_B E(X|\mathscr{G}) d\mathbb{P} = \int_B E(X|\mathscr{H}) d\mathbb{P}$$

for every $B \in \mathscr{H}$. Therefore, the definition of conditional expectation implies that

$$E(E(X|\mathscr{G})|\mathscr{H}) = E(X|\mathscr{H}).$$

6. For any $n \in \mathbb{Z}^+$, we put

$$A_n = \left\{ E(X|\mathscr{G}) \leq -\frac{1}{n} \right\}.$$

Since $E(X|\mathscr{G})$ is a $\mathscr{G}$ measurable r.v., we have $A_n \in \mathscr{G}$. If $X \geq 0$ a.s., then

$$0 \leq \int_{A_n} X d\mathbb{P} \overset{(4.6)}{=} \int_{A_n} E(X|\mathscr{G}) d\mathbb{P} \leq -\frac{1}{n} \mathbb{P}(A_n),$$

which implies $\mathbb{P}(A_n) = 0$. Since

$$\{E(X|\mathscr{G}) < 0\} = \bigcup_{n=1}^{\infty} A_n,$$

it follows that $\mathbb{P}\{E(X|\mathscr{G}) < 0\} = \mathbb{P}(\bigcup_{n=1}^{\infty} A_n) \le \sum_{n=1}^{\infty} \mathbb{P}(A_n) = 0$, completing the proof. □

Example 4.16. **(Independence and Zero-Correlation)** Recall that $X \in \mathscr{L}^1$ if $E(|X|) < \infty$. If X_1 and X_2 are independent, then they are **uncorrelated**, i.e., $E(X_1X_2) = E(X_1)E(X_2)$. To see this, note that

$$E(X_1X_2) \stackrel{(2)}{=} E(E(X_1X_2|\sigma(X_2))) \stackrel{(3)}{=} E(X_2E(X_1|\sigma(X_2))) \stackrel{(4)}{=} E(X_2E(X_1)) \stackrel{(1)}{=} E(X_1)E(X_2)\,, \tag{4.20}$$

where the four equalities are consequences of Properties 2), 3), 4), 1) of Theorem 4.6 respectively. Repeating the argument in (4.20), we have that, if $X_1, \ldots, X_n \in \mathscr{L}^1$ are independent, then they are uncorrelated, i.e.,

$$E(X_1X_2\cdots X_n) = E(X_1)E(X_2)\cdots E(X_n)$$

provided that the product $X_1X_2\cdots X_n \in \mathscr{L}^1$. However, the opposite direction is not true. See Exercise 4.18.

4.6 Useful Inequalities

This section collects some inequalities that are frequently used in probability theory and statistical inference.

Theorem 4.7. (Markov's Inequality) *If $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ is a monotonically increasing function, and X is a random variable on $(\Omega, \mathscr{F}, \mathbb{P})$ satisfying $E(\psi(X)) < \infty$, then*

$$\mathbb{P}(|X| \ge a) \le \frac{E(\psi(|X|))}{\psi(a)}\,,$$

for any $a > 0$.

Proof. Since ψ is increasing, the events $\{|X| \ge a\}$ and $\{\psi(|X|) \ge \psi(a)\}$ are the same. Thus,

$$\begin{aligned} E(\psi(|X|)) &\ge E\left(\psi(|X|)1_{\{\psi(|X|)\ge\psi(a)\}}\right) \ge E\left(\psi(a)1_{\{\psi(|X|)\ge\psi(a)\}}\right) \\ &= \psi(a)\mathbb{P}\{\psi(|X|) \ge \psi(a)\} = \psi(a)\mathbb{P}\{|X| \ge a\}\,, \end{aligned}$$

which completes the proof. □

Corollary 4.8 (Chebyshev's Inequality) *Putting $\psi(x) = x^2$ and replacing X by $X - E(X)$, Markov's inequality in Theorem 4.7 becomes*

$$\mathbb{P}(|X-E(X)| \geq a) \leq \frac{Var(X)}{a^2},$$

for any $a > 0$*.*

Example 4.17. **Law of large numbers** If $X_1, \ldots, X_n$ are i.i.d. random variables on $(\Omega, \mathscr{F}, \mathbb{P})$ with mean μ and variance σ^2, then Chebyshev's inequality implies

$$\begin{aligned}\mathbb{P}\left(\left|\frac{\sum_{i=1}^n X_i}{n} - \mu\right| > \varepsilon\right) &\leq \frac{\operatorname{Var}\left(\frac{\sum_{i=1}^n (X_i-\mu)}{n}\right)}{\varepsilon^2} \\ &= \frac{\sigma^2}{n\varepsilon^2} \\ &\to 0, \quad \text{as } n \to \infty.\end{aligned}$$

Hence, we have $\frac{\sum_{i=1}^n X_i}{n} \xrightarrow{p} \mu$, which is the classical **law of large numbers**.

Definition 4.11. (Convex Function) A function $\varphi : \mathbb{R} \to \mathbb{R}$ is **convex** if for any $x, y \in \mathbb{R}$ and any $\lambda \in [0,1]$

$$\varphi(\lambda x + (1-\lambda)y) \leq \lambda \varphi(x) + (1-\lambda)\varphi(y).$$

Graphically, a function φ is convex if it lies below the line segment from the point $(x, \varphi(x))$ to the point $(y, \varphi(y))$ on any interval $[x, y]$. See Figure 4.3. □

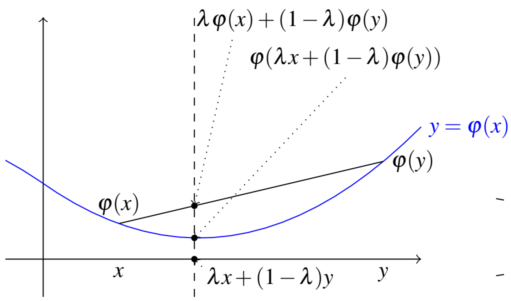


Fig. 4.3 Convex function

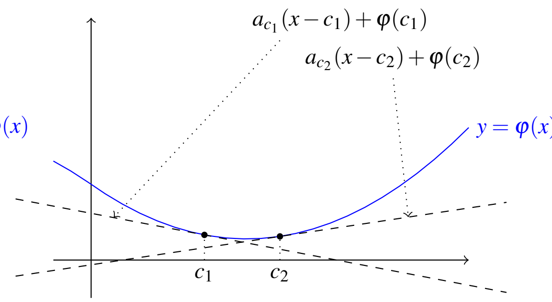


Fig. 4.4 lines below a convex function

Theorem 4.9. (Jensen's Inequality) *Let* $\varphi : \mathbb{R} \to \mathbb{R}$ *be a convex function and let* $X \in \mathscr{L}^1$ *on* $(\Omega, \mathscr{F}, \mathbb{P})$ *such that* $\varphi(X) \in \mathscr{L}^1$*. Then*

$$\varphi(E(X|\mathscr{G})) \leq E(\varphi(X)|\mathscr{G}) \quad a.s. \tag{4.21}$$

for any σ*-field* $\mathscr{G}$ *on* Ω *contained in* $\mathscr{F}$*.*

Proof. By the convexity of φ, for every fixed $c \in \mathbb{R}$, there exists an a_c such that $\varphi(x) \geq a_c(x-c) + \varphi(c)$ for all x (See Figure 4.4 and Exercise 4.32). Replacing x

by a r.v. X and taking conditional expectation on both sides, Theorem 4.6 1) and 6) imply

$$E(\varphi(X)|\mathscr{G}) \geq a_c(E(X|\mathscr{G}) - c) + \varphi(c),$$

almost surely. Since c is arbitrary, put $c = E(X|\mathscr{G})$ yields (4.21). □

Theorem 4.10. (Hölder's Inequality) *Let $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$ and X, Y be r.v.s on $(\Omega, \mathscr{F}, \mathbb{P})$. If both $E(|X|^p)$ and $E(|Y|^q)$ are in $(0, \infty)$, then $E(|XY|) < \infty$ and*

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

Proof. First, from Figures 4.5 and 4.6, for any $\alpha > 0$ and $\beta > 0$, the sum of the shaded areas is greater than $\alpha\beta$. That is, we have the inequality

$$\alpha\beta \leq \int_0^\alpha t^{p-1}\,dt + \int_0^\beta u^{q-1}\,du = \frac{\alpha^p}{p} + \frac{\beta^q}{q}. \tag{4.22}$$

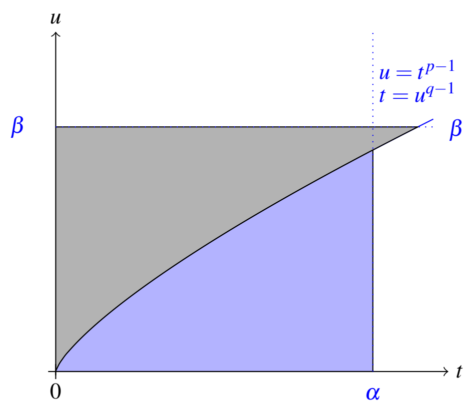


Fig. 4.5 Graphical illustration of (4.22) with $p = 0.75$. Note that $\frac{1}{p-1} = q - 1$

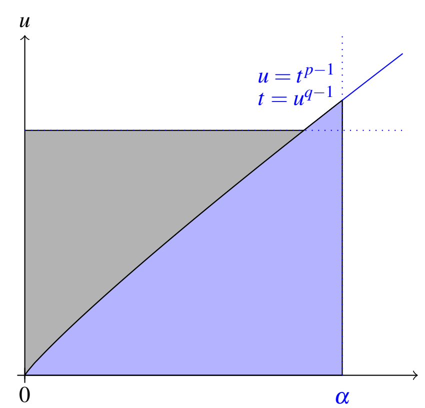


Fig. 4.6 Graphical illustration of (4.22) with $p = 0.9$.

Substituting $\alpha = |X|/E(|X|^p)^{\frac{1}{p}}$ and $\beta = |Y|/E(|Y|^q)^{\frac{1}{q}}$ in (4.22), we have

$$\frac{|XY|}{(E(|X|^p))^{\frac{1}{p}}(E(|Y|^q))^{\frac{1}{q}}} \leq \frac{|X|^p}{pE(|X|^p)} + \frac{|Y|^q}{qE(|Y|^q)}.$$

Taking expected values on both sides and rearranging terms completes the proof. □

The following is the triangular inequality for the $\mathscr{L}^p$ norm $\|X\|_p = (E|X|^p)^{1/p}$, which is the $\mathscr{L}^p$ distance between X and 0.

Theorem 4.11. (Minkowski's Inequality) *Let $p > 1$ and X, Y be r.v.s on $(\Omega, \mathscr{F}, \mathbb{P})$. If $E(X^p) < \infty$, then*

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

Proof. We focus on $p > 1$ as $p = 1$ corresponds to the usual triangular inequality. Note that

$$|X+Y|^p \leq |X||X+Y|^{p-1} + |Y||X+Y|^{p-1}. \tag{4.23}$$

Set $q = \frac{p}{p-1}$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Apply Hölder's inequality on $|X||X+Y|^{p-1}$ yields

$$\int_\Omega |X||X+Y|^{p-1}\, d\mathbb{P} \leq \|X\|_p (E|X+Y|^{(p-1)q})^{\frac{1}{q}} = \|X\|_p \|X+Y\|_p^{\frac{p}{q}}. \tag{4.24}$$

Similarly, we have

$$\int_\Omega |Y||X+Y|^{p-1}\, d\mathbb{P} \leq \|Y\|_p \|X+Y\|_p^{\frac{p}{q}}. \tag{4.25}$$

The result follows from combining (4.23), (4.24) and (4.25). $\square$

Given a sequence of random variables $X_1, \ldots, X_n$ on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$, we have the following generalization of MCT, Fatou's lemma and DCT:

Theorem 4.12. *Let $\mathscr{G}$ be a sub σ-field of $\mathscr{F}$.*

i) **(Monotone Convergence Theorem (MCT))** *If $0 \leq X_n \uparrow X$, then $E(X_n|\mathscr{G}) \uparrow E(X|\mathscr{G})$ a.s..*
ii) **(Fatou's Lemma)** *If $X_n \geq 0$, then $E(\liminf X_n|\mathscr{G}) \leq \liminf E(X_n|\mathscr{G})$ a.s..*
iii) **(Dominated Convergence Theorem (DCT))** *If $X_n \overset{a.s.}{\to} X$, and for some r.v. M with $E(M) \leq \infty$, $X_n \leq M$ a.s. for all $n \geq 1$, then*

$$E(X_n|\mathscr{G}) \overset{a.s.}{\to} E(X|\mathscr{G}).$$

4.7 Exercises

Exercise 4.13 *Consider the dice example $(\Omega, \mathscr{F}, \mathbb{P})$ where $\Omega = \{1,2,3,4,5,6\}$ and $\mathscr{F} = 2^\Omega$. Let $\mathbb{P}(\{\omega\}) = 1/6$ for $\omega = 1, \ldots, 6$, and*

$$\begin{aligned}
A_1 &= \{1,2,3\}, A_2 = \{4,5,6\}, A_3 = \{2,5\}, A_4 = \{2,6\},\\
X_i &= 1_{A_i}, \mathscr{F}_i = \sigma(\{A_i\}) \text{ for } i = 1, \ldots, 4,\\
X_5 &= X_5(\omega) = \omega, \mathscr{F}_5 = \sigma(\{A_1, A_3\}), \mathscr{F}_6 = \sigma(\{A_3, A_4\}), \mathscr{F}_7 = 2^\Omega.
\end{aligned}$$

a) Which events are independent?
b) Which random variables are independent?
c) Which σ-fields are independent?
d) Repeat a)-c) using the measure $\mathbb{P}(\{i\}) = \frac{|i-3.5|}{9}$ for $i = 1, \ldots, 6$.

Exercise 4.14 *Under the setting of Exercise 4.13,*

a) Find $\sigma(X_i)$ for $i = 1, \ldots, 5$.
b) What are $E(X_1|X_5)$, $E(X_5|X_1)$, $E(X_3|X_4)$ and $E(X_5|X_2)$?
c) What are $E(X_1|\mathscr{F}_2)$ and $E(X_5|\mathscr{F}_1)$?

Exercise 4.15 *Explain the meaning of independence between two events and mutually exclusive between two events. Can two events be both independent and mutually exclusive? Explain.*

Exercise 4.16 *How to define the independence between an event and a random variable? How to define the independence between an event and a σ-field? How to define the independence between a random variable and a σ-field?*

Exercise 4.17 *Show that two random variables X and Y are independent if and only if the σ-fields $\sigma(X)$ and $\sigma(Y)$ are independent.*

Exercise 4.18 *Let $X \sim N(0,1)$ and $Y = X^2$. Show that X and Y are uncorrelated but dependent.*

Exercise 4.19 *Let $\mathbb{I}_A$ be the* **indicator function** *of an event A, i.e.*

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

Show that $E(\mathbb{I}_A|B) = \mathbb{P}(A|B)$ for any event B with $\mathbb{P}(B) \neq 0$.

Exercise 4.20 *For the following r.v.s, write down the probability space and the induced probability space, and find the expectation of X.*

i) (Constant r.v.) $X(\omega) = a$ for all ω.
ii) $X : [0,1] \to \mathbb{R}$ given by $X(\omega) = \min\{\omega, 1-\omega\}$ (The distance to the nearest endpoint of the interval $[0,1]$)
iii)$X : [0,1]^2 \to \mathbb{R}$, the distance to the nearest endpoint of the unit square.

Exercise 4.21 *Let $\Omega = [0,1]$, $\mathscr{F} = \mathscr{B}_{[0,1]}$ and $\mathbb{P} = \lambda_{[0,1]}$. Let $X(\omega) = \omega(1-\omega)$ for $\omega \in \Omega$. For any random variable Y on $(\Omega, \mathscr{F}, \mathbb{P})$,*

i) Show that $Y(\omega) + Y(1-\omega)$ is $\sigma(X)$ measurable. (Hint: Draw the graph Y against ω)
ii) Show that $E(Y|X)(\omega) = \frac{Y(1-\omega)+Y(\omega)}{2}$ a.s..

Exercise 4.22 *Take $\Omega = [0,1]$, $\mathscr{F} = \mathscr{B}_{[0,1]}$ and $\mathbb{P} = \lambda_{[0,1]}$, the Lebesgue measure on $[0,1]$. Find $E(X|Z)$ if $X(\omega) = 2\omega^2$ and $Z(\omega) = 1_{\{\omega \in [0,1/3]\}} + 2 \times 1_{\{\omega \in (2/3,1]\}}$.*

Exercise 4.23 *Show that if Z is a constant function, then $E(X|Z)$ is constant and equals to $E(X)$ a.s..*

Exercise 4.24 *On $(\Omega, \mathscr{F}, \mathbb{P})$, show that for A, $B \in \mathscr{F}$ such that $\mathbb{P}(B) \neq 0$,*

$$E(1_A|1_B)(\omega) = \begin{cases} \mathbb{P}(A|B) & \text{if } \omega \in B, \\ \mathbb{P}(A|\Omega \backslash B) & \text{if } \omega \notin B. \end{cases}$$

Exercise 4.25 *Recall that if $Z : \Omega \to \mathbb{R}$ is a discrete r.v. taking n values $\{z_i\}_{i=1,\ldots,n}$, then there exists disjoint A_i such that $\Omega = \cup_{i=1}^n A_i$ and $Z(\omega) = z_i$ whenever $\omega \in A_i$. Show that every element in $\sigma(Z)$ is a union of some A_is. Generalize to the case where Z is a discrete r.v. taking infinite values $\{z_i\}_{i=1,\ldots}$.*

Exercise 4.26 *Show that if X is $\mathscr{G}$ measurable, then $E(X|\mathscr{G}) = X$ a.s.*

Exercise 4.27 *Refer to Example 4.12. If $B \notin \mathscr{G}$, does the result still hold? Prove it or give a counter example.*

Exercise 4.28 *Show that, if $\mathbb{Q} \ll \mathbb{P}$, then for any $\varepsilon > 0$, we can find a $\delta > 0$ such that $\mathbb{P}(F) < \delta$ implies $\mathbb{Q}(F) \leq \varepsilon$.*

Exercise 4.29 *Show that, if $\mathbb{P}$ and $\mathbb{Q}$ satisfy* (4.16) *for any $A \in \mathscr{F}$., then $\mathbb{Q} \ll \mathbb{P}$.*

Exercise 4.30 *Let $\mathbb{Q}_1$, $\mathbb{Q}_2$ and $\mathbb{P}$ be measures on $(\Omega, \mathscr{F})$. Show that if $\mathbb{Q}_1 \ll \mathbb{P}$ and $\mathbb{Q}_2 \ll \mathbb{P}$, then $\mathbb{Q}_1 + \mathbb{Q}_2 \ll \mathbb{P}$.*

Exercise 4.31 *Consider the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with $\Omega = (1,2,3,4,5,6)$ and $\mathscr{F} = 2^{\Omega}$. If $\mathbb{P}(\{i\}) = i/21$ and $\mathbb{Q}(\{i\}) = (i-1)/15$. Is $\mathbb{P} \ll \mathbb{Q}$? Is $\mathbb{Q} \ll \mathbb{P}$? Find the Radon-Nikodym derivative $\frac{d\mathbb{P}}{d\mathbb{Q}}$ and $\frac{d\mathbb{Q}}{d\mathbb{P}}$ if they exist.*

Exercise 4.32 *Suppose that $y = \varphi(x)$ is a convex function.*

i) Using Definition 4.11, show that for any $c \leq d \leq x$, $\frac{\varphi(d)-\varphi(c)}{d-c} \leq \frac{\varphi(x)-\varphi(d)}{x-d}$.
ii) Show that for any $c \leq d \leq x$, $\frac{\varphi(d)-\varphi(c)}{d-c} \leq \frac{\varphi(x)-\varphi(c)}{x-c}$.
iii)Show that for any $x \leq e \leq c$, $\frac{\varphi(c)-\varphi(e)}{c-e} \geq \frac{\varphi(x)-\varphi(c)}{x-c}$.
iv) Show that for $e \leq c \leq d$, $\frac{\varphi(c)-\varphi(e)}{c-e} \leq \frac{\varphi(d)-\varphi(c)}{d-c}$. Hence, deduce that $\lim_{e\uparrow c} \frac{\varphi(c)-\varphi(e)}{c-e} \triangleq l_1$ and $\lim_{d\downarrow c} \frac{\varphi(d)-\varphi(c)}{d-c} \triangleq l_2$ exist and satisfy $l_1 \leq l_2$.
v) Using the above results, show that, for every fixed $c \in \mathbb{R}$, there exists an a_c such that $\varphi(x) \geq a_c(x-c) + \varphi(c)$ for all x.

Exercise 4.33 *Using the notation in the proof of Theorem 4.4.*

i) From the existence of $l = \lim_{n\to\infty} \int_\Omega Y_n^2\, d\mathbb{P}$, show that we can find a subsequence $\{n_k\}_{k=1,\ldots}$ such that $l - \frac{1}{4^k} < \int_\Omega Y_{n_k}^2\, d\mathbb{P} \leq l$, for all $k \in \mathbb{Z}$.
ii) Using the definition of Y_k, show that $\int_\Omega (Y_{n_{k+1}} - Y_{n_k})^2\, d\mathbb{P} = \int_\Omega (Y_{n_{k+1}}^2 - Y_{n_k}^2)\, d\mathbb{P} < \frac{1}{4^k}$.
iii)By Cauchy Schwartz inequality, show that $\int_\Omega |Y_{n_{k+1}} - Y_{n_k}|\, d\mathbb{P} < \frac{1}{2^k}$.
iv) Using (iii) and MCT, show that $\int_\Omega \sum_{k=1}^\infty |Y_{n_{k+1}} - Y_{n_k}|\, d\mathbb{P} = \sum_{k=1}^\infty \int_\Omega |Y_{n_{k+1}} - Y_{n_k}|\, d\mathbb{P} < \infty$.
v) Using (iv), argue that $\sum_{k=1}^\infty |Y_{n_{k+1}} - Y_{n_k}| < \infty$ almost surely. (Hint: proof by contradiction). Then, show that $\sum_{k=1}^\infty (Y_{n_{k+1}} - Y_{n_k}) < \infty$ almost surely.
vi) Using (v), show that $\lim_{k\to\infty} Y_{n_k}$ exists almost surely.
vii) Noting that $l - \frac{1}{4^k} < \int_\Omega Y_{n_k}^2 \leq \int_\Omega Y_{\mathscr{R}_{n_k}}^2 \leq l$, repeating the arguments (ii)-(iii), show that $\left| \int_A |Y_{\mathscr{R}_{n_k}} - Y_{n_k}|\, d\mathbb{P} \right| < 1/2^k$.

viii) Using (vi) and (vii), show that $\int_A Y_{\mathscr{R}_{n_k}}\, d\mathbb{P} \to \int_A Y\, d\mathbb{P}$.

ix) Using $\int_A Y_{\mathbb{P}}\, d\mathbb{S} = \mathbb{P}(A) = \int_A d\mathbb{P}$, *show that* $E_{\mathbb{S}}(YY_{\mathbb{P}}) = E_{\mathbb{P}}(Y)$ *for any simple function* Y *on* $\mathscr{F}$. *Extend to the case where* Y *is* $\mathscr{F}$ *measurable.*

Chapter 5
Stochastic Processes and Martingales

In Chapter 2 we have discussed the formal definition of random variable. In practice, one random variable is seldom sufficient to model real world phenomena. For example, a random variable S_t may be used to model the closing price of a stock at time t. However, to describe the evolution of the closing prices, we need a sequence of random variables $\{S_t\}_{t=1,2...}$. If we wish to investigate closely the tick-by-tick evolution of the stock prices, then we may even need to use a continuous process $\{S_t\}_{t\in[0,\infty)}$. In this chapter we study how to define discrete time and continuous time processes. Moreover, we will study one particular class of stochastic process, **martingale**, which is of fundamental importance in mathematical finance.

5.1 Sequences of Random Variables

Discrete time stochastic process is a sequence of random variables $\{X_t\}_{t=1,2,...}$ indexed by a subset of integers. It represents the outcomes of a series of random phenomena, for example, a sequence of coin tosses, or the evolution of daily stock closing prices. Note that 'discrete time' $t = 1,2,3,\ldots$ is used to keep track of the order of events, and may not necessarily be related to the physical time of event occurrences. For example, the indexes for stock prices are recorded only on business days, but not on Saturdays, Sundays or public holidays.

Definition 5.1. (Discrete Time Stochastic Process) A **Discrete Time Stochastic Process** is an infinite dimensional random vector $\mathbf{X} = \{X_t\}_{t=1,2,...}$ from some abstract measurable space $(\Omega,\mathscr{F})$ to the measurable space $(\mathbb{R}^\infty,\mathscr{B}^\infty)$, where $\mathbb{R}^\infty = \mathbb{R}\times\mathbb{R}\times\cdots$ and $\mathscr{B}^\infty$ is $\sigma(\mathscr{B}\times\mathscr{B}\times\cdots)$.

Definition 5.2. (Sample Path) For any fixed $\omega\in\Omega$, the image of the stochastic process is a sequence of numbers $X_1(\omega),X_2(\omega),\ldots$, which is called a **sample path**.

Example 5.1. Consider three consecutive coin tosses X_i, $i = 1,2,3$, with $X_i = 1$ for head and 0 for tail. Setting $S = \{H,T\}$, we can take $\Omega = S\times S\times S$, $\mathscr{F} =$

$\sigma(2^S \times 2^S \times 2^S)$. Note that Ω contains 8 possible elements, and each element corresponds to a sample path. For example, if $\omega_1 = HHH, \omega_2 = HHT, \omega_8 = TTT$, then $X_1(\omega_1) = X_2(\omega_1) = X_3(\omega_1) = 1$; $X_1(\omega_2) = X_2(\omega_2) = 1, X_3(\omega_2) = 0$; $X_1(\omega_8) = X_2(\omega_8) = X_3(\omega_8) = 0$. A probability measure may be defined by $\mathbb{P}(\{\omega_i\}) = \frac{1}{8}$ for $i = 1, 2, \ldots, 8$. □

5.1.1 Filtrations

Intuitively, the σ-field $\mathscr{B}^\infty$ in Definition 5.1 contains all possible evolutions of the whole process $X_1, X_2, \ldots$. As the time n goes on sequentially, we can define the **filtration**, which is an *increasing sequence* of σ-fields that contains all possible evolutions up to different time points.

Definition 5.3. (Filtration) A sequence of σ-fields $\mathscr{F}_1, \mathscr{F}_2, \ldots$ on Ω such that

$$\mathscr{F}_1 \subset \mathscr{F}_2 \subset \cdots$$

is called a **filtration**.

Definition 5.4. (Random Variables adapted to a Filtration) A sequence of random variables $X_1, X_2, \ldots$ is **adapted** to a filtration $\mathscr{F}_1, \mathscr{F}_2, \ldots$ if X_n is $\mathscr{F}_n$-measurable for each $n = 1, 2, \ldots$, i.e., $X_n^{-1}(B) \in \mathscr{F}_n$ for all $B \in \mathscr{B}$.

Example 5.2. **(Natural Filtration)** If $\mathscr{F}_n = \sigma(X_1, \ldots, X_n)$ is the σ-field generated by $X_1, \ldots, X_n$, then $X_1, X_2, \ldots$ are adapted to $\mathscr{F}_1, \mathscr{F}_2, \ldots$. Such a filtration $\{\mathscr{F}_n\}_{n=1,\ldots,}$ is known as the **natural filtration**. □

Remark 5.1. **(σ-field = Information)** The σ-field $\mathscr{F}_n$ may represent our knowledge up to time n of everything related to $X_1, \ldots, X_n$ (In general, it may contain more information). To understand it mathematically, note that if $A \in \mathscr{F}_n$, then $\mathbb{P}(A|\mathscr{F}_n) = E(1_A|\mathscr{F}_n) = 1_A$, since 1_A is $\mathscr{F}_n$ measurable. That is, at time n, if $A \in \mathscr{F}_n$, then we know that whether A has occurred or not. Mathematically,

$$\mathbb{P}(A|\mathscr{F}_n)(\omega) = 1_A(\omega) = \begin{cases} 1 & \omega \in A \quad (A \text{ has occurred}) \\ 0 & \omega \notin A \quad (A \text{ has not occurred}) \end{cases}.$$

□

Example 5.3. For a sequence $X_1, X_2, \ldots$ of coin tosses we take $\mathscr{F}_n$ to be the σ-field generated by $X_1, \ldots, X_n$, i.e.,

$$\mathscr{F}_n = \sigma(X_1, \ldots, X_n).$$

Let

$$A = \{\text{the first 5 tosses give at least 2 heads}\}.$$

At time $n = 5$, the coin has been tossed five times. It will be possible to decide whether A has occurred or not. This means that $A \in \mathscr{F}_5$. However, at $n = 4$ it is not always possible to tell if A has occurred or not. If the outcomes of the first four tosses are, say,

$$\text{tail, tail, head, tail,} \tag{5.1}$$

then the event A remains undecided. Therefore, $A \notin \mathscr{F}_4$. □

Remark 5.2. Suppose that the outcomes of the first four coin tosses are

$$\text{tail, head, tail, head.}$$

In this case we know that A has occurred already at $n = 4$, regardless of the outcome of the fifth toss. It does not mean, however, that A belongs to $\mathscr{F}_4$. Intuitively, in order for A to belong to $\mathscr{F}_4$, it must be possible to tell whether A has occurred or not after the first four tosses, *no matter what the first four outcomes are*. This is clearly not so in view of (5.1).

Mathematically, the event A can be formulated as $A = \{HHTTT, HTHTT, \ldots\}$ ($C_2^5 + C_3^5 + C_4^5 + C_5^5$ elements), which is clearly not in $\mathscr{F}_4$. On the other hand, the statement "A must occur given that the first four coin tosses were $\{THTH\}$" is equivalent to the statement "$\mathbb{P}(A|\{THTH\}) = 1$", which is not the same as the statement "A belongs to $\mathscr{F}_4$", i.e., "$\mathbb{P}(A|\mathscr{F}_4)(\omega) = 1_A(\omega)$". Note that the last statement is not correct as $A \notin \mathscr{F}_4$. □

5.2 Martingales

The concept of a martingale has its origin in gambling for describing a fair game of chance. In fact, martingales reach well beyond gambling and appear in various areas of modern probability and stochastic analysis. First we introduce the basic definition and properties for **martingales in discrete time**.

Definition 5.5. (Martingale) A sequence of random variables $X_1, X_2, \ldots$ is called a **martingale** with respect to a filtration $\mathscr{F}_1, \mathscr{F}_2, \ldots$ if

1. X_n is integrable for each $n = 1, 2, \ldots$, i.e., $X_n \in \mathscr{L}^1$, or $E|X_n| < \infty$;
2. $X_1, X_2, \ldots$ is adapted to $\mathscr{F}_1, \mathscr{F}_2, \ldots$;
3. $E(X_{n+1}|\mathscr{F}_n) = X_n$ a.s. for each $n = 1, 2, \ldots$.

Example 5.4. Let $Z_1, Z_2, \ldots$ be a sequence of independent integrable random variables such that $E(Z_n) = 0$ for all $n = 1, 2, \ldots$. Set

$$X_n = Z_1 + \ldots + Z_n\,,$$

$$\mathscr{F}_n = \sigma(Z_1, \ldots, Z_n)\,.$$

As X_n is $\mathscr{F}_n$ measurable for each $n \in \mathbb{Z}^+$, $\{X_n\}_{n=1,2,\dots}$ is adapted to the filtration $\{\mathscr{F}_n\}_{n=1,2,\dots}$. Also, as Z_n is assumed to be integrable,

$$\begin{aligned} E(|X_n|) &= E(|Z_1 + \cdots + Z_n|) \\ &\le E(|Z_1|) + \cdots + E(|Z_n|) \\ &< \infty, \end{aligned}$$

i.e., X_n is integrable. Moreover,

$$\begin{aligned} E(X_{n+1}|\mathscr{F}_n) &= E(Z_{n+1}|\mathscr{F}_n) + E(X_n|\mathscr{F}_n) \\ &= E(Z_{n+1}) + X_n \\ &= X_n, \end{aligned}$$

since Z_{n+1} is independent of $\mathscr{F}_n$ (condition on independent σ-field reduces to uncondition) and X_n is $\mathscr{F}_n$-measurable (taking out what is known); see Theorem 4.6. Therefore, $\{X_n\}$ is a martingale with respect to $\{\mathscr{F}_n\}$. □

Example 5.5. Let X be an integrable random variable and let $\mathscr{F}_1, \mathscr{F}_2, \dots$ be a filtration. For $n = 1, 2, \dots$, put

$$X_n = E(X|\mathscr{F}_n).$$

By construction, X_n is $\mathscr{F}_n$-measurable. Next, since $|X_n| = |E(X|\mathscr{F}_n)| \le E(|X|\,|\mathscr{F}_n)$, we have by tower property that

$$E(|X_n|) \le E(E(|X|\,|\mathscr{F}_n)) = E(|X|) < \infty.$$

Lastly, by $\mathscr{F}_n \subset \mathscr{F}_{n+1}$ and tower property of conditional expectation,

$$E(X_{n+1}|\mathscr{F}_n) = E(E(X|\mathscr{F}_{n+1})|\mathscr{F}_n) = E(X|\mathscr{F}_n) = X_n.$$

Therefore $\{X_n\}$ is a martingale with respect to $\{\mathscr{F}_n\}$. □

Example 5.6. **(Symmetric Random Walk)** Let X_n be a symmetric random walk, that is,

$$X_n = Z_1 + \cdots + Z_n,$$

where $Z_1, Z_2, \dots$ is a sequence of independent identically distributed random variables such that

$$\mathbb{P}\{Z_n = 1\} = \mathbb{P}\{Z_n = -1\} = \frac{1}{2}$$

(a sequence of coin tosses, for example). Then $\{X_n^2 - n\}$ is a martingale with respect to the filtration $\{\mathscr{F}_n\}$, where

$$\mathscr{F}_n = \sigma(Z_1, \dots, Z_n).$$

It is standard to check the measurability and integrability of X_n similar to Example 5.4 (See Exercise 5.15). Therefore, it remains to verify that $E(X_{n+1}^2 - (n+1)|\mathscr{F}_n) = X_n^2 - n$. Write

$$X_{n+1}^2 = (X_n + Z_{n+1})^2 = Z_{n+1}^2 + 2Z_{n+1}X_n + X_n^2 .$$

Since X_n is $\mathscr{F}_n$-measurable and Z_{n+1} is independent of $\mathscr{F}_n$, we have

$$\begin{aligned} E(X_{n+1}^2|\mathscr{F}_n) - (n+1) &= E(Z_{n+1}^2|\mathscr{F}_n) + 2E(Z_{n+1}X_n|\mathscr{F}_n) + E(X_n^2|\mathscr{F}_n) - (n+1) \\ &= E(Z_{n+1}^2) + 2X_n E(Z_{n+1}) + X_n^2 - (n+1) \\ &= X_n^2 - n , \end{aligned}$$

as desired. □

Since the real world is never fair, to model favorable and unfavorable games of chance, submartingale and supermartingale are introduced as a generalization of martingale.

Definition 5.6. We say that $X_1, X_2, \ldots$ is a **supermartingale (submartingale)** with respect to a filtration $\mathscr{F}_1, \mathscr{F}_2, \ldots$ if

1) X_n is integrable for each $n = 1,2,\ldots$;
2) $X_1, X_2, \ldots$ is adapted to $\mathscr{F}_1, \mathscr{F}_2, \ldots$;
3) $E(X_{n+1}|\mathscr{F}_n) \le X_n$ (respectively, $E(X_{n+1}|\mathscr{F}_n) \ge X_n$) a.s. for each $n = 1,2,\ldots$.

Example 5.7. Let X_n be a sequence of square integrable random variables. If $\{X_n\}$ is a martingale with respect to a filtration $\{\mathscr{F}_n\}$, then $\{X_n^2\}$ is a submartingale with respect to the same filtration. To see this, note that

1) $E|X_n^2| < \infty$ by the square integrability assumption.
2) X_n^2 is $\mathscr{F}_n$ measurable since X_n is. So $\{X_n^2\}$ is adapted to $\{\mathscr{F}_n\}$.
3) Using Jensen's inequality with the convex function $\varphi(x) = x^2$ yields

$$E(X_{n+1}^2|\mathscr{F}_n) \ge [E(X_{n+1}|\mathscr{F}_n)]^2 = X_n^2 .$$

Thus the definition of a submartingale is verified. □

5.3 Trading Strategies and Martingales

Suppose that we take part in a game with Z_n being the *profits (or losses) per unit of investment* in period n. Let $Z_1, Z_2, \ldots$ be a sequence of integrable random variables. If we keep our investment to be \$1 in each period, then our *total profit* X_n after n periods is

$$X_n = Z_1 + \cdots + Z_n . \tag{5.2}$$

Consider the filtration $\{\mathscr{F}_n\}_{n=0,1,\ldots}$ defined by $\mathscr{F}_0 = \{\emptyset, \Omega\}$ and

$$\mathscr{F}_n = \sigma(Z_1, \ldots, Z_n).$$

If we are at the end of the $(n-1)$-th period, our accumulated knowledge will be represented by the σ-field $\mathscr{F}_{n-1}$. Note that $\mathscr{F}_0 = \{\emptyset, \Omega\}$ indicates that nothing is known before the first game. Assuming that the risk free rate is 0, the market is fair if X_n is a **martingale** w.r.t. $\{\mathscr{F}_n\}$. In particular, for $X_0 = 0$ and $n \geq 1$, we have

$$E(X_n|\mathscr{F}_{n-1}) = X_{n-1}\,.$$

That is, we expect that our total profit at time n is on average the same as that at time $n-1$. The market will be favourable to us if X_n is a **submartingale** w.r.t. $\{\mathscr{F}_n\}$,

$$E(X_n|\mathscr{F}_{n-1}) \geq X_{n-1},$$

and unfavourable to us if X_n is a **supermartingale** w.r.t. $\{\mathscr{F}_n\}$,

$$E(X_n|\mathscr{F}_{n-1}) \leq X_{n-1}\,,$$

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

Suppose that we keep the investment to be α_n at period n, $n = 1, 2, \ldots$. (α_n may be zero if we refrain from investing or negative if we do short selling.) At the beginning of the n-th period, we decide the investment α_n with the knowledge about the outcomes in the first $n-1$ periods. Therefore it is reasonable to assume that α_n is $\mathscr{F}_{n-1}$-measurable. Such a sequence $\{\alpha_i\}_{i=1,2,\ldots}$ is called a **previsible** sequence.

Definition 5.7. (Trading Strategy) A **trading strategy** $\alpha_1, \alpha_2, \ldots$ (with respect to a filtration $\mathscr{F}_1, \mathscr{F}_2, \ldots$) is a sequence of previsible random variables, i.e.,α_n is $\mathscr{F}_{n-1}$-measurable for each $n = 1, 2, \ldots$, where $\mathscr{F}_0 = \{\emptyset, \Omega\}$.

If one follows a strategy $\alpha_1, \alpha_2, \ldots$, then α_k unit of investment is made in period k. The profit in the k-th period is $\alpha_k(X_k - X_{k-1})$, and the **total profit** after n periods will be

$$Y_n = \alpha_1(X_1 - X_0) + \cdots + \alpha_n(X_n - X_{n-1}) = \sum_{k=1}^{n} \alpha_k(X_k - X_{k-1})\,. \tag{5.3}$$

We also put $Y_0 = 0$ for convenience.

The following proposition has important consequences for traders. It means that no matter which trading strategy is used,

- a fair market will always be a fair one if the available capital or credit limit (i.e., α_ns) is bounded.
- it is impossible to turn an unfavorable market into a favorable (or vice versa) one if α_ns are non-negative.

In other words, one cannot beat the system!

Proposition 5.1 *Let $\alpha_1, \alpha_2, \ldots$ be a previsible trading strategy, Z_n be the profit of game n, $X_n = \sum_{i=1}^{n} Z_i$, and $Y_n = \sum_{i=1}^{n} \alpha_i Z_i$ be the total profit defined in* (5.3).

1) If $\alpha_1, \alpha_2, \ldots$ is a bounded sequence and $X_0, X_1, X_2, \ldots$ is a martingale, then $Y_0, Y_1, Y_2, \ldots$ is a martingale (a fair market is always fair);

2) *If* $\alpha_1, \alpha_2, \ldots$ *is a non-negative bounded sequence and* $X_0, X_1, X_2, \ldots$ *is a supermartingale, then* $Y_0, Y_1, Y_2, \ldots$ *is a supermartingale (an unfavourable market is always unfavourable no matter which strategy is used).*
3) *If* $\alpha_1, \alpha_2, \ldots$ *is a non-negative bounded sequence and* $X_0, X_1, X_2, \ldots$ *is a submartingale, then* $Y_0, Y_1, Y_2, \ldots$ *is a submartingale (a favourable market is always favourable no matter what strategy is used).*

Proof. First, by the boundedness of α_ks and the integrability and measurability of X_ns, it is standard to check that Y_n is integrable and $\mathscr{F}_n$ measurable.

Since α_n and Y_{n-1} are $\mathscr{F}_{n-1}$-measurable, we can take them out of the expectation conditioned on $\mathscr{F}_{n-1}$ ('taking out what is known'). Thus, we obtain

$$\begin{aligned} E(Y_n|\mathscr{F}_{n-1}) &= E(Y_{n-1} + \alpha_n(X_n - X_{n-1})|\mathscr{F}_{n-1}) \\ &= Y_{n-1} + \alpha_n(E(X_n|\mathscr{F}_{n-1}) - X_{n-1}). \end{aligned}$$

If X_n is a martingale, then

$$\alpha_n(E(X_n|\mathscr{F}_{n-1}) - X_{n-1}) = 0\,,$$

which proves 1). If X_n is a supermartingale and $\alpha_n \geq 0$, then

$$\alpha_n(E(X_n|\mathscr{F}_{n-1}) - X_{n-1}) \leq 0\,,$$

proving 2). Finally, if X_n is a submartingale and $\alpha_n \geq 0$, then

$$\alpha_n(E(X_n|\mathscr{F}_{n-1}) - X_{n-1}) \geq 0\,.$$

and 3) follows.

5.4 Stopping Times

In any market, investors may quit at any time. Let τ be the number of periods an investor engaged before quitting the market. It can be fixed to be $\tau = 10$ (say) if one decides in advance to stop investing after 10 periods no matter what happens. But in general the decision can be made after each period, depending on the knowledge accumulated so far. Therefore τ can be assumed to be a random variable taking values from the set $\{1, 2\ldots\} \cup \{\infty\}$. Infinity is included to cover the theoretical possibility that the investment never stops. At each step n one should be able to decide whether to stop trading or not, i.e. whether $\tau = n$. Therefore the event $\{\tau = n\}$ should be in the σ-field $\mathscr{F}_n$ that represents our knowledge at time n. This gives rise to the following definition.

Definition 5.8. (Stopping Time) A random variable τ with values in the set $\{1, 2\ldots\}$ $\cup\{\infty\}$ is called a **stopping time** (with respect to a filtration $\{\mathscr{F}_n\}$) if for each $n = 1, 2, \ldots$

$$\{\tau = n\} \in \mathscr{F}_n\,.$$

Remark 5.3. From Definition 5.8, if τ is a stopping time, then $\{\tau \le n\} = \bigcup_{k=1}^{n}\{\tau = k\} \in \mathscr{F}_n$ and $\{\tau > n\} = \{\tau \le n\}^c \in \mathscr{F}_n$.

Example 5.8. **(First Hitting Time)** Suppose that a coin is tossed repeatedly and you win (lose) \$1, if a head (tail) is shown. Suppose that you start investing with, say, \$5 in your pocket and decide to play until you have \$10 or you lose everything. If X_n is the amount you have at step n, then the time when you stop the game is

$$\tau = \inf\{n : X_n = 10 \text{ or } 0\},$$

and is called the **first hitting time** of the random sequence X_n. This τ is a stopping time with respect to the filtration $\mathscr{F}_n = \sigma(X_1,\ldots,X_n)$. To see this, note that

$$\{\tau = n\} = \{0 < X_1 < 10\} \cap \cdots \cap \{0 < X_{n-1} < 10\} \cap \{X_n = 10 \text{ or } 0\}.$$

Now, each of the sets on the right-hand side belongs to $\mathscr{F}_n$, so does their intersection. This implies that

$$\{\tau = n\} \in \mathscr{F}_n$$

for each n, i.e., τ is a stopping time. □

Example 5.9. Define the *last* hitting time $\tau = \sup\{n : X_n = 10 \text{ or } 0\}$. Intuitively, τ is not a stopping time since for any fixed time n we never know if X will hit 0 or 10 in the future. Formally, $\{\tau = n\} = \{X_n = 0 \text{ or } 10\} \cap (\bigcap_{k=1}^{\infty}\{X_{n+k} \neq 0 \text{ or } 10\})$ involves $\{X_{n+k}\}_{k=1,2,\ldots}$, which is is clearly not in $\mathscr{F}_n$. □

Let $\{X_n\}$ be a sequence of random variables adapted to a filtration $\{\mathscr{F}_n\}$ and τ be a stopping time (with respect to the same filtration). Suppose that X_n represents your profits (or losses) after n periods. If you decide to quit after τ periods, then your total profit will be X_τ. In this case your profits after n rounds will in fact be $X_{\tau\wedge n}$, where $a \wedge b = \min(a,b)$ denotes the smaller one of the two numbers a and b.

Definition 5.9. (Stopped sequence/process) We call $X_{\tau\wedge n}$ the sequence **stopped** at τ. It is often denoted by X_n^τ. Precisely, for each $\omega \in \Omega$, we have

$$X_{\tau\wedge n}(\omega) = X_{\tau(\omega)\wedge n}(\omega).$$

□

Theorem 5.2. (Stopped process is adapted) *If X_n is a sequence of random variables adapted to a filtration $\{\mathscr{F}_n\}$, then so is the sequence $X_{\tau\wedge n}$.*

Proof. We need to show that $X_{\tau\wedge n}$ is adapted to $\{\mathscr{F}_n\}$, i.e., $X_{\tau\wedge n}$ is $\mathscr{F}_n$ measurable. Let $B \subseteq \mathbb{R}$ be a Borel set. We can write

$$\{X_{\tau\wedge n} \in B\} = \{X_n \in B, \tau > n\} \cup \bigcup_{k=1}^{n}\{X_k \in B, \tau = k\},$$

where

$$\{X_n \in B, \tau > n\} = \{X_n \in B\} \cap \{\tau > n\} \in \mathscr{F}_n ,$$

and for each $k = 1, \ldots, n$,

$$\{X_k \in B, \tau = k\} = \{X_k \in B\} \cap \{\tau = k\} \in \mathscr{F}_k \subseteq \mathscr{F}_n .$$

It follows that for each n, $\{X_{\tau\wedge n} \in B\} \in \mathscr{F}_n$, as required. □

We know from Proposition 5.1 that it is impossible to turn a fair market into an unfair market, an unfavourable market into a favourable one, or vice versa, no matter which trading strategy is used. The next proposition shows that this cannot be achieved using any stopping time (the idea is that stopping is also a trading strategy).

Proposition 5.3 (Stopped Martingale is a Martingale) *Let τ be a stopping time.*

1. *If X_n is a martingale, then so is $X_{\tau\wedge n}$.*
2. *If X_n is a supermartingale, then so is $X_{\tau\wedge n}$.*
3. *If X_n is a submartingale, then so is $X_{\tau\wedge n}$.*

Proof. This is in fact a consequence of Proposition 5.1. Given a stopping time τ, we set

$$\alpha_t = 1_{\{\tau \geqslant t\}} = \begin{cases} 1 & \text{if } \tau \geq t \\ 0 & \text{if } \tau < t \end{cases} .$$

Notice that α_t is previsible (that is, α_t is $\mathscr{F}_{t-1}$-measurable). This is because the inverse image $\{\omega : \alpha_t(\omega) \in B\} = \{\alpha_t \in B\}$ of any Borel set $B \in \mathscr{B}$ is equal to

$$\{\alpha_t \in B\} = \begin{cases} \emptyset & \text{if} \quad 0,1 \notin B \\ \Omega & \text{if} \quad 0,1 \in B \\ \{\tau \geq t\} = \{\tau > t-1\} & \text{if } 1 \in B, 0 \notin B \\ \{\tau < t\} = \{\tau \leq t-1\} & \text{if } 0 \in B, 1 \notin B \end{cases} ,$$

and that $\emptyset$, Ω, $\{\tau > t-1\}$ and $\{\tau \leq t-1\}$ belong to $\mathscr{F}_{t-1}$. Setting $X_0 = 0$, we can express $X_{\tau\wedge n}$ as

$$X_{\tau\wedge n} = \alpha_1(X_1 - X_0) + \cdots + \alpha_n(X_n - X_{n-1}) = \sum_{t=1}^{n} \alpha_t Z_t ,$$

where $Z_t = X_t - X_{t-1}$. Therefore, Proposition 5.3 follows from Proposition 5.1 with $Y_n = X_{\tau\wedge n}$.

Example 5.10. **(Doubling Strategy)** (You could try to beat the system if you had unlimited capital!) Suppose the profit of a unit of investment at the n-th trading day is denoted by Z_n, which takes value +1 (up) or -1 (down). First, you invest \$1. If you win (the market goes up), you quit. If you lose (the market goes down), you invest twice as much as last time, and so on until you win. Mathematically, your trading strategy is

$$\alpha_t = 2^{t-1} 1_{\{\tau \geq t\}},$$

where

$$\tau = \min\{n : Z_n = 1 \quad \text{(up)}\}$$

is your stopping time. Since $\{\tau \geq t\} = \{\tau \leq t-1\}^c \in \mathscr{F}_{t-1} \triangleq \sigma(Z_1, \ldots, Z_{t-1})$, $\{\alpha_t\}$ is previsible, and thus a valid trading strategy. It can be shown that $\mathbb{P}(\{\tau < \infty\}) = 1$, that is, the market will go up someday ($Z_n = 1$ for some n) with probability one (See Exercise 5.19).

Let $X_0 = 0$ and

$$X_n = Z_1 + 2Z_2 + \cdots + 2^{n-1} Z_n,$$

for $n \geq 1$. Your profit after n periods is

$$Y_n = \sum_{t=1}^{n} \alpha_t Z_t = \sum_{t=1}^{n} 2^{t-1} 1_{\{\tau \geq t\}} Z_t = \sum_{t=1}^{n \wedge \tau} 2^{t-1} Z_t = X_{\tau \wedge n}.$$

From Propositions 5.1 and 5.3, it can be shown that $\{X_n\}$ and $\{X_{\tau \wedge n}\}$ are martingales (See Exercise 5.19). Therefore, we have (see Exercise 5.12)

$$E(Y_n) = E(X_{\tau \wedge n}) = E(X_0) = 0, \tag{5.4}$$

which comes with no surprise because we cannot turn a fair game into an unfair game.

It turns out to be surprising if we consider Y_τ, the total profit if we continue to play the game until the first head appears. You would be bound to win eventually because

$$Y_\tau = (Z_1 + 2Z_2 + \cdots + 2^{\tau-2} Z_{\tau-1}) + 2^{\tau-1} Z_\tau = -1 - 2 - \cdots - 2^{\tau-2} + 2^{\tau-1} = 1,$$

for any τ. That is, the profit in the last period is able to cover the loss in the previous $\tau - 1$ periods, with a net gain of \$1. Note that this results does not contradicts Propositions 5.1, which only cover the case of fixed time Y_n but not random time Y_τ.

Although $E(Y_\tau) = Y_\tau = 1 > 0 = E(Y_n)$, it still doesn't mean that we can turn a fair game to a favorable game. Indeed, the expected loss just before the ultimate win is infinite, i.e.,

$$E(Y_{\tau-1}) = -\infty.$$

To see this, note that the probability that the trading terminates at step n is

$$\mathbb{P}(\{\tau = n\}) = \mathbb{P}(n-1 \text{ downs followed by an up at step } n) = \frac{1}{2^n}.$$

Therefore,

$$
\begin{aligned}
E(Y_{\tau-1}) &= \sum_{n=1}^{\infty} E(Y_{n-1}|\{\tau=n\})\mathbb{P}(\{\tau=n\}) \\
&= \sum_{n=2}^{\infty} (-1-2-\cdots-2^{n-2})\frac{1}{2^n} \qquad (5.5) \\
&= -\sum_{n=2}^{\infty} \frac{2^{n-1}-1}{2^n} = \sum_{n=2}^{\infty} \frac{1}{2^n} - \sum_{n=2}^{\infty} \frac{1}{2} \\
&= \frac{1}{2} - \sum_{n=2}^{\infty} \frac{1}{2} = -\infty,
\end{aligned}
$$

where (5.5) follows from the fact that $Y_{\tau-1} = Y_{n-1} = -1-2-\cdots-2^{n-2}$ if $\tau = n$. In other words, the game is still *fair* in the sense that, you are expected to hold ∞ amount of capital in order to win merely \$1. $\square$

5.5 Optional Stopping Theorem

If $\{X_n\}$ is a martingale, then it can be shown that (see Exercise 5.12)

$$E(X_n) = E(X_1)$$

for each n. In Example 5.10, the martingale $\{Y_n\}$ satisfies $E(Y_\tau) = Y_\tau = 1$, since the profit Y_τ is always 1 when we stop at τ. On the other hand, $E(Y_1) = E(Z_1) = 0$ is clearly not equal to $E(Y_\tau)$. Thus, in general we don't have $E(X_\tau) = E(X_1)$ for any martingale $\{X_n\}$. However, if the equality

$$E(X_\tau) = E(X_1) \qquad (5.6)$$

does hold, it can be very useful. The optional stopping theorem provides sufficient conditions for (5.6) to hold.

Theorem 5.4. (Optional Stopping Theorem) *Let X_n be a martingale and τ be a stopping time with respect to a filtration $\{\mathscr{F}_n\}_{n=1,2,\ldots}$ such that the following conditions hold:*

1) $\tau < \infty$ a.s.,
2) X_τ is integrable,
3) $|X_{\tau\wedge n}| \leq Y$ a.s. for some $Y \in \mathscr{L}^1$ (i.e., $E(|Y|) < \infty$) and all $n \in \mathbb{Z}^+$.

Then

$$E(X_\tau) = E(X_1).$$

Proof. Note that X_τ is undefined when $\tau = \infty$. Therefore $\tau < \infty$ a.s. is assumed so that X_τ makes sense. Also, 2) must hold in order to define $E(X_\tau)$. Finally, as $\tau < \infty$ a.s, $X_{\tau\wedge n} \to X_\tau$ a.s. as $n \to \infty$. By 3), the Dominated Convergence Theorem (DCT) implies that as $n \to \infty$,

$$\lim_{n\to\infty} E(X_{\tau\wedge n}) = E(X_\tau)\,. \tag{5.7}$$

On the other hand, since $X_{\tau\wedge n}$ is a martingale by Proposition 5.3, we have

$$E(X_{\tau\wedge n}) = E(X_{\tau\wedge 1}) = E(X_1)\,. \tag{5.8}$$

Combining (5.7) and (5.8), the result follows. □

Example 5.11. **(Expected first hitting time of a random walk)** Let X_n be a symmetric random walk as in Example 5.6 and let K be a positive integer. Define the first hitting time (of $\pm K$ by X_n) to be

$$\tau = \min\{n : |X_n| = K\}.$$

It can be checked that τ is a stopping time (see Exercise 5.17). It has been shown in Example 5.6 that $X_n^2 - n$ is a martingale. If the optional stopping theorem can be applied, then

$$E(X_\tau^2 - \tau) = E(X_1^2 - 1) = 0.$$

This allows us to find the expectation

$$E(\tau) = E(X_\tau^2) = K^2,$$

since $|X_\tau| = K$.

Next we verify Conditions 1-3 of the optional stopping theorem on the martingale $Y_n \triangleq X_n^2 - n$ and stopping time τ.

1) First we show that $\mathbb{P}(\{\tau < \infty\}) = 1$. Before the random walk is stopped, it must be within $[-(K-1), K-1]$. Consider the events

$$\begin{aligned} A_n &= \{\textbf{after} \text{ time } 2Kn, \text{ there are consecutive +1s of length } 2K\} \\ &= \{Z_j = +1,\ j = 2Kn+1, \ldots, 2K(n+1)\}\,. \end{aligned}$$

Note that $P(A_n) = \frac{1}{2^{2K}}$ for all integer $n \geq 0$. Also, $\{A_n\}_{n=0,1,2,\ldots}$ are independent events since they involve different Z_js.
If event A_n occurs, then the random walk must exceed K and thus stop before time $2K(n+1)$. In other words, if $\tau > 2K(n+1)$, then none of $A_0, A_1, \ldots, A_n$ happens, i.e., $\{\tau > 2K(n+1)\} \subseteq \bigcap_{j=1}^n A_j^c$. Thus,

$$\begin{aligned} \mathbb{P}\{\tau > 2K(n+1)\} &\leq \mathbb{P}\left\{\bigcap_{i=0}^n A_i^c\right\} = \prod_{i=0}^n \left(1 - \frac{1}{2^{2K}}\right) \qquad \text{(independence of } \{A_i\}\text{)} \\ &= \left(1 - \frac{1}{2^{2K}}\right)^{n+1}. \end{aligned} \tag{5.9}$$

As $\{\tau > 2K(n+1)\}$ is a decreasing sequence in n, Theorem 3.5 implies that

$$\mathbb{P}\{\tau=\infty\} \;=\; \mathbb{P}\left(\bigcap_{n=0}^{\infty}\{\tau>2K(n+1)\}\right)=\lim_{n\to\infty}\mathbb{P}\{\tau>2K(n+1)\}$$
$$\stackrel{(5.9)}{=}\lim_{n\to\infty}\left(1-\frac{1}{2^{2K}}\right)^{n+1}=0\,,$$

giving $\mathbb{P}(\{\tau<\infty\})=1$.

2) We need to show that

$$E(|Y_\tau|)=E(|X_\tau^2-\tau|)<\infty.$$

Indeed, writing $n=2Kj+k$, we have

$$\begin{aligned}
E(\tau) &= \sum_{n=1}^{\infty} n\mathbb{P}\{\tau=n\} && (5.10)\\
&= \sum_{j=0}^{\infty}\sum_{k=1}^{2K}\{2Kj+k\}\mathbb{P}\{\tau=2Kj+k\}\\
&\le \sum_{j=0}^{\infty}\sum_{k=1}^{2K}2K(j+1)\mathbb{P}\{\tau>2Kj\}\\
&\le \sum_{j=0}^{\infty}\sum_{k=1}^{2K}2K(j+1)\left(1-\frac{1}{2^{2K}}\right)^{j} && \text{(by (5.9))}\\
&= 4K^2\sum_{j=0}^{\infty}(j+1)\left(1-\frac{1}{2^{2K}}\right)^{j}<\infty, && \text{(inner summation has no } k\text{)}
\end{aligned}$$

since the series $\sum_{j=1}^{\infty}(j+1)q^j$ is convergent for any $q\in(-1,1)$. Moreover, $X_\tau^2=K^2$, so

$$\begin{aligned}
E(|X_\tau^2-\tau|) &\le E(X_\tau^2)+E(\tau)\\
&= K^2+E(\tau)\\
&< \infty.
\end{aligned}$$

3) As $|X_{\tau\wedge n}|<K$ and $\tau\wedge n\le\tau$ a.s., we have $|Y_{\tau\wedge n}|=|X_{\tau\wedge n}^2+\tau\wedge n|\le K^2+\tau$ a.s. Thus, 3) is satisfied because $E(K^2+\tau)<\infty$ from (5.10). □

Example 5.12. For the Doubling Strategy in Example 5.10, we do not have $E(X_\tau)=E(X_1)$. On the other hand, the first two conditions of the optional stopping theorem are satisfied. Specifically, Condition 1 is shown in Exercise 5.20; Condition 2 follows from the fact that $X_\tau=1$ and that any constant is integrable. However, Condition 3 of the OST is not satisfied for $X_{\tau\wedge n}$. To see this, note that if $\tau>n$, then $X_{\tau\wedge n}=X_n=-1-2-\cdots-2^{n-1}=1-2^n$. In other words, $X_{\tau\wedge n}$ can take arbitrarily negative values with positive probability ($\mathbb{P}\{\tau>n\}>0$ for all n). Thus there cannot be any integrable Y that bounds $|X_{\tau\wedge n}|$ a.s. for all n.

□

Example 5.13. Let X_n be a symmetric random walk and $\{\mathscr{F}_n\}_{n\geq 1}$ be the filtration defined in Example 5.6. Denote by τ the smallest n such that $|X_n| = K$ as in Example 5.11. It is standard to verify that

$$Y_n = (-1)^n \cos[\pi(X_n + K)]$$

is a martingale (see Exercise 5.21). We will show that Y_n and τ satisfy the conditions of the optional stopping theorem and apply the theorem to find $E[(-1)^\tau]$.

If optional stopping theorem is applicable, then

$$E(Y_\tau) = E(Y_1). \tag{5.11}$$

Since $X_\tau = K$ or $-K$, we have

$$Y_\tau = (-1)^\tau \cos[\pi(K + X_\tau)] = (-1)^\tau. \tag{5.12}$$

Also, we have

$$E(Y_1) = -\frac{1}{2}(\cos[\pi(1+K)] + \cos[\pi(-1+K)]) = \cos(\pi K) = (-1)^K. \tag{5.13}$$

Combining (5.11), (5.12) and (5.13), it follows that

$$E[(-1)^\tau] = (-1)^K.$$

Thus it remains to verify that $\{Y_n\}$ and τ satisfy Conditions 1-3 of the optional stopping theorem.
Condition 1 has been verified in Example 5.11. Conditions 2 and 3 hold since the cosine function $|Y_n| \leq 1$; thus $E(|Y_\tau|) \leq 1 < \infty$ and $|Y_{\tau\wedge n}| \leq 1$ for all $n \in \mathbb{Z}^+$. □

Example 5.14. **(Asymmetric Simple Random Walk)** The asymmetric simple random walk is defined by

$$X_n = Z_1 + \cdots + Z_n\,,$$

where $\mathbb{P}(Z_i = 1) = p$ and $\mathbb{P}(Z_i = -1) = 1 - p$ for all i. W.L.O.G. assume that $p > 1/2$. Let $\tau_x = \inf\{n : X_n = x\}$ and $T = \min(\tau_a, \tau_b)$, $a < 0 < b$, be the first time of hitting the boundary of the interval (a,b). Define

$$\varphi(x) = \{(1-p)/p\}^x.$$

We have the following results by the OST. We only illustrate the applications of OST and the verification of conditions is left to Exercise 5.26. We have

1)

$$\mathbb{P}(\tau_a < \tau_b) = \frac{\varphi(b) - 1}{\varphi(b) - \varphi(a)}\,. \tag{5.14}$$

2) If $a < 0$, then $\mathbb{P}(\tau_a < \infty) = \varphi(-a)$.

3) If $b > 0$, then $E(\tau_b) = b/(2p-1)$.

Proof. 1. First, it can be checked that $\varphi(X_n)$ is a martingale (Exercise 5.26). Using OST on $\varphi(X_T)$, we have

$$1 = \varphi(0) = E(\varphi(X_T)) = \mathbb{P}(\tau_a < \tau_b)\varphi(a) + \mathbb{P}(\tau_b < \tau_a)\varphi(b)\,. \quad (5.15)$$

Note that $\mathbb{P}(\tau_a = \tau_b) = 0$ as the martingale cannot stop at a and b at the same time. Together with $\mathbb{P}(\tau_a < \tau_b) + \mathbb{P}(\tau_b < \tau_a) = 1$, solving for $\mathbb{P}(\tau_a < \tau_b)$ gives (5.14).

2. Note that $\varphi(b) = [(1-p)/p]^b \to 0$ as $b \to \infty$, and $1/\varphi(a) = \varphi(-a)$. Thus 2) follows from taking $b \to \infty$ in (5.14).
3. Applying OST to $X_{\tau_b} - (2p-1)\tau_b$ (see Exercise 5.26 for the fact that $X_n - (2p-1)n$ is a martingale), we have

$$b - (2p-1)E(\tau_b) = E(X_{\tau_b} - (2p-1)\tau_b) = E(X_1 - (2p-1)) = 0,$$

i.e., $E(\tau_b) = b/(2p-1)$. □

Example 5.15. **(Symmetric Simple Random Walk; one-sided boundary)** Consider $E(\tau_b) = b/(2p-1)$ from Example 5.14 3). If we take $p \to 1/2$, then

$$E(\tau_b) \to \infty\,, \quad (5.16)$$

suggesting that a symmetric random walk, on average, takes ∞ amount of time to hit b. However, the convergence (5.16) is only heuristic but not rigorously justified. Indeed, it is not easy to justify (5.16) by MCT or DCT since now it is not τ_b that depends on p, but the probability measure underlying the expectation depends on p.

For a more rigorous argument, consider the simple random walk $X_n = Z_1 + \cdots + Z_n$, where $\mathbb{P}(Z_i = 1) = \mathbb{P}(Z_i = -1) = 1/2$ for all i, and $T = \min(\tau_a, \tau_b)$, $a < 0 < b$. Since $\{X_n\}$ is a martingale and both X_T and $X_{T\wedge n}$ are bounded in the interval $[a, b]$, we can use OST on X_T to yield (again, $\mathbb{P}(\tau_a = \tau_b) = 0$)

$$0 = E(X_T) = a\mathbb{P}(\tau_a < \tau_b) + b\mathbb{P}(\tau_a > \tau_b)\,.$$

Combining with $\mathbb{P}(\tau_a < \tau_b) + \mathbb{P}(\tau_a > \tau_b) = 1$, we obtain (two equations two unknowns)

$$\mathbb{P}(\tau_a < \tau_b) = \frac{b}{b-a}\,, \quad \mathbb{P}(\tau_a > \tau_b) = \frac{-a}{b-a}\,. \quad (5.17)$$

Next, using OST on $X_T^2 - T$ (the integrability of T can be justified similar to (5.10)), we have $E(X_T^2 - T) = E(X_1^2 - 1) = 0$, implying that

$$E(T) = E(X_T^2) = a^2\mathbb{P}(\tau_a < \tau_b) + b^2\mathbb{P}(\tau_a > \tau_b) \overset{(5.17)}{=} \frac{a^2 b}{b-a} + \frac{-ab^2}{b-a} = -ab\,. \quad (5.18)$$

Since τ_a increases to ∞ a.s. as $a \to -\infty$, we have $T = \min(\tau_a, \tau_b)$ increases to τ_b as $a \to -\infty$. Therefore, MCT implies

$$E(\tau_b) = E\left(\lim_{a\to-\infty} T\right) \stackrel{(MCT)}{=} \lim_{a\to-\infty} E(T) \stackrel{(5.18)}{=} -\lim_{a\to-\infty} ab = \infty\,,$$

justifying (5.16). □

5.6 Martingale Convergence Theorem

Definition 5.10. (Upcrossing Strategy) Given an adapted sequence of r.v. $X_1, X_2, \ldots$, and two real numbers $a < b$, the **upcrossing strategy** $\alpha_1, \alpha_2, \ldots$ is given by

$$\alpha_1 = 0\,, \quad \alpha_{n+1} = \begin{cases} 1 \text{ if } \alpha_n = 0 \text{ and } X_n < a \\ 1 \text{ if } \alpha_n = 1 \text{ and } X_n \le b \\ 0 \qquad \text{otherwise} \end{cases} \tag{5.19}$$

for $n = 1, 2 \ldots$. When $\alpha_k = 1$ and $\alpha_{k+1} = 0$, we say that there is an **upcrossing** at time k. Let the **upcrossing sequence** $u_1, u_2, \ldots$ be the time indices where upcrossings occur. The number of upcrossings made up to time n is defined by

$$U_n[a, b] = \max\{k : u_k \le n\}\,.$$

Remark 5.4.
The upcrossing strategy is related to the trading principle of *buy-low-sell-high* in finance. Imagine

- X_n is the stock price at time n.
- $\alpha_n = 1$ means holding one unit of the stock, 0 otherwise.

To be specific, when the stock price X falls below \$ a, we hold the stock. We wait until X reaches \$ b, where $b > a$, and sell the stock. Then we wait for the next chance that X falls below \$ a, and so forth. The upcrossing corresponds to selling the stock and completing a cycle of trading. Note that each upcrossing increases the total wealth by at least \$ $b - a$.

Theorem 5.5. *The upcrossing strategy* $\{\alpha_n\}$ *in Definition 5.10 is previsible.*

Proof. Intuitively, $\{\alpha_n\}$ is previsible because following the strategy (5.19), α_{n+1} is governed by the knowledge up to time n. To formally prove that α_n is $\mathscr{F}_{n-1}$ measurable for each n, we use mathematical induction. First, note that by definition $\mathscr{F}_0 = \{\emptyset, \Omega\}$ and the constant $\alpha_1 = 0$ is $\mathscr{F}_0$ measurable since for any Borel set $B \in \mathscr{B}$, we have

$$\{\alpha_1 \in B\} = \begin{cases} \Omega \text{ if } 0 \in B \\ \emptyset \text{ if } 0 \notin B\,. \end{cases}$$

Next, suppose that α_k is $\mathscr{F}_{k-1}$ measurable, we have for any Borel set $B \in \mathscr{B}$ that

$$\{\alpha_{k+1} \in B\} = \begin{cases} \Omega & \text{if } \{0,1\} \subset B \\ \big(\{\alpha_k = 0\} \cap \{X_k < a\}\big) \cup \big(\{\alpha_k = 1\} \cap \{X_k \le b\}\big) & \text{if } 1 \in B, 0 \notin B \\ \text{the complement of the above} & \text{if } 1 \notin B, 0 \in B \\ \emptyset & \text{if } 0 \notin B, 1 \notin B\,. \end{cases}$$

Since $\{\alpha_k = 0\}$, $\{\alpha_k = 1\}$, $\{X_k < a\}$ and $\{X_k \le b\}$ are $\mathscr{F}_k$ measurable by induction assumption, it follows that α_{k+1} is $\mathscr{F}_k$ measurable. □

Theorem 5.6. (Upcrossings Inequality) *Let $\{X_n\}_{n=1,2,\ldots}$ be a supermartingale and $a < b$. Then,*

$$(b-a)E(U_n[a,b]) \le E((X_n - a)^-)\,, \tag{5.20}$$

where $x^- = \max(0, -x)$.

Proof. Let $X_0 = 0$ and

$$Y_n = \alpha_1(X_1 - X_0) + \cdots + \alpha_n(X_n - X_{n-1})\,,$$

be the total profit at time n. Note from Proposition 5.1 that Y_n is a supermartingale.

Suppose that $U_n[a,b] = k$ at time n, with $0 < u_1 < \cdots < u_k \le n$. Since each upcrossing increases the total profit by $b - a$, we have

$$Y_{u_i} - Y_{u_{i-1}} \ge b - a\,, \tag{5.21}$$

for $i = 1, \ldots, k$, where $u_0 \triangleq 0$. At time n, if $X_n \ge a$, then $Y_n - Y_{u_k} \ge 0 = -(X_n - a)^-$. On the other hand, if $X_n < a$, then $Y_n - Y_{u_k} \ge X_n - a = -(X_n - a)^-$. Putting together, we have

$$Y_n - Y_{u_k} \ge -(X_n - a)^-\,. \tag{5.22}$$

Summing over (5.22) and (5.21) for $i = 1, \ldots, k$, we have

$$Y_n \ge (b-a)U_n[a,b] - (X_n - a)^-\,. \tag{5.23}$$

Taking expectation on both sides, we get

$$E(Y_n) \ge (b-a)E(U_n[a,b]) - E((X_n - a)^-)\,.$$

Lastly, since Y_n is a supermartingale, we have $0 \ge E(Y_1) \ge E(Y_n)$, thus

$$E((X_n - a)^-) \ge (b-a)E(U_n[a,b])\,,$$

completing the proof. □

Example 5.16. For a random walk $X_n = Z_1 + \ldots + Z_n$, $Z_i \overset{a.s.}{\sim} N(0,1)$. We have $X_n \sim N(0,n)$. Suppose that the upcrossing strategy is performed with $a = -1$ and $b = 1$. Using (5.20) the expected number of upcrossing by time n satisfies

$$
\begin{aligned}
(1-(-1))E(U_n[-1,1]) &\leq \int (x-(-1))^{-}\phi(x,0,n)\,dx \\
\Leftrightarrow E(U_n[-1,1]) &\leq -\frac{1}{2}\int_{-\infty}^{-1}(x+1)\phi(x,0,n)\,dx \\
&= -\frac{1}{2}\left(\int_{-\infty}^{-1} xe^{-x^2/2n}/\sqrt{2\pi n}\,dx + \int_{-\infty}^{-1} e^{-x^2/2n}/\sqrt{2\pi n}\,dx\right) \\
&= \frac{\sqrt{n}e^{-\frac{1}{2n}}}{2\sqrt{2\pi}} - \frac{1}{2}\Phi(-\frac{1}{\sqrt{n}}),
\end{aligned}
$$

where $\phi(x,0,n) = e^{-x^2/2n}/\sqrt{2\pi n}$ is the p.d.f. of the Normal random variable with mean 0, and variance n and Φ is the c.d.f. of the standard Normal random variable. □

Theorem 5.7. (Doob's Martingale Convergence Theorem) *Suppose $\{X_n\}_{n=1,\ldots}$ is a supermartingale with respective to the filtration $\{\mathscr{F}_n\}_{n=1,\ldots,}$ such that*

$$\sup_n E(|X_n|) = M < \infty .$$

There is an integrable random variable X such that

$$X_n \to X\,, \quad a.s. \quad \text{as } n \to \infty .$$

Remark 5.5. Note that this theorem is also valid for martingales and submartingales since

- every martingale is a supermartingale.
- $-X_n$ is a supermartingale if X_n is a submartingale.

Proof. **(Theorem 5.7)** First we show that $\mathbb{P}(\lim_{n\to\infty} U_n[a,b] < \infty) = 1$. By the Upcrossing Inequality (5.20), for every $-\infty < a < b < \infty$,

$$E(U_n[a,b]) \leq \frac{E((X_n-a)^-)}{b-a} \leq \frac{M+|a|}{b-a} < \infty .$$

Since $U_n[a,b]$ is non-decreasing in n, the MCT implies

$$E\left(\lim_{n\to\infty} U_n[a,b]\right) = \lim_{n\to\infty} E(U_n[a,b]) \leq \frac{M+|a|}{b-a} < \infty .$$

This implies that

$$\mathbb{P}\left(\lim_{n\to\infty} U_n[a,b] < \infty\right) = 1\,, \tag{5.24}$$

otherwise the expectation must be infinity. To show that $X_n \overset{a.s.}{\to} X$, we will use (5.24) to establish

$$\mathbb{P}\left(\omega : \liminf_n X_n(\omega) = \limsup_n X_n(\omega)\right) = 1\,. \tag{5.25}$$

It is equivalent to showing that $\mathbb{P}(B) = 0$ where

$$B = \left\{ \liminf_n X_n < \limsup_n X_n \right\} \subset \Omega \,. \tag{5.26}$$

Note that if $\omega \in B$ then for some rational numbers a_ω and b_ω,

$$\liminf_n X_n(\omega) < a_\omega < b_\omega < \limsup_n X_n(\omega) \,. \tag{5.27}$$

As (5.26) implies that $\{X_n\}$ has upcrossing from a_ω to b_ω infinity many times, we have

$$B \subseteq \bigcup_{a,b \in \mathbb{Q}} \left\{ \lim_{n\to\infty} U_n[a,b] = \infty \right\} \,.$$

Therefore, we have from (5.24) that

$$\mathbb{P}(B) \le \sum_{a,b\in\mathbb{Q}} \mathbb{P}\left(\lim_{n\to\infty} U_n[a,b] = \infty \right) = 0 \,. \tag{5.28}$$

Note that introducing the countable rational numbers a_ω, b_ω is crucial to obtain a countable sum of zeros in (5.28), which remains zero.

It remains to show that the limit X is an integrable random variable. By Fatou's lemma,

$$\begin{aligned} E(|X|) &= E(\lim_n |X_n|) = E(\liminf_n |X_n|) \quad \text{(since the limit exists)} \\ &\le \liminf_n E(|X_n|) \qquad\qquad\qquad\qquad \text{(Fatou's lemma)} \\ &\le \sup_n E(|X_n|) = M < \infty \,, \end{aligned}$$

completing the proof. □

Example 5.17. Let $Z_1, Z_2, \ldots$ be i.i.d. random variables with $E(Z_m) = 1$ and $\mathbb{P}(Z_m = 1) < 1$. It is easy to verify that $X_n = \prod_{m\le n} Z_m$ is a martingale and $E(X_n) = 1$ for all n. Therefore, Theorem 5.7 implies that $X_n \overset{a.s.}{\to} X$ for some r.v. X. Note that for positive A and B, if $A > \delta$ and $B > \varepsilon$, then $AB > \delta\varepsilon$. Speaking in set theory, we have

$$\begin{aligned} \{A > \delta\} \cap \{B > \varepsilon\} \subset \{AB > \delta\varepsilon\} \,, \quad \text{or} \\ \mathbb{P}(AB > \delta\varepsilon) \ge \mathbb{P}(\{A > \delta\} \cap \{B > \varepsilon\}) \,. \end{aligned} \tag{5.29}$$

Now, putting $A = X_n$ and $B = |Z_{n+1} - 1|$, we have

$$\begin{aligned} \mathbb{P}(|X_{n+1} - X_n| > \delta\varepsilon) &= \mathbb{P}(X_n |Z_{n+1} - 1| > \delta\varepsilon) \\ &\ge \mathbb{P}(\{X_n > \delta\} \cap \{|Z_{n+1} - 1| > \varepsilon\}) \qquad \text{(by (5.29))} \\ &= \mathbb{P}(X_n > \delta)\mathbb{P}(|Z_{n+1} - 1| > \varepsilon) \qquad \text{(by independence)} \end{aligned}$$

For any $\delta > 0$ and $\varepsilon > 0$, the L.H.S. of the above converges to 0 by the existence of limit of X_n. For R.H.S, $\mathbb{P}(|Z_{n+1} - 1| > \varepsilon)$ is strictly positive for some $\varepsilon > 0$ by the assumption $\mathbb{P}(Z_m = 1) < 1$. Therefore, $\mathbb{P}(X_n > \delta) \to 0$ for any $\delta > 0$. As by construction, X_n is positive, we conclude that

$$X_n \overset{a.s.}{\to} 0\,.$$

Therefore, besides $X_n(\omega) = n \times 1_{\{\omega < 1/n\}}$ in Example (3.9), we have another example of $E(X_n) = 1$ for all n but $X_n \overset{a.s.}{\to} 0$. $\square$

5.7 Continuous Time Processes, Martingales and Stopping Times

We have introduced martingales and stopping times in the discrete time context. In fact, similar notions can be defined in continuous time. In this section we define continuous time processes, martingales and stopping times, and state some useful results. In Chapter 6 we will study in detail an important continuous time stochastic process, Brownian Motion.

Definition 5.11. (Continuous Time Filtration) The sequence of σ-fields $\{\mathscr{F}_t\}_{t\geq 0}$ is called a filtration if $\mathscr{F}_s \subseteq \mathscr{F}_t$ for all $s \leq t$.

Definition 5.12. (Continuous Time Stochastic Process) A **continuous time Stochastic Process** is a random function $\{X_t\}_{t\geq 0}$ from some abstract measurable space $(\Omega, \mathscr{F})$ to the measurable space $(\mathscr{C}, \sigma(\mathscr{C}))$, where $\mathscr{C}$ is some space of functions. In other words, given $\omega \in \Omega$, $X_t(\omega)$ a real-valued function of t. Some commonly used space of function includes the space of k-th time differentiable functions $\mathscr{C}^k$ $(k \geq 1)$, the space of continuous functions $\mathscr{C}^0$, and the space of right continuous functions with left limits, $\mathscr{D}$.

Definition 5.13. (Natural Filtration) If the stochastic process $\{X_t\}$ satisfies $X_t \in \mathscr{F}_t$ for all $t \geq 0$, then $\{X_t\}$ is **adapted** to $\{\mathscr{F}_t\}$. If $\mathscr{F}_t^X = \sigma\{X_s, s \leq t\}$, then the stochastic process $\{X_t\}$ is adapted to $\{\mathscr{F}_t^X\}$ by construction. Hence, $\{\mathscr{F}_t^X\}$ is called the **natural filtration** of X_t.

Example 5.18. i) If $Z_t = \int_0^t X_s\, ds$, then $\{Z_t\}$ is adapted to $\mathscr{F}_t^X$.
ii) If $M_t = \max_{0\leq s\leq t} W_s$, then $\{M_t\}$ is adapted to $\mathscr{F}_t^W$.
iii)If $Z_t = W_{t+1}^2 - W_t^2$, then $\{Z_t\}$ is NOT adapted to $\mathscr{F}_t^W$.

Definition 5.14. (Predictable Process) Given a filtration $\{\mathscr{F}_t\}_{t\geq 0}$, the stochastic process $\{X_t\}_{t\geq 0}$ is $\{\mathscr{F}_t\}$ predictable if X_t is $\mathscr{F}_{t-}$ measurable, where

$$\mathscr{F}_{t-} = \bigcup_{s<t} \mathscr{F}_s\,.$$

Definition 5.15. (Continuous Time Martingale) Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space with filtration $\{\mathscr{F}_t\}_{t\geq 0}$. A stochastic process $\{M_t\}$ is called a $(\mathbb{P}, \{\mathscr{F}_t\})$ **martingale** if

i) $\{M_t\}$ is adapted to $\{\mathscr{F}_t\}$;
ii) $E(|M_t|) < \infty$;
iii)$E(M_t|\mathscr{F}_s) = M_s$ if $s < t$.

Definition 5.16. (Stopping Time) Given a stochastic process $\{X_t\}$ adapted to a filtration $\{\mathscr{F}_t\}$, a random variable τ is called a stopping time if

$$\{\tau \leq t\} \in \mathscr{F}_t \,.$$

In other words, with information $\mathscr{F}_t$ up till time t, we can determine whether $\tau \leq t$ or not.

Similar to the discrete time case, we have the following theorems about continuous time martingale and stopping time.

Theorem 5.8. (Stopped Martingale is a Martingale) *If X_t is a $\{\mathbb{P}, \{\mathscr{F}_t\}\}$ martingale and τ is a stopping time adapted to $\{\mathscr{F}_t\}$, then $X_{t\wedge\tau}$ is also a $\{\mathbb{P}, \{\mathscr{F}_t\}\}$ martingale.*

Theorem 5.9. (Optional Stopping Theorem) *Suppose that X_t is a $\{\mathbb{P}, \{\mathscr{F}_t\}\}$ martingale and τ is a stopping time adapted to $\{\mathscr{F}_t\}$ such that the following conditions hold:*

1. *$\tau < \infty$ a.s.;*
2. *X_τ is integrable, i.e., $E(|X_\tau|) < \infty$;*
3. *$|X_{\tau\wedge t}| \leq Y$ a.s. for some $Y \in \mathscr{L}^1$ (i.e., $E(|Y|) < \infty$) and all $t > 0$.*

Then

$$E(X_\tau) = E(X_0)\,.$$

5.8 Exercises

Exercise 5.10 *Let $X_1, X_2, \ldots$ be a sequence of coin tosses and let $\mathscr{F}_n$ be the σ-field generated by $X_1, \ldots, X_n$. Write down all the elements in each of $\mathscr{F}_i$, $i = 1, 2$. What is the number of elements in $\mathscr{F}_3$?*

Exercise 5.11 *Let $X_1, X_2, \ldots$ be a sequence of coin tosses and let $\mathscr{F}_n$ be the σ-field generated by $X_1, \ldots, X_n$. For each of the following events find the smallest n such that the event belongs to $\mathscr{F}_n$:*
A = {the first occurrence of heads is preceded by no more than 10 tails},
B = {there is at least 1 head in the sequence $X_1, X_2, \ldots$},
C = {the first 100 tosses produce the same outcome},
D = {there are no more than 2 heads and 2 tails among the first 5 tosses}.

Exercise 5.12 *Show that if $\{X_n\}$ is a martingale with respect to $\{\mathscr{F}_n\}$, then*

$$E(X_1) = E(X_2) = \cdots$$

Hint: What is the expectation of $E(X_{n+1}|\mathscr{F}_n)$?

Exercise 5.13 *Suppose that $\{X_n\}$ is a martingale with respect to a filtration $\{\mathscr{F}_n\}$. Show that $\{X_n\}$ is a martingale with respect to the filtration $\{\mathscr{G}_n\}$, where*

$$\mathscr{G}_n = \sigma(X_1, \dots, X_n).$$

Hint: Observe that $\mathscr{G}_n \subset \mathscr{F}_n$ and use tower property of conditional expectation.

Exercise 5.14 *Let X_n be a symmetric random walk and $\{\mathscr{F}_n\}_{n\geq 1}$ be the filtration defined in Example 5.6. Show that if*

$$Y_n = (-1)^n \cos(\pi X_n)\,,$$

then $\{Y_n\}$ is a martingale with respect to $\{\mathscr{F}_n\}$.

Hint: Try to manipulate $E((-1)^{n+1}\cos(\pi X_{n+1})|\mathscr{F}_n)$ to obtain $(-1)^n\cos(\pi X_n)$. Use a similar argument as in Example 5.6 to achieve this. But, first of all, make sure that Y_n is integrable and adapted to $\{\mathscr{F}_n\}_{n\geq 1}$.

Exercise 5.15 *Recall Example 5.6. Verify the measurability and integrability of the martingale $X_n^2 - n$.*

Exercise 5.16 *Show that the following conditions are equivalent:*

1. *$\{\tau \leq n\} \in \mathscr{F}_n$ for each $n = 1, 2, \dots$*
2. *$\{\tau = n\} \in \mathscr{F}_n$ for each $n = 1, 2, \dots$*

Hint: Express $\{\tau \leq n\}$ in terms of the events $\{\tau = k\}$, where $k = 1, \dots, n$. Express $\{\tau = n\}$ in terms of the events $\{\tau \leq k\}$, where $k = 1, \dots, n$.

Exercise 5.17 *Let X_n be a sequence of random variables adapted to a filtration $\mathscr{F}_n$ and let $B \subset \mathbb{R}$ be a Borel set. Show that the **time of first entry** of X_n into B,*

$$\tau = \min\{n : X_n \in B\}\,,$$

is a stopping time.

Hint: Example 5.8 covers the case when $B = (-\infty, 0] \cup [10, \infty)$. Extend the argument to an arbitrary Borel set B.

Exercise 5.18 *If X_n is a martingale w.r.t. $\mathscr{F}_n$, then is X_n^3 a martingale? If $X_n = \sum_{i=1}^n Z_i$, where $Z_i = 1$ or -1 with equal probabilities, can you construct a martingale involving X_n^3?*

Exercise 5.19 *Verify that the process $\{Y_n\}$ and the stopped process $\{Y_{\tau\wedge n}\}$ in Example 5.10 are martingales. Show also that $\mathbb{P}(\tau < \infty) = 1$.*

Exercise 5.20 *Find the expected length of a game, $E(\tau)$, using the doubling strategy in Example 5.10*

Exercise 5.21 *In Example 5.13, show that Y_n is a martingale.*

Exercise 5.22 *Show that a previsible martingale is a constant.*

Exercise 5.23 *Let X_n be the symmetric random walk defined in Example 5.6. Let $Y_n = X_n + 1$ be a random walk started from 1 at time 0. Let $\tau = \inf\{n : Y_n = 0\}$.*

- *Show that τ is a stopping time.*
- *Find $E(\tau)$*
- *Show that Y_n and $Y_{\tau\wedge n}$ are martingales.*
- *Is $E(Y_\tau) = E(Y_1)$? Explain by verifying the optional stopping theorem.*

Exercise 5.24 *By writing $X_\tau = X_{\tau\wedge n} + (X_\tau - X_n)1_{\{\tau>n\}}$, show the Optional Stopping Theorem 5.4 by replacing condition 3) by the condition $E(X_n 1_{\{\tau>n\}}) \to 0$ as $n \to \infty$.*

Exercise 5.25 *Consider the random walk $X_n = Z_1 + \cdots + Z_n$, where $\mathbb{P}(Z_i = 1) = \mathbb{P}(Z_i = -1) = 1/2$ for all i, and $\tau_a = \inf\{n : X_n = a\}$, $a > 0$. Arguing in terms of each $\omega \in \Omega$, show that τ_a increases to ∞ a.s. as $a \to \infty$.*

Exercise 5.26 *In Example 5.14, verifying that $\varphi(X_n)$ and $X_n - (2p-1)n$ are martingales. Justify the applications of OST to the two martingales.*

Exercise 5.27 *In the upcrossing strategy, if $\{X_n\}_{n=1,\ldots,14} = (3,4,6,1,3,4,6,7,2,1,3,8,3,3)$, write down the paths $\{\alpha_n\}$ and the final total profit for the cases: i)$(a,b) = (3.1,3.9)$, ii) $(a,b) = (2.9,4.1)$.*

Exercise 5.28 *For a drifted random walk $X_n = Z_1 + \ldots + Z_n$, $Z_i \overset{a.s.}{\sim} N(\mu,1)$. We have $X_n \sim N(n\mu,n)$. Suppose that the upcrossing strategy is performed with $a = -1$ and $b = 1$. In terms of μ and n, find an upper bound for the expected number of upcrossing by time n. Give the values for $\mu = 0,-1,-2$ and $n = 10,100,1000$ (9 combinations).*

Exercise 5.29 *If $M_t^{(1)}$ and $M_t^{(2)}$ are martingales, then is $Y_t = M_t^{(1)} M_t^{(2)}$ a martingale?*

Chapter 6
Brownian Motion

Brownian motion originated from a model proposed by Robert Brown in 1828 for the phenomenon of "continual swarming motion" of pollen grains suspended in water. In 1900, Bachelier employed Brownian motion as a continuous time model for stock price fluctuations, which improves the discrete time models by allowing the stock price to change at any instant. The mathematical theory of Brownian motion was later developed by Norbert Wiener. Therefore, Brownian motion is also known as Wiener process.

6.1 Definition of Brownian motion

Brownian motion is an important example of a continuous time stochastic process. Formally defining a continuous time stochastic process $\{X_t : t \geqslant 0\}$ requires a probability triple $(\Omega, \mathscr{F}, \mathbb{P})$ such that X_t is $\mathscr{F}$-measurable for all t. As in the discrete case, we shall rarely specify $(\Omega, \mathscr{F}, \mathbb{P})$ explicitly. On the other hand, the probability space $(\Omega_B, \mathscr{F}_B, \mathbb{P}_B)$ induced by the Brownian motion can be defined as follows.

Definition 6.1. (Probability Space of Brownian Motion) The probability space induced by the Brownian motion, $(\Omega_B, \mathscr{F}_B, \mathbb{P}_B)$, is given by

- $\Omega_B = \{\text{all continous functions } \ \omega : [0,\infty) \to \mathbb{R}\}$
- $\mathscr{F}_B$ is the σ-field generated by the *finite dimensional sets*, i.e.,

$$\mathscr{F}_B = \sigma\left\{\{\omega : \omega(t_i) \in B_i, i = 1,\ldots,n\}, \forall n \in \mathbb{N}, \forall B_i \in \mathscr{B}, \forall t_i \in [0,\infty)\right\}. \tag{6.1}$$

- $\mathbb{P}_B$ is a probability measure on $(\Omega_B, \mathscr{F}_B)$ such that $\mathbb{P}_B(\{\omega : \omega(0) = 0\}) = 1$ and

$$\mathbb{P}_B\{\omega : \omega(t_i) \in B_i, i = 1,\ldots,n\} = \int_{B_1} \cdots \int_{B_n} \phi_{t_1\ldots t_n}(x_1,\ldots,x_n)\, dx_n \ldots dx_1 , \tag{6.2}$$

  where $\phi_{t_1\ldots t_n}(x_1, x_2, \ldots, x_n)$ is the joint p.d.f. of a Gaussian random vector with zero mean and covariance matrix $(\min(t_i,t_j))_{i,j}$.

A Brownian motion is defined as an element of Ω_B that is distributed according to $\mathbb{P}_B$. As the mathematical theory of Brownian motion is developed by Wiener, the notation W_t is commonly used to denote a Brownian motion. Precisely, W_t is a *random function* $W_t(\omega) = \omega(t)$ which maps an abstract $\omega \in \Omega$ to a function $\omega(t) \in \Omega_B$. However, since we are seldom interested in ω, W_t is often understood as an element of Ω_B. To simplify notations, in the following we denote the probability space $(\Omega_B, \mathscr{F}_B, \mathbb{P}_B)$ by $(\Omega, \mathscr{F}, \mathbb{P})$.

To check whether a stochastic process is a Brownian motion, it is often not easy to verify (6.2). Alternatively, the below definition offers an easier way for the verification.

Definition 6.2. (Brownian motion) A real-valued stochastic process $\{W_t : t \geqslant 0\}$ is a $\mathbb{P}$-Brownian motion (or a $\mathbb{P}$-Wiener process) if for some $\sigma \in \mathbb{R}^+$, under $\mathbb{P}$,

1) **(Stationary increment)** for each $s \geqslant 0$ and $t > 0$ the random variable $W_{t+s} - W_s$ is normally distributed with mean zero and variance $\sigma^2 t$,
2) **(Independent increment)** for each $n \geqslant 1$ and any time indexes $0 \leqslant t_0 \leqslant t_1 \leqslant \cdots \leqslant t_n$, the random variables $\left\{W_{t_r} - W_{t_{r-1}}\right\}_{r=1,\ldots,n}$ are independent.
3) **(Starts at 0)** $\mathbb{P}(W_0 = 0) = 1$.

Theorem 6.1. *Definitions 6.1 and 6.2 are equivalent.*

Proof. First we show that Definition 6.1 implies Definition 6.2. Without loss of generality, let $n = 3$ and $\sigma = 1$. From (6.2), the joint p.d.f. of W_{t_1}, W_{t_2} and W_{t_3} is

$$\phi_{t_1,t_2,t_3}(x_1,x_2,x_3) = \frac{1}{(2\pi)^{3/2}|\Sigma|} e^{-\frac{1}{2}\mathbf{x}'\Sigma^{-1}\mathbf{x}}, \quad \text{where } \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \ \Sigma = \begin{pmatrix} t_1 & t_1 & t_1 \\ t_1 & t_2 & t_2 \\ t_1 & t_2 & t_3 \end{pmatrix}.$$

Using the transformation

$$\begin{pmatrix} W_{t_1} \\ W_{t_2} - W_{t_1} \\ W_{t_3} - W_{t_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} W_{t_1} \\ W_{t_2} \\ W_{t_3} \end{pmatrix} =: M \begin{pmatrix} W_{t_1} \\ W_{t_2} \\ W_{t_3} \end{pmatrix}, \tag{6.3}$$

and the fact that $|M| = 1$, the joint p.d.f. of $W_{t_1}, W_{t_2} - W_{t_1}$ and $W_{t_3} - W_{t_2}$ is

$$\begin{aligned} f_{W_{t_1}, W_{t_2}-W_{t_1}, W_{t_3}-W_{t_2}}(y_1,y_2,y_3) &= \frac{1}{(2\pi)^{3/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(M^{-1}\mathbf{y})'\Sigma^{-1}M^{-1}\mathbf{y}} = \frac{1}{(2\pi)^{3/2}|\widetilde{\Sigma}|^{1/2}} e^{-\frac{1}{2}\mathbf{y}'\widetilde{\Sigma}^{-1}\mathbf{y}} \\ &= \prod_{i=1}^{3} \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} e^{-\frac{y_i^2}{2(t_i - t_{i-1})}}, \end{aligned} \tag{6.4}$$

where $t_0 := 0$ and

$$\widetilde{\Sigma} = M\Sigma M' = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} t_1 & t_1 & t_1 \\ t_1 & t_2 & t_2 \\ t_1 & t_2 & t_3 \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} t_1 & 0 & 0 \\ 0 & t_2 - t_1 & 0 \\ 0 & 0 & t_3 - t_2 \end{pmatrix}.$$

Since the joint p.d.f. factorizes into a product of Normal p.d.f.s, the independent increment condition 2) holds. Substituting $t_1 = s$ and $t_2 = t + s$, we have $t_2 - t_1 = s$ and thus $W_{t+s} - W_s \sim N(0, \sigma^2 t)$, i.e., the stationary increment condition 1) follows. Finally, 3) is already assumed in Definition 6.1.

Conversely, to show Definition 6.2 implies Definition 6.1, note that the stationary and independent increment conditions 1)2) guarantee (6.4). Then, inverting the transformation in (6.3) gives (6.2). What remains is the continuity of the process, which is assumed in Definition 6.1 but not in Definition 6.2. For illustration, we show that Brownian motion is continuous in probability at any t, i.e., for any $\varepsilon > 0$,

$$\lim_{u\to 0} \mathbb{P}(|W_{t+u} - W_t| \geqslant \varepsilon) = 0\,. \tag{6.5}$$

Specifically, for any $\varepsilon > 0$, by the stationary increment condition 1), as $u \to 0$,

$$\mathbb{P}(|W_{t+u} - W_t| \geqslant \varepsilon) \stackrel{1)}{=} \mathbb{P}(|N(0,u)| \geqslant \varepsilon) = 2\left(1 - \Phi\left(\frac{\varepsilon}{\sqrt{u}}\right)\right) \to 2(1 - \Phi(\infty)) = 0\,,$$

giving (6.5). Here $\Phi(\cdot)$ is the c.d.f. of a standard Normal distribution. In fact, Brownian motion is almost surely continuous, i.e., $\mathbb{P}(\lim_{u\to 0} |W_{t+u} - W_t| = 0) = 1$, but the proof is much more involved. □

Remark 6.1. **(Some terminologies)**

1. **(Standard Brownian Motion)** The process $\{W_t : t \geqslant 0\}$ with $\sigma^2 = 1$ is called a standard Brownian motion. The parameter σ^2 is known as the variance parameter. For simplicity, unless otherwise stated we shall assume that $\sigma^2 = 1$.
2. **(Finite dimensional distribution)** The joint distributions of $W_{t_1}, \cdots, W_{t_n}$ for each $n \geqslant 1$ and all $t_1, \cdots, t_n$ are known as the finite dimensional distributions of the process.

The following examples demonstrate how Definition 6.2 can be used to check whether a process is a Brownian motion.

Example 6.1. For any positive constant c, the process $\widetilde{W}_t = c^{-1/2} W_{ct}$ is a Brownian motion. To see this, we verify the three conditions in Definition 6.2.

1. **(Stationary increment)** For each $s \geqslant 0$ and $t > 0$ the random variable
$$\widetilde{W}_{t+s} - \widetilde{W}_s \stackrel{D}{=} c^{-1/2}(W_{c(t+s)} - W_{cs}) \stackrel{D}{=} c^{-1/2} N(0, ct) = N(0,t)\,.$$
2. **(Independent increment)** For each $n \geqslant 1$ and any time indexes $0 = t_0 < t_1 < \cdots < t_n$, the random variables $\left\{\widetilde{W}_{t_r} - \widetilde{W}_{t_{r-1}}\right\}_{r=1,\ldots,n} = \left\{c^{-1/2}(W_{ct_r} - W_{ct_{r-1}})\right\}_{r=1,\ldots,n}$ are independent since $\left\{W_{ct_r} - W_{ct_{r-1}}\right\}_{r=1,\ldots,n}$ are independent.
3. **(Starts at 0)** $\widetilde{W}_0 = c^{-1/2} W_{c0} = 0$ since $W_0 = 0$.

Remark 6.2. **(Fractal)** Example 6.1 shows that Brownian motion is a **fractal**, i.e., a set of points that shows **self-similar pattern**. For example, if $c = 0.01$, the graph of the rescaled process $\widetilde{W}_t = c^{-1/2}W_{ct} = 10W_{0.01t}$ on $t \in [0,1]$ is the graph of 10 times of the original Brownian motion W_t from $t \in [0, 0.01]$. However, the shapes of $\{\widetilde{W}_t\}_{t\in[0,1]}$ and $\{W_t\}_{t\in[0,1]}$ are similar, since they are both standard Brownian motions. In other words, it does't matter what scale you examine the Brownian motions – they look just the same.

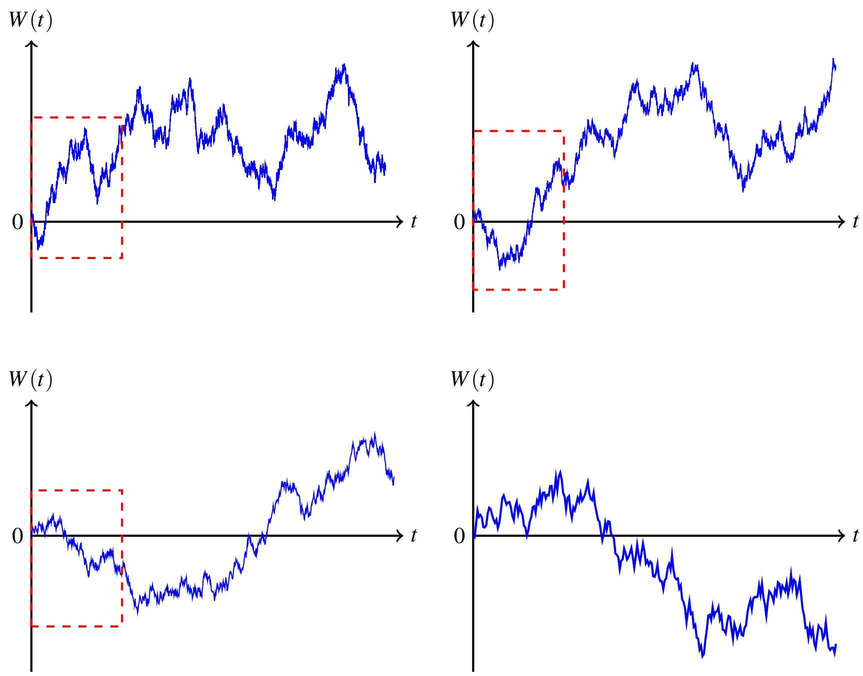


Fig. 6.1 Fractal property: zooming in Brownian motion.

Example 6.2. **(Translational Invariance)** For any $s \geqslant 0$, the process $\widetilde{W}_t = W_{t+s} - W_s$ is a standard Brownian motion. The verification is left to Exercise 6.1

From the stationary increment property in Definition 6.2, we can define the transition probabilities of the standard Brownian motion.

Definition 6.3. (Transition Probabilities) The **transition probability** of a standard Brownian motion from x to y in time t is the conditional probability density of W_{t+s} given $W_s = x$, which is given by

$$p(t,x,y) := \frac{1}{\sqrt{2\pi t}} \exp\left\{-\frac{(y-x)^2}{2t}\right\}.$$

Transition probability is a useful tool in establishing many probabilistic properties of Brownian motions. For example, as a by-product of the transformation arguments in the proof of Theorem 6.1, a simple formula for the joint density can be deduced as

$$\begin{aligned} f_{W_{t_1},\dots,W_{t_n}}(x_1,\cdots,x_n) &= f_{W_{t_1},W_{t_2}-W_{t_1},\dots,W_{t_n}-W_{t_{n-1}}}(x_1,x_2-x_1,\dots,x_n-x_{n-1}) \\ &= \prod_{j=1}^{n} p\left(t_j-t_{j-1},x_{j-1},x_j\right) . \end{aligned}$$

The conditional mean and variance can be computed easily using the stationary and independent increment of Brownian motions (See Exercise 6.6).

Lemma 6.2 *For any $s,t>0$,*

1) $E\left(W_{t+s}-W_s|\sigma\{W_r : 0\leqslant r\leqslant s\}\right)=0$.
2) $\operatorname{Cov}(W_s,W_t)=s\wedge t$.

Example 6.3. **(Martingales constructed from Brownian motions)** Let $\{W_t\}$ be a standard Brownian motion and $\{\mathscr{F}_t\}$ be the natural filtration of the B.M., i.e. $\mathscr{F}_t=\sigma\{\omega\in\Omega_B:\omega(t_i)\in\mathscr{B}_i,B_i\in\mathscr{B},0\leqslant t_1<\cdots<t_n\leqslant t\}$, then

i) W_t is a $(\mathbb{P},\{\mathscr{F}_t\})$ martingale.
ii) W_t^2-t is a $(\mathbb{P},\{\mathscr{F}_t\})$ martingale.
iii)$e^{W_t-\frac{1}{2}t}$ is a $(\mathbb{P},\{\mathscr{F}_t\})$ martingale.

See Exercise 6.7.

Example 6.4. Given a standard Brownian motion W_t, consider the first hitting time $T_a=\inf\{t\geqslant 0: W_t=a\}$. It is easy to see that T_a is a stopping time because

$$\{T_a\leqslant t\}=\{\exists s\in[0,t]: W_s=a\}\in\mathscr{F}_t\,,$$

(the set in the middle depends only on $\{W_s : 0\leqslant s\leqslant t\}$). Notice also that, by definition, if $T_a<\infty$, then $W_{T_a}=a$.

Other examples of stopping times include $T_{|a|}:=\inf\{t\geqslant 0: |W_t|=a\}$ and $T_{(a,b)}:=\inf\{t\geqslant 0: |W_t|\notin(a,b)\}$. Analogous to random walks, an example of a random time that is *not* a stopping time is the *last* time that a process hits some level. □

From the property of independent increment, Brownian motions have no memory. That is, if $\{W_t : t\geqslant 0\}$ is a Brownian motion and $s\geqslant 0$ is any fixed time, then $\{W_{t+s}-W_s : t\geqslant 0\}$ is also a Brownian motion, independent of $\{W_r : 0\leqslant r\leqslant s\}$. This is known as the **Markov property**. In fact, the same conclusion holds even when the constant s is replaced by a stopping time τ. This is known as the **strong Markov property**.

Theorem 6.3. (Strong Markov Property) *If $W=\{W_t\}$ is a Brownian motion and τ is a stopping time adapted to $\{\mathscr{F}_t\}$, then*

$$\widetilde{W}_t=W_{t+\tau}-W_\tau\,,$$

for $t \geqslant 0$ is a Brownian motion independent of $\{W_r : 0 \leqslant r \leqslant \tau\}$.

Proof. First we approximate the stopping time τ by defining a discretized version $T_n = \sum_{m=0}^{\infty}(m+1)2^{-n}1_{\{\tau \in [m2^{-n},(m+1)2^{-n})\}}$. Since $\{T_n \leqslant t\} = \{\tau < t\} \in \mathscr{F}_t$, T_n is a stopping time. Let $\widetilde{W}_k(t) = W_{t+k/2^n} - W_{k/2^n}$. The independent increment condition guarantees that $\{\widetilde{W}_k(t),\ t \in [0,\infty)\}_{k\in\mathbb{Z}^+}$ are Brownian motions. Next, approximate $\widetilde{W}$ by $\widetilde{W}^{(n)}$ with $\widetilde{W}_t^{(n)} = W_{t+T_n} - W_{T_n}$. Suppose that $E \in \mathscr{F}_{T_n}$, then for every $B \in \mathscr{F}$ ($\mathscr{F} = \mathscr{F}_B$ defined in (6.1)), we have

$$
\begin{aligned}
\mathbb{P}(\{\widetilde{W}^{(n)} \in B\} \cap E) &= \sum_{k=0}^{\infty} \mathbb{P}(\{\widetilde{W}_k \in B\} \cap E \cap \{T_n = k2^{-n}\}) \qquad (\widetilde{W}^{(n)} = \widetilde{W}_k \text{ if } T_n = k2^{-n}) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(\{\widetilde{W}_k \in B\})\mathbb{P}(E \cap \{T_n = k2^{-n}\}) \qquad \text{(independent increment)} \\
&= \mathbb{P}(\{\widetilde{W}_1 \in B\}) \sum_{k=0}^{\infty} \mathbb{P}(E \cap \{T_n = k2^{-n}\}) \qquad \text{(stationary increment)} \\
&= \mathbb{P}(\{\widetilde{W}_1 \in B\})\mathbb{P}(E),
\end{aligned}
$$

implying that $\{\widetilde{W}^{(n)}\}$ is a Brownian motion and is independent of E. Since E can be any subset of $\mathscr{F}_{T_n}$, $\{\widetilde{W}^{(n)}\}$ is independent of $\mathscr{F}_{T_n}$.

Finally, since $T_n \downarrow \tau$ as $n \to \infty$, and Brownian motion is continuous a.s., $\widetilde{W}_t = \lim_{n\to\infty} \widetilde{W}_t^{(n)}$ for all $t > 0$ a.s. and $\lim_{n\to\infty} \mathscr{F}_{T_n} = \mathscr{F}_\tau$. Thus $\widetilde{W}$ is a Brownian motion independent of $\mathscr{F}_\tau$, and the proof is complete. □

6.2 Non-Differentiability of Brownian Motions

We know that Brownian motion is continuous. In this section we show that Brownian motion is non-differentiable at any time point. In other words, Brownian motion has a **rough path**.

Theorem 6.4. (Non-differentiability of Brownian Motions) *A Brownian motion W_t is almost surely non-differentiable at any t.*

Proof. By the translation invariance property, $W_{t+s} - W_s \overset{D}{=} W_t$ for any $s \geqslant 0$. Thus it suffices to focus on the case $t = 0$. If W_t is differentiable at $t = 0$, then $\frac{W_t}{t}$ converges as $t \to 0$. Conversely, if for any n, there exists a $t \in \left(0, \frac{1}{n^4}\right]$ such that $\frac{|W_t|}{t} > n$, then W_t is not differentiable at $t = 0$. That is, $\bigcap_{n=1}^{\infty} A_n \subseteq \{W_t \text{ is not differentiable at } t = 0\}$, where

$$A_n = \left\{ \frac{|W_t|}{t} > n \text{ for some } t \in \left(0, \frac{1}{n^4}\right] \right\}.$$

Hence, the almost sure non-differentiability follows if we can show that $\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = 1$. Now, since $\left\{ \frac{|W_t|}{t} > n \text{ at } t = \frac{1}{n^4} \right\} \subseteq A_n$, we have

$$
\begin{aligned}
\mathbb{P}(A_n) &\geqslant \mathbb{P}\left(\left\{\frac{|W_t|}{t} > n \text{ at } t = \frac{1}{n^4}\right\}\right) = \mathbb{P}\left(\frac{|W_{1/n^4}|}{1/n^4} > n\right) \\
&= \mathbb{P}\left(|n^2 W_{1/n^4}| > \frac{1}{n}\right) && \text{(Simple algebra)} \\
&= \mathbb{P}\left(|\widetilde{W}_1| > \frac{1}{n}\right) && (\text{Example } 6.1 : \widetilde{W}_t = n^2 W_{t/n^4} \text{ is a B.M.}) \\
&\to 1\,,
\end{aligned}
$$

as $n \to \infty$. As $\{A_n\}$ is a contracting sequence of events, Theorem 3.5 implies

$$
\mathbb{P}(W_t \text{ is non-differentiable at } 0) \geqslant \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n) = 1\,.
$$

□

Indeed, whether such a bizarre process actually *exists* is far from obvious. We will devote the next section to show the existence of B.M. by explicitly constructing a stochastic process that satisfies Definition 6.2.

6.3 Lévy's construction of Brownian motions

In this section we present a construction of B.M. due to Lévy's polygonal approximation.

6.3.1 A polygonal approximation

Levy's idea is to construct a path of Brownian motion by a limit of polygonal interpolation. Particularly, by specifying the endpoints of the process and then infilling the middle with appropriate values, we can ensure the almost sure convergence of the approximation. We require the following lemma.

Lemma 6.5 *Suppose that $\{W_t : t \geqslant 0\}$ is a Brownian motion starting at x_0. Conditioned on $W_t = x_1$, the probability density function of $W_{\frac{t}{2}}$ is*

$$
p_{\frac{t}{2}}(x) := \sqrt{\frac{2}{\pi t}} \exp\left\{-\frac{1}{2}\frac{\left(x - \frac{1}{2}(x_0 + x_1)\right)^2}{\frac{t}{4}}\right\}.
$$

In other words, $W_{t/2}|\{W_0 = x_0, W_t = x_1\} \sim N(\frac{x_0+x_1}{2}, \frac{t}{4})$.

Proof. By the property of B.M., $(W_{t/2}, W_t) \sim N((x_0, x_0), \Sigma)$, where $\Sigma = \begin{pmatrix} t/2 & t/2 \\ t/2 & t \end{pmatrix}$. Direct calculations give $|\Sigma| = t^2/4$ and $\Sigma^{-1} = \begin{pmatrix} 4/t & -2/t \\ -2/t & 2/t \end{pmatrix}$. Thus, the joint p.d.f. can be expressed as

$$\begin{aligned} f_{W_{t/2},W_t}(x,y) &= \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}(x-x_0,\ y-x_0)\Sigma^{-1}(x-x_0,\ y-x_0)'} \\ &= \frac{1}{\pi t} e^{-\frac{1}{2t}(4(x-x_0)^2-4(x-x_0)(y-x_0)+2(y-x_0)^2)} \end{aligned}$$

Since the marginal distribution of W_t is $f_{W_t}(y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x_0)^2}{2t}}$, the conditional distribution of $W_{t/2}$ given W_t is

$$\begin{aligned} f_{W_{t/2}|W_t=x_1}(x) &= f_{W_{t/2},W_t}(x,x_1)/f_{W_t}(x_1) \\ &= \frac{1}{\sqrt{2\pi(t/4)}} e^{-\frac{1}{2(t/4)}\left(x-\frac{x_0+x_1}{2}\right)^2}, \end{aligned}$$

completing the proof. $\square$

Without loss of generality we take the range of t to be $[0, 1]$. Lévy's construction builds (inductively) a polygonal approximation to the Brownian motion by infilling the middle point with normal random variables using Lemma 6.5. In particular, in the n-th step of the approximation, a process $X_n = X_n(t)$ on $t = [0,1]$ is defined on the grid $G_n = \{k/2^n\}_{k=0,1,\dots,2^n}$ based on normal random variables at each grid point. As the grid G_n becomes finer as $n \to \infty$, the B.M. is defined as the limit of X_n. The construction is illustrated in Figure 6.2.

Levy's Construction of Brownian motions: Notation: $\{Z_n^k\}_{k,n\in\mathbb{Z}^+}$ are independent standard normal random variables; $X_n^k = X_n(k/2^n)$.

- **Initialization.** Set $X_0(0) = 0$, $X_0(1) = Z_1^1$ and

$$X_0(t) = tZ_1^1,$$

i.e., X_0 is a linear function on $[0,1]$ with vertices on the grid $G_0 = \{k/2^0\}_{k=0,1}$.

- **The inductive Step.** Given X_n with vertices on the grid G_n, we define X_{n+1} with vertices on G_{n+1} by infilling more points in the set $G_{n+1}\backslash G_n$. On the common grid-points $G_{n+1}\cap G_n = G_n$, set X_{n+1} to equal X_n, that is

$$X_{n+1}^{2k} = X_n^k \ \text{ for } \ k = 1,\ldots,2^n. \tag{6.6}$$

It remains to infill the middle X_{n+1}^{2k-1} for $k = 1,\ldots,2^n$. Using Lemma 6.5 with $(x_0,x_1) = (X_{n+1}^{2k-2}, X_{n+1}^{2k})$ and $t = 1/2^n$, the conditional distribution of X_{n+1}^{2k-1} is $N\left(\frac{1}{2}(X_{n+1}^{2k-2}+X_{n+1}^{2k}), 2^{-n-2}\right)$. Therefore, we can generate X_{n+1}^{2k-1} by

$$X_{n+1}^{2k-1} = \frac{1}{2}(X_{n+1}^{2k-2}+X_{n+1}^{2k}) + 2^{-(n/2+1)}Z_{n+1}^{2k-1}. \tag{6.7}$$

Thus, we obtain $X_{n+1}(t)$ on the grid G_{n+1}. Next, define $X_{n+1}(t)$ on $t\in[0,1]$ by interpolating the points, i.e., for $k = 0,\ldots,2^{n+1}-1$,

$$X_{n+1}(t) = X_{n+1}^k + (2^{n+1}t-k)(X_{n+1}^{k+1} - X_{n+1}^k) \ \text{ for } \ t \in \left[\frac{k}{2^{n+1}}, \frac{k+1}{2^{n+1}}\right].$$

- Repeat the induction until $n\to\infty$.

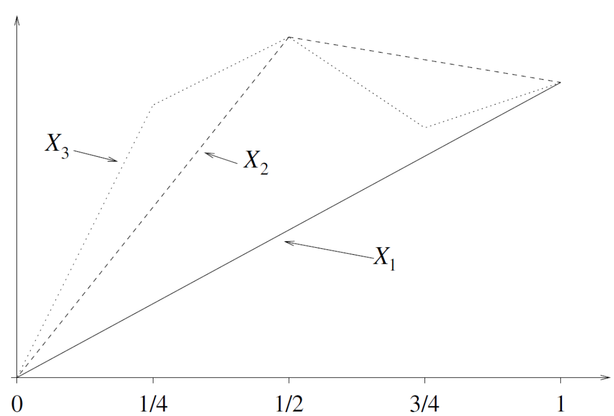


Fig. 6.2 Lévy's sequence of polygonal approximations to Brownian motion.

Remark 6.3. In the n-th iteration of Levy's construction, $X_n(t)$, linear in each interval $[(k-1)2^{-n}, k2^{-n}]$, is continuous in t and satisfies $X_n(0)=0$. It is thus determined by the values $\left\{X_n^k : k=1,\cdots,2^n\right\}$. By (6.6) and the linear property in the construction, we have

$$\frac{1}{2}(X_{n+1}^{2k-2}+X_{n+1}^{2k})=\frac{1}{2}(X_n^{k-1}+X_n^k)=X_n((2k-1)/2^{n+1})\,.$$

Thus (6.7) can be written as $X_{n+1}^{2k-1}=X_n((2k-1)/2^{n+1})+2^{-(n/2+1)}Z_{n+1}^{2k-1}$, that is,

$$X_{n+1}((2k-1)/2^{n+1})-X_n((2k-1)/2^{n+1})=2^{-(n/2+1)}Z_{n+1}^{2k-1}\,. \tag{6.8}$$

The representation (6.8) will be useful in the next subsection.

6.3.2 Convergence to Brownian motion

To justify the above construction of B.M., it is important to show the existence of $\lim_{n\to\infty}X_n(t)$. Since $X_n(t)$ is a random function, the limit should be understood as a random function $X(t)$ on $(\Omega_B,\mathscr{F}_B)$. We will show that X_n converges uniformly to X almost surely, i.e.,

$$P\left(\lim_{n\to\infty}\max_{t\in[0,1]}|X_n(t)-X(t)|=0\right)=1\,. \tag{6.9}$$

Note that the term "uniform" refers to the operator $\max_{t\in[0,1]}$. Moreover, we need to verify that the limit $X(t)$ satisfies Definition 6.2, hence a Brownian motion. This is the content of the following theorem.

Theorem 6.6. (Convergence of the Polygonal Construction of Brownian Motions)

1) There exists a random function $X(t)$ on $(\Omega_B,\mathscr{F}_B)$ satisfying (6.9).
2) The limit $X(t)$ satisfies Definition 6.2.

Proof. 1. **(Existence of Limit).** Note that $\max_{t\in[0,1]}|X_{n+1}(t)-X_n(t)|$ must be attained at one of the vertex $t\in\left\{(2k-1)2^{-(n+1)} : k=1,2,\cdots,2^n\right\}$. Using (6.8), we have

$$\begin{aligned}\mathbb{P}\left\{\max_{t\in[0,1]}|X_{n+1}(t)-X_n(t)|\geqslant 2^{-\frac{n}{4}}\right\}&=\mathbb{P}\left\{\max_{1\leqslant k\leqslant 2^n}Z_{n+1}^{2k-1}\geqslant 2^{\frac{n}{4}+1}\right\}\\&\leqslant 2^n\mathbb{P}\left\{N(0,1)\geqslant 2^{\frac{n}{4}+1}\right\}.\end{aligned} \tag{6.10}$$

Direct calculations (see Exercise 6.3) suggest that, for $x>0$,

$$\mathbb{P}\{N(0,1) \geqslant x\} \leqslant \frac{1}{x\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}, \tag{6.11}$$

and for $n \geqslant 4$,

$$\exp\left\{-2^{\frac{n}{2}+1}\right\} < 2^{-2n+2}. \tag{6.12}$$

Combining (6.10), (6.11) and (6.12), for $n \geqslant 4$, we have

$$\begin{aligned}
\mathbb{P}\left\{\max_{t\in[0,1]} |X_{n+1}(t) - X_n(t)| \geqslant 2^{-\frac{n}{4}}\right\} &\leqslant 2^n \mathbb{P}\left\{N(0,1) \geqslant 2^{\frac{n}{4}+1}\right\} \\
&\leqslant \frac{2^n}{2^{\frac{n}{4}+1}} \frac{1}{\sqrt{2\pi}} \exp\left\{-2^{\frac{n}{2}+1}\right\} \\
&\leqslant \frac{2^n}{2^{\frac{n}{4}+1}} 2^{-2n+2} < 2^{-n}.
\end{aligned} \tag{6.13}$$

Consider now for $k > n \geqslant 4$,

$$\begin{aligned}
&\mathbb{P}\left\{\max_{t\in[0,1]} |X_k(t) - X_n(t)| \leqslant 2^{-\frac{n}{4}+3}\right\} \\
\geqslant\ & \mathbb{P}\left\{\sum_{j=n}^{k-1} \max_{t\in[0,1]} \left|X_{j+1}(t) - X_j(t)\right| \leqslant 2^{-\frac{n}{4}+3}\right\} \\
\geqslant\ & \mathbb{P}\left\{\max_{t\in[0,1]} \left|X_{j+1}(t) - X_j(t)\right| \leqslant 2^{-\frac{j}{4}},\ j = n, \cdots, k-1\right\} \\
\overset{(6.13)}{\geqslant}\ & 1 - \sum_{j=n}^{k-1} 2^{-j} \geqslant 1 - 2^{-n+1},
\end{aligned}$$

where the second inequality follows from the fact that if $m_j \leqslant 2^{-j/4}$ for all $j = n, \ldots, k-1$, then $\sum_{j=n}^{k-1} m_j \leqslant \sum_{j=n}^{k-1} 2^{-j/4} \leqslant 2^{-n/4+3}$. In other words, we have

$$\mathbb{P}\left\{\max_{t\in[0,1]} |X_k(t) - X_n(t)| \geqslant 2^{-\frac{n}{4}+3}\right\} \leqslant 2^{-n+1} \tag{6.14}$$

for all $k \geqslant n$ (the case $k = n$ is trivial). Since the maximum is attained only at the vertices, the maximum of $|X_k(t) - X_n(t)|$ can only be increased by the addition of new vertices. Thus, the event on the left of (6.14) is increasing with k, and Theorem 3.5 implies that

$$\begin{aligned}
&\mathbb{P}\left[\bigcup_{k\geqslant n} \left\{\max_{t\in[0,1]} |X_k(t) - X_n(t)| \geqslant 2^{-\frac{n}{4}+3}\right\}\right] \\
=\ & \lim_{k\to\infty} \mathbb{P}\left[\max_{t\in[0,1]} |X_k(t) - X_n(t)| \geqslant 2^{-\frac{n}{4}+3}\right] \leqslant 2^{-n+1}.
\end{aligned}$$

Let $A_n = \bigcup_{k\geqslant n}\left\{\max_{t\in[0,1]}|X_k(t)-X_n(t)| \geqslant 2^{-\frac{n}{4}+3}\right\}$. As $\sum_{n\geqslant 1}2^{-n+1}<\infty$, Borel Cantelli Lemma 3.6 implies that $\mathbb{P}(\limsup_{n\to\infty}A_n)=0$, i.e.,

$$\mathbb{P}\left\{\liminf_{n\to\infty} A_n^c\right\} = \mathbb{P}\{\omega : \omega \in A_n^c \text{ for all but finitely many } n\} = 1\,. \quad (6.15)$$

For each $\omega\in\liminf_{n\to\infty}A_n^c$, there exists some $N>0$ such that for all $n>N$, $\max_{t\in[0,1]}|X_k(t)-X_n(t)|<2^{-\frac{n}{4}+3}$ for all $k>n$. This in turn implies that for each $\varepsilon>0$, there exists some $N>0$ such that $d_{\text{sup}}(X_k,X_n)=\max_{t\in[0,t]}|X_k(t)-X_n(t)|$ $<\varepsilon$ for all $k,n>N$. Thus, for each $\omega\in\liminf_{n\to\infty}A_n^c$, $\{X_n\}$ is a Cauchy sequence in supremum distance d_{sup}. Hence, the limit X exists and $X_n\to X$ uniformly (note that this convergence has no randomness involved because we have fixed a sample path ω). Equation (6.15) guarantees that the uniform convergence holds almost surely, and the a.s. existence of limit is established.

2. **(The Limit is B.M.)** In the first step of the construction, $X_0(0)=0$ and $X_0(1)=Z(1)$ have the same distributional property of B.M., i.e., $W_0=0$ and $W_1\sim N(0,1)$. In each induction step, a new vertex is constructed using the covariance structure of the B.M.. Thus Properties 1-3 in Definition 6.2 are automatically satisfied for $X_n(t)$ restricted to the grid $G_n=\{k2^{-n}\}_{k=0,1,\cdots,2^n}$. Since we don't change the values of $X_k(t)$ on $t\in G_n$ for $k>n$, the same must be true for the limit X on $\cup_{n=1}^{\infty}G_n$. Before we show Properties 1-3 on the whole interval $[0,1]$, we show the continuity of $X(t)$: For any $\varepsilon>0$ and $n\geqslant 0$,

$$\begin{aligned}&\mathbb{P}(|X(t+\delta)-X(t)|>\varepsilon)\\ \leq\ &\mathbb{P}(|X(t+\delta)-X_n(t+\delta)|>\frac{\varepsilon}{3})+\mathbb{P}(|X_n(t+\delta)-X_n(t)|>\frac{\varepsilon}{3})+\mathbb{P}(|X_n(t)-X(t)|>\frac{\varepsilon}{3})\,.\end{aligned}$$

Note that the middle term goes to zero as $\delta\to 0$ by the continuity of X_n. Also, from (6.9) the first and the third term can be arbitrarily small because n can be arbitrarily large. Thus $\mathbb{P}(|X(t+\delta)-X(t)|>\varepsilon)$ converges to 0 as $\delta\to 0$, yielding the continuity of the limit.
Finally, note that any real number on $[0,1]$ can be arbitrarily closely approximated by some elements in $\cup_{n=1}^{\infty}G_n$. Thus by approximating any $0<t_1<t_2<\cdots<t_n<1$ from $\cup_{n=1}^{\infty}G_n$, the continuity of $X(t)$ implies that Properties 1-3 in Definition 6.2 hold on the whole interval $t\in[0,\,1]$. $\square$

6.4 The Reflection Principle and Hitting Times

Having proved the existence of Brownian motions, we explore more properties related to B.M in this and the next sections.

6.4.1 The reflection principle

Many useful results can be obtained from exploiting the symmetry inherent in Brownian motions. As a warm-up, we calculate the distribution function of $T_a = \min\{t : W_t = a\}$.

Lemma 6.7 (Distribution function of first hitting time) *Let $\{W_t : t \geqslant 0\}$ be a $\mathbb{P}$-Brownian motion started from $W_0 = 0$, then for $a > 0$, we have*

$$\mathbb{P}\{T_a < t\} = 2\mathbb{P}\{W_t > a\} .$$

Proof. If $W_t > a$, then by the continuity of the Brownian path, $T_a < t$. Moreover, from Theorem 6.3, $\{W_{t+T_a} - W_{T_a} : t \geqslant 0\}$ is a Brownian motion. So, by symmetry,

$$\mathbb{P}\{W_t - W_{T_a} > 0 | T_a < t\} = \frac{1}{2}.$$

Thus

$$\begin{aligned} \mathbb{P}\{W_t > a\} &= \mathbb{P}\{\{T_a < t\} \cap \{W_t - W_{T_a} > 0\}\} \\ &= \mathbb{P}\{T_a < t\}\mathbb{P}\{W_t - W_{T_a} > 0 | T_a < t\} \\ &= \frac{1}{2}\mathbb{P}\{T_a < t\}. \end{aligned}$$

□

By verifying the definition of Brownian motions (with the help of strong Markov property in Theorem 6.3), we have a more refined version of the above idea.

Lemma 6.8 (The reflection principle) *Let $\{W_t : t \geqslant 0\}$ be a standard Brownian motion and T be a stopping time. Define*

$$\widetilde{W}_t := \begin{cases} W_t, & \text{if } t \leqslant T; \\ 2W_T - W_t, & \text{if } t > T. \end{cases}$$

then $\left\{\widetilde{W}_t : t \geqslant 0\right\}$ is also a standard Brownian motion.

Notice that if $T = T_a$ (the first hitting time on a), then $\widetilde{W}_t$ reflects the portion of the path after T_a about the line $x = a$ (see Figure 6.3). The reflection principle is the key to prove the following result, which is useful to price certain barrier options (see Example 7.13).

Lemma 6.9 (Joint distribution of Brownian motion and its maximum) *Let $M_t := \max_{s\in[0,t]} W_s$, the maximum level reached by Brownian motion in the time interval $[0,t]$. Then for $a > 0$, $a \geqslant x$ and all $t \geqslant 0$,*

$$\mathbb{P}(\{M_t \geqslant a\} \cap \{W_t \leqslant x\}) = 1 - \Phi\left(\frac{2a - x}{\sqrt{t}}\right),$$

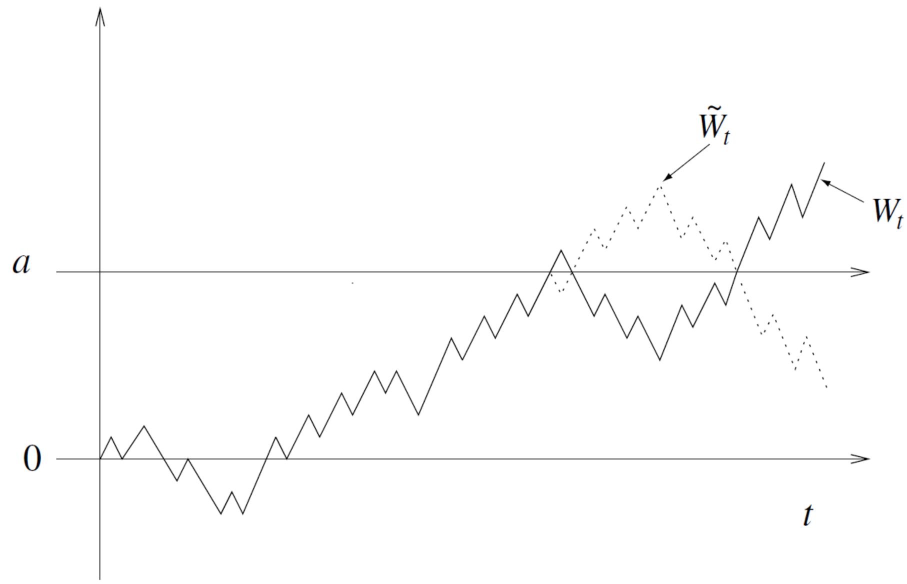


Fig. 6.3 The reflection principle when $T = T_a$.

where $\Phi(x)$ is the c.d.f. of the standard normal distribution $N(0,1)$.

Proof. Notice that M_t is non-negative and non-decreasing in t. Also, if, for $a > 0$, T_a is defined to be the first hitting time of level a, then $\{M_t \geqslant a\} = \{T_a \leqslant t\}$. On $\{T_a \leqslant t\}$, we can take $T = T_a$ in Lemma 6.8 (reflection principle) to obtain $\{W_t \leqslant x\} = \{2W_{T_a} - \widetilde{W}_t \leqslant x\} = \{2a - x \leqslant \widetilde{W}_t\}$. Thus we have

$$\begin{aligned}
\mathbb{P}(\{M_t \geqslant a\} \cap \{W_t \leqslant x\}) &= \mathbb{P}(\{T_a \leqslant t\} \cap \{W_t \leqslant x\}) \\
&= \mathbb{P}\left(\{T_a \leqslant t\} \cap \left\{2a - x \leqslant \widetilde{W}_t\right\}\right) \\
&= \mathbb{P}\left(2a - x \leqslant \widetilde{W}_t\right) \qquad \text{(see explanations below)} \\
&= 1 - \Phi\left(\frac{2a - x}{\sqrt{t}}\right).
\end{aligned}$$

Note that the Reflection Principle is cleverly used in the third equality to facilitate the derivation of the joint density: As we focus on $a \geqslant x$, we have $2a - x \geqslant a$. If $2a - x \leqslant \widetilde{W}_t$, then necessarily $\{W_t\}$ has hit a before time t. (If $\{W_t\}$ hasn't hit a, then from definition, $\widetilde{W}_t < a \leqslant 2a - x$). Now, since $\{W_t\}$ has hit level a, $\{T_a \leqslant t\}$ must have happened. So we only need to consider the probability of the event $\left\{2a - x \leqslant \widetilde{W}_t\right\}$. □

With the distributional result on the first hitting time T_a in Lemma 6.7, we have:

Theorem 6.10. (Divergence and Return) *Brownian motion will almost surely eventually hit any and every real value (no matter how large or negative), i.e.,*

$$\mathbb{P}\left(\sup_{s\geqslant 0} W_s = \infty\right) = \mathbb{P}\left(\inf_{s\geqslant 0} W_s = -\infty\right) = 1\,.$$

Also, no matter how far above the axis, the B.M. will almost surely be back down to zero at some later time.

Proof. Note the equivalence of the events $\{T_a \leqslant t\} = \{\sup_{0\leqslant s\leqslant t} W_s \geqslant a\}$. Since $\{\sup_{s\leqslant t} W_s \geqslant a\}$ is a contracting sequence of events as a increases, and an increasing sequence of events as t increases, we have from Theorem 3.5 that

$$\begin{aligned}
\mathbb{P}\left(\sup_{s\geqslant 0} W_s = \infty\right) &= \lim_{a\to\infty} \mathbb{P}\left(\sup_{s\geqslant 0} W_s \geqslant a\right) \\
&= \lim_{a\to\infty}\lim_{t\to\infty} \mathbb{P}\left(\sup_{0\leqslant s\leqslant t} W_s \geqslant a\right) \\
&= \lim_{a\to\infty}\lim_{t\to\infty} \mathbb{P}\left(T_a \leqslant t\right) \\
&= \lim_{a\to\infty}\lim_{t\to\infty} 2\mathbb{P}\left(W_t \geqslant a\right) \quad \text{(Lemma 6.7)} \\
&= \lim_{a\to\infty}\lim_{t\to\infty} 2\left(1-\Phi(a/\sqrt{t})\right) \\
&= \lim_{a\to\infty} 2\left(1-\Phi(0)\right) \\
&= 1\,.
\end{aligned}$$

The equality $\mathbb{P}\left(\inf_{s\geqslant 0} W_s = -\infty\right) = 1$ can be shown similarly (Exercise 6.5).

Lastly we show the second statement. Given any s, from Example 6.2, $\widetilde{W}_t \triangleq W_{t+s} - W_s$ is a B.M. As we just showed, $\widetilde{W}_t$ hits any $u \in (-\infty,\infty)$ in finite time. No matter how large or negative u is, $\widetilde{W}_t \triangleq W_{t+T_u} - W_{T_u}$ is again a Brownian motion. The fact that $\tilde{W}_t$ hits any value implies that W_t must return to 0. □

6.4.2 *Hitting a level a*

For pricing a perpetual American put option,we shall use the following result.

Lemma 6.11 *Let W_t be a Brownian motion and $T_a := \inf\{t \geqslant 0 : W_t = a\}$. Then for $\theta > 0$, $a > 0$*

$$\mathbb{E}e^{-\theta T_a} = e^{-a\sqrt{2\theta}}\,.$$

Proof. We apply the optional stopping theorem (Theorem 5.9) to the martingale

$$M_t = e^{\sigma W_t - \frac{1}{2}\sigma^2 t}$$

with the stopping time T_a. Since Theorem 6.10 implies $T_a < \infty$ a.s., and $|M_{T_a \wedge t}| < |M_{T_a}| \le e^{\sigma a} < \infty$ a.s., the conditions of Theorem 5.9 are satisfied. Thus, we can conclude that

$$1 = E(M_0) = E(M_{T_a}) = e^{\sigma a}\mathrm{E}(e^{-\frac{1}{2}\sigma^2 T_a}) .$$

The proof is completed by taking $\theta = \frac{1}{2}\sigma^2$. □

6.5 Variation of Brownian Motions

The notion of **variation** of a process will be useful in defining stochastic integrals in the next chapter.

Definition 6.4. (Variation of a function) Let $\Pi = \{t_0, t_1, \dots, t_{N(\Pi)}\}$ ($t_0 = 0 < t_1 < \cdots < t_{N(\Pi)} = T$) be a partition of the interval $[0,T]$, $N(\Pi)$ be the number of intervals partitioned by Π, and $\delta(\Pi)$ be the **mesh** of Π, i.e., the length of the longest interval in Π,

$$\delta(\Pi) = \max_{j=1,\dots,N(\Pi)} |t_j - t_{j-1}| .$$

Then, the p**-variation** of a function f is defined as

$$\lim_{\varepsilon \to 0} \left\{ \sup_{\Pi:\delta(\Pi)=\varepsilon} \sum_{j=1}^{N(\Pi)} |f(t_j) - f(t_{j-1})|^p \right\} . \tag{6.16}$$

First we show the convergence of the second variation, or quadratic variation, of Brownian motions.

Theorem 6.12. (Quadratic Variation of Brownian Motions) *Let W_t be a Brownian motion. Define the quadratic variation (2-variation) w.r.t. Π by*

$$S(\Pi) = \sum_{j=1}^{N(\Pi)} \left| W_{t_j} - W_{t_{j-1}} \right|^2 ,$$

If Π_n is a sequence of partitions of $[0,T]$ such that $\delta(\Pi_n) \to 0$, then

$$E(|S(\Pi_n) - T|^2) \to 0 .$$

Proof. To stress on the dependence of t_js on Π_n, we use $t_{n,j}$ to denote the t_j of Π_n. Since $T = \sum_{j=1}^{N(\Pi_n)} t_{n,j} - t_{n,j-1}$, we have

$$
\begin{aligned}
|S(\Pi_n) - T|^2 &= \left| \sum_{j=1}^{N(\Pi_n)} \left\{ |W_{t_{n,j}} - W_{t_{n,j-1}}|^2 - (t_{n,j} - t_{n,j-1}) \right\} \right|^2 \\
&= \sum_{j=1}^{N(\Pi_n)} \delta_{n,j}^2 + 2 \sum_{j<k} \delta_{n,j} \delta_{n,k}\,, \qquad (6.17)
\end{aligned}
$$

where

$$\delta_{n,j} = |W_{t_{n,j}} - W_{t_{n,j-1}}|^2 - (t_{n,j} - t_{n,j-1})\,.$$

Since Brownian motions have independent increment, we have

$$\mathrm{E}(\delta_{n,j}\delta_{n,k}) = \mathrm{E}(\delta_{n,j})\mathrm{E}(\delta_{n,k}) = 0\,, \quad \text{if} \;\; j \neq k\,. \qquad (6.18)$$

Also, using the fact that $E(X^2) = \sigma^2$ and $E(X^4) = 3\sigma^4$ for $X \sim N(0, \sigma^2)$ (Exercise 6.12), we have

$$
\begin{aligned}
\mathrm{E}(\delta_{n,j}^2) &= \mathrm{E}\left[|W_{t_{n,j}} - W_{t_{n,j-1}}|^4 - 2|W_{t_{n,j}} - W_{t_{n,j-1}}|^2 (t_{n,j} - t_{n,j-1}) + (t_{n,j} - t_{n,j-1})^2 \right] \\
&= 3(t_{n,j} - t_{n,j-1})^2 - 2(t_{n,j} - t_{n,j-1})^2 + (t_{n,j} - t_{n,j-1})^2 \\
&= 2(t_{n,j} - t_{n,j-1})^2\,. \qquad (6.19)
\end{aligned}
$$

Combing (6.18) and (6.19), taking expectation on both sides of (6.17) gives

$$
\begin{aligned}
\mathrm{E}|S(\Pi_n) - T|^2 &= \sum_{j=1}^{N(\Pi_n)} \mathrm{E}(\delta_{n,j}^2) + 2 \sum_{j<k} \mathrm{E}(\delta_{n,j}\delta_{n,k}) \\
&= \sum_{j=1}^{N(\Pi_n)} 2(t_{n,j} - t_{n,j-1})^2 \\
&\leqslant 2\delta(\Pi_n) \sum_{j=1}^{N(\Pi_n)} (t_{n,j} - t_{n,j-1}) \\
&= 2\delta(\Pi_n) T \to 0\,,
\end{aligned}
$$

since T is fixed and $\delta(\Pi_n) \to 0$. Thus the proof is complete. □

Motivated by the above theorem, we have the following general definition.

Definition 6.5. (Quadratic Variation Process) Suppose that $\{M_t\}$ is a martingale. The **quadratic variation process** associated with $\{M_t\}_{t \geqslant 0}$ is the process $\{[M]_t\}_{t \geqslant 0}$ such that for any sequence of partitions $\{\Pi_n\}$ of $[0, T]$ with $\delta(\Pi_n) \to 0$,

$$\mathrm{E}\left[\left| \sum_{j=1}^{N(\Pi_n)} |M_{t_j} - M_{t_{j-1}}|^2 - [M]_T \right|^2 \right] \to 0\,,$$

as $n \to \infty$.

Corollary 6.13 *Theorem 6.12 shows that the quadratic variation process of the standard Brownian motion W_t is*

$$[W]_T = T\,.$$

Corollary 6.14 (First Variation of Brownian Motions) *The first variation of Brownian motions is ∞ for any interval $[0,T]$.*

Proof. Suppose on the contrary that the first variation is not ∞. Then, there exist some T and some K such that

$$\lim_{\delta\to 0}\left\{\sup_{\Pi:\delta(\Pi)=\delta}\sum_{j=1}^{N(\Pi)}|W_{t_j}-W_{t_{j-1}}|\right\}=K<\infty\,,$$

where Π is a partition of $[0,T]$. Then

$$\begin{aligned}\lim_{\delta\to 0}\left\{\sup_{\Pi:\delta(\Pi)=\delta}\sum_{j=1}^{N(\Pi)}|W_{t_j}-W_{t_{j-1}}|^2\right\} &\leq \lim_{\delta\to 0}\left\{\sup_{\substack{\Pi:\delta(\Pi)=\delta\\ j=1,\dots,N(\Pi)}}|W_{t_j}-W_{t_{j-1}}|\sup_{\Pi:\delta(\Pi)=\delta}\sum_{j=1}^{N(\Pi)}|W_{t_j}-W_{t_{j-1}}|\right\}\\ &\leqslant K\lim_{\delta\to 0}\left\{\sup_{\Pi:\delta(\Pi)=\delta}|W_{t_j}-W_{t_{j-1}}|\right\}\\ &\to 0 \qquad\qquad (\text{continuity of } W_t)\,,\end{aligned}$$

contradicting Corollary 6.13 that the quadratic variation of B.M. is T. Thus the first variation of B.M. is ∞.

6.6 Exercises

Exercise 6.1. Show that the process $\widetilde{W}_t$ in Example 6.2 is a standard Brownian motion.

Exercise 6.2. Based on the probability measure in Definition 6.1, can you deduce property 1 and 2 in Definition 6.2 and the transition probability in Definition 6.3?

Exercise 6.3. Using integration by parts, prove that if $\{W_t : t \geqslant 0\}$ is a standard Brownian motion under $\mathbb{P}$, then for $x>0$,

$$\mathbb{P}\{W_t \geqslant x\} \equiv \int_x^{\infty}\frac{1}{\sqrt{2\pi t}}\exp\left\{-\frac{y^2}{2t}\right\}dy \leqslant \frac{\sqrt{t}}{x\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2t}\right\}.$$

Also, show that, for $n\geqslant 4$,

$$\exp\left\{-2^{\frac{n}{2}+1}\right\} < 2^{-2n+2}\,.$$

Exercise 6.4. Find the third variation of a standard Brownian motion.

Exercise 6.5. If W_t is a Brownian motion, show that $\mathbb{P}\left(\inf_{s\geqslant 0} W_s = -\infty\right) = 1$.

Exercise 6.6. Prove Lemma 6.2.

Exercise 6.7. Provide details for Example 6.3.

Exercise 6.8. Let Z be normally distributed with mean zero and variance one under the measure $\mathbb{P}$. What is the distribution of $\sqrt{t}Z$? Is the process $X_t := \sqrt{t}Z$ a Brownian motion?

Exercise 6.9. Suppose that W_t and $\widetilde{W}_t$ are independent Brownian motions under the measure $\mathbb{P}$ and let $\rho \in [-1,1]$ be a constant. Is the process $X_t := \rho W_t + \sqrt{1-\rho^2}\widetilde{W}_t$ a Brownian motion?

Exercise 6.10. Let $\{W_t : t \geqslant 0\}$ be standard Brownian motion under the measure $\mathbb{P}$. Which of the following are $\mathbb{P}$-Brownian motions?

1. $\{-W_t : t \geqslant 0\}$,
2. $\left\{cW_{t/c^2} : t \geqslant 0\right\}$, where c is some non-zero constant,
3. $\left\{\sqrt{t}W_1 : t \geqslant 0\right\}$,
4. $\{W_{2t} - W_t : t \geqslant 0\}$.

Justify your answers.

Exercise 6.11. Let $\{W_t : t \geqslant 0\}$ be standard Brownian motion under the measure $\mathbb{P}$, and let $\{\mathscr{F}_t\}$, $t \geqslant 0$, be filtration for this Brownian motion. Show that $W_t^2 - t$ is a $\{\mathbb{P}, \{\mathscr{F}_t\}\}$ martingale.

Exercise 6.12. Suppose that X is normally distributed with mean μ and variance σ^2. Calculate $\mathbb{E}\exp\{\theta X\}$ and hence evaluate $\mathbb{E}X^2$ and $\mathbb{E}X^4$.

Exercise 6.13. Brownian motion is not adequate as a stock market model. First, it has constant mean, whereas the company stocks usually grow at some rates. Moreover, it may be too 'noisy' (variance of the path increments is large) or not noisy enough. We can scale to change the 'noisiness' and we can artificially introduce a drift, but this is still not a good model, since it may take negative values. Suppose that $\{W_t : t \geqslant 0\}$ is standard Brownian motion under $\mathbb{P}$. Define a new process$\{S_t : t \geqslant 0\}$ by $S_t := \mu t + \sigma W_t$ where $\sigma > 0$ and $\mu \in \mathbb{R}$ are constants. Show that for all values of $\sigma > 0$, $\mu \in \mathbb{R}$ and $T > 0$ there is a positive probability that S_T is negative.

Exercise 6.14. Let $\{W_t\}$ be a standard Brownian motion and $\{\mathscr{F}_t\}$ be the natural filtration of the B.M.. Show that $e^{\sigma W_t - \frac{\sigma^2}{2}t}$ is a $(\mathbb{P}, \{\mathscr{F}_t\})$ martingale.

Exercise 6.15. Let $\{W_t : t \geqslant 0\}$ be standard Brownian motion under $\mathbb{P}$. For $a > 0$, let T_a be the 'hitting time of level a', i.e., $T_a = \inf\{t \geqslant 0 : W_t = a\}$. Use the fact $\mathbb{E}\exp\{-\theta T_a\} = \exp\left\{-a\sqrt{2\theta}\right\}$ to calculate

1. $\mathbb{E}T_a$,
2. $\mathbb{P}\{T_a < \infty\}$.

Exercise 6.16. Let $\{W_t : t \geqslant 0\}$ be standard Brownian motion under $\mathbb{P}$ and define $\{M_t : t \geqslant 0\}$ by $M_t := \max_{s\in[0,t]} W_s$ Suppose that $x \geqslant a$. Calculate

1. $\mathbb{P}(\{M_t \geqslant a\} \cap \{W_t \geqslant x\})$,
2. $\mathbb{P}(\{M_t \geqslant a\} \cap \{W_t \leqslant x\})$.

Exercise 6.17. Let $\{W_t : t \geqslant 0\}$ be standard Brownian motion under $\mathbb{P}$ and define $\{M_t : t \geqslant 0\}$ by $M_t := \max_{s\in[0,t]} W_s$.

1. Show that M_t, $|W_t|$ and $M_t - W_t$ have the same marginal distribution, with density
$$f_{M_t}(z) = \frac{2}{\sqrt{t}}\phi\left(\frac{z}{\sqrt{t}}\right)\mathbf{1}_{(0,\infty)}(z),$$
where $\phi(\cdot)$ is the p.d.f. of $N(0,1)$.
2. $\mathbb{E}M_t = \sqrt{\frac{2t}{\pi}}$.

Exercise 6.18. Find the probability density function of the stopping time $T_a = \inf\{t \geqslant 0 : W_t = a\}$.

Exercise 6.19. Let $\{W_t : t \geqslant 0\}$ be standard Brownian motion under $\mathbb{P}$. Derive the joint distribution of W_t and $m_t := \min_{s\in[0,t]} W_s$. (Hint: Consider $-W_t$.)

Exercise 6.20. Let $\{W_t : t \geqslant 0\}$ be standard Brownian motion under $\mathbb{P}$, and $s \in [0,t)$. Show that the conditional distribution of W_s given $W_t = b$ is Normal and give its mean and variance.

Exercise 6.21. Let (X,Y) be a pair of random variables with joint density function
$$f_{X,Y}(x,y) = \frac{2|x|+y}{\sqrt{2\pi}}\exp\left\{-\frac{(2|x|+y)^2}{2}\right\}\mathbf{1}_{\{y\geqslant -|x|\}}.$$
Show that X and Y are standard normal random variables and that they are uncorrelated but not independent.

Exercise 6.22. Let $\{W_t : t \geqslant 0\}$ be standard Brownian motion under $\mathbb{P}$. For the Gaussian process
$$Y_t = e^{-\alpha t} W_{\exp\{2\alpha t\}},$$
find its mean and covariance functions, $\mu(t) = \mathrm{E}(Y_t)$ and $\gamma_k = \mathrm{Cov}(Y_t, Y_{t+k})$.

Exercise 6.23. In attempt to show the almost surely continuity of Brownian motions, Keith provided the following argument:

$$
\begin{aligned}
&\mathbb{P}\left(\lim_{t\to 0} W_t = 0\right) \\
&= \mathbb{P}\left(\bigcap_{n\geqslant 1}\left\{\lim_{t\to 0}|W_t| < \frac{1}{n}\right\}\right) \\
&= \lim_{n\to\infty}\mathbb{P}\left(\left\{\lim_{t\to 0}|W_t| < \frac{1}{n}\right\}\right) \\
&= \lim_{n\to\infty}\mathbb{P}\left(\left\{\text{some } t^*, \forall t \leqslant t^*, |W_t| < \frac{1}{n}\right\}\right) \\
&= \lim_{n\to\infty}\mathbb{P}\left(\bigcup_{t^*\in\mathbb{R}}\bigcap_{t\leqslant t^*}\left\{|W_t| < \frac{1}{n}\right\}\right) \\
&\geqslant \lim_{n\to\infty}\mathbb{P}\left(\bigcup_{m=1}\bigcap_{t\leqslant 1/m}\left\{|W_t| < \frac{1}{n}\right\}\right) \\
&= \lim_{n\to\infty}\lim_{m\to\infty}\mathbb{P}\left(\bigcap_{t\leqslant 1/m}\left\{|W_t| < \frac{1}{n}\right\}\right) \\
&\geqslant \lim_{n\to\infty}\lim_{m\to\infty}\mathbb{P}\left(\left\{|W_{1/m}| < \frac{1}{n}\right\}\right) \\
&\geqslant \lim_{n\to\infty}\lim_{m\to\infty}[\Phi(\sqrt{m}/n) - \Phi(-\sqrt{m}/n)] \\
&\geqslant 1\,.
\end{aligned}
$$

Is the argument correct? Explain.

Exercise 6.24. In this exercise we describe a second proof of the fact $\mathbb{P}(\sup_{t\geqslant 0} W_t = \infty) = 1$. Let $Z := \sup_{t\in[0,\infty)} W_t$ and $\tilde{Z}$ is an independent and identical copy of Z.

1. Show that for any $c > 0$, cZ has the same distribution as Z.
2. Using the above result, show that, $P(Z \in [1,n)) = 0$ and $P(Z \in (1/n, 1]) = 0$ for all $n \geqslant 1$. Thus show that with probability one, $Z \in \{0, \infty\}$.
3. By noting that $\{Z = 0\} \subset \{W_1 \leq 0\} \cap \{\sup_{t\geqslant 0} W_{t+1} - W_1 = 0\}$ and that $W_{t+1} - W_1$ is a B.M., argue that
$$P(Z=0) \leqslant P(W_1 \leq 0)P(\tilde{Z}=0) = P(W_1 \leq 0)P(Z=0)\,. \tag{6.20}$$
Hence deduce that $P(Z = \infty) = 1$.
4. Using similar argument, show that $P(\inf_{t\in[0,\infty)} W_t = -\infty) = 1$.

Exercise 6.25. Show that if a continuous function $f(x)$ satisfies $f(a+b) = f(a)f(b)$ and $f(1) \neq 0$, then $f(x) = e^{cx}$ for some c.

Chapter 7
Ito's Stochastic Calculus

7.1 Introduction

When Bachelier first applied Wiener processes to model the fluctuation of asset prices, the price of an asset at time t, X_t, has an infinitesimal increment dX_t proportional to the increment dW_t of the Wiener process, i.e.,

$$dX_t = \sigma\, dW_t,$$

where σ is a positive constant. As a result, an asset with initial price $X_0 = x$ is worth

$$X_t = x + \sigma W_t$$

at time t. This model suffers from one serious flaw: for any $t > 0$ the price X_t can be negative with non-zero probability, contradicting that the actual stock prices are never negative. To tackle this problem, successors assumed that the relative price dX_t/X_t of an asset is proportional to dW_t, i.e.,

$$dX_t = \sigma X_t\, dW_t. \tag{7.1}$$

Although this equation looks like a differential equation, traditional methods are no longer applicable because the paths of W_t are not differentiable (Theorem 6.4). A way to tackle the obstacle was found in the 1940s by Ito, who gave a rigorous meaning to (7.1) by writing it as

$$X_t = x + \sigma \int_0^t X_s\, dW_s, \tag{7.2}$$

where the integral with respect to W_t on the right-hand side is called the *Ito stochastic integral*. In this section we discuss the definition and properties of stochastic integral, and explore its applications to pricing financial products.

7.2 Ito Stochastic Integral

7.2.1 Motivations

Motivated by (7.2), we construct the Ito stochastic integral of the form $\int_0^T f(t)\,dW_t$ for some stochastic process/random function f. For simplicity we denote the stochastic process as f or $f(t)$, $t \in \mathbb{R}$, instead of the precise form, $f(t,\omega)$, which indicates the randomness. We follow an approach similar to constructing Riemann integral, i.e., define the integral by the limit of the discretized version

$$\sum_{i=0}^{n-1} f(s_i)(W_{t_{i+1}} - W_{t_i})\,, \tag{7.3}$$

where $s_i \in [t_i, t_{i+1}]$, $0 = t_0 < t_1 < \cdots < t_n = T$. The major differences between Riemann and Ito integrals are:

1) Riemann integration gives a real number, but Ito integration gives a random variable (since $f(t)$ and W_t are random). Thus, while defining Riemann integral involves convergence of real numbers, defining Ito integral in (7.3) requires convergence of random variables, which is considerably more difficult.
2) If the Riemann integral exists, then s_i can be an arbitrary point in $[t_i, t_{i+1}]$ since the upper and lower Riemann sum converge. However, in Ito integral, the limit will be different depending on the choice of s_i. This is due to the non-zero quadratic variation of the Brownian motion W_t; see Exercise 7.1.

To avoid the ambiguity in 2), the definition of stochastic integral fixes the choice $s_i = t_i$ for each i in the approximating sum (7.3). The choice $s_i = t_i$ is natural if we regard $f(t)$ as the trading strategy and W_t as the stock price: For the $i+1$-th period $[t_i, t_{i+1}]$, the trading strategy should only depend on the information up to time t_i. Hence, $f(t_i)$ units are invested and a profit of $f(t_i)(W_{t_{i+1}} - W_{t_i})$ is made. Therefore, $\sum_{i=0}^{n-1} f(t_i)(W_{t_{i+1}} - W_{t_i})$ represents the total profit.

Remark 7.1. **(Previsible)** To be precise, in defining stochastic integral we require the integrand f to be **previsible** or **predictable**, i.e., $f(t)$ is $\mathscr{F}_{t-}$ for all t, where $\mathscr{F}_{t-} = \sigma(\bigcup_{s<t} \mathscr{F}_s)$ and $\mathscr{F}_t = \sigma(\{W_s, s \le t\})$. Indeed, it can be shown that if f is **continuous** and adapted to $\mathscr{F}_t$, then f is automatically previsible. Since we mainly deal with continuous integrands (e.g. functions of Brownian motions), we do not distinguish between previsible and adapted process.

7.2.2 Ito Integral

Ito integral $\int_0^T f(t)\,dW_t$ is a random variable since W_t and the integrand $f(t)$ are random. To ensure the regularity of the Ito integral (such as the existence of the first and second moments), we restrict $f(t)$ to the following class of stochastic processes.

Definition 7.1. ($\mathscr{M}_T^2$ and $\mathscr{M}^2$ Stochastic Processes) Denote $\mathscr{M}_T^2$ as the class of stochastic processes $f(t)$, $t \geq 0$, such that

$$\mathrm{E}\left(\int_0^T |f(t)|^2\,dt\right) < \infty .$$

Let $\mathscr{M}^2$ be the class of stochastic processes $f(t)$ such that $f(t) \in \mathscr{M}_T^2$ for any $T > 0$. Recall that a random variable X is in $\mathscr{L}^2$, or $X \in \mathscr{L}^2$, if $\mathrm{E}|X|^2 < \infty$. Both $\mathscr{M}^2$ and $\mathscr{L}^2$ are related to the existence of second moment, but $\mathscr{M}^2$ is for a stochastic process and $\mathscr{L}^2$ is for a random variable. Later in Theorem 7.2 b), we will see that the restriction $f \in \mathscr{M}_T^2$ guarantees the existence of the second moment of the Ito integral $\int_0^T f(t)\,dW_t$.

Since the Ito integral is defined as a limit of random variables of the form (7.3), we need to specify a measure of distance and a mode of convergence. As the second moment of Brownian motions is easy to compute, we choose the $\mathscr{L}^2$ distance $d_2(X,Y) = \sqrt{E(|X-Y|^2)}$. For notational simplicity, we define the $\mathscr{L}^2$ **norm** as

$$\|X\|_{\mathscr{L}^2} = d_2(X,0) = \sqrt{E\left(X^2\right)}, \tag{7.4}$$

so that $d_2(X,Y) = \|X-Y\|_{\mathscr{L}^2}$.

Since the integrand f is a stochastic process, we analogously define a measure of distance, norm, and a mode of convergence for stochastic processes.

Definition 7.2. ($\mathscr{M}_T^2$ distance, $\mathscr{M}_T^2$ norms) For two **stochastic processes** f and g, the $\mathscr{M}_T^2$ norm and distance are given respectively by

$$\|f\|_{\mathscr{M}_T^2} = \sqrt{E\left(\int_0^T |f(t)|^2\,dt\right)} \quad \text{and} \quad \|f-g\|_{\mathscr{M}_T^2} = \sqrt{E\left(\int_0^T |f(t)-g(t)|^2\,dt\right)}.$$

□

Definition 7.3. ($\mathscr{M}_T^2$ and $\mathscr{M}^2$ Convergences) A sequence of **stochastic processes** $\{f_n\}$ **converges in** $\mathscr{M}_T^2$ to f if for any $T > 0$,

$$\|f_n - f\|_{\mathscr{M}_T^2} = \sqrt{\mathrm{E}\left(\int_0^T |f_n(t)-f(t)|^2\,dt\right)} \to 0 .$$

Similarly, $\{f_n\}$ **converges in** $\mathscr{M}^2$ to f if $\{f_n\}$ converges to f in $\mathscr{M}_T^2$ for all T. □

Using Definitions 7.1 and 7.3, we define the Ito Integral on the class of stochastic process in $\mathscr{M}^2$ as follows:

Definition 7.4. (Ito Integral) For any $T > 0$ and any stochastic process $f \in \mathscr{M}^2$, the **stochastic integral** of f on $[0,T]$ is defined by

$$I_T(f) = \lim_{n\to\infty} \sum_{j=0}^{n-1} f(t_j)\left(W_{t_{j+1}} - W_{t_j}\right),$$

where $(0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T)$ is any partition of $[0,T]$ with $\max_j |t_j - t_{j-1}| \to 0$ as $n \to \infty$. We write $I_T(f)$ as

$$I_T(f) = \int_0^T f(t)\,dW_t\,, \qquad \left(\text{or } \int_0^T f\,dW\right).$$

The following theorem justifies the above definition of Ito integral. In particular, it shows the existence of the limit, and that the limit does not depend on how the partition is chosen (uniqueness).

Theorem 7.1. (Existence and Uniqueness of Ito Integral) *Suppose that a function $f \in \mathscr{M}^2$ satisfies the following assumptions: For all $t \geq 0$,*

A1) f is almost surely continuous, i.e., $\mathbb{P}(\lim_{\varepsilon\to 0} |f(t+\varepsilon) - f(t)| = 0) = 1$,
A2) f is adapted to the filtration $\{\mathscr{F}_t\}_{t\geq 0}$, where $\mathscr{F}_t = \sigma(\{W_s, s \leq t\})$.

Then, for any $T > 0$, the Ito integral

$$I_T(f) = \int_0^T f(t)\,dW_t$$

in Definition 7.4 exists and is unique, almost surely.

Proof. The proof proceeds in three steps:

1) Construct a sequence of adapted stochastic processes $\{f_n\}$ such that $\|f - f_n\|_{\mathscr{M}^2} \to 0$.
2) Using Cauchy criterion, show that $\{I_T(f_n)\}$ has a $\mathscr{L}^2$ limit. That is, there exists a random variable I_∞ such that $\|I_T(f_n) - I_\infty\|_{\mathscr{L}^2} \to 0$.
3) Define $I_T(f) = I_\infty$ and show the a.s. uniqueness of the limit $I_T(f)$.

The details of the three steps are provided as follows.

1) First we find a sequence of $\mathscr{M}_T^2$ functions $f_1, f_2, \ldots$ such that $\|f - f_n\|_{\mathscr{M}_T^2} = \mathrm{E}(\int_0^T |f_n(t) - f(t)|^2\,dt) \to 0$. From Assumption A1 we can assume that f is continuous on $\Omega_c \subseteq \Omega$ where $\mathbb{P}(\Omega_c) = 1$. Define, for each $\omega \in \Omega_c$ (recall $f_n(t) = f_n(t,\omega)$ and $f(t) = f(t,\omega)$),

$$f_n(t) = \begin{cases} n\int_{\frac{k-1}{n}}^{\frac{k}{n}} f(s)\,ds\,, & \text{if } t \in [\frac{k}{n}, \frac{k+1}{n}),\ k = 1,2,\ldots,[Tn]-1\,,\\ 0\,, & \text{otherwise}\,. \end{cases} \tag{7.5}$$

By construction, $f_n(t)$ is adapted to $\mathscr{F}_t$, and is a step function for each ω. (Function of this kind is called **random step function**). More importantly, (7.5) and Hölder's Inequality implies

$$\int_{\frac{k}{n}}^{\frac{k+1}{n}} |f_n(t)|^2\,dt \overset{(7.5)}{=} n\left|\int_{\frac{k-1}{n}}^{\frac{k}{n}} f(t)\,dt\right|^2 \overset{Hölder}{\leq n} \int_{\frac{k-1}{n}}^{\frac{k}{n}} 1^2\,dt \int_{\frac{k-1}{n}}^{\frac{k}{n}} |f(t)|^2\,dt = \int_{\frac{k-1}{n}}^{\frac{k}{n}} |f(t)|^2\,dt\,. \tag{7.6}$$

As (7.6) holds for all $\omega \in \Omega_c$, we can say that (7.6) holds almost surely.

Next we show $\|f - f_n\|^2_{\mathscr{M}_T^2} = \mathrm{E}(\int_0^T |f_n(t) - f(t)|^2\, dt) \to 0$. By (7.5) and the a.s. continuity of $f(t)$, we have pointwise convergence $f_n(t) - f(t) \overset{a.s}{\to} 0$ as $n \to \infty$. As the domain $[0,T]$ is bounded, for sufficiently large n, we have $|f_n(t) - f(t)| \leq 1$ for all $t \in [0,T]$, a.s. Hence, the Dominant Convergence Theorem (DCT) implies

$$\lim_{n\to\infty} \int_0^T |f_n(t) - f(t)|^2\, dt = 0 \quad a.s. \tag{7.7}$$

Let $Y_n = \int_0^T |f_n(t) - f(t)|^2\, dt$. Note that (7.7) means that $\lim_{n\to\infty} Y_n \overset{a.s.}{=} 0$. Also, note that

$$Y_n \leq 2\int_0^T \left(|f(t)|^2 + |f_n(t)|^2\right)\, dt \leq 4\int_0^T |f(t)|^2\, dt \triangleq \bar{Y}\,, \tag{7.8}$$

where the second inequality follows from summing over all k in (7.6). Since $Y_n \overset{a.s.}{\to} 0$, $|Y_n| \leq \bar{Y}$ a.s. and $E\bar{Y} < \infty$ by the definition of $\mathscr{M}_T^2$, we can apply Dominant Convergence Theorem (DCT) to obtain $\mathrm{E}(Y_n) \to 0$, i.e.,

$$\|f - f_n\|^2_{\mathscr{M}_T^2} = \mathrm{E}\left(\int_0^T |f_n(t) - f(t)|^2\, dt\right) \to 0\,. \tag{7.9}$$

2) To show the existence of Ito Integral, we need to show that $I_T(f_n) = \int_0^T f_n(t)\, dW_t$ converges to an element in $\mathscr{L}^2$. Since $f_n(t)$ is a step function taking constant values in each interval $[\frac{k}{n}, \frac{k+1}{n})$, we can write

$$I_T(f_n) = \sum_{k\geq 1} f_n\left(\frac{k}{n}\right)\left(W_{\frac{k+1}{n}} - W_{\frac{k}{n}}\right)\,.$$

Note that

$$
\begin{aligned}
&\|I_T(f_n)\|^2_{\mathscr{L}^2} \\
&= \mathrm{E}\left(\sum_{k=0}^{[Tn]-1} f_n\left(\frac{k}{n}\right)\left(W_{\frac{k+1}{n}} - W_{\frac{k}{n}}\right)\right)^2 \\
&= \mathrm{E}\left(\sum_{k=0}^{[Tn]-1}\sum_{j=0}^{[Tn]-1} f_n\left(\frac{k}{n}\right) f_n\left(\frac{j}{n}\right)\left(W_{\frac{k+1}{n}} - W_{\frac{k}{n}}\right)\left(W_{\frac{j+1}{n}} - W_{\frac{j}{n}}\right)\right) \\
&= \sum_{k=0}^{[Tn]-1} \mathrm{E}\left(f_n\left(\frac{k}{n}\right)^2\right) \mathrm{E}\left(W_{\frac{k+1}{n}} - W_{\frac{k}{n}}\right)^2 \quad \left(\text{Independent increment and } f_n\left(\frac{k}{n}\right) \in \mathscr{F}_{\frac{k}{n}}\right) \\
&= \sum_{k=0}^{[Tn]-1} \mathrm{E}\left(f_n\left(\frac{k}{n}\right)^2\right) \frac{1}{n} \qquad (\text{Stationary increment and } \mathrm{E}(W_t^2) = t) \\
&= \mathrm{E}\int_0^T f_n(t)^2\,dt \qquad (\text{Riemann integral}) \\
&= \|f_n\|^2_{\mathscr{M}_T^2}\,.
\end{aligned} \tag{7.10}
$$

Similarly, it can be shown that

$$
\|I_T(f_n) - I_T(f_m)\|^2_{\mathscr{L}^2} = \|f_n - f_m\|^2_{\mathscr{M}_T^2}\,. \tag{7.11}
$$

From (7.9), for any $\varepsilon > 0$, there is an N such that $\|f - f_n\|^2_{\mathscr{M}_T^2} < \frac{\varepsilon}{2}$ for all $n > N$. Thus for $n, m > N$,

$$
\begin{aligned}
\|I_T(f_n) - I_T(f_m)\|^2_{\mathscr{L}^2} &= \|f_n - f_m\|^2_{\mathscr{M}_T^2} \\
&\le \|f_n - f\|^2_{\mathscr{M}_T^2} + \|f_m - f\|^2_{\mathscr{M}_T^2} \quad (\text{Triangular Inequality}) \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\,.
\end{aligned}
$$

The sequence $\{I_T(f_n)\}_{n\ge 1}$, satisfying $\|I_T(f_n) - I_T(f_m)\|^2_{\mathscr{L}^2} \to 0$ as $n, m \to \infty$, is thus a **Cauchy sequence**. From Theorem 3.7, $\{I_T(f_n)\}_{n\ge 1}$ has a limit, which we call $I_T(f)$.

3) Finally, we show the uniqueness of the limit. Suppose that there are two sequences of stochastic processes $\{f_i^{(1)}\}_{i\ge 1}$ and $\{f_i^{(2)}\}_{i\ge 1}$ satisfying $\mathrm{E}(\int_0^T |f(t) - f_n^{(j)}|^2\,dt) \xrightarrow{n\to\infty} 0$ for $j = 1, 2$. We need to show that $I_T(f_n^{(j)})$, $j = 1, 2$, converge to the same limit. Introduce a new sequence

$$
\{g_i\}_{i\ge 1} = \{f_1^{(1)}, f_1^{(2)}, f_2^{(1)}, f_2^{(2)}, f_3^{(1)}, \ldots\}\,.
$$

By construction, $\{g_i\}_{i\ge 0}$ satisfies $\mathrm{E}(\int_0^T |f(t) - g_n|^2\,dt) \xrightarrow{n\to\infty} 0$. Therefore, the arguments in 2) show that $I_T(g_n)$ converges to some limit. Note that if a sequence converges, then every subsequence converges to the same limit (Exercise 7.2).

Thus $I_T(f_n^{(1)})$ and $I_T(f_n^{(2)})$ do converge to the same limit, completing the proof of uniqueness. □

Remark 7.2. 1) The construction (7.5) is essential to yield (7.6), which in turn yields the bound $\bar{Y}$ in (7.8) for DCT. Therefore, we cannot replace $n\int^{\frac{k}{n}} f(s)\,ds$ in (7.5) by $f\left(\frac{k}{n}\right)$, which gives a simpler adapted step function approximation of f.

2) The proof of uniqueness was in $\mathscr{L}^2$ sense, i.e., if $I_T(f_n^{(j)}) \xrightarrow{\mathscr{L}^2} I_T(f^{(j)})$ for $j = 1,2$, then $\|I_T(f^{(1)}) - I_T(f^{(2)})\|^2_{\mathscr{L}^2} = 0$. However, $\|I_T(f^{(1)}) - I_T(f^{(2)})\|^2_{\mathscr{L}^2} = 0$ implies that $I_T(f^{(1)}) = I_T(f^{(2)})$ almost surely (See Exercise 7.13). Thus, the definition of Ito integral is indeed unique a.s..

Example 7.1. To show the existence of $\int_0^T W_t\,dW_t$, from Theorem 7.1 we need to show that the Wiener process $\{W_t\}$ belongs to $\mathscr{M}^2$. Since for all $T > 0$,

$$\mathrm{E}\left(\int_0^T |W_t|^2\,dt\right) = \int_0^T \mathrm{E}(|W_t|^2)\,dt = \int_0^T t\,dt < \infty\,.$$

Thus $\{W_t\}$ belongs to $\mathscr{M}^2$. Also, from the a.s. continuity and adaptedness of W_t, Assumptions A1 and A2 of Theorem 7.1 hold. Hence the existence of the Ito integral $\int_0^T W_t\,dW_t$ is justified.

Example 7.2. We derive the formula $\int_0^T W_t\,dW_t = \frac{1}{2}W_T^2 - \frac{1}{2}T$ directly from definition, i.e., by approximating the integrand by random step functions. Fix $T > 0$ and $t_i^n = \frac{iT}{n}$. Set

$$f_n(t) = \sum_{i=0}^{n-1} W_{t_i^n}\mathbf{1}_{[t_i^n, t_{i+1}^n)}(t)\,.$$

Then the sequence $f_1, f_2, \cdots \in \mathscr{M}_T^2$ approximates f, since

$$\begin{aligned}
\mathbb{E}\left(\int_0^T |f(t) - f_n(t)|^2\,dt\right) &= \sum_{i=0}^{n-1}\int_{t_i^n}^{t_{i+1}^n} \mathbb{E}\left(|W_t - W_{t_i^n}|^2\right)\,dt \\
&= \sum_{i=0}^{n-1}\int_{t_i^n}^{t_{i+1}^n} (t - t_i^n)\,dt \\
&= \frac{1}{2}\sum_{i=0}^{n-1}\left(t_{i+1}^n - t_i^n\right)^2 \\
&= \frac{1}{2}\frac{T^2}{n} \to 0 \quad \text{as } n \to \infty.
\end{aligned}$$

From Theorem 7.1, $\lim_{n\to\infty} I_T(f_n)$ exists in $\mathscr{L}^2$. To find an explicit formula of the limit, note from the equation $a(b-a) = \frac{1}{2}(b^2 - a^2) - \frac{1}{2}(b-a)^2$ that

$$\begin{aligned}
I_T(f_n) &= \sum_{i=0}^{n-1} W_{t_i^n}\left(W_{t_{i+1}^n} - W_{t_i^n}\right) \\
&= \frac{1}{2}\left[\sum_{i=0}^{n-1}\left(W_{t_{i+1}^n}^2 - W_{t_i^n}^2\right) - \sum_{i=0}^{n-1}(W_{t_{i+1}^n} - W_{t_i^n})^2\right] \\
&= \frac{1}{2}W_T^2 - \frac{1}{2}\sum_{i=0}^{n-1}(W_{t_{i+1}^n} - W_{t_i^n})^2 \to \frac{1}{2}W_T^2 - \frac{1}{2}T
\end{aligned}$$

in $\mathscr{L}^2$ as $n \to \infty$. The last convergence follows from the fact that the quadratic variation of Wiener processes is T. Therefore, we conclude that

$$\int_0^T W_t\, dW_t = I_T(f) = \frac{1}{2}W_T^2 - \frac{1}{2}T\,.$$

□

7.3 Properties of the Stochastic Integral

The basic properties of the Ito integral are summarized in the following theorem:

Theorem 7.2. *For any $f, g \in \mathscr{M}^2$, any $\alpha, \beta \in \mathbb{R}$ and any $0 \leqslant s < T$, the Ito integral satisfies:*

a) **Linearity***:*

$$\int_0^T (\alpha f(t) + \beta g(t))\, dW_t = \alpha \int_0^T f(t)\, dW_t + \beta \int_0^T g(t)\, dW_t; \tag{7.12}$$

b) **Isometry***:*

$$\mathbb{E}\left(\left|\int_0^T f(t)\, dW_t\right|^2\right) = \mathbb{E}\left(\int_0^T |f(t)|^2\, dt\right); \tag{7.13}$$

c) **Martingale Property***:*

$$\mathbb{E}\left(\int_0^T f(t)\, dW_t \middle| \mathscr{F}_s\right) = \int_0^s f(t)\, dW(t).$$

In particular, $\mathbb{E}\left(\int_0^T f(t)\, dW_t\right) = 0$.

Proof. a) If f and g belong to $\mathscr{M}^2$, then from the proof of Theorem 7.1, they can be approximated by some sequences $f_1, f_2, \cdots$ and $g_1, g_2, \cdots$. From Definition 7.4, it is clear that, for each $n \in \mathbb{Z}^+$,

$$I_T(\alpha f_n + \beta g_n) = \sum_{j=0}^{n-1} (\alpha f(t_j^n) + \beta g(t_j^n)) \left(W_{t_{j+1}^n} - W_{t_j^n} \right) = \alpha I_T(f_n) + \beta I_T(g_n)\,, \tag{7.14}$$

where $\{t_j^n\}_{j=0,1,\ldots,n}$ is the "common" grid where both f_n and g_n are constant in each interval $[t_{j-1}^n, t_j^n)$. Taking $n \to \infty$ on both sides of (7.14), we obtain

$$I(\alpha f + \beta g) = \alpha I(f) + \beta I(g)\,,$$

which is (7.12).

b) The word **isometry** means an equality under different metrics. Note that (7.13) can be written as $\|I_T(f)\|_{\mathscr{L}^2} = \|f\|_{\mathscr{M}_T^2}$. Hence, the Ito integral connects the two metrics $\mathscr{L}^2$ and $\mathscr{M}_T^2$.
Now we show $\|I_T(f)\|_{\mathscr{L}^2} = \|f\|_{\mathscr{M}_T^2}$. For any $f \in \mathscr{M}^2$, from the proof of Theorem 7.1, we can construct f_n using (7.5) such that $\|I_T(f_n) - I_T(f)\|_{\mathscr{L}^2} \to 0$ and $\|f_n - f\|_{\mathscr{M}_T^2} \to 0$. Combining with (7.10), the result follows.

c) Following the proof of Theorem 7.1, we can find a random step function f_n approximating f such that $\|I_T(f_n) - I_T(f)\|_{\mathscr{L}^2} \to 0$. W.L.O.G., we can assume $s = t_k$ for some k. (Adding an extra point still gives a step function.) Thus

$$\begin{aligned}
&\mathrm{E}(I_T(f_n)|\mathscr{F}_s) \\
&= \mathrm{E}\left(\sum_{j=1}^{n} f(t_{j-1})(W_{t_j} - W_{t_{j-1}}) \middle| \mathscr{F}_s \right) \\
&= \sum_{j=1}^{k} f(t_{j-1})(W_{t_j} - W_{t_{j-1}}) + \mathrm{E}\left(\sum_{j=k+1}^{n} f(t_{j-1})(W_{t_j} - W_{t_{j-1}}) \middle| \mathscr{F}_s \right) \\
&= \sum_{j=1}^{k} f(t_{j-1})(W_{t_j} - W_{t_{j-1}}) \qquad \text{(by independent increment and } t_k = s\text{)} \\
&= I_s(f_n)\,.
\end{aligned} \tag{7.15}$$

The same argument in the proof of Theorem 7.1 yields $\|I_s(f_n) - I_s(f)\|_{\mathscr{L}^2} \to 0$. On the other hand, we have $\|\mathrm{E}(I_T(f_n)|\mathscr{F}_s) - \mathrm{E}(I_T(f)|\mathscr{F}_s)\|_{\mathscr{L}^2} \to 0$, since

$$\begin{aligned}
&\mathrm{E}\left[(\mathrm{E}(I_T(f_n)|\mathscr{F}_s) - \mathrm{E}(I_T(f)|\mathscr{F}_s))^2\right] \\
=\ &\mathrm{E}\left[(\mathrm{E}(I_T(f_n) - I_T(f)|\mathscr{F}_s))^2\right] \\
\le\ &\mathrm{E}\left(\mathrm{E}((I_T(f_n) - I_T(f))^2|\mathscr{F}_s)\right) \qquad \text{(Jensen's Inequality)} \\
=\ &\mathrm{E}((I_T(f_n) - I_T(f))^2) \qquad \text{(Tower Property)} \\
\to\ &0\,.
\end{aligned}$$

By taking limit on both sides of Equation (7.15), we have

$$\mathrm{E}(I_T(f)|\mathscr{F}_s) = I_s(f)\,,$$

completing the proof.

7.4 Ito's Lemma

In this section we prove the Ito's Lemma which is of fundamental importance to mathematical finance and stochastic calculus.

7.4.1 The case $F(t,W_t)$

Theorem 7.3. (Ito's Lemma) *Suppose that $F(t,x)$ is a real valued function with continuous partial derivatives $F_t(t,x)$, $F_x(t,x)$ and $F_{xx}(t,x)$ for all $t \geqslant 0$ and $x \in \mathbb{R}$. Assume also that the process $F_x(t,W_t)$ belongs to $\mathscr{M}^2$. Then $F(t,W_t)$ satisfies*

$$\begin{aligned} &F(T,W_T) - F(0,W_0) \\ &= \int_0^T \left[F_t(t,W_t) + \frac{1}{2}F_{xx}(t,W_t)\right] dt + \int_0^T F_x(t,W_t)\, dW_t, \quad \text{a.s.} \end{aligned} \tag{7.16}$$

In differential notation, (7.16) *can be written as*

$$dF(t,W_t) = \left[F_t(t,W_t) + \frac{1}{2}F_{xx}(t,W_t)\right] dt + F_x(t,W_t)\, dW_t\,. \tag{7.17}$$

Remark 7.4

a) Comparing (7.17) with the usual chain rule

$$dF(t,x_t) = F_t(t,x_t)\, dt + F_x(t,x_t)\, dx_t$$

for a differentiable function x_t, the additional term $\frac{1}{2}F_{xx}(t,W_t)\,dt$ in (7.17) is called the Ito correction.

b) Equation (7.17) is often written in the abbreviated form

$$dF = \left(F_t + \frac{1}{2}F_{xx}\right) dt + F_x\, dW_t\,. \tag{7.18}$$

Strictly speaking, the differential notation (7.17) *does not make sense due to the non-differentiability of Brownian motion (dW_t is undefined).*

Proof. We first prove the case where F_t, F_x, F_{xx} are all bounded by some $C > 0$. Consider a partition $0 = t_0^n < t_1^n < \cdots < t_n^n = T$ of $[0,T]$, where $t_i^n = \frac{iT}{n}$. Denote $W_{t_i^n}$ by W_i^n; the increments $W_{i+1}^n - W_i^n$ by $\Delta_i^n W$; and $t_{i+1}^n - t_i^n$ by $\Delta_i^n t$.

Using Taylor's expansion, for each $i = 1\ldots,n$, there is a $\widetilde{W}_i^n$ between W_i^n and W_{i+1}^n, and a $\widetilde{t}_i^n \in [t_i^n, t_{i+1}^n]$ such that

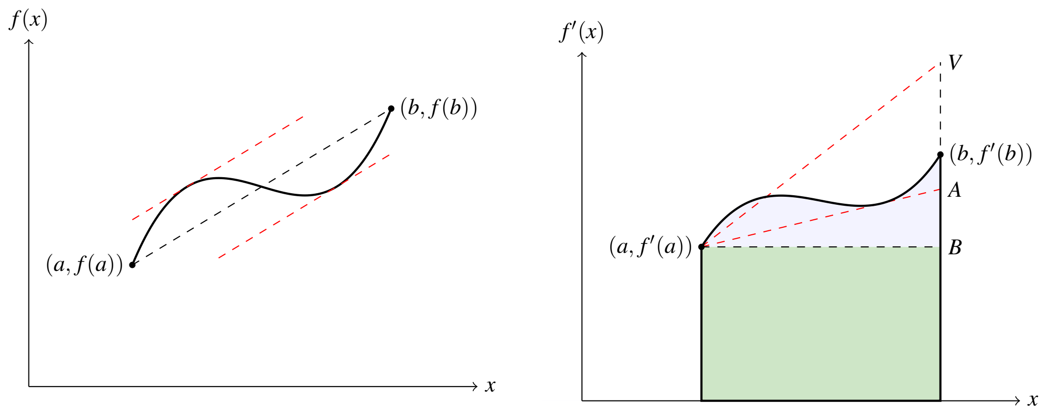


Fig. 7.1 Mean value theorem. Left: $f(b) = f(a) + (b-a)f'(x)$. Right: $f(b) = f(a) + f'(a)(b-a) + \frac{1}{2}f''(x)(b-a)^2$. Note that the green area is $f'(a)(b-a)$ and the purple area is $\frac{1}{2}f''(x)(b-a)^2$.

$$
\begin{aligned}
&F(T, W_T) - F(0, W_0) \\
&= \sum_{i=0}^{n-1} \left[F(t_{i+1}^n, W_{i+1}^n) - F(t_i^n, W_i^n) \right] \\
&= \sum_{i=0}^{n-1} \left[F(t_{i+1}^n, W_{i+1}^n) - F(t_i^n, W_{i+1}^n) \right] + \sum_{i=0}^{n-1} \left[F(t_i^n, W_{i+1}^n) - F(t_i^n, W_i^n) \right] \\
&= \sum_{i=0}^{n-1} F_t(\widetilde{t}_i^n, W_{i+1}^n)\Delta_i^n t + \sum_{i=0}^{n-1} F_x(t_i^n, W_i^n)\Delta_i^n W + \frac{1}{2}\sum_{i=0}^{n-1} F_{xx}(t_i^n, \widetilde{W}_i^n)\,(\Delta_i^n W)^2 \\
&= \sum_{i=0}^{n-1} F_t(\widetilde{t}_i^n, W_{i+1}^n)\Delta_i^n t + \frac{1}{2}\sum_{i=0}^{n-1} F_{xx}(t_i^n, W_i^n)\Delta_i^n t + \sum_{i=0}^{n-1} F_x(t_i^n, W_i^n)\Delta_i^n W \\
&\qquad + \frac{1}{2}\sum_{i=0}^{n-1} F_{xx}(t_i^n, W_i^n)\left[(\Delta_i^n W)^2 - \Delta_i^n t \right] + \frac{1}{2}\sum_{i=0}^{n-1} \left[F_{xx}(t_i^n, \widetilde{W}_i^n) - F_{xx}(t_i^n, W_i^n) \right](\Delta_i^n W)^2 \\
&= A_{1,n} + A_{2,n} + A_{3,n} + A_{4,n} + A_{5,n}\,, \qquad (7.19)
\end{aligned}
$$

say. Note that as F_t, F_x and F_{xx} are **continuous** and **bounded** functions, we have

$$\lim_{n\to\infty} \sup_{i=0,\ldots,n-1} \sup_{t\in[t_i^n, t_{i+1}^n]} |F_t(\widetilde{t}_i^n, W_{i+1}^n) - F_t(t, W_t)| \to 0 \quad a.s.\,, \tag{7.20}$$

$$\lim_{n\to\infty} \sup_{i=0,\ldots,n-1} \sup_{t\in[t_i^n, t_{i+1}^n]} |F_{xx}(t_i^n, W_i^n) - F_{xx}(t, W_t)| \to 0\,, \quad a.s., \tag{7.21}$$

$$\lim_{n\to\infty} \sup_{i=1,\ldots,n} |F_{xx}(t_i^n, \widetilde{W}_i^n) - F_{xx}(t_i^n, W_i^n)| \to 0 \quad a.s.\,. \tag{7.22}$$

Now we separately deal with each sum in (7.19):

i) From the continuity of F_t and F_{xx} in (7.20) and (7.21), we have the convergence of the Riemann integral

$$\lim_{n\to\infty} A_{1,n} = \lim_{n\to\infty} \sum_{i=0}^{n-1} F_t(\widetilde{t}_i^n, W_{i+1}^n)\Delta_i^n t = \int_0^T F_t(t,W_t)\,dt \quad \text{a.s.}, \quad \text{and}$$

$$\lim_{n\to\infty} A_{2,n} = \lim_{n\to\infty} \sum_{i=0}^{n-1} F_{xx}(t_i^n, W_i^n)\Delta_i^n t = \int_0^T F_{xx}(t,W_t)\,dt \quad \text{a.s.}$$

ii) From the assumption $F_x \in \mathscr{M}^2$, we have

$$\lim_{n\to\infty} A_{3,n} = \lim_{n\to\infty} \sum_{i=0}^{n-1} F_x(t_i^n, W_i^n)\Delta_i^n W = \int_0^T F_x(t,W_t)\,dW_t$$

in $\mathscr{L}^2$ from Theorem 7.1.

iii) We show the $\mathscr{L}^2$ convergence $\mathrm{E}(A_{4,n}^2) \to 0$. To be specific,

$$\begin{aligned}
\mathbb{E}(A_{4,n}^2) &= \mathbb{E}\left(\sum_{i=0}^{n-1} F_{xx}(t_i^n, W_i^n)\left[(\Delta_i^n W)^2 - \Delta_i^n t\right]\right)^2 \\
&= \sum_{i=0}^{n-1} \mathbb{E}\left(F_{xx}(t_i^n, W_i^n)\left[(\Delta_i^n W)^2 - \Delta_i^n t\right]\right)^2 \qquad \text{(cross terms have expectation 0)} \\
&= \sum_{i=0}^{n-1} \mathbb{E}\,[F_{xx}(t_i^n, W_i^n)]^2\,\mathbb{E}\left[(\Delta_i^n W)^2 - \Delta_i^n t\right]^2 \qquad \text{(independent increment)} \\
&\le C^2 \sum_{i=0}^{n-1} \mathbb{E}\left[(\Delta_i^n W)^2 - \Delta_i^n t\right]^2 \qquad \text{(boundedness of } F_{xx}) \\
&= 2C^2 \sum_{i=0}^{n-1} (\Delta_i^n t)^2 = 2C^2 \sum_{i=0}^{n-1} \frac{T^2}{n^2} = 2C^2 \frac{T^2}{n} \to 0 \quad \text{as}\, n \to \infty\,.
\end{aligned}$$

iv) Similar to the proof of iii), $\sum_{i=0}^{n-1} (\Delta_i^n W)^2 \to t$ in $\mathscr{L}^2$ and thus in probability (See Theorem 3.1). (Indeed, the left quantity is the quadratic variation of Brownian motion.) Together with the a.s. continuity result (7.21), we have the following convergence (in probability):

$$\begin{aligned}
|A_{5,n}| &= \left|\sum_{i=0}^{n-1}\left[F_{xx}(t_i^n, \widetilde{W}_i^n) - F_{xx}(t_i^n, W_i^n)\right](\Delta_i^n W)^2\right| \\
&\leqslant \sup_{i=0,1,\ldots,n-1}\left|F_{xx}(t_i^n, \widetilde{W}_{i+1}^n) - F_{xx}(t_i^n, W_{i+1}^n)\right| \sum_{i=0}^{n-1} (\Delta_i^n W)^2 \xrightarrow{P} 0\,.
\end{aligned}$$

Note that the convergences of $A_{i,n}$, $i = 1,\ldots,5$ involve different modes: $A_{1,n}$ and $A_{2,n}$ converge almost surely, $A_{3,n}$, $A_{4,n}$ converge in $\mathscr{L}^2$, and $A_{5,n}$ converges in probability. To combine the results, note from Theorem 3.1 that convergence in $\mathscr{L}^2$

implies convergence in probability. Thus, all $A_{3,n}$, $A_{4,n}$ and $A_{5,n}$ converge in probability. Note also from Exercise 7.15 that there is a subsequence $\{n_k\}_{k=1,2,\ldots}$ such that $\{A_{3,n_k}\}_{k=1,2,\ldots}$ converge a.s.. Along this subsequence, we can find a further subsequence n_{k_l} such that $A_{4,n_{k_l}}$ converges a.s., and so forth. Therefore, all $A_{j,n}$, $j = 1,\ldots,5$ converge a.s. with respect to some subsequence $m_1 < m_2 < \cdots$, say. Then

$$
\begin{aligned}
&F(T,W_T) - F(0,W_0) \\
&= \lim_{k\to\infty}\Bigg\{ \sum_{i=0}^{m_k-1} F_t(\widetilde{t}_i^{m_k}, W_{i+1}^{m_k})\Delta_i^{m_k} t + \frac{1}{2}\sum_{i=0}^{m_k-1} F_{xx}(t_i^{m_k}, W_i^{m_k})\Delta_i^{m_k} t + \sum_{i=0}^{m_k-1} F_x(t_i^{m_k}, W_i^{m_k})\Delta_i^{m_k} W \\
&\qquad + \frac{1}{2}\sum_{i=0}^{m_k-1} F_{xx}(t_i^{m_k}, W_i^{m_k})\left[\left(\Delta_i^{m_k} W\right)^2 - \Delta_i^{m_k} t\right] \\
&\qquad + \frac{1}{2}\sum_{i=0}^{m_k-1}\left[F_{xx}(t_i^{m_k}, \widetilde{W}_i^{m_k}) - F_{xx}(t_i^{m_k}, W_i^{m_k})\right]\left(\Delta_i^{m_k} W\right)^2 \Bigg\}. \\
&= \int_0^T \left[F_t(t,W_t) + \frac{1}{2}F_{xx}(t,W_t)\right] dt + \int_0^T F_x(t,W_t)\, dW_t, \quad \text{a.s.},
\end{aligned}
$$

completing the proof. The general case where F_t, F_x and F_{xx} are not bounded is left as an exercise (Exercise 7.8).

Example 7.3. For $F(t,x) = x^2$, we have $F_t(t,x) = 0$, $F_x(t,x) = 2x$ and $F_{xx}(t,x) = 2$. Ito formula gives $dW_t^2 = dt + 2W_t\, dW_t$, provided that $2W_t \in \mathscr{M}^2$ (which has been verified in Example 7.1). □

7.4.2 General Case

Note that Ito's Lemma aims at providing a first order approximation for $dF(t,W_t)$ using the terms dt and dW_t. In the Taylor's expansion of $F(t,W_t)$ up to the second order terms, the quantities $dt, dW_t, dtdW_t, (dt)^2, (dW_t)^2, \ldots$ are involved. Since the quadratic variation of W_t is $[W]_t = t$, it follows that the term $(dW_t)^2$ contributes an additional dt term. Other terms such as $dtdW_t, (dt)^2$ and $(dW_t)^3$ are smaller than dt, and thus can be omitted. It is convenient to remember the results using the following *Ito multiplication table*:

$\times$	dt	dW_t
dt	0	0
dW_t	0	dt

Table 7.1 Ito multiplication table.

Looking closely into the proof of Theorem 7.3, it can be seen that Ito's Lemma can be extended from $F(t,W_t)$ to $F(t,X_t)$ where X_t is an arbitrary process with quadratic variation $[X]_t$ satisfying $d[X]_t = g(t)dt$ for some $\sqrt{g(t)} \in \mathscr{M}^2$. For example, if X_t is an **Ito Process** given by

$$dX_t = a_t\,dt + b_t\,dW_t, \tag{7.23}$$

where a_t and $b_t \in \mathscr{M}^2$, then it can be checked (Exercise 7.16) that $d[X]_t = b_t^2 dt$. (Informally, we can use Table 7.4.2 to see that $d[X]_t = (dX_t)^2 = (a_t dt + b_t dW_t)^2 = a_t^2(dt)^2 + 2a_t b_t (dt)(dW_t) + b_t^2(dW_t)^2 = b_t^2 dt$). The following Theorem gives the general case of Ito's Lemma for $F(t,X_t)$, where X_t is an arbitrary process instead of a Brownian motion.

Theorem 7.5. (Ito formula, general case) *Let $X_t = a_t\,dt + b_t\,dW_t$ be an Ito process with quadratic variation $[X]_t$ satisfying $d[X]_t = b_t^2 dt$ where $b_t \in \mathscr{M}^2$. Suppose that $F(t,x)$, $F_t(t,x)$, $F_x(t,x)$ and $F_{xx}(t,x)$ are continuous for all $t \geqslant 0$ and $x \in \mathbb{R}$. Also, assume the process $b_t F_x(t,X_t) \in \mathscr{M}^2$. Then, $F(t,X_t)$ can be expressed as*

$$dF(t,X_t) = F_t(t,X_t)\,dt + F_x(t,X_t)dX_t + \frac{1}{2}F_{xx}(t,X_t)d[X]_t\,. \tag{7.24}$$

Example 7.4. If X_t is an Ito process satisfying (7.23), then (7.24) reduces to (writing $F(t,X_t)$ as F)

$$\begin{aligned} dF &= F_t\,dt + F_x\,dX_t + \frac{1}{2}F_{xx}\,d[X]_t \\ &= F_t\,dt + F_x(a_t\,dt + b_t\,dW_t) + \frac{1}{2}F_{xx}b_t^2 dt \\ &= \left(F_t + F_x a_t + \frac{1}{2}F_{xx}b_t^2\right)dt + F_x b_t\,dW_t\,. \end{aligned}$$

We end this section with the analog of the **product rule** $df(x)g(x) = f(x)dg(x) + g(x)df(x)$ in ordinary calculus. Again, there is an additional term involving dt contributed by the quadratic variation process.

Corollary 7.1. (Product Rule) *If X_t and Y_t are processes satisfying $dX_t = a_x(t)dt + b_x(t)dW_t$ and $dY_t = a_y(t)dt + b_y(t)dW_t$, then*

$$dX_tY_t = X_t dY_t + Y_t dX_t + b_x(t)b_y(t)\,dt\,.$$

Proof. Intuitively, for a small Δt,

$$\begin{aligned} &X_{t+\Delta t}Y_{t+\Delta t} - X_tY_t \\ &= X_t(Y_{t+\Delta t} - Y_t) + Y_t(X_{t+\Delta t} - X_t) + (Y_{t+\Delta t} - Y_t)(X_{t+\Delta t} - X_t) \end{aligned}$$

As $\Delta \to 0$, by definition, $X_{t+\Delta t}Y_{t+\Delta t} - X_tY_t \to dX_tY_t$, $X_{t+\Delta t} - X_t \to dX_t$ and $Y_{t+\Delta t} - Y_t \to dY_t$. Thus Corollary 7.1 follows by noting from Table 7.4.2 that

$$(Y_{t+\Delta t} - Y_t)(X_{t+\Delta t} - X_t) \to (a_y(t)dt + b_y(t)dW_t)(a_x(t)dt + b_x(t)dW_t) = b_x(t)b_y(t)dt\,.$$

A formal proof can be obtained by similar argument as in Theorem 7.1. □

Example 7.5. Let $dX_t = \mu X_t dt + \sigma X_t dW_t$. Using Ito's Lemma,

$$dX_t^3 = 3X_t^2 dX_t + 3X_t (dX_t)^2\,. \tag{7.25}$$

Writing $Y_t = X_t^2$,

$$\begin{aligned} dX_t^3 &= d(X_t Y_t) = X_t^2 dX_t + X_t(dX_t^2) + dX_t(dX_t^2) \\ &= X_t^2 dX_t + X_t(2X_t dX_t + (dX_t)^2) + dX_t(2X_t dX_t + (dX_t)^2) \\ &= 3X_t^2 dX_t + 3X_t(dX_t)^2 + (dX_t)^3\,, \end{aligned}$$

which agrees with (7.25) since $(dX_t)^3$ is negligible. □

7.5 Stochastic Differential Equations

In this section, we consider *stochastic differential equation* of the form

$$dX_t = f(X_t)\,dt + g(X_t)\,dW_t\,,$$

with *initial condition* $X_0 = x_0$. The goal is to obtain an explicit formula for X_t.

Example 7.6. Consider the initial value problem

$$\begin{cases} dX_t = -\alpha X_t\,dt + \sigma\,dW_t\,, \\ X_0 = x_0 \in \mathbb{R}\,. \end{cases}$$

Let $F(t,x) = e^{\alpha t}x$, and thus $F(0,X_0) = x_0$. By Ito's Lemma, we have

$$\begin{aligned} dF(t,X_t) &= \left[\alpha e^{\alpha t} X_t - \alpha e^{\alpha t} X_t\right] dt + \sigma e^{\alpha t}\,dW_t \\ &= \sigma e^{\alpha t}\,dW_t\,. \end{aligned}$$

It is easily seen that $e^{\alpha t} \in \mathscr{M}^2$. It follows that $F(t,X_t) = x_0 + \sigma \int_0^t e^{\alpha s}\,dW_s$ is well defined. Thus,

$$X_t = e^{-\alpha t} F(t,X_t) = e^{-\alpha t} x_0 + \sigma e^{-\alpha t} \int_0^t e^{\alpha s}\,dW_s\,.$$

□

Example 7.7. Suppose $X_t = x_0 e^{at + bW_t}$ for $t \geqslant 0$. By Ito's Lemma,

$$\begin{aligned} dX_t &= d\left(x_0 e^{at+bW_t}\right) \\ &= \left(a x_0 e^{at+bW_t} + \frac{b^2}{2} x_0 e^{at+bW_t}\right) dt + b x_0 e^{at+bW_t}\, dW_t \\ &= \left(a + \frac{b^2}{2}\right) X_t\, dt + b X_t\, dW_t. \end{aligned}$$

The problem of checking that $bX_t \in \mathscr{M}^2$ for all $T > 0$ is left to the reader in Exercise 7.17. This implies the solution of the initial value problem

$$\begin{cases} dX_t = \left(a + \frac{b^2}{2}\right) X_t\, dt + b X_t\, dW_t, \\ X_0 = x_0 \end{cases}$$

is $X_t = x_0 e^{at+bW_t}$. □

From the above examples, to solve a SDE, one may first guess a solution and then use Ito's Lemma to verify that the solution satisfies the SDE. The following table summarizes some examples of SDE where closed form solutions are available. In the table, c stands for a constant and $\sinh(x) = (e^x - e^{-x})/2$.

Name	SDE	Solution (X_t)
Ornstein-Uhlenbeck(OU) process	$dX_t = -\alpha X_t dt + \sigma dW_t$	$ce^{-\alpha t} + \sigma e^{-\alpha t}\int_0^t e^{\alpha s} dW_s$
Mean reverting OU	$dX_t = (m - \alpha X_t)dt + \sigma dW_t$	$\frac{m}{\alpha} + \left(c - \frac{m}{\alpha}\right) e^{-\alpha t} + \sigma \int_0^t e^{\alpha(s-t)} dW_s$
Geometric Brownian motion	$dX_t = aX_t dt + bX_t dW_t$	$ce^{(a-b^2/2)t+bW_t}$
Brownian bridge	$dX_t = \frac{b-X_t}{1-t} dt + dW_t$	$a(1-t) + bt + (1-t)\int_0^t \frac{dW_s}{1-s}$
	$dX_t = \left(\sqrt{1+X_t^2} + \frac{1}{2}X_t\right) dt + \sqrt{1+X_t^2} dW_t$	$\sinh(c+t+W_t)$
	$dX_t = X_t^3 dt + X_t^2 dW_t$	$\frac{1}{c-W_t}$
	$dX_t = -\frac{1}{2}X_t dt + \sqrt{1-X_t^2} dW_t$	$\sin(c+W_t)$
	$dX_t = -\frac{1}{1+t}X_t dt + \frac{1}{1+t} dW_t$	$(c+W_t)/(1+t)$
	$dX_t = r dt + \alpha X_t dW_t$	$ce^{\alpha W_t - \frac{1}{2}\alpha^2 t} + r\int_0^t e^{\alpha(W_t - W_s) - \frac{1}{2}\alpha^2(t-s)} ds$

7.6 Three Technical Results

In this section, we discuss three technical results that constitute the proof of the **Fundamental Theorem of Asset Pricing**:

- **Levy's Characteristic of Brownian Motion**,
- **Girsanov Theorem**,
- **Brownian Martingale Representation Theorem**.

As the proofs do not help the understanding of the Fundamental Theorem of Asset Pricing, one may skip the proofs in the first reading.

7.6.1 Levy's Characteristic of Brownian Motion

Levy's characterization allows one to verify whether a process is a Brownian motion by investigating its quadratic variation.

Theorem 7.6. (Levy's Characteristic of Brownian Motion) *Let $\{W_t\}$ be a stochastic process with natural filtration $\{\mathscr{F}_t\}_{t\geq 0}$. Then, W_t is a Brownian motion if and only if the following conditions hold:*

1) $W_0 = 0$ a.s.;
2) $\{W_t\}$ is a martingale with respect to the filtration $\{\mathscr{F}_t\}_{t\geq 0}$.
3) The quadratic variation $[W]_t = t$ for all $t \geq 0$.

Proof. Since $W_0 = 0$ a.s. is assumed, it suffices to check the stationary and independent increment properties of Brownian motions in Definition 6.2. For a constant $\theta > 0$, consider the stochastic process

$$M_t = e^{\theta W_t - \frac{\theta^2}{2} t}. \tag{7.26}$$

Applying Ito's Lemma on M_t, together with the assumption $[W]_t = t$, we have

$$\begin{aligned} dM_t &= -\frac{\theta^2}{2} M_t dt + \theta M_t dW_t + \frac{\theta^2}{2} M_t dt \\ &= \theta M_t dW_t , \end{aligned}$$

or $M_t = \int_0^t \theta M_s dW_s$. Since W_t is a martingale, the proof of Theorem 7.2(c) implies that M_t is a martingale, i.e., for $s \leq t$,

$$\mathrm{E}(M_t | \mathscr{F}_s) = M_s . \tag{7.27}$$

Substituting (7.26) to (7.27), we have

$$\mathrm{E}\left(e^{\theta(W_t - W_s)} \middle| \mathscr{F}_s \right) = e^{\frac{1}{2}(t-s)\theta^2} . \tag{7.28}$$

Since the RHS is the m.g.f. of the normal distribution $N(0, t-s)$, we have $W_t - W_s \sim N(0, t-s)$, proving the stationary increment property.

To show independent increment, note that for $0 = t_0 < t_1 < \cdots < t_n = T$, we have from (7.28) and repeated applications of the Tower property that, for constants θ_is,

$$\begin{aligned} \mathrm{E}\left(e^{\sum_{i=1}^n \theta_i (W_{t_i} - W_{t_{i-1}})} \right) &= \mathrm{E}\left[\mathrm{E}\left(e^{\sum_{i=1}^n \theta_i (W_{t_i} - W_{t_{i-1}})} | \mathscr{F}_{t_{n-1}} \right) \right] \\ &= \mathrm{E}\left[e^{\sum_{i=1}^{n-1} \theta_i (W_{t_i} - W_{t_{i-1}})} \mathrm{E}\left(e^{\theta_n (W_{t_n} - W_{t_{n-1}})} \middle| \mathscr{F}_{t_{n-1}} \right) \right] \\ &= e^{\frac{1}{2}(t_n - t_{n-1})\theta_n^2} \mathrm{E}\left(e^{\sum_{i=1}^{n-1} \theta_i (W_{t_i} - W_{t_{i-1}})} \right) \quad \text{(by(7.28))} \\ &= \prod_{i=1}^n e^{\frac{1}{2}(t_i - t_{i-1})\theta_i^2} . \qquad \text{(Repeating the same argument)} \end{aligned}$$

Since the joint m.g.f. can be factorized into product of normal m.g.f.s, the sequence $\{W_{t_j} - W_{t_{j-1}}\}_{j=1,\dots,n}$ are independent normal random variables. Thus, W_t satisfies the three conditions of Brownian motion in Definition 6.2.

Example 7.8. We have verified in Example 6.1 that $\widetilde{W}_t = c^{-1/2} W_{ct}$ is a Brownian motion. Alternatively, using Theorem 7.6, the verification can be achieved by

1) showing the martingale property;
2) showing that $[\widetilde{W}]_t = t$.

First, the martingale property is inherited from the stationary and independent increment properties of Brownian motion W_t. Next, the quadratic variation can be computed using the same argument in Theorem 6.12: Set $Q_n = c^{-1}\sum_{i=1}^{n}(W_{ct_i} - W_{ct_{i-1}})^2$ for partitions $\Pi_n = \{t_i\}_{i=1,\dots,n}$ of $[0,t]$ with mesh $\delta(\Pi_n) \to 0$. Then show that $\mathrm{E}(Q_n) \to t$ and $\mathrm{Var}(Q_n) \to 0$. □

7.6.2 Girsanov Theorem

Recall that in the discrete time world, the physical probability measure is not useful for pricing. Instead, we use the asset values to deduce a risk neutral probability measure for pricing. In the continuous time world, it is no longer possible to find the risk neutral probability by linear algebra. We need the Girsanov Theorem to change the physical measure to the risk neutral probability measure for pricing.

Theorem 7.7. (Girsanov Theorem) *Suppose that $\{W_t\}_{t\ge 0}$ is a $\mathbb{P}$-Brownian motion with natural filtration $\{\mathscr{F}_t\}_{t\ge 0}$, and $\{\theta_t\}_{t\ge 0}$ is an $\{\mathscr{F}_t\}$-adapted process with*

$$E\left(e^{\frac{1}{2}\int_0^T \theta_t^2\,dt}\right) < \infty .$$

Then, for any "drifted" Brownian motion

$$W_t^{\mathbb{Q}} = W_t + \int_0^t \theta_s\,ds\,,$$

there exists a measure $\mathbb{Q}$ such that $W_t^{\mathbb{Q}}$ is a standard Brownian motion under $\mathbb{Q}$. The measure $\mathbb{Q}$ is given by

$$\mathbb{Q}(A) = \int_A L_t(\omega)\,\mathbb{P}(d\omega)\,, \tag{7.29}$$

for all $A \in \mathscr{F}_t$ and

$$\left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_t \triangleq L_t = e^{-\int_0^t \theta_s\,dW_s - \frac{1}{2}\int_0^t \theta_s^2\,ds} \tag{7.30}$$

is the Radon-Nikodym Derivative of $\mathbb{Q}$ w.r.t. $\mathbb{P}$, with both $\mathbb{P}$ and $\mathbb{Q}$ are restricted on $(\Omega, \mathscr{F}_t)$. Particularly, for any $\mathscr{F}_t$ measurable random variable X_t,

$$
\begin{aligned}
E_{\mathbb{P}}(X_t) &= E_{\mathbb{Q}}\left(X_t \frac{1}{L_t}\right) = E_{\mathbb{Q}}\left(X_t e^{\int_0^t \theta_s\, dW_s + \frac{1}{2}\int_0^t \theta_s^2\, ds}\right) \\
&= E_{\mathbb{Q}}\left(X_t e^{\int_0^t \theta_s\, dW_s^{\mathbb{Q}} - \frac{1}{2}\int_0^t \theta_s^2\, ds}\right). \qquad (7.31)
\end{aligned}
$$

Proof. **(Heuristic)** First we give a heuristic derivation to understand the Girsanov Theorem. Let X be a r.v. on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and $\mathbb{P}(X < x) = \Phi(x)$, i.e., $X \sim N(0, \sigma^2)$ under $\mathbb{P}$. In this case, the p.d.f. of X is given by $f_{\mathbb{P}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}$. Obviously, the random variable $X + \mu$ follows $N(\mu, \sigma^2)$ under $\mathbb{P}$. If we want a measure $\mathbb{Q}$ such that $X + \mu$ follows $N(0, \sigma^2)$ under $\mathbb{Q}$, then $X \sim N(-\mu, \sigma^2)$, and the p.d.f. of X under $\mathbb{Q}$ is

$$
f_{\mathbb{Q}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x+\mu)^2}{2\sigma^2}} = e^{-\frac{1}{\sigma^2}\left(x\mu + \frac{1}{2}\mu^2\right)} f_{\mathbb{P}}(x).
$$

In particular, the Radon-Nikodym Derivative is given by

$$
\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{f_{\mathbb{Q}}(X)}{f_{\mathbb{P}}(X)} = e^{-\frac{1}{\sigma^2}\left(X\mu + \frac{1}{2}\mu^2\right)}. \qquad (7.32)
$$

When $\mu = \sigma^2 = t$, (7.32) is exactly (7.30) with $\theta_s \equiv 1$.

Consider the general case where W_t is a Brownian motion under $\mathbb{P}$. Let $t_i = i/n$ and $\mathbb{P}^{(n)}$ be the finite dimensional probability measure for $\{W_{t_i}\}_{i=1,\ldots,n}$. Suppose that we want a measure $\mathbb{Q}^{(n)}$ such that the drifted Brownian motion $\widetilde{W}_t = W_t + \int_0^t \theta_s\, ds$ behaves the same as a standard Brownian motion at $\{t_i\}_{i=1,\ldots,n}$.

Using transition probabilities, the joint p.d.f of the B.M. at the time points $\{t_i\}_{i=1,\ldots,n}$ is the same as the joint p.d.f. of $\{W_{t_i} - W_{t_{i-1}}\}_{i=1,2,\ldots,n}$, i.e.,

$$
f_{\mathbb{P}^{(n)}}(x_1, \cdots, x_n) = \prod_{j=1}^{n} \phi\left(x_j - x_{j-1}, 0, \frac{1}{n}\right), \qquad (7.33)
$$

where $x_0 = 0$ and $\phi(\cdot, \mu, \sigma^2)$ is the p.d.f. of a $N(\mu, \sigma^2)$ distribution. On the other hand, since

$$
\widetilde{W}_{t_i} - \widetilde{W}_{t_{i-1}} = W_{t_i} - W_{t_{i-1}} + \int_{t_{i-1}}^{t_i} \theta_s\, ds \approx W_{t_i} - W_{t_{i-1}} + \frac{1}{n}\theta_{t_{i-1}},
$$

the joint p.d.f. of $\{\widetilde{W}_{t_i} - \widetilde{W}_{t_{i-1}}\}_{i=1,2,\ldots,n}$ under $\mathbb{P}^{(n)}$ is the product of p.d.f.s of normal random variable plus a constant (similar to $X + \mu$).

Thus, to find a measure $\mathbb{Q}^{(n)}$ such that $\{\widetilde{W}_{t_i} - \widetilde{W}_{t_{i-1}}\}_{i=1,\ldots,n}$ has a joint p.d.f. equal to the RHS of (7.33) under $\mathbb{Q}^{(n)}$, the previous argument (7.32) can be employed with $X = W_{t_j} - W_{t_{j-1}}$, $\mu = \frac{\theta_{t_{j-1}}}{n}$ and $\sigma^2 = \frac{1}{n}$. This yields the Radon-Nikodym Derivative (RND) of the $\mathbb{Q}^{(n)}$ w.r.t. $\mathbb{P}^{(n)}$,

$$\begin{aligned}
\frac{d\mathbb{Q}^{(n)}}{d\mathbb{P}^{(n)}} &= \prod_{j=1}^{n} e^{\left(-(W_{t_j}-W_{t_{j-1}})\theta_{t_{j-1}}-\frac{1}{2n}\theta_{t_{j-1}}^2\right)} \\
&= e^{\left(-\sum_{j=1}^{n}(W_{t_j}-W_{t_{j-1}})\theta_{t_{j-1}}-\sum_{j=1}^{n}\frac{1}{2n}\theta_{t_{j-1}}^2\right)} \\
&\approx e^{-\int_0^1 \theta_s dW_s-\frac{1}{2}\int_0^1 \theta_s^2\,ds},
\end{aligned}$$

which motivates the form of the RND in (7.30).

Proof. **(Formal Proof)** First we need to verify that the measure $\mathbb{Q}$ on $\{\Omega,\mathscr{F}_t\}$ defined by (7.29) is indeed a probability measure, i.e. (Definition 2.6)

i) $\mathbb{Q}(A)\geq 0$ for $A\in\mathscr{F}_t$
ii) $\mathbb{Q}(\emptyset)=0$
iii) (**Countable additivity**) If $A_i\in\mathscr{F}_t$ $(i=1,2,\cdots)$ are **disjoint**, then $\mathbb{Q}\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}\mathbb{Q}(A_i)$.
iv) $\mathbb{Q}(\Omega)=1$

Since the $\mathbb{Q}(A)=\int_A L_t\,\mathbb{P}$ is defined as an integral over a set $A\in\mathscr{F}_t$ and $L_t\geq 0$ by definition, i) to iii) hold, and it remains to show iv).

Note that L_t is a $(\mathbb{P},\{\mathscr{F}_t\}_{t\geq 0})$ martingale. To see this, let

$$L_t=e^{Z_t}\,,\quad \text{where}\quad Z_t=-\int_0^t \theta_s\,dW_s-\frac{1}{2}\int_0^t \theta_s^2\,ds\,.$$

Using Ito's Lemma (7.24) on L_t, we have

$$\begin{aligned}
dL_t &= \frac{\partial L_t}{\partial t}dt+\frac{\partial L_t}{\partial Z_t}dZ_t+\frac{1}{2}\frac{\partial^2 L_t}{\partial Z_t^2}d[Z]_t \\
&= 0+L_t dZ_t+\frac{1}{2}L_t\theta_t^2 dt \\
&= -\theta_t L_t dW_t\,, \qquad \left(\text{since } dZ_t=-\theta_t dW_t-\frac{1}{2}\theta_t^2 dt\right), \qquad (7.34)
\end{aligned}$$

which implies that L_t is a $\mathbb{P}$-martingale by Theorem 7.2 (c). The martingale property implies that

$$\mathbb{Q}(\Omega)\equiv\int_\Omega L_t\,d\mathbb{P}=\mathrm{E}_{\mathbb{P}}(L_t)=L_0=1\,,$$

i.e., $\mathbb{Q}$ is a *probability* measure. In other words, $\left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_t\equiv L_t$ is the Radon Nikodym Derivative (R.N.D.) of $\mathbb{Q}$ w.r.t. $\mathbb{P}$ on $(\Omega,\mathscr{F}_t)$.

Next, to show that $\{W_t^{\mathbb{Q}}\}$ is a Brownian Motion under $\mathbb{Q}$, we use Levy's Characteristic Theorem 7.6: We need to show that

1) $\{W_t^{\mathbb{Q}}\}$ is a $\mathbb{Q}$-martingale;
2) the quadratic variation process of $\{W_t^{\mathbb{Q}}\}$ is $[W^{\mathbb{Q}}]_t=t$.

1) The idea is to show that $W_t^{\mathbb{Q}} L_t$ is a $\mathbb{P}$-martingale, and then use the property of R.N.D. to show that $W_t^{\mathbb{Q}}$ is a $\mathbb{Q}$-martingale. From $W_t^{\mathbb{Q}} = W_t + \int_0^t \theta_s ds$, we have $dW_t^{\mathbb{Q}} = dW_t + \theta_t dt$. Also, by Corollary 7.1 (product rule), we have

$$\begin{aligned} d\left(W_t^{\mathbb{Q}} L_t\right) &= W_t^{\mathbb{Q}} dL_t + L_t dW_t^{\mathbb{Q}} + dW_t^{\mathbb{Q}} dL_t \\ &= W_t^{\mathbb{Q}}(-\theta_t L_t dW_t) + L_t(\theta_t dt + dW_t) + (\theta_t dt + dW_t)(-\theta_t L_t dW_t) \\ &= (L_t - \theta_t L_t W_t^{\mathbb{Q}}) dW_t \,. \end{aligned}$$

Thus, from Theorem 7.2(3),

$$W_t^{\mathbb{Q}} L_t \ \text{ is a } \mathbb{P} \text{ martingale.} \tag{7.35}$$

It follows that $W_t^{\mathbb{Q}}$ is a $\mathbb{Q}$-martingale. To see this, note from (7.35) that $\mathrm{E}_{\mathbb{P}}\left(W_t^{\mathbb{Q}} L_t | \mathscr{F}_s\right) = W_s^{\mathbb{Q}} L_s$, which is the same as

$$\int_S W_t^{\mathbb{Q}} L_t d\mathbb{P} = \int_S \mathrm{E}_{\mathbb{P}}(W_t^{\mathbb{Q}} L_t | \mathscr{F}_s) d\mathbb{P} = \int_S W_s^{\mathbb{Q}} L_s d\mathbb{P} \,, \tag{7.36}$$

for all $S \in \mathscr{F}_s$ (definition of conditional expectation). Since $L_t = \frac{d\mathbb{Q}}{d\mathbb{P}}\Big|_t$ is the R.N.D. of $\mathbb{Q}$ w.r.t. $\mathbb{P}$ on $(\Omega, \mathscr{F}_t)$, Equation (7.36) can be expressed as

$$\int_S W_t^{\mathbb{Q}} d\mathbb{Q} = \int_S W_s^{\mathbb{Q}} d\mathbb{Q} \,,$$

for all $S \in \mathscr{F}_s \subset \mathscr{F}_t$, which means $\mathrm{E}_{\mathbb{Q}}(W_t^{\mathbb{Q}} | \mathscr{F}_s) = W_s^{\mathbb{Q}}$, i.e., $W_t^{\mathbb{Q}}$ is a $\mathbb{Q}$-martingale.

2) Since $W_t^{\mathbb{Q}}$ is a drifted version of W_t, similar arguments as in the proof of Theorem 6.12 show that the quadratic variation of $W_t^{\mathbb{Q}}$ under $\mathbb{Q}$ is t. (Exercise 7.16). $\square$

Example 7.9. Consider a $(\mathbb{P}, \{\mathscr{F}_t\}_{t \geq 0})$ Brownian motion $\{W_t\}$. Recall from Lemma 6.9 the joint probability density of $M_t = \max_{t>u>0} W_u$ and W_t, say $f_{M,W}(m,w)$. Suppose we want to find

$$\mathrm{E}_{\mathbb{P}}(S_t 1_{\{\max_{t>u>0} S_u > K\}}) \,, \quad \text{where } S_t = e^{at+W_t} \,. \tag{7.37}$$

The difficulty here is that the set $\{\max_{t>u>0} S_u > K\} = \{\max_{t>u>0} e^{au+W_u} > K\}$ is not directly related to $M_t = \max_{t>u>0} W_u$. To tackle this, using Girsanov Theorem (7.31) with a measure $\mathbb{Q}$ such that $W_t^{\mathbb{Q}} = at + W_t$ is a $\mathbb{Q}$-B.M., (7.37) reduces to

$$\mathrm{E}_{\mathbb{P}}\left(e^{at+W_t} 1_{\left\{\max_{t>u>0} e^{au+W_u} > K\right\}}\right) = \mathrm{E}_{\mathbb{Q}}\left(e^{W_t^{\mathbb{Q}}} 1_{\left\{\max_{t>u>0} W_u^{\mathbb{Q}} > \log K\right\}} e^{aW_t^{\mathbb{Q}} - \frac{1}{2}a^2 t}\right) \tag{7.38}$$

Now, in (7.38), only $W_t^{\mathbb{Q}}$ and $\max_{t>u>0} W_u^{\mathbb{Q}}$ are involved, and $W_t^{\mathbb{Q}}$ is a standard B.M. under $\mathbb{Q}$. Thus $f_{M,W}(m,w)$ can be employed to compute the RHS of (7.38) by

$$\iint_{m>w} 1_{\{m>\log K\}} e^{(1+a)w-\frac{1}{2}a^2 t} f_{M,W}(m,w)\, dm\, dw\,.$$

Example 7.10. **(Hitting a slopping line)** In Lemma 6.11 we found the m.g.f of the first hitting time T_a of a Brownian motion hitting a level a. We now employ the Girsanov Theorem to find the m.g.f of the first hitting time $T_{a,b} := \inf\{t \geqslant 0 : W_t = a+bt\}$ of a Brownian motion hitting a slopping line ($a \geqslant 0, b \in \mathbb{R}$).

Note that "a standard Brownian motion W_t hitting a sloping line $a+bt$" is equivalent to "a drifted Brownian motion $W_t - bt$ hitting level a". Therefore, to find the stopping time $T_{a,b} := \inf\{t \geqslant 0 : W_t = a+bt\}$, we may use Girsanov Theorem with $W_t^{\mathbb{Q}} = W_t - bt$ so that $T_{a,b}$ becomes $\inf\{t \geqslant 0 : W_t^{\mathbb{Q}} = a\}$. Then, the result in Lemma 6.11 about stopping time of hitting a horizontal level, i.e., $\mathbb{E}\exp\{-\theta T_a\} = \exp\left\{-a\sqrt{2\theta}\right\}$, can be applied.

To use Girsanov Theorem (7.31), the random variable in the expectation must be $\mathscr{F}_t$ measurable for some fixed $t>0$. However, the random stopping time $T_{a,b}$ is unbounded and is not $\mathscr{F}_t$ measurable for any $t>0$. To solve this problem, we introduce the indicator $1_{\{T_{a,b}\leq t\}}$ so that $1_{\{T_{a,b}\leq t\}} e^{-\theta T_{a,b}}$ is $\mathscr{F}_t$ measurable:

$$\mathrm{E}\left(e^{-\theta T_{a,b}}\right) = \mathrm{E}\left(1_{\{T_{a,b}\leq t\}} e^{-\theta T_{a,b}}\right) + \mathrm{E}\left(1_{\{T_{a,b}>t\}} e^{-\theta T_{a,b}}\right)\,. \tag{7.39}$$

The first term in (7.39) can be handled using Girsanov Theorem:

$$\begin{aligned}
\mathrm{E}\left(1_{\{T_{a,b}\leq t\}} e^{-\theta T_{a,b}}\right) &= \mathrm{E}\left(1_{\{T_{a,b}\leq t\}} e^{-\theta \inf\{t\geqslant 0:\, W_t = a+bt\}}\right) \\
&= \mathrm{E}_{\mathbb{Q}}\left(1_{\{T_{a,b}\leq t\}} e^{-\theta \inf\{t\geqslant 0:\, W_t^{\mathbb{Q}} = a\}} e^{-bW_t^{\mathbb{Q}} - \frac{b^2 t}{2}}\right) \qquad (\text{Girsanov:} W_t^{\mathbb{Q}} = W_t - bt \text{ is a } \widetilde{\mathbb{Q}}\ B.M.) \\
&= \mathrm{E}_{\mathbb{Q}}\left(1_{\{\tau\leq t\}} e^{-\theta\tau} e^{-b(W_t^{\mathbb{Q}} - W_\tau^{\mathbb{Q}}) - W_\tau^{\mathbb{Q}} b - \frac{b^2 t}{2}}\right) \qquad (\tau \triangleq T_{a,b} \triangleq \inf\{t\geqslant 0 : W_t^{\mathbb{Q}} = a\}) \\
&= \mathrm{E}_{\mathbb{Q}}\left(1_{\{\tau\leq t\}} e^{-\theta\tau} e^{-b(W_t^{\mathbb{Q}} - W_\tau^{\mathbb{Q}}) - ab - \frac{b^2 t}{2}}\right) \qquad (W_\tau^{\mathbb{Q}} = a \quad \text{by definition}) \\
&= e^{-\frac{b^2 t}{2} - ab}\mathrm{E}_{\mathbb{Q}}\left(1_{\{\tau\leq t\}} e^{-\theta\tau} \mathrm{E}_{\mathbb{Q}}\left(e^{-b(W_t^{\mathbb{Q}} - W_\tau^{\mathbb{Q}})} \,\middle|\, \tau\right)\right) \qquad (\text{Tower property}) \\
&= e^{-\frac{b^2 t}{2} - ab}\mathrm{E}_{\mathbb{Q}}\left(1_{\{\tau\leq t\}} e^{-\theta\tau}\left(e^{\frac{b^2}{2}(t-\tau)}\right)\right) \qquad (\text{M.G.F. of Brownian motion}) \\
&= e^{-ab}\mathrm{E}_{\mathbb{Q}}\left(1_{\{\tau\leq t\}} e^{-(\theta+\frac{b^2}{2})\tau}\right) \\
&= e^{-ab}\mathrm{E}_{\mathbb{Q}}\left(e^{-(\theta+\frac{b^2}{2})\tau}\right) - e^{-ab}\mathrm{E}_{\mathbb{Q}}\left(1_{\{\tau>t\}} e^{-(\theta+\frac{b^2}{2})\tau}\right) \\
&= e^{-ab} e^{-a\sqrt{b^2+2\theta}} - e^{-ab}\mathrm{E}_{\mathbb{Q}}\left(1_{\{\tau>t\}} e^{-(\theta+\frac{b^2}{2})\tau}\right) \qquad (\text{Lemma 6.11}) \\
&= e^{-a(b+\sqrt{b^2+2\theta})} - e^{-ab}\mathrm{E}_{\mathbb{Q}}\left(1_{\{\tau>t\}} e^{-(\theta+\frac{b^2}{2})\tau}\right)\,.
\end{aligned} \tag{7.40}$$

Since t is arbitrary, if $\theta > 0$, then $1_{\{T_{a,b}>t\}}e^{-\theta T_{a,b}} \leqslant e^{-\theta t}$ and $1_{\{\tau>t\}}e^{-(\theta+\frac{b^2}{2})\tau} \leqslant e^{-(\theta+\frac{b^2}{2})t}$ converge to zero as $t \to \infty$. Hence, the DCT guarantees that the second terms of both (7.39) and (7.40) are negligible. In summary, we have, for $\theta > 0$ and $a > 0$,

$$\mathrm{E}\left(e^{-\theta T_{a,b}}\right) = e^{-a(b+\sqrt{b^2+2\theta})}. \tag{7.41}$$

$\square$

7.6.3 Brownian Martingale Representation Theorem

Finally, the Brownian Martingale Representation Theorem shows that any martingale can be represented by a stochastic integral involving a Brownian motion.

Theorem 7.8. (Brownian Martingale Representation Theorem) *Let $\{\mathscr{F}_t\}_{t\geq 0}$ denote the natural filtration of the $\mathbb{P}$-Brownian motion $\{W_t\}_{t\geq 0}$. Let $\{M_t\}_{t\geq 0}$ be a square-integrable $(\mathbb{P}, \{\mathscr{F}_t\}_{t\geq 0})$ martingale, i.e., for each $t > 0$,*

$$E(|M_t|^2) < \infty.$$

Then, there exists an $\{\mathscr{F}_t\}_{t\geq 0}$ predictable process $\{\theta_t\}_{t\geq 0}$ such that with $\mathbb{P}$-probability one,

$$M_t = M_0 + \int_0^t \theta_s \, dW_s. \tag{7.42}$$

Proof. First, we create a class of stochastic processes $\mathscr{G}$ of the form

$$Z_t^f = e^{\int_0^t f(s)\,dW_s - \frac{1}{2}\int_0^t f(s)^2\,ds}.$$

where $f(s) = \sum_{i=1}^n \beta_i 1_{(t_{i-1},t_i]}(s)$, $\beta_i \in \mathbb{R}$ and $\{t_i\}_{i=0,1,\ldots,n}$ is a partition of $[0,T]$. Using Ito's formula, it can be checked that (Exercise 7.12)

$$Z_t^f = 1 + \int_0^t f(s) Z_s^f \, dW_s. \tag{7.43}$$

Using theory in Hilbert space, it can be shown that any random variable in $\mathscr{M}^2$ can be represented by a linear combination of elements in $\mathscr{G}$ (for details, see Chapter 12 of Steele (2001)). Specifically, any martingale M_t has a representation $M_t = \sum_{j=1}^\infty \alpha_j Z_t^{f_j}$ for some constant α_i. Hence, from (7.43) we have

$$
\begin{aligned}
M_t &= \sum_{j=1}^{\infty} \alpha_j + \int_0^t \sum_{j=1}^{\infty} \alpha_j f_j(s) Z_s^{f_j} \, dW_s \\
&= \theta_0 + \int_0^t \theta_s \, dW_s \,,
\end{aligned}
$$

where $\theta_0 = \sum_{j=1}^{\infty} \alpha_j$ and $\theta_s = \sum_{j=1}^{\infty} \alpha_j f_j(s) Z_s^{f_j}$. As $f_j(t)$ is non-random, $\{Z_t^{f_j}\}$ is adapted to $\{\mathscr{F}_t\}$ by construction. Also, as $\int_0^t f(s)\, dW_s$ and $\frac{1}{2}\int_0^t f(s)^2\, ds$ are continuous in t a.s., so does $Z_t^{f_j}$. Thus, $\{\theta_t\}_{t\geq 0}$ is predictable with respect to $\{\mathscr{F}_t\}_{t\geq 0}$. Finally, note from the martingale property of stochastic integral (Theorem 7.2(3)), we have $\mathrm{E}(M_t|\mathscr{F}_0) = \theta_0 + \mathrm{E}\int_0^t \theta_s\, dW_s = \theta_0$. On the other hand, $\mathrm{E}(M_t|\mathscr{F}_0) = M_0$ by the definition of martingale. Thus $\theta_0 = M_0$, and (7.42) follows. □

7.7 Fundamental Theorem of Asset Pricing

In this section, we apply the results in the Section 7.6 to asset pricing.

7.7.1 Self-Financing Portfolio

Similar to the discrete time financial model, we focus on a simple situation that there are only two assets in the market: the bond $B_t = e^{rt}$ and the stock S_t. A trading strategy is denoted by a pair of predictable processes $\{\psi_t, \phi_t\}_{t\geq 0}$ that indicates the holding amount of each asset, i.e, investing ψ_t units of bond B_t and ϕ_t units of stock S_t at time t. This gives a portfolio V which is worth

$$
V_t = \psi_t B_t + \phi_t S_t \,. \tag{7.44}
$$

Using the principle of no arbitrage pricing, we price a financial product by constructing a replicating portfolio. To be specific, suppose that we are at time t and we want to price a product with payoff C_T (a random variable) at time T. We seek for a trading strategy $\{\psi_t, \phi_t\}_{t\geq 0}$ such that the value of the replicating portfolio at maturity, V_T, agrees with C_T in all cases (i.e., $V_T = C_T$ a.s.). Then the price of the product C at t is given by V_t.

It is important to note that, as the initial value V_0 of the replicating portfolio V is the amount required to reproduce $V_T = C_T$ at T, the portfolio should be **self-financing**, i.e., **no external input/output** for V during $[0,T]$. In particular, the changes in value of V is only due to the changes in the bond and stock values. As $\psi_t dB_t + \phi_t dS_t$ is the infinitesimal change of the replicating portfolio at time t, this leads to the formula

$$
dV_t = \psi_t dB_t + \phi_t dS_t \,, \tag{7.45}
$$

for $t \in [0,T]$. Equations (7.44) and (7.45) together motivate the following **self-financing** conditions:

Definition 7.5. (Self-financing Strategy) A **Self-financing strategy** is defined by a pair of adapted process $\{\psi_t\}_{t\geq 0}$, $\{\phi_t\}_{t\geq 0}$, satisfying (7.44),

$$\int_0^t |\psi_u| du + \int_0^t |\phi_u|^2 du < \infty, \;\; a.s., \tag{7.46}$$

for all $t > 0$, and

$$V_t = \psi_t B_t + \phi_t S_t = \psi_0 B_0 + \phi_0 S_0 + \int_0^t \psi_u dB_u + \int_0^t \phi_u dS_u\,, \;\; a.s., \tag{7.47}$$

for all $t > 0$.

Remark 7.3. Equation (7.46) is required so that the integrals in (7.47) make sense. Note that the existence of $\int_0^t \psi_u dB_u = \int_0^t \psi_u r e^{ru} du \leqslant r e^{rt} \int_0^t |\psi_u| du$ requires integrability of ψ_u, while the existence of $\int_0^t \phi_u dS_u$ requires integrability of ϕ_u^2. □

We end this subsection with the following lemma which represents the discounted portfolio value by the discounted stock price. When equipped with the martingale measure (to be introduced in the next subsection), such representation provides a simple way for pricing.

Lemma 7.9 (Self-financing Strategy and Discounted Portfolio Value) *Let* $\{\psi_t\}_{t\geq 0}$ *and* $\{\phi_t\}_{t\geq 0}$ *be predictable processes satisfying* (7.46). *Let* $\widetilde{V}_t = e^{-rt} V_t = e^{-rt}(\psi_t B_t + \phi_t S_t)$ *and* $\widetilde{S}_t = e^{-rt} S_t$, *then* $\{\psi_t, \phi_t\}_{t\geq 0}$ *defines a self-financing strategy if and only if*

$$d\widetilde{V}_t = \phi_t \, d\widetilde{S}_t \quad a.s.,$$

for all $t \leq T$.

Proof. Using Ito's Lemma, we have

$$\begin{aligned}
d\widetilde{V}_t &= d e^{-rt} V_t \\
&= -r e^{-rt} V_t \, dt + e^{-rt} dV_t && \text{(product rule)} \\
&= -r e^{-rt} (\psi_t B_t + \phi_t S_t) \, dt + e^{-rt} (\psi_t dB_t + \phi_t dS_t) \\
&\quad + e^{-rt} (dV_t - \psi_t dB_t - \phi_t dS_t) && \text{((7.44) and "add and subtract")} \\
&= \psi_t (-r e^{-rt} B_t \, dt + e^{-rt} dB_t) + \phi_t (-r e^{-rt} S_t \, dt + e^{-rt} dS_t) \\
&\quad + e^{-rt} (dV_t - \psi_t dB_t - \phi_t dS_t) && \text{(grouping } \psi_t \text{ and } \phi_t\text{)} \\
&= 0 + \phi_t d(e^{-rt} S_t) + e^{-rt} (dV_t - \psi_t dB_t - \phi_t dS_t) && (B_t = e^{rt} \text{ and product rule)} \\
&= \phi_t d\widetilde{S}_t + e^{-rt} (dV_t - \psi_t dB_t - \phi_t dS_t)\,.
\end{aligned}$$

In other words, $d\widetilde{V}_t = \phi_t d\widetilde{S}_t$ if and only if $dV_t - \psi_t dB_t - \phi_t dS_t = 0$, i.e., $\{\psi_t, \phi_t\}_{t\geq 0}$ is a self-financing strategy. □

7.7.2 Self-Financing Replicating portfolio and Martingale Measure

In Lemma 7.9, for a self-financing portfolio with trading strategy $\{\psi_t, \phi_t\}$, we have obtained the representation $d\widetilde{V}_t = \phi_t d\widetilde{S}_t$, or

$$\widetilde{V}_t = \widetilde{V}_s + \int_s^t \phi_u d\widetilde{S}_u\,, \tag{7.48}$$

for $t \geq s \geq 0$. If we only know the distribution of V_t but not the whole path $\{\psi_t, \phi_t\}_{t\geq 0}$, the following Lemma provides a way to find the value of a portfolio at time s, V_s, by simply taking expectation.

Lemma 7.10 (Self-financing Strategy and Martingale Measure) *Let $\widetilde{V}_t(\psi,\phi) = e^{-rt}V_t(\psi,\phi)$ be a discounted value of a* **self-financing** *portfolio with some strategy (possibly unknown) $\{\psi_t, \phi_t\}_{t\geq 0}$. If there exists a probability measure $\mathbb{Q}$ such that the process $\{\widetilde{S}_u\}$ is a martingale under $\mathbb{Q}$, then for any $t \geq s \geq 0$,*

$$V_s = E_{\mathbb{Q}}(e^{-r(t-s)}V_t|\mathscr{F}_s)\,.$$

Proof. Using the arguments in Theorem 7.2 c) and the fact that $\widetilde{S}_u$ is a martingale under $\mathbb{Q}$, we have for some partition $s = t_0 < \cdots < t_n = t$ that

$$\begin{aligned}
\mathrm{E}_{\mathbb{Q}}\left(\int_s^t \phi_u d\widetilde{S}_u \middle| \mathscr{F}_s\right) &= \lim_{n\to\infty} \mathrm{E}_{\mathbb{Q}}\left(\sum_{j=1}^n \phi_{t_j}(\widetilde{S}_{t_{j+1}} - \widetilde{S}_{t_j}) \middle| \mathscr{F}_s\right) \\
&= \lim_{n\to\infty} \mathrm{E}_{\mathbb{Q}}\left(\sum_{j=1}^n \phi_{t_j}\mathrm{E}_{\mathbb{Q}}(\widetilde{S}_{t_{j+1}} - \widetilde{S}_{t_j}|\mathscr{F}_{t_j}) \middle| \mathscr{F}_s\right) \\
&= 0\,. \qquad\qquad (\text{since } \mathrm{E}_{\mathbb{Q}}(\widetilde{S}_{t_{j+1}} - \widetilde{S}_{t_j}|\mathscr{F}_{t_j}) = 0)
\end{aligned}$$

Thus, the proof is completed by taking conditional expectation $\mathrm{E}_{\mathbb{Q}}(\cdot|\mathscr{F}_s)$ on both sides of (7.48). □

Remark 7.4. The measure $\mathbb{Q}$ is known as the **Martingale Measure** or the **Risk Neutral Probability Measure**. Under this measure, all discounted portfolio value is a martingale. The term *risk neutral* may be interpreted from the martingale property $\mathrm{E}(e^{-rt}V_t|\mathscr{F}_s) = e^{-rs}V_s$, which asserts that all portfolio on average behave the same as the bond e^{rt}.

7.7.3 Fundamental Theorem of Asset Pricing

Consider pricing a financial product with payoff C_T at maturity. If we can find a self-financing replicating portfolio $\{\psi_t, \phi_t\}_{t\geq 0}$ such that $C_T = V_T$ a.s., then from the principle of no arbitrage, the price of the financial product at time t is V_t. From Lemma 7.10, we have $V_t = \mathrm{E}_{\mathbb{Q}}(e^{-r(T-t)}V_T|\mathscr{F}_t) = \mathrm{E}_{\mathbb{Q}}(e^{-r(T-t)}C_T|\mathscr{F}_t)$, which is the price of C at time t. The Fundamental Theorem of Asset Pricing asserts that if C_t is $\mathscr{F}_t$ measurable ($\{\mathscr{F}_t\}$ is the natural filtration generated by $\{S_t\}$), then there exists a self-financing replicating portfolio $\{\psi_t, \phi_t\}_{t\geq 0}$ such that $C_T = V_T$, a.s.. This offers a general way of pricing financial derivatives.

Theorem 7.11. (Fundamental Theorem of Asset Pricing) *Let $\mathbb{Q}$ be the measure given by Lemma 7.10. Suppose that a financial derivative C at time T is given by the non-negative $\mathscr{F}_T$ measurable random variable C_T. If*

$$E_{\mathbb{Q}}(C_T^2) < \infty ,$$

then C is replicable and the value at time t of any replicating portfolio is given by

$$V_t = E_{\mathbb{Q}}\left(e^{-r(T-t)}C_T|\mathscr{F}_t\right) . \tag{7.49}$$

Proof. Note that from Lemma 7.10, the price of any self-financing portfolio can be computed by taking expectation under the martingale measure. Thus it suffices to show that there exists a self-financing replicating portfolio V with strategy $\{\psi_t, \phi_t\}_{t\geq 0}$ such that $C_T = V_T$ a.s.. Let $\widetilde{V}_t = e^{-rt}V_t$ and $\widetilde{C}_t = e^{-rt}C_t$. From Lemma 7.9, it suffices to find a V_t such that $\widetilde{V}_T = \widetilde{C}_T$ a.s. and $\widetilde{V}_t$ has the representation

$$d\widetilde{V}_t = \phi_t d\widetilde{S}_t ,$$

for some $\{\mathscr{F}_t\}$ predictable process ϕ_t.

The key to the proof is the clever construction $\widetilde{V}_t = \mathrm{E}_{\mathbb{Q}}(\widetilde{C}_T|\mathscr{F}_t)$, where $\mathbb{Q}$ is the martingale measure given in Lemma 7.10 that makes $\widetilde{S}_t$ a martingale. This construction satisfies our requirements that

1) $\widetilde{V}_T = \widetilde{C}_T$ a.s.
2) V_t is self-financing.

First, 1) is trivial since $\widetilde{C}_T$ is $\mathscr{F}_T$ measurable.

Next, 2) follows from the fact that $\widetilde{V}_t$ is a martingale (see Example 5.5) and the Brownian Martingale Representation Theorem 7.8. To be specific, from Theorem 7.8, there exists a $\{\mathscr{F}_t\}_{t\geq 0}$ predictable process $\{\phi_t\}_{t\geq 0}$ such that with $\mathbb{Q}$-probability one, $\widetilde{V}_t = \widetilde{V}_0 + \int_0^t \phi_s \, dW_s^{\mathbb{Q}}$, i.e.,

$$d\widetilde{V}_t = \phi_t \, dW_t^{\mathbb{Q}} , \tag{7.50}$$

where $W_t^{\mathbb{Q}}$ is a Brownian motion under $\mathbb{Q}$. Since $\widetilde{S}_t$ is a $\mathbb{Q}$-martingale by construction, the same argument can be repeated to yield $d\widetilde{S}_u = \theta_u dW_u^{\mathbb{Q}}$ for some θ_u. Thus (7.50) can be expressed as $d\widetilde{V}_t = \frac{\phi_t}{\theta_t} d\widetilde{S}_t$. Hence, Lemma 7.9 implies that V_t is a self-financing portfolio.

In summary, we have constructed a portfolio $V_t = e^{rt}\mathrm{E}_{\mathbb{Q}}(\widetilde{C}_T|\mathscr{F}_t) = \mathrm{E}_{\mathbb{Q}}(e^{-r(T-t)}C_T|\mathscr{F}_t)$ which is self-financing and satisfies $V_T = C_T$ a.s.. Thus, V_t gives the price of C at time t. □

Below is a summary of the general way of pricing in a continuous time financial model:

Pricing via Risk Neutral Probability Measure
To price a financial derivative C with final payoff C_T, first define a probability measure $\mathbb{Q}$ under which the discounted stock process $\widetilde{S}_t = e^{-rt}S_t$ takes the form $d\widetilde{S}_t = \theta_t dW_t^{\mathbb{Q}}$, where $W_t^{\mathbb{Q}}$ is a Brownian motion under $\mathbb{Q}$. This ensures that $\widetilde{S}_t$ is a $\mathbb{Q}$-martingale. Finally, the price of C at time t, V_t, is obtained by

$$V_t = \mathrm{E}_{\mathbb{Q}}(e^{-r(T-t)}C_T|\mathscr{F}_t)\,.$$

Remark 7.5. In the Fundamental Theorem of Asset Pricing, C_t needs to be $\mathscr{F}_t$ measurable, where $\{\mathscr{F}_t\}_{t\geq 0}$ is the natural filtration of the stock S_t. Otherwise, the replication of C_t using S_t will not be possible (Requirement 1) in the proof of Theorem 7.11 fails. For example, if we want to price a European call option C on HSBC stock $S_t^{(5)}$, we can form a replicating portfolio using the bond B_t and the stock $S_t^{(5)}$, but not the bond and China mobile stock $S_t^{(941)}$. In this case, C_t is $\sigma(\{S_s^{(5)}\}_{s\leq t})$ measurable but not $\sigma(\{S_s^{(941)}\}_{s\leq t})$ measurable. □

One of the key ingredient of Fundamental Theorem of Asset Pricing is to find the martingale measure $\mathbb{Q}$ in Lemma 7.10. To achieve this, we need to assume a model for S_t. In the next section, we describe the geometric Brownian motion model which allows $\mathbb{Q}$ to be found via the Girsanov Theorem.

7.8 Pricing Under Geometric Brownian Motion Model

7.8.1 Geometric Brownian Motion (GBM) Model

The **Geometric Brownian Motion** model assumes that the stock price S_t follows an Ito process, i.e.,

$$dS_t = \mu S_t\,dt + \sigma S_t\,dW_t\,, \tag{7.51}$$

with S_0 being current price of the stock. Here, μ is called the *drift* and σ is called the *volatility*. Both μ and σ are assumed to be constant.

Applying Ito's Lemma with $F(t,x) = \ln x$, we have

$$d\ln S_t = \left(\mu - \frac{\sigma^2}{2}\right)dt + \sigma\,dW_t,$$

or

$$S_t = S_0 \exp\left\{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right\}. \tag{7.52}$$

Since S_t is the exponential function of a Brownian motion, this motives the name **Geometric Brownian Motion**.

7.8.2 Martingale Measure for GBM

Recall that in the Fundamental Theorem of Asset Pricing we need to find a martingale measure $\mathbb{Q}$ such that the discounted stock process $\widetilde{S}_t = e^{-rt}S_t$ satisfies

$$d\widetilde{S}_t = \theta_t dW_t^{\mathbb{Q}}, \tag{7.53}$$

where $W_t^{\mathbb{Q}}$ is a Brownian motion under $\mathbb{Q}$ and $\{\theta_t\}_{t\geq 0}$ is a predictable process w.r.t. $\{\mathscr{F}_t\}_{t\geq 0}$. Now, under the GBM model (7.51), we have from Ito's Lemma that

$$\begin{aligned} d\widetilde{S}_t &= (\mu - r)\widetilde{S}_t dt + \sigma\widetilde{S}_t dW_t \\ &= \sigma\widetilde{S}_t d\left(W_t + \frac{\mu - r}{\sigma}t\right). \end{aligned}$$

Therefore, we need a measure $\mathbb{Q}$ such that

$$W_t^{\mathbb{Q}} \equiv W_t + \frac{\mu - r}{\sigma}t \tag{7.54}$$

is a B.M. under $\mathbb{Q}$. The solution follows directly from the Girsanov Theorem 7.7 with $\theta_s = \frac{\mu - r}{\sigma}$, i.e., for $A \in \mathscr{F}_t$,

$$\mathbb{Q}(A) = \int_A L_t d\mathbb{P}, \tag{7.55}$$

where

$$L_t = e^{-\frac{\mu - r}{\sigma}W_t - \frac{1}{2}\left(\frac{\mu - r}{\sigma}\right)^2 t}$$

is the Radon Nikodym Derivative.

7.8.3 Pricing In Action

With the martingale measure for the geometric Brownian Motion model, in this section we demonstrate the use of the Fundamental Theorem of Asset Pricing (Theorem 7.11) in pricing financial derivatives.

Example 7.11. **(European Call Option)** An European call option offers a choice for the holder to buy a stock S at the strike price K. Thus the payoff at maturity T is

$$C_T = (S_T - K)^+ ,$$

where $X^+ = X \cdot I_{\{X>0\}}$. From (7.49), the price of this option at time 0 is given by

$$\begin{aligned}
V_0 &= \mathrm{E}_{\mathbb{Q}}\left(e^{-rT}(S_T - K)^+\right) \\
&= \mathrm{E}_{\mathbb{Q}}\left(e^{-rT}\left(S_0 e^{\left(\mu-\frac{\sigma^2}{2}\right)T+\sigma W_T} - K\right)^+\right) && \text{(from (7.52))} \\
&= \mathrm{E}_{\mathbb{Q}}\left(e^{-rT}\left(S_0 e^{\left(r-\frac{\sigma^2}{2}\right)T+\sigma W_T^{\mathbb{Q}}} - K\right)^+\right) && \text{(from (7.54))} \\
&= e^{-rT}\int_{-d_2\sqrt{T}}^{\infty}(S_0 e^{\left(r-\frac{\sigma^2}{2}\right)T+\sigma x} - K)\phi(x,0,T)\,dx && \left(\text{Let } d_2 = \frac{\ln\frac{S_0}{K} + (r-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) \\
&= \int_{-d_2}^{\infty}\left((S_0 e^{-\frac{\sigma^2}{2}T+\sigma u\sqrt{T}} - Ke^{-rT})\phi(u)\right)du && \left(\text{change of variables} : u = \frac{x}{\sqrt{T}}\right) \\
&= S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2),
\end{aligned}$$

where $\phi(x,0,T)$ is the p.d.f. of $N(0,T)$, $\phi(x)$ and $\Phi(x)$ are the p.d.f. and c.d.f. of $N(0,1)$ distribution, and $d_1 = d_2 + \sigma\sqrt{T}$.

Example 7.12. **(Digital Options)** A **digital call option**, also known as **binary call option** or a **cash or nothing call option**, has payoff structure

$$C_T = 1_{\{S_T \geq K\}} .$$

From (7.49), the price of this option at time 0 is given by

$$
\begin{aligned}
V_0 &= \mathrm{E}_{\mathbb{Q}}\left(e^{-rT}1_{\{S_T \geq K\}}\right) \\
&= \mathrm{E}_{\mathbb{Q}}\left(e^{-rT}1_{\left\{S_0 e^{\left(\mu-\frac{\sigma^2}{2}\right)T+\sigma W_T} \geq K\right\}}\right) && \text{(from (7.52))} \\
&= \mathrm{E}_{\mathbb{Q}}\left(e^{-rT}1_{\left\{S_0 e^{\left(r-\frac{\sigma^2}{2}\right)T+\sigma W_T^{\mathbb{Q}}} \geq K\right\}}\right) && \text{(from (7.54))} \\
&= e^{-rT}\int_{-d_2\sqrt{T}}^{\infty} \phi(x,0,T)\,dx \qquad \left(\text{Let } d_2 = \frac{\ln\frac{S_0}{K}+(r-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) \\
&= e^{-rT}\int_{-d_2}^{\infty} \phi(x)\,dx = e^{-rT}\Phi(d_2)\,,
\end{aligned}
$$

where $\phi(x,0,T)$, $\phi(x)$ and $\Phi(x)$ are defined in Example 7.11.

Example 7.13. **(Barrier Options)** A barrier option can be *activated* or *deactivated* if the asset price crosses a barrier before the maturity. By considering the combinations of the following two features, there are four types of barrier options:

i) Direction of hitting the barrier:
 Up: the barrier is hit from below; **Down**: the barrier is hit from above.
ii) Action when the barrier is hit:
 In: the option is activated; **Out**: the option is deactivated.

Note that these two features can be applied to any options. For example, we have down-and-in European call options, up-and-out digital put options, or down-and-out Asian call options, etc...

Since the maximum or minimum of the stock price process determines whether the barrier is hit, its distribution will be useful in pricing barrier options. Recall from Lemma 6.9 the joint distributions of W_t and $M_t = \max_{t>s>0} W_s$, where W_t is a $\mathbb{P}$-B.M.:

$$\mathbb{P}(M_t > a, W_t \in dx) = \phi\left(\frac{2a-x}{\sqrt{t}}\right)\frac{1}{\sqrt{t}} \tag{7.56}$$

$$\mathbb{P}(M_t \in da, W_t \in dx) = \phi\left(\frac{2a-x}{\sqrt{t}}\right)\frac{2(2a-x)}{t^{3/2}}\,, \tag{7.57}$$

for $a \geq x$, where (7.57) is obtained from (7.56) by differentiation w.r.t. a. Note that $\mathbb{P}(M_t \in da, W_t \in dx) = 0$ for $a < x$ since the running maximum is never less than the current value. Therefore, we can see that

$$\mathbb{P}(M_t > a, W_t \in dx) = \mathbb{P}(M_t > x, W_t \in dx)$$

for $a < x$. We illustrate the pricing procedure using the following **up-and-in digital option** with payoff:

$$C_T = 1_{\{M_T \geq K\}}\,,$$

where $M_T \equiv \max_{T>t>0} S_t$ and $K > S_0$. Again, from (7.49), the price of this option at time 0 is given by

$$\begin{aligned}
V_0 &= \mathrm{E}_{\mathbb{Q}}\left(e^{-rT} 1_{\{M_T \geq K\}}\right) \\
&= \mathrm{E}_{\mathbb{Q}}\left(e^{-rT} 1_{\left\{\max_{T>t>0} S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t} \geq K\right\}}\right) \qquad \text{(from (7.52))} \\
&= \mathrm{E}_{\mathbb{Q}}\left(e^{-rT} 1_{\left\{\max_{T>t>0} S_0 e^{\left(r - \frac{\sigma^2}{2}\right)t + \sigma W_t^{\mathbb{Q}}} \geq K\right\}}\right) \qquad \text{(from (7.54))} \\
&= \mathrm{E}_{\widetilde{\mathbb{Q}}}\left(e^{-rT} 1_{\left\{\max_{T>t>0} S_0 e^{\sigma W_t^{\widetilde{\mathbb{Q}}}} \geq K\right\}} e^{bW_T^{\widetilde{\mathbb{Q}}} - \frac{1}{2}b^2 T}\right) \qquad (\text{Girsanov:} W_t^{\widetilde{\mathbb{Q}}} = W_t^{\mathbb{Q}} + bt \text{ is a } \widetilde{\mathbb{Q}} - B.M.,\ b = \frac{r}{\sigma} - \frac{\sigma}{2}) \\
&= \int\int_{a>x} e^{-rT} 1_{\{S_0 e^{\sigma a} \geq K\}} e^{bx - \frac{1}{2}b^2 T} \widetilde{\mathbb{Q}}(\widetilde{M}_T \in da, W_T^{\widetilde{\mathbb{Q}}} \in dx)\, da\, dx \quad (\text{from (7.57), where } \widetilde{M}_T = \max_{T>t>0} W_t^{\widetilde{\mathbb{Q}}}) \\
&= \int_{-\infty}^{\infty}\int_{c\vee x}^{\infty} e^{-rT} e^{bx - \frac{1}{2}b^2 T} \widetilde{\mathbb{Q}}(\widetilde{M}_T \in da, W_T^{\widetilde{\mathbb{Q}}} \in dx)\, da\, dx \qquad (\text{where } c \triangleq \log(K/S_0)/\sigma) \\
&= \int_{-\infty}^{\infty} e^{-rT} e^{bx - \frac{1}{2}b^2 T} \widetilde{\mathbb{Q}}(\widetilde{M}_T > c \vee x, W_T^{\widetilde{\mathbb{Q}}} \in dx)\, dx \qquad \text{(evaluate the inner integral)} \\
&= \int_{-\infty}^{c} e^{-rT} e^{bx - \frac{1}{2}b^2 T} \phi\left(\frac{2c - x}{\sqrt{T}}\right) \frac{1}{\sqrt{T}}\, dx + \int_{c}^{\infty} e^{-rT} e^{bx - \frac{1}{2}b^2 T} \phi\left(\frac{x}{\sqrt{T}}\right) \frac{1}{\sqrt{T}}\, dx \\
&= \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi T}} e^{-rT} e^{bx - \frac{1}{2}b^2 T} e^{-\frac{(2c-x)^2}{2T}}\, dx + \int_{c}^{\infty} \frac{1}{\sqrt{2\pi T}} e^{-rT} e^{bx - \frac{1}{2}b^2 T} e^{-\frac{x^2}{2T}}\, dx \\
&\qquad \left(\text{Normal pdf: } \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\right) \\
&= e^{2bc - rT} \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi T}} e^{-\frac{(x - 2c - bT)^2}{2T}}\, dx + e^{-rT} \int_{c}^{\infty} \frac{1}{\sqrt{2\pi T}} e^{-\frac{(x - bT)^2}{2T}}\, dx \\
&= e^{2bc - rT} \Phi\left(\frac{-c - bT}{\sqrt{T}}\right) + e^{-rT}\left(1 - \Phi\left(\frac{c - bT}{\sqrt{T}}\right)\right).
\end{aligned}$$

Example 7.14. **(American Options)** The key feature of American options is that the time of exercising the option is not determined. Therefore, the "Maturity" has to be modeled by a stopping time τ. The probabilistic properties about the stopping time developed in Chapter 6 will be useful in pricing American options.

Consider the American option where \$ K is paid at the first time that the stock price S_t reaches a, $a > S_0$. Let $T_a = \inf\{t \geqslant 0 : S_t = a\} = \inf\{t \geqslant 0 : S_0 e^{(r - \frac{\sigma^2}{2})t + \sigma W_t^{\mathbb{Q}}} = a\}$, the payoff is given by

$$C_{T_a} = K\,.$$

From (7.49), the price of this option at time 0 is given by

$$\begin{aligned}
V_0 &= \mathrm{E}_{\mathbb{Q}}\left(Ke^{-rT_a}\right)\\
&= \mathrm{E}_{\mathbb{Q}}\left(Ke^{-r\inf\left\{t\geqslant 0:\, S_0 e^{(r-\frac{\sigma^2}{2})t+\sigma W_t^{\mathbb{Q}}}=a\right\}}\right)\\
&= \mathrm{E}_{\mathbb{Q}}\left(Ke^{-r\inf\{t\geqslant 0:\, W_t^{\mathbb{Q}}=c-bt\}}\right) \qquad \left(\text{set } c=\frac{1}{\sigma}\log\frac{a}{S_0},\ b=\frac{r}{\sigma}-\frac{\sigma}{2}\right)\\
&= Ke^{-c(-b+\sqrt{2r+b^2})}\,. \qquad (\text{Example 7.10})
\end{aligned}$$

7.9 Exercises

Exercise 7.1. Let $0 = t_0^n < t_1^n < \cdots < t_n^n = T$, where $t_j^n = \frac{jT}{n}$, be a partition of the interval $[0,T]$ into n equal parts. Using the identity $a(b-a) = \frac{1}{2}\left(b^2-a^2\right) - \frac{1}{2}(a-b)^2$ and Theorem 6.12, show that

$$\lim_{n\to\infty}\sum_{j=0}^{n-1} W(t_j^n)\left(W(t_{j+1}^n)-W(t_j^n)\right) = \frac{1}{2}W_T^2 - \frac{T}{2}\,. \tag{7.58}$$

Using the identity $b(b-a) = \frac{1}{2}\left(b^2-a^2\right) + \frac{1}{2}(a-b)^2$ and Theorem 6.12, show that

$$\lim_{n\to\infty}\sum_{j=0}^{n-1} W(t_{j+1}^n)\left(W(t_{j+1}^n)-W(t_j^n)\right) = \frac{1}{2}W_T^2 + \frac{T}{2}\,. \tag{7.59}$$

Find the limit

$$\lim_{n\to\infty}\sum_{j=0}^{n-1} W(pt_{j+1}^n + (1-p)t_j^n)\left(W(t_{j+1}^n)-W(t_j^n)\right)\,, \tag{7.60}$$

for $p \in (0,1)$.

Exercise 7.2. Show that if a sequence converges to some limit, then every subsequence converges to the same limit.

Exercise 7.3. Show that W_t^2 belongs to $\mathscr{M}^2$, where W_t is a Wiener process.

Exercise 7.4. Suppose that τ is a stopping time w.r.t. the filtration $\{\mathscr{F}_t\}_{t\geq 0}$. Is there any t such that $\tau \in \mathscr{F}_t$?

Exercise 7.5. Using the notations in Theorem 7.7, show that if $A \in \mathscr{F}_t$, then $\mathbb{Q}(A) := \int_A L_t(\omega)\mathbb{P}(d\omega) = \int_A L_T(\omega)\mathbb{P}(d\omega)$ for any $T > t$. If $s < t$, is $\mathbb{Q}(A) := \int_A L_t(\omega)\mathbb{P}(d\omega) = \int_A L_s(\omega)\mathbb{P}(d\omega)$ true? Prove it or give a counter example.

Exercise 7.6. Verify the equalities

$$\int_0^T t\,dW_t = TW_T - \int_0^T W_t\,dt, \tag{7.61}$$

and

$$\int_0^T W_t^2\,dW_t = \frac{1}{3}W_T^3 - \int_0^T W_t\,dt. \tag{7.62}$$

Note that the integral on the right-hand side is a Riemann integral defined path-wise, i.e. defined separately for each $\omega \in \Omega$.

Exercise 7.7. Suppose that $g(x)$ is a continuous, and with $t_i^n = iT/n$, $0 = t_0 < t_1 < \cdot < t_n = T$ is a partition of $[0,T]$. Show that $\lim_{n\to\infty} \sup_{i=1,\ldots,n} \sup_{t\in[t_i^n, t_{i+1}^n]} |g(t_i^n) - g(t)| \to 0$. If $T = \infty$, does the convergence hold? Prove it or give a counter example.

Exercise 7.8. (General Case of Ito's Lemma) Suppose that $\psi_n(x)$ is a smooth function satisfying $\psi_n(x) = 1$ for any $x \in [-n,n]$ and $\psi_n(x) = 0$ for $x \notin [-n-1, n+1]$. Let $F_n(t,x) = \psi_n(x)F(t,x)$, show that $F_n(t,x)$ satisfies the conditions of Theorem 7.3 and has bounded partial derivatives $(F_n)_x$ and $(F_n)_{xx}$ for each n.

Since $F(t,x) = F_n(t,x)$ for every $t \in [0,T]$ and $x \in [-n,n]$, the Ito's Lemma holds for F on the set $A_n = \{\sup_{t\in[0,T]} |W_t| < n\}$. By show that $\lim_{n\to} P(A_n) = 1$, deduce that Ito's Lemma holds for F almost surely.

Exercise 7.9. For $F(t,x) = x^3$ we have $F_t(t,x) = 0$, $F_x(t,x) = 3x^2$ and $F_{xx}(t,x) = 6x$. Verify that $3W_t^2 \in \mathscr{M}_T^2$ and use Ito's Lemma to show the equality

$$d\left(W_t^3\right) = 3W_t\,dt + 3W_t^2\,dW_t\,.$$

Exercise 7.10. Show that the exponential martingale $X_t = \exp\left\{W_t - \frac{t}{2}\right\}$ satisfies

$$dX_t = X_t\,dW_t\,.$$

Exercise 7.11. Show that the initial value problem

$$\begin{cases} dX_t = aX_t\,dt + bX_t\,dW_t, \\ X(0) = x_0 \end{cases}$$

has a solution given by $X_t = x_0 \exp\left\{\left(a - \frac{b^2}{2}\right)t + bW_t\right\}$.

Exercise 7.12. Show Equation (7.43).

Exercise 7.13. Show that for any two random variables X and Y, $\|X - Y\|_{\mathscr{L}^2} = 0$ implies $X = Y$ a.s..

Consider a sequence of random variables X_n. If $\|X_n - X\|_{\mathscr{L}^2} \to 0$, does it imply $X_n \to X$ a.s.? Prove or give counter examples.

Exercise 7.14. Use Ito's Lemma to compute $\mathrm{E}(W_t^6)$.

Exercise 7.15. Recall the First Borel-Cantelli Lemma that $\mathbb{P}(\bigcap_{i=1}^{\infty}\bigcup_{j=i}^{\infty}A_j) = 0$ if $\sum_{i=1}^{\infty}\mathbb{P}(A_i) < \infty$.

1. Let $A_j = \{|X_j - X| > 2^{-j}\}$. Show that $\mathbb{P}(\bigcap_{i=1}^{\infty}\bigcup_{j=i}^{\infty}A_j) = 0$ implies $X_j \stackrel{a.s.}{\to} X$.
2. Suppose that $X_n \stackrel{p}{\to} X$. Show that there is a subsequence X_{n_j} of X_n such that $\mathbb{P}(|X_{n_j} - X| > 2^{-j}) < 2^{-j}$ along the subsequence.
3. Show that, if X_n converges to X in probability, then there is some subsequence $\{X_{n_j}\}$ which converges to X a.s..

Exercise 7.16. Find the quadratic variation of X_t where X_t is an Ito's process satisfying (7.23). Show that $d[X]_t = b^2(t)dt$.

Exercise 7.17. Given $X_t = x_0e^{at+bW_t}$, where $a, b \in \mathbb{R}$, verify that $X_t \in \mathscr{M}^2$.

Exercise 7.18. Let S_t follow the GBM model (7.51), prove that

1. the density function $f(x)$ of $\frac{S_t}{S_0}$ is given by

$$f(x) = \frac{1}{\sqrt{2\pi t}\sigma x}\exp\left\{-\frac{\left[\ln x - \left(\mu - \frac{\sigma^2}{2}\right)t\right]^2}{2\sigma^2 t}\right\}\mathbf{1}_{\mathbb{R}^+}(x).$$

2. $\mathbb{E}S_t = S_0\exp\{\mu t\}$.
3. $\mathrm{Var}(S_t) = S_0^2\exp\{2\mu t\}\left(\exp\left\{\sigma^2 t\right\} - 1\right)$.

Exercise 7.19. Complete the calculation of $\mathrm{E}_{\mathbb{P}}(S_t 1_{\{\max_{t>u>0} S_u > K\}})$ in Example 7.9.

Exercise 7.20. Let $\{X_t\}$ be a $\{\mathscr{F}_t\}$ adapted process and $\{L_t\}$ be the Radon-Nikodym process such that $L_t = \left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_t$ is the R.N.D. of $\mathbb{Q}$ w.r.t. $\mathbb{P}$ when both are defined on $(\Omega, \mathscr{F}_t)$. Let $t > x$, show that

1. L_t is a $\mathbb{P}$-martingale.
2. For $A \in \mathscr{F}_s$, $\int_A X_s L_t\, d\mathbb{P} = \int_A X_s L_s\, d\mathbb{P} = \int_A X_s\, d\mathbb{Q}$.

For $A \in \mathscr{F}_s$, is $\int_A X_t L_t\, d\mathbb{P} = \int_A X_t L_s\, d\mathbb{P} = \int_A X_t\, d\mathbb{Q}$?

End of the book. Thank you for taking RMSC 4005.