Introduction to Financial Mathematics

Concepts and Computational Methods

Arash Fahim

Introduction to Financial Mathematics

Concepts and Computational Methods

Arash Fahim
Florida State University

July 8, 2019

This publication was made possible by an Alternative Textbook Grant issued by Florida State University Libraries. This work is published by Florida State University Libraries, 116 Honors Way, Tallahassee, FL 32306.

DOI: 10.33009/financialmath1

Dedicated to Nasrin, Sevda and Idin.

Notations

Table 1: Table of Notation

$A \cup B$	$\triangleq$	union of two disjoint set A and B
$\mathbb{R}^d$	$\triangleq$	d-dimensional Euclidean space
$\mathbb{R}^d_+$	$\triangleq$	All points in $\mathbb{R}^d$ with nonnegative coordinates
$x \in \mathbb{R}^d$	$\triangleq$	d-dimensional column vector
$\min(x, y) = (\min(x_1, y_1), ..., \min(x_n, y_n))$	$\triangleq$	component-wise minimum of two vectors
$x_+ = \min(x, 0)$	$\triangleq$	component-wise minimum of a vector with zero vector
A^T	$\triangleq$	transpose of matrix A
$x \cdot y = x^\mathsf{T} y$	$\triangleq$	inner (dot) product in Euclidean space
$A = 0$	$\triangleq$	all the entities of the matrix (vector) A are zero
$A \neq 0$	$\triangleq$	at least of the entities of the matrix (vector) A is nonzero
$A \geqslant 0$ ($A > 0$)	$\triangleq$	all the entities of the matrix (vector) A are non-negative (positive)
∇f	$\triangleq$	column vector of gradient for a differentiable function $f : \mathbb{R}^d \to \mathbb{R}$
$\nabla^2 f$	$\triangleq$	Hessian matrix for a twice differentiable function $f : \mathbb{R}^d \to \mathbb{R}$
$\mathbb{P}$ or $\hat{\mathbb{P}}$	$\triangleq$	Probability
$\mathbb{P}(A)$ (resp. $\hat{\mathbb{P}}(A)$)	$\triangleq$	Probability of an event A under probability $\mathbb{P}$ (resp. $\hat{\mathbb{P}}$)
$\mathbb{E}[X]$ (resp. $\hat{\mathbb{E}}[X]$)	$\triangleq$	Expected value of a random variable X under probability $\mathbb{P}$ (resp. $\hat{\mathbb{P}}$)

Preface

The story of this book started when I was assigned to teach an introductory financial mathematics course at Florida State University. Originally, this course was all measure theory, integration and stochastic analysis. Then, it evolved to cover theory of measures, some probability theory, and option pricing in the binomial model. When I took over this course, I was not sure what I was going to do. However, I had a vision to educate students about some new topics in financial mathematics, while keeping the classical risk management material. My vision was to include some fundamental ideas that are shared between all models in financial mathematics, such as martingale property, Makrovian property, time-homogeneity, and the like, rather than studying a comprehensive list of models. To start, I decided to seek advise from a colleague to use a textbook by two authors, a quantitative financial analysts and a mathematician. The textbook was a little different and covered various models that quants utilize in practice. The semester started, and as I was going through the first couple of sections from the textbook, I realized that the book was unusable; many grave mistakes and wrong theorems, sloppy format, and coherency issues made it impossible to learn from this textbook. It was my fault that I only skimmed the book before the start of semester. A few months later, I learned that another school had had the same experience with the book as they invited one of the authors to teach a similar course. Therefore, I urgently needed a plan to save my course. So, I decided to write my own lecture notes based on my vision, and, over the past three years, these lecture notes grew and grew to include topics that I consider useful for students to learn. In 2018, the Florida State University libraries awarded me the "Alternative Textbook Grant" to help me make my lecture notes into an open access free textbook. This current first edition is the result of many hours of effort by my library colleagues and myself.

Many successful textbooks on financial mathematics have been developed in the recent decades. My favorite ones are the two volumes by Steven Shreve, *Stochastic Calculus for Finance I and II*; [27, 28]. They cover a large variety of topics in financial mathematics with emphasis on the option pricing, the classical practice of quantitative financial analysts (quants). It also covers a great deal of stochastic calculus which is a basis for modeling almost all financial assets. Option pricing remains a must-know for every quant and stochastic calculus is the language of the quantitative finance. However over time, a variety of other subjects have been added to the list of what quants need to learn, including efficient computer programming, machine learning, data mining, big data, and so on. Many of these topics were irrelevant in 70s, when the quantitative finance was initially introduced. Since then, financial markets has changed in the tools that the traders use, and the speed of transactions. This is a common feature of many disciplines that the amount of data that can be used to make business decisions is too large to be handled by classical statistical techniques. Also, financial regulations has been adjusted to the new market environment. They now require financial institutions to provide structured measurements of some of their risks that is not included in the classical risk management theory. For instance, after the financial meltdown in 2007, systemic risk and central clearing became important research areas for the regulator. In addition, a demand for more robust evaluation of risks led to researches in the robust risk management and model risk evaluations.

As the financial mathematics career grows to cover the above-mentioned topics, the prospect of the financial mathematics master's programs must also become broader in topics. In the current book, I tried to include some new topics in an introductory level. Since this is an open access book, it has the ability to include more of the new topics in financial mathematics.

One of the major challenges in teaching financial mathematics is the diverse background of students, at least in some institutions such as Florida State University. For example, some students whom I observed during the last five years, have broad finance background but lack the necessary mathematical background. They very much want to learn the mathematical aspects, but with fewer details and stepping more quickly into the implementational aspects. Other students have majors in mathematics, engineering or computer science who need more basic knowledge in finance. One thing that both groups need is to develop their problem-solving abilities. Current job market favors employees who can work independently and solve hard problems, rather than those who simply take instructions and implement them. Therefore, I designed this book to serve as an introductory course in financial mathematics with focus on conceptual understanding of the models and problem solving, in contrast to textbooks that include more details of the specific models. It includes the mathematical background needed for risk management, such as probability theory, optimization, and the like. The goal of the book is to expose the reader to a wide range of basic problems, some of which emphasize analytic ability, some requiring programming techniques and others focusing on statistical data analysis. In addition, it covers some areas which are outside the scope of mainstream financial mathematics textbooks. For example, it presents marginal account setting by the CCP and systemic risk, and a brief overview of the model risk.

One of the main drawbacks of commercial textbooks in financial mathematics is the lack of flexibility to keep up with changes of the discipline. New editions often come far apart and with few changes. Also, it is not possible to modify them into the course needed for a specific program. The current book is a free, open textbook under a creative common license with attribution. This allows instructors to use parts of this book to design their own course in their own program, while adding new parts to keep up with the changes and the institutional goals of their program.

The first two chapters of this book only require calculus and introductory probability and can be taught to senior undergraduate students. There is also a brief review of these topics in Sections A.1 and B of the appendix. I tried to be as brief as possible in the appendix; many books, including *Stochastic Calculus for Finance I* ([27, 28]) and *Convex Optimization* ([8]), cover these topics extensively. My goal to include these topics is only to make the current book self-sufficient. The main goal of Chapter 1 is to familiarize the reader with the basic concepts of risk management in financial mathematics. All these concepts are first introduced in a relatively nontechnical framework of one-period such as Markowitz portfolio diversification or the Arrow-Debreu market model. Chapter 2 generalizes the crucial results of the Arrow-Debreu market model to the multiperiod case and introduces the multiperiod binomial model and the numerical methods based on it. Chapter 3 discusses more advanced subjects in probability, which are presented in the remainder of Section B and Section C of the appendix. This chapter is more appropriate for graduate students. In Section 3.2, we first build important concepts and computational methods in continuous-time through the Bachelier model. Then, we provide the outline for the more realistic Black-Scholes model in Section 3.3. Chapter 4 deals with pricing a specific type of financial derivative: American options. Sections 4.0.1 and 4.1 can be studied directly after finishing Chapter 2. The rest of this section requires an understanding of Section 3.3 as a prerequisite. The inline exercises and various examples can help students to prepare for the exams on this book. Many of the exercises and the examples are brand new and are specifically created for the assignments and exams during the three last years of teaching the course.

Acknowledgements

I am grateful to the staff of Florida State University libraries who awarded me the *Alternative Textbook Grant* and guided me through the process of creating this book from my lecture notes; especially my thanks goes to Devin Soper who introduced me to the grant and walked me through the process. I learned all I know about different licenses of intellectual material from him. The copyedit was also managed by Devin. Laura Miller, a library intern, has been very helpful with Pandoc, a software that converts LaTeXinto MS Word. She also created some designs of the verso page for me to chose. Matthew Hunter helped me to find the right option when I needed to do indexing and other thing that I never did before.

Table of Contents

1

Preliminaries of finance and risk management

A *risky asset* is an asset with an uncertain future price, e.g. the stock of a company in an exchange market. Unlike a risky asset, a bank account with a fixed interest rate has an absolutely predictable value, and is called a *risk-free* asset. Risky assets are classified in many different categories. Among them, *financial securities*, which constitute the largest body of risky assets, are traded in the exchange markets and are divided into three subcategories: *equity*, *debt*, and *derivatives*.

An equity is a claim of ownership of a company. If it is issued by a corporation, it is called common stock, stock, or share. Debt, sometimes referred to as a *fixed-income instrument*, promises a fixed cash flow until a time called *maturity* and is issued by an entity as a means of borrowing through its sale. The cash flow from a fixed-income security is the return of the borrowed cash plus interest and is subject to default of the issuer, i.e., if the issuer is not able to pay the cash flow at any of the promised dates. A derivative is an asset whose price depends on a certain event. For example, a derivative can promise a payment (payoff) dependent on the price of a stock, the price of a fixed-income instrument, the default of a company, or a climate event.

An important class of assets that are not financial securities are described as *commodities*. Broadly speaking, a commodity is an asset which is not a financial security but is still traded in a market, for example crops, energy, metals, and the like. Commodities are in particular important because our daily life depends on them. Some of them are storable such as crops, which some others, such as electricity, are not. Some of them are subject to seasonality, such as crops or oil. The other have a constant demand throughout the year, for instance aluminum or copper. These various features of commodities introduce challenges in modeling commodity markets. There are other assets that are not usually included in any of the above classes, for instance real estate.

If the asset is easily traded in an exchange market, it is called *liquid*. Equities are the most liquid of assets; fixed-income instruments and derivatives are less liquid. Commodities have become very liquid, partly due to the introduction of emerging economies in the global marketplace. Real estate is one of the most illiquid of assets.

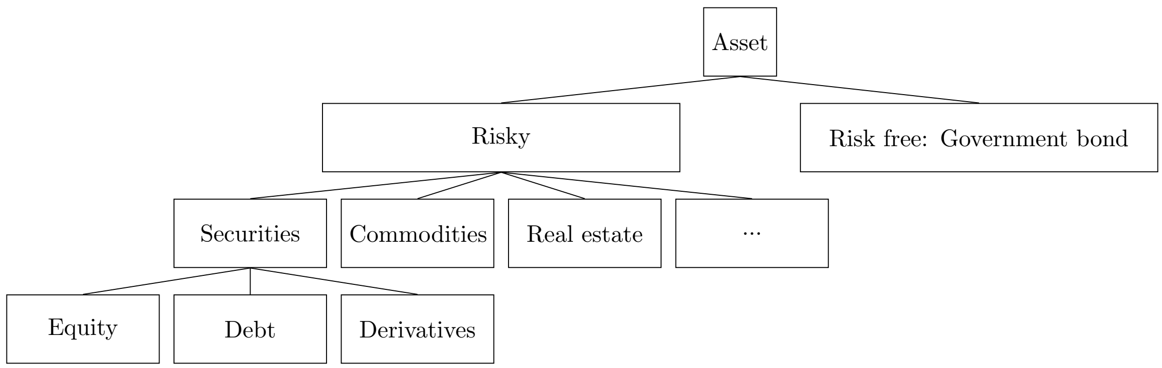


Figure 1.0.1: Classification of assets

1.1 Basic financial derivatives

Financial risk is defined as the risk of loss of investment in financial markets. Two of the main categories of financial risk are market risk, causes by the changes in the price of market equities, and credit risk, caused by the default of a party in meeting its obligations. Financial derivatives are designed to cover the loss caused by the market risk and the credit risk. There are other important forms of financial risk such as operational risk and systemic risk. However, these are irrelevant to the study of financial derivatives. Therefore, in this section, we cover the basics of some simple financial derivatives on the market risk, bonds and some credit derivatives. In practice, bonds are not considered as derivative. However, theoretically, a bond is a derivative on the interest rate.

1.1.1 Futures and forward contracts

Forward and futures contracts are the same in principle, but they differ in operational aspects. In both contracts, the two parties are obliged to exchange a specific asset at a specific date in the future at a fair price that they have agreed upon. The asset subject to exchange is called the *underlying asset*; the date of exchange is called the *maturity date*; and the price is called the *forward/futures price*. In other words, futures and forward contracts lock the price at the moment of a deal in the future.

Forward contracts are simpler than futures. They are nontradable contracts between two specific parties, one of whom is the buyer of the underlying asset, or the *long position*, and the other is the seller of the underlying asset, or the *short position*. The buyer (seller) is

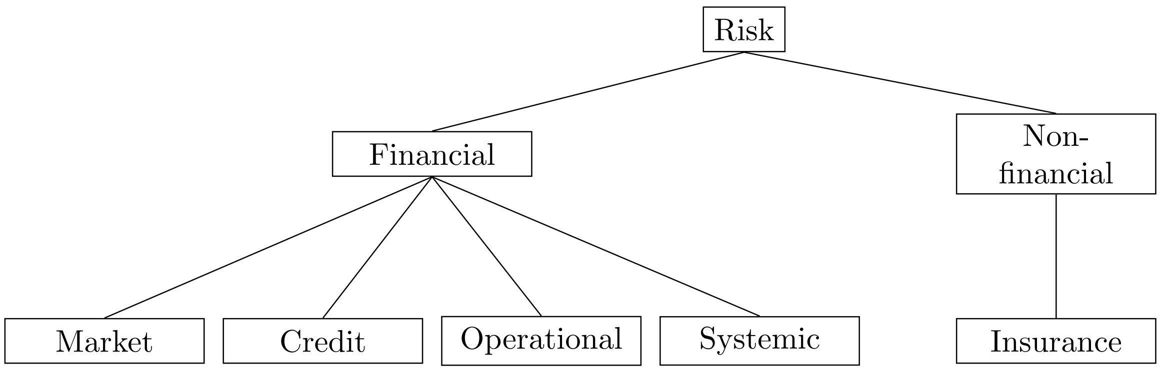


Figure 1.1.1: A classification of risks

obliged to buy (sell) a determined number of units of the underlying asset from the seller (to the buyer) at a price specified in the forward contract, called the *forward price*. The forward price is usually agreed upon between two parties at the initiation of the contract. The forward contract price is not universal and depend upon what the two parties agree upon. Two forward contracts with the same maturity on the same underlying asset can have two different forward prices. Usually, one party is the issuer of the forward contract and quotes the forward price to the other party, or the *holder*, who faces a decision to agree or decline to enter the deal. Generally, the issuer is a financial firm and the holder is a financial or industrial firm. Unlike forward contracts, futures contracts are tradable

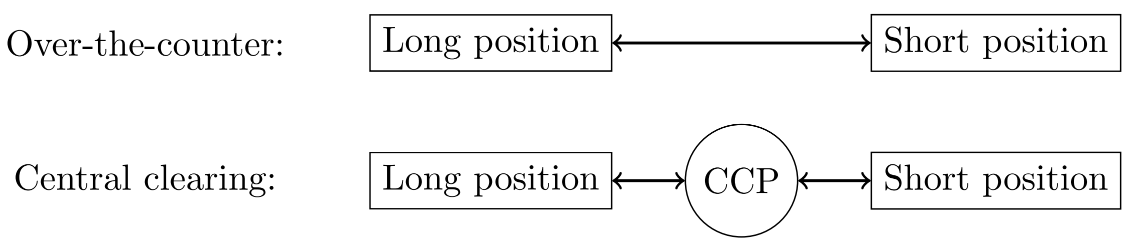


Figure 1.1.2: Forward (top) versus futures (bottom). In futures markets, the CCP regulates the contracts to eliminate the counterparty risk.

in specialized markets. Therefore, given a fixed underlying asset and a fixed maturity T, across the market there is only one *futures price*, a price listed in the futures market. In other words, the futures price at time t for delivery date (maturity) T is not agreed upon between two parties only; rather, it reflects the cumulative attitude of all investors toward the price of the underlying asset at maturity T. The futures price is different from the current price of the underlying, the *spot price*. For a specific underlying asset, we denote by $F_t(T)$ the futures price at time t for delivery at T and by S_t the spot price at time t .

$F_t(T)$ and S_t are related through

$$\lim_{t \to T} F_t(T) = S_T.$$

To avoid counterparty risk, i.e., the risk that either of the parties might be unable to meet their obligation on the futures, the market has a *central counterparty clearinghouse* (CCP). When a trader enters the futures market, per the CCP regulation, he or she is required to open a *marginal account* that is managed by the CCP. The marginal account works as collateral; if the holder cannot meet her obligation, the CCP closes the account to cover the failure of the party. A holder of a futures contract is supposed to keep the amount of money in the marginal account above a level variable with changes in the futures price. To understand the operation of a marginal account, consider a long position, the party who is obliged to buy the underlying at time T at price $F_0(T)$. The financial gain/loss from a derivative is called the *payoff* of the derivative. Her financial payoff at time T is $S_T - F_0(T)$, because she is obliged to buy the underlying at price $F_0(T)$ while the market price is S_T. See Figure 1.1.3. While $F_0(T)$ is fixed and remains unchanged over the term of the contract, the underlying price S_T is unknown. If at time T, $F_0(T) > S_T$, then she loses the amount of $F_0(T) - S_T$. The aim of the CCP is to make sure that the long position

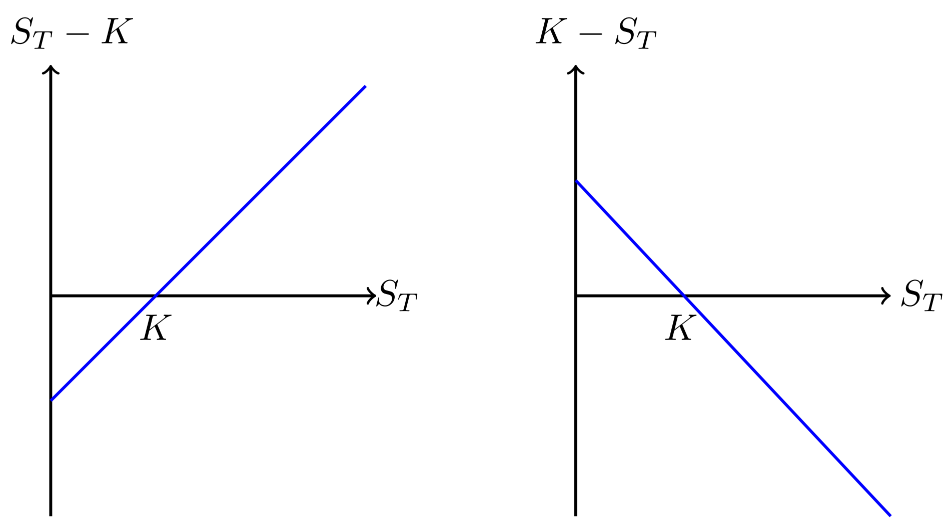


Figure 1.1.3: The payoff of forward/futures at maturity T as a function of the price of underlying S_T. K is the forward/futures price, i.e., $K = F_0(T)$. Left: long position. Right: short position.

holds at least $F_0(T) - S_T$ in her marginal account when is in the loosing position, i.e., $F_0(T) > S_T$. To do this, the CCP asks her to always rebalance her marginal account to keep it above $(F_0(T) - S_t)_+$[1] at any day $t = 0, ..., T$ to cover the possible future loss. An example of marginal account rebalancing is shown in Table 1.1.

Time t	0	1	2	3	4
Underlying asset price S_t	87.80	87.85	88.01	88.5	87.90
Marginal account	.20	.15	0	0	.10
Changes to the marginal account	–	-.05	-.15	0	+.10

Table 1.1: Rebalancing the marginal account of a long position in futures with $F_0(T)$=\$88 in four days.

The marginal account can be subject to several regulations, including minimum cash holdings. In this case, the marginal account holds the amount of $(F_0(T) - S_t)_+$ plus the minimum cash requirement. For more information of the mechanism of futures markets, see [16, Chapter 2].

The existence of the marginal account creates an opportunity cost; the fund in the marginal account can alternatively be invested somewhere else for profit, at least in a risk-free account with a fixed interest rate. The following example illustrates the opportunity cost.

Example 1.1.1 (Futures opportunity cost)**.** *Consider a futures contract with maturity T of 2 days, a futures price equal to \$99.95, and a forward contract with the same maturity but a forward price of \$100. Both contracts are written on the same risky asset with spot price S_0 = \$99.94. The marginal account for the futures contract has a \$10 minimum cash requirement and should be rebalanced daily thereafter according to the closing price. We denote the day-end price by S_1 and S_2 for day one and day two, respectively. Given that the risk-free daily compound interest rate is 0.2%, we want to find out for which values of the spot price of the underlying asset, (S_1, S_2), the forward contract is more interesting than the futures contract for long position.*

Time t	*0*	*1*	*2*
Underlying asset price S_t	*99.94*	S_1	S_2
Marginal account	*10.01*	*10+*$(99.95 - S_1)_+$	*closed*

The payoff of the forward contract for the long position is $S_T - 100$, while the same quantity for the futures is $S_T - 99.95$. Therefore, the payoff of futures is worth .05 more than the payoff of the forward on the maturity date $T = 2$.

However, there is an opportunity cost associated with futures contract. On day one, the marginal account must have \$10.01; the opportunity cost of holding \$10.01 in the marginal

[1] $(x)_+ := \max\{x, 0\}$

account for day one is

$$(10.01)(1 + .002) - 10.01 = (10.01)(.002) = .02002,$$

which is equivalent to $(.02002)(1.002) = .02006004$ *at the end of day two. On day two, we have to keep \$10 plus* $(99.95 - S_1)_+$ *in the marginal account which creates an opportunity cost of* $(.002)(10 + (99.95 - S_1)_+)$*. Therefore, the actual payoff of the futures contract is calculated at maturity as*

$$\begin{aligned} S_T - 99.95 - (1.002) \underbrace{(.002)(10.01)}_{\textit{Opportunity cost of day 1}} &- (.002)\underbrace{(10 + (99.95 - S_1)_+)}_{\textit{Opportunity cost of day 2}} \\ &= S_T - 99.99006004 - (.002)(99.95 - S_1)_+. \end{aligned}$$

The total gain/loss of futures minus forward is

$$\begin{aligned} f(S_1) :=& S_T - 99.99006004 - (.002)(99.95 - S_1)_+ - S_T + 100 \\ =& .00993996 - (.002)(99.95 - S_1)_+ \end{aligned}$$

shown in Figure 1.1.4,

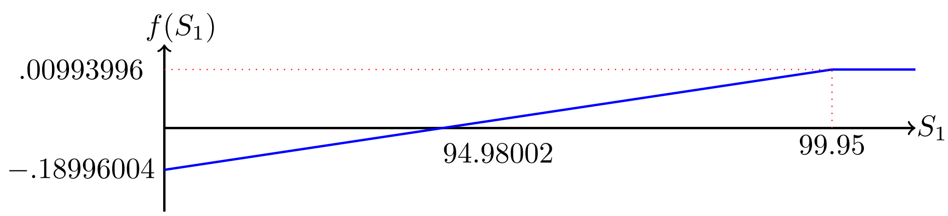


Figure 1.1.4: The difference between the gain of the futures and forward in Example 1.1.1.

Exercise 1.1.1. *Consider a futures contract with maturity* $T = 2$ *days and futures price equal to* \$100*, and a forward contract with the same maturity and forward price of* \$99*; both are written on a risky asset with price* $S_0 = \$99$*. The marginal account for the futures contract needs at least* \$20 *upon entering the contract and should be rebalanced thereafter according to the spot price at the beginning of the day. Given that the risk-free daily compound interest rate is* 0.2%*, for which values of the spot price of the underlying asset,* (S_1, S_2)*, is the forward contract is more interesting than the futures contract for the short position?*

A futures market provides easy access to futures contracts for a variety of products and for different maturities. In addition, it makes termination of a contract possible. A long position in a futures can even out his position by entering a short position of the same contract.

One of the practices in the futures market is *rolling over*. Imagine a trader who needs to have a contract on a product at a maturity T far in the future, such that there is no contract in the futures market with such maturities. However, there is one with maturity $T_1 < T$. The trader can enter the futures with maturity T_1 and later close it when a maturity $T_2 > T_1$ become available. Then, he can continue thereafter until he reaches certain maturity T.

It is well known that in an ideal situation, the spot price S_t of the underlying is less than or equal to the forward/futures price $F_t(T)$ which is referred to as *contango*; see Proposition 1.3.1 for a logical explanation of this phenomena. In reality, this result can no longer be true; especially for futures and forwards on commodities which typically incur storage cost or may not even be storable. A situation in which the futures price of a commodity is less than the spot price is called *normal backwardation* or simply backwardation. Backwardation is more common in commodities with relatively high storage cost; therefore, a low futures price provides an incentive to go into a futures contract. In contrast, when the storage cost is negligible, then contango occurs. We close the discussion on futures and forward with

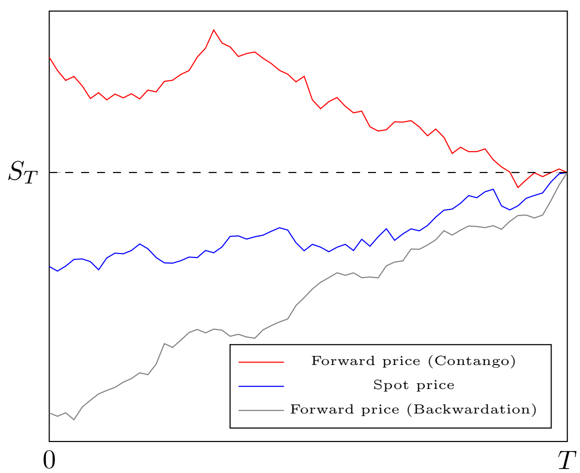


Figure 1.1.5: Contango vs backwardation. Recall that $\lim_{t\to T} F_t(T) = S_T$ implies that the forward price and spot price must converge at maturity.

a model for CCP to determine the marginal account. Notice that the material in 1.1.2 is not restricted to futures market CCP and can be generalized to any market monitored by the CCP.

1.1.2 Eisenberg-Noe model CCP

The CCP can be a useful tool to control and manage systemic risk by setting capital requirements for the entities in a network of liabilities. Because if the marginal account is not set properly, default of one entity can lead directly to a cascade of defaults. Legal action against a defaulted entity is the last thing the CCP wants to do.

Assume that there are N entities in a market and that L_{ij} represents the amount that entity i owes to entity j. Entity i has an equity (cash) balance of $c_i \geqslant 0$ in a marginal account with the CCP. The CCP wants to mediate by billing entity i amount p_i as a *clearing payment* and use it to pay the debt of entity i to other entities. We introduce the liability matrix L consisting of all mutual liabilities L_{ij}. It is obvious that $L_{ii} = 0$, since there is no self-liability.

$$L = \begin{bmatrix} 0 & L_{1,2} & \cdots & L_{1,N-1} & L_{1,N} \\ L_{2,1} & 0 & \cdots & \cdots & L_{2,N} \\ \vdots & & \ddots & & \vdots \\ L_{N-1,1} & \cdots & \cdots & 0 & L_{N-1,N} \\ L_{N,1} & L_{N,2} & \cdots & L_{N,N-1} & 0 \end{bmatrix}$$

We define $L_j := \sum_{i=1}^{N} L_{ji}$ to be the total liability of entity j and define the weights $\pi_{ji} := \frac{L_{ji}}{L_j}$ as the portion of total liability of entity j that is due to entity i. If a clearing payment p_j is made by entity j to the CCP, then entity i receives $\pi_{ji} p_j$ from the CCP. Therefore, after all clearing payments $p_1, ..., p_N$ are made, the entity i receives total of $\sum_{j=1}^{N} \pi_{ji} p_j$. In [11], the authors argue that a clearing payment p_i must not exceed either the total liability L_i or the total amount of cash available by entity i.

First, the Eisenberg-Noe model assume that the payment vector cannot increase the liability, i.e., $p_i \leqslant L_i$ for all i. Secondly, if the equity (cash) of entity i is given by c_i, then after the clearing payment, the balance of the entity i, i.e., $c_i + \sum_{j=1}^{N} \pi_{ji} p_j - p_i$ must remain nonnegative. Therefore, the model suggests that the clearing payment p satisfies

$$p_i = \min \left\{ L_i \ , \ c_i + \sum_{j=1}^{N} \pi_{ji} p_j \right\} \quad \text{for all} \quad i.$$

In other word, the clearing payment vector $p = (p_1, ..., p_N)$ is therefore a *fixed point* of the map $\Phi(p) = (\Phi_1(p), ..., \Phi_N(p_N))$, where the function $\Phi : [0, L_1] \times \cdots \times [0, L_N] \to [0, L_1] \times \cdots \times [0, L_N]$ is given by

$$\Phi_i(p) := \min \left\{ L_i \ , \ c_i + \sum_{j=1}^{N} \pi_{ji} p_j \right\} \qquad \text{for} \quad i = 1, .., N.$$

In general, the clearing payment vector $p = (p_1, ..., p_N)$ is not unique. The following theorem characterizes important properties of the clearing vector.

Theorem 1.1.1 ([11]). *There are two clearing payment vectors p^{max} and p^{min} such that for any clearing payment p, we have $p_i^{min} \leqslant p_i \leqslant p_i^{max}$ for all $i = 1, ..., N$.*

In addition, the value of the equity of each entity remains unaffected by the choice of clearing vector payment; i.e., for all $i = 1, ..., N$ and for all clearing vector payments p,

$$c_i + \sum_{j=1}^{N} \pi_{ji} p_j - p_i = c_i + \sum_{j=1}^{N} \pi_{ji} p_j^{max} - p_i^{max} = c_i + \sum_{j=1}^{N} \pi_{ji} p_j^{min} - p_i^{min}$$

Remark 1.1.1. *The above theorem asserts that if by changing the clearing payment, entity i pays more to other entities, it is going to receive more so the total balance of the equity remains the same.*

If an entity cannot clear all its liability with a payment vector, i.e., $L_i > p_i$, then we say that the entity *has defaulted*. Obviously, the equity of a defaulted entity vanishes. The vanishing of an equity can also happen without default when $c_i + \sum_{j=1}^{N} \pi_{ji} p_j = p_i = L_i$.

A condition for the uniqueness of the clearing payment vector is provided in the original work of Eisenberg-Noe,[11]. However, the condition is restrictive and often hard to check in a massive network of liabilities. On the other hand, the equity of each entity does not depend on the choice of the payment vector. In a massive network, the problem of finding at least one payment vector can also be challenging. One of the ways to find a payment vector is through solving a linear programming problem.

Theorem 1.1.2 ([11])**.** *Let $f : \mathbb{R}^N \to \mathbb{R}$ be a strictly increasing function. Then, the minimizer of the following linear programming problem is a clearing payment vector.*

$$\max f(p) \;\; \textit{subject to} \;\; p \geqslant 0, p \leqslant L \;\; \textit{and} \;\; p \leqslant c + p\Pi, \tag{1.1.1}$$

where $c = (c_1, ..., c_N)$ is the vector of the equities of the entities, $L = (L_1, ..., L_N)$ is the vector of the total liability of the entities, and Π is given by

$$\Pi = \begin{bmatrix} 0 & \pi_{1,2} & \cdots & \pi_{1,N-1} & \pi_{1,N} \\ \pi_{2,1} & 0 & \cdots & \cdots & \pi_{2,N} \\ \vdots & & \ddots & & \vdots \\ \pi_{N-1,1} & \cdots & \cdots & 0 & \pi_{N-1,N} \\ \pi_{N,1} & \pi_{N,2} & \cdots & \pi_{N,N-1} & 0 \end{bmatrix}$$

Exercise 1.1.2. *Consider the liability matrix below by solving the linear programming problem (1.1.1) with $f(p) = \sum_{i=1}^{N} p_i$. Each row/column is an entity.*

$$L = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 2 & 0 & 0 & 1 \\ 0 & 5 & 0 & 0 \end{bmatrix}$$

The initial equities are given by $c_1 = 1$, $c_2 = 2$, $c_3 = 0$, and $c_4 = 0$.

a) *By using a linear programming package such as* `linprog` *in* `MatLab`*, find a clearing payment vector.*

b) *Mark the entities that default after applying the clearing payment vector found in part (a).*

c) *Increase the value of the equity of the defaulted entities just as much as they do not default anymore.*

1.1.3 Vanilla call and put options

A call option gives the holder the right but not the obligation to buy a certain asset at a specified time in the future at a predetermined price. The specified time is called the *maturityoption!maturity* and often is denoted by T, and the predetermined price is called the *strike price* and is denoted by K. Therefore, a call option protects its owner against any increase in the price of the *underlying asset* above the strike price at maturity. The asset price at time t is denoted by S_t and at the maturity by S_T. Call options are available in the specialized options markets at a price that depends, among other factors, on time t, T, K, and spot price at current time $S_t = S$. To simplify, we denote the price of a call option by $C(T, K, S, t)$[2] to emphasize the main factors, i.e., t, T, K, and spot price at current time S. Another type of vanilla option, the put option, protects its owner against any increase in the price of the underlying asset above the strike price at the maturity; i.e., it promises the seller of the underlying asset at least the strike price at maturity. The price of put option is denoted by $P(T, K, S, t)$, or simply P when appropriate.

The payoff of an option is the owner's gain in a dollar amount. For instance, the payoff of a call option is $(S_T - K)_+$. This is because, when the market price at maturity is S_T and the strike price is K, the holder of the option is buying the underlying asset at lower price K and gains $S_T - K$, provided $S_T > K$. Otherwise, when $S_T \leqslant K$, the holder does not exercise the option and buys the asset from the market directly. Similarly, the payoff of a put option is $(K - S_T)_+$.

Similar to futures, options are also traded in specialized markets. You can see *option chain* for Tesla in Figure 1.1.7. The columns"bid" and "ask" indicate the best buy and sell prices in the outstanding orders, and column"Open Int" (open interest) shows the total volume of outstanding orders.

When the spot price S_t of the underlying asset is greater than K, we say that the call options are *in-the-money* and the puts options are *outof-the-money* . Otherwise, when $S_t < K$, the put options are *in-the-money* and the call options are *out-of-the-money*. If the strike price K is (approximately) the same as spot price S_t, we call the option *at-the-money* (or ATM).

Far in-the-money call or put options are behave like forward contracts but with a wrong forward price! Similarly, far out-of-the-money call or put options have negligible worth.

[2]We will see later that C only depends on $T - t$ in many models.

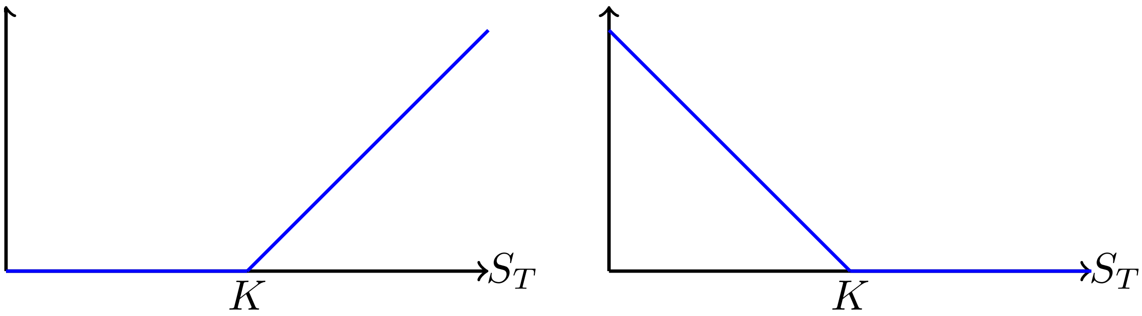


Figure 1.1.6: Left: the payoff $(S_T - K)_+$ of a call option with strike K. Right: the payoff $(K - S_T)_+$ of a put option with strike K.

Tesla Motors Inc (NASDAQ:TSLA) Add to portfolio

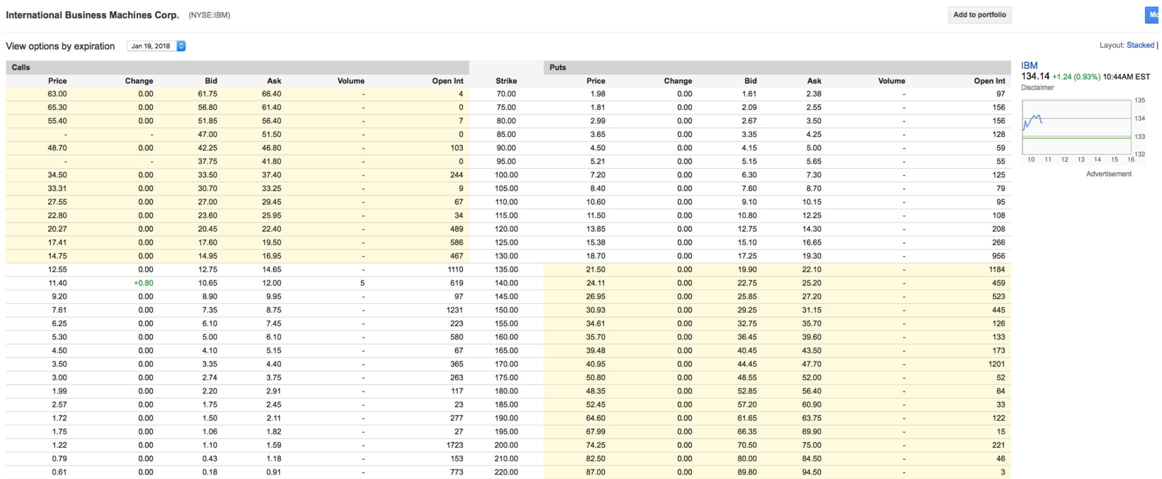
View options by expiration Jan 15, 2016

Calls							Puts					
Price	Change	Bid	Ask	Volume	Open Int	Strike	Price	Change	Bid	Ask	Volume	Open Int
209.40	0.00	-	-	-	5	12.50	0.01	0.00	-	0.01	-	2061
242.68	0.00	194.15	197.70	-	9	15.00	0.05	0.00	-	0.01	-	1762
196.00	0.00	191.65	195.15	-	1	17.50	0.03	0.00	0.02	0.10	-	215
210.68	0.00	-	-	-	16	20.00	0.01	0.00	-	-	-	12525
203.70	0.00	186.60	190.55	-	0	22.50	0.05	0.00	-	0.14	-	860
202.00	0.00	184.15	187.65	-	1	25.00	0.01	0.00	-	-	-	1941
196.00	0.00	-	-	-	91	30.00	0.01	0.00	-	-	-	1641

Figure 1.1.7: Call and put option quotes on Tesla stocks on January 11, 2016. The first column is the price of the underlying asset (NASDAQ:TSLA). The bid price is the price at which trades are willing to buy the options and the ask price is the price at which trades are willing to sell. The spot price at the time was \$206.11. Source: Google Finance.

The holder of an option is called a *long position*, and the issuer of the option is called a *short position*. While the holder has the privilege of exercising the option when profitable, the issuer has the obligation to pay the holder the amount of payoff upon exercise.

A *European option* is an option whose payoff is a function $g(S_T)$ of the asset price at maturity S_T. The function $g : \mathbb{R}_+ \to \mathbb{R}$ is called a European payoff function. Call and put options are particularly important because any piecewise linear continuous European payoff can be written as a linear combination (possibly infinite!) of call option payoffs with possibly different strikes but the same maturity and a constant cash amount; or, equivalently, a linear combination of call option payoffs and a put option payoff. Therefore,



International Business Machines Corp. (NYSE:IBM)

Add to portfolio

View options by expiration Jan 19, 2018

Layout: Stacked |

Calls							Puts					
Price	Change	Bid	Ask	Volume	Open Int	Strike	Price	Change	Bid	Ask	Volume	Open Int
63.00	0.00	61.75	66.40	-	4	70.00	1.98	0.00	1.61	2.38	-	97
65.30	0.00	56.80	61.40	-	0	75.00	1.81	0.00	2.09	2.55	-	156
55.40	0.00	51.85	56.40	-	7	80.00	2.99	0.00	2.67	3.50	-	156
-	-	47.00	51.50	-	0	85.00	3.65	0.00	3.35	4.25	-	128
48.70	0.00	42.25	46.80	-	103	90.00	4.50	0.00	4.15	5.00	-	59
-	-	37.75	41.80	-	0	95.00	5.21	0.00	5.15	5.65	-	55
34.50	0.00	33.50	37.40	-	244	100.00	7.20	0.00	6.30	7.30	-	125
33.31	0.00	30.70	33.25	-	9	105.00	8.40	0.00	7.60	8.70	-	79
27.55	0.00	27.00	29.45	-	67	110.00	10.60	0.00	9.10	10.15	-	95
22.80	0.00	23.60	25.95	-	34	115.00	11.50	0.00	10.80	12.25	-	108
20.27	0.00	20.45	22.40	-	489	120.00	13.85	0.00	12.75	14.30	-	208
17.41	0.00	17.60	19.50	-	586	125.00	15.38	0.00	15.10	16.65	-	266
14.75	0.00	14.95	16.95	-	467	130.00	18.70	0.00	17.25	19.30	-	956
12.55	0.00	12.75	14.65	-	1110	135.00	21.50	0.00	19.90	22.10	-	1184
11.40	+0.80	10.65	12.00	5	619	140.00	24.11	0.00	22.75	25.20	-	459
9.20	0.00	8.90	9.95	-	97	145.00	26.95	0.00	25.85	27.20	-	523
7.61	0.00	7.35	8.75	-	1231	150.00	30.93	0.00	29.25	31.15	-	445
6.25	0.00	6.10	7.45	-	223	155.00	34.61	0.00	32.75	35.70	-	126
5.30	0.00	5.00	6.10	-	580	160.00	35.70	0.00	36.45	39.60	-	133
4.50	0.00	4.10	5.15	-	67	165.00	39.48	0.00	40.45	43.50	-	173
3.50	0.00	3.35	4.40	-	365	170.00	40.95	0.00	44.45	47.70	-	1201
3.00	0.00	2.74	3.75	-	263	175.00	50.80	0.00	48.55	52.00	-	52
1.99	0.00	2.20	2.91	-	117	180.00	48.35	0.00	52.85	56.40	-	64
2.57	0.00	1.75	2.45	-	23	185.00	52.45	0.00	57.20	60.90	-	33
1.72	0.00	1.50	2.11	-	277	190.00	64.60	0.00	61.65	63.75	-	122
1.75	0.00	1.06	1.82	-	27	195.00	67.99	0.00	66.35	69.90	-	15
1.22	0.00	1.10	1.59	-	1723	200.00	74.25	0.00	70.50	75.00	-	221
0.79	0.00	0.43	1.18	-	153	210.00	82.50	0.00	80.00	84.50	-	46
0.61	0.00	0.18	0.91	-	773	220.00	87.00	0.00	89.80	94.50	-	3

IBM
134.14 +1.24 (0.93%) 10:44AM EST
Disclaimer

Advertisement

Figure 1.1.8: Call and put option quotes on IBM stocks on January 13, 2016. In-the-money options are highlighted. Source: Google Finance.

the price of payoff $g(S_T) = a_0 + \sum_i a_i(S_T - K_i)_+$ is given by

$$B_t(T)a_0 + \sum_i a_i C(T, K_i, S_t, t).$$

In the above, a_0 is the constant cash amount, and for each i, $(S_T - K_i)_+$ is the payoff of a call option with the strike price K_i and the maturity T.

Remark 1.1.2. *The underlying asset S_T at time T is a call option with strike $K = 0$.*

Example 1.1.2. *Put-call parity suggests that the payoff $(K - S_T)_+$ of a put option can be written as the summation of payoffs of a long position in K amount of cash, a long position in a call option with payoff $(S_T - K)_+$, and a short position in an underlying asset. By Remark 1.1.2, a long position in the underlying asset is a call option with strike 0. See Figure 1.1.9.*

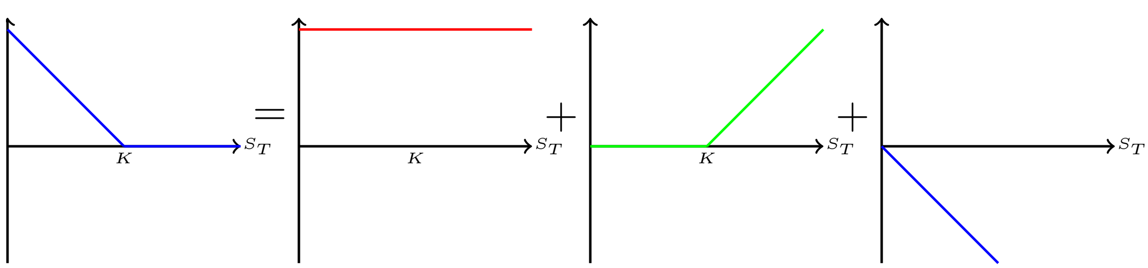


Figure 1.1.9: $(K - S_T)_+ = K + (S_T - K)_+ - S_T$.

Example 1.1.3. *An option that promises the payoff* $g(S_T) := |S_T - K|$, *as shown in Figure 1.1.10, is made of a long position in a call option with strike* K, *and long position in a put with strike* K, *both with the same maturity. Equivalently, this payoff can be written as* K *amount of cash, a short position in underlying, and two long positions in a call option with strike* K, *all with the same maturity.*

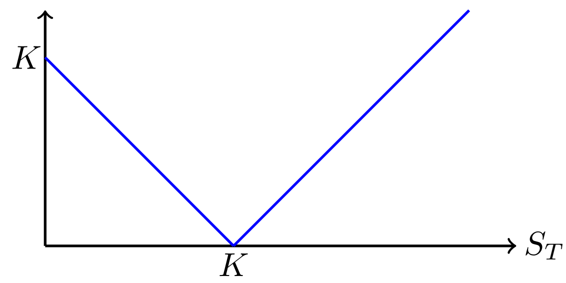


Figure 1.1.10: Payoff $g(S_T) = |S_T - K|$ from Example 1.1.3

Example 1.1.4. *A put option with payoff* $(K - S_T)_+$ *can be written as* K *amount of cash, a short position in a call option with strike* 0, *and a long position on a call option with strike* K.

Exercise 1.1.3. *Consider the payoff* $g(S_T)$ *shown in Figure 1.1.6.*

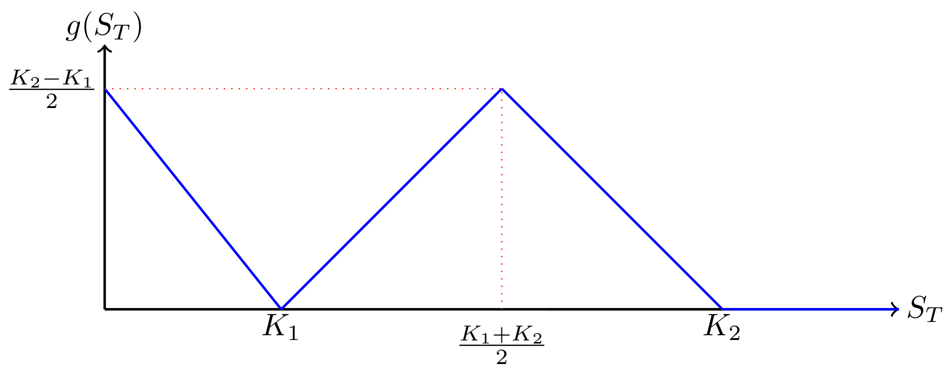


Figure 1.1.11: Payoff for Exercise 1.1.3

a) *Write this payoff as a linear combination of the payoffs of some call options and a put option with different strikes and the same maturity.*

b) *Repeat part (a) with call options and cash. (No put option is allowed.)*

1.1.4 American options

European options can only be exercised at a maturity for the payoff $g(S_T)$. An American option gives its owner the right but not the obligation to exercise at any date before or at maturity. Therefore, at the time of exercise $\tau \in [0, T]$, an American call option has an exercise value equal to $(S_\tau - K)_+$. We use the notation $C_{\text{Am}}(T, K, S, t)$ and $P_{\text{Am}}(T, K, S, t)$ to denote the price of an American call and an American put, respectively.

The exercise time τ is not necessarily deterministic. More precisely, it can be a random time that depends upon the occurrence of certain events in the market. An optimal exercise time can be found among those of threshold type; the option is exercised before maturity the first time the market price of option becomes equal to the exercise value, whereas the option usually has a higher market value than the exercise value. Notice that both of these quantities behave randomly over time.

1.1.5 Bond and forward rate agreements

A zero-coupon bond (or simply zero bond) is a *fixed-income* security that promises a fixed amount of cash in a specified currency at a certain time in the future, e.g., \$100 on January 30. The promised cash is called the *principle*, *face value* or *bond!face value* and the time of delivery is called the *maturity*. All bonds are traded in specialized markets at a price often lower than the principle[3].

For simplicity, throughout this book, a zero bond means a zero bond with principle of \$1, unless the principle is specified; for example, a zero bond with principle of \$10 is ten zero bond s. At a time t, we denote the price of a zero bond maturing at T by $B_t(T)$.

We can use the price of a zero bond to calculate the *present value* of a future payment or cashflow. For example, if an amount of \$$x$ at time T is worth $\frac{x}{B_t(T)}$ at an earlier time t. This is because, if we invest \$$\frac{x}{B_t(T)}$ in a zero bond with maturity T, at the maturity we receive a dollar amount of $\frac{x}{B_t(T)} B_t(T) = x$.

The price of the zero bond is the main indicator of the interest rate. While the term "interest rate" is used frequently in news and daily conversations, the precise definition of the interest rate depends on the time horizon and the frequency of compounding. An interest rate compounded yearly is simply related to the zero bond price by $1+R^{(\text{yr})} = \frac{1}{B_0(1)}$, while for an interest rate compounded monthly, we have $\left(1 + \frac{R^{(\text{mo})}}{12}\right)^{12} = \frac{1}{B_0(1)}$. Therefore,

$$1 + R^{(\text{yr})} = \left(1 + \frac{R^{(\text{mo})}}{12}\right)^{12}.$$

Generally, an n-times compounded interest rate during the time interval $[t, T]$, denoted by $R_t^{(n)}(T)$, satisfies $\left(1 + \frac{R_t^{(n)}(T)}{n}\right)^n = \frac{1}{B_t(T)}$. When the frequency of compounding n

[3] There have been instances when this has not held, e.g., the financial crisis of 2007.

approaches infinity, we obtain

$$B_t(T) = e^{-R_t^{(\infty)}(T)}.$$

This motivates the definition of the zero-bond *yield*. The yield (or yield curve) $R_t(T)$ of the zero bond $B_t(T)$ is defined by

$$B_t(T) = e^{-(T-t)R_t(T)} \quad \text{or} \quad R_t(T) := -\frac{1}{T-t} \ln B_t(T). \tag{1.1.2}$$

The yield is a bivariate function $R : (t,T) \in D \to \mathbb{R}_{\geqslant 0}$ where D is given by $\{(t,T) : T > 0 \text{ and } t < T\}$. Yield is sometimes referred to as the term structure of the interest rate since both variables t and T are time.

If the yield curve is a constant, i.e., $R_t(T) = r$ for all $(t,T) \in D$, then, $B_t(T) = e^{-r(T-t)}$. In this case, r is called the *continuously compounded*, *instantaneous*, *spot*, or *short* rate. However, the short rate does not need to be constant. A time-dependent short interest rate is a function $r : [0,T] \to \mathbb{R}_+$ such that for any $T > 0$, and $t \in [0,T]$ we have $B_t(T) = e^{-\int_t^T r_s ds}$; or equivalently, short rate can be defined as

$$r_t := -\frac{\partial \ln B_t(T)}{\partial T}\Big|_{T=t} = \frac{\partial \ln B_t(T)}{\partial t}.$$

The short rate r or r_s is an abstract concept; it exists because it is easier to model the short rate than the yield curve. In practice, the interest rate is usually given by the yield curve.

Besides zero bonds, there are other bonds that pay *coupons* on a regular basis, for example a bond that pays the principle of \$100 in 12 months and \$20 every quarter. A *coupon-carrying* bond, or simply, a coupon bond, can often be described as a linear combination of zero bonds; i.e., a bond with coupon payments of \$ c_i at date T_i with $T_1 < ... < T_{n-1}$ and principle payment P at maturity $T_n = T$ is the same as a portfolio of zero bonds with principle c_i and maturity T_i for $i = 1, ..., n$ and is worth

$$\sum_{i=1}^{n-1} c_i B_t(T_i) + P B_t(T_n) = \sum_{i=1}^{n-1} c_i e^{-(T_i - t)R_t(T_i)} + P e^{-(T_n - t)R_t(T_n)}.$$

Therefore, zero bonds are the building blocks of all bonds, and the yield curve is the main factor in determining the price of all bonds.

Example 1.1.5. *A risk-free 1-year zero bond with \$20 principle is priced* $B_0(1) = \$19$ *and a risk-free 2-year zero bond with \$20 principle is priced* $B_0(2) = \$17$*. Then, the yield* $R_1(2)$ *is given by*

$$R_1(2) = \ln B_0(1) - \ln B_0(2) \ln B_0(1) = \ln 19 - \ln 17 \approx 0.1112,$$

and the price of a risk-free zero bond that start in one year and ends at in two years with *principle $20 is given by*

$$B_1(2) = 100e^{-R_1(2)} = \frac{1700}{19} \approx 89.47.$$

The price of a bond that pays a $30 coupon at the end of the current year and $100 as the principle in two years equals to

$$30\frac{B_0(1)}{100} + 100\frac{B_0(2)}{100} = 5.9 + 17 = 22.9.$$

Exercise 1.1.4. *a) If a risk-free 1-year zero bond with $100 principle is priced* $B_0(1) = \$96$ *and a risk-free 2-year zero bond with $100 principle is priced* $B_0(2) = \$92$*, find the price of a risk-free zero bond* $B_1(2)$ *and yield curve* $R_1(2)$*.*

b) What is the price of a bond that pays a $30 coupon at the end of the current year and $100 as the principle in two years?

In the above discussion, we implicitly assumed that the issuer of the bond is not subject to default on payment of coupons or principle. This type of bonds are called *sovereign* bonds and are often issued by the Federal Reserve or central bank of a given country in that country's own currency. For example, sovereign bonds in the United States are T-bills, T-notes, and T-bonds. T-bills are bonds that have a maturity of less than a year, T-notes have a maturity of more than a year up to ten years, and T-bonds have maturity more than ten years. Bonds issued by other entities or governments in a foreign currency are usually called *corporate* bonds. The word "corporate" emphasizes the default risk of the issuer on the payments. In addition, sovereign bonds in a foreign currency are subject to the market risk that is caused by fluctuating exchange rates in the foreign exchange market[4]. Therefore, what is considered a risk-free bond in the United States is not risk-free in the European Union.

The zero bond price $B_t(T)$ can directly be used to discount a payment or a cashflow at time T without appealing to a specific short rate model. For example, a cashflow of $10 at time $T = 1$ is worth \$ $10B_0(1)$ now.

Similar to the yield curve, the *forward rate* $F_t(T)$[5] of a zero bond is defined by

$$B_t(T) = e^{-\int_t^T F_t(u)du} \quad \text{or} \quad F_t(T) := -\frac{\partial \ln B_t(T)}{\partial T}.$$

The forward rate reflects the current perception among traders about the future fluctuations of the interest rates. More precisely, at time t, we foresee the continuously compounded

[4]The foreign exchange market is a decentralized over-the-counter market where traders across the world use to trade currencies.

[5]The notation for forward rate is the same as the notation for futures price or forward price in Section 1.1.1.

interest rates for N time intervals $[t_0, t_1]$, $[t_1, t_2]$, ..., $[t_{N-1}, T]$ in the future as $F_t(t)$, $F_t(t_1)$, ..., $F_t(t_{N-1})$, respectively. Here, for $n = 0, ..., N$, $t_n = t + n\delta$ and $\delta = \frac{T-t}{N}$. Then, \$1 at time T is worth $e^{-F_t(t_{N-1})\delta)}$ at time t_{N-1}, $e^{-(F_t(t_{N-1})+F_t(t_{N-2}))\delta}$ at time t_{N-2}, $e^{-(F_t(t_{N-1})+...+F_t(t_n))\delta}$ at time t_n, and $e^{-(F_t(t_{N-1})+...+F_t(t_1)+F_t(t))\delta}$ at time t. As n goes to infinity, the value $B_t(T)$ of the zero bond converges to

$$\lim_{n\to\infty} e^{-\sum_{n=0}^{N-1} F_t(t_n)\delta} = e^{-\int_t^T F_t(u)du}.$$

The forward rates are related to so-called forward rate agreements. A forward rate agreement is a contract between two parties both committed to exchanging a specific loan (a zero bond with a specified principle and a specific maturity) in a future time (called the delivery date) with a specific interest rate. We can denote the agreed rate by $f(t_0, t, T)$ where t_0 is the current time, t is the delivery date, and T is the maturity of the bond. As always, we take principle to be \$1. Then, the price of the underlying bond at time t should equal $B_t(T) = \frac{B_{t_0}(T)}{B_{t_0}(t)}$. This is because, if we invest \$ x in $B_{t_0}(t)$ at time t_0, we have $\frac{x}{B_{t_0}(t)}$ at time t. Then, at time t, we reinvest this amount in $B_t(T)$. At time T, we have $\frac{x}{B_{t_0}(t)B_t(T)}$. Alternatively, if we invest in $B_{t_0}(T)$ from the beginning, we obtain $\frac{x}{B_{t_0}(T)}$, which must be the same as the value of the two-step investment described above[6]. Therefore, the fair forward rate in a forward rate agreement must satisfy

$$f_{t_0}(t, T) = -\frac{\ln B_{t_0}(T) - \ln B_{t_0}(t)}{T - t}.$$

If we let $T \downarrow t$, we obtain $\lim_{T\downarrow t} f_{t_0}(t, T) = F_{t_0}(t)$. In terms of yield, we have

$$f_{t_0}(t, T) = \frac{(T - t_0)R_{t_0}(T) - (t - t_0)R_{t_0}(t)}{T - t}.$$

$F_{t_0}(t)$ is the *instantaneous* forward rate. However, $f_{t_0}(t, T)$ is the forward rate at time t_0 for time interval $[t, T]$ and is related to the instantaneous forward rate by

$$f_{t_0}(t, T) = \int_t^T F_{t_0}(u)du.$$

Unlike the forward rate and short rate r_t, yield curve $R_t(T)$ is accessible through market data. For example, LIBOR[7] is the rate at which banks worldwide agree to lend to each other and is considered more or less a benchmark interest rate for international trade. Or, the United States treasury yield curve is considered a risk-free rate for domestic transactions within the United States. The quotes of yield curve $R_t(T)$ for LIBOR and the United

[6]A more rigorous argument is provided in Section 1.3 Example 1.3.2

[7]London Interbank Offered Rate

States treasuries at different maturities are given in Tables 1.2 and 1.3, respectively.

Time-to-maturity $T-t$	Current date t				
	02-19-2016	02-18-2016	02-17-2016	02-16-2016	02-15-2016
USD LIBOR - overnight	0.37090 %	0.37140 %	0.37000%	0.37100 %	-
USD LIBOR - 1 week	0.39305 %	0.39200 %	0.39160%	0.39050 %	0.39340 %
USD LIBOR - 1 month	0.43350 %	0.43200 %	0.43005 %	0.42950 %	0.42925 %
USD LIBOR - 2 months	0.51720 %	0.51895 %	0.51675 %	0.51605 %	0.51580 %
USD LIBOR - 3 months	0.61820 %	0.61820 %	0.61940 %	0.61820 %	0.61820 %
USD LIBOR - 6 months	0.86790 %	0.87040 %	0.86660 %	0.86585 %	0.86360 %
USD LIBOR - 12 months	1.13975 %	1.14200 %	1.13465 %	1.13215 %	1.12825 %

Table 1.2: LIBOR yield curve for US dollars. Source: www.global-rates.com. The the date format in the table is DD-MM-YYYY, contrary to the date format MM/DD/YYYY in the United States.

Date t	Time-to-maturity $T-t$							
	1 Mo	3 Mo	6 Mo	1 Yr	5 Yr	10 Yr	20 Yr	30 Yr
02/22/16	0.28%	0.33%	0.46%	0.55%	1.25%	1.77%	2.18%	2.62%
02/23/16	0.28%	0.32%	0.47%	0.55%	1.23%	1.74%	2.16%	2.60%
02/24/16	0.28%	0.33%	0.46%	0.55%	1.21%	1.75%	2.16%	2.61%

Table 1.3: Treasury yield curve for US dollar. Source: https://www.treasury.gov.

The forward rates $f_{t_0}(t,T)$ can also be obtained from the data on forward rate agreements, but this data is not publicly available.

We should clarify that there is a slight difference between the yield curve defined by (1.1.2) and the yield curve data in Tables 1.2 and 1.3. The recorded data on the yield curve comes from

$$B_t(T) = (1+\hat{R_t(T)})^{T-t}, \quad \text{or} \quad \hat{R}_t(T) := \left(\frac{1}{B_t(T)}\right)^{\frac{1}{T-t}} - 1, \tag{1.1.3}$$

where the time-to-maturity $T-t$ measured in years. Therefore,

$$\hat{R}_t(T) = \left(\exp\left(R_t(T)(T-t)\right)\right)^{\frac{1}{T-t}} - 1 = \exp\left(R_t(T)\right) - 1 \approx R_t(T),$$

when $R_t(T)$ is small. For example, in Table 1.3, the yield of a zero bond that expires in one month is given by $0.28\% = .0028$ and the price of such a bond is equal to $(1.0028)^{-1/12} = 0.9996$.

It is also important to know that an interpolation method is used to generate some of the yield curve data in Tables 1.2 and 1.3. This is because a bond that expires exactly in one month, three months, six months, or some other time does not necessarily always exist. A bond that expires in a month will be a three-week maturity bond after a week. Therefore, after calculating the yield for the available maturities, an interpolation gives us the interpolated yields of the standard maturities in the yield charts. For example, in Table 1.3, The Treasury Department uses the cubic Hermite spline method to generate daily yield curve quotes.

Remark 1.1.3. *While the market data on the yield of bonds with different maturity is provided in the discrete-time sense, i.e., (1.1.3), the task of modeling a yield curve in financial mathematics is often performed in continuous-time. Therefore, it is important to learn both frameworks and the relation between them.*

Sensitivity analysis of the bond price

We measure the sensitivity of the zero bond price with respect to changes in the yield or errors in the estimation of the yields by $\frac{dB_t(T)}{dR_t(T)} = -(T-t)B_t(T)$. As expected, the sensitivity is negative, which means that the increase (decrease) in yield is detrimental (beneficial) to the bond price. It is also proportional to the time-to-maturity of the bond; i.e., the *duration* of the zero bond is equal to $-\frac{dB_t(T)/dR_t(T)}{B_t(T)}$. For a coupon bond, the sensitivity is measured after defining the yield of the bond; the yield of a coupon bond with coupons payments of \$ c_i at date T_i with $T_1 < ... < T_{n-1}$ and principle payment $P = c_n$ at maturity $T_n = T$ is a number $\hat{y}$ such that $p(\hat{y})$ equals the price of the bond, and $p(y)$ is the function defined below:

$$p(y) := \sum_{i=1}^{n} c_i e^{-(T_i-t)y}.$$

Therefore, the yield of a bond is the number y that satisfies

$$\sum_{i=1}^{n} c_i e^{-(T_i-t)\hat{y}} = \sum_{i=1}^{n} c_i B_t(T_i).$$

The function $p(y)$ is a strictly decreasing function with $p(-\infty) = \infty$, $p(0) = \sum_{i=1}^{n-1} c_i + P$, and $p(\infty) = 0$. Therefore, for a bond with a positive price, the yield of the bond exists as a real number. In addition, if the price of the bond is in the range $(0, p(0))$, the yield of the bond is a positive number.

Therefore, the yield $\hat{y}$ depends on all parameters t, c_i, T_i, and $R_t(T_i)$, for $i = 1, ..., n$. Motivated from the zero bond, the *duration* the coupon bond is given by

$$D := -\frac{dp/dy}{p}(\hat{y}) = \sum_{i=1}^{n}(T_i - t)\frac{c_i e^{-(T_i-t)\hat{y}}}{p(\hat{y})},$$

a naturally weighted average of the duration of payments, with weights $\frac{c_i e^{-(T_i-t)\hat{y}}}{p(\hat{y})}$, for $i = 1, ..., n$.

The *convexity* of a zero bond is defined by using the second derivative $C := \frac{d^2 B_t(T)/dR_t(T)^2}{B_t(T)} = (T-t)^2$, and for a coupon bond is expressed as

$$C := \frac{d^2p/dy^2}{p}(\hat{y}) = \sum_{i=1}^{n} (T_i - t)^2 \frac{c_i e^{-(T_i-t)\hat{y}}}{p(\hat{y})}.$$

Recall that a function f is convex if and only if $f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda) f(y)$ for all $\lambda \in (0,1)$. If the function is twice differentiable, convexity is equivalent to $f'' \geqslant 0$. For more details on convexity, see Section A.1.

While the duration indicates a negative relation between changes in yield $\hat{y}$ and price of the bond $p(\hat{y})$, the convexity tells more about the magnitude of this change. For example, considering two bonds with the same duration, the one with higher convexity is more sensitive to changes in the yield. See Figure 1.1.12.

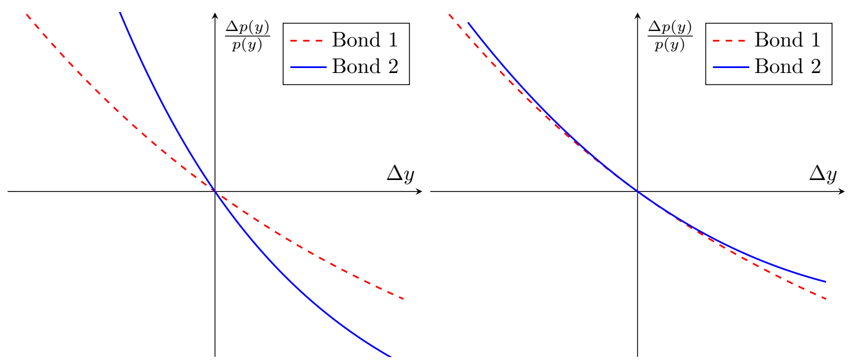


Figure 1.1.12: The relative price change of the bond $\frac{\Delta p(y)}{p(\hat{y})} = \frac{p(y)-p(\hat{y})}{p(\hat{y})}$ with change in the yield $\Delta y = y - \hat{y}$. Left: Bond 1 is longer in duration and therefore less sensitive than bond 2. Right: Both bonds are the same duration, but bond 1 is less convex and more sensitive.

Remark 1.1.4. *While in practice bonds are not considered derivatives, they are bets on the interest rate and therefore can mathematically be considered as derivatives. Forward interest agreements are derivatives on bond.*

1.1.6 Credit derivatives

Financial instruments are issued by financial companies such as banks. There is always a risk that the issuer will go bankrupt or at least default on some payments and be unable to

Rating	Moody's	S&P	Fitch
Prime	Aaa	AAA	AAA
Subprime	Ba1 and lower	BB+ and lower	BB+ and lower
Default	C	D	D

Table 1.4: A brief table of rating by Moody, S&P and Fitch.

meet its obligation. The same situation holds when a debt such as a mortgage is issued. In such cases, the beneficiary of the issued security is exposed to credit risk. Therefore, it is important to know about the creditworthiness of individual and corporate loan applicants. There are three major credit rating companies for corporations and other institutions, including governments: Moody's, Standard and Poor's (S&P), and Fitch. The rates are usually shown by letters similar to letter grading of a university course. Regardless of the notation used, the highest level of creditworthiness is called *prime rating* and the lowest is given to a *defaulted* entity. Lower-half rates are usually referred to *subprime* which indicates higher exposure to credit risk. See Table 1.4 for a sample of a ratings table and its notations. To cover credit risk, financial institutes issue credit derivatives. There are two well-known credit derivatives in the market: credit default swap (CDS) and collateralized debt obligation (CDO). Both derivatives are written on defaultable loans (such as bonds). For simplicity, we only consider CDOs and CDSs on defaultable zero bonds, i.e., the simplest of all defaultable assets. First, we introduce defaultable zero bonds and explain how the yield of a defaultable zero bond is calculated.

Defaultable zero bond

Consider a zero bond with a face value \$1. If the bond is sovereign with the yield $R_t(T)$, then the value of the bond is

$$B_t(T) = e^{-R_t(T)(T-t)}.$$

We assume that in case of default, the value of the bond vanishes instantly. The default of a company and the time of default are random. If we denote the (random) time of the default by τ and assume that the default has not occurred yet, i.e., $\tau > t$, we define the *survival rate* of the defaultable bond by $\lambda_t(T)$, which satisfies

$$\mathbb{P}(\tau > T \mid \tau > t) = 1 - \mathbb{P}(\tau \leqslant T \mid \tau > t) = 1 - e^{-\lambda_t(T)(T-t)}.$$

In the above $\mathbb{P}(\cdot \mid \tau > t)$ represents the probability measure (function) conditional on $\tau > t$, i.e., the default has not occurred until time t. Notice that $\lambda_t(T)$ always exists as a nonnegative number, or $+\infty$, and is given by

$$\lambda_t(T) = -\frac{1}{T-t} \ln \mathbb{P}(\tau \leqslant T \mid \tau > t).$$

If $\lambda_t(T) = 0$, the bond is sovereign and never defaults. Otherwise, if $\lambda_t(T) = +\infty$, the probability of default is 1 and the default is a certain event. The payoff of a defaultable bond is given by the indicator random variable below:

$$1_{\{\tau > T\}} := \begin{cases} 1 & \text{when } \tau > T, \\ 0 & \text{when } \tau \leqslant T. \end{cases}$$

A common formalism in pricing financial securities with a random payoff is to take expectation from the discounted payoff. More precisely, the value of the defaultable bond is given by the expected value of the discounted payoff, i.e.,

$$B_t^\lambda(T) := \mathbb{E}\left[B_t(T)1_{\{\tau>T\}} \mid \tau > t\right] = B_t(T)\mathbb{P}(\tau > T \mid \tau > t) = e^{-R_t(T)(T-t)}\Big(1-e^{-\lambda_t(T)(T-t)}\Big).$$

The *risk-adjusted yield* of a defaultable bond is defined by the value $R_t^\lambda(T)$ such that

$$e^{-R_t^\lambda(T)(T-t)} = B_t^\lambda(T) = e^{-R_t(T)(T-t)}\Big(1 - e^{-\lambda_t(T)(T-t)}\Big).$$

Equivalently,

$$R_t^\lambda(T) = R_t(T) - \frac{1}{T-t} \ln\Big(1 - e^{-\lambda_t(T)(T-t)}\Big)$$

For a defaultable bond, we always have $R_t^\lambda(T) > R_t(T)$. Notice that when $\lambda_t(T) \uparrow \infty$, $R_t^\lambda(T) \downarrow R_t(T)$, and when $\lambda_t(T) \downarrow 0$, $R_t^\lambda(T) \uparrow \infty$. .

The higher the probability of default, the higher the adjusted yield of the bond.

Exercise 1.1.5. *Consider a defaultable zero bond with $T = 1$, face value \$1, and survival rate 0.5. If the current risk-free yield for maturity $T = 1$ is 0.2, find the adjusted yield of this bond.*

Credit default swap (CDS)

A CDS is a swap that protects the holder of a defaultable asset against default before a certain maturity time T by recovering a percentage of the nominal value specified in the contract in case the default happens before maturity. Usually, some percentage of the loss can be covered by collateral or other assets of the defaulted party, i.e., a recovery rate denoted by R. The recovery rate R is normally a percentage of the face value of the defaultable bond and is evaluated prior to the time of issue. Therefore, the CDS covers $1 - R$ percent of the value of the asset at the time of default. In return, the holder makes regular, constant premium payments κ until the time of default or maturity, whichever happens first. The maturity of a CDS is often the same as the maturity of the defaultable asset, if there is any. For example, a CDS on a bond with maturity T also expires at time T.

To find out the fair premium payments κ for the CDS, let's denote the time of the default by τ and the face value of the defaultable bond by P. If the default happens at $\tau \leqslant T$, the CDS pays the holder an amount of $(1-R)P$. The present value of this amount is obtained by discounting it with a sovereign zero bond, i.e., $(1-R)PB_0(T \wedge \tau)$. If the issuer of the bond defaults after time T, the CDS does not pay any amount. Thus, the payment of the CDS is a random variable expressed as

$$(1-R)PB_0(T \wedge \tau)1_{\{\tau \leqslant T\}}.$$

The holder of the CDS makes regular payments of amount κ at points $0 = T_0 < T_1 < \cdots < T_n$ in time with $T_n < T$. Then, the present value of payment of amount κ, paid at time T_i is

$$\kappa 1_{\{T_i < \tau\}} B_0(T_i).$$

The total number of premium payments is $N := \max\{i+1 : T_i < \tau \wedge T\}$, which is also a random variable with values $1, ..., n+1$. Therefore, the present value of all premium payments is given by

$$\kappa \sum_{i=1}^{N} B_0(T_i).$$

The discounted payoff of the CDS starting at time t is given by

$$(1-R)PB_0(T \wedge \tau)1_{\{\tau \leqslant T\}} - \kappa \sum_{i=1}^{N} B_0(T_i) \tag{1.1.4}$$

The only source of randomness in the above payoff is the default time, i.e., τ. This makes the terms $B_0(T \wedge \tau)1_{\{\tau \leqslant T\}}$ and $\sum_{i=1}^{N} B_0(T_i)$ random variables. Notice that, although each individual term in the summation $\sum_{i=1}^{N} B_0(T_i)$ is not random, the number of terms N in the summation is.

Because of the presence of randomness, we follow the formalism that evaluates the price of an asset with random payoff by taking the expected value of the discounted payoff. In case of a CDS, the price is known to be zero; either party in a CDS does not pay or receive any amount by entering a CDS contract. Therefore, the premium payments κ should be such that the expected value of the payoff (1.1.4) vanishes. To do so, first we need to know the probability distribution of the time of default. The task of finding the distribution of default can be performed through modeling the survival rate, which is defined in Section 1.1.6. Provided that the distribution of default is known, κ can be determined by taking the expected value as follows:

$$\kappa = (1 - R)P \frac{\mathbb{E}\left[B_0(T \wedge \tau)1_{\{\tau \leqslant T\}}\right]}{\mathbb{E}\left[\sum_{0 \leqslant T_i < \tau} B_0(T_i)\right]}. \tag{1.1.5}$$

Exercise 1.1.6. *Consider a CDS on a defaultable bond with maturity $T = 1$ year and recovery rate R at 90%. Let the default time τ be a random variable with the Poisson distribution with mean 6 months. Assume that the yield of a risk-free zero bond is a constant 1 for all maturities within a year. Find the monthly premium payments of the CDS in terms of the principle of the defaultable bond $P = \$1$.*

Collateralized debt obligation (CDO)

A CDO is a complicated financial instrument. For illustrative purposes, we present a simplified structure of a CDO in this section. One leg of CDO is a special-purpose entity (SPE) that holds a portfolio of defaultable assets such as mortgage-backed securities, commercial real estate bonds, and corporate loans. These defaultable assets serve as collateral; therefore we call the portfolio of these assets a *collateral portfolio*. Then, SPE issues bonds, which pay the cashflow of the assets to investors in these bonds. The holders of these special bonds do not uniformly receive the cashflow. There are four types of bonds in four *trenches*: senior, mezzanine, junior, and equity. The cashflow is distributed among investors first to the holders of senior bonds, then mezzanine bond holders, then junior bond holders, and finally equity bond holders. In case of default of some of the collateral assets in the portfolio, equity holders are the first to lose income. Therefore, a senior trench bond is the most expensive and an equity trench bond is the cheapest. CDOs are traded in specialized debt markets, derivative markets, or over-the-counter (OTC).

A CDO can be structurally very complicated. For illustration purpose, in the next example we focus on a CDO that is written only on zero bonds.

Example 1.1.6. *Consider a collateral portfolio of 100 different defaultable zero bonds with the same maturity. Let's trenchize the CDO in four equally sized trenches as shown in Figure 1.1.13. If none of the bonds in the collateral portfolio default, the total \$100 cashflow will be evenly distributed among CDO bond holders. However, if ten bonds default, then total cashflow is \$90; an amount of \$75 to be evenly distributed among the junior, mezzanine, and senior holders, and the remaining amount of \$15 dollars will be evenly distributed among equity holders. If 30 bonds default, then total cashflow is \$70; an amount of \$50 to be evenly distributed among the mezzanine and senior holders, and the remaining amount of \$5 dollars will be evenly distributed among mezzanine holders. Equity holders receive \$0. If there are at least 50 defaults, equity and junior holders receive nothing. The mezzanine trench loses cashflow, if and only if the number of defaults exceed 50. The senior trench receives full payment, if and only if the number of defaults remains at 75 or below.*

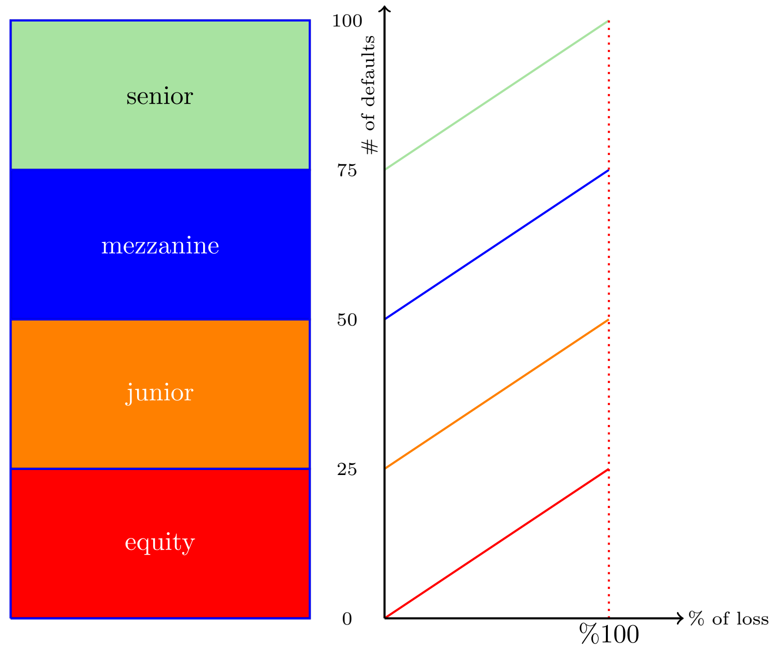


Figure 1.1.13: Simplification of CDO structure

Above 75 defaults, equity, junior, and mezzanine trenches totally lose their cashflow, and senior trench experiences a partial loss.

Exercise 1.1.7. *Consider a credit derivative on two independently defaultable zero bonds with the same face value at the same maturity, which pays the face value of either bond in case of default of that bond. A credit derivative of this type is written on two independently bonds: one of the bonds has a risk-adjusted yield of 5% and the other has a risk-adjusted yield of 15%. Another credit derivative of this type is written on two other independently defaulted bonds, both with a risk-adjusted yield of 10%. If both credit derivatives are offered at the same price, which one is better? Hint: The probability of default is* $\mathbb{P}(\tau \leqslant T \mid \tau > t) = e^{-\lambda_t(T)(T-t)}$ *and the risk-adjusted yield satisfies* $R_t^{\lambda}(T) = R_t(T) - \frac{1}{T-t} \ln\left(1 - e^{-\lambda_t(T)(T-t)}\right)$. *Therefore, the probability of default satisfies* $\mathbb{P}(\tau \leqslant T \mid \tau > t) = 1 - e^{-(R_t^{\lambda}(T) - R_t(T))(T-t)}$.

Loss distribution and systemic risk

We learned from the 2007 financial crisis that even a senior trench bond of a CDO can yield yield an unexpectedly low cashflow caused by an unexpectedly large number of defaults in the collateral portfolio, especially when the structure of the collateral portfolio creates a *systemic risk*. To explain the systemic risk, consider a collateral portfolio, which is made up of mortgages and mortgage-based securities. These assets are linked through several common risk factors, some are related to the real estate market, and the others are related to the overall situation of the economy. These risk factors create a correlation between defaults of these assets. Most of the risk factor that were known before the 2007 financial crisis can only cause a relatively small group of assets within the collateral portfolio to default. If risk factor can significantly increase the chance of default of a large group of assets within the collateral portfolio, then it is a systemic risk factor. If we only look at the correlation between the defaults of the assets in the collateral portfolio, we can handle the nonsystemic risk factors. However, a systemic risk factor can only be found by studying the structure of the collateral portfolio beyond the correlation between the defaults.

A difference between systemic and nonsystemic risk can be illustrated by the severity of loss. In Figure 1.1.14, we show the distribution of loss in three different cases: independent defaultable assets, dependent defaultable assets without a systemic factor, and dependent defaultable assets with systemic factor. The loss distribution, when a systemic risk factor exists has at least one spike at a large loss level. It is important to emphasize that the empirical distribution of loss does not show the above-mentioned spike and the systemic risk factor does not leave a trace in a calm situation of a market. Relying on only a period of market data, in which systemic losses have not occurred, leads us into a dangerous territory, such as the financial meltdown of 2007–2008. Even having the data from a systemic event may not help predict the next systemic event, unless we have a sound understanding of the financial environment. Therefore, we can only find systemic risk factors through studying the structure of a market.

Take the following example, as extreme and hypothetical as it is, as an illustration of systemic risk. Consider a CDO made up of a thousand derivatives on a single defaultable asset. If the asset does not default, all trenches collect even shares of the payoff of the derivatives. However, in case of default, all trenches become worthless. Even if you increase the number of assets to, ten, it only takes a few simultaneously defaulted assets to blow up the CDO. Even when the number of assets becomes large, their default may only depend on few factors; i.e., when a few things go wrong, the CDO can become worthless.

Exercise 1.1.8. *Consider a portfolio of 1000 defaultable asset with the same future value \$1. Let Z_i represent the loss from asset i which is 1 when asset i defaults and 0 otherwise. Therefore, the total loss of the portfolio is equal to $L := \sum_{i=1}^{1000} Z_i$. Plot the probability density function (pdf) of L in the following three cases.*

a) *$Z_1, ..., Z_{1000}$ are i.i.d. Bernoulli random variables with probability $p = .01$.*

b) *$Z_1, ..., Z_{1000}$ are correlated in the following way. Given the number of defaults $N \geqslant 0$, the defaulted assets can with equal likelihood be any combination of N out of 100 assets, and N is distributed as a negative binomial with parameters $(r, p) = (90, .1)$. When $N \geqslant 1000$, all assets have defaulted. See Example B.14. Plot the pdf of L.*

c) *Now let X be a Bernoulli random variable with probability $p = .005$. Given $X = 0$, the new set of random variables $Z_1, ..., Z_{1000}$ are i.i.d. Bernoulli random variables*

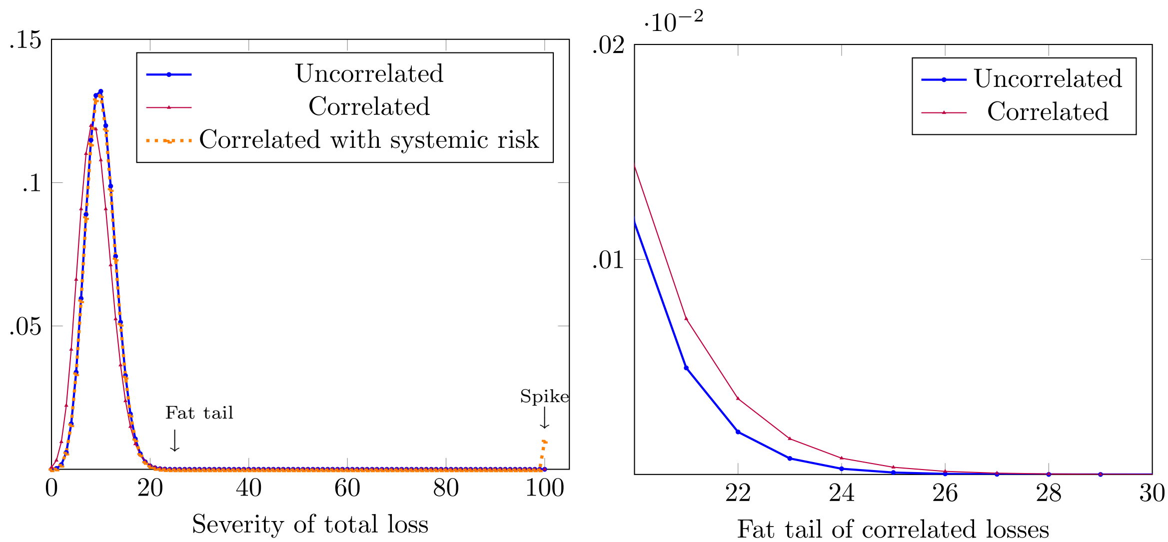


Figure 1.1.14: Distribution of loss: Correlation increases the probability at the tail of the distribution of loss. Systemic risk adds a spike to the loss distribution. All distributions have the same mean. The fat-tailed loss distribution and the systemic risk loss distribution have the same correlation of default.

with probability $p = .01$, and given $X = 1$, random variables $Z_1, ..., Z_{1000}$ are i.i.d. Bernoulli random variables with probability $p = .3$. The structure is illustrated in Figure 1.1.15.
Plot the distribution of the loss.

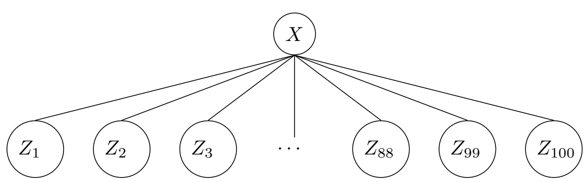


Figure 1.1.15: Variable X represents a systemic factor for variables $Z_1, ..., Z_{100}$. When $X = 1$, the chance of $Z_i = 1$ increases drastically. Since $Z_i = 1$ represents the loss from asset i, when the systemic factor X is passive, i.e., $X = 0$, the loss distribution is similar to one for a portfolio of independent defaultable assets.

There are two main approaches to modeling a financial environment. Some studies, such as (Acemoglu, Ozdaglar, and Tahbaz-Salehi 2015; Cont, Moussa, and Santos 2011; Amini, Filipović, and Minca 2015), model a complex financial network of loans by a random graph. Others, such as (Garnier, Papanicolaou, and Yang 2013), use the theory of mean field games to model the structure of a financial environment. While the former emphasizes the contribution of the heterogeneity of the network in systemic risk, the latter shows that systemic risk can also happen in a homogeneous environment. An example of such a network method, that most central banks and central clearinghouses use to assess systemic risk to the financial networks they oversee is the Eisenberg-Noe model, which is discussed in Section 1.1.2.

1.2 Optimization in finance

Optimization is a regular practice in finance: a hedge fund wants to increase its profit, a retirement fund wants to increase its long-term capital gain, a public company wants to increase its share value, and so on. One of the early applications of optimization in finance is the Markowitz mean-variance analysis on diversification; [20]. This leads to quadratic programming and linear optimization with quadratic constraint. Once we define the condition for the optimality of a portfolio in a reasonable sense, we can build an optimal portfolio, or an *efficient portfolio*. Then, the efficient portfolio can be used to analyze other investment strategies or price new assets. For example, in the capital asset pricing model (CAPM), we evaluate an asset based on its correlation with the efficient market. In this section, we present a mean-variance portfolio selection problem as a classical use of optimization methods in finance.

Consider a market with N assets. Assume that we measure the profit of the asset over a period $[0,1]$ by its *return*:

$$\mathbf{R}_i = \frac{S_1^{(i)} - S_0^{(i)}}{S_0^{(i)}}.$$

Here, $S_0^{(i)}$ and $S_1^{(i)}$ are the current price and the future price of the asset i, respectively. The return on an asset is the relative gain of the asset. For instance, if the price of an asset increases by 10%, the return is 0.1. Since the future price is unknown, we take return as a random variable and define the expected return by the expected value of the return, i.e.,

$$R_i := \mathbb{E}[\mathbf{R}_i].$$

The risk of an asset is defined as the standard deviation of the return σ_i, where

$$\sigma_i^2 := \mathrm{var}(\mathbf{R}_i).$$

Expected return and risk are two important factors in investment decisions. An investor with a fixed amount of money wants to distribute her wealth over different assets to make an *investment portfolio*. In other words, she wants to choose weights $(\theta_1, ..., \theta_N) \in \mathbb{R}_+^N$ such that $\sum_{i=1}^N \theta_i = 1$ and invest θ_i fraction of her wealth on asset i. Then, her expected return on this portfolio choice is given by

$$R_\theta := \sum_{i=1}^{N} \theta_i R_i.$$

However, the risk of her portfolio is a little more complicated and depends on the correlation between the assets

$$\sigma_\theta^2 := \sum_{i=1}^{N} \theta_i^2 \sigma_i^2 + 2 \sum_{1 \leqslant i < j \leqslant N} \theta_i \theta_j \varrho_{ij} \sigma_i \sigma_2.$$

Here ϱ_{ij} is the correlation between the returns of assets i and j.

Example 1.2.1 (Two assets)**.** *Assume that $N = 2$ and $\mathbf{R}_1$ and $\mathbf{R}_2$ are assets with correlated returns, and that correlation is given by ϱ_{12}. Thus, for $\theta \in [0,1]$, we invest θ portion of the wealth in asset 1 and the rest in asset 2. Then, the expected return and the risk of the portfolio as a function of θ are given by*

$$R_\theta = \theta R_1 + (1-\theta) R_2 \;\; \text{and} \;\; \sigma_\theta^2 = \theta^2 \sigma_1^2 + (1-\theta)^2 \sigma_2^2 + 2\theta(1-\theta)\varrho_{12}\sigma_1\sigma_2.$$

Therefore, by eliminating θ, σ_θ^2 becomes a quadratic function of R_θ; see the red parabola in Figure 1.2.1.

Exercise 1.2.1. *Show that when $\theta = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$, the portfolio with two uncorrelated assets in Example 1.2.1 takes the minimum risk σ_θ^2. Find the minimum value. Repeat the result for the positively and negatively correlated assets.*

As seen in Figure 1.2.1 and shown in Exercise 1.2.1, there is a portfolio with minimum risk $\sigma_{\min}$ and return R^* that is higher than the minimum return between the two assets. If the goal is to minimize risk regardless of the return of the portfolio, there is a better option than fully investing in the lower-risk asset. Even if the assets are correlated, negatively or positively, the minimum risk portfolio exists. The only exception is when the two assets are positively correlated with $\varrho_{12} = 1$, where the least risky option is to invest fully in the asset with lower risk. Notice that in the dotted parts of the red and green curves, all portfolios are worse than the minimum risk portfolio. In other words, the minimum risk portfolio has higher return than all dotted portfolios while it maintains the lowest risk. By choosing a portfolio in the solid part of the curve, we accept to take higher risk than the minimum risk portfolio. In return, the return of the chosen portfolio also is higher than the return of the minimum risk portfolio. The solid part of the curve is called the efficient frontier.

The collection of all portfolios made up of more than two assets is not represented simply by a one-dimensional curve; such a portfolio is represented by a point in a two-dimensional region that is not always easy to find. However, Robert Merton in [21] shows that the

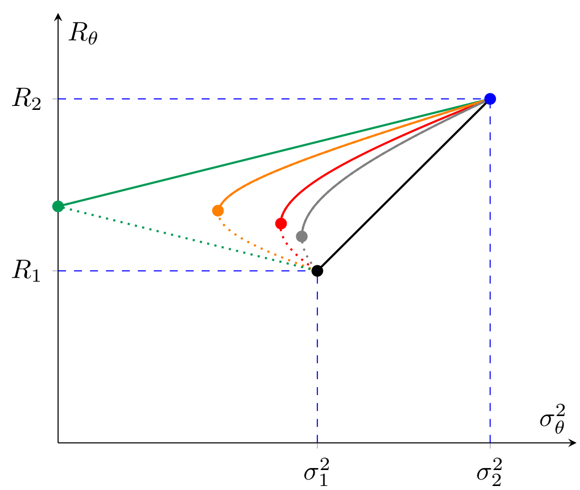


Figure 1.2.1: The risk of a portfolio with two assets as a function of expected return. σ_θ^2 as a function of R_θ is a quadratic curve. The green curve indicates when the returns of assets are negatively correlated with correlation -1, the green curve indicates when the returns of assets are negatively correlated with correlation less than -1, the red curve indicates uncorrelated assets, the gray curve indicates when they are positively correlated, and the black curve indicates when the returns of assets are positively correlated with correlation 1.

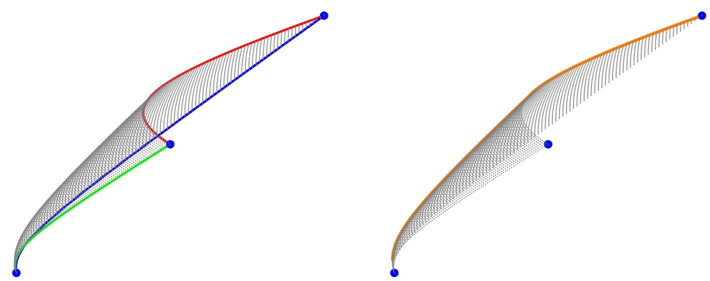


Figure 1.2.2: Left: the risk vs return of all portfolios on three assets. The assets are marked by blue dots. The bi-asset risk-return curves are plotted with blue, green and red colors. Right: The efficient frontier for three assets is shown in orange.

efficient frontier always exists regardless of the number of assets. In Figure 1.2.2, we sketched the risk-return region for three assets and marked the efficient frontier for them.

To define the efficient frontier, we impose a natural partial order among all portfolios based on their risk and return: $\theta > \hat{\theta}$ if and only if $R_\theta > R_{\hat{\theta}}$ and $\sigma_\theta \leqslant \sigma_{\hat{\theta}}$, or $R_\theta \geqslant R_{\hat{\theta}}$ and $\sigma_\theta < \sigma_{\hat{\theta}}$. In other words, one portfolio is better than another if it has either a lower risk with at least the same return, or a higher return with at most the same risk. The efficient frontier is the set of all maximal portfolios under this order; i.e., there is no portfolio that is better.

Minimum risk portfolio

The portfolio with the least risk $\sigma_{\min}$ is located in the lowest end of the efficient frontier. To find the lowest-risk portfolio, we solve the following problem.

$$\min \sigma_\theta \ \text{ subject to } \ \theta \geqslant 0 \ \text{ and } \ \sum_i \theta_i = 1.$$

Notice that the above optimization problem is equivalent to solving the quadratic programming problem.

$$\min \sigma_\theta^2 \ \text{ subject to } \ \theta \geqslant 0 \ \text{ and } \ \sum_i \theta_i = 1. \tag{1.2.1}$$

Recall that

$$\sigma_\theta^2 := \sum_{i=1}^{N} \theta_i^2 \sigma_i^2 + 2 \sum_{1 \leqslant i < j \leqslant N} \theta_i \theta_j \varrho_{ij} \sigma_i \sigma_2$$

which can be written in matrix form by $\theta^\mathsf{T} C \theta$. Here, the matrix C is the variance-covariance matrix between assets given by

$$C_{ij} = \varrho_{ij} \sigma_i \sigma_j.$$

If all assets are linearly uncorrelated, then C is positive-definite. Therefore, problem (1.2.1) is similar to those studied in Section A.2.

Minimizing risk subject to return constraint

If we want to have a return higher than the return from the portfolio, we have to take more risk than $\sigma_{\min}$. This can be achieved by adding the constraint of a minimum return.

$$\min \sigma_\theta \text{ subject to } \theta \geqslant 0 \ \sum_i \theta_i = 1 \text{ and } R_\theta \geqslant R_0. \tag{1.2.2}$$

The constant R_0 is the desired return from the portfolio.

Exercise 1.2.2. *Consider a portfolio of ten assets with the expected return given by*

$$\begin{bmatrix} .1 & .2 & .3 & .5 & .2 & .1 & .05 & .1 & .2 & .1 \end{bmatrix}$$

and the variance-covariance matrix by

$$\begin{bmatrix}
1 & .2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
.2 & 1 & .2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & .2 & 1 & .2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & .2 & 1 & .2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & .2 & 2 & .1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & .1 & 2 & .1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & .1 & 2 & .1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .1 & 2 & .1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & .1 & 2 & .1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .1 & 2
\end{bmatrix}$$

Use `CVX` under `MatLab` or `CVXOPT` under `Python`, introduced in Section A.2, to numerically solve problems (1.2.1) and (1.2.2).

Maximizing return subject to risk constraint

It is obvious that the highest return comes from the asset with the highest return. However, it may be too risky to invest all one's resources in one risky asset. Therefore, there is usually a risk constraint:

$$\max R_\theta \text{ subject to } \theta \geqslant 0 \ \sum_i \theta_i = 1 \text{ and } \sigma_\theta^2 \leqslant \sigma_0. \tag{1.2.3}$$

The constant σ_0 is the maximum risk of the portfolio.

Exercise 1.2.3. *Assume that all assets are uncorrelated. Use the Cauchy-Schwartz inequality to show that an optimal solution for problem* (1.2.3) *is obtained when* $\theta_i = \frac{R_i}{a\sigma_i^2}$ *where* $a = \sum_i \frac{R_i}{\sigma_i^2}$.

Now consider the general case where the assets are correlated with a positive-definite variance-covariance matrix. Use the Cholesky decomposition for the positive-definite matrices to find an optimal solution to the problem. Is the optimal solution unique?

1.3 No-dominance principle and model-independent arbitrage

In this section, we focus on the market properties that hold true objectively regardless of the choice of the model. Therefore, here we do not make any assumption about the dynamics of the assets, the yield of zero bonds, and the like. Instead, we only impose two basic assumptions and ignore any friction in the market such as transaction cost, liquidity restriction, and nontradability of assets. Some of the contents of this section can be found in [30, Section 1.2].

We consider a sample space Ω that includes the collection of all possible events in the market at future time T, and let χ be given set of portfolios (a collection of assets and strategies on how to trade them dynamically) in the market. For any portfolio in χ, the payoff of the portfolio is what it is worth at time T, and is a random variable from Ω to $\mathbb{R}$. We assume that there exists a pricing function $\Pi : \chi \to \mathbb{R}$; i.e., the price of portfolio P is given by $\Pi(P)$. We denote the payoff of portfolio P at the event $\omega \in \Omega$ by $P(\omega)$.

The first assumption that we impose is the following:

Assumption 1.3.1 (No-dominance)**.** *If* $P \in \chi$ *has a nonnegative payoff, i.e.,* $P(\omega) \geqslant 0$ *for all* $\omega \in \Omega$*, then the price* $\Pi(P)$ *of portfolio* P *is nonnegative.*

Remark 1.3.1. *No-dominance principle implies that two portfolios* P_1 *and* P_2 *with* $P_1(\omega) = P_2(\omega)$ *for all* $\omega \in \Omega$ *have the same price.*

The second assumption is the linearity of the pricing function. For any two portfolios P_1 and P_2, define $P_1 \oplus P_2$ is the portfolio made by combining the two. We impose the following natural assumption of linearity.

Assumption 1.3.2. *For* $P_1, P_2 \in \chi$*, we have*

$$\Pi(P_1 \oplus P_2) = \Pi(P_1) + \Pi(P_2).$$

Let's first fix some terminology that we'll be using. By "being in a long position" in an asset, a bond, or the like, we mean that we hold the asset, bond, etc. Similarly, by "being in a short position" in an asset, a bond, or the like, we mean that we owe the asset, bond, etc. For a bond or an option, the issuer is in the short position. Having a short position in an asset means borrowing that asset and then selling it for cash or keeping it for other reasons. This is often referred to as *short selling*, which is a common practice in the market. Usually, the borrower is obliged to pay the short-sold security back upon the request of the lender or in an agreed time.

Example 1.3.1. *The forward price K satisfies*

$$F_t(T) = \frac{S_t}{B_t(T)}.$$

To see this, consider a portfolio made of a long position in the underlying asset and a short position in $a = F_t(T)$ zero bonds at time t. The value of this portfolio at time T is $S_T - F_t(T)$. This value is the same as the payoff of a long position in forward contract. Therefore, by the no-dominance principle (Remark 1.3.1), we obtain that the value of the portfolio is the same as the price of the forward contract, which is zero; $S_t - B_t(T)F_t(T) = 0$. On the other hand, the price of the forward is zero and therefore we have the result.

Remark 1.3.2. *Proposition 1.3.1 explains a market condition called contango in the futures markets; the futures price is larger than the spot price. If holding an underlying asset incurs storage cost, then the result of Proposition 1.3.1 may be violated and we have $F_t(T) < S_t$. This market condition is called backwardation. Backwardation can also occur if the underlying of the futures contract is not even storable, for instance electricity.*

Example 1.3.2. *Forward rate $f_0(t,T)$ for delivery at time t of a zero bond with maturity T satisfies*

$$f_0(t,T) = \frac{\ln B_0(t) - \ln B_0(T)}{T - t}.$$

To see this, consider a portfolio made of a zero bond with a principle of \$1 at time T. The price of this portfolio is $B_0(T)$. On the other hand, consider another portfolio made of a forward rate agreement on a bond with the principle of \$1 starting at time t, maturing at time T, and with a forward rate $f_0(t,T)$; and a bond with a principle of \$$e^{-f_0(t,T)(T-t)}$ maturing at time t. The price of such a portfolio is \$$B_0(t)e^{-f_0(t,T)(T-t)}$. Since both portfolios have the same payoff of \$1, they must have the same price. Thus, $B_0(t)e^{-f_0(t,T)(T-t)} = B_0(T)$.

Exercise 1.3.1. *Consider a zero bond $B_0^d(T)$ in the domestic currency and another zero bond $B_0^f(T)$ in a foreign currency. At the current time $t = 0$, the domestic-to-foreign exchange rate is denoted by $r_0^{d/f}$, and the forward domestic-to-foreign exchange rate*[8] *for time T is denoted by $f_0^{d/f}(T)$. Show that*

$$B_0^d(T) r_0^{d/f} = B_0^f(T) f_0^{d/f}(T).$$

Proposition 1.3.1. *The price of an American option is always greater than or equal to the price of a European option with the same payoff.*

[8] The forward exchange rate is guaranteed at maturity.

Proof. An American option can always be exercised, but not necessarily optimally, at maturity, and generates the same payoff as a European option. □

Proposition 1.3.2. *The price of vanilla options satisfies*

$$C(T, K_1, S, t) \leqslant C(T, K_2, S, t) \quad \text{and} \quad P(T, K_1, S, t) \geqslant P(T, K_2, S, t),$$

where $K_1 \geqslant K_2$.

Proof. Consider a portfolio that consists of a long position in a call option with strike price K_2 and a short position in a call option with strike K_1, both maturing at T. Then, the terminal value of the portfolio is $(S_T - K_2)_+ - (S_T - K_1)_+$, which is nonnegative. By the no-dominance principle, we have $C(T, K_2, S, t) - C(T, K_1, S, t) \geqslant 0$. For a put options, a similar argument works. □

Exercise 1.3.2. *Show that the price of an American call or put option is an increasing function of maturity* T.

Exercise 1.3.3. *Let* $\lambda \in (0, 1)$. *Then,*

$$C(T, \lambda K_1 + (1-\lambda)K_2, S, t) \leqslant \lambda C(T, K_1, S, t) + (1-\lambda)C(T, K_2, S, t).$$

In other words, the price of a call option is convex in K.
Show the same claim for the price of a put option, an American call option, and an American put option.

Exercise 1.3.4. *It is well known that a convex function has right and left derivatives at all points. From the above exercise it follows that the right and the left derivatives of a call option price with respect to strike price,* $\partial_{K\pm} C(T, K, S, t)$, *exists. Use no-dominance to show that*

$$-B_t(T) \leqslant \partial_{K\pm} C(T, K, S, t) \leqslant 0$$

Hint: Consider a portfolio made of a long position in a call with strike K_2, *a long position in* $K_2 - K_1$ *bonds, and a short position in a call option with strike* K_1.

Proposition 1.3.3 (Put-call parity)**.** *The price of a call option and the price of a put option with the same strike and maturity satisfy*

$$C(T, K, S, t) + KB_t(T) = S + P(T, K, S, t).$$

Proof. Since $(S_T - K)_+ + K = S_T + (K - S_T)_+$, according to no-dominance principle, a portfolio consisting of a call option and K units of zero bond $B_t(T)$ is worth as much as a portfolio made of a put option and one unit of underlying asset. □

Exercise 1.3.5. *A portfolio of long positions in call options with the same maturity and strikes on different assets is worth more than a call option on a portfolio of the same assets with the same weight; i.e.,*

$$\sum_{i=1}^{n} \lambda_i C(T, K_i, S^{(i)}, t) \geqslant C(T, \hat{K}, \hat{S}, t), \tag{1.3.1}$$

where $\lambda_i \geqslant 0$ is the number of units invested in the call option on the ith underlying asset, $K_i \geqslant 0$ is the strike of the call option on the ith underlying asset , $S^{(i)}$ denotes the current price of the ith asset, $\hat{K} = \sum_{i=1}^{n} \lambda_i K_i$, and $\hat{S} = \sum_{i=1}^{n} \lambda_i S^{(i)}$ is the value of a portfolio that has λ_i units of ith asset for $i = 1, ..., n$.

Remark 1.3.3. *Exercise 1.3.5 demonstrates an important implication about the risk of a portfolio. A portfolio made of different assets can be insured against the risk of price increase in two ways: by purchasing a call option for each unit of each asset or by purchasing a call option on the whole portfolio. It follows from (1.3.1) that the latter choice is cheaper than the former. An option on a portfolio is called a basket option.*

Proposition 1.3.4 (Arbitrage bounds for the price of a call option)**.** *The price of a call option should satisfy*

$$(S - B_t(T)K)_+ \leqslant C(T, K, S, t) \leqslant S.$$

Proof. Since $(S_T - K)_+ \leqslant S_T$, no-dominance implies that $C(T, K, S, t) \leqslant S$. To see the right-hand side inequality, first notice that since $0 \leqslant (S_T - K)_+$, $0 \leqslant C(T, K, S, t)$. On the other hand, a portfolio of along position in the underlying asset and a short position in K units of zero bond has a payoff $S_T - K$ which is less than or equal to the payoff of call option $(S_T - K)_+$. Therefore, $S - KB_0(T) \leqslant C(T, K, S, t)$. As a result, $(S - B_t(T)K)_+ = \max\{0, S - B_t(T)K\} \leqslant C(T, K, S, t)$. □

The notion of model-specific arbitrage will later be explored in Section 2.1. However, in this section, we present *model-independent arbitrage*, which is in the same context as the no-dominance principle.

Definition 1.3.1. *A portfolio P with a positive payoff and a zero price is called a model-independent arbitrage. In other words, a portfolio P is called model-independent arbitrage if $P(\omega) > 0$ for all $\omega \in \Omega$, and $\Pi(P) = 0$.*

By Definition 1.3.1, a portfolio with $\chi(\omega') = 0$ for some $\omega' \in \Omega$ and $P(\omega) > 0$ $\omega \neq \omega'$ is not a model-independent arbitrage. This leads to a weaker notion of arbitrage.

Definition 1.3.2. *A portfolio P with a nonnegative payoff such that for some $\omega \in \Omega$ $P(\omega) > 0$ and zero price is called a weak arbitrage. In other words, a portfolio P is called a weak arbitrage if $P(\omega) \geqslant 0$ for all $\omega \in \Omega$, $P(\omega') > 0$ for some $\omega' \in \Omega$ and $\Pi(P) = 0$.*

Example 1.3.3. *Let $K_1 > K_2$ and $C(T, K_1, S, t) > C(T, K_2, S, t)$. A portfolio that consists of a short position in a call with strike K_1, a long position in a call with strike K_2, and the difference of$C(T, K_1, S, t) - C(T, K_2, S, t)$ in cash is a model-independent arbitrage. This is because, the value of such a portfolio is zero. However, the payoff is $(S_T - K_2) - (S_T - K_1)_+ + C(T, K_1, S, t) - C(T, K_2, S, t)$, which is strictly positive, and therefore we have a model-independent arbitrage.*

On the other hand, if $C(T, K_1, S, t) = C(T, K_2, S, t)$, the same portfolio described in the previous paragraph has a positive value whenever $S_T > K_2$, and a zero value otherwise. Therefore, it is only a weak arbitrage.

Example 1.3.4. *As a result of Exercise 1.3.4, if the price of the option is smaller that $(S - B_t(T)K)_+$, or larger than the asset price S, then there is a model-independent arbitrage. For example, if $C(T, K, S, t) > S$, one can have a portfolio of a short position in a call and a long position in the underlying asset, and the difference $C(T, K, S, t) - S$ in cash. The value of this portfolio is zero. However, the payoff is strictly positive, i.e., $S_T - (S_T - K)_+ + C(T, K, S, t) - S > 0$.*

Remark 1.3.4. *When $C(T, K, S, t) = S$ and $K > 0$, Example 1.3.4 suggests that we still have a model-independent arbitrage, unless the event $S_T = 0$ is legitimate. Therefore, choice between the model-independent arbitrage or weak arbitrage depends on whether the event $S_T = 0$ is included in Ω or not.*

Proposition 1.3.5. *If there is no model-independent arbitrage, then the no-dominance principle holds.*

Proof. Assume that no-dominance does not hold; i.e., there is a portfolio P with nonnegative payoff with a negative price $\Pi(P)$. Then, consider a new portfolio made of a long position in the old portfolio P and holding a zero bond with face value $-\Pi(P)$. The new portfolio has a positive payoff of $P(\omega) - \Pi(P)$ for each $\omega \in \Omega$ and a zero price. Therefore, it is a model-independent arbitrage. ☐

Exercise 1.3.6. *For $0 < t < T$, show that if $B_0(t)B_t(T) > B_0(T)$ (equiv. $B_0(t)B_t(T) < B_0(T)$), there is a model-independent arbitrage.*

Exercise 1.3.7. *Consider a zero bond $B_0^d(T)$ on the domestic currency and another zero bond $B_0^f(T)$ on a foreign currency. At the current time $t = 0$, the domestic-to-foreign exchange rate is denoted by $r_0^{d/f}$ and the forward domestic-to-foreign exchange rate for time T is denoted by $f_0^{d/f}(T)$. Show that if*

$$B_0^d(T) r_0^{d/f} > B_0^f(T) f_0^{d/f}(T),$$

then there exists a model-independent arbitrage.

2

Modeling financial assets in discrete-time markets

Section 2.1 deals with a single-period market in which there are only two trading dates, one at the beginning of the period and one at maturity. In this section, an introductory knowledge of probability theory is required; more specifically, the reader needs the theory of probability on discrete sample spaces that is provided in Section B.1. A few times in Section 2.2, we mention concepts form the general theory of probability. However, these cases are not crucial for understanding of the content of this sections and can simply be skipped. In Section 2.3, we extend the results of Section 2.1 to a multiperiod market with a focus on the binomial model. This is also important in our later study of continuous-time markets, which can be seen as the limit of discrete-time markets. A key concept from the appendix is the notion of martingales that is explained in Section B.3. The last section, Section 2.4, deals with the problem of tuning the parameters of a model to match the data in the specific context of binomial model.

2.1 Arrow-Debreu market model

The ideas and concepts behind pricing derivatives are easier to explain in a single-period framework with finite number of outcomes, i.e., Arrow-Debreu market model. According to the Arrow-Debreu market model, an asset has a given price and a set of possible values. There are N assets with prices arranged in a column vector $p = (p_1, ..., p_N)^\intercal$[1]. For each $i = 1, ..., N$, the possible future values or payoff of asset i is given by $\{P_{i,j} : j = 1, ..., M\}$. $P_{i,j}$ is the jth state of future value of asset i and M is a universal number for all assets.

[1] $A^\intercal$ is the transpose of matrix A.

Then, one can encode the payoffs of all assets into a N-by-M matrix

$$\mathbf{P} := [P_{i,j}]_{i=1,\dots,N,j=1,\dots,M} = \begin{bmatrix} P_{1,1} & \cdots & P_{1,M} \\ \vdots & \ddots & \vdots \\ P_{N,1} & \cdots & P_{N,M} \end{bmatrix}.$$

In the Arrow-Debreu market model, row $\mathbf{P}_{i,\cdot}$ of the matrix $\mathbf{P}$ represents all future prices of asset i for different states of the market, and column $\mathbf{P}_{\cdot,j}$ represents prices of all assets in future state j of the market.

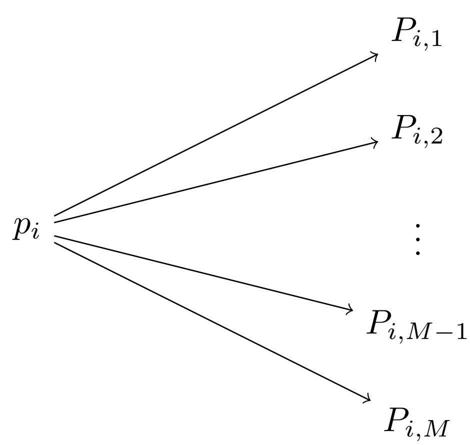


Figure 2.1.1: Arrow-Debreu market model

Example 2.1.1 (Game of chance)**.** *Let $N = 1$, $M = 2$ and $P_{1,1} = -P_{1,2} = 1$. In other words, there is a fee p_1 to enter a game of chance in which the player either gains or loses a dollar based on the outcome of flipping a coin. Notice that for now we do not specify the heads-tails probability for the coin. This probability determines whether the price of the game p_1 is a fair price or not.*

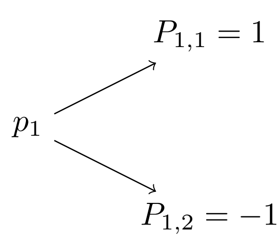


Figure 2.1.2: Game of chance described in the Arrow-Debreu market model

2.1.1 Arbitrage portfolio and the fundamental theorem of asset pricing

A *portfolio* is a row vector $\theta = (\theta_1, \cdots, \theta_N)$ where $\theta_i \in \mathbb{R}$ which represents the number of units of asset i in the portfolio. The total price of the portfolio is then given by

$$\theta p = \sum_{i=1}^{N} \theta_i p_i.$$

Here, the notation of product is simply the matrix product. Notice that if $\theta_i > 0$, the position of the portfolio in asset i is called *long*, and otherwise if $\theta_i < 0$, it is called *short*.

Arbitrage is a portfolio that costs no money but gives a nonnegative future value and for some states positive values. More precisely, we have the following definition.

Definition 2.1.1. *θ is called a weak arbitrage portfolio or weak arbitrage opportunity if*

a) $\theta p = 0$

b) $\theta \mathbf{P}_{\cdot,j} \geqslant 0$ *for all* $j = 1, ..., M$

c) $\theta \mathbf{P}_{\cdot,j} > 0$ *for at least one* j.

Notice that for a given $\theta = (\theta_1, \cdots, \theta_N)$, the portfolio represented by θ is itself an asset with value $\theta \mathbf{P}_{\cdot j}$ at the state j of the market.

We say that a market model is *free of weak arbitrage* or that it satisfies *no weak arbitrage condition* (NWA for short), if there is no weak arbitrage opportunity in this model.

Sometimes, an arbitrage opportunity starts with a zero-valued portfolio and ends up with positive values at all states of the market. This defined an strong arbitrage:

Definition 2.1.2. *θ is called a strong arbitrage portfolio or arbitrage opportunity if*

a) $\theta p < 0$

b) $\theta \mathbf{P}_{\cdot,j} \geqslant 0$ *for all* $j = 1, ..., M$

Remark 2.1.1. *Notice that if we remove some of the states of the market, then weak arbitrage can disappear. However, strong arbitrage does not.*

Notice that model-independent arbitrage as defined in Definition 1.3.1 is even stronger than strong arbitrage.

The following theorem is the most important in financial mathematics that characterizes the notion of arbitrage in a simple way. Basically, it presents a simple criterion to determine if a market model is free of weak or strong arbitrage.

Theorem 2.1.1 (Fundamental theorem of asset pricing (FTAP))**.** *The Arrow-Debreu market model is free of weak (respectively strong) arbitrage opportunity if and only if there exist a column vector of positive (respectively nonnegative) numbers* $\pi = (\pi_1, ..., \pi_M)^\top$ *such that*

$$p = \mathbf{P}\pi. \tag{2.1.1}$$

Notice that Theorem 2.1.1 does not claim that the vector π is unique, and therefore there can be several solutions π such that (2.1.1) holds.

Proof. The proof for the strong arbitrage case is the result of the so-called Farkas' lemma which asserts that

> Given an $n \times m$ matrix $\mathbf{A}$ and a vector $\mathbf{b} \in \mathbb{R}^m$. Then, exactly one of the following two statements is true:
>
> 1) There exists a $\pi \in \mathbb{R}^n$ such that $\mathbf{A}\pi = \mathbf{b}$ and $\pi \geqslant 0$.
>
> 2) There exists a $\theta \in \mathbb{R}^m$ such that $\theta\mathbf{A} \geqslant 0$ and $\theta\mathbf{b} < 0$.

We apply Farkas' lemma for $\mathbf{A} = \mathbf{P}$ and $\mathbf{b} = p$. Strong arbitrage is equivalent to (2) in Farkas' lemma. Therefore, lack of strong arbitrage is equivalent to (1).

To show the result for weak arbitrage, we consider the following extension of Farkas's lemma.

> Given an $n \times m$ matrix $\mathbf{A}$ and a vector $\mathbf{b} \in \mathbb{R}^m$. Then, exactly one of the following two statements is true:
>
> 1') There exists a $\pi \in \mathbb{R}^n$ such that $\mathbf{A}\pi = \mathbf{b}$ and $\pi > 0$.
>
> 2') There exists a $\theta \in \mathbb{R}^m$ such that $\theta\mathbf{A} \geqslant 0$, $\theta\mathbf{A} \neq 0$ and $\theta\mathbf{b} \leqslant 0$.

For $\mathbf{A} = \mathbf{P}$ and $\mathbf{b} = p$, weak arbitrage is equivalent to (2'). Therefore, lack of weak arbitrage is equivalent to (1'). □

Remark 2.1.2. *The Farkas' lemma and its extension that are used in the proof of Theorem 2.1.1 are corollaries of the separating hyperplane theorem. For more details about separating hyperplane theorem and Farkas' lemma see [8, Section 5.8.3, Section 2.5.1 and Exercise 2.20].*

Lack of strong arbitrage does not imply that the vector π in Theorem 2.1.1 has some positive entities. For instance, when $p = 0$, i.e., all the assets in the market have zero price, there is no strong arbitrage. In such a case, a trivila solution to equation (2.1.1) is $\pi = 0$. If $p \neq 0$, then no-arbitrage condition implies that π has at least a positive entity. Therefore, one can normalize it by

$$\hat{\pi} = \frac{1}{\sum_{j=1}^{M} \pi_j}\pi.$$

$\hat{\pi}$ is a probability. The probability vector $\hat{\pi} = (\hat{\pi}_1, ..., \hat{\pi}_M)^\mathsf{T}$ is referred to as a *risk-neutral* or *risk-adjusted* probability.

Example 2.1.2. *Consider a game in which you pay a fee of \$.20 to enter. Then, a coin flip determines whether you win or lose a dollar. We also include a zero bond with a face value of \$1 and a price \$.80. See Figure 2.1.3. It follows from the FTAP that the no-arbitrage condition is equivalent to the existence of a positive vector* $(\pi_1, \pi_2)^\top$ *satisfying*

$$\begin{cases} \pi_1 - \pi_2 = .2, \\ \pi_1 + \pi_2 = .8. \end{cases}$$

In fact, such a vector (uniquely) exists and is given by $(5/8, 3/8)$. *If we exclude the zero*

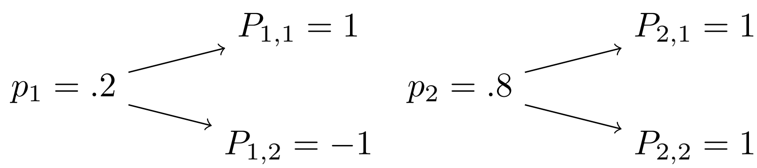


Figure 2.1.3: An Arrow-Debreu market model with two assets and two states representing a coin game and a zero bond

bond from the market, the FTAP criterion for no-arbitrage is reduced to the existence of a positive solution to $\pi_1 - \pi_2 = .2$. *There are obviously infinitely many positive solutions.*

Example 2.1.3. *Consider a game of chance using two coins; to enter the game, you pay a \$1 fee. If both coins turn heads (tails), you win (lose) a dollar. Otherwise, the gain is zero. There is also a zero bond with a face value of \$1 and a price of \$.80. It follows from FTAP that there is an arbitrage opportunity. Notice that the system of equation*

$$\begin{cases} \pi_1 - \pi_4 = 1 \\ \pi_1 + \pi_2 + \pi_3 + \pi_4 = .8 \end{cases}$$

implies $\pi_3 = -\pi_2 - .2 - 2\pi_4$, *which excludes the possibility of a positive solution. Therefore, there is an arbitrage in this game.*

FTAP predicts the existence of arbitrage but does not provide any. Knowing an arbitrage opportunity exists, finding one is sometimes a challenging problem, even in the toy Arrow-Debreu market model.

Exercise 2.1.1. *Find an arbitrage opportunity in Example 2.1.3.*

Remark 2.1.3. *The weak and strong arbitrage are model specific. If we change the model, the arbitrage opportunity can disappear. However, model-independent arbitrage remains a strong arbitrage opportunity in any possible model.*

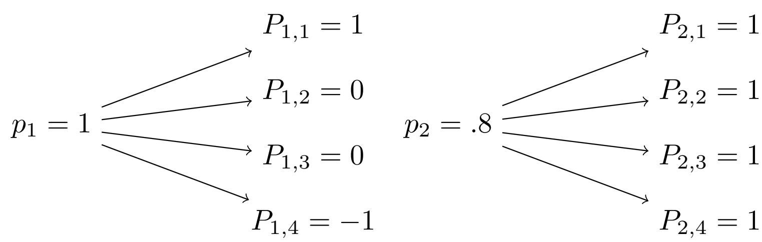


Figure 2.1.4: An Arrow-Debreu market model with two assets and four states representing a coin game and a zero bond

2.1.2 Arrow-Debreu market model with a risk-free bond

If we intend to add a new asset with price p' and values $P'_1, ..., P'_M$ to the current Arrow-Debreu market model, the new asset does not create an arbitrage opportunity if and only if at least one of the existing positive solutions π of (2.1.1) satisfies

$$p' = \sum_{j=1}^{M} P'_j \pi_j.$$

More specifically, assume that the new asset is a zero bond with yield (interest rate) R and face value 1. Therefore, its price is given by $p_0 = \frac{1}{1+R}$. See Figure 2.1.5. For no-

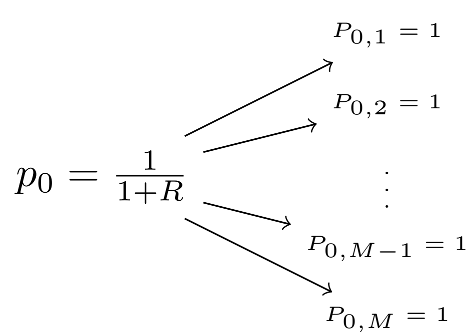


Figure 2.1.5: Risk-free asset in Arrow-Debreu market model

arbitrage condition to hold for the market with the new bond, at least one of the positive solutions of (2.1.1) implies that $p_0 = \frac{1}{1+R} = \sum_{j=1}^{M} \pi_j$. Therefore, no-arbitrage condition for an Arrow-Debreu market model with a zero bond is equivalent to existence of a positive vector π such that (2.1.1) holds and

$$\sum_{j=1}^{M} \pi_j = \frac{1}{1+R},$$

holds for all such π. Therefore, any positive vector π can be normalized to a risk-neutral probability through

$$\hat{\pi} = \frac{\pi}{\sum_{j=1}^{M} \pi_j} = (1+R)\pi.$$

Then, the price p_i of each asset is given by

$$p_i = \sum_{j=1}^{M} P_{i,j}\pi_j = \frac{1}{1+R}\sum_{j=1}^{N} P_{i,j}\hat{\pi}_j. \tag{2.1.2}$$

The right-hand side above has an important interpretation: provided that the no-arbitrage condition holds, the price of the asset is equal to the discounted expected payoff with respect to risk-neutral probability, i.e.,

$$p_i = \frac{1}{1+R}\hat{\mathbb{E}}[\mathbf{P}_{i\cdot}].$$

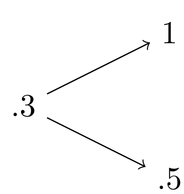
Here, $\hat{\mathbb{E}}[\mathbf{P}_{i\cdot}] := \sum_{j=1}^{M} \hat{\pi}_j P_{i,j}$ is the expected payoff of asset i with respect to the risk-neutral probability $\hat{\pi}$. Factor $\frac{1}{1+R}$ is the discount factor.

By rearranging (2.1.2), one obtains

$$R = \sum_{j=1}^{M} \frac{P_{i,j} - p_i}{p_i}\hat{\pi}_j.$$

The term $\frac{P_{i,j}}{p_i} - 1$ in RHS is the *realized return* of asset i if the state j of the market occurs. Therefore, the interpretation of the above equality is that the expected return of each asset under the risk-neutral probability $\hat{\pi}$ is equal to the risk-free interest rate R.

Example 2.1.4. *Consider an Arrow-Debreu market model with a risky asset shown below and a risk-free asset with interest rate $R = .5$. To see if there is no arbitrage in this model,*

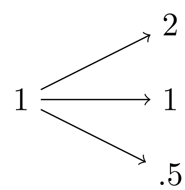


we should investigate the solutions to the system

$$\begin{cases} 1 = \hat{\pi}_1 + \hat{\pi}_2 \\ .3 = \frac{1}{1.5}\left(\hat{\pi}_1 + .5\hat{\pi}_2\right) \end{cases}$$

The first equation accounts for that $\hat{\pi}$ is a probability vector, and the second equation comes from (2.1.2). However, the only solution is $\hat{\pi} = (-.1, 1.1)^{\mathsf{T}}$, which is not a probability.

Example 2.1.5. *Consider an Arrow-Debreu market model with a risky asset shown below and a risk-free asset with interest rate $R = .5$. To see if there is no arbitrage in this model,*

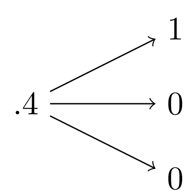


we should investigate the solutions to the system

$$\begin{aligned} 1 &= \hat{\pi}_1 + \hat{\pi}_2 + \hat{\pi}_3 \\ 1 &= \frac{1}{1.5}\left(2\hat{\pi}_1 + \hat{\pi}_2 + .5\hat{\pi}_3\right) \end{aligned}$$

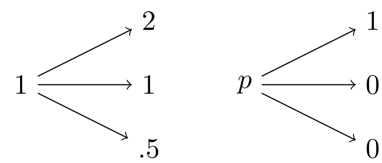
One of the infinitely many solutions to the above system is $\hat{\pi} = (.6, .2, .2)^{\mathsf{T}}$, which implies no-arbitrage condition.

If a second risky asset, shown below, is added to the market, we still do not have arbitrage because vector $\hat{\pi} = (.6, .2, .2)^{\mathsf{T}}$ works for the new market.



Exercise 2.1.2. *Consider an Arrow-Debreu market model with two risky assets shown below and a risk-free asset with interest rate $R = .5$. Find all the values for p such that the market is arbitrage free.*

Remark 2.1.4. *A zero bond is a risk-free asset in the currency of reference. For example, a zero bond that pays \$1 is risk-free under the dollar. However, it is not risk-free if the currency of reference is the euro. In the latter case, a euro zero bond is subject to the risk caused by the foreign exchange rate and is a risky asset. See Exercise (2.1.3) below.*



Exercise 2.1.3. *Consider an Arrow-Debreu market model with two assets and two states; one is a zero bond in the domestic currency with interest rate R_d under the domestic currency, and the other is a zero bond in a foreign currency with interest rate R_f under foreign currency.*

a) Given that the domestic-to-foreign exchange rate at time 0 is F_0[2], and at time 1 takes nonnegative values F_1 and F_2, what is the Arrow-Debreu market model description of a foreign zero bond in the domestic currency?

*b) A **currency swap** is a contract that guarantees a fixed domestic-to-foreign exchange rate, or a **forward exchange rate** for maturity. The forward exchange rate is agreed upon between two parties such that the value of the contract is zero. Express the forward exchange rate of a currency swap maturing at 1 in terms of F_0, R_d, F_1, R_f, and F_2.*

Remark 2.1.5. *The risk-neutral probability $\hat{\pi}$ has little to do with the actual probability (or **physical probability**) with which each state of the market happens. The probabilities $f_j := \mathbb{P}($ state j occurs$)$ can be obtained through statistical analysis on historical market data. However, the risk-neutral probability $\hat{\pi}$ depends only on the matrix $\mathbf{P}$ and vector p and not on historical market data. The one and only genuine relevance between physical probability $f = (f_1, ..., f_m)^\mathsf{T}$ and risk-neutral probability $\hat{\pi} = (\hat{\pi}_1, ..., \hat{\pi}_M)^\mathsf{T}$ is that they both assign nonzero probability to each of the states of the market; i.e., $f_j > 0$ if and only if $\hat{\pi}_j > 0$ for at least one risk-neutral probabilities $\hat{\pi}$.*

One can interpret $\hat{\pi}$ as an investor's preference toward the different states of the market. To see this, let's rewrite (2.1.2) as the following

$$p_i = \frac{1}{1+R}\sum_{j=1}^{M}\left(\frac{\hat{\pi}_j}{f_j}\right) f_j P_{i,j} = \frac{1}{1+R}\mathbb{E}\Big[\Big(\frac{\hat{\pi}_\cdot}{f_\cdot}\Big)\mathbf{P}_{i,\cdot}\Big]. \tag{2.1.3}$$

Here, $\mathbb{E}$ is the expectation with respect to the physical probability. The quotient $\frac{\hat{\pi}_j}{f_j}$ is the risk preference of the investor toward the state j of the market, which is also referred to as the state-price deflator. The state-price deflator shows that, apart from the physical probability of a certain state, an investor may be keen or averse toward the appearance of that state. For instance, f_j can be a very high probability, but state-price deflator $\frac{\hat{\pi}_j}{f_j}$ can be

[2] 1 unit of domestic is worth F_0 units of foreign.

small, which means that the probable event of appearance of state j has little value to the investor.

One can analyze the Arrow-Debreu market model by introducing M new *elementary securities* to the market; for $i = N+1, ..., N+M$, asset $N+j$ pays \$1 if market state j happens and pays \$0 otherwise. See Figure 2.1.6. Then, it is straightforward to see that $\hat{\pi}_j$ is the arbitrage free price of asset $N+j$. Therefore, the cashflow from asset i is equivalent to the cashflow of a basket of $P_{i,1}$ units of asset $N+1$, $P_{i,2}$ units of asset $N+2$, ..., and $P_{i,M}$ asset s_{N+M}. Recall from (2.1.2) that

$$p_i = \frac{1}{1+R} \sum_{j=1}^{M} P_{i,j} \hat{\pi}_j.$$

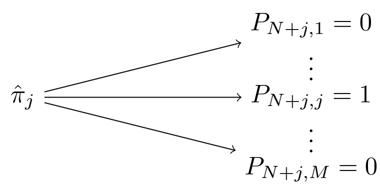
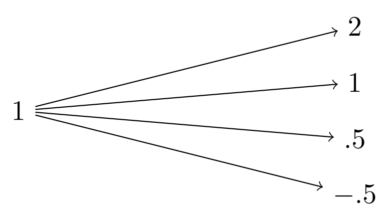


Figure 2.1.6: Elementary asset s_{N+j}

Example 2.1.6. *Consider an Arrow-Debreu market model with a risky asset shown below and a zero bond with interest rate $R = .5$.*



a) A risk-neutral probability $\hat{\pi} = (\hat{\pi}_1, \hat{\pi}_2, \hat{\pi}_3, \hat{\pi}_4)^\top$ must satisfy

$$\begin{cases} \hat{\pi}_1 + \hat{\pi}_2 + \hat{\pi}_3 + \hat{\pi}_4 = 1 \\ 2\hat{\pi}_1 + \hat{\pi}_2 + \frac{1}{2}\hat{\pi}_3 - \frac{1}{2}\hat{\pi}_4 = \frac{3}{2} \end{cases}.$$

Notice that we have two equations and four unknown; there are two more variables than equations. Therefore, we shall represent all risk-neutral probabilities as a parametrized surface with two parameters. For example, we take $\hat{\pi}_1 = s$ and $\hat{\pi}_2 = t$.

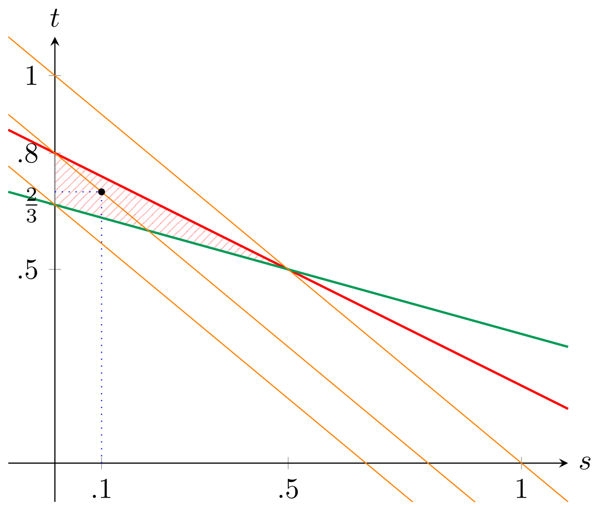


Figure 2.1.7: The hatched region represents the values for t and s which generate all risk-neutral probabilities in Example 2.1.6. The specific risk-neutral probability $(0.7, 0.1, 0.1, 0.1)^\mathsf{T}$ is shown as a black dot. Each orange line corresponds to a arbitrage-free price for the new asset that is described in Part (b) of the example.

Then, we write $\hat{\pi}_3$ and $\hat{\pi}_4$ is terms of t and s as follows:

$$\begin{cases} \hat{\pi}_3 + \hat{\pi}_4 = 1 - t - s \\ \hat{\pi}_3 - \hat{\pi}_4 = 3 - 4t - 2s. \end{cases}$$

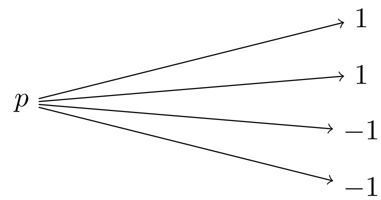
You also have to specify the suitable range for the parameters t and s such that vector $\hat{\pi}$ is nonnegative. More specifically, we must have the following four inequalities

$$\begin{cases} 0 \leqslant t \\ 0 \leqslant s \\ 0 \leqslant 2 - \frac{5}{2}t - \frac{3}{2}s \\ 0 \leqslant -1 + \frac{3}{2}t + \frac{1}{2}s \end{cases},$$

which specify the region plotted in Figure 2.1.7. The region can also be specified in the following simpler way:

$$\begin{cases} 0 \leqslant s \leqslant \frac{1}{2} \\ \frac{2}{3} - \frac{1}{3}s \leqslant t \leqslant \frac{4}{5} - \frac{3}{5}s \end{cases}. \tag{2.1.4}$$

b) We introduce the new asset below.



We shall find the range for the price p of the new asset such that the market remains free of arbitrage; strong or weak. By FTAP, Theorem 2.1.1, p must satisfy

$$p = \frac{1}{1.5}\Big(t + s - 2 + \frac{5}{2}t + \frac{3}{2}s + 1 - \frac{3}{2}t - \frac{1}{2}s\Big) = \frac{2}{3}\Big(2t + 2s - 1\Big),$$

for at least one value of (t, s) in the interior of the region found in Part (a). Therefore, valid range for p is (A, B), where

$$A := \min \frac{2}{3}\Big(2t + 2s - 1\Big) \ \textit{subject to constraints} \ (2.1.4)$$

and

$$B := \max \frac{2}{3}\Big(2t + 2s - 1\Big) \ \textit{subject to constraints} \ (2.1.4).$$

Both of the above values are the values of linear programming problems which can be solved by comparing the values of $(2/3)(2t + 2s - 1)$ at the three nodes of the hatched triangle that represents all risk-neutral probabilities, i.e., $(.5, .5)$, $(0, .7)$, and $(0, 2/3)$. The smallest value is $2/9$ which is attained at $(0, 2/3)$, and the greatest value is $2/3$ which is attained at $(.5, .5)$. Therefore, (weak) no-arbitrage for the new asset is equivalent to $p \in (2/9, 2/3)$.

c) *Lack of strong arbitrage is when the asset price is such that all the risk-neutral probabilities have some zero component. In this case, price p should be such that the risk-neutral probabilities are only on one of the edges of the hatched triangle and does not include any interior point. In Figure 2.1.7, the orange lines are $p = \frac{2}{3}\Big(2t+2s-1\Big)$ for different values of p. The only values of p which does not have an intersection with the interior of the triangle are $p = 2/9$ and $p = 2/3$. Therefore, the lack of strong arbitrage implies that $p \in [2/9, 2/3]$.*

Exercise 2.1.4. *Consider an Arrow-Debreu market model with $N = 3$ and $M = 4$ shown in Figure 2.1.8 and take the bond yield $R = 0$, where v_1 and v_2 are two (distinct) real numbers.*

a) *Find all risk-neutral probabilities.*

b) *Recall the notion of independent random variables. Find a risk-neutral probability that makes the random variables of the prices of two assets independent.*

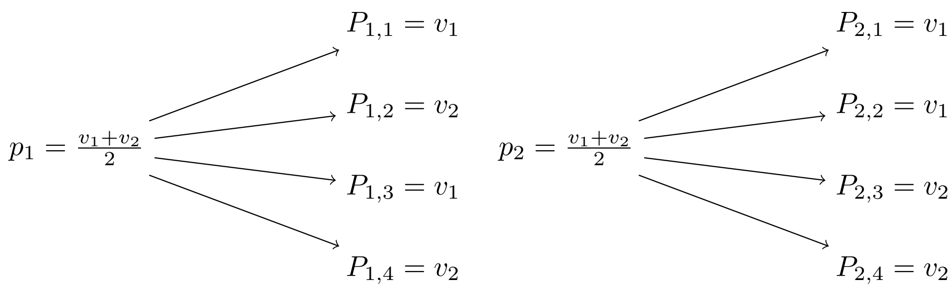


Figure 2.1.8: Exercise 2.1.4

2.1.3 One-period binomial model

Let $M = 2$ and $N = 2$ with one risk-free zero bond and a risky asset with price $S_0 := p_1$, and future cash flow given by$P_{1,1} = S_0 u$ and $P_{1,2} = S_0 \ell$ where S_0 and $\ell < u$ are all positive real numbers. By Theorem 2.1.1, in a one-period binomial model, no-arbitrage condition

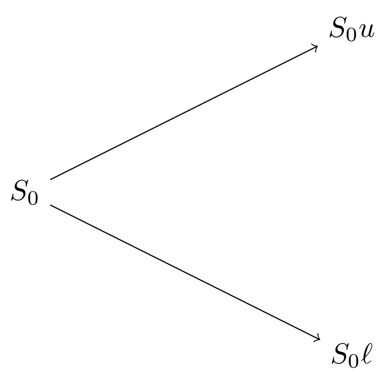


Figure 2.1.9: One-period binomial model

is equivalent to $\ell < 1 + R < u$ and $\pi = (\pi_\ell, \pi_u)^\intercal$ with $\pi_u = \frac{1+R-\ell}{(u-\ell)(1+R)}$ and $\pi_\ell = \frac{u-1-R}{(u-\ell)(1+R)}$. The risk-neutral probability is then given by $\hat{\pi}_u = \frac{1+R-\ell}{u-\ell}$ and $\hat{\pi}_\ell = \frac{u-1-R}{u-\ell}$.

Exercise 2.1.5. *Show the above claims.*

From FTAP, we know that $\ell < 1 + R < u$ is equivalent to the no-arbitrage condition. But, it is often insightful to construct an arbitrage portfolio when $\ell < 1 + R < u$ is violated. For example, consider the case when $u \leqslant 1 + R$. Then, consider a portfolio with a short position in one unit of the asset and a long position in S_0 units of bonds. To construct this portfolio, no cash is needed, and it is worth zero. However, the two possible future outcomes are either $S_0(1 + R) - S_0 u \geqslant 0$ or $S_0(1 + R) - S_0 \ell > 0$, which matches with the definition of (weak) arbitrage in Definition 2.1.1. If we assume the strict inequality $u < 1 + R$, then the arbitrage is strong.

Next, we consider the addition of a new asset into the market with payoff P_1 and P_2 in states 1 and 2, respectively. Then, no-arbitrage condition implies that the price p of this asset is must be given by

$$p = \frac{1}{1+R}\left(\hat{\pi}_u P_1 + \hat{\pi}_\ell P_2\right).$$

In particular, if the new asset is an option with payoff function $g(S_1)$, the no arbitrage price $V(S_0)$ of the option is given by

$$V(S_0) := \frac{1}{1+R}\Big(\hat{\pi}_u g(S_0 u) + \hat{\pi}_\ell g(S_0 \ell)\Big) = \frac{1}{1+R}\hat{\mathbb{E}}[g(S_1)]. \tag{2.1.5}$$

Here, $\hat{\mathbb{E}}$ is the expectation under probability $\hat{\pi}$, and S_1 is a random variable of the price of asset at time $t = 1$ that takes the values $S_0\ell$ and S_0u. For instance, a call option with payoff $(S-K)_+$ with $\ell S_0 \leqslant K < uS_0$, shown in figure 2.1.10, has a "no-arbitrage price":

$$C(K) = \frac{1+R-\ell}{(u-\ell)(1+R)}(uS_0 - K).$$

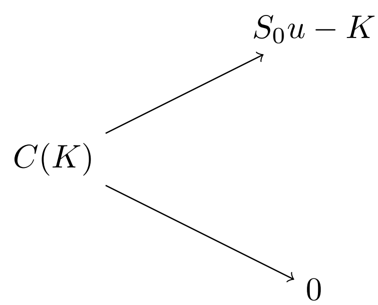


Figure 2.1.10: Cashflow of a call option in a one-period binomial model

We shall now see why any price other than $\frac{(uS_0-K)}{(1+R)}\hat{\pi}_u$ for the call option causes arbitrage in the binomial market with a zero bond, a risky asset and a call option on the risky asset. For this reason, we need to first introduce the notion of a *replicating portfolio*. Consider a portfolio with θ_0 investment in a zero bond and θ_1 units of risky asset. Then, this portfolio generates the cashflow shown in Figure 2.1.11. We want to choose (θ_0, θ_1) such that the payoff of this portfolio matches the payoff of the call option, i.e.

$$\begin{cases} \theta_0 + \theta_1 S_0 u & = S_0 u - K \\ \theta_0 + \theta_1 S_0 \ell & = 0 \end{cases} \tag{2.1.6}$$

Therefore, we have to choose

$$\theta_0 = -\frac{l(S_0u - K)}{u - \ell} \quad \text{and} \quad \theta_1 = \frac{S_0u - K}{S_0(u - \ell)}.$$

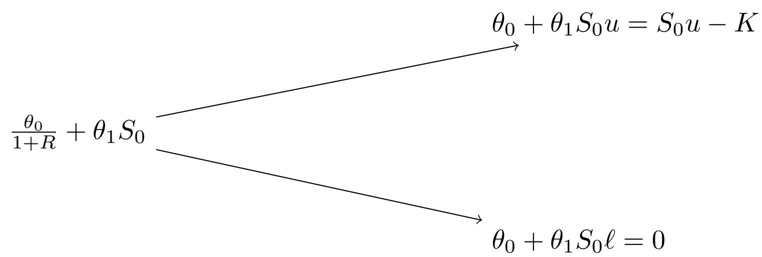


Figure 2.1.11: Replicating portfolio in a one-period binomial model

Then, one can see that the value of the replicating portfolio is equal to the price of the call option, i.e.,

$$\frac{\theta_0}{1+R} + \theta_1 S_0 = -\frac{l(S_0u - K)}{(1+R)(u - \ell)} + \frac{S_0u - K}{u - \ell} = \frac{1}{1+R}(S_0 - K)\hat{\pi}_u.$$

Now, we return to building an arbitrage in the case where the price of call option C is different from $\frac{(uS_0 - K)}{(1+R)}\hat{\pi}_u$. We only cover the case $C < \frac{(uS_0 - K)}{(1+R)}\hat{\pi}_u$. Consider a portfolio that consists of a long position in a call option and a short position in a replicating portfolio on the same call option. Shorting a replicating portfolio is equivalent to a $-\theta_0$ position in cash, and a $-\theta_1$ position in the underlying asset. Then, the value of such a portfolio is equal to $C - \frac{(uS_0 - K)}{(1+R)}\hat{\pi}_u < 0$. This means that there is some extra cash in the pocket, while the payoff of the call option can be used to clear off the shorted replicating portfolio in full. Here, the arbitrage is in the strong sense of Definition 2.1.2.

A replicating portfolio can be built for any payoff $g(S_1)$ by solving the system of equations

$$\begin{cases} \theta_0 + \theta_1 S_0 u &= g(S_0 u) \\ \theta_0 + \theta_1 S_0 \ell &= g(S_0 \ell) \end{cases} \tag{2.1.7}$$

to obtain

$$\theta_0 = \frac{ug(S_0\ell) - \ell g(S_0u)}{u - \ell} \quad \text{and} \quad \theta_1 = \frac{g(S_0u) - g(S_0\ell)}{S_0(u - \ell)}.$$

The value of the replicating portfolio is given by

$$\frac{\theta_0}{1+R} + \theta_1 S_0 = \frac{ug(S_0\ell) - \ell g(S_0u)}{(1+R)(u - \ell)} + \frac{g(S_0u) - g(S_0\ell)}{u - \ell} = \frac{1}{1+R}(\hat{\pi}_u g(S_0u) + \hat{\pi}_\ell g(S_0\ell)),$$

which is equal to the expected value of the discounted payoff under risk-neutral probability.

Exercise 2.1.6. *Consider a one-period binomial model with parameters ℓ, u and R and let $K \in (S_0\ell, S_0u]$. Find a replicating portfolio for a put option with strike K. Verify that the value of the replicating portfolio is equal to the no-arbitrage price of the put option $P(K) = \pi_\ell(K - S_0\ell)$. Then, find an arbitrage portfolio when the price of the put option with strike K is less than $P(K)$.*

Exercise 2.1.7. *Consider an Arrow-Debreu market model with $M = 2$ that consists of a risk-free bond with interest rate $R = .01$ and a forward contract[3] on a nonstorable asset[4] with forward price K and maturity of one period. Given that the payoff of the forward contract for the long position takes values $P_{1,1} = 4$, and $P_{1,2} = -2$ respective to the state of the market at maturity, is there any arbitrage?*
Now assume that the underlying asset is storable and has price $p = 10$. Given that there is no arbitrage, find K, and binomial model parameters u and d for the underlying asset.

Example 2.1.7. *Consider the binomial model with $S_0 = 4$, $R = .05$, $u = 1.45$, and $\ell = .85$. We shall price and replicate the payoff g in Figure 2.1.12 . To find the replicating portfolio, we solve the system of equations (2.1.7)*

$$\begin{cases} \theta_0 + 5.8\theta_1 & = g(5.8) = .8 \\ \theta_0 + 3.4\theta_1 & = g(3.4) = .6 \end{cases}$$

to obtain $\theta_0 = \frac{1.9}{6}$ and $\theta_1 = \frac{1}{12}$. The price can be found in two ways: either by using the

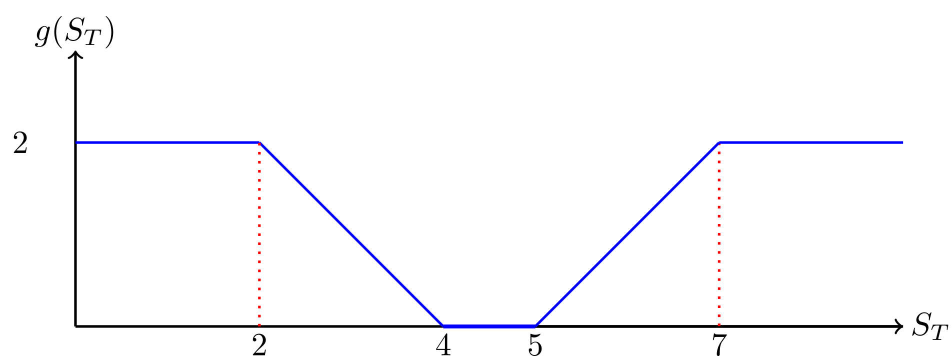


Figure 2.1.12: Payoff of Example 2.1.7

replicating portfolio or by the risk-neutral probability. The former gives the price by the

[3]In the context of this exercise, the forward can be replaced by a futures contract.
[4]E.g. electricity.

value of the replicating portfolio:

$$\frac{\theta_0}{1+R} + \theta_1 S_0 = \frac{1.9}{6(1.05)} + (4)\frac{1}{12} = \frac{2}{3.15}.$$

For the latter, we use the risk-neutral probability $\hat{\pi}_u = \frac{1}{3}$ *and* $\hat{\pi}_\ell = \frac{2}{3}$ *to obtain*

$$\frac{1}{1+R}\hat{\mathbb{E}}[g(S_1)] = \frac{1}{1+R}\left(\hat{\pi}_u g(S_0 u) + \hat{\pi}_\ell g(S_0 \ell)\right) = \frac{1}{1.05}\left(\frac{1}{3}(.8) + \frac{2}{3}(.6)\right) = \frac{2}{3.15}.$$

2.1.4 One-period trinomial model

In a one-period trinomial model, $M = 3$, $N = 2$, $S_0 := p_1$, $u := P_{1,1}/S_0$, $m := P_{1,2}/S_0$ and $\ell := P_{1,3}/S_0$, where S_0, $P_{1,1}$, $P_{1,2}$ and $P_{1,3}$ are all positive real numbers. By Theorem

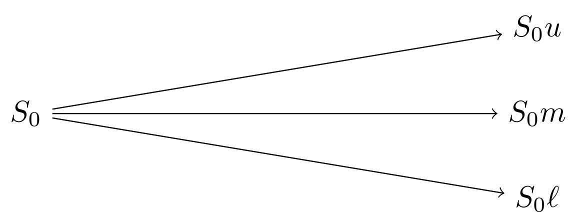


Figure 2.1.13: One-period trinomial model

2.1.1, no-arbitrage condition is equivalent to the existence of a positive probability vector $\hat{\pi} = (\hat{\pi}_\ell, \hat{\pi}_m, \hat{\pi}_u)$ such that

$$\begin{cases} \ell\hat{\pi}_\ell + m\hat{\pi}_m + u\hat{\pi}_u = 1+R \\ \hat{\pi}_\ell + \hat{\pi}_m + \hat{\pi}_u = 1. \end{cases} \tag{2.1.8}$$

It is not hard to see that the no-arbitrage condition has the same condition as the one-period binomial model, i.e., $\ell < 1 + R < u$. The intersection of two planes with equations (2.1.8) in $\mathbb{R}^3$ is a line parametrized by

$$\begin{cases} \hat{\pi}_\ell = \frac{u-(1+R)}{u-\ell} - \frac{m-\ell}{u-\ell}t \\ \hat{\pi}_m = t \\ \hat{\pi}_u = \frac{1+R-\ell}{u-\ell} - \frac{u-m}{u-\ell}t \end{cases}$$

The no-arbitrage condition is equivalent to the existence of a segment of this line in the positive octane; i.e., there exists a t such that

$$\begin{cases} \hat{\pi}_\ell = \frac{u-(1+R)}{u-\ell} - \frac{m-\ell}{u-\ell} t > 0 \\ \hat{\pi}_m = t > 0 \\ \hat{\pi}_u = \frac{1+R-\ell}{u-\ell} - \frac{u-m}{u-\ell} t > 0 \end{cases},$$

which is guaranteed if and only if $0 < \min\left\{\frac{1+R-\ell}{u-m}, \frac{u-(1+R)}{m-\ell}\right\}$. Or equivalently, $\ell < 1+R < u$. Notice that the positive segment is given by $0 < t < \min\left\{\frac{1+R-\ell}{u-m}, \frac{u-(1+R)}{m-\ell}\right\}$.

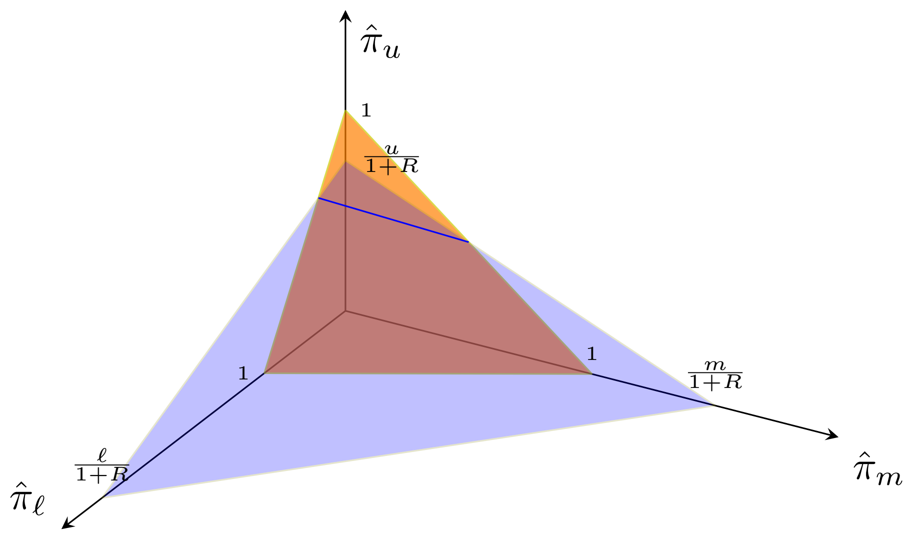


Figure 2.1.14: The positive segment of line given by (2.1.8) when $\ell < 1 + R < u$

Exercise 2.1.8. *Derive the no-arbitrage condition for the multinomial model in Figure 2.1.15. Here, $u_1 < u_2 < \cdots < u_M$ are positive numbers.*

2.1.5 Replication and complete market

A *contingent claim* (or simply a claim) on an underlying asset S is a new asset with a payoff given by a function $g : \mathbb{R}_+ \to \mathbb{R}$ on the price S_T of the underlying asset at maturity T; i.e., the payoff is $g(S_T)$. Call and put options are examples of contingent claims with payoff functions $g(S_T) := (S_T - K)_+$ and $g(S_T) := (K - S_T)_+$, respectively. A replicating

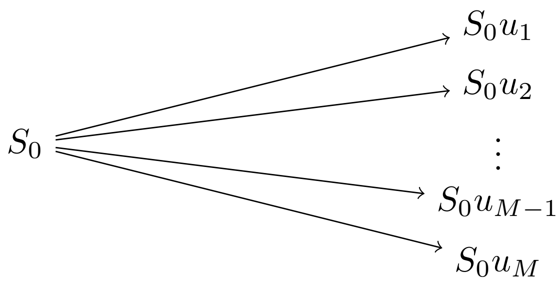


Figure 2.1.15: The description of the asset price in the multinomial model

(or hedging) portfolio for a contingent claim is a portfolio with the same future value as the payoff of the claim at **all** states of the market. For example, in the binomial model in Section 2.1.3, any arbitrary claim can be replicated. More precisely, a payoff $g(S_0u)$ and $g(S_0\ell)$ for states u and ℓ, respectively, is replicated by a portfolio (θ_0, θ_1), given by

$$\theta_0 = \frac{ug(S_0\ell) - \ell g(S_0u)}{u - \ell} \quad \text{and} \quad \theta_1 = \frac{g(S_0u) - g(S_0\ell)}{S_0(u - \ell)}.$$

Contrary to the binomial model, in the trinomial model, several claims may not be replicable. For instance, the replication of a call option with $K \in [S_0m, S_0u)$ leads to the following system of three equations and two unknowns:

$$\begin{cases} \theta_0(1+R) + \theta_1 S_0 u & = S_0u - K \\ \theta_0(1+R) + \theta_1 S_0 m & = 0 \\ \theta_0(1+R) + \theta_1 S_0 \ell & = 0 \end{cases},$$

which obviously does not have any solutions for (θ_0, θ_1). A market model in which **every** claim is replicable is called a *complete market*. A binomial model is a complete market, whereas a trinomial model is an *incomplete market*.

For a general Arrow-Debreu market model, the condition of completeness is expressed in the following theorem.

Theorem 2.1.2. *Assume that there is no arbitrage, i.e., there exits a risk-neutral probability $\hat{\pi}$. Then, the market is complete if and only if there is a* ***unique*** *risk-neutral probability; i.e., if and only if the system of linear equation* (2.1.1)

$$p = \mathbf{P}\pi$$

has a unique positive solution.

While in the binomial model there is only one risk-neutral probability and therefore the market is complete, in the trinomial model there are infinitely many risk-neutral probabilities and therefore the market is incomplete.

2.1.6 Superreplication and model risk

Replication (or hedging) is a normal practice for the issuer of an option to manage the risk of issuing that option. When the market is not complete, one cannot *perfectly* replicate all claims and the issuer of a claim should take another approach: a nonperfect replication. In practice, the replication starts only after pricing the claim. The issuer first picks up a *pricing model*, i.e., a risk-neutral probability $\hat{\pi}^* = (\hat{\pi}_1^*, ..., \hat{\pi}_M^*)^\intercal$, to price the claim by $\frac{1}{1+R}\hat{\mathbb{E}}^*[g(S_T)]$. Then, she tries to use the fund raised by selling the option to find a nonperfect replication strategy. Genuinely, the higher the price of the claim, the less the issuer is exposed to the risk. Therefore, the chosen risk-neutral probability $\hat{\pi}^*$ to price the claims represents some level of exposure to the risk. In this section, we would like to provide a method to measure this risk, namely *model risk*.

The choice of a replication strategy usually depends on many variables, including the risk preference of the issuer, which are outside the context of this section. However, for the purpose of model risk, we introduce one specific choice of a nonperfect replication strategy, namely *superreplication*. A superreplication strategy prepares the issuer for the worst-case scenario. The *superreplication price* of an option is defined as the cheapest price of a portfolio that generates a payoff greater than or equal to the payoff the option for **all** states of the market. In the Arrow-Debreu market model for asset shown in Figure 2.1.15, we want to make a portfolio (θ_0, θ_1) such that

$$\theta_0 + \theta_1 S_0 u_j \geqslant g(S_0 u_j) \quad \text{for all} \quad j = 1, ..., M.$$

Then, among all such portfolios we want to choose the one that has the least cost, i.e., $\min \frac{\theta_0}{1+R} + \theta_1 S_0$. For instance, in the trinomial model, the superreplication price of an option with payoff $g(S_1)$ is defined by

$$\min \frac{\theta_0}{1+R} + \theta_1 S_0 \tag{2.1.9}$$

over all θ_0 and θ_1 subject to the constraints

$$\begin{cases} \theta_0 + \theta_1 S_0 u \geqslant g(S_0 u) \\ \theta_0 + \theta_1 S_0 m \geqslant g(S_0 m) \\ \theta_0 + \theta_1 S_0 \ell \geqslant g(S_0 \ell) \end{cases} . \tag{2.1.10}$$

The superreplication price of a claim is the **smallest value** that enables the issuer to build a portfolio which **dominates** the payoff of the claim, in other words, to remove all risk exposure from issuing the claim.

For simplicity, we build the rest of this section in the context of the trinomial model. However, extension to multinomial model is straightforward. Subreplication price can be defined similarly by

$$\max \frac{\theta_0}{1+R} + \theta_1 S_0 \tag{2.1.11}$$

over all θ_0 and θ_1 subject to the constraints

$$\begin{cases} \theta_0 + \theta_1 S_0 u \leqslant g(S_0 u) \\ \theta_0 + \theta_1 S_0 m \leqslant g(S_0 m) \\ \theta_0 + \theta_1 S_0 \ell \leqslant g(S_0 \ell) \end{cases} . \tag{2.1.12}$$

Sub or superreplication is a linear programming problem that can be solved using some standard algorithms. However, for the trinomial model, the solution is simple: for superreplication the minimum is attained in one of the at most three points of intersection between the lines

$$\begin{cases} \theta_0 + \theta_1 S_0 u = g(S_0 u) \\ \theta_0 + \theta_1 S_0 m = g(S_0 m) \\ \theta_0 + \theta_1 S_0 \ell = g(S_0 \ell) \end{cases} .$$

As shown in Figure 2.1.16, we only need to

(1) find these three points,

(2) exclude those that do not satisfy the inequalities (2.1.10), and

(3) check which one of the remaining yields the smallest value for $\frac{\theta_0}{1+R} + \theta_1 S_0$.

Example 2.1.8. *In trinomial model, let $S_0 = 1$, $R = .5$, $u = 2$, $m = 1$, and $\ell = 1/2$. Consider an option with payoff g shown in Figure 2.1.17.*

a) Notice that payoff $g(2) = 0$ and $g(1) = g(1/2) = 1$. We shall find the superreplication price for this option.

$$\min \frac{2\theta_0}{3} + \theta_1$$

over all θ_0 and θ_1 subject to the constraints

$$\begin{cases} \theta_0 + 2\theta_1 \geqslant 0 \\ \theta_0 + \theta_1 \geqslant 1 \\ \theta_0 + \frac{\theta_1}{2} \geqslant 1 \end{cases} .$$

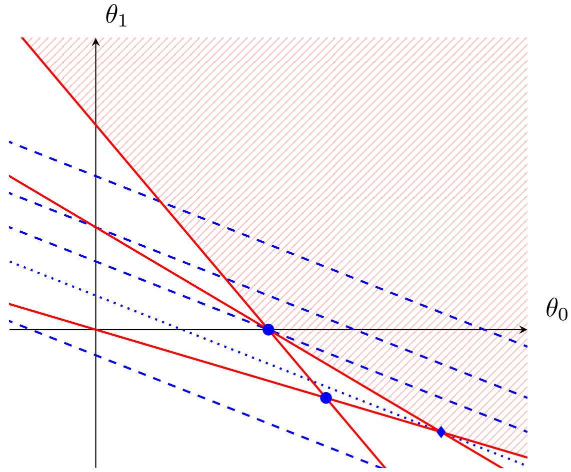


Figure 2.1.16: The linear programing problem for superhedging. The hatched region is determined by constraints (2.1.10). The dashed lines are the contours of the linear function $\frac{\theta_0}{1+R} + \theta_1 S_0$ in (2.1.9). The point marked by $\diamond$ is where the minimum is attained.

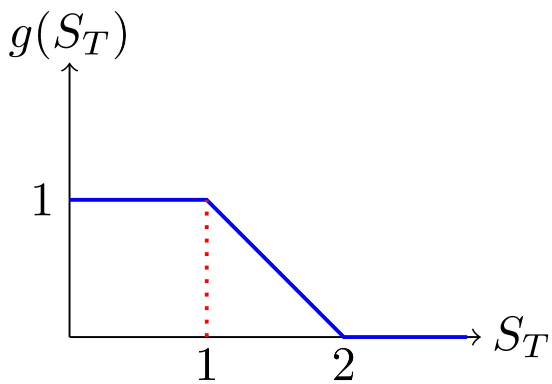


Figure 2.1.17: Payoff of Example 2.1.8

This linear programing problem matches the one sketched in Figure 2.1.16 with the minimizer given by $(2, -1)$. Therefore, the superreplication price is given by $\frac{1}{3}$.

b) *We next find all risk-neutral probabilities. Notice that any risk-neutral probability $\hat{\pi} = (\hat{\pi}_u, \hat{\pi}_m, \hat{\pi}_\ell)^\mathsf{T}$ satisfies*

$$\begin{cases} 2\hat{\pi}_u + \hat{\pi}_m + \frac{\hat{\pi}_\ell}{2} = 3/2 \\ \hat{\pi}_u + \hat{\pi}_m + \hat{\pi}_\ell = 1 \\ \hat{\pi}_u, \hat{\pi}_m, \hat{\pi}_\ell > 0 \end{cases}$$

By eliminating $\hat{\pi}_m$, we obtain $\hat{\pi}_u = \frac{1}{2} + \frac{\hat{\pi}_\ell}{2}$. Then, we parametrize the line of intersection of the two planes $2\pi_u + \pi_m + \frac{\pi_\ell}{2} = 3/2$ and $\hat{\pi}_u + \hat{\pi}_m + \hat{\pi}_\ell = 1$.

$$\begin{cases} \hat{\pi}_u = \frac{1}{2} + \frac{t}{2} \\ \hat{\pi}_m = \frac{1}{2} - \frac{3t}{2} \\ \hat{\pi}_\ell = t \end{cases}$$

It is easy to see that $\pi_u, \pi_m, \pi_\ell > 0$ if and only if $t \in (0, \frac{1}{3})$. The positive segment is shown in Figure 2.1.18.

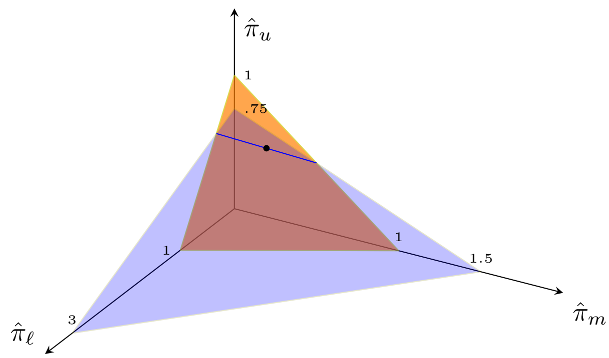


Figure 2.1.18: The positive segment of the intersection of two lines given by $\hat{\pi}_u u + \hat{\pi}_m m + \hat{\pi}_\ell \ell = 1 + R$ and $\hat{\pi}_u + \hat{\pi}_m + \hat{\pi}_\ell = 1$. The mark on the segment represents a chosen pricing model $\hat{\pi}^*$.

c) *Next, we find the range of prices for the option described in Part (a) generated by different risk-neutral probabilities. Notice that in Part (b), all risk-neutral probabilities are generated by a single parameter $t \in (0, \frac{1}{3})$. Thus, the risk-neutral price of this option is given by*

$$\frac{2}{3}\left(\left(\frac{1}{2} + \frac{t}{2}\right)g(2) + \left(\frac{1}{2} - \frac{3t}{2}\right)g(1) + tg(1/2)\right) = \frac{1}{3} - t + \frac{2t}{3} = \frac{1}{3} - \frac{t}{3}$$

As t changes in $(0, \frac{1}{3})$, the price changes in $(\frac{2}{9}, \frac{1}{3})$.

Notice that in Example 2.1.8, the superreplication price $\frac{1}{3}$ is the same as the supremum of the price range given by risk-neutral probabilities $(\frac{2}{9}, \frac{1}{3})$. This is not a coincidence. One can see the relation between the superreplication problem and risk neural pricing through the linear programming duality in Theorem A.3.

The problem of superreplication is a linear programming problem (A.3) that has a dual problem given by (A.4). Theorem A.3 suggests that both problems have the same value. In the context of superreplication for the trinomial model, the dual problem is given by

$$\max \pi_u g(S_0 u) + \pi_m g(S_0 m) + \pi_\ell g(S_0 m)$$

over all π_ℓ, π_m, and π_u subject to the constraints

$$\begin{cases} \pi_u S_0 u + \pi_m S_0 m + \pi_\ell S_0 m = S_0 \\ \pi_u + \pi_m + \pi_\ell = \frac{1}{1+R} \\ \pi_u, \pi_m, \pi_\ell \geqslant 0 \end{cases}$$

By change of variable $\hat{\pi} = (1+R)\pi$, the dual problem turns into

$$\mathcal{P}(g) := \frac{1}{1+R} \max \ \hat{\pi}_u g(S_0 u) + \hat{\pi}_m g(S_0 m) + \hat{\pi}_\ell g(S_0 m), \quad \text{subject to} \quad \begin{cases} \hat{\pi}_u u + \hat{\pi}_m m + \hat{\pi}_\ell \ell = 1+R \\ \hat{\pi}_u + \hat{\pi}_m + \hat{\pi}_\ell = 1 \\ \hat{\pi}_u, \hat{\pi}_m, \hat{\pi}_\ell \geqslant 0 \end{cases}. \tag{2.1.13}$$

Given no-arbitrage condition, the line of intersection of two planes $\hat{\pi}_u u + \hat{\pi}_m m + \hat{\pi}_\ell \ell = 1+R$ and $\hat{\pi}_u + \hat{\pi}_m + \hat{\pi}_\ell = 1$ has a segment in the positive octane, shown in Figure 2.1.14. Then, the value of $\mathcal{P}(g)$ in (2.1.13) is attained at one of endpoints of this segment. This segment represents the set of all risk-neutral probabilities. According to Theorem 2.1.1, the lack of weak arbitrage implies that this segment is inside the first octane with to endpoints on two different coordinate planes, and the lack of strong arbitrage only implies that the endpoints of the segment are on the coordinate planes, possibly the same coordinate plane.

Let's recall what we presented at the beginning of this section: the issuer of the option chooses a risk-neutral probability $\hat{\pi}^*$ inside the positive segment to price the option. The endpoints of this segment, one of which maximizes and the other minimizes the value $\frac{1}{1+R} \max \ \hat{\pi}_u g(S_0 u) + \hat{\pi}_m g(S_0 m) + \hat{\pi}_\ell g(S_0 m)$, provide the super and subreplication price of the option with payoff g, respectively. While the issuer has priced the claim by

$$\frac{1}{1+R} \hat{\mathbb{E}}^*[g(S_1)],$$

to completely cover the risk, he needs $\mathcal{P}$ in (2.1.13). Therefore, he is short as much as

$$\Theta(\hat{\pi}^*) := \mathcal{P}(g) - \frac{1}{1+R} \hat{\mathbb{E}}^*[g(S_1)]$$

in order to cover the risk of issuing the claim. The value $\Theta(\hat{\pi}^*)$ is called *superreplication* *model risk measure*indexsuperreplication model risk measure.

Example 2.1.9. *In Example 2.1.8, if the pricing probability* $\hat{\pi}^* = (7/12, 1/4, 1/6)^\mathsf{T}$ *is chosen, then the superreplication model risk is measured by*

$$\mathcal{P}(g) - \frac{1}{1+R}\hat{\mathbb{E}}^*[g(S_1)] = \frac{1}{3} - \frac{2}{3}\left(\frac{7}{12}g(2) + \frac{1}{4}g(1) + \frac{1}{6}g(1/2)\right) = \frac{1}{3} - \frac{5}{18} = \frac{1}{18}.$$

Exercise 2.1.9. *In the trinomial model, let* $S_0 = 1$*,* $R = 0$*,* $u = 2$*,* $m = 1$*, and* $\ell = 1/2$*.*

a) *Find all the risk-neutral probabilities and the range of prices generated by them for a call option with strike* $K = 1$*.*

b) *Find the superreplication price and sub-replication price for this call option and compare them to the lowest and highest prices in Part (a).*

Example 2.1.10. *In Example 2.1.8, if we modify the yield by setting* $R = 1$*, then there will be an arbitrage. This can be seen through the absence of risk-neutral probabilities. On the other hand, the superreplication problem is still feasible, i.e., problem (2.1.11) with constraint (2.1.10) still has a finite value. But this value is zero. This is because we minimize* $\frac{\theta_0}{2} + \theta_1$ *subject to* $\theta_0 + 2\theta_1 \geqslant 0$*,* $\theta_0 + \theta_1 \geqslant 1$ *and* $\theta_0 + \frac{\theta_1}{2} \geqslant 1$*, which obviously takes minimum value zero.*

On the other hand, if we set $R > 1$*, then the superreplication problem is not feasible anymore and the minimum is* $-\infty$*.*

Example 2.1.11. *Consider the trinomial model with* $S_0 = 4$*,* $R = .05$*,* $u = 1.45$*,* $m = 1.25$ *and* $\ell = .85$*. To price payoff* g *in Figure 2.1.12, the risk-neutral probability* $\hat{\pi} = (\hat{\pi}_u, \hat{\pi}_m, \hat{\pi}_\ell)^T = (5/18, 1/12, 23/36)^T$ *has been chosen. In order to find the model risk, we find the superreplication price by solving the following linear programming problem.*

$$\min \frac{\theta_0}{1.05} + 4\theta_1 \quad \textit{subject to} \quad \begin{cases} \theta_0 + 5.8\theta_1 & \geqslant g(5.8) = .8 \\ \theta_0 + 5\theta_1 & \geqslant g(5) = 0 \\ \theta_0 + 3.4\theta_1 & \geqslant g(3.4) = .6 \end{cases}$$

The minimizer is given by $\theta_0 = \frac{19}{6}$ *and* $\theta_1 = \frac{1}{12}$*, and the superreplication price is given by* $\frac{2}{3.15}$*. Then, the model risk is* $\Theta(g) := \frac{1}{1+R}\hat{\mathbb{E}}[g(S_1)] - \mathcal{P}$*.*

$$\Theta(g) = \frac{2}{3.15} - \frac{1}{1+R}\left(\hat{\pi}_u g(S_0 u) + \hat{\pi}_m g(S_0 m) + \hat{\pi}_\ell g(S_0 \ell)\right) = \frac{2}{3.15} - \frac{1}{1.05}\left(\frac{5}{18}(.8) + \frac{1}{12}(0) + \frac{23}{36}(.6)\right)$$

We leave the treatment of the general multinomial model as an exercise.

Exercise 2.1.10. *Write the linear programming problem associated with the superreplication of a contingent claim in the multinomial model in Exercise 2.1.8 and its dual.*

Remark 2.1.6. *The superhedging price and therefore the model risk is not linear in the payoff. The superreplication price of payoff $g_1 + g_2$ is less than the sum of the superreplication price of payoff g_1 and the superreplication price of payoff g_2. To see this, recall that, by Theorem A.3, the superreplication price of a payoff g is*

$$\mathcal{P}(g) := \sup\left\{\frac{1}{1+R}\hat{\mathbb{E}}[g(S_1)] \ : \ \text{over all risk-neutral probabilities} \quad \hat{\pi}\right\}.$$

Because $\sup\{f(x) + g(x)\} \leqslant \sup f(x) + \sup g(x)$, *we have*

$$\mathcal{P}(g_1 + g_2) \leqslant \mathcal{P}(g_1) + \mathcal{P}(g_2).$$

2.2 Multiperiod discrete-time markets

2.2.1 A discussion on the sample space for multiperiod asset price process

Consider a discrete-time market with time horizon T in which trading occurs only at time $t = 0, ..., T$, and there are $d + 1$ assets. We shall denote the price of asset i at time $t = 0, ..., T$ by $S_t^{(i)}$, for $i = 0, ..., d$. The price of each risky asset in a multiperiod market model is a *stochastic process*; this means for each $t = 1, ..., T$, the price at time t, denoted by S_t is a random variable. In this section, we set up a sample space to host stochastic processes that represent the price of assets in a discrete-time market.

For a single period market with one asset whose price is a random variable with discrete values, a finite or countably infinite sample space which matches the states to the values of a random variable may suffice. For instance, if the future price S_1 of an asset is given by three values 2, 1, and 0.5, one can assume a sample space with three outcomes, namely $\Omega = \{\omega_1, \Omega_2, \omega_3\}$ with $S_1(\omega_1) = 2$, $S_1(\omega_2) = 1$, and $S_1(\omega_3) = 0.5$. When we have a multiperiod market with a single asset, the price of the asset is a stochastic process, more than one random variable, and we need a more complicated construction for sample space. The following examples elaborate on this situation.

Example 2.2.1. *Let the price S_t of an asset at time $t = 1, 2$ be given by the diagram in the figure below. The random variable S_1 takes values 1 and -1 and the random variable*

$$S_{1,1} = 1 \xrightarrow{\omega_1} S_{2,1} = 2$$

$$S_{1,2} = 1 \xrightarrow{\omega_2} S_{2,2} = 0$$

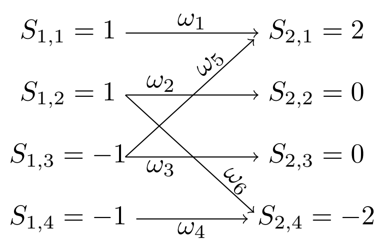
$$S_{1,3} = -1 \xrightarrow{\omega_3} S_{2,3} = 0$$

$$S_{1,4} = -1 \xrightarrow{\omega_4} S_{2,4} = -2$$

S_1 takes values 2, 0, and -1. However, a suitable sample to host both random variables is one with only four members, $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ with $S_1(\omega_1) = 1$, $S_2(\omega_1) = 2$, $S_1(\omega_2) = 1$, $S_2(\omega_2) = 0$, $S_1(\omega_3) = -1$, $S_2(\omega_3) = 0$, and $S_1(\omega_4) = -1$, $S_2(\omega_4) = -2$. This sample space is made up of all possible states of the market; here there are four states that are shown with arrows. These states are also called the paths of the price process, which represents the evolution of the price process in time for each outcome.

The above example is a standard random walk with two periods; see Example B.13.

Example 2.2.2. *We modify Example 2.2.1 as shown below. Unlike Example 2.2.1, a*



sample space needs more than four members, exactly six. There are six different paths that the price of the asset can evolve in time. $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$ with $S_1(\omega_1) = 1$, $S_2(\omega_1) = 2$, $S_1(\omega_2) = 1$, $S_2(\omega_2) = 0$, $S_1(\omega_3) = -1$, $S_2(\omega_3) = 0$, and $S_1(\omega_4) = -1$, $S_2(\omega_4) = -2$, $S_1(\omega_5) = -1$, $S_2(\omega_5) = 2$, and $S_1(\omega_6) = 1$, $S_2(\omega_6) = -2$.

In a more general case, a sample space needs to at least have the set of all probable *sample paths*superreplication model risk measure. of the price process. In the Arrow-Debreu model, the set of all sample paths is the set of all states of the market, regardless of the number of assets. In a market with T periods and one asset that takes values V_t at time t with $t = 0, ...T$, we can have the sample space

$$\Omega = \big\{(x_1, ..., x_T) \,:\, x_t \in V_t, \;\; t = 1, ..., T\big\}.$$

This sample space have $\prod_{t=1}^{T} V_t$ samples paths. We can use the sample-path methodology to write one single sample space for both Example 2.2.1 and Example 2.2.2, namely

$$\Omega = \big\{(a, b) \,:\, a = -1 \text{ or } 1, \text{ and } b = -2, 0, \text{ or } 2\big\}.$$

For a market with T periods and $d + 1$ assets such that at time t the vector of assets $(S_t^{(0)}, ..., S_t^{(d)})$ takes values in the set $V_t \subseteq \mathbb{R}_+^{d+1}$ for each $t = 0, ...T$, the set of all sample paths is the collection of all functions of the form

$$\omega : \{1, ..., T\} \to V_1 \times \cdots \times V_T$$

Notice that some of the sample paths may have probability zero. In Example 2.2.1, the sample path $(-1, 2)$ has probability zero, but in Example 2.2.2, no sample path has probability zero.

A unifying approach to set a sample space for the asset price process is by extending the set of values of the price to include all positive real numbers, $V_t = \mathbb{R}_+^{d+1}$, $\Omega = \prod_{t=1}^T V_t = (\mathbb{R}_+^{d+1})^T$[5] equipped with the Borel σ-field $\mathcal{B}((\mathbb{R}_+^{d+1})^T)$. The random variable $S_t^{(i)}$ which represents the price of asset i at time t is then defined by the *canonical mapping*

$$S_t^{(i)} : \omega = (\omega_1, ..., \omega_T) \in (\mathbb{R}_+^{d+1})^T \mapsto \omega_t^{(i)}, \text{ where } \omega_t = (\omega_t^{(0)}, ..., \omega_t^{(d)}) \in \mathbb{R}_+^{d+1}.$$

Here, we assume that the price of an asset only takes positive values $\omega_t^{(i)} \in \mathbb{R}_+$. If it takes negative values, then we can extend the sample space to $\Omega := (\mathbb{R}^{d+1})^T$. If asset 0 is a risk-free asset, then we can remove its contribution in the sample space and write $\Omega := (\mathbb{R}_+^d)^T$.

The choice of canonical space allows us to cover all types of models for asset price with discrete or nondiscrete distribution. For the rest of this section, we do not need to emphasize on choice the sample space, **we assume that the sample space is finite**. Some of the result can be generalized to countably infinite or even uncountable sample spaces. But, the treatment of such cases needs more advance tools form martingale theory.

2.2.2 Arbitrage and trading in multiperiod discrete-time markets

Consider a market with multiple assets. There are three ways to represent a portfolio that is made up of these assets: based on the proportion of each asset in the portfolio, the number of units of each asset in the portfolio, and the value invested in each asset. We start to define a portfolio based on the proportion, because it is easier to understand, then we provide the equivalent representations on the number of units of each asset in the portfolio and the value invested in each asset.

Portfolio

A *self-financing portfolio*, or simply a portfolio, is represented by the sequence of vectors $\theta_t = (\theta_t^{(0)}, ..., \theta_t^{(d)})^\intercal$ for $t = 0, ..., T-1$ with $\sum_{i=0}^d \theta_t^{(i)} = 1$, for $t = 0, ..., T-1$. Here, $\theta_t^{(i)}$ is the ratio of the value of portfolio invested in asset i at time t; $\theta_t^{(i)} = \frac{W^{(i)}}{W_t}$, where $W_t^{(i)}$ is the proportion of the value of portfolio invested in asset i at time t and W_t is the total value of the portfolio at time t. Equivalently, $W_t^{(i)} = \Delta_t^{(i)} S_t^{(i)}$, where $\Delta_t^{(i)}$ is the number of

[5] We prefer to write $(\mathbb{R}_+^{d+1})^T$ and not $\mathbb{R}_+^{(d+1)T}$.

shares of asset i in the portfolio and $S_t^{(i)}$ is the current price of asset i. More precisely,

$$W_t = \sum_{i=0}^{d} W_t^{(i)} = \sum_{i=0}^{d} \Delta_t^{(i)} S_t^{(i)} \text{ and } W_t^{(i)} = \theta_t^{(i)} W_t = \Delta_t^{(i)} S_t^{(i)}.$$

$\theta_t^{(i)}$, $\Delta_t^{(i)}$, and $W_t^{(i)}$ can be positive, negative, or zero, representing long, short, or no-investment positions in asset i, respectively. In addition, they do not need to be deterministic; in general, a portfolio strategy can be a sequence of random vector that depends on the information obtained from the assets' prices before or at time t. In other words, $\theta_t^{(i)}$ is a function of random variables $S_i^{(j)}$ for $i = 1, ..., t$ and $j = 1, ..., d$. More rigorously:

Definition 2.2.1. *A self-financing portfolio strategy is given by a sequence of vector functions* $\theta_t = (\theta_t^{(0)}, ..., \theta_t^{(d)})^\top$ *such that* $\theta_0 \in \mathbb{R}^d$ *is a real vector and for any* $t = 1, ..., T-1$, $\theta_t = (\theta_t^{(0)}, ..., \theta_t^{(d)})^\top$ *is a function that maps*

$$\mathbf{S}_t := \begin{bmatrix} S_1^0 & \cdots & S_t^0 \\ \vdots & \ddots & \vdots \\ S_1^d & \cdots & S_t^d \end{bmatrix} \tag{2.2.1}$$

into a vector in $\mathbb{R}^d$ *that satisfies*

$$\sum_{i=0}^{d} \theta_t^{(i)} = 1, \quad \textit{for} \quad t = 0, ..., T-1.$$

In other words, the function $\theta_t^{(i)}$ depends only on the prices of all assets from time $t = 0$ until time t, not the future prices at points $t+1, ..., T$ in time. This is in line with the intuition that a portfolio strategy can only depend on the information gathered up to the present time.

In different applications, different representation of a portfolio strategy are proved useful. For example in the portfolio theory, it is easier to write the portfolio in terms of the ratio, while in the replicating portfolio of an option, the number of shares in each asset happens to provide a more convenient representation. If Δ_t is known, the value of portfolio at time t is given by

$$W_t = \sum_{i=0}^{d} \Delta_t^{(i)} S_t^{(i)}.$$

Then, the value invested in asset i changes by $\Delta_t^{(i)}(S_{t+1}^{(i)} - S_t^{(i)})$ from time t to time $t+1$.

Time	# of units of asset (j)	Value of the portfolio
t	$\Delta_t^{(j)} = \frac{\theta_t^{(j)} W_t}{S_t^{(j)}}$	$W_t = \Delta_t^{(0)} S_t^{(0)} + \cdots + \Delta_t^{(d)} S_t^{(d)}$
$t+1$ before rebalancing	$\frac{\theta_t^{(j)} W_{t+1}}{S_{t+1}^{(j)}}$	$W_{t+1} = \Delta_t^{(0)} S_{t+1}^{(0)} + \cdots + \Delta_t^{(d)} S_{t+1}^{(d)}$
$t+1$ after rebalancing	$\Delta_{t+1}^{(j)} = \frac{\theta_{t+1}^{(j)} W_{t+1}}{S_{t+1}^{(j)}}$	$W_{t+1} = \Delta_{t+1}^{(0)} S_{t+1}^{(0)} + \cdots + \Delta_{t+1}^{(d)} S_{t+1}^{(d)}$

Table 2.1: Rebalancing a portfolio strategy from time t to time $t+1$.

Therefore, the total wealth at time $t+1$ changes to

$$W_{t+1} = \sum_{i=0}^{d} \Delta_t^{(i)} S_{t+1}^{(i)},$$

and the change in the value of portfolio is given by

$$W_{t+1} - W_t = \sum_{i=0}^{d} \Delta_t^{(i)} (S_{t+1}^{(i)} - S_t^{(i)}), \tag{2.2.2}$$

However, at time $t+1$, we need to change the investment ratio θ_t to a different value θ_{t+1}, and as a consequence, Δ_t should be changed to Δ_{t+1} accordingly. Since the sum of ratios is always one,

$$\sum_{i=0}^{d} \theta_{t+1}^{(i)} = 1, \quad \text{for} \quad t = 0, ..., T-1,$$

and $\theta_t^{(i)} = \frac{\Delta_t^{(i)} S_t^{(i)}}{W_t}$, the vector Δ_{t+1} must satisfy

$$\sum_{i=0}^{d} \Delta_{t+1}^{(i)} S_{t+1}^{(i)} = W_{t+1} = \sum_{i=0}^{d} \Delta_t^{(i)} S_{t+1}^{(i)}, \quad \text{for} \quad t = 0, ..., T-1.$$

Stochastic integral: discrete-time markets

For ease of calculation, we consider the case $d = 1$, i.e., $i = 0$ and 1. Recall from (2.2.2) that the change in the value of the portfolio satisfies

$$\underbrace{W_{t+1} - W_t}_{\text{change in the value of the portfolio}} = \underbrace{\Delta_t^{(0)} (S_{t+1}^{(0)} - S_t^{(0)})}_{\text{change due to risky asset (0)}} + \underbrace{\Delta_t^{(1)} (S_{t+1}^{(1)} - S_t^{(1)})}_{\text{change due to risky asset (1)}}.$$

Given initial wealth W_0, we sum up the above telescopic summation to obtain

$$W_t = W_0 + \sum_{i=0}^{t-1} \Delta_i^{(0)}(S_{i+1}^{(0)} - S_i^{(0)}) + \sum_{i=0}^{t-1} \Delta_i^{(1)}(S_{i+1}^{(1)} - S_i^{(1)}).$$

In the right-hand side above, either of the summations corresponds to the cumulative changes in the value of the portfolio due to investment in one of the assets.

If asset $S^{(0)}$ denotes a risk-free asset with discrete yield R in the time period from i to $i+1$, then $S_{i+1}^{(0)} - S_i^{(0)} = RS_i^{(1)}$. On the other hand, $\Delta_i^{(0)} S_i^{(0)} = W_i - \Delta_i^{(1)} S_i^{(1)}$. Therefore,

$$\Delta_i^{(0)}(S_{i+1}^{(0)} - S_i^{(0)}) = R(W_i - \Delta_i^{(1)} S_i^{(1)}).$$

In this case, we can simply drop the superscript of the risky asset to write

$$\underbrace{W_{t+1} - W_t}_{\text{change in the value of the portfolio}} = \underbrace{R(W_t - \Delta_t S_t)}_{\text{change due to risk-free asset}} + \underbrace{\Delta_t(S_{t+1} - S_t)}_{\text{change due to risky asset}}.$$

Therefore, the total wealth satisfies

$$W_t = W_0 + R\sum_{i=0}^{t-1}(W_i - \Delta_i S_i) + \sum_{i=0}^{t-1} \Delta_i(S_{i+1} - S_i). \tag{2.2.3}$$

In (2.2.3), the first summation is the cumulative changes in the value of the portfolio due to investment in the risk-free asset, and the second summation is the cumulative investment in the portfolio due to changes in the risky asset.

An important consequence of this formula is that a self-financing portfolio is only characterized by trading strategy $\Delta = (\Delta_0, ..., \Delta_{T-1})$ and the initial wealth W_0, since there is no inflow and outflow of cash to or from the portfolio. The term

$$(\Delta \cdot S)_t := \sum_{i=0}^{t-1} \Delta_i(S_{i+1} - S_i)$$

is called a discrete stochastic integral[6].

Exercise 2.2.1. *Let $\tilde{W}_t := (1+R)^{-t} W_t$ and $\tilde{S}_t := (1+R)^{-t} S_t$ be, respectively, discounted wealth process and discounted asset price. Then, show that*

$$\tilde{W}_t = W_0 + \sum_{i=0}^{t-1} \Delta_i(\tilde{S}_{i+1} - \tilde{S}_i), \quad \hat{W}_0 = W_0.$$

[6] According to Philip Protter, this notation was devised by the prominent French probabilist Paul-André Meyer to simplify the task of typing with old-fashioned typewriters.

To understand the meaning of stochastic integral $(\Delta \cdot S)_t$, we provide the following example.

Example 2.2.3. *Recall from Example B.13 that a random walk W is the wealth of a player in a game of chance in which he wins or loses \$1 in each round based on the outcome of flipping a coin. If we denote the amount of bet of the player in round i by Δ_{i-1}, then the total wealth $\mathcal{W}^\Delta$ from the betting strategy $\Delta = (\Delta_0, \Delta_1, ...)$ in t rounds is given by*

$$\mathcal{W}_t^\Delta = \mathcal{W}_0^\Delta + \sum_{i=0}^{t-1} \Delta_i \xi_{i+1}$$

where $\{\xi_i\}_i$ is a sequence of i.i.d. random variables with values 1 and -1. Since $\xi_i = W_{i+1} - W_i$, we have

$$\mathcal{W}_t^\Delta = \sum_{i=0}^{t-1} \Delta_i (W_{i+1} - W_i) = (\Delta \cdot W)_t.$$

In particular, if $\Delta_i \equiv 1$, $\mathcal{W}^\Delta = W$ is merely a random walk.

Example 2.2.4 (Saint Petersburg paradox and doubling strategy)**.** *In the setting of Example 2.2.3, we consider the following strategy: $\Delta_0 = 1$ and $\Delta_i = 2^i$, for $i > 0$, if the player has lost all the past rounds from 1 to $i-1$. Otherwise, if the first winning occurs at round i, we set $\Delta_j = 0$ for $j \geqslant i$. For example, if the player's outcome in the first five rounds are given by "loss, loss, loss, loss, win", then his bets are given by "1, 2, 4, 8, 16, 0 , ...", respectively. Then, the wealth of the player after five rounds is given by*

$$\mathcal{W}_5 = 1(-1) + 2(-1) + 4(-1) + 8(-1) + 16(1) = 1.$$

However, his wealth before the fifth round is always negative.

$$\mathcal{W}_1 = -1, \quad \mathcal{W}_2 = -3, \quad \mathcal{W}_3 = -7, \quad \mathcal{W}_4 = -15.$$

More generally, if the player loses first $i-1$st rounds and win the ith round, the wealth of the player satisfies

$$\mathcal{W}_1 = -1, ..., \mathcal{W}_{i-1} = \sum_{j=0}^{i-1} 2^j = 2^i - 1 \text{ and } \mathcal{W}_{i-1} = 1.$$

Exercise 2.2.2. *In Example 2.2.4, we showed that if the player has the opportunity to borrow with no limitation and continue the game until the first win, he will always end up with terminal wealth equal to \$1.*

a) *Assume that the player has a credit line, denoted by C. He stops playing if either he reaches his credit limit or he wins for the first time. Find the possible values for the terminal wealth of the player.*

b) *Find the expected value of the terminal wealth of the player, given that the probability of winning is $p \in (0,1)$.*

Arbitrage strategy

In this section, we present the definition of an arbitrage opportunity and a version of the FTAP[7] for multistep discrete-time markets. In order to define arbitrage, we first fix the sample space Ω of all samples paths of the price process and we define a σ-field $\mathcal{F}$ of all events on the sample paths space. When Ω is a finite or countably infinite set, we can choose $\mathcal{F}$ to be the σ-field of all subsets of Ω. When $\Omega = (\mathbb{R}_+^{d+1})^T$, we set $\mathcal{F} = \mathcal{B}((\mathbb{R}_+^{d+1})^T)$. To define arbitrage, it is crucial to to determine the set of all events on the sample paths space that are believed to have a chance to occur. This is a part of modeling a financial market. Relevant events are those that basically represent our beliefs about the market behavior. Equivalently, one can determine the set of all events on the sample paths space that are deemed impossible to occur. We define such events below.

Definition 2.2.2. *We call a collection of events $\mathcal{N} \subseteq \mathcal{F}$ a polar collection if it satisfies*

a) $\varnothing \in \mathcal{N}$.

b) *If $B \in \mathcal{N}$ and $A \subseteq B$, then $A \in \mathcal{N}$.*

c) *If $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{N}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{N}$.*

The members of a polar collection are called polar sets .

As discussed in examples below, the polar collection depends on the choice of the space sample paths.

Example 2.2.5 (polar set)**.** *In the binomial model in Section 2.1.3, if we set the sample space to be $\Omega = \mathbb{R}_+$, the polar collection $\mathcal{N}$ is given by all subsets of Ω that do not contain any of the points S_0u or $S_0\ell$. However, if we set the sample space to $\Omega = \{u, \ell\}$, then $\mathcal{N} = \{\varnothing\}$. Similarly, in the trinomial model in Section 2.1.4, $\Omega = \mathbb{R}_+$ is the canonical space and the polar collection $\mathcal{N}$ is the collection of all events A that do not contain any of the points S_0u, S_0m, or $S_0\ell$. If we set the sample space to $\Omega = \{u, m, \ell\}$, then $\mathcal{N} = \{\varnothing\}$.*

Example 2.2.6. *In Example 2.2.1, consider the sample paths space given by $\Omega = \{(x, y) : x = -1 \text{ or } +1 \text{ and } y = -2,\ 0,\ \text{or } 2\}$. Then, the set of polar collection is*

$$\Big\{\varnothing, \{(1,-2)\}, \{(-1,2)\}, \{(1,-2),(-1,2)\}\Big\}.$$

The arbitrage relative to the polar collection $\mathcal{N}$ is defined below.

[7] A fundamental theorem of asset pricing

Definition 2.2.3. *A (weak) arbitrage opportunity is a portfolio* Δ *such that*

a) $W_0 = 0$,

b) $\{W_T < 0\} \in \mathcal{N}$, *and*

c) $\{W_T > 0\} \notin \mathcal{N}$.

A strong arbitrage opportunity is a portfolio Δ *such that*

a) $W_0 < 0$, *and*

b) $\{W_T < 0\} \in \mathcal{N}$.

By $\{W_T < 0\} \in \mathcal{N}$, we mean that the event $\{W_T \geqslant 0\}$ will surely happen. $\{W_T > 0\} \notin \mathcal{N}$ means that $\{W_T > 0\}$ is likely to happen with a possibly small chance.

In the above, we did not assign any probability to the events, except the polar collection; the polar collection is the collection of all events that are believe to have probability zero. Outside polar collection, all the events have nonzero probability, which may not be known.

To extend FTAP to discrete-time multiperiod market models, we need the probabilistic notion of *martingale*, which is introduced in Section B.3. Risk-neutral probability in a multiperiod market can be defined in terms of the martingale property for *the discounted asset price*: $\hat{S}_t := (1+R)^{-t} S_t$; we assume implicitly that there is a zero bond with yield R in the market.

Definition 2.2.4. *We call a probability* $\hat{\mathbb{P}}$ *a risk-neutral probability if the discounted asset price is a martingale with respect to the* σ*-field generated by the price process* $\{S_t : t \geqslant 0\}$ *under the probability measure* $\hat{\mathbb{P}}$*; i.e.,*

$$\hat{\mathbb{E}}[\hat{S}_{t+1} \mid S_t, S_{t-1}, ..., S_1] = \hat{S}_t. \tag{2.2.4}$$

Here, $\hat{\mathbb{E}}$ *is the expectation with respect to* $\hat{\mathbb{P}}$.

We can also use the notion of σ-field $\mathcal{F}_t^S := \sigma(S_t, ..., S_0)$ generated by $S_t, ..., S_0$ to write

$$\hat{\mathbb{E}}[\hat{S}_{t+1} \mid \mathcal{F}_t^S] = \hat{S}_t.$$

For a single-period binomial model, the martingale property with respect to risk-neutral probability is expressed and verified in Section 2.1.3; see (2.1.5).

Exercise 2.2.3. *Let* $\{\Delta_t\}_{t=0}^{\infty}$ *be a bounded portfolio strategy; i.e., there exists a number* C *such that* $|\Delta_t| < C$ *for all* $t = 0, 1, ...$ *Show that if the discounted price* $\hat{S}_t = \frac{S_t}{(1+R)^t}$ *is a martingale with respect to probability* $\hat{\mathbb{P}}$*, then the stochastic integral* $(\Delta \cdot \hat{S})_t$ *and the discounted wealth process* $\tilde{W}_t$ *are martingales with respect to* $\hat{\mathbb{P}}$.

The following result extends Theorem 2.1.1 into multiple periods. In order to have a fundamental theorem of asset pricing in a general form, we need to impose the following assumption.

Assumption 2.2.1 (Dominating probability)**.** *There exists a probability* $\mathbb{P}$ *such that* $A \in \mathcal{N}$ *if and only if* $\mathbb{P}(A) = 0$*.*

Assumption 2.2.1 holds trivially if the sample space of all paths are finite or countably infinite. It is a nontrivial assumption when the sample path space is uncountable.

The probability $\mathbb{P}$ can be regarded as the physical probability in the market. Therefore, one can equivalently define the collection of all polar sets as the set of all events that have zero probability under $\mathbb{P}$. However, as seen in Theorem 2.2.1, the only relevant information about probability $\mathbb{P}$ is the no-arbitrage condition is the polar collection; the actual value of the probability of an event does not matter as long as it has a nonzero probability. In this case, we say that the polar sets are *generated* by $\mathbb{P}$.

Definition 2.2.5. *Two probabilities* $\mathbb{P}$ *and* $\hat{\mathbb{P}}$ *are called equivalent if they generate the same polar sets. We denote the equivalency by* $\mathbb{P} \equiv \hat{\mathbb{P}}$*.*

Theorem 2.2.1 (Fundamental theorem of asset pricing (FTAP))**.** *Let Assumption 2.2.1 hold. Then, there is no weak arbitrage opportunity in the discrete-time market model if and only if there exists a probability measure* $\hat{\mathbb{P}}$ *such that*

a) $\hat{\mathbb{P}} \equiv \mathbb{P}$*, and*

b) the discounted asset price $\hat{S}_t$ *is a (local) martingale(local) martingale*[8] *under* $\hat{\mathbb{P}}$*.*

The probability measure $\hat{\mathbb{P}}$ is called a risk-neutral probability.

A market model with no arbitrage is called a complete market model if any contingent claim is replicable. In other words, for a contingent claim with payoff $g : (\mathbb{R}_+^{d+1})^T \to \mathbb{R}$ which maps the history of an underlying asset price, $S_0, S_1, ..., S_T$ into $g(S_0, S_1, ..., S_T)$, there exists a portfolio $\Delta_0, ..., \Delta_{T-1}$ such that

$$(\Delta \cdot S)_T = g(S_0, S_1, ..., S_T).$$

As an extension to Theorem 2.2.1, we have the following condition for the completeness of a market.

Corollary 2.2.1. *Let Assumption 2.2.1 and no-weak-arbitrage condition hold. Then, The market is complete if and only if there is a unique risk-neutral probability measure.*

[8] local martingale is roughly a martingale without condition (a) in Definition B.15.

One direction of Theorem 2.2.1 is easy to prove. Here is a glimpse of the proof. Assume that a risk-neutral probability $\hat{\mathbb{P}}$ exists, and consider an arbitrage strategy Δ with the corresponding discounted wealth process W_t integrable. Then,

$$\hat{W}_t = (\Delta \cdot \hat{S})_t. \quad \text{(Recall that} \quad W_0 = 0.)$$

Since $\hat{S}_t$ is a $\hat{\mathbb{P}}$-martingale, then by Exercise 2.2.1, $\hat{W}_t$ is a martingale and we have

$$\hat{\mathbb{E}}[\hat{W}_t] = W_0 = 0.$$

This is in contradiction to condition (c) in the definition of arbitrage 2.2.3. For a complete proof of this result, see [13, Pg. 7, Theorem 1.7]. A very general form of this theorem can be found in a seminal paper by Delbaen and Schachermayer [9].

Remark 2.2.1. *Assumption 2.2.1 can be relaxed by assuming that the polar sets are generate by a convex collection of probabilities $\mathcal{P}$. More precisely $A \in \mathcal{N}$ if and only if $\mathbb{P}(A) = 0$ for all $\mathbb{P} \in \mathcal{P}$. Then, the fundamental theorem of asset pricing should be modified: there is no weak arbitrage opportunity in the discrete-time model if and only if*

a) $\mathcal{Q} := \{\hat{\mathbb{P}} : \hat{S}_t$ *is a* $\hat{\mathbb{P}}$*-martingale*$\}$ *is nonempty, and*

b) $\mathcal{P}$ *and* $\mathcal{Q}$ *generate the same polar sets.*

For more on the relaxation of Assumption 2.2.1, see [6].

By Theorem 2.2.1, the existence of risk-neutral probability eliminates the possibility of arbitrage. However when $T = \infty$, the Saint Petersburg Paradox still holds even though a risk-neutral probability exists. This is a different issue and is related to the notion of an *admissible portfolio strategy*. The following corollary suggest a practical way around this paradox.

Corollary 2.2.2. *Let $\{M_t\}_{t=0}^{\infty}$ be a martingale, $\{\Delta_t\}_{t=0}^{\infty}$ a portfolio strategy, and C be a constant such that $(\Delta \cdot M)_t \geqslant C$ for all $t \geqslant 0$. Then, for any stopping time τ such that $\tau < \infty$ a.s., we have*

$$\mathbb{E}[(\Delta \cdot M)_\tau \mid \mathcal{F}_t] = (\Delta \cdot M)_{\tau \wedge t}.$$

In particular, $\mathbb{E}[(\Delta \cdot M)_\tau] \leqslant 0$.

Definition 2.2.6 (Admissible strategy)**.** *A portfolio strategy is called admissible if there exists a C be a constant such that $(\Delta \cdot S)_t \geqslant C$ for all $t \geqslant 0$.*

restriction to admissible strategies is not necessary when there are only finite number of periods, $T < \infty$. For $T = \infty$, we need to restrict the portfolio strategy choices to admissible strategies. In addition, this happens to be necessary when we pass to a limit from a discrete-time market to a continuous-time market by sending the number of periods to infinity.

2.3 Binomial model

Let $H_1, H_2, ...$ be an i.i.d.[9] sequence of random variables with values u and ℓ[10]. Let T be the maturity, and let the time variable t take values $0, 1, ..., T$. At time 0, the price of the asset is S_0. At time $t = 1, ..., T$, the price of the asset satisfies $S_t = S_{t-1}H_t$. The binomial model is shown in Figure 2.3.1.

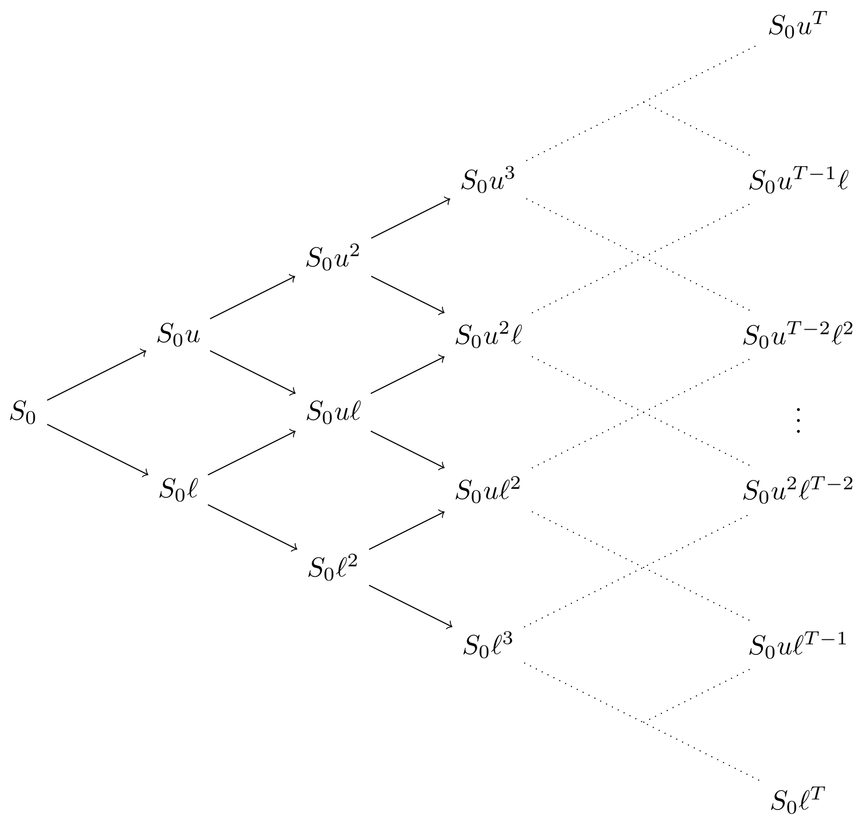


Figure 2.3.1: Asset price in the binomial model

We label the nodes of the binomial model by the time and the state of the asset price. For example, at time t when the asset price is equal to $S_{t,j+1} := S_0u^j\ell^{t-j}$, the node is labeled $(t, j+1)$. The only node at time 0 is labeled 0 for simplicity. See Figure 2.3.2.

[9]Independent identically distributed

[10]The probabilities of these values are irrelevant at this moment.

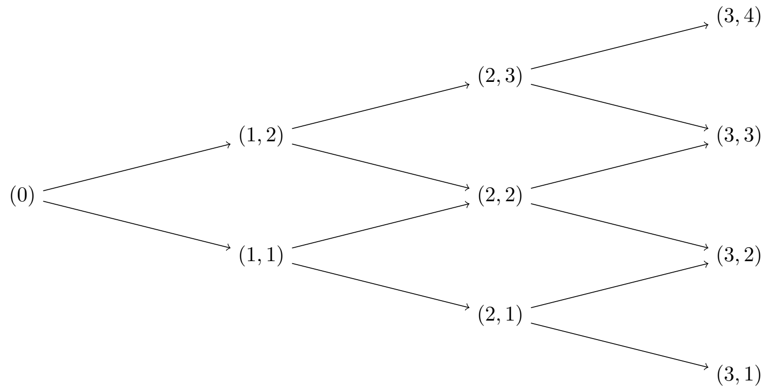


Figure 2.3.2: Label of the nodes in a three-period binomial model

2.3.1 No-arbitrage condition

The no-arbitrage condition for the multiperiod binomial model is the same as for the single-period. There is no arbitrage in the multiperiod binomial model if and only if there is no arbitrage for the single-period model with the same parameters.

Proposition 2.3.1. *There is no arbitrage in the multiperiod binomial model if and only if $\ell < 1 + R < u$. In this case, the multiperiod binomial market is complete, and the risk-neutral probability is given by assigning the following distribution to each H_i.*

$$\hat{\mathbb{P}}(H_i = u) = \frac{1 + R - \ell}{u - \ell} \quad \text{and} \quad \hat{\mathbb{P}}(H_i = \ell) = \frac{u - 1 - R}{u - \ell}.$$

Proof. By Theorem 2.2.1 (FTAP), no-arbitrage condition is equivalent to the existence of a risk-neutral probability. We first show that given $\ell < 1 + R < u$, the probability $\hat{\mathbb{P}}$ defined in the theorem is a risk-neutral probability. In other words, we shall show that the discounted asset price is a martingale. Notice that since $S_{t+1} = H_{t+1}S_t$, we have $\hat{S}_{t+1} = \frac{H_{t+1}}{1+R}\hat{S}_t$. Therefore,

$$\hat{\mathbb{E}}[\hat{S}_{t+1} \mid \mathcal{F}_t^S] = \frac{1}{1+R}\hat{\mathbb{E}}[\hat{S}_t H_{t+1} \mid \mathcal{F}_t^S],$$

where $\mathcal{F}_t^S = \sigma(\hat{S}_t, ..., \hat{S}_0)$. Since $\hat{S}_t$ is known given $\mathcal{F}_t^S$, if follows from Corollary B.6 that

$$\hat{\mathbb{E}}[\hat{S}_t H_{t+1} \mid \mathcal{F}_t^S] = \frac{\hat{S}_t}{1+R}\hat{\mathbb{E}}[\hat{H}_{t+1} \mid \mathcal{F}_t^S].$$

On the other hand, since $H_1, H_2, ...$ is a sequence of independent random variables, H_{t+1}

is independent of $\mathcal{F}_t^S$, and we have

$$\hat{\mathbb{E}}[\hat{H}_{t+1} \mid \mathcal{F}_t^S] = \hat{\mathbb{E}}[\hat{H}_{t+1}] = \hat{\pi}_u u + \hat{\pi}_\ell \ell = 1 + R.$$

Thus,

$$\hat{\mathbb{E}}[\hat{S}_{t+1} \mid \mathcal{F}_t^S] = \hat{S}_t.$$

For the other direction, assume by contraposition that either $u \leqslant 1 + R$ or $\ell \geqslant 1 + R$. Then, Section 2.1.3 shows that there is an arbitrage in the first period. Then, one can liquidate the position to cash after the first period to carry over the arbitrage until time T. □

The above proof is also presented in Example B.41 in a different way.

As shown in Figure 2.3.1, the random variable S_n only takes values $S_0 u^{n-k} \ell^k$ for $k = 0, ..., n$. Under the risk-neutral probability,

$$\hat{\mathbb{P}}(S_n = S_0 u^{n-k} \ell^k) = \binom{n}{k} (\hat{\pi}_u)^{n-k} (\hat{\pi}_\ell)^k.$$

To see this, notice that in the binomial model in Figure 2.3.1, there are $\binom{n}{k}$ paths from the node S_0 to node $S_0 u^{n-k} \ell^k$, and the probability of each path is $(\hat{\pi}_u)^{n-k} (\hat{\pi}_\ell)^k$[11]. For simplicity, we denote $S_0 u^{t-i} \ell^i$ by $S_t(i)$.

Remark 2.3.1 (Recombination in the binomialmodel)**.** *The binomial model has a feature that allows reduce the number of values that S_n can take. For example, consider a two period market with one asset whose price process is given below:*

$$S_0 = 1, S_1 = H_1, \text{ and } S_2 = H_1 H_2,$$

where H_i takes values u_i and ℓ_i for $i = 1, 2$. At time $t = 2$, the S_2 takes values $u_1 u_2$, $u_1 \ell_2$, $u_2 \ell_1$, and $\ell_1 \ell_2$. All these values are distinct unless $u_1 \ell_2 = u_2 \ell_1$. In the binomial model $u_1 = u_2 = u$ and $\ell_1 = \ell_2 = \ell$. Therefore, $u_1 \ell_2 = u_2 \ell_1 = u\ell$. Therefore, the values for S_2 binomial model reduces to three, because the values $u_1 \ell_2$ and $u_2 \ell_1$ recombine. See Figure 2.3.3. In general for a T period binomial market model, the recombination allows that the price S_t takes only $t + 1$ values, whereas in a nonrecombining market model, there are potentially 2^t distinct values for S_t.

2.3.2 Basic properties of the binomial model

The binomial model described above has some properties that are the common features in many models in finance. These features allows to perform risk management evaluations in a reasonable time. In this section, we discuss these properties.

[11] This is an elementary combinatorics problem.

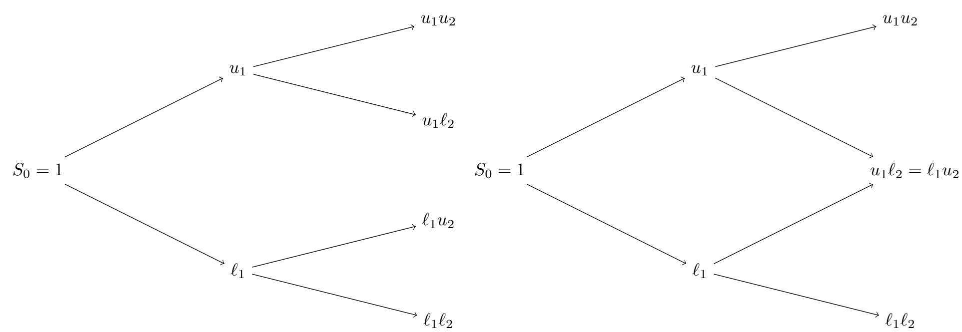


Figure 2.3.3: Recombining binomial market model (right) versus a nonrecombining one (left).

Time homogeneity

Since $\{H_i\}_{t=1}^{\infty}$ is an **i.i.d.** sequence of random variables, then for $t > s$, $\prod_{i=s+1}^{t} H_i$ has the same distribution as $\prod_{i=1}^{t-s} H_i$. Therefore, given $S_s = S$, $S_t = S_s \prod_{i=s+1}^{t} H_i$ has the same distribution as $S_{t-s} = S \prod_{i=1}^{t-s} H_i$. In other words, the conditional distribution of S_t given $S_s = S$ is the same as conditional distribution of $S_{t-s} = S_0 \prod_{i=1}^{t-s} H_i$ given $S_0 = S$.

Markovian property

The *Markovian* property for a stochastic process asserts that in order to determine the probability of future scenarios of the value of the process, for example the value at a time in the future, *the only relevant information from the past history of the price process is the most recent one.* In other words,

Definition 2.3.1. *A stochastic process $\{X_t : t \geqslant 0\}$ with values in $\mathbb{R}^d$ is called Markovian if for any $A \subset \mathbb{R}$ and $s > t$, we have*

$$\mathbb{P}(X_s \in A \mid X_t, ..., X_0) = \mathbb{P}(X_s \in A \mid X_t].$$

Equivalently, one can write the Markovian property of in terms of conditional expectation:

$$\mathbb{E}[g(X_s) \mid X_t, ..., X_0] = \mathbb{E}[g(X_s) \mid X_t]. \tag{2.3.1}$$

The binomial model is *Markovian* under risk-neutral probability; given $S_t, ..., S_1, S_0$

$$\hat{\mathbb{E}}[g(S_s) \mid S_t, ..., S_0] = \hat{\mathbb{E}}\Big[g\Big(S_t \prod_{i=t+1}^{s} H_i\Big) \mid S_t, ..., S_0\Big].$$

Since $\prod_{i=t+1}^{s} H_i$ is independent of $S_t, ..., S_0$, by Proposition B.6, we have

$$\hat{\mathbb{E}}\Big[g\Big(S_t \prod_{i=t+1}^{s} H_i\Big) \mid S_t = s_t, ..., S_0 = s_0\Big] = \hat{\mathbb{E}}\Big[g\Big(s_t \prod_{i=t+1}^{s} H_i\Big)\Big] = \hat{\mathbb{E}}\Big[g\Big(S_t \prod_{i=t+1}^{s} H_i\Big) \mid S_t = s_t\Big].$$

Therefore,

$$\hat{\mathbb{E}}[g(S_s) \mid S_t, ..., S_0] = \hat{\mathbb{E}}[g(S_s) \mid S_t].$$

See Figure2.3.4 for the illustration of the Markovian property in the binomial model in one period. Given $S_t = S_t(i)$, the probability that $S_{t+1} = S_{t+1}(i+1)$ is $\hat{\pi}_u$, and the probability that $S_{t+1} = S_{t+1}(i)$ is $\hat{\pi}_\ell$. Given $S_t \neq S_t(i)$, both of the probabilities are 0.

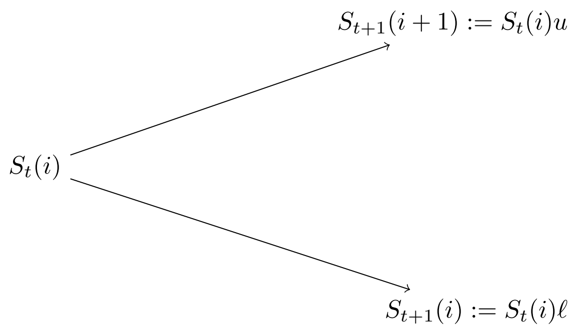


Figure 2.3.4: Conditioning of the binomial model

Why is Markovian property important? The Markovian property is often useful in reducing the computational effort, and models with the Markovian property are computationally efficient. A reason for this reduction lies in the solution to the exercise below.

Exercise 2.3.1. *How many paths are there in a binomial model from time $t = 0$ to time $t = n$? How many nodes (values of asset price process at all points in time) are there?*

If we don't have Markovian property, we need to evaluate the conditional expectation $\hat{\mathbb{E}}[g(S_s) \mid S_t, ..., S_0]$ once for each sample path; $\hat{\mathbb{E}}[g(S_s) \mid S_t, ..., S_0]$ is a random variable that has as many values as the process $(S_0, ..., S_t)$ does. However, Markov property allows us to reduce the conditional expectation to $\hat{\mathbb{E}}[g(S_s) \mid S_t]$, and therefore, the number of values that $\hat{\mathbb{E}}[g(S_s) \mid S_t]$ can take is as many as the number of values of the ranvom variable S_t.

2.3.3 Arbitrage pricing and replicating European contingent claims in the binomial model

As in Section 2.1, the introduction of a new asset can create arbitrage if and only if the discounted price of the new asset does not satisfy

$$\frac{1}{1+R}\hat{\mathbb{E}}[\mathbf{P}'_{\cdot j}] = p',$$

for at least one risk-neutral probability. In the one-period binomial model, since there is only one risk-neutral probability provided that there is no arbitrage, one can readily find the no-arbitrage price of any new asset including all derivatives.

The same methodology applies to the multiperiod binomial model, with a slight difference: the discounted price of the newly introduced asset must be a martingale with respect to the risk-neutral probability.

We start by illustrating the idea in a two-period binomial model in the following example.

Example 2.3.1. *Consider a two-period binomial model with* $S_0 = 1$, $u = 2$, $\ell = \frac{1}{2}$, *and* $R = .5$ *(for simplicity). We consider a European call option with strike* $K = .8$*; the payoff is* $g(S_2) = (S_2 - .8)_+$*. Therefore,* $\hat{\pi}_u = \frac{2}{3}$ *and* $\hat{\pi}_\ell = \frac{1}{3}$.

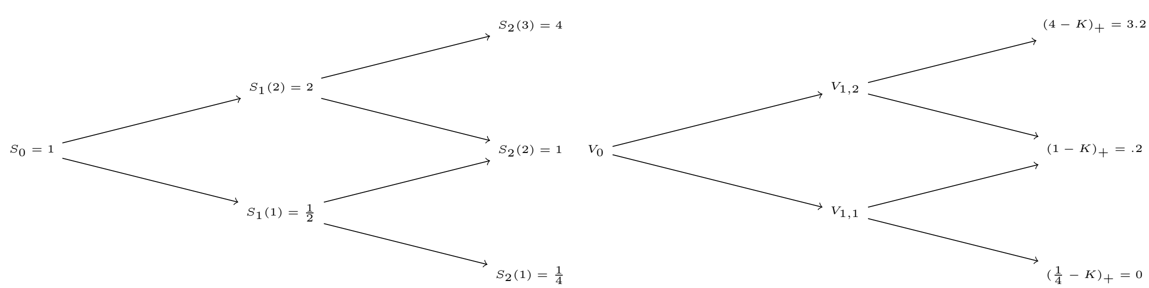


Figure 2.3.5: European call option in a two-period binomial model. Left: asset price. Right: option price

We first argue that the price of the European call option mimics the binomial model for the asset and takes a similar form shown on the right-hand-side in Figure 2.3.5. At the maturity, the value of the option is given by $V_2 = (S_2 - K)_+$*. Since* S_2 *takes three values, so does* V_2*. At time* $t = 1$,

We assume that there is no arbitrage. Therefore, the discounted price of the option must be a martingale with respect to the asset price under the risk-neutral probability:

$$V_1 = \frac{1}{1+R}\hat{\mathbb{E}}[V_2 \mid S_1, S_0] = \frac{1}{1+R}\hat{\mathbb{E}}[(S_2 - K)_+ \mid S_1, S_0].$$

By the Markovian property of the asset price, we have,

$$V_1 = \frac{1}{1+R}\hat{\mathbb{E}}[(S_2 - K)_+ \mid S_1].$$

Therefore, V_1 is a function of S_1, and since S_1 takes two values, the value V_1 of the option takes two values $V_{1,2}$ and $V_{1,1}$ when S_1 takes values $S_1(2) = S_0 u$ and $S_1(1) = S_0\ell$, respectively. More precisely,

$$\begin{aligned} V_{1,2} &= \frac{1}{1+R}\hat{\mathbb{E}}[(S_2 - K)_+ \mid S_1 = S_0 u] = \frac{1}{1+R}\left(\hat{\pi}_u (S_1 u - K)_+ + \hat{\pi}_\ell (S_1 \ell - K)_+\right)\Big|_{S_1 = S_0 u} \\ &= \frac{1}{1.5}\left(\frac{2}{3}(3.2) + \frac{1}{3}(.2)\right) = \frac{4.4}{3} \approx 1.4666. \end{aligned}$$

Similarly at node $(1,1)$, where $t = 1$ and state is 1, we have

$$\begin{aligned} V_{1,1} &= \frac{1}{1+R}\hat{\mathbb{E}}[(S_2 - K)_+ \mid S_1 = S_0 \ell] = \frac{1}{1+R}\left(\hat{\pi}_u (S_1 u - K)_+ + \hat{\pi}_\ell (S_1 l - K)_+\right)\Big|_{S_1 = S_0 \ell} \\ &= \frac{1}{1.5}\left(\frac{2}{3}(.2) + \frac{1}{3}(0)\right) = \frac{.8}{9} \approx 0.8888. \end{aligned}$$

To evaluate the option price V_0 at time $t = 0$, we use that the no-arbitrage implies the martingale property for the option:

$$V_0 = \frac{1}{1+R}\hat{\mathbb{E}}[V_1] = \frac{1}{1+R}\left(\hat{\pi}_u V_{1,2} + \hat{\pi}_\ell V_{1,1}\right) = \frac{2}{3}\left(\frac{2}{3}(\frac{4.4}{3}) + \frac{1}{3}(\frac{.8}{9})\right) \approx .6716.$$

To replicate the option, we need to solve the same system of equations as in 2.1.7 at each node of the binomial model in a backward manner. At node $(1,2)$,

$$\begin{cases} \theta_0(1+R) + \theta_1 S_1(2) u &= (4-K)_+ \\ \theta_0(1+R) + \theta_1 S_1(2)\ell &= (1-K)_+ \end{cases} \quad or \quad \begin{cases} 1.5\theta_0 + 4\theta_1 &= 3.2 \\ 1.5\theta_0 + \theta_1 &= .2 \end{cases}.$$

*Thus, $\theta_1 = 1$ and $\theta_0 = -\frac{1.6}{3}$. In other words, to replicate the claim at node $(1,2)$ we need to keep one unit of the risky asset and **borrow** $\frac{1.6}{3}$ units of the risk-free zero bond. This leads to the price $\theta_0 + \theta_1 S_1(2) = 2 - \frac{1.6}{3} = \frac{4.4}{3}$, the same price we found with risk-neutral probability.*

The same method should be used in the other node, $(1,1)$, to obtain the system of equation

$$\begin{cases} \theta_0(1+R) + \theta_1 &= .2 \\ \theta_0(1+R) + \frac{1}{4}\theta_1 &= 0 \end{cases}.$$

Thus, $\theta_1 = \frac{.8}{3}$ and $\theta_0 = -\frac{.4}{9}$. The price $\theta_0\theta_1 S_1(1) = \frac{1}{2}\left(\frac{.8}{3}\right) - \frac{.4}{9} = \frac{.8}{9}$ is again the same as *the risk-neutral price.*

At node 0*, the replicating portfolio needs to reach the target prices of the claim at time* $t = 1$*, i.e.*

$$\begin{cases} \theta_0 + \theta_1 S_0 u & = V_{1,2} = \frac{4.4}{3} \\ \theta_0 + \theta_1 S_0 \ell & = V_{1,1} = \frac{.8}{9} \end{cases} \quad \text{or} \quad \begin{cases} \theta_0(1+R) + 2\theta_1 & = \frac{4.4}{3} \\ \theta_0(1+R) + \frac{1}{2}\theta_1 & = \frac{.8}{9} \end{cases}.$$

By solving the above system, we obtain $\theta_1 = \frac{24.8}{27}$ *units of the risky asset and* $\theta_0 = -\frac{20}{81}$ *units of the risk-free bond in the replicating portfolio. This is the structure of the replicating portfolio at the beginning of the replication. If the price moves up, we have to restructure the portfolio to keep one unit of the risky asset and* $-\frac{1.6}{3}$ *units of the risk-free bond. If it moves down, we need to readjust the position to* $\frac{.8}{3}$ *units of the risky asset and* $-\frac{.4}{9}$ *units of the risk-free bond.*

Next, consider a general European claim with payoff $g(S_T)$, where $g : \mathbb{R}_+ \to \mathbb{R}$ is a function that assigns a value to the payoff based on the price S_T of the asset at terminal time T. Such European claims are also called *Markovian claims*. Let $V_0, ..., V_T$ be random variables representing the price of this European claim at time $t = 0, ..., T$, respectively. Then, in order to avoid arbitrage, the derivative price must remain martingale, i.e.

$$V_t = \frac{1}{1+R}\hat{\mathbb{E}}[V_{t+1} \mid S_t, ..., S_0].$$

Let's assume that at time $t+1$, V_{t+1} is a function $V(t+1, \cdot)$ and S_{t+1}. This assumption is true for T, where $V_T = g(S_T)$. We use induction to show that V_t is a function of S_t. It follows from the Markovian property of the binomial model that $\hat{\mathbb{E}}[V_{t+1} \mid S_t, ..., S_0] = \hat{\mathbb{E}}[V(t+1, S_{t+1}) \mid S_t, ..., S_0] = \hat{\mathbb{E}}[V(t+1, S_{t+1}) \mid S_t]$, and therefore

$$V(t, S_t) := \frac{1}{1+R}\hat{\mathbb{E}}[V(t+1, S_{t+1}) \mid S_t]. \tag{2.3.2}$$

Therefore, one needs to evaluate function $V(t, S)$ over the binomial model, as shown in Figure 2.3.6.

We can also show bu induction that

$$V(t, S_t) = \frac{1}{(1+R)^{T-t}}\hat{\mathbb{E}}[\hat{\mathbb{E}}[g(S_T) \mid S_t].$$

This holds for $t = T - 1$:

$$V(T-1, S_{T-1}) = \frac{1}{1+R}\hat{\mathbb{E}}[g(S_T) \mid S_{T-1}].$$

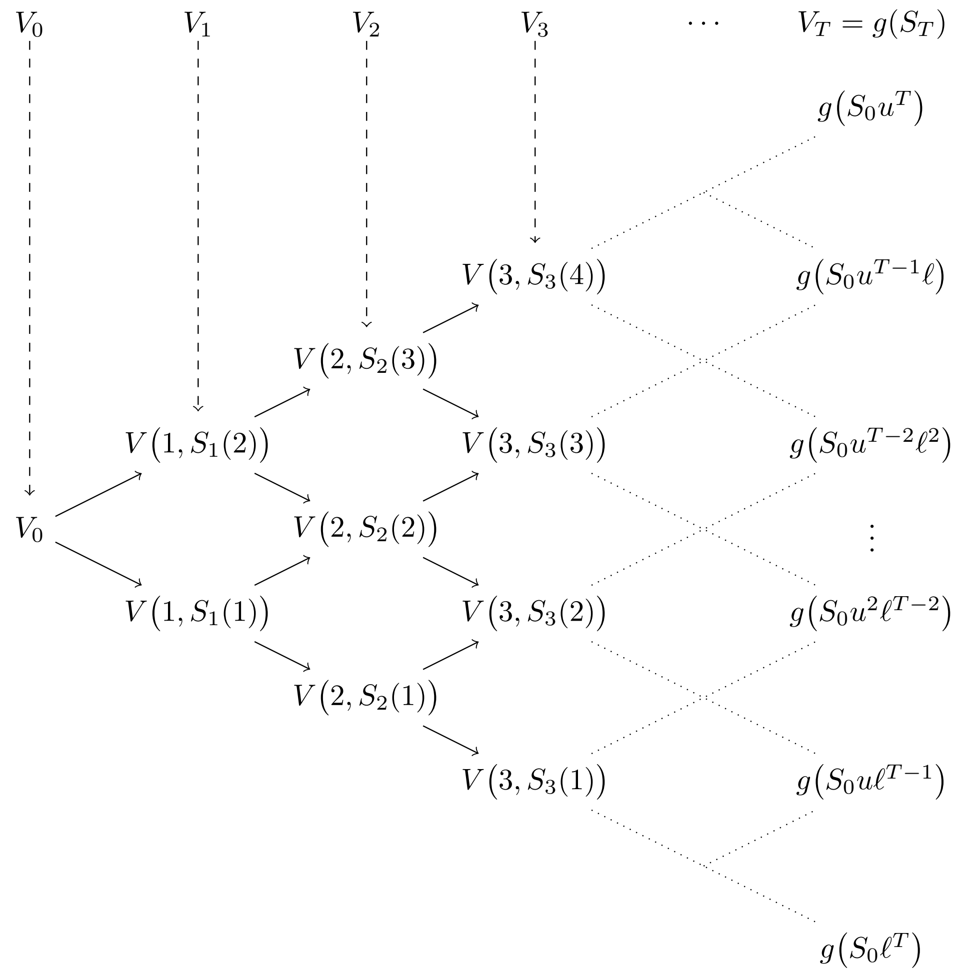


Figure 2.3.6: The price of a Markovian European contingent claim in the binomialmodel

Now assume that

$$V(t+1, S_{t+1}) = \frac{1}{(1+R)^{T-(t+1)}}\hat{\mathbb{E}}[\hat{\mathbb{E}}[g(S_T) \mid S_{t+1}].$$

Then, by the tower property of the conditional expectation, we can write

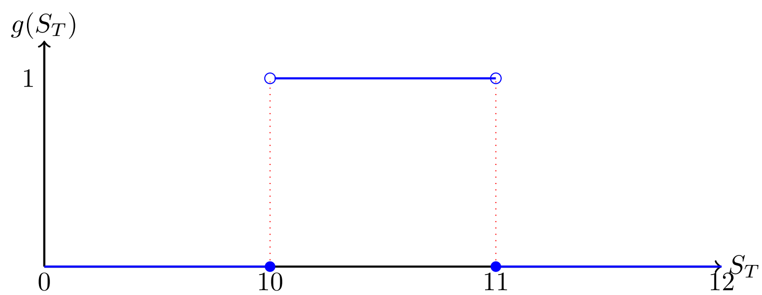
$$\begin{aligned} V(t, S_t) &:= \frac{1}{1+R}\hat{\mathbb{E}}[V(t+1, S_{t+1}) \mid S_t] \\ &= \frac{1}{1+R}\hat{\mathbb{E}}\Big[\frac{1}{(1+R)^{T-(t+1)}}\hat{\mathbb{E}}[g(S_T) \mid S_{t+1}, ..., S_0] \mid S_t\Big] = \frac{1}{(1+R)^{T-t}}\hat{\mathbb{E}}[g(S_T) \mid S_t]. \end{aligned}$$

Therefore, for any $t = 0, ..., T-1$, we have

$$V(t, S_t) = \frac{1}{(1+R)^{T-t}} \hat{\mathbb{E}}[g(S_T) \mid S_t]. \tag{2.3.3}$$

Remark 2.3.2. *Since the binomial model is time homogeneous, $\hat{\mathbb{E}}[g(S_T) \mid S_t = S]$ in (2.3.3) is equal to $\hat{\mathbb{E}}[g(S_{T-t}) \mid S_0 = S]$. This suggests that the price function $V(t, S)$ is a function of S and* ***time-to-maturity*** *$\tau := T - t$, $V(\tau, S)$. Time-to-maturity is often used instead of time in financial literature regarding the evaluation of contingent claims.*

Example 2.3.2. *Consider a four-period binomial model for a risky asset with each period equal to a year, and take $S_0 = \$10$, $u = 1.06$, $\ell = 0.98$, and $R = .02$. We shall find the price V_0 of the option with the payoff shown in figure below. By 2.3.3, the value $V(0, S_0)$*



of the option is the expected value of the discounted payoff under risk-neutral probability;

$$V(0, S_0) = \frac{1}{(1+R)^4} \hat{\mathbb{E}}[g(S_4)].$$

Since the random variable S_4 takes values 9.2236816, 9.9766352, 10.7910544, 11.6719568, and 12.6247696. Therefore, the only nonzero value of the payoff is obtained when $S_4 = 10.7910544$ and is $g(10.7910544) = 1$. The risk-neutral probability of $S_T = 10.7910544$ is simply $\binom{2}{4}(\hat{\pi}_u)^2(\hat{\pi}_\ell)^2 = \frac{3}{8}$. Thus,

$$V(0, S_0) = \frac{3}{8(1.02)^4} \approx 0.34644203476.$$

Replication of European option in the binomial model

We show that any European contingent claim (even non-Markovian ones) are perfectly replicable in the binomial model. The argument follows inductively: let the replicating portfolio is built at each node of the binomial model at all points $t+1$ or later in time. As a result of this assumption, at each node the value of the replicating portfolio is the same as the value of the option. We continue by replicating the price of the option at time $t+1$;

i.e., we need to solve the following system of equations for each $i = 1, ..., t+1$:

$$\begin{cases} \theta_0(1+R) + \theta_1 S_{t+1}(i+1) & = V(t+1, S_{t+1}(i+1)) \\ \theta_0(1+R) + \theta_1 S_{t+1}(i) & = V(t+1, S_{t+1}(i+1)) \end{cases}.$$

The solution is given by

$$\theta_0 = \frac{uV(t+1, S_t(i)l) - lV(t+1, S_t(i)u)}{(u-\ell)(1+R)} \text{ and } \theta_1 = \frac{V(t+1, S_t(i)u) - V(t+1, S_t(i)l)}{S_t(i)(u-\ell)}.$$

θ_1 is the number of units of the risky asset in the replicating portfolio, and θ_0 is the number of units of the risk-free bond in the replicating portfolio. In other words, the replicating portfolio is a self-financing portfolio given by (2.2.3) with initial wealth V_0 and portfolio strategy given by $\{\Delta(t, S_t)\}_{t=0}^{T-1}$.

$$\Delta(t, S) := \frac{V(t+1, Su) - V(t+1, S\ell)}{S(u-\ell)}. \tag{2.3.4}$$

The number of units of the risky asset in the replicating portfolio, given by (2.3.4), is called the *Delta* of the contingent claim at time t. Basically, (2.3.4) suggests that the Delta of a European Markovian contingent claim is a function of time t and the price of the underlying asset at time t.

Remark 2.3.3. *As you can see from* (2.3.4)*, the Delta of the contingent claim at time t measures the sensitivity of the value of the contingent claim with respect to changes in the price of the underlying asset, i.e., changes in the price of the option due to changes in the price of the underlying asset.*

By (2.2.3), the replicating portfolio for the binomial model takes the form

$$V(t, S_t) = V(0, S_0) + R\sum_{i=0}^{t-1}(V(i, S_i) - \Delta(i, S_i)S_i) + \sum_{i=0}^{t-1}\Delta(i, S_i)(S_{i+1} - S_i),$$

where $V(i, S_i)$ is the price of the contingent claim at time i when the underlying price is S_i. The term $R\sum_{i=0}^{t-1}(V(i, S_i) - \Delta_i(S_i)S_i)$ represents accumulated changes in the risk-free zero bond in the replicating portfolio caused by compounding of the interest, and $\sum_{i=0}^{t-1}\Delta_i(S_i)(S_{i+1} - S_i)$ represents the accumulated changes in the replicating portfolio caused by changes in the risky asset price. The act of constructing a replicating portfolio for a contingent claim is ofter referred to as *Delta hedging*.

The above discussion is summarized in the following algorithm.

Remark 2.3.4. *Given that functions Δ and V are calculated, one has to plug time t and asset price S_t into the function to find the price and adjust the replicating portfolio*

Backward pricing and replicating European options in the binomialmodel

```
At time $T$, the value of the option is $g(S_T(j))$.
for each $t = T-1, ..., 0$ do
    for each $j = 1, ..., t+1$ do
The value of the option $V(t, S_t(j)) = \frac{1}{1+R}\hat{\mathbb{E}}[V(t+1, S_{t+1}) \mid S_t = S_t(j)]$.
The replicating portfolio is made of $\Delta(t, S_t(j))$ units of the risky asset and $V(t, S_t(j)) - S_t(j)\Delta(t, S_t(j))$ is the risk-free bond.
    end for
end for
```

of the contingent claim. However, there is no guarantee that quoted prices in the market will match the prices in the binomial (or any other) model. In such case, interpolation techniques can be exploited to find the price and adjust the replicating portfolio.

Exercise 2.3.2. *In the binomial model, show that the Delta of a call option Δ^{call} and the Delta of a put option Δ^{put} with the same maturity and strike satisfy*

$$\Delta_t^{call} - \Delta_t^{put} = 1, \quad \textit{for all} \quad t = 0, ..., T-1.$$

Is this result model-independent? Hint: consider the put-call parity.

Exercise 2.3.3. *Consider a two-period binomial model for a risky asset with each period equal to a year and take $S_0 = \$1$, $u = 1.03$ and $\ell = 0.98$.*

a) If the interest rate for both periods is $R = 0.01$, find the price of the option with the payoff shown in Figure 2.3.7 with $K_1 = 1.00$ and $K_2 = 1.05$ at all nodes of the binomial model.

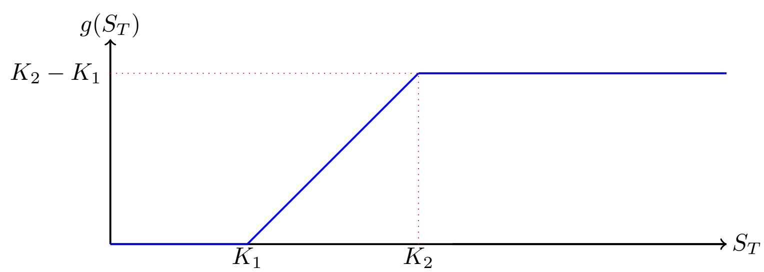


Figure 2.3.7: Payoff of Exercise 2.3.3

b) Find the replicating portfolio at each node of the binomial model.

Exercise 2.3.4. *Consider a two-period binomial model for a risky asset with each period equal to a year and take $S_0 = \$1$ $u = 1.05$ and $\ell = 1.00$. Each year's interest rate comes from Exercise 1.1.4.*

a) Is there any arbitrage? Why or why not? Give an arbitrage portfolio if you find out there is one.

b) Now consider a two-period binomial model for a risky asset with each period equal to a year and take $S_0 = \$1$ $u = 1.05$ and $\ell = 0.95$.

Find the price of the option with the payoff shown in Figure 2.3.7 with $K_1 = 1.00$ and $K_2 = 1.05$ at all nodes of the binomial model. Find the interest rates for each period from information in Part (a).

c) Find the replicating portfolio and specifically Δ at all nodes of the binomial model.

Remark 2.3.5. *2.3.2 suggests that the price of a Markovian claim in a binomial model does not depend on past movements of the price and only depends on the current price S. This is not indeed true for non-Markovian claims. For example, a look-back option with payoff $\left(\max_{t=0,\dots,T} S_t - K\right)_+$ or an Asian option $\left(\frac{1}{T+1}\sum_{t=0}^T S_t - K\right)_+$ are non-Markovian options with the price depending to some extent on the past history of the price movement rather than only the current price of the underlying. Evaluation of these type of non-Markovian payoffs, namely path-dependent payoffs, cannot benefit fully from the Markovian property of the model.*

Example 2.3.3. *Consider the setting of Exercise 2.3.3: $R = 0.01$, $S_0 = \$1$, $u = 1.03$ and $\ell = 0.98$ To price a look-back option with payoff $\left(\max_{t=0,1,2} S_t - 1\right)_+$, first notice that the payoff of the option at time $T = 2$ is path-dependent: the paths $(S_0, S_0u, S_0u\ell)$ and $(S_0, S_0\ell, S_0u\ell)$ generate the same value $S_0u\ell$ for S_2. However, the payoff for the former is*

$$\left(\max\{S_0, S_0u, S_0u\ell\} - 1\right)_+ = \left(\max\{1, 1.03, 1.0094\} - 1\right)_+ = 0.03;$$

while the latter has payoff

$$\left(\max\{S_0, S_0\ell, S_0u\ell\} - 1\right)_+ = \left(\max\{1, 0.98, 1.0094\} - 1\right)_+ = 0.0094.$$

Therefore, the binomial model should be shown as in Figure 2.3.8 It follows from no arbi-

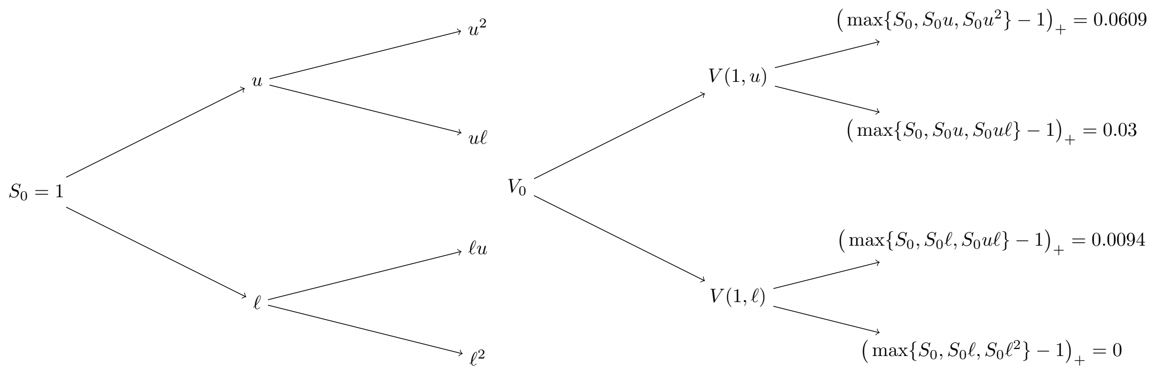


Figure 2.3.8: Look-back option in the binomial model in Example 2.3.3

trage that the price of the look-back option must be a martingale, and therefore,

$$\begin{aligned}
V(1,u) &= \frac{1}{1+R}\hat{\mathbb{E}}\big[g(S_0,S_1,S_2) \mid S_1 = u, S_0 = 1\big] \\
&= \frac{1}{1.01}\Big(\hat{\pi}_u(0.0609) + \hat{\pi}_\ell(0.03)\Big) \approx 0.0480594, \\
V(1,\ell) &= \frac{1}{1+R}\hat{\mathbb{E}}\big[g(S_0,S_1,S_2) \mid S_1 = \ell, S_0 = 1\big] \\
&= \frac{1}{1.01}\Big(\hat{\pi}_u(0.0094) + \hat{\pi}_\ell(0)\Big) \approx 0.0055841, \text{ and} \\
V_0 &= \frac{1}{(1+R)^2}\hat{\mathbb{E}}\big[g(S_0,S_1,S_2) \mid S_0 = 1\big] \\
&= \frac{1}{(1.01)^2}\Big(\hat{\pi}_u^2(0.0609) + \hat{\pi}_u\hat{\pi}_\ell(0.03) + \hat{\pi}_\ell\hat{\pi}_u(0.0094) + \hat{\pi}_\ell^2(0)\Big) \approx 0.0307617.
\end{aligned}$$

Replication is similar to the markovian case. For instance, to replicated the look-back option at $S_1 = u$, we first need to solve the system of equations

$$\begin{cases} \theta_0(1+R) + \theta_1 u &= 0.0609 \\ \theta_0(1+R) + \theta_1 \ell &= 0.03 \end{cases}.$$

However, since the binomial model does not recombine, the number such systems of equations to solve grows exponentially in the number of periods; whereas, in Markovian case the number of such systems of equations grows quadratic.

Exercise 2.3.5. *Consider the setting of Exercise 2.3.3. Price and replicate an Asian option with payoff $\big(\frac{1}{T+1}\sum_{t=0}^{T} S_t - K\big)_+$ with $K = 10$. Hint: The price is path dependent and each path has a payoff.*

Example 2.3.4 (Call and put option)**.** *Consider a call option with strike K and maturity T. Let the nonnegative integer k_0 be such that $S_0 u^{k_0-1}\ell^{T-k_0+1} < K \leqslant S_0 u^{k_0}\ell^{T-k_0}$. Then,*

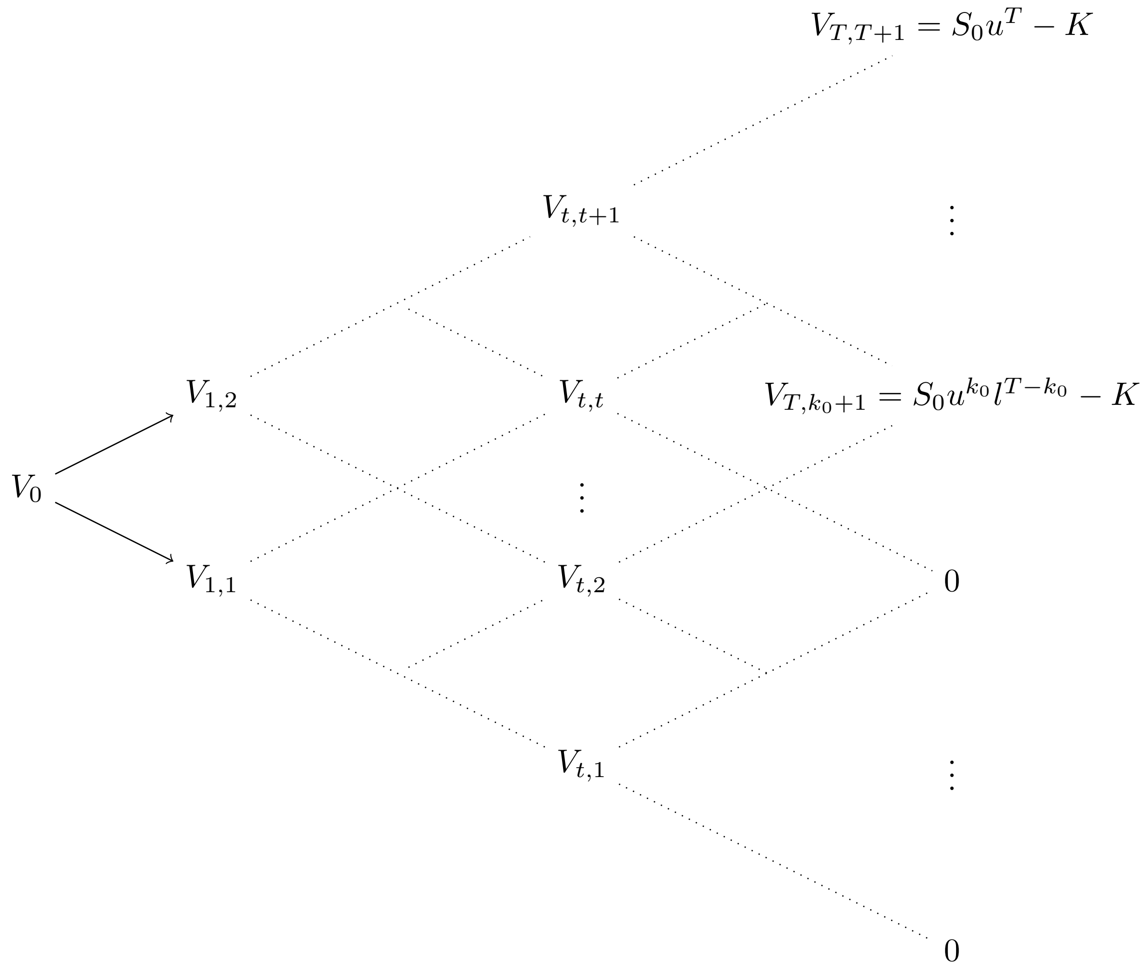


Figure 2.3.9: Payoff of a call option in a binomial model; k_0 is such that $S_0 u^{k_0-1}\ell^{T-k_0+1} < K \leqslant S_0 u^{k_0}\ell^{T-k_0}$.

by (2.3.3)*, we have*

$$V_0^{call} = \frac{1}{(1+R)^T} \sum_{k=k_0}^{T} \binom{n}{k} (\hat{\pi}_u)^{n-k} (\hat{\pi}_\ell)^k (S_T^k - K).$$

Similarly, one can use (2.3.3) *to obtain the price of a put option. However, given that we already have the price of a call option in the above, one can find the price of a put option* *by using the put-call parity (Proposition 1.3.3):*

$$\begin{aligned}
V_0^{put} &= V_0^{call} + \frac{K}{(1+R)^T} - S_0 \\
&= \frac{1}{(1+R)^T} \sum_{k=k_0}^{T} \binom{n}{k} (\hat{\pi}_u)^{n-k} (\hat{\pi}_\ell)^k (S_T^k - K) + \frac{K - \hat{\mathbb{E}}[S_T]}{(1+R)^T} \\
&= \sum_{k=0}^{k_0-1} \binom{n}{k} (\hat{\pi}_u)^{n-k} (\hat{\pi}_\ell)^k (K - S_T^k).
\end{aligned}$$

In the above, we used the martingale property of the discounted asset price, (2.2.4), to write $S_0 = \hat{\mathbb{E}}[\hat{S}_T] = \frac{\hat{\mathbb{E}}[S_T]}{(1+R)^T}$*. Then, we expanded the expectation to write*

$$K - \hat{\mathbb{E}}[S_T] = \hat{\mathbb{E}}[K - S_T] = \sum_{k=0}^{T} \binom{n}{k} (\hat{\pi}_u)^{n-k} (\hat{\pi}_\ell)^k (K - S_T^k).$$

2.3.4 Dividend-paying stock

Stocks usually pay cash dividends to the shareholders. Then, it is up to the individual shareholders to decide whether to consume the cash dividend or invest it back into the market. The dividend policy is determined by the management of the company, but it is also influenced by the preference of the shareholders. Dividends are usually announced in advanced and are paid in a regular basis, quarterly, semiannually, annually and the like. However, when the company announces unexpected high earnings, a special dividend can be paid. Also, a regular dividend can be stopped if the earnings are unexpectedly low. Dividends are announced as a cash amounts; however, for the ease of calculation, we model them as a percentage of the asset price, which is referred to as *dividend yield* and is a number in $[0,1)$. If the asset price at time t is S_t, then, after paying a dividend yield of $d_t \in [0,1)$, the asset price is reduced to $(1-d_t)S_t$. Therefore, under the dividend policy $d_1, ..., d_n$, the asset price dynamics in the binomial model follows

$$S_t = S_0 H_1 ... H_t \prod_{i=1}^{t} (1 - d_i),$$

where H_i is a sequence of i.i.d. random variables with distribution

$$\hat{\mathbb{P}}(H_i = u) = \frac{1 + R - \ell}{u - \ell} \quad \text{and} \quad \hat{\mathbb{P}}(H_i = \ell) = \frac{u - 1 - R}{u - \ell}.$$

If we define $\tilde{H}_i = H_i(1-d_i)$, the dividend policy makes the binomial model look like

$$S_t = S_0\tilde{H}_1...\tilde{H}_t.$$

See Figure 2.3.10. Then, $\tilde{H}_i$ is distributed as

$$\hat{\mathbb{P}}(\tilde{H}_i = u(1-d_i)) = \frac{1+R-\ell}{u-\ell} \text{ and } \hat{\mathbb{P}}(\tilde{H}_i = \ell(1-d_i)) = \frac{u-1-R}{u-\ell}.$$

Remark 2.3.6. *It is important to notice that in this case, the no-arbitrage condition remains the same as $\ell < 1+R < u$, the case where the underlying asset does not pay any dividend. Regarding the effect of dividend on the replicating portfolio, each dividend payment transfers the value invested in the risky asset into the risk-free zero bond. A period before a dividend yield of $d \in [0,1)$, if the asset price is S, the investment in the risk-free zero bond is θ_0, and the investment in the risky asset is θ_1, the replicating portfolio solves a similar system of equations as for the case of no-dividend:*

$$\begin{cases} \underbrace{\theta_0(1+R)+\theta_1 dSu}_{\textit{risk-free}} + \underbrace{\theta_1(1-d)Su}_{\textit{risky}} = V(t+1, Su(1-d)) \\ \underbrace{\theta_0(1+R)+\theta_1 dS\ell}_{\textit{risk-free}} + \underbrace{\theta_1(1-d)S\ell}_{\textit{risky}} = V(t+1, S\ell(1-d)) \end{cases}.$$

One way of pricing contingent claims on dividend-paying assets is to introduce an adjusted asset price with no dividend and transform the payoff in terms of the adjusted asset price. If we define the adjusted asset price by $\tilde{S}_t := S_0H_1...H_t$, then $S_t = \tilde{S}_t\prod_{i=1}^{T}(1-d_i)$. Then, the payoff of a European contingent claim given by $g(S_T)$ is given in terms of $\tilde{S}_T$ by $g\big(\tilde{S}_T\prod_{i=1}^{T}(1-d_i)\big)$ on the adjusted binomial asset $\tilde{S}_T$. Then, the pricing of a contingent claim with payoff $g(S_T)$ at time t given $S_t = S$ is given by

$$\begin{aligned} V(t,S) &:= \frac{1}{(1+R)^{T-t}}\hat{\mathbb{E}}[g(S_T) \mid S_t = S] = \frac{1}{(1+R)^{T-t}}\hat{\mathbb{E}}\Big[g\big(\tilde{S}_T\prod_{j=t+1}^{T}(1-d_i)\big) \mid S_t = S\Big] \\ &= \frac{1}{(1+R)^{T-t}}\sum_{i=t}^{T}\binom{T-t}{i}\hat{\pi}_u^i\hat{\pi}_\ell^{T-t-i}g\Big(Su^i\ell^{T-t-i}\prod_{j=t+1}^{T}(1-d_j)\Big). \end{aligned} \tag{2.3.5}$$

Remark 2.3.7. *Notice that the price $V(t,S)$ of a European contingent claim $g(S_T)$ on a dividend-paying underlying asset S is no longer a function of time-to-maturity $\tau = T-t$. This is due to the term $\prod_{j=t+1}^{T}(1-d_i)$ in (2.3.5), which cannot be expressed as a function of τ, unless $d_i = d$ for all $i = 1,...,T$; then, $\prod_{j=t+1}^{T}(1-d_i) = (1-d)^\tau$.*

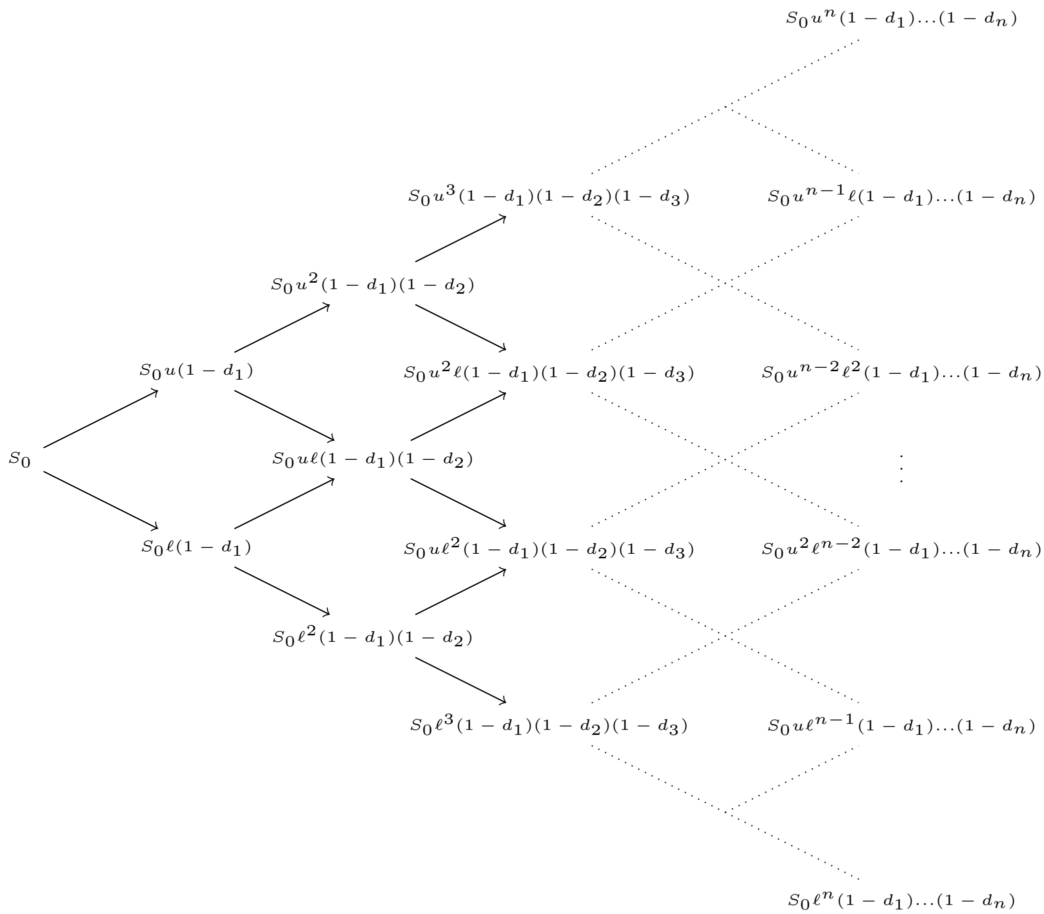


Figure 2.3.10: Dividend-paying asset in a binomial model

Remark 2.3.8. *Recall that a regular dividend policy is announced in cash and not the dividend yield. This means for a high (low) asset price the cash dividend is equivalent to a small (large) dividend yield. A dividend cash of \$D corresponds to a dividend yield of* $\frac{D}{S}$*, where* S *is the predividend asset price. In addition, a bigger picture of dividend payments also suggests that the companies can change their dividend policies based on certain random events. Therefore, it is natural to assume that the divided policy is random. If the dividend policy is Markovian, at each node of the binomial model at time* t *and price* $S_t(j)$*, the divided is a random variable* $d(t, S_t(j))$*, for* $j = 1, ..., t+1$*, the pricing methodology is similar to the Markovian European option. In the two-period binomial model in of Exercise 2.3.6, Part (b) and Part (c) are special instances of Markovian dividends.*

If the dividend policy is not Markovian, even for a Markovian asset-price model and a Markovian contingent claim with payoff $g(S_T)$*, the pricing is similar to pricing methodology* *of path-dependent options. This is because the modified payoff takes the form* $g\big(\tilde{S}_T \prod_{t=1}^{T}(1-d_t)\big)$*, where* d_t *depends on* $S_t, ..., S_0$*. Therefore, pricing and hedging must be conducted similar to Remark 2.3.5, Example 2.3.3 and Exercise 2.3.5.*

Exercise 2.3.6. *Consider a two-period binomial model for a risky asset with each period equal to a year and take* $S_0 = \$10$*,* $u = 1.15$ *and* $\ell = 0.95$*. The interest rate for both periods is* $R = .05$*.*

a) If the asset pays a 10% dividend yield in the first period and 20% in the second period, find the price of a call option with strike $K = 8$*.*

b) Consider a more complicated dividend policy that pays a10% dividend yield only if the price moves up and no dividend if the price moves down in each period. Find the price of a call option with strike $K = 8$*.*

c) Finally, consider a dividend policy that pays \$1 divided in the first period in each period. Find the price of a call option with strike $K = 8$*.*

2.4 Calibrating the parameters of the model to market data: the binomial model

Calibration is the practice of matching the parameters of a model to data. In the binomial model, the parameters are interest rate R, u, and ℓ. Calibrating the risk-free interest rate R is a separate job and usually uses the price quotes of risk-free (sovereign) zero bonds. For the purposes of this section, we assume that the yield of a zero bond is already calibrated and satisfies $R_t(t+\delta) = r\delta + o(\delta)$, where short rate r is constant and δ represents the duration of one period in the binomial market. δ is usually small relative to the maturity T.

Data: price quote and return

Assume that the asset price quotes are collected at δ time lapse; $S_{-m\delta}, ..., S_0$ tabulates the past quotes of the asset price from time $-m\delta$ until the present time $t = 0$. We denote the quoted price by S to distinguish it from the random variable S for future price.

The arithmetic return and the logarithmic return at time t are defined by

$$\mathbf{R}_t^{\text{arth}} := \frac{S_{t+\delta} - S_t}{S_t},$$

and

$$\mathbf{R}_t^{\log} := \ln\left(\frac{S_{t+\delta}}{S_t}\right),$$

respectively. If the time step δ is small, then the price movement $S_{t+\delta} - S_t$ is also small, and $\mathbf{R}_t^{\log}$ and $\mathbf{R}_t^{\text{arth}}$ are very close[12]. However, we will see that the small difference between the two returns will show up in the parameter estimation. For the moment, we focus on the arithmetic return and drop the superscript "arth" for simplicity. We discuss the logarithmic return (or log return) at the end of this section.

From the data points $S_{-m\delta}, ..., S_0$, we obtain data points for the return, $\mathbf{R}_{-m}, ..., \mathbf{R}_{-1}$ given by

$$\mathbf{R}_{-k} := \frac{S_{(-k+1)\delta} - S_{-k\delta}}{S_{-k\delta}}.$$

Next, we will use this data to estimate some parameters of the market that are important in the calibration process.

Binomial model with physical probability

The data on the quoted prices comes from the physical probability and not the risk-neutral probability; see Remark 2.1.5. Therefore, for calibration, we need to present the binomial model with physical probability rather than the risk-neutral probability. While the nodes of the binomial model will not change, the probability must change; $S_{t+1} = S_t H_{t+1}$ and $\{H_t\}_{t=1}^{\infty}$ is a sequence of i.i.d. random variables with the following distribution under physical measure

$$H_t = \begin{cases} u & \text{with probability } p \\ \ell & \text{with probability } 1-p \end{cases}, \quad \text{for all } \; t = 0, 1, ... \tag{2.4.1}$$

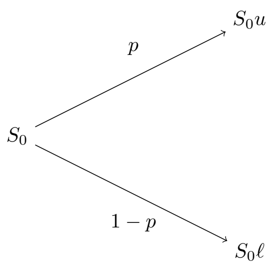
In the binomial model under physical probability measure, the sequence of returns $\{\mathbf{R}_t\}_{t=0}^{\infty}$ also makes a sequence of i.i.d. random variables. So, to proceed with calibration, we need to impose the same assumption on the data.

Assumption 2.4.1. *The return $\{\mathbf{R}_t\}_{t=0}^{\infty}$ is a sequence of i.i.d. random variables with the mean and the variance given by by $\mu\delta + o(\delta)$ and $\sigma^2\delta + o(\delta)$, respectively.*

The dimensionless quantities σ and μ are respectively called the *volatility* and *mean return rate* of the price.

Remark 2.4.1. *The assumption that volatility is a constant is not very realistic. However, this assumption, which was widely used in practice in the 1970s and 1980s, makes the problems more tractable. We will try later to test some approaches that relax this assumption in different directions.*

[12] $\ln(1+x) \approx x$ for small x



$$\mathbf{R}_t = H_t - 1 = \begin{cases} U := u-1 & \text{with probability } p \\ \\ L := \ell - 1 & \text{with probability } 1-p \end{cases}$$

Figure 2.4.1: Left: binomial model under physical probability. Right: arithmetic return $\mathbf{R}_t$.

Statistical estimation of return and volatility

We proceed by introducing some basic statistical methods to estimate μ and σ. The simplest estimators for these parameters come from Assumption 2.4.1:

$$\hat{\mu} = \frac{1}{\delta m}\sum_{k=1}^{m}\hat{\mathbf{R}}_{-k} \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{\delta(m-1)}\sum_{k=1}^{m}(\hat{\mathbf{R}}_{-k} - \hat{\mu}\delta)^2.$$

Exercise 2.4.1 (Project)**.** *Go to Google Finance, Yahoo Finance, or any other free database that provides free asset price quotes. Download a spreadsheet giving the daily price of a highly liquid asset such as IBM, Apple, Alphabet, etc. Assuming Assumption 2.4.1, find the volatility σ and the average return rate μ of the asset. Then, use these quantities to find the daily, weekly, and yearly standard deviation and mean of the return.*

A calibration of the model parameters u, ℓ, and p

We match the first and second momentum of the binomial model with the mean return rate and volatility:

$$\begin{cases} pU + (1-p)L = \mu\delta + o(\delta) \\ pU^2 + (1-p)L^2 - (pU + (1-p)L)^2 = \sigma^2\delta + o(\delta) \end{cases}$$

Or, equivalently,

$$\begin{cases} pU + (1-p)L = \mu\delta + o(\delta) \\ p(1-p)(U-L)^2 = \sigma^2\delta + o(\delta) \end{cases} \tag{2.4.2}$$

In the above system of two equations, there are three unknowns U, L, and p, which give us one degree of freedom. We are going to use this degree of freedom by assuming that **the variance of return $\mathbf{R}_t$ under risk-neutral probability is also** $\sigma^2\delta + o(\delta)$;

$$\hat{\pi}_u\hat{\pi}_\ell(U-L)^2 = \sigma^2\delta + o(\delta), \tag{2.4.3}$$

where $\hat{\pi}_u = \frac{1+R-\ell}{u-\ell} = \frac{R-L}{U-L}$.

Remark 2.4.2. *This assumption suggests that the variance of return under the risk-neutral probability measure does not deviate significantly from the physical probability measure. Whether we are allowed to make such a strong assumption or not is debatable. However, evidences from financial economics as well as statistical modeling of financial markets suggest that this assumption is practical and significantly useful. In addition, in the continuous-time modeling of financial markets, the assertion of this assumption becomes a conceptually deep result. The reason this assumption will become clear later when we study continuous-time models, specifically the Black-Scholes model.*

To simplify further, we also drop the $o(\delta)$ term from the equations. Therefore, we have a system of three equations and three unknowns.

$$\begin{cases} pU + (1-p)L = \mu\delta \\ p(1-p)(U-L)^2 = \sigma^2\delta \\ (U-R)(R-L) = \sigma^2\delta \end{cases}$$

For the sake of simplicity, we set new variables

$$\alpha = \frac{U-R}{\sqrt{\delta}\sigma}, \quad \beta = \frac{R-L}{\sqrt{\delta}\sigma}, \quad \text{and} \quad r = \frac{R}{\delta}, \tag{2.4.4}$$

where r is the *annual interest rate (APR) calculated at periods* δ. Thus, we have

$$\begin{cases} p\alpha - (1-p)\beta = \lambda\sqrt{\delta} \\ \sqrt{p(1-p)}(\alpha+\beta) = 1 \\ \alpha\beta = 1 \end{cases} \tag{2.4.5}$$

Here $\lambda := \frac{\mu-r}{\sigma}$.

Remark 2.4.3 (Risk premium)**.** *The quantity* $\frac{\mu-r}{\sigma}$ *is referred to as the risk premium of the asset and measures the excess mean return of the asset adjusted with its level of riskiness measured by its volatility.*

The solution to (2.4.5) is given by

$$\alpha = \sqrt{\frac{1-p}{p}} + \lambda\sqrt{\delta}, \quad \beta = \sqrt{\frac{p}{1-p}} - \lambda\sqrt{\delta}, \quad \text{and} \quad p = \frac{1}{1+x_0^2}, \tag{2.4.6}$$

where x_0 is the unique positive solution to the equation (see Figure 2.4.2)

$$x - \frac{1}{x} = \lambda\sqrt{\delta}. \tag{2.4.7}$$

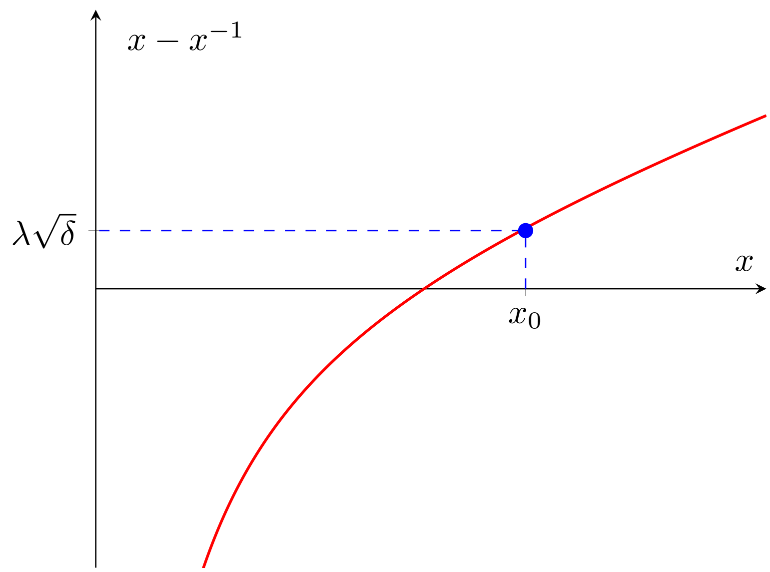


Figure 2.4.2: Positive solution to equation $x - \frac{1}{x} = \lambda\sqrt{\delta}$.

Then, u and ℓ are given by

$$u = 1 + \delta r + \sqrt{\delta}\sigma\alpha, \quad \text{and} \quad \ell = 1 + \delta r - \sqrt{\delta}\sigma\beta. \tag{2.4.8}$$

Remark 2.4.4. *The above calibration is well posed in the sense that, for any possible value of parameters $\sigma > 0$, μ, and r, one can find proper u and ℓ such that $\ell < 1 + R < u$ in a unique fashion. Notice that in (2.3.3), the pricing of contingent claims is not affected by p, and thus p is the least important parameter in this context.*

Example 2.4.1 (Symmetric probabilities)**.** *We shall show that the following choice of parameters for the binomial model also provides a calibration; it satisfies the Assumption 2.4.1.*

$$u = e^{\delta(r-\frac{\sigma^2}{2})+\sqrt{\delta}\sigma}, \quad \ell = e^{\delta(r-\frac{\sigma^2}{2})-\sqrt{\delta}\sigma} \quad \textit{and} \quad p = \frac{1}{2}\Big(1 + \lambda\sqrt{\delta}\Big).$$

First notice that

$$\mathbf{R}_t = \begin{cases} e^{\delta(r-\frac{\sigma^2}{2})+\sqrt{\delta}\sigma} - 1 & \textit{with probability} \quad \frac{1}{2}\Big(1 + \lambda\sqrt{\delta}\Big) \\ e^{\delta(r-\frac{\sigma^2}{2})-\sqrt{\delta}\sigma} - 1 & \textit{with probability} \quad \frac{1}{2}\Big(1 - \lambda\sqrt{\delta}\Big) \end{cases}$$

Therefore,

$$\mathbb{E}[\mathbf{R}_t] = \frac{1}{2}\bigg(\Big(e^{\delta(r-\frac{\sigma^2}{2})+\sqrt{\delta}\sigma} - 1\Big)\Big(1 + \lambda\sqrt{\delta}\Big) + \Big(e^{\delta(r-\frac{\sigma^2}{2})-\sqrt{\delta}\sigma} - 1\Big)\Big(1 - \lambda\sqrt{\delta}\Big)\bigg).$$

We use $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$, *to write*

$$\begin{aligned}\mathbb{E}[\mathbf{R}_t] =&\frac{1}{2}\bigg(\Big(\delta(r - \frac{\sigma^2}{2}) + \sqrt{\delta}\sigma + \frac{\delta\sigma^2}{2} + o(\delta)\Big)\Big(1 + \lambda\sqrt{\delta}\Big)\\ &+ \Big(\delta(r - \frac{\sigma^2}{2}) - \sqrt{\delta}\sigma + \frac{\delta\sigma^2}{2} + o(\delta)\Big)\Big(1 - \lambda\sqrt{\delta}\Big)\bigg)\\ =&\delta(r - \frac{\sigma^2}{2} + \frac{\sigma^2}{2} + \lambda\sigma) + o(\delta) = \mu\delta + o(\delta)\end{aligned}$$

Recall that $\lambda\sigma = \mu - r$. *On the other hand,*

$$\begin{aligned}\mathbb{E}[\mathbf{R}_t^2] =&\frac{1}{2}\bigg(\Big(e^{\delta(r-\frac{\sigma^2}{2})+\sqrt{\delta}\sigma} - 1\Big)^2\Big(1 + \lambda\sqrt{\delta}\Big) + \Big(e^{\delta(r-\frac{\sigma^2}{2})-\sqrt{\delta}\sigma} - 1\Big)^2\Big(1 - \lambda\sqrt{\delta}\Big)\bigg)\\ =&\frac{1}{2}\bigg(\Big(\delta\sigma^2 + o(\delta)\Big)\Big(1 + \lambda\sqrt{\delta}\Big) + \Big(\delta\sigma^2 + o(\delta)\Big)\Big(1 - \lambda\sqrt{\delta}\Big)\bigg) = \sigma^2\delta + o(\delta).\end{aligned}$$

Exercise 2.4.2 (Subjective return)**.** *Show that the following choice of parameters for the binomial model also provides a calibration; it satisfies the Assumption 2.4.1.*

$$u = e^{\delta\nu+\sqrt{\delta}\sigma}, \quad \ell = e^{\delta\nu-\sqrt{\delta}\sigma} \quad \text{and} \quad p = \frac{1}{2}\Big(1 + (\frac{\mu-\nu}{\sigma} - \frac{\sigma}{2})\sqrt{\delta}\Big),$$

where ν *is a real number. Find the range of* ν *in terms of other parameters such that there is no arbitrage.*

Calibration the binomial model for the purpose of pricing contingent claims

Estimating rate of return is often a more difficult task than estimating volatility. Methods such as the CAPM[13] have been developed to approximate the rate of return. However, if the calibration is only used for option pricing, we often do not need the rate of return. Recall that under risk-neutral probability, the return of a binomial asset is equal to the yield of a zero bond, which is already estimated through the bond market data;

$$\hat{\mathbb{E}}[\mathbf{R}_t] = \hat{\mathbb{E}}[H_t - 1] = (u-1)\hat{\pi}_u + (\ell-1)\hat{\pi}_\ell = \frac{(u-1)(1+R-\ell) + (\ell-1)(u-1-R)}{u-\ell} = R.$$

Therefore, if we only calibrate the binomial model under risk-neutral probability, the rate of return under physical probability is irrelevant.

We impose the following assumption on the distribution of the return process under the risk-neutral probability and use it to calibrate the binomial model under risk-neutral probability using data to estimate the volatility only.

[13] the capital asset pricing model

Assumption 2.4.2. *The arithmetic return* $\{\mathbf{R}_t\}_{t=0}^{\infty}$ *is a sequence of independent random variables under risk-neutral probability* $\hat{\mathbb{P}}$ *with mean* $\hat{\mathbb{E}}[\mathbf{R}_t] = r\delta + o(\delta)$ *and* $\hat{\text{var}}(\mathbf{R}_t) = \sigma^2\delta + o(\delta)$*.*

This assumption implies the assumption that we imposed earlier in (2.4.3).
Given r and $R = r\delta + o(\delta)$, the only equation that we obtain is

$$(U - R)(R - L) = \sigma^2\delta.$$

Here, we drop the $o(\delta)$ term for simplicity. If we define $\alpha = \frac{U-R}{\sqrt{\delta}\sigma}$ and $\beta = \frac{R-L}{\sqrt{\delta}\sigma}$, we have $\alpha\beta = 1$. Any choice for α leads to $\beta = \frac{1}{\alpha}$, and therefore, is a calibration of the binomial model. We obtain a calibration similar to (2.4.8):

$$u = 1 + \delta r + \sqrt{\delta}\sigma\alpha, \quad \text{and} \quad \ell = 1 + \delta r - \sqrt{\delta}\sigma\beta.$$

Here, α and β are different than in 2.4.8.

Remark 2.4.5. *In practice, Assumption 2.4.2 is too good to be true. In fact, most of the arguments for calibration of the parameters hold without appealing to such a strong assumption. The main use of this assumption in this section is to estimate parameters* μ *and* σ*, the mean return and volatility. One can find estimation of these parameters by assuming that there exists a martingale* $\{M_t\}_{t=0}^{\infty}$ *such that* $\mathbf{R}_t - r\delta = M_{t+1} - M_t$*.*

Time-varying return and volatility

Assumption 2.4.1, which asserts that $\{\mathbf{R}_t\}_t$ is a sequence of i.i.d. random variables, is not realistic in some situations and must be relaxed. Several empirical studies show that the volatility is not constant. This removes the "identical distribution" of the return sequence. The independence condition also does not have an empirical basis. In this section, we keep the independence assumption but remove the part that says $\{\mathbf{R}_t\}_t$ is identically distributed. We also allow for the interest rate to depend on time $R_t = r_t\delta$. Under either of the following assumptions, we can derive a calibration of the form

$$u_t = 1 + \delta r_t + \sqrt{\delta}\sigma_t\alpha_t, \quad \text{and} \quad \ell_t = 1 + \delta r_t - \sqrt{\delta}\sigma_t\beta_t. \tag{2.4.9}$$

Assumption 2.4.3. *The arithmetic return* $\{\mathbf{R}_t\}_t$ *is a sequence of independent random variables with mean* $\mathbb{E}[\mathbf{R}_t] = \mu_t\delta + o(\delta)$ *and* $\text{var}(\mathbf{R}_t) = \sigma_t^2\delta + o(\delta)$*.*

Assumption 2.4.4. *The arithmetic return* $\{\mathbf{R}_t\}_t$ *is a sequence of independent random variables under risk-neutral probability* $\hat{\mathbb{P}}$ *with mean* $\hat{\mathbb{E}}[\mathbf{R}_t] = \mu_t\delta + o(\delta)$ *and* $\hat{\text{var}}(\mathbf{R}_t) = \sigma_t^2\delta + o(\delta)$*.*

Assumption 2.4.3 allows for the parameters μ and σ to vary over time. Therefore, the calibration in Section 2.4, should be modified such that α, β, and p vary in time and satisfy

$$\begin{cases} p_t\alpha_t - (1-p_t)\beta_t = \lambda_t\sqrt{\delta} \\ \sqrt{p_t(1-p_t)}(\alpha_t+\beta_t) = 1 \\ \alpha_t\beta_t = 1 \end{cases}.$$

Here $\lambda_t := \frac{\mu_t - r_t}{\sigma_t}$. If μ, σ, and r vary with time but λ remain does not, the calibration in (2.4.9), becomes slightly simpler, because α and β does not depend on time.

$$u_t = 1 + \delta r_t + \sqrt{\delta}\sigma_t\alpha, \quad \text{and} \quad \ell_t = 1 + \delta r_t - \sqrt{\delta}\sigma_t\beta. \tag{2.4.10}$$

Assumption 2.4.4 provides a different calibration of the same form as (2.4.10) but without assuming that λ is time-invariant. Here, we can arbitrarily choose α and $\beta = \frac{1}{\alpha}$.

Estimating time-varying parameters falls into the time series analysis which is beyond the scope of this book. For more details on the estimation of financial time series, see [14].

3

Modeling financial assets in continuous-time

In 1900, Louis Bachelier introduced the first asset price model and pricing method for derivatives in his Ph.D. dissertation titled "The theory of speculation" [2]. Although this model is now considered impractical, its educational implications of this model are still important. Bachelier modeled the discounted asset price by a Brownian motion. As seen in Section B.5, the Brownian motion is the weak limit of a normalized random walk. In Bachelier's time, the Brownian motion had not yet been rigorously defined. However, many of its properties were well understood. Bachelier's contribution to the theory of probability and stochastic processes was to use heat equation in derivative pricing. However, this contribution was neglected for about thirty years, until Andrey Nikolaevich Kolmogorov employed partial differential equations to describe a class of stochastic processes called *diffusion processes*. Kolmogorov is the first mathematician to bring probability into rigor by establishing its mathematical foundation. Other mathematicians who built upon Kolmogorov's work include Norbert Wiener, the first to discover the path properties of Brownian motion, Paul Lévy, who provided a simple characterization of Brownian motion, and Kiyosi Itô[1], who introduced a simple representation of diffusion processes in terms of Brownian motion.

The Bachelier model has a major drawback: asset price in this model can take negative values, which will be discussed in more details in Section 3.2.3. Almost fifty years after Bachelier, Paul Samuelson, an economist, suggested to use geometric Brownian motion (GBM) to model the price of assets. GBM never takes nonpositive values for the asset price, and therefore, does not suffer from the major drawback of Bachelier model. GBM is also known as Black-Scholes model, named after Fischer Black and Myron Scholes. Black and Scholes in [5] and Robert Merton in [22] independently developed a pricing method for derivatives under the GBM. For a through review of Bachelier's efforts and contribution

[1] You may find different romanization of Kiyosi such as Kiyoshi, and different romanization Itô such as Itō, Itoh, or Ito. According to Wikipedia, he himself ued the spelling Kiyosi Itô.

see [26]. For a brief history of asset price models, see [25]. For Bachelier's biography of see [29].

3.1 Trading and arbitrage in continuous-time markets

Recall from Section II.2.2.2 that in a discrete-time market, where trading occurs at points $t_0 = 0 < t_1 < ... < t_N = T$ in time, the value of the portfolio generated by the strategy $\Delta_0, \Delta_{t_1}, ..., \Delta_{t_{N-1}}$ is given by

$$W_{t_n} = W_0 + \sum_{i=0}^{n-1} R_i(W_{t_i} - \Delta_{t_i} S_{t_i}) + \sum_{i=0}^{n-1} \Delta_{t_i}(S_{t_{i+1}} - S_{t_i}).$$

Here, R_i is the interest rate for the period of time $[t_i, t_{i+1}]$ which can be taken to be $r(t_{i+1} - t_i)$, where r is the short rate. For $i = 0, ..., N-1$, the strategy Δ_{t_i} is a function of the history of the asset price, $\{S_u : u \leqslant t_i\}$. Therefore,

$$W_{t_n} = W_0 + r\sum_{i=0}^{n-1}(W_{t_i} - \Delta_{t_i} S_{t_i})(t_{i+1} - t_i) + \sum_{i=0}^{n-1} \Delta_{t_i}(S_{t_{i+1}} - S_{t_i}).$$

If we take $t_{i+1} - t_i := \delta = \frac{T}{N}$ and let $\delta \to 0$, we obtain the Riemann integral

$$\lim_{\delta \to 0} r \sum_{i=0}^{N-1}(W_{t_i} - \Delta_{t_i} S_{t_i})(t_{i+1} - t_i) = r\int_0^T (W_t - \Delta_t S_t)dt,$$

which is the accumulated net change in the portfolio due to investment in the risk-free asset. The limit of the second term

$$\sum_{i=0}^{n-1} \Delta_{t_i}(S_{t_{i+1}} - S_{t_i})$$

does not necessarily exist unless we enforce proper assumptions on the asset price S. For instance if the asset price follows Brownian motion (the Bachelier model) or GBM (the Black-Scholes model), then the limit exists and is interpreted as a stochastic Itô integral. If the stochastic integral above is well defined, the wealth generated by the trading strategy $\{\Delta\}_{t\geqslant 0}$ follows

$$W_t = W_0 + r\int_0^t (W_s - \Delta_s S_s)ds + \int_0^t \Delta_s dS_s. \tag{3.1.1}$$

We choose the same notion $(\Delta \cdot S)_t := \int_0^t \Delta_s dS_s$ for the stochastic integral in continuous time. Notice that in the discrete-time setting, the strategy Δ at time t_i is a function of the past history of the asset price $S_0, ..., S_{t_i}$. When we pass to the limit, the trading strategy

Δ_t at time t, depends only on the paths of the asset until time t, $\{S_u : u \leqslant t\}$. In other words, a trading strategy does not use any information from the future.

Following the discussion in Section C.1, specifically (C.1), the stochastic integral is defined only in the *almost surely* sense and can only be defined on the sample paths of the asset price model S. Therefore, a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is needed. It is convenient to take the sample path space with a proper σ-field and a probability measure:

$$\Omega := \{S : t \to S_t(\omega) : \text{ for each outcome } \omega\}$$

Then, the notion of arbitrage is defined as follows:

Definition 3.1.1. *A (weak) arbitrage opportunity is a portfolio Δ such that*

a) $W_0 = 0$,

b) $W_T \geqslant 0$ a.s., and

c) $W_T > 0$ on a set of sample paths with positive $\mathbb{P}$ probability.

A strong arbitrage opportunity is a portfolio Δ such that

a) $W_0 < 0$ a.s., and

b) $W_T \geqslant 0$ a.s.

Given that the stochastic integral $(\Delta \cdot S)_t = \int_0^t \Delta_s dS_s$ is defined in a probability space $(\Omega, \mathbb{P})$, the fundamental theorem of asset pricing (FTAP) for continuous time is as follows. Two probabilities $\hat{\mathbb{P}}$ and $\mathbb{P}$ are called equivalent if any event with probability zero under one of them has probability zero under the other, i.e., $\hat{\mathbb{P}}(A) = 0$ if and only if $\mathbb{P}(A) = 0$.

Theorem 3.1.1 (Fundamental theorem of asset pricing (FTAP))**.** *There is no weak arbitrage opportunity in a continuous-time model if and only if there exists a probability $\hat{\mathbb{P}}$, namely risk-neutral probability or martingale probability, equivalent to $\mathbb{P}$ such that S is a $\hat{\mathbb{P}}$-(local) martingale*[2]*.*

Recall that a market model is called complete if any contingent claim is replicable, i.e., for any $\mathcal{F}$-measurable payoff X, there exists a strategy $\Delta := \{\Delta_t\}_{t=0}^T$ such that the wealth W generated by Δ in (3.1.1) satisfies $W_T = X$.

Corollary 3.1.1. *Under the same setting as in Theorem 3.1.1, the market is complete if and only if there is a unique risk-neutral probability.*

[2] local martingale is roughly a martingale without condition (a) in Definition B.15.

3.2 Bachelier's continuous-time market

We start by recalling the properties of a Brownian motion from Section B.5. A standard Brownian motion is a *stochastic process* such that $B_0 = x \in \mathbb{R}^d$ and characterized by the following properties.

1) B has continuous sample paths.

2) When $s < t$, the increment $B_t - B_s$ is a normally distributed random variable with mean 0 and variance $t - s$ and is independent of $\{B_u \ : \ \text{for all} \ \ u \leqslant s\}$.

The Bachelier model is based on an assumption made by Bachelier himself in his PhD dissertation"Théorie de la spèculation" ([2]):

$$\text{L'espèrance mathématique de l'acheteur de prime est nulle (page 33)} \tag{3.2.1}$$

which translates as "The mathematical expectation of the buyer of the asset is zero". In the modern probabilistic language, what Bachelier meant is that the discounted asset price is a **martingale** under a unique risk-neutral probability, which is equivalent to no arbitrage condition by the FTAP in [9]. Without any discussion on the physical probability measure, the Bachelier model simply takes the discounted asset price under the risk-neutral probability as a factor of a standard Brownian motion, i.e.

$$\hat{S}_t = e^{-rt} S_t = S_0 + \sigma B_t, \quad \sigma > 0. \tag{3.2.2}$$

Therefore under risk-neutral probability, $\hat{S}_t$ has a Gaussian distribution with mean S_0 and variance $\sigma^2 t$, and the pdf of $\hat{S}_t$ is given by

$$f_{\hat{S}_t}(x) = \frac{1}{\sigma\sqrt{2\pi t}} \exp\left(-\frac{(x - S_0)^2}{2\sigma^2 t}\right) \quad \text{for} \ \ x \in \mathbb{R} \quad \text{and} \quad t > 0. \tag{3.2.3}$$

The price of the underlying asset in the Bachelier model is given by $S_t = e^{rt}(S_0 + \sigma B_t)$. Under risk-neutral probability, S_t is also a Gaussian random variable with

$$\hat{\mathbb{E}}[S_t] = S_0 e^{rt}, \quad \text{and} \quad \text{var}(S_t) = \sigma^2 e^{2rt} t.$$

Therefore, the risk-neutral expected value of the asset price increases in a similar fashion as a risk-free asset. The variance of S_t increases exponentially quickly in time, too. It follows from applying the Itô formula (C.4) to $f(t, x) = e^{rt}(S_0 + \sigma x)$ that S_t satisfies the SDE

$$dS_t = rS_t dt + \sigma e^{rt} dB_t. \tag{3.2.4}$$

Inherited from the Brownian motion, the Bachelier model possesses the same properties as the binomial model in Section 2.3.2:

1) time homogeneity and

The conditional distribution of $\hat{S}_t$ given $\hat{S}_s = S$ equals the distribution of $\hat{S}_{t-s}$ with $\hat{S}_0 = S$

2) Markovian.

In using the available information up to time s to make projections about the future, the only relevant part is the asset price at current time $\hat{S}_s$ (or, equivalently S_s). $\hat{\mathbb{E}}[f(\hat{S}_t) \mid \hat{S}_u,\ u \leqslant s] = \mathbb{E}[f(\hat{S}_t) \mid \hat{S}_s], \qquad \forall 0 \leqslant s < t.$

The condition $\{\hat{S}_u,\ u \leqslant s\}$ (or $\hat{S}_s$) in the above conditional expectations can simply be replaced by $\{S_u,\ u \leqslant s\}$ (or S_s). Because both sets of random variables generate the same σ-field $\mathcal{F}_s := \sigma\big(S_u\ :\ u \leqslant s\big)$:

$$\hat{\mathbb{E}}[\cdot \mid \mathcal{F}_s] = \hat{\mathbb{E}}[\cdot \mid S_s].$$

It is important to know that the Bachelier model is not practically interesting in modeling financial markets. However, for educational purposes, it has all the basic components of the more practical Black-Scholes model. Pricing a European option in the Bachelier model is equivalent to solving a heat equation, a simple parabolic partial differential equation. On one hand, the heat equation also appears in the simplification of the Black-Scholes model. On the other hand, most of the knowledge and techniques used in the heat equation can also be applied to other forms of parabolic partial differential equations that appear in more general models.

3.2.1 Pricing and replicating contingent claims in the Bachelier model

As a result of quote (3.2.1), Bachelier concluded that the price of a European contingent claim with payoff $g(S_T)$ is simply the discounted expectation of the payoff under the risk-neutral probability, i.e.

$$V_0 = e^{-rT}\hat{\mathbb{E}}[g(S_T)]. \tag{3.2.5}$$

In addition, given the past history of asset price $\mathcal{F}_t := \sigma\big(S_u\ :\ u \leqslant t\big)$, the price of this option at time t is given by

$$V_t = e^{-r(T-t)}\hat{\mathbb{E}}[g(S_T) \mid \mathcal{F}_t]. \tag{3.2.6}$$

This is inline with the fundamental theorem of asset pricing that asserts the equivalency of no arbitrage condition with the existence of a risk-neutral probability. Precisely, the discounted price of an option with payoff $g(S_T)$ must be a martingale:

$$e^{-rt}V_t = e^{-rT}\hat{\mathbb{E}}[V_T \mid \mathcal{F}_t] = e^{-rT}\hat{\mathbb{E}}[g(S_T) \mid \mathcal{F}_t]. \tag{3.2.7}$$

Price of a contingent claim at time zero

Since under risk-neutral probability $S_T = e^{rT}\hat{S}_T$ is a Gaussian random variable with mean $e^{rT}S_0$ and variance $e^{2rT}\sigma^2 T$, one can explicitly calculate V_0 in cases where the following integral can be given in a closed-form:

$$V_0 = \frac{e^{-rT}}{\sigma\sqrt{2\pi T}} \int_{-\infty}^{\infty} g\left(e^{rT}x\right) e^{-\frac{(x-S_0)^2}{2\sigma^2 T}} dx. \tag{3.2.8}$$

We start off by providing a closed-form solution for the Bachelier price of vanilla options.

Example 3.2.1 (Price of call and put in the Bachelier model). *Let $g(S_T) = (S_T - K)_+ = e^{rT}(\hat{S}_T - \hat{K})_+$, where $\hat{K} = e^{-rT}K$. Since $\hat{S}_T \sim N(S_0, \sigma^2 T)$, the price V_0^{call} can be calculated in closed form.*

$$\begin{aligned} V_0^{call} &= \frac{e^{-rT}}{\sigma\sqrt{2\pi T}} \int_{-\infty}^{\infty} \left(e^{rT}x - K\right)_+ e^{-\frac{(x-S_0)^2}{2\sigma^2 T}} dx \\ &= \frac{1}{\sigma\sqrt{2\pi T}} \int_{-\infty}^{\infty} \left(x - \hat{K}\right)_+ e^{-\frac{(x-S_0)^2}{2\sigma^2 T}} dx \\ &= \frac{1}{\sigma\sqrt{2\pi T}} \int_{\hat{K}}^{\infty} \left(x - \hat{K}\right) e^{-\frac{(x-S_0)^2}{2\sigma^2 T}} dx. \end{aligned}$$

By the change of variable $y = \frac{x-S_0}{\sigma\sqrt{T}}$, we obtain

$$\begin{aligned} V_0^{call} &= \frac{1}{\sqrt{2\pi}} \int_{(\hat{K}-S_0)/(\sigma\sqrt{T})}^{\infty} \left(\sigma\sqrt{T}y + S_0 - \hat{K}\right) e^{-\frac{y^2}{2}} dy \\ &= \frac{\sigma\sqrt{T}}{\sqrt{2\pi}} \int_{(\hat{K}-S_0)/(\sigma\sqrt{T})}^{\infty} y e^{-\frac{y^2}{2}} dy + \frac{S_0 - \hat{K}}{\sqrt{2\pi}} \int_{(\hat{K}-S_0)/(\sigma\sqrt{T})}^{\infty} e^{-\frac{y^2}{2}} dy \end{aligned}$$

The second integral above can be calculated in terms of the standard Gaussian cdf $\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy$:

$$\frac{1}{\sigma\sqrt{2\pi T}} \int_{(\hat{K}-S_0)/(\sigma\sqrt{T})}^{\infty} e^{-\frac{y^2}{2}} dy = 1 - \Phi\Big(\frac{\hat{K} - S_0}{\sigma\sqrt{T}}\Big) = \Phi\Big(\frac{S_0 - \hat{K}}{\sigma\sqrt{T}}\Big).$$

Here, we used $\Phi(x) = 1 - \Phi(-x)$.

The first integral can be evaluated by the change of variable $u = \frac{y^2}{2}$.

$$\frac{1}{\sqrt{2\pi}} \int_{(\hat{K}-S_0)/(\sigma\sqrt{T})}^{\infty} ye^{-\frac{y^2}{2}} dy = \frac{1}{\sqrt{2\pi}} \int_{|S_0-\hat{K}|/(\sigma\sqrt{T})}^{\infty} ye^{-\frac{y^2}{2}} dy$$

$$= -\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \Big|_{|S_0-\hat{K}|/(\sigma\sqrt{T})}^{\infty} = \frac{1}{\sqrt{2\pi}} e^{-\frac{(S_0-\hat{K})^2}{2\sigma^2 T}} = \Phi'\Big(\frac{S_0 - \hat{K}}{\sigma\sqrt{T}}\Big),$$

where $\Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ *is the standard Gaussian pdf.*

To summarize, we have

$$V_0^{call} = \sigma\sqrt{T}\left(\Phi'(d) + d\Phi(d)\right), \tag{3.2.9}$$

where $d = \frac{S_0-\hat{K}}{\sigma\sqrt{T}}$. *By the put-call parity, we have*

$$V_0^{put} = \hat{K} - S_0 + V_0^{call} = \sigma\sqrt{T}\left(\Phi'(d) - d\Phi(-d)\right).$$

Example 3.2.2. *A digital call option is an option with payoff*

$$g(S_T) = 1_{\{S_T \geqslant K\}} = \begin{cases} 1 & S_T \geqslant K \\ 0 & S_T < K \end{cases}.$$

A digital put has payoff $1_{\{S_T \leqslant K\}}$.

The Bachelier price of a digital call option with strike K *is given by*

$$V_0^{digit\text{-}c} = e^{-rT}\hat{\mathbb{E}}\Big[1_{\{S_T \geqslant K\}}\Big] = e^{-rT}\hat{\mathbb{P}}\Big(S_T \geqslant K\Big).$$

Notice that $S_T \geqslant K$ *is equivalent to* $\frac{B_t}{\sqrt{t}} = \frac{\hat{S}_t - S_0}{\sigma\sqrt{t}} \geqslant \frac{e^{-rt}K - S_0}{\sigma\sqrt{t}}$, *where* B_t *is the Brownian motion at time* t. *Therefore,* $\frac{\hat{S}_t - S_0}{\sqrt{t}} = \frac{B_t}{\sqrt{t}}$ *is a standard Gaussian random variable. We can write*

$$V_0^{digit\text{-}c} = e^{-rT}\Big(1 - \Phi\big(-d\big)\Big) = e^{-rT}\Phi\big(d\big).$$

Here, $d = \frac{S_0-\hat{K}}{\sigma\sqrt{T}}$ *is the same as in Example 3.2.1.*

Exercise 3.2.1. *Find a closed-form solution for the Bachelier price of a digital put option with strike* K.

Exercise 3.2.2. *Find a closed-form solution for the Bachelier price of a European option with payoff in Figure* (3.2.1)

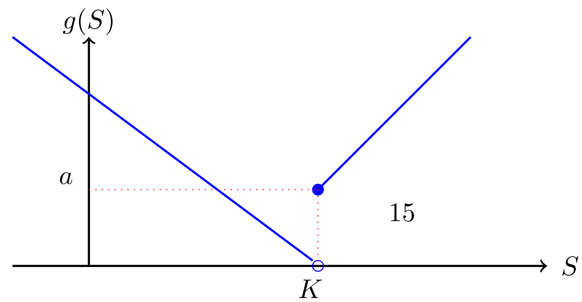


Figure 3.2.1: All slopes are -1 or 1.

The price at an arbitrary point in time: the Markovian property of the option price

Recall from (3.2.6) that the option price V_t at time t is a random variable given by

$$V_t = e^{-r(T-t)}\hat{\mathbb{E}}[g(e^{rT}\hat{S}_T) \mid \mathcal{F}_t] = e^{-r(T-t)}\hat{\mathbb{E}}[g(e^{rT}\hat{S}_T) \mid S_u \,:\, u \leqslant t].$$

Since Brownian motion B is a Markovian process, the only relevant information from the past is the most recent asset price, S_t. Therefore,

$$V_t = e^{-r(T-t)}\hat{\mathbb{E}}[g(e^{rT}\hat{S}_T) \mid \hat{S}_t] =: V(t, \hat{S}_t),$$

where the function $V(t,x)$ is given by

$$\begin{aligned} V(t,x) &= e^{-r(T-t)}\hat{\mathbb{E}}[g(e^{rT}\hat{S}_T) \mid \hat{S}_t = x] \\ &= \frac{e^{-r\tau}}{\sqrt{2\pi(T-t)}}\int_{-\infty}^{\infty} g\big(e^{rT}y\big)e^{-\frac{(x-y)^2}{2\sigma^2\tau}}\,dy. \end{aligned} \tag{3.2.10}$$

The second equality in the above is because given $\hat{S}_t = x$, $\hat{S}_T$ is a Gaussian random variable with pdf

$$f_{\hat{S}_T}(y \mid \hat{S}_t = x) = \frac{1}{\sqrt{2\pi(T-t)}}e^{-\frac{(x-y)^2}{2\sigma^2(T-t)}}.$$

The function $V(t,x)$ is called a pricing function, which provides the price of the contingent claim in terms of time t and the discounted price S_t.

Remark 3.2.1. *It is simpler in the Bachelier model to write the pricing function V as a function of the discounted asset price, which is a Brownian motion, rather than the asset price.*

Time homogeneity of the option price

By the time homogeneity of the Brownian motion, the pricing function is actually a function of time-to-maturity $\tau = T - t$ and the **discounted** underlying price $\hat{S}_t = x$. Because, $\hat{S}_T$ conditional on $\hat{S}_t = x$ has the same distribution as $\hat{S}_\tau$ conditional on $\hat{S}_0 = x$, then

$$V(t,x) = e^{-r\tau}\hat{\mathbb{E}}[g(e^{rT}\hat{S}_\tau) \mid \hat{S}_0 = x] =: U(\tau, x). \tag{3.2.11}$$

Example 3.2.3 (Vanilla options). *Recall that in the Bachelier model, we have $\hat{S}_T = \hat{S}_t + \sigma(B_T - B_t)$. The payoff of the call option can be written as*

$$(S_T - K)_+ = e^{rT}(\hat{S}_T - \hat{K})_+ = e^{rT}(\hat{S}_t + \sigma(B_T - B_t) - \hat{K})_+.$$

Here, $\hat{K} = e^{-rT}K$. Given $\hat{S}_t = x$,

$$U^{call}(\tau,x) = e^{-r\tau}\hat{\mathbb{E}}[e^{rT}(\hat{S}_t + \sigma(B_T - B_t) - \hat{K})_+ \mid \hat{S}_t = x].$$

Since $B_T - B_t$ and $\hat{S}_t = \sigma B_t$ are independent random variables, we have

$$U^{call}(\tau,x) = e^{-r\tau}\hat{\mathbb{E}}[e^{rT}(x + \sigma(B_T - B_t) - \hat{K})_+].$$

Because $B_T - B_t$ is a Gaussian random variable with mean zero and variance $\tau = T - t$, we have

$$\begin{aligned} U^{call}(\tau,x) &= \frac{e^{-r\tau}}{\sigma\sqrt{2\pi\tau}}\int_{-\infty}^{\infty} e^{rT}\left(x + y - \hat{K}\right)_+ e^{-\frac{y^2}{2\sigma^2\tau}}dy \\ &= \frac{e^{-r\tau}}{\sigma\sqrt{2\pi\tau}}\int_{\hat{K}-x}^{\infty} e^{rT}\left(x + y - \hat{K}\right)_+ e^{-\frac{y^2}{2\sigma^2\tau}}dy \\ &= e^{rT}\left((x - \hat{K})\frac{e^{-r\tau}}{\sigma\sqrt{2\pi\tau}}\int_{\hat{K}-x}^{\infty} e^{-\frac{y^2}{2\sigma^2\tau}}dy + \int_{\hat{K}-x}^{\infty} ye^{-\frac{y^2}{2\sigma^2\tau}}dy.\right) \end{aligned}$$

Similar to the calculations in Example 3.2.1, a closed-form solution for the Bachelier price of a call option with strike K and maturity T at time t as a function of $\tau = T - t$ and $\hat{S}_t = x$ is given by

$$U^{call}(\tau,x) = e^{rT}e^{-r\tau}\sigma\sqrt{\tau}\left(\Phi'\big(d(\tau,x)\big) + d(\tau,x)\Phi\big(d(\tau,x)\big)\right),$$

where

$$d(\tau,x) := \frac{x - \hat{K}}{\sigma\sqrt{\tau}}.$$

Exercise 3.2.3. *By mimicking the method in Example 3.2.3, show that*

$$U^{digit\text{-}c}(\tau, x) = e^{rT}e^{-r\tau}\Phi\big(d(\tau, x)\big)$$

is a closed form for the Bachelier price of a digital call option.

Exercise 3.2.4. *Use the put-call parity to find a closed form for the Bachelier price $U^{put}(\tau, x)$ of a put option with strike K and maturity T.*

Exercise 3.2.5. *What is the Bachelier price of at-the-money put option ($K = S_0$) with $T = 10$, $\sigma = .5$, $R_0(10) = .025$ (yield), and $S_0 = 1$? What is the probability that the asset price takes a negative value at T?*

Exercise 3.2.6. *What is the Bachelier price of the payoff in Figure 3.2.6 with $T = 1$, $\sigma = .1$, $R_0(1) = .2$ (yield), and $S_0 = 2$? What is the probability that the option ends up out of the money?*

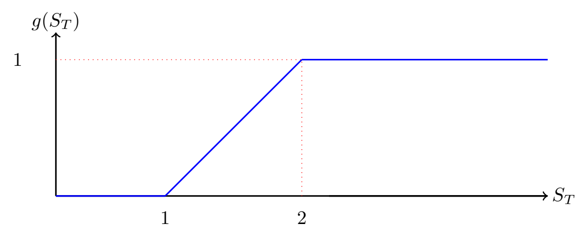


Figure 3.2.2: Payoff of Exercise 3.2.6.

Martingale property of the option price and heat equation

No arbitrage condition assures that the discounted price of an option is also a martingale, $e^{-rt}V(t, \hat{S}_t)$ is a martingale. If we assume that the pricing function $V(t, x)$ is continuously differentiable on t and twice continuously differentiable on x, by the Itô formula we obtain

$$\begin{aligned} d\left(e^{-rt}V(t, \hat{S}_t)\right) &= e^{-rt}\left(\partial_t V + \frac{\sigma^2}{2}\partial_{xx}V - rV\right)(t, \hat{S}_t)dt + e^{-rt}\partial_x V(t, \hat{S}_t)d\hat{S}_t \\ &= e^{-rt}\left(\partial_t V + \frac{\sigma^2}{2}\partial_{xx}V - rV\right)(t, \hat{S}_t)dt + \sigma e^{-rt}\partial_x V(t, \hat{S}_t)dB_t \end{aligned} \tag{3.2.12}$$

Notice that in the above, we used $\partial_t\Big(e^{-rt}V\Big) = e^{-rt}\Big(\partial_t V - rV\Big)$ and $d\hat{S}_t = \sigma dB_t$.

Then, it follows from Section C.3 that $e^{-rt}V(t, \hat{S}_t)$ is a martingale if and only if V satisfied the following partial differential equation (PDE):

$$\partial_t V + \frac{\sigma^2}{2}\partial_{xx}V - rV = 0.$$

This is because $d\left(e^{-rt}V(t,\hat{S}_t)\right)$ reduces to the stochastic integral $\sigma e^{-rt}\partial_x V(t,\hat{S}_t)dB_t$.

A PDE needs appropriate boundary conditions be well-posed. The boundary conditions here include the terminal condition given by the payoff g of the contingent claim

$$V(T,x) = g\big(e^{rT}x\big)$$

and proper growth conditions as $x \to \pm\infty$[3]. The PDE above, always hold regardless of the payoff of the option. The option payoff only appears as the terminal condition. Therefore, the problem of finding the pricing function $V(t,x)$ reduces to solving the boundary value problem (BVP) below.

$$\begin{cases} \partial_t V + \frac{\sigma^2}{2}\partial_{xx}V - rV = 0 \\ V(T,x) = g\big(e^{rT}x\big) \end{cases} \tag{3.2.13}$$

and the growth condition at infinity: $|V(t,x)| \leqslant C|g(x)|$ for some constant C, as $x \to \pm\infty$. The BVP (3.2.13) is a backward heat equation; i.e., we video record the evolution of the heat over time and play it back in reverse. If we do the change of variable $\tau = T - t$ and $U(\tau,x) = V(t,x)$, then U satisfies the forward heat equation

$$\begin{cases} \partial_\tau U = \frac{\sigma^2}{2}\partial_{xx}U - rU \\ U(0,x) = g\big(e^{rT}x\big) \end{cases} \tag{3.2.14}$$

Therefore, the price of a contingent claim at any time can be obtained by solving the BVP (3.2.14).

Example 3.2.4. *By bare-handed calculations, we can show that the function*

$$U(\tau,x) = e^{r(T-\tau)}\sigma\sqrt{\tau}\left(\Phi'\big(d(\tau,x)\big) + d(\tau,x)\Phi\big(d(\tau,x)\big)\right),$$

where

$$d(\tau,x) := \frac{x-\hat{K}}{\sigma\sqrt{\tau}}.$$

satisfies

$$\partial_\tau U = \frac{\sigma^2}{2}\partial_{xx}U - rU.$$

[3]In order for a PDE to have a unique solution, it is necessary to impose proper boundary conditions. The terminal condition is not sufficient to make the boundary value problem well posed. We always need boundary conditions at other boundaries; here they are growth conditions at infinity.

Exercise 3.2.7. *Show that the pricing function on nondiscounted price* $S_t := e^{r(T-t)}\hat{S}_t$*,* $\tilde{U}(\tau, x) := U(\tau, e^{r\tau}x)$*, satisfies*

$$\begin{cases} \partial_\tau \tilde{U} = rx\partial_x \tilde{U} + \frac{e^{2rt}\sigma^2}{2}\partial_{xx}\tilde{U} - r\tilde{U} \\ \tilde{U}(0,x) = g(x) \end{cases}$$

Exercise 3.2.8. *Show that the discounted pricing function* $u(\tau, x) = e^{-r\tau}U(\tau, x)$ *satisfies the standard form of the heat equation below*

$$\begin{cases} \partial_\tau u = \frac{\sigma^2}{2}\partial_{xx}u \\ u(0,x) = g\big(e^{rT}x\big) \end{cases} . \tag{3.2.15}$$

Remark 3.2.2 (On regularity of the pricing function)**.** *To be able to apply Itô formula in* (3.2.12)*, the pricing function* $V(t, x)$ *needs to be continuously differentiable on* t *and twice continuously differentiable on* x*. While the payoff of the option may not be differentiable or value function n continuous, the* $V(t, x)$ *is infinitely differentiable for all* $t < T$ *and all* x*.*

Replication in the Bachelier model: Delta hedging

By (3.1.1), the dynamics of a portfolio in the Bachelier model are given by

$$W_t = W_0 + r\int_0^t (W_s - \Delta_s S_s)ds + \int_0^t \Delta_s dS_s.$$

Similar to Exercise I.2.2.1, the discounted wealth from a portfolio strategy Δ satisfies

$$\hat{W}_t = e^{-rt}W_t = W_0 + \int_0^t \Delta_s d\hat{S}_s,$$

and is a martingale. On the other hand, by applying the Itô formula to the discounted option price $e^{-rt}V(t, \hat{S}_t)$, we obtain (3.2.12)

$$\begin{aligned} e^{-rt}V(t, \hat{S}_t) =& V(0, S_0) + \int_0^t e^{-rs}\left(\partial_t V + \frac{\sigma^2}{2}\partial_{xx}V - rV\right)(s, \hat{S}_s)ds \\ &+ \int_0^t e^{-rs}\partial_x V(s, \hat{S}_s)d\hat{S}_s \\ =& V(0, S_0) + \int_0^t e^{-rs}\partial_x V(s, \hat{S}_s)d\hat{S}_s. \end{aligned}$$

The last inequality above comes from the martingale property of discounted option price and (3.2.13).

A replicating portfolio is a portfolio that generates the terminal wealth W_T equal to the payoff $V(T, \hat{S}_T) = g\big(e^{rT}\hat{S}\big)$. Since both $\hat{W}_t$ and $e^{-rt}V(t, \hat{S}_t)$ are martingales with $\hat{W}_T = e^{-rT}V(T, \hat{S}_T) = e^{-rT}g\big(e^{rT}\hat{S}\big)$, then we must have $\hat{W}_t = e^{-rt}V(t, \hat{S}_t) =: \hat{V}(t, \hat{S}_t)$ for all $t \in [0, T]$, $V(0, S_0) = W_0$, and $\Delta_t = e^{-rt}\partial_x V(t, \hat{S}_t) = \partial_x \hat{V}(t, \hat{S}_t)$.

Δ_t represents the number of units of the underlying in the replicating portfolio. It follows from (3.2.12) that Δ_t is a function of t and $\hat{S}_t$ and is given by

$$\Delta(t, \hat{S}_t) = e^{-rt}\partial_x V(t, \hat{S}_t).$$

Notice that since V is a function of $\tau = T - t$, so is Δ:

$$\Delta(\tau, \hat{S}_t) = e^{-rt}\partial_x V(t, \hat{S}_t) = e^{-r(T-\tau)}\partial_x U(\tau, \hat{S}_t).$$

To summarize, the issuer of the option must trade continuously in time to keep exactly $\Delta_t = e^{-rt}\partial_x V(t, \hat{S}_t)$ number of units of the underlying asset at time t in the replicating portfolio. Δ_t also accounts for the sensitivity of the option price with respect to the change in the price of the underlying.

Example 3.2.5. *The replicating portfolio for a call option in the Bachelier model is obtained by taking the partial derivative ∂_x of the function*

$$V^{call}(t, x) = U^{call}(\tau, x) = e^{r(T-\tau)}\sigma\sqrt{\tau}\left(\Phi'\big(d(\tau, x)\big) + d(\tau, x)\Phi\big(d(\tau, x)\big)\right),$$

with

$$d(\tau, x) := \frac{x - \hat{K}}{\sigma\sqrt{\tau}}.$$

We have

$$\begin{aligned}
\Delta(\tau, x) &= e^{r(T-\tau)}\partial_x U^{call}(\tau, x) \\
&= \sigma\sqrt{\tau}\Big(\partial_x d(\tau, x)\Phi''\big(d(\tau, x)\big) + \partial_x d(\tau, x)\Phi\big(d(\tau, x)\big) + d(\tau, x)\partial_x d(\tau, x)\Phi'\big(d(\tau, x)\big)\Big) \\
&= \Phi''\big(d(\tau, x)\big) + \Phi\big(d(\tau, x)\big) + d(\tau, x)\Phi'\big(d(\tau, x)\big) \\
&= \Phi\big(d(\tau, x)\big).
\end{aligned} \tag{3.2.16}$$

Here, $\Phi'(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ is the pdf of the standard Gaussian, and we used $\Phi''(x) = \frac{-x}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} = -x\Phi'(x)$ and $\partial_x d(\tau, x) = \frac{1}{\sigma\sqrt{\tau}}$.

Example 3.2.6. *To find the Bachelier price of an option with payoff $g(x) = e^x$, we need* *to find the the pricing function U, which solves the following BVP:*

$$\begin{cases} \partial_\tau U = \frac{\sigma^2}{2}\partial_{xx}U - rU \\ U(0,x) = \exp\left(e^{rT}x\right) \end{cases}.$$

We shall verify that the solution to this problem is if the form $U(\tau,x) = e^{\lambda\tau}\exp\left(e^{rT}x\right)$, and find the constant λ by plugging it into the equation:

$$\partial_\tau U - \frac{\sigma^2}{2}\partial_{xx}U + rU = \left(\lambda - e^{2rT}\frac{\sigma^2}{2} + r\right)U(\tau,x) = 0.$$

Thus, for $\lambda = e^{2rT}\frac{\sigma^2}{2} - r$, $U(\tau,x)$ satisfies the equation and the initial condition. Delta hedging is obtained by

$$\Delta(\tau,x) = e^{-rT}e^{r\tau}\partial_x U(\tau,x) = e^{(\lambda+r)\tau}\exp\left(e^{rT}x\right) = \exp\left(e^{2rT}\frac{\sigma^2}{2}\tau\right)\exp\left(e^{rT}x\right).$$

Exercise 3.2.9. *Find a closed-form solution for the Bachelier price of an option with payoff $g(x) = 2\cos(\sqrt{2}x) - 3\sin(-x)$. Hint: Search for the solution of the form $U(\tau,x) = \alpha_1 e^{\lambda_1\tau}\cos(\sqrt{2}x) + \alpha_2 e^{\lambda_2\tau}\sin(-x)$.*

Example 3.2.7. *Let $S_0 = \$10$, $\sigma = .03$, and $r = 0.03$. The Bachelier Delta of the following portfolio of vanilla options given in the table below is the linear combination of the Deltas, $3\Delta^{call}(\tau = .5, K = 10) - 3\Delta^{put}(\tau = 1, K = 10) - \Delta^{call}(\tau = 2, K = 8)$.*

position	*units*	*type*	*strike*	*maturity*
long	*3*	*call*	*\$8*	*.5*
short	*3*	*put*	*\$10*	*1*
short	*1*	*call*	*\$8*	*2*

The maturities are given in years. Then, (3.2.16) for the Delta of the call option in Example 3.2.5 should be used to evaluate $\Delta^{call}(\tau = .5, K = 10)$, $3\Delta^{put}(\tau = 1, K = 10)$ and $\Delta^{call}(\tau = 2, K = 8)$.

Exercise 3.2.10. *Let $S_0 = 10$, $\sigma = .03$, and $r = 0.03$. Consider the portfolio below.*

position	*units*	*type*	*strike*	*maturity*
long	*3*	*call*	*\$10*	*.25 yrs*
long	*4*	*put*	*\$5*	*.5 yrs*

How many units x of the underlying are required to eliminate any sensitivity in the portfolio with respect to changes in the price of the underlying?

Example 3.2.8. *Let $S_0 = 10$, $r = .01$, $\sigma = .02$, and $T = 1$. Consider the payoff in Figure 3.2.3.*

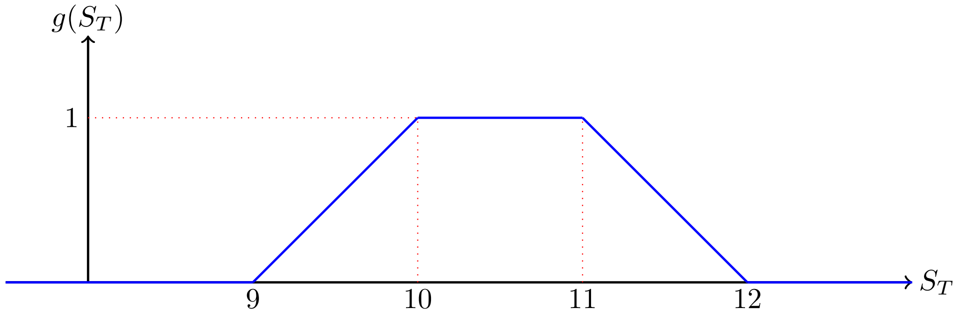


Figure 3.2.3: Payoff in Exercise 3.2.8.

a) Find the Delta of the payoff $g(S_T)$ at $t = 0$.

b) Find ***an*** *appropriate portfolio of call options, put options and a short position in an option with payoff $g(S_T)$ such that the portfolio has a constant Δ at all points in time.*

(a) The payoff $g(S_T)$ can be written as the following combination of call options

$$g(S_T) = (S_T - 9)_+ - (S_T - 10)_+ - (S_T - 11)_+ + (S_T - 12)_+.$$

Therefore,

$$\Delta^g(t = 0, x = 10) = \Delta^{call}(\tau = 1, K = 9) - \Delta^{call}(\tau = 1, K = 10) - \Delta^{call}(\tau = 1, K = 11) + \Delta^{call}(\tau = 1, K = 12).$$

Then, (3.2.16) for the Delta of the call option in Example 3.2.5 should be used to evaluate $\Delta^{call}(\tau = 1, K)$, for $K = 9, 10, 11,$ and 12.

(b) By part (a), if we add a long position in a put option with strike $K = 9$, a long position in a put option with strike $K = 12$, a short positions in a put option with strike $K = 10$, and a short positions in a put option with strike $K = 11$ all with maturity $T = 1$, then the total payoff of the portfolio will be

$$(9-S_T)_+-(10-S_T)_+-(11-S_T)_++(12-S_T)_+-g(S_T) = 9-S_T+S_T-10+S_T-11+12-S_T = 0.$$

Thus, the above portfolio is equivalent to zero position in cash and zero position in the underlying over time, which has a Delta of zero.

Exercise 3.2.11. *Let $S_0 = 9$, $r = .01$, $\sigma = .05$ and $T = 1$. Consider the payoff $g(S_T)$ shown in Figure 3.2.4.*

a) Find the Delta of the payoff $g(S_T)$ at time $t = 0$.

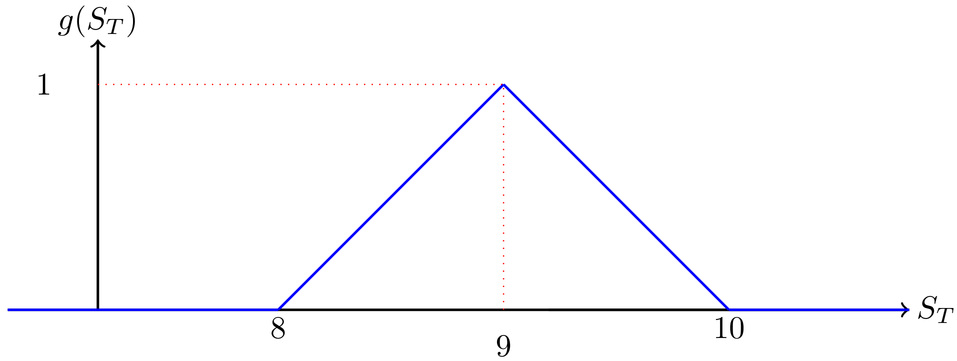


Figure 3.2.4: Payoff in Exercise 3.2.11.

*b) Find **an** appropriate portfolio of call options, put options, and a short position in an option with payoff $g(S_T)$ such that the portfolio has Δ of 0 at all points in time.*

Example 3.2.9. *Consider the payoff $g(S_T)$ shown in Figure 3.2.5. Take $T = 10$, $\sigma = 0.05$,*

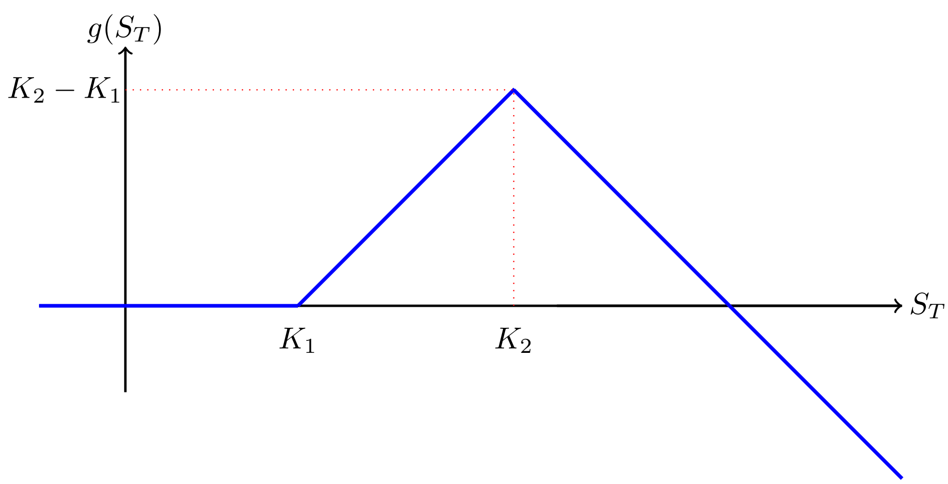


Figure 3.2.5: Payoff in Example 3.2.9.

$R_0(10) = 0.01$ (yield), and $S_0 = 1$. In addition, we assume that $K_1 = 0.8$, but K_2 is unknown. However, assume that the Bachelier Delta of the contingent claim at time 0 is (approximately) equal to -0.385. From this we shall find K_2. Notice that

$$g(S_T) = (S_T - K_1)_+ - 2(S_T - K_2)_+.$$

Therefore,

$$\Delta^g(\tau = 10, x = 1) = \Delta^{call}(\tau = 10, K = K_1, x = 1) - 2\Delta^{call}(\tau = 10, K = K_2, x = 1)$$
$$= \Phi\big(\frac{1 - 0.8e^{-0.1}}{0.05\sqrt{10}})\big) - 2\Phi\big(\frac{1 - K_2 e^{-0.1}}{0.05\sqrt{10}}\big) = 0.9596 - 2\Phi\big(\frac{1 - K_2 e^{-0.1}}{0.05\sqrt{10}}\big).$$

Thus,

$$K_2 = 1.0963.$$

To find the Bachelier price of the contingent claim at time 0*, we simply use*

$$U(\tau = 10, x = 1) = U^{call}(\tau = 10, K = 0.8, x = 1) - 2U^{call}(\tau = 10, K = 1.0963, x = 1).$$

Exercise 3.2.12. *Consider the payoff $g(S_T)$ shown in Figure 3.2.6.*

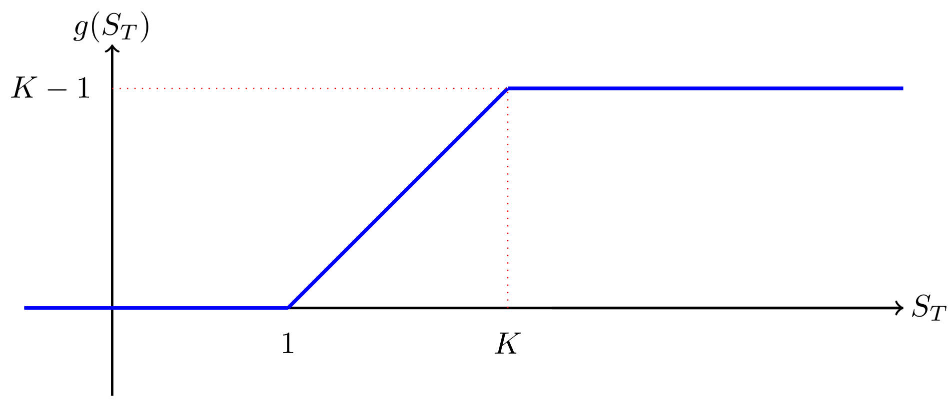


Figure 3.2.6: Payoff in Exercise 3.2.12. All the slopes are 0, 1, or −1.

Part a) *Take $T = 10$, $\sigma = .05$, $R_0(10) = .01$ (yield), and $S_0 = 1$. Find K such that the Bachelier Delta of the contingent claim at time* 0 *is (approximately) equal to* -0.385

Part b) *Find the Bachelier price of the contingent claim at time* 0*.*

3.2.2 Numerical methods for option pricing in the Bachelier model

The BVP for the heat equation in (3.2.13), or, equivalently, (3.2.14) or (3.2.15), generates closed-form solutions in some specific cases, such as a linear combination of call or put options, an exponential payoff in Example 3.2.6, or a sin-cos payoff in Exercise 3.2.9. In general, a closed-form solution can be obtained if the integral in (3.2.8) can precisely be evaluated, or, equivalently, if the BVP for the heat equation has a closed-form solution. The class of payoffs with closed-form solutions is narrow, and therefore one needs to study numerical methods for solving the heat equation in the Bachelier model. Even though the Bachelier model is very far from being practically valuable, the numerical methods presented in this section can be applied indirectly to the Black-Scholes model in the next section. In addition, studying this numerical models for the BVP for the heat equation establishes the methodology for evaluations of more complicated models. Therefore, the reader is recommended to read this section to obtain a background on different methods of the evaluation for the BVPs in finance.

Fourier transform

In this section, we interpret that (3.2.11) as the Fourier transform of a function. The advantage of such interpretation is that a class of methods called *fast Fourier transform* (FFT) algorithms can be deployed to efficiently approximate the Fourier transform and its inverse with highly accurately. The Fourier transform of a function $u(x)$ is defined by

$$\widehat{u}(\theta) := \mathscr{F}[u](\theta) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u(x) e^{-x\theta i} dx,$$

and the inverse Fourier transform of a function $\widehat{u}(\theta)$ is given by

$$\mathscr{F}^{-1}[\widehat{u}](x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \widehat{u}(\theta) e^{x\theta i} dx,$$

If u is integrable, i.e., $\int_{-\infty}^{\infty} |u(x)| dx < \infty$, then the Fourier transform $\mathscr{F}[u](\theta)$ exists and is bounded uniformly on θ. However, the inverse Fourier transform of a bounded function does not necessarily exists. If, in addition, we assume that u is square integrable, i.e., $\int_{-\infty}^{\infty} |u(x)|^2 dx < \infty$, then the Fourier transform $\mathscr{F}[u](\theta)$ is also square integrable and $\mathscr{F}^{-1}[\mathscr{F}[u]](x) = u(x)$ and $\mathscr{F}[\mathscr{F}^{-1}[\widehat{u}]](\theta) = \widehat{u}(\theta)$. For a twice continuously differentiable square integrable function u, the Fourier transform of $\partial_{xx} u(x)$ equals $-\theta^2 \widehat{u}(\theta)$.

In specific payoffs, one can apply Fourier transform and then inverse Fourier transform to find a closed-from solution to the heat equation. However, for a wider class of payoffs, one can only obtain a closed form only for the Fourier transform of the solution to the heat equation. Then, the inverse Fourier transform can be achieved numerically through a class of efficient algorithms, namely *fast Fourier transform*. Since these algorithms are numerically highly efficient, a closed-form for the Fourier transform of the solution is as good as a closed-form for the solution. Assuming that $U(\tau, x)$ is twice continuously differentiable and square integrable on x, the Fourier transform of $\partial_{xx} U(\tau, x)$ equals $-(\theta)^2 \widehat{U}(\tau, \theta)$ and $\widehat{U}(\tau, \theta)$ satisfies the ordinary differential equation (ODE)

$$\begin{cases} \frac{d\widehat{U}}{d\tau} = -\sigma^2 \theta^2 \widehat{U} - r\widehat{U} \\ \widehat{U}(0, \theta) = e^{-rT} \widehat{g}(e^{-rT}\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(e^{rT}x) e^{-x\theta i} dx \end{cases}.$$

Notice that in the above ODE, θ is a parameter and therefore, the solution is given by

$$\widehat{U}(\tau,\theta) = e^{-rT}e^{-r\tau}e^{-\sigma^2\theta^2\tau}\widehat{g}(e^{-rT}\theta).$$

If $U(\tau,\cdot)$ is square integrable, the inverse Fourier transform of $\widehat{U}(\tau,\theta)$ recovers the solution $U(\tau,x)$, i.e.

$$U(\tau,x) = \mathscr{F}^{-1}[\widehat{U}(\tau,\cdot)](x) = e^{-rT}e^{-r\tau}\mathscr{F}^{-1}\left[e^{-\sigma^2\theta^2\tau}\widehat{g}(e^{-rT}\theta)\right](x). \tag{3.2.17}$$

It follows from the convolution rule in Fourier transform that,

$$\mathscr{F}^{-1}\left[e^{-\sigma^2\theta^2\tau}e^{-rT}\widehat{g}(e^{-rT}\theta)\right] = e^{-rT}\left(\mathscr{F}^{-1}\Big[e^{-\sigma^2\theta^2\tau}\Big]\right) * \left(\mathscr{F}^{-1}\Big[\widehat{g}(e^{-rT}\theta)\Big]\right).$$

Here $*$ is the convolution operator defined by

$$f * g(x) := \int_{-\infty}^{\infty} f(y)g(x-y)dy.$$

Since $\mathscr{F}^{-1}\Big[e^{-\sigma^2\theta^2\tau}\Big](x) = \frac{1}{\sigma\sqrt{2\pi\tau}}e^{-\frac{x^2}{2\sigma^2\tau}}$, we derive the formula

$$U(\tau,x) = \frac{e^{-r\tau}}{\sqrt{2\pi\tau}}\int_{-\infty}^{\infty} g\big(e^{rT}y\big)e^{-\frac{(x-y)^2}{2\sigma^2\tau}}dy,$$

which is the same as (3.2.10).

Finite-difference scheme for the heat equation

In this section, we introduce the finite-difference method for the classical heat equation (3.2.15), that is

$$\begin{cases} \partial_\tau u = \frac{\sigma^2}{2}\partial_{xx}u \\ u(0,x) = \tilde{g}(x) := g\Big(e^{rT}x\Big) \end{cases}$$

For educational purposes, despite the availability of analytical formulas, we restrict the discussion to call and put options only. Other types of payoffs should be treated with a similar but yet different analysis. We denote the price of a call (put) option with strike K as a function of time-to-maturity τ and the current discounted price of underlying x as $u(\tau,x) = C(\tau,x)$ $(u(\tau,x) = P(\tau,x))$, which is the solution to the heat equation (3.2.14) with initial condition $u(0,x) = \big(e^{rT}x - K\big)_+$ $(u(0,x) = \big(K - e^{rT}x\big)_+)$. Finally, we set $S_0 = 0$ by considering the change of variable $X_t = \hat{S}_t - S_0$, i.e., the shifted price equal to the difference between the discounted price $\hat{S}_t$ and the initial price S_0.

As the actual domain of the heat equation is infinite, in order to apply the finite-difference scheme, first we need to choose a finite *computational domain* $(\tau, x) \in [0, T] \times [-x_{\max}, x_{\max}]$, for a suitable choice of $x_{\max} > 0$. Now, since the computational domain is bounded, it induces more boundary conditions to the problem at the boundaries $x = x_{\max}$ and $x = -x_{\max}$. Recall that to solve a BVP analytically or numerically, the boundary condition is necessary at all the boundaries. We should find appropriate boundary conditions at both points $x_{\max}$ and $-x_{\max}$, which usually rely on the terminal payoff of the option. These types of boundary conditions, which are induced by the computational domain and do not exist in the original problem are called *artificial boundary conditions*, or in short ABC.

To learn how to set the ABC, let's study the case of a call option with payoff $g(x) = \big(e^{rT}x - (K + e^{rT}S_0)\big)+$, i.e., $u(0, x) = (e^{rT}x - (K + e^{rT}S_0))_+$. The idea is simply as follows. If x is a very small negative number, then $u(0, x) = 0$. If the current discounted price of the underlying is a sufficiently small negative number, the probability that the price at maturity enters the in-the-money region $[(K + e^{rT}S_0), \infty)$ is significantly small. For instance, since the discounted price $S_t = S_0 + \sigma B_t$ is a Gaussian random variable, for $A > 0$ the probability that $S_t \geqslant A$ (or equivalently $S_t \leqslant -A$) is given by

$$\frac{1}{\sigma\sqrt{2\tau\pi}} \int_A^\infty e^{-\frac{y^2}{2\tau\sigma^2}}\, dy \sim \frac{1}{2} e^{-\frac{A^2}{2\tau\sigma^2}}, \quad \text{as} \quad A \to \infty,$$

which is smaller than .006 for $A > 3\sigma\sqrt{T}$. In other words, far out-of-money options should have almost zero price. On the other hand, when $S_t = x$ is sufficiently large, the probability that the discounted price of the underlying will drop below $K + e^{rT}S_0$ at maturity T (out-of-money) is significantly small, and therefore $(e^{rT}\hat{S}_T - K)_+ \approx e^{rT}\hat{S}_T - K$. Far in-the-money options should have almost the same price as the price of payoff $S_T - K$, that is one unit of asset minus K units of cash.

More rigorously, we need the following estimation:

$$\frac{1}{\sigma\sqrt{2\tau\pi}} \int_A^\infty y e^{-\frac{y^2}{2\tau\sigma^2}}\, dy \sim \frac{\sigma\sqrt{\tau}}{\sqrt{2\pi}} e^{-\frac{A^2}{2\tau\sigma^2}}, \quad \text{as} \quad A \to \infty.$$

If we set $A := e^{-rT}K + S_0 + e^{-rT}x_{\max}$, for sufficiently large $x_{\max}$, we have

$$\begin{aligned}
u(\tau, -x_{\max}) &= \frac{e^{-r\tau}}{\sigma\sqrt{2\tau\pi}} \int_{-\infty}^\infty (e^{rT}y - (K + e^{rT}S_0))_+ e^{-\frac{(y + x_{\max})^2}{2\tau\sigma^2}}\, dy \\
&= \frac{e^{-r\tau}}{\sigma\sqrt{2\tau\pi}} \int_{e^{-rT}K + S_0}^\infty (e^{rT}y - (K + e^{rT}S_0)) e^{-\frac{(y + x_{\max})^2}{2\tau\sigma^2}}\, dy \\
&= \frac{e^{r(T-\tau)}}{\sigma\sqrt{2\tau\pi}} \int_A^\infty (y - A) e^{-\frac{y^2}{2\tau\sigma^2}}\, dy \\
&\leqslant \frac{e^{r(T-\tau)}}{\sigma\sqrt{2\tau\pi}} \int_A^\infty y e^{-\frac{y^2}{2\tau\sigma^2}}\, dy \sim \frac{e^{r(T-\tau)}\sigma\sqrt{\tau}}{\sqrt{2\pi}} e^{-\frac{A^2}{2\tau\sigma^2}}, \quad \text{as} \quad A \to \infty.
\end{aligned}$$

In other words, far out-of-money options should have zero price. This suggests that $u(\tau, -x_{\max}) \approx 0$ for sufficiently large $x_{\max}$.

On the other hand, if we set $B := e^{-rT}K + S_0 - e^{-rT}x_{\max}$, we obtain

$$\begin{aligned} u(\tau, x_{\max}) &= \frac{e^{-r\tau}}{\sigma\sqrt{2\tau\pi}} \int_{-\infty}^{\infty} (e^{rT}y - (K + e^{rT}S_0))_+ e^{-\frac{(y-x_{\max})^2}{2\tau\sigma^2}} dy \\ &= \frac{e^{r(T-\tau)}}{\sigma\sqrt{2\tau\pi}} \int_{B}^{\infty} (y - B)e^{-\frac{y^2}{2\tau\sigma^2}} dy. \end{aligned}$$

Notice that since

$$\frac{1}{\sigma\sqrt{2\tau\pi}} \int_{B}^{\infty} ye^{-\frac{y^2}{2\tau\sigma^2}} dy = \frac{1}{\sigma\sqrt{2\tau\pi}} \int_{-\infty}^{B} ye^{-\frac{y^2}{2\tau\sigma^2}} dy \sim \frac{\sigma\sqrt{\tau}}{\sqrt{2\pi}} e^{-\frac{B^2}{2\tau\sigma^2}},$$

and

$$\frac{1}{\sigma\sqrt{2\tau\pi}} \int_{B}^{\infty} e^{-\frac{y^2}{2\tau\sigma^2}} dy \sim 1$$

as $B \to -\infty$, we have

$$u(\tau, x_{\max}) \sim -e^{r(T-\tau)}B = e^{-r\tau}(x_{\max} - (K + e^{rT}S_0)).$$

Following this observation, we choose ABC for (3.2.18) for the call option given by

$$u(\tau, x_{\max}) = e^{-r\tau}(e^{rT}x_{\max} - (K + e^{rT}S_0)) \text{ and } u(\tau, -x_{\max}) = 0.$$

For put option, put-call parity Proposition 1.3.3 implies that the ABC is given by $u(\tau, x_{\max}) = 0$ and $u(\tau, -x_{\max}) = e^{-r\tau}((K + e^{rT}S_0) - e^{rT}x_{\max})$.

To summarize, we must solve the following BVP to numerically price a call option.

$$\begin{cases} \partial_\tau u(\tau, x) &= \frac{\sigma^2}{2}\partial_{xx}u(\tau, x) \quad \text{for} \quad x \in (-x_{\text{amx}}, x_{\max}), t > 0 \\ u(0, x) &= (e^{rT}x - (K + e^{rT}S_0))_+ \quad \text{for} \quad x \in (-x_{\text{amx}}, x_{\max}) \\ u(\tau, x_{\max}) &= e^{-r\tau}(e^{rT}x_{\max} - (K + e^{rT}S_0)) \quad \text{for} \quad t > 0 \\ u(\tau, -x_{\max}) &= 0 \quad \text{for} \quad \tau > 0 \end{cases}. \tag{3.2.18}$$

The next step is to *discretize* the BVP (3.2.18) in time and space. For time discretization, we choose N as the number of time intervals and introduce the time step $h := \frac{T}{N}$ and discrete points τ_i in time for $i = 0, ..., N-1, N$. Then, we choose a computational domain $[-x_{\max}, x_{\max}]$. We discretize the computational domain by $x_j := kj$ with $k := \frac{x_{\max}}{M}$ for $j = -M, ..., M$. The discretization leads to a *grid* including points (t_i, x_j) for $i = 0, ..., N$ and $j = -M, ..., M$, shown in Figure 3.2.7.

Next, we need to introduce derivative approximation. There are two ways to do this: *explicit* and *implicit*. In both methods, the first derivative of a function $u(\tau_i, x_j)$ with

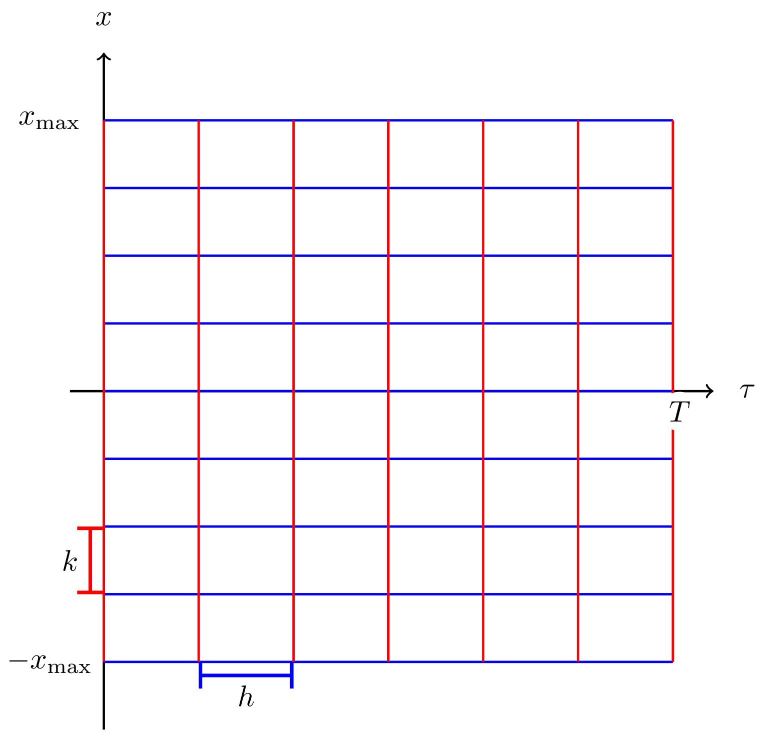


Figure 3.2.7: Finite difference grid for the heat equation. In the explicit scheme the CFL condition should be satisfied, i.e., $\frac{h}{k^2} \leqslant \frac{1}{\sigma}$. Artificial boundary conditions are necessary on both $x_{\max}$ and $-x_{\max}$.

respect to time τ at$any discrete time (\tau_i, x_j)$, is approximated by

$$\partial_t u(\tau_i, x_j) \approx \frac{u(\tau_{i+1}, x_j) - u(\tau_i, x_j)}{h}$$

Then, the second derivative with respect to x can be approximated by

$$\partial_{xx} u(\tau_i, x_j) \approx \frac{u(\tau_i, x_{j+1}) + u(\tau_i, x_{j-1}) - 2u(\tau_i, x_j)}{k^2}.$$

Now, we have all the ingredients to present the explicit scheme for the heat equation. The scheme is obtained from the heat equation (3.3.25) by simply plugging the above approximations for derivatives, i.e.

$$\frac{u(\tau_{i+1}, x_j) - u(\tau_i, x_j)}{h} = \frac{\sigma^2}{2} \cdot \frac{u(\tau_i, x_{j+1}) + u(\tau_i, x_{j-1}) - 2u(\tau_i, x_j)}{k^2}.$$

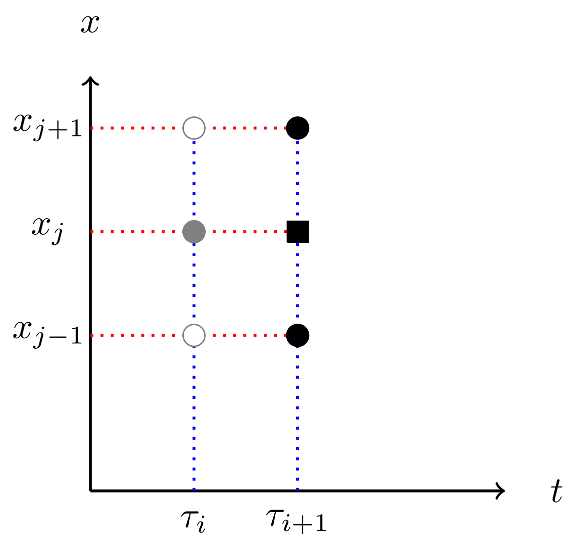


Figure 3.2.8: Possible active points in the finite-difference scheme to evaluate $u(\tau_{i+1}, x_j)$, marked with a square. The function u is unknown at dark nodes and known at light nodes. All six nodes are active for implicit scheme with $\theta \neq 1$. For explicit scheme, $\theta = 1$, only filled-in nodes are active.

We can simplify the scheme by writing

$$u(\tau_{i+1}, x_j) = \left(1 - \frac{h\sigma^2}{k^2}\right) u(\tau_i, x_j) + \frac{h\sigma^2}{2k^2} \left(u(\tau_i, x_{j+1}) + u(\tau_i, x_{j-1})\right). \tag{3.2.19}$$

In order to use explicit finite-difference scheme in (3.2.19), we need to have the CFL[4] condition

$$\frac{h}{k^2} \leqslant \frac{1}{\sigma^2}.$$

Otherwise, the scheme does not converge. The right-hand side of the CFL condition is always $\frac{1}{2}$ times the inverse of the coefficient of the second derivative in the equation. For implicit schemes, this condition can be relaxed.

Notice that in problem (3.2.18), at $\tau_0 = 0$, the initial condition is known. Therefore, we set

$$u(0, x_j) := e^{-rT} g(e^{rT} x_j) \quad \text{for} \quad j = -M, ..., M.$$

Then, if $u(\tau_i, x_j)$ is known for all $j = -M, ..., M$, the explicit scheme (3.2.19) suggests that $u(\tau_{j+1}, x_j)$ can be found for all $j = -M+1, ..., M-1$. For $j = -M$ and M, one can use ABC to assign values to $u(\tau_i, x_{-M})$ and $u(\tau_i, x_M)$.

[4] Courant-Friedrichs-Lewy

The implicit scheme is a little more difficult than explicit scheme to implement. But it has its own advantages; e.g., the CFL condition is not necessary. To present the implicit method we need to modify the approximation of the second derivative as follows.

$$\begin{aligned}\partial_{xx}u(\tau_i,x_j) \approx (1-\theta)&\frac{u(\tau_i,x_{j+1})+u(\tau_i,x_{j-1})-2u(\tau_i,x_j)}{k^2}\\ &+\theta\frac{u(\tau_{i+1},x_{j+1})+u(\tau_{i+1},x_{j-1})-2u(\tau_{i+1},x_j)}{k^2}.\end{aligned}$$

In the above, $\theta \in [0,1]$ is a parameter. If $\theta = 0$, then the scheme is the same as the explicit scheme. If $\theta = 1$, we call it a pure implicit scheme. Then for $\theta \neq 0$, we can present the implicit scheme as follows.

$$\begin{aligned}&\left(1+\theta\frac{h\sigma^2}{k^2}\right)u(\tau_{i+1},x_j)-\theta\frac{h\sigma^2}{2k^2}\left(u(\tau_{i+1},x_{j+1})+u(\tau_{i+1},x_{j-1})\right)=\\ &\qquad\left(1-(1-\theta)\frac{h\sigma^2}{k^2}\right)u(\tau_i,x_j)+(1-\theta)\frac{h\sigma^2}{2k^2}\left(u(\tau_i,x_{j+1})+u(\tau_i,x_{j-1})\right).\end{aligned} \tag{3.2.20}$$

If $u(\tau_i,x_j)$ is known for all $j=-M,...,M$, then the right-hand side above is known. Lets denote the right-hand side by

$$R(\tau_i,x_j) := \left(1-(1-\theta)\frac{h\sigma^2}{k^2}\right)u(\tau_i,x_j)+(1-\theta)\frac{h\sigma^2}{2k^2}\left(u(\tau_i,x_{j+1})+u(\tau_i,x_{j-1})\right).$$

For $j=M-1$, $u(\tau_{i+1},x_{j+1})$ on the left-hand side is known. Thus, we move this term to the other side

$$\begin{aligned}&\left(1+\theta\frac{h\sigma^2}{k^2}\right)u(\tau_{i+1},x_{M-1})-\theta\frac{h\sigma^2}{2k^2}u(\tau_{i+1},x_{M-2})=\\ &\qquad R(\tau_i,x_{M-1})+\theta\frac{h\sigma^2}{2k^2}u(\tau_{i+1},x_M).\end{aligned}$$

Similarly for $j=-M+1$ we have

$$\begin{aligned}&\left(1+\theta\frac{h\sigma^2}{k^2}\right)u(\tau_{i+1},x_{-M+1})-\theta\frac{h\sigma^2}{2k^2}u(\tau_{i+1},x_{-M+2})=\\ &\qquad R(\tau_i,x_{-M+1})+\theta\frac{h\sigma^2}{2k^2}u(\tau_{i+1},x_{-M}).\end{aligned}$$

To find $u(\tau_{i+1},x_j)$, one needs to solve the following tridiagonal equation for $u(\tau_{i+1},x_j)$,

$j = -M+1, ..., M-1$.

$$A\mathbf{U}_{i+1} = \mathbf{R}_i - \mathbf{Y}_i \tag{3.2.21}$$

where A is a $2M-1$-by-$2M-1$ matrix given by

$$A := \begin{bmatrix} 1+\theta\frac{h\sigma^2}{k^2} & -\theta\frac{h\sigma^2}{2k^2} & 0 & 0 & \cdots & 0 \\ -\theta\frac{h\sigma^2}{2k^2} & 1+\theta\frac{h\sigma^2}{k^2} & -\theta\frac{h\sigma^2}{2k^2} & 0 & \cdots & 0 \\ 0 & -\theta\frac{h\sigma^2}{2k^2} & 1+\theta\frac{h\sigma^2}{k^2} & -\theta\frac{h\sigma^2}{2k^2} & \cdots & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & \cdots & -\theta\frac{h\sigma^2}{2k^2} & 1+\theta\frac{h\sigma^2}{k^2} & -\theta\frac{h\sigma^2}{2k^2} & 0 \\ 0 & \cdots & 0 & -\theta\frac{h\sigma^2}{2k^2} & 1+\theta\frac{h\sigma^2}{k^2} & -\theta\frac{h\sigma^2}{2k^2} \\ 0 & \cdots & 0 & 0 & -\theta\frac{h\sigma^2}{2k^2} & 1+\theta\frac{h\sigma^2}{k^2} \end{bmatrix},$$

$\mathbf{Y}_i$ is the column $2M-1$-vector

$$\mathbf{Y}_i := \theta\frac{h\sigma^2}{2k^2}(u(\tau_{i+1}, x_{-M}), 0, \cdots, 0, u(\tau_{i+1}, x_M))^{\mathsf{T}},$$

$\mathbf{R}_i$ is the column $2M-1$-vector

$$\mathbf{R}_i := (R(\tau_i, x_{-M+1}), \cdots, R(\tau_i, x_{M-1}))^{\mathsf{T}},$$

and the unknown is the column $2M-1$-vector

$$\mathbf{U}_{i+1} := (u(\tau_{i+1}, x_{-M+1}), \cdots, u(\tau_{i+1}, x_{M-1}))^{\mathsf{T}}.$$

Notice that the endpoints $u(\tau_{i+1}, x_{-M})$ and $u(\tau_{i+1}, x_M)$ are given by the ABC:

$$u(\tau_{i+1}, x_M) = e^{-r\tau}(e^{rT}x_M - (K + e^{rT}S_0)) \quad \text{and} \quad u(t, x_{-M}) = 0.$$

The CFL condition for the implicit scheme with $\theta \in [0, 1)$ is given by

$$\frac{h}{k^2} \leqslant \frac{1}{(1-\theta)\sigma^2}.$$

For the pure implicit scheme ($\theta = 1$), no condition is necessary for convergence.

Monte Carlo approximation for the Bachelier model

Recall that in the Bachelier model Gaussian distribution with mean S_0 and variance $\sigma^2 T$, $\hat{S}_\tau$ is $\mathcal{N}(x, \sigma^2\tau)$, and the pricing formula is given by

$$U(\tau, x) = e^{-r\tau}\hat{\mathbb{E}}[g(e^{rT}\hat{S}_\tau) \mid \hat{S}_0 = x].$$

In the Monte Carlo method, we generate samples based on the underlying probability distribution to approximate the above expectation. Let the i.i.d. samples $\{x_j : j = 1, ..., N\}$ be taken from $\mathcal{N}(0,1)$. Then, the expectation $\hat{\mathbb{E}}[g(e^{rT}\hat{S}_\tau) \mid \hat{S}_0 = x]$ can be approximated by

$$\frac{1}{N}\sum_{j=1}^{N} g(e^{rT}(x + \sigma\sqrt{\tau}x_i)).$$

Hence, the price of the option $U(\tau, t, x)$ can be approximated by

$$U^{\text{approx}}(\tau, x) = \frac{e^{-r(T-t)}}{N}\sum_{j=1}^{N} g(e^{rT}(x + \sigma\sqrt{\tau}x_i)). \tag{3.2.22}$$

The larger the number of samples N is, the more accurate the approximation $U^{\text{approx}}(\tau, x)$ is obtained. The plain Monte Carlo method is not as efficient as the finite-difference, at least when there is only a single risky asset. However, some methods such as variance reduction or quasi Monte Carlo can be used to increase its performance.

Quadrature approximation for the Bachelier model

Quadrature methods are based on a deterministic (nonrandom) approximation of the integral. In the Bachelier model the price of the option is given by

$$U(\tau, x) = \frac{e^{-r\tau}}{\sigma\sqrt{2\pi\tau}}\int_{-\infty}^{\infty} g\left(e^{rT}y\right) e^{-\frac{(y-x)^2}{2\sigma^2\tau}}\,dy.$$

As an example of the quadrature method, one can first approximate the improper integral above with the proper integral

$$\frac{e^{-r\tau}}{\sigma\sqrt{2\pi\tau}}\int_{x-A}^{x+A} g\left(e^{rT}y\right) e^{-\frac{(y-x)^2}{2\sigma^2\tau}}\,dy$$

and then use Riemann sums to approximate the price of option by

$$U^{\text{approx}}(\tau, x) = \frac{e^{-r\tau}}{\sigma\sqrt{2\pi\tau}}\sum_{j=0}^{N-1} g\left(e^{rT}y_j^*\right) e^{-\frac{(y_j^*-x)^2}{2\sigma^2\tau}}(y_{j+1} - y_j),$$

where $y_0 = x - A < y_1 < \cdots < y_N = x + A$. If the discrete points y_j for $j = 0, ..., N$ are carefully chosen, the quadrature method outperforms the plain Monte Carlo method.

Exercise 3.2.13 (Project)**.** *Consider the initial price S_0, σ, r and payoff assigned to your group.*

Group 1	T	S_0	r	σ
1	*1*	*10*	*.2*	*1*
2	*10*	*100*	*.1*	*5*
3	*1*	*2*	*.2*	*.5*
4	*2*	*2*	*.5*	*.5*
5	*2*	*2*	*.5*	*.5*
6	*1*	*.5*	*.1*	*.001*

Step 1. *Choose a computational domain around the initial price,* $[S_0 - x_{\max}, S_0 + x_{\max}]$.

Step 2. *Set appropriate artificial boundary conditions (ABC) at the boundary points* $S_0 - x_{\max}$ *and* $S_0 + x_{\max}$.

Step 3. *Write a program that implements the implicit finite-difference code. The time and space discretization parameters* (h, k) *must be set to satisfy*

$$\frac{h}{k^2} \leqslant \frac{1}{(1-\theta)\sigma^2}.$$

Step 4. *To make sure that your code is correct, run the program for a call (or a put) option and compare it to the closed-form solution in* (3.2.9).

Step 5. *Run the program for* $\theta = 0$ *(explicit),* $\theta = \frac{1}{2}$ *(semi-implicit) and* $\theta = 1$ *(implicit) and record the results.*

Step 6. *Simulate a discrete sample path of the price of the underlying asset with the same time discretization parameter* h. *Use the following algorithm:*

Simulating a sample paths of underlying in the Bachelier model

1: Discretize time by $t_0 = 0$, $t_i = ih$, and $T = hN$.
2: **for** each $j = 1, ..., N$ **do**
3: Generate a random number w_j from standard Gaussian distribution $N(0,1)$ to represent $(B_{t_j} - B_j)/\sqrt{h}$
4: $\hat{S}_j = \hat{S}_{j-1} + \sigma\sqrt{h}w_j$
5: $s_j = e^{rt_j}\hat{S}_j$
6: **end for**

Output: vector $(s_0 = S_0, s_{t_1}, ..., s_{t_{N-1}}, s_T)$ is a discretely generated sample path of the Bachelier model.

Step 7. *Recall that the hedging is given by the derivative* $\partial_x \bar{V}(t, \hat{S}_t) = e^{rt}\partial_x V(t, e^{rt}S_t)$. *Evaluate the hedging strategy discretely at each node of the discretely generated sample* *path* $(s_0 = S_0, s_{t_1}, ..., s_{t_{N-1}})$ *in **Step 7**, i.e.,*

$$(\partial_x V(0, e^{rt} s_0), \partial_x V(t_1, e^{rt_1} s_{t_1}), ..., e^{rt_{N-1}} \partial_x V(t_{N-1}, e^{rt_{N-1}} s_{t_{N-1}})).$$

Note that some interpolation may be needed in this step.

Submit the following outputs:

Output 1. *The program*

Output 2. *A comparison with the closed-form solution for a call at* $(0, S_0)$.

Output 3. *The price and hedging strategy at the points of the discrete grid.*

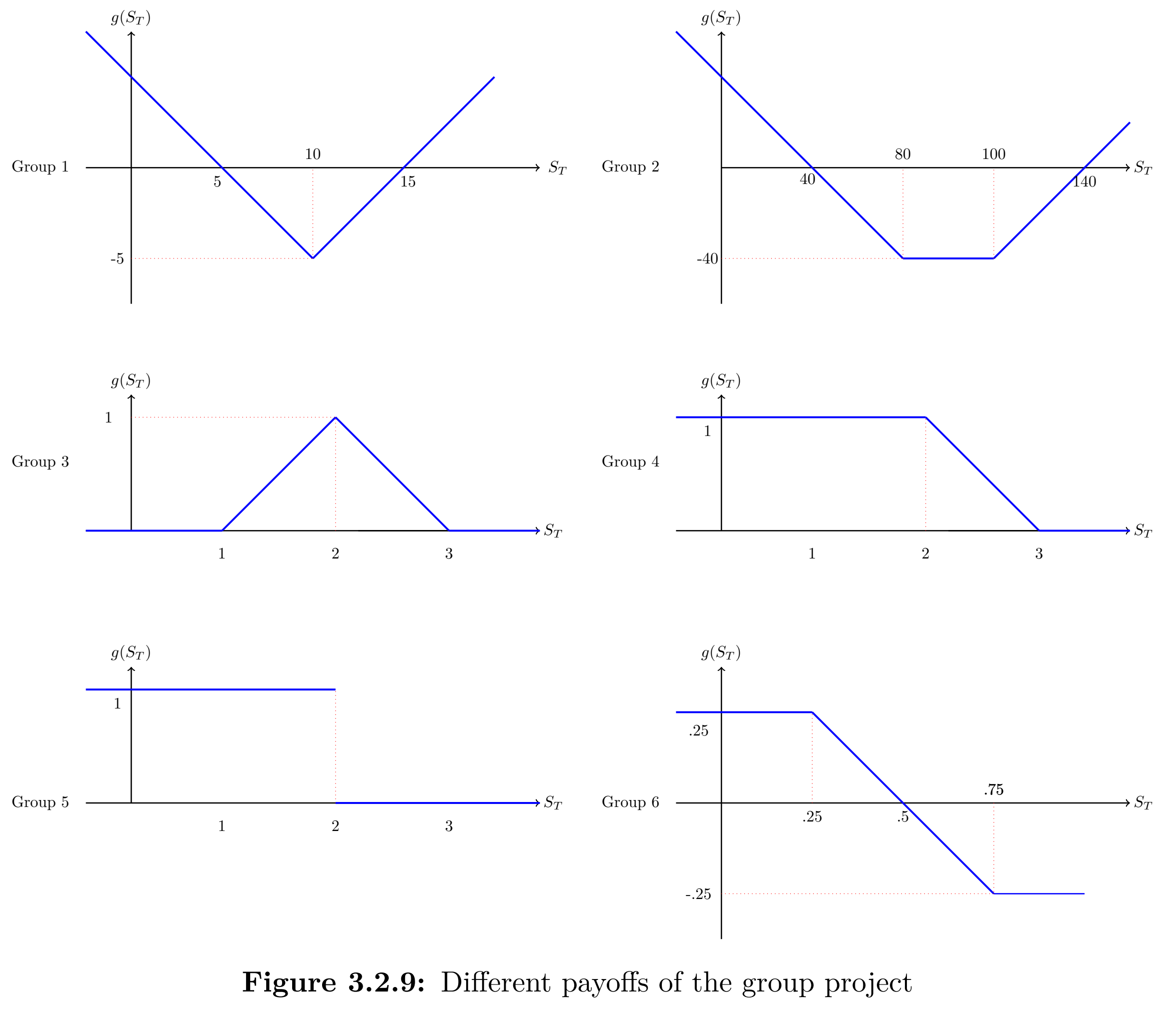


Figure 3.2.9: Different payoffs of the group project

3.2.3 Discussion: drawbacks of the Bachelier model

One of the drawbacks of the Bachelier model is the possibility of negative realization of the asset price. However, this is not the main concern, as in many other applications; Gaussian random variables are used to model positive quantities such as human weight. The negative asset price can be problematic if the ratio $\frac{S_0}{\sigma\sqrt{T}}$ is not sufficiently large. For instance, if the ratio $\frac{S_0}{\sigma\sqrt{T}} = 1$, then the chance of a negative price at maturity T is significant, i.e., higher than .3.

In addition, the return of an asset in the Bachelier model is not integrable. The return of an asset in the Bachelier model is given by

$$\mathbf{R}_t^{\text{arth}} := \frac{e^{r(t+\delta)}\hat{S}_{t+\delta} - e^{rt}\hat{S}_t}{e^{rt}\hat{S}_t} = \frac{e^{r\delta}B_{t+\delta} - B_t}{B_t} = \frac{e^{r\delta}(B_{t+\delta} - B_t)}{B_t} + e^{r\delta} - 1.$$

Given B_t, the return in the Bachelier model is a Gaussian random variable. Some empirical studies support the Gaussian distribution for the return. However, the random variable $\frac{1}{B_t}$ is not integrable:

$$\hat{\mathbb{E}}\left[\frac{1}{|B_t|}\right] = \frac{1}{\sqrt{2t\pi}}\int_{\mathbb{R}} \frac{1}{|x|} e^{-\frac{x^2}{2t}} dx = \infty.$$

3.3 Continuous-time market of Black-Scholes

The Black-Scholes model can be obtained by asymptotic methods from the binomial model. To do this, we first present some asymptotic properties of the binomial model.

3.3.1 The Black-Scholes model: limit of binomial under risk-neutral probability

Let $T > 0$ be a real number and N be a positive integer. We divide T units of time into N time intervals, each of size $\delta := \frac{T}{N}$[5]. Then, consider a binomial model with N periods given by the times $t_0 = 0 < t_1 = \delta, ..., t_k = k\delta, ...t_N = T$. Recall from binomial model that

$$S_{(k+1)\delta} = S_{k\delta}H_{k+1}, \quad \text{for} \quad k = 0, ..., N-1$$

where, in accordance with Assumption 2.4.1, $\{H_k\}_{k=1}^N$ is a sequence of i.i.d. random variables with the following distribution

$$H_k = \begin{cases} u & \text{with probability } \hat{\pi} \\ \ell & \text{with probability } 1-\hat{\pi} \end{cases}$$

[5] Each time unit is divided into $1/\delta$ small time intervals.

Asymptotics of parameters u, ℓ, and p

The goal of this section is to show the following approximation:

$$\begin{aligned} u &= 1 + \delta r + \sqrt{\delta}\sigma\alpha = e^{\delta(r-\frac{\sigma^2}{2})+\sqrt{\delta}\sigma\alpha} + o(\delta), \\ \ell &= 1 + \delta r - \sqrt{\delta}\sigma\beta = e^{\delta(r-\frac{\sigma^2}{2})-\sqrt{\delta}\sigma\beta} + o(\delta). \end{aligned} \tag{3.3.1}$$

To obtain this approximation, one should notice that from (2.4.7), we have $x_0 = 1 + \frac{\lambda\sqrt{\delta}}{2} + o(\sqrt{\delta})$. Thus, it follows from (2.4.6) that

$$\frac{1-p}{p} = x_0^2 = 1 + \lambda\sqrt{\delta} + o(\sqrt{\delta}) \quad \text{and} \quad p = \frac{1}{2} - \frac{\lambda\sqrt{\delta}}{4} + o(\sqrt{\delta}).$$

Then,

$$\alpha^2 = \frac{1-p}{p} + 2\lambda\sqrt{\delta}\sqrt{\frac{1-p}{p}} + o(\sqrt{\delta}) = x_0^2 = 1 + 3\lambda\sqrt{\delta} + o(\sqrt{\delta}). \tag{3.3.2}$$

On the other hand, one can easily see that[6]

$$\begin{aligned} e^{\delta(r-\frac{\sigma^2}{2})+\sqrt{\delta}\sigma\alpha} &= 1 + \delta(r - \frac{\sigma^2}{2}) + \sqrt{\delta}\sigma\alpha + \frac{(\delta(r-\frac{\sigma^2}{2}) + \sqrt{\delta}\sigma\alpha)^2}{2} + o(\delta) \\ &= 1 + \delta(r - \frac{\sigma^2}{2}) + \sqrt{\delta}\sigma\alpha + \frac{\delta\sigma^2\alpha^2}{2} + o(\delta) \\ &= 1 + \delta r + \sqrt{\delta}\sigma\alpha + \frac{\delta(\sigma^2-1)\alpha^2}{2} + o(\delta) \\ &= 1 + \delta r + \sqrt{\delta}\sigma\alpha + o(\delta). \end{aligned}$$

In the above we used (3.3.2), i.e., $\sigma^2 - 1 = O(\sqrt{\delta})$.
Similarly, we have $\beta^2 = 1 - 3\lambda\sqrt{\delta} + o(\sqrt{\delta})$ and

$$e^{\delta(r-\frac{\sigma^2}{2})-\sqrt{\delta}\sigma\beta} = 1 + \delta r - \sqrt{\delta}\sigma\beta + o(\delta).$$

Arithmetic return versus log return

This asymptotics yields to the relation between the arithmetic return and the log return in the binomial model. While the arithmetic return is given by

$$\mathbf{R}_t^{\text{arth}} = \begin{cases} \delta r + \sqrt{\delta}\sigma\alpha \\ \delta r - \sqrt{\delta}\sigma\beta \end{cases}$$

[6] $e^x = 1 + x + \frac{x^2}{2} + o(x^2)$

the log return is given by

$$\mathbf{R}_t^{\log} = \begin{cases} \delta(r - \frac{\sigma^2}{2}) + \sqrt{\delta}\sigma\alpha + o(\delta) \\ \delta(r - \frac{\sigma^2}{2}) - \sqrt{\delta}\sigma\beta + o(\delta) \end{cases} .$$

The probabilities of the values in both returns are given by $(\hat{\pi}, 1-\hat{\pi})$ for risk-neutral probability and $(p, 1-p)$ for physical probability. In particular, if $\hat{\mathbb{E}}$ and $\mathbb{E}$ are respectively expectations with respect to risk-neutral probability and physical probability, then we have

$$\begin{aligned} \hat{\mathbb{E}}[\mathbf{R}_t^{\text{arth}}] &= r\delta, \quad \text{and} \quad \hat{\mathbb{E}}[(\mathbf{R}_t^{\text{arth}})^2] = \sigma^2\delta + o(\delta) \\ \mathbb{E}[\mathbf{R}_t^{\text{arth}}] &= \mu\delta, \quad \text{and} \quad \mathbb{E}[(\mathbf{R}_t^{\text{arth}})^2] = \sigma^2\delta + o(\delta) \\ \hat{\mathbb{E}}[\mathbf{R}_t^{\log}] &= (r - \frac{1}{2}\sigma^2)\delta, \quad \text{and} \quad \hat{\mathbb{E}}[(\mathbf{R}_t^{\log})^2] = \sigma^2\delta + o(\delta) \\ \mathbb{E}[\mathbf{R}_t^{\log}] &= (\mu - \frac{1}{2}\sigma^2)\delta, \quad \text{and} \quad \mathbb{E}[(\mathbf{R}_t^{\log})^2] = \sigma^2\delta + o(\delta). \end{aligned} \tag{3.3.3}$$

Weak convergence of the binomial model: the geometric Brownian motion

From the asymptotics in (3.3.3), the log return of the calibrated binomial model is given by

$$\mathbf{R}_t^{\log} = \ln(H_k) = \begin{cases} \delta(r - \frac{\sigma^2}{2}) + \sqrt{\delta}\sigma\alpha & \text{with probability } \hat{\pi} \\ \delta(r - \frac{\sigma^2}{2}) - \sqrt{\delta}\sigma\beta & \text{with probability } 1-\hat{\pi} \end{cases}$$

Indeed,

$$\ln(S_{(k+1)\delta}) = \ln(S_{k\delta}) + \ln(H_{k+1}),$$

or

$$\ln(S_t) = \ln(S_0) + \sum_{k=1}^{N} \ln(H_k).$$

Let $\{Z_k\}_{k=1}^N$ be a sequence of i.i.d. random variables with the following distribution

$$Z_k = \begin{cases} \alpha & \text{with probability } \hat{\pi} \\ -\beta & \text{with probability } 1-\hat{\pi} \end{cases}$$

Then, we have

$$\ln(S_T) = \ln(S_0) + (r - \frac{\sigma^2}{2})T + \sigma\sqrt{T} \cdot \frac{1}{\sqrt{N}} \sum_{k=1}^{N} Z_k. \tag{3.3.4}$$

Next, we want to show that the normalized summation $\frac{1}{\sqrt{N}} \sum_{k=1}^{N} Z_k$ converges weakly to a random variable (distribution) as the number of time intervals N approaches infinity[7]. To show this, from Theorem B.9 from the appendix, we only need to find

$$\lim_{N\to\infty} \chi_{\frac{1}{\sqrt{N}}\sum_{k=1}^{N} Z_k}(\theta).$$

Notice that here the characteristic function is under the risk-neutral probability, $\chi_X(\theta) = \hat{\mathbb{E}}[e^{i\theta X}]$.

Since $\{Z_t\}_{k=1}^N$ is a sequence of i.i.d. random variables, we have

$$\chi_{\frac{1}{\sqrt{N}}\sum_{k=1}^{N} Z_k}(\theta) = \prod_{k=1}^{N} \chi_{Z_1}\left(\frac{\theta}{\sqrt{N}}\right).$$

On the other hand,

$$\chi_{Z_1}(\theta) = \hat{\mathbb{E}}[e^{i\theta Z_1}] = 1 + i\theta\hat{\mathbb{E}}[Z_1] - \frac{\theta^2\hat{\mathbb{E}}[Z_1^2]}{2} + o(\theta^2).$$

Notice that by (2.4.4), we can write

$$\hat{\pi} = \frac{R-L}{U-L} = \frac{\beta}{\alpha+\beta}. \tag{3.3.5}$$

Therefore, straightforward calculations show that $\hat{\mathbb{E}}[Z_1] = 0$, and $\hat{\mathbb{E}}[Z_1^2] = 1$. By using the Taylor expansion of the characteristic function, we obtain

$$\chi_{Z_1}\left(\frac{\theta}{\sqrt{N}}\right) = 1 - \frac{\theta^2}{2N} + o(\frac{1}{N}),$$

and

$$\chi_{\frac{1}{\sqrt{N}}\sum_{k=1}^{N} Z_k}(\theta) = \left(1 - \frac{\theta^2}{2N} + o(\frac{1}{N})\right)^N.$$

By sending $n \to \infty$, we obtain

$$\lim_{N\to\infty} \chi_{\frac{1}{\sqrt{N}}\sum_{k=1}^{N} Z_k}(\theta) = \lim_{N\to\infty} \left(1 - \frac{\theta^2}{2N} + o(\frac{1}{N})\right)^N = e^{-\frac{\theta^2}{2}}.$$

Since $e^{-\frac{\theta^2}{2}}$ is the characteristic function of a standard Gaussian random variable, we conclude that $\frac{1}{\sqrt{N}}\sum_{k=1}^{N} Z_k$ converges weakly to $\mathcal{N}(0,1)$.

Exercise 3.3.1. *In the above calculations, explain why we cannot apply the central limit theorem (Theorem B.7) directly.*

[7] Or equivalently $\delta \to 0$.

(3.3.4) suggests that we define a continuous-time model for price S_t of the asset at time T as the weak limit of the binomial model by

$$\ln(S_T) = \ln(S_0) + (r - \frac{\sigma^2}{2})T + \sigma\mathcal{N}(0,T)$$

or equivalently

$$S_T = S_0 \exp\left((r - \frac{\sigma^2}{2})T + \sigma\mathcal{N}(0,T)\right).$$

Calibrating binomial model: revised

In the above, we see that the choice of parameters u, ℓ, and R leads to a perfect choice of $\hat{\mathbb{E}}[Z_1] = 0$ and $\hat{\mathbb{E}}[Z_1^2] = 1$. However, in (3.3.4), the only criteria for the convergence of binomial model to Black-Scholes model is that the random variables Z_k, $k = 1, ..., N$ must satisfy $\hat{\mathbb{E}}[Z_1] = o(\delta)$ and $\hat{\mathbb{E}}[Z_1^2] = 1 + o(1)$;

$$\begin{aligned} &\text{If } \hat{\mathbb{E}}[Z_1] = o(\delta), \quad \text{and} \quad \hat{\mathbb{E}}[Z_1^2] = 1 + o(1), \\ &\text{then } \frac{1}{\sqrt{N}}\sum_{k=1}^{N} Z_k \quad \text{converges to} \quad \mathcal{N}(0,1) \quad \text{weakly.} \end{aligned} \tag{3.3.6}$$

To avoid the calculation of α and β, one can choose different parameters for the binomial model. Notice that the binomial model has three parameters u, ℓ and R while the Black-Scholes parameters are only two. This degree of freedom provides us with some modifications of the binomial model which still converges to the Black-Scholes formula. This also simplifies the calibration process in Section 2.4 significantly. Here are some choices:

1) Symmetric probabilities:

$$u = e^{\delta(r - \frac{\sigma^2}{2}) + \sqrt{\delta}\sigma}, \quad \ell = e^{\delta(r - \frac{\sigma^2}{2}) - \sqrt{\delta}\sigma}, \quad \text{and} \quad R = r\delta,$$

Then

$$\hat{\pi}_u = \hat{\pi}_\ell = \frac{1}{2}.$$

Notice that (3.3.4) should be modified by setting the$\{Z_k\}_{k=1}^N$ distribution of the i.i.d. sequence

$$Z_1 = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases}$$

and $\hat{\mathbb{E}}[Z_1] = 0$ and $\hat{\mathbb{E}}[Z_1^2] = 1$.

2) Subjective return:

$$u = e^{\delta\nu + \sqrt{\delta}\sigma}, \quad \ell = e^{\delta\nu - \sqrt{\delta}\sigma}, \quad \text{and} \quad R = r\delta,$$

Then

$$\hat{\pi}_u = \frac{1}{2}\left(1 + \sqrt{\delta}\frac{r - \nu - \frac{1}{2}\sigma^2}{\sigma}\right) \quad \text{and} \quad \hat{\pi}_\ell = \frac{1}{2}\left(1 - \sqrt{\delta}\frac{r - \nu - \frac{1}{2}\sigma^2}{\sigma}\right).$$

In this case, (3.3.4) should be modified by setting the$\{Z_k\}_{k=1}^N$ distribution of the i.i.d. sequence

$$Z_1 = \begin{cases} \sqrt{\delta}\frac{\nu - r + \frac{1}{2}\sigma^2}{\sigma} + 1 & \text{with probability } \hat{\pi}_u \\ \sqrt{\delta}\frac{\nu - r + \frac{1}{2}\sigma^2}{\sigma} - 1 & \text{with probability } \hat{\pi}_\ell \end{cases}$$

and $\hat{\mathbb{E}}[Z_1] = 0$ and $\hat{\mathbb{E}}[Z_1^2] = 1 + \left(\frac{\nu - r + \frac{1}{2}\sigma^2}{\sigma}\right)^2 \delta$.

Exercise 3.3.2. *Show*

$$\hat{\mathbb{E}}[Z_1] = o(\delta), \quad \text{and} \quad \hat{\mathbb{E}}[Z_1^2] = 1 + o(1)$$

in the following cases:

a) symmetric probabilities

b) subjective return

Exercise 3.3.3. *Consider a risk-neutral trinomial model with N periods presented by*

$$S_{(k+1)\delta} = S_{k\delta} H_{k+1}, \quad \text{for} \quad k = 0, ..., N-1$$

where $\delta := \frac{T}{N}$ and $\{H_k\}_{k=1}^N$ is a sequence of i.i.d. random variables with distribution

$$H_k = \begin{cases} e^{\delta(r - \frac{\sigma^2}{2}) + \sqrt{3\delta}\sigma} & \text{with probability } \hat{\pi} = \frac{1}{6} \\ e^{\delta(r - \frac{\sigma^2}{2})} & \text{with probability } 1 - 2\hat{\pi} = \frac{2}{3} \\ e^{\delta(r - \frac{\sigma^2}{2}) - \sqrt{3\delta}\sigma} & \text{with probability } \hat{\pi} = \frac{1}{6} \end{cases}$$

and $\hat{\pi} < \frac{1}{2}$. Show that as $\delta \to 0$, this trinomial model converges to the Black-Scholes model in the weak sense.
Hint: Find Z_k such that $\ln(H_k) = (r - \frac{\sigma^2}{2})\delta + \sigma\sqrt{\delta}Z_k$. Then, show (3.3.6)

3.3.2 Pricing contingent claims in the Black-Scholes model

Recall from the last section that the limit of the binomial model under risk-neutral probability yields the geometric Brownian motion (GBM)

$$S_t = S_0 \exp\left((r - \frac{\sigma^2}{2})t + \sigma B_t\right). \tag{3.3.7}$$

We can use this random variable S_t to price an option in this continuous setting. We start with a call option with maturity T and strike price K.

Inspired by the geometric Brownian motion market model, the price of this call option is the discounted expected value of $(S_T - K)_+$ under risk-neutral probability. To calculate this price, we only need to know the distribution of S_t under risk-neutral probability, which is given by (3.3.7). Since S_t is a function of a standard Gaussian random variable, we obtain

$$\hat{\mathbb{E}}[(S_T - K)_+] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(S_0 \exp\left((r - \frac{\sigma^2}{2})T + \sigma\sqrt{T}x\right) - K\right)_+ \exp\left(-\frac{x^2}{2}\right) dx.$$

Notice that when $x \leqslant x^* := \frac{1}{\sigma\sqrt{T}}\left(\ln(K/S_0) - (r - \frac{\sigma^2}{2})T\right)$, the integrand is zero and otherwise $(S_T - K)_+ = S_t - K$. Therefore,

$$\begin{aligned}
\hat{\mathbb{E}}[(S_T - K)_+] &= S_0 e^{(r - \frac{\sigma^2}{2})T} \int_{x^*}^{\infty} \frac{e^{\sigma\sqrt{T}x - \frac{x^2}{2}}}{\sqrt{2\pi}} dx - K \int_{x^*}^{\infty} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx \\
&= S_0 e^{rT} \int_{x^*}^{\infty} \frac{e^{-\frac{(x-\sigma\sqrt{T})^2}{2}}}{\sqrt{2\pi}} dx - K \int_{x^*}^{\infty} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx \\
&= S_0 e^{rT} \int_{x^*+\sigma\sqrt{T}}^{\infty} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx - K \int_{x^*}^{\infty} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx
\end{aligned}$$

Notice that $\int_{x^*}^{\infty} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx$ is the probability that a standard Gaussian random variable is greater than x^* and $\int_{x^*+\sigma\sqrt{T}}^{\infty} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx$ is the probability that a standard Gaussian random variable is greater than $x^* + \sigma\sqrt{T}$. Simple calculation shows that

$$x^* + \sigma\sqrt{T} = \frac{1}{\sigma\sqrt{T}}\left(\ln(K/S_0) - (r + \frac{\sigma^2}{2})T\right).$$

In other words, the price of a European call option is given by

$$C(T, K, S_0, 0) := e^{-rT}\hat{\mathbb{E}}[(S_t - K)+] = S_0\Phi(d_1) - e^{-rT}K\Phi(d_2), \tag{3.3.8}$$

where

$$d_1 = \frac{1}{\sigma\sqrt{T}}\Big(\ln(S_0/K) + \big(r + \frac{\sigma^2}{2}\big)T\Big) \text{ and } d_2 = \frac{1}{\sigma\sqrt{T}}\Big(\ln(S_0/K) + \big(r - \frac{\sigma^2}{2}\big)T\Big). \quad (3.3.9)$$

Here $\Phi(x) = \int_{-\infty}^{x} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy$ is the standard Gaussian distribution function.[8]

Exercise 3.3.4. *Use put-call parity to show that the Black-Scholes formula for the price of a put option with maturity T and strike K is given by*

$$P(T, K, S_0, 0) := e^{-rT} K\Phi(-d_2) - S_0\Phi(-d_1).$$

For a general contingent claim, the price of a derivative with payoff $g(S_t)$ in the Black-Scholes model is given by

$$\begin{aligned} V_0(S_0) &:= e^{-rT}\hat{\mathbb{E}}[g(S_T)] = e^{-rT}\hat{\mathbb{E}}\left[g\left(e^{(r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}\mathcal{N}(0,1)}\right)\right] \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g\left(e^{(r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}x}\right)e^{-\frac{x^2}{2}}dx. \end{aligned}$$

As a consequence of Assumption 2.4.1, if we repeat the calculations in Section 3.3.1, we obtain

$$S_t = S_0 \exp\left((r - \frac{\sigma^2}{2})t + \sigma\mathcal{N}(0,t)\right) \quad \text{and} \quad S_T = S_t \exp\left((r - \frac{\sigma^2}{2})(T-t) + \sigma\mathcal{N}(0,T-t)\right).$$

Now we would like to explain the relation between the two normal random variables $\mathcal{N}(0, T-t)$ and $\mathcal{N}(0,t)$ in the above. In the Black-Scholes model, $\frac{S_T}{S_t}$ is independent of S_t. This, in particular, implies that the **Black-Scholes model is Markovian**[9] and given that the price of the underlying asset at time t is equal to S_t, the price of a call option with strike K and maturity T is a function of S_t and t but not S_u for $u < t$. As a result, the price of a call option at time t given $S_t = S$ is given by

$$\begin{aligned} C(T, K, S, t) &:= e^{-r(T-t)}\hat{\mathbb{E}}[(S_T - K)_+ \mid S_t = S] = S\Phi(d_1) - e^{-r(T-t)}K\Phi(d_2), \\ d_1 &= \frac{1}{\sigma\sqrt{T-t}}\left(\ln(S/K) + (r + \frac{\sigma^2}{2})(T-t)\right) \text{ and} \\ d_2 &= \frac{1}{\sigma\sqrt{T-t}}\left(\ln(S/K) + (r - \frac{\sigma^2}{2})(T-t)\right). \end{aligned}$$

In general, the Markovian property implies that for a general contingent claim with payoff

[8] In the above calculations, we use $1 - \Phi(x) = \Phi(-x)$.

[9] Given the current asset price, future movements of the price are independent of past movements.

$g(S_t)$, the Black-Scholes price at time t is $V(t, S_t) := e^{-r(T-t)}\hat{\mathbb{E}}[g(S_T) \mid S_t]$. We will study some more properties of this function $V : [0, T] \times \mathbb{R}_+ \to \mathbb{R}$ later in the chapter. In the Black-Scholes market, a contingent claim that has payoff $g(S_T)$, a function of the price of the underlying asset at time T, is called a *Markovian claim*, and the price of a Markovian claim is given by

$$\begin{aligned} V(t, S) &:= e^{-r(T-t)}\hat{\mathbb{E}}[g(S_T) \mid S_t = S] = e^{-r(T-t)}\hat{\mathbb{E}}\left[g\left(e^{(r-\frac{\sigma^2}{2})(T-t)+\sigma\sqrt{T-t}\mathcal{N}(0,1)}\right)\right] \\ &= \frac{e^{-r(T-t)}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g\left(S\big(e^{(r-\frac{\sigma^2}{2})(T-t)+\sigma\sqrt{T-t}x}\big)\right)e^{-\frac{x^2}{2}}dx. \end{aligned} \tag{3.3.10}$$

Remark 3.3.1. *The price of a Markovian claim in the Black-Scholes model does not depend on past movements of the price; it only depends on the current price S_t. This is not true for non-Markovian claims. For example, a look-back option with payoff $(\max_{0\leqslant t\leqslant T} S_t - K)_+$ or an Asian option $\left(\frac{1}{T}\int_0^T S_t - K\right)_+$ are non-Markovian options with a price that depends to some extent on the history of price movements rather than only on the current price of the underlying.*

As seen in (3.3.10), the Black-Scholes price of a Markovian European option is always a function of $T - t$ rather than t and T separately. Therefore, we can introduce a new variable $\tau := T - t$, *time-to-maturity*. Then, one can write the value of the Markovian European option as a function of τ and S by

$$V(\tau, S) := e^{-r\tau}\hat{\mathbb{E}}[g(S_\tau) \mid S_0 = S] = \frac{e^{-r\tau}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g\left(S\big(e^{(r-\frac{\sigma^2}{2})\tau+\sigma\sqrt{\tau}x}\big)\right)e^{-\frac{x^2}{2}}dx.$$

For call option the Black-Scholes formula in terms of τ is given by

$$\begin{gathered} C(\tau, K, S) = S\Phi(d_1) - e^{-r\tau}K\Phi(d_2), \\ d_1 = \frac{1}{\sigma\sqrt{\tau}}\left(\ln(S/K) + (r + \frac{\sigma^2}{2})\tau\right) \quad \text{and} \quad d_2 = \frac{1}{\sigma\sqrt{\tau}}\left(\ln(S/K) + (r - \frac{\sigma^2}{2})\tau\right). \end{gathered} \tag{3.3.11}$$

3.3.3 Delta hedging

As seen in the binomial model, to hedge the risk of issuing an option, one has to construct a replicating portfolio. The replicating portfolio contains $\Delta_t(S_t)$ units of risky asset at time t. If the price of the asset at time t is given by $S_t = S$, the $\Delta_t(S_t)$ in the binomial model is given by (2.3.4);

$$\Delta_t(S) := \frac{V(t+\delta, Su) - V(t+\delta, S\ell)}{S(u-\ell)}.$$

Using the asymptotic formula for u and ℓ from (3.3.1), we have

$$\Delta_t(S) = \frac{V(t+\delta, S + S(\delta r + \sqrt{\delta}\sigma\alpha)) - V(t+\delta, S + S(1 + \delta r - \sqrt{\delta}\sigma\beta))}{S\sqrt{\delta}\sigma(\alpha+\beta)}$$

By expanding $V(t+\delta, S + S(\delta r + \sqrt{\delta}\sigma\alpha))$ and $V(t+\delta, S + S(1 + \delta r - \sqrt{\delta}\sigma\beta))$ about the point (t, S) and sending $\delta \to 0$, we obtain the Delta of the Black-Scholes model as

$$\Delta_t^{\mathrm{BS}}(S) = \partial_S V(t, S), \tag{3.3.12}$$

where $V(t, S)$ is the Black-Scholes price of a general contingent claim with any given payoff.
As for a call option with maturity T and strike K, by taking derivative with respect to S from (3.3.11), we obtain

$$\Delta_t^{\mathrm{BS}}(S) = \Phi(d_1).$$

Here we used

$$S\partial_S d_1 \Phi'(d_1) - e^{-r(T-t)} K \partial_S d_2 \Phi'(d_2) = 0.$$

Exercise 3.3.5. *Let $S_0 = \$10$, $\sigma = .03$, and $r = 0.03$. Find the Black-Scholes Delta of the following portfolio.*

position	*units*	*type*	*strike*	*maturity*
long	*3*	*call*	*\$10*	*60 days*
short	*3*	*put*	*\$10*	*90 days*
short	*1*	*call*	*\$10*	*120 days*

The maturities are given in business days.

Exercise 3.3.6. *Let $S_0 = 10$, $\sigma = .03$, and $r = 0.03$. Consider the portfolio below.*

position	*units*	*type*	*strike*	*maturity*
long	*3*	*call*	*\$10*	*60 days*
long	*4*	*put*	*\$5*	*90 days*
?	*x*	*underlying*	*NA*	*NA*

How many units x of the underlying are required to eliminate any sensitivity of the portfolio with respect to the changes in the price of the underlying?

Another derivation of the Delta in the Black-Scholes model

Another heuristic derivation of this result is as follows. In the binomial model, we can write

$$\hat{\mathbb{E}}[g(S_T) \mid S_{t+\delta} = Su] = \hat{\mathbb{E}}[g(S_{T-\delta}u) \mid S_t = S] \quad \text{and}$$
$$\hat{\mathbb{E}}[g(S_T) \mid S_{t+\delta} = S\ell] = \hat{\mathbb{E}}[g(S_{T-\delta}l) \mid S_t = S].$$

This is because the binomial model is time homogeneous. Therefore,

$$\begin{aligned}
\Delta_t(S) &= \frac{V(t+\delta, Su) - V(t+\delta, S\ell)}{S(u-\ell)} \\
&= e^{-r(T-t)} \frac{\hat{\mathbb{E}}[g(S_T) \mid S_{t+\delta} = Su] - \hat{\mathbb{E}}[g(S_T) \mid S_{t+\delta} = S\ell]}{S(u-\ell)} + O(\delta) \\
&= e^{-r(T-t)} \frac{\hat{\mathbb{E}}[g(S_{T-\delta}u) \mid S_t = S] - \hat{\mathbb{E}}[g(S_{T-\delta}l) \mid S_t = S]}{S(u-\ell)} + O(\delta) \\
&= e^{-r(T-t)} \frac{\hat{\mathbb{E}}[S_{T-\delta}g'(S_{T-\delta}) \mid S_t = S](u-\ell)}{S(u-\ell)} + O(\delta).
\end{aligned}$$

In the above, we used (3.3.1) to obtain $(u-\ell)^2 = O(\delta)$, and we used

$$g(x) - g(y) = g'(x)(y-x) + O((x-y)^2)$$

to obtain the last equality. Hence,

$$\Delta_t(S) = \frac{e^{-r(T-t)}}{S} \hat{\mathbb{E}}[S_{T-\delta}g'(S_{T-\delta}) \mid S_t = S] + O(\delta)$$

Now, by the weak convergence of the underlying asset price in the binomial model to the asset price in the Black-Scholes model, as $\delta \to 0$, we obtain

$$\Delta_t^{\mathrm{BS}}(S) = \frac{e^{-r(T-t)}}{S} \hat{\mathbb{E}}[S_T g'(S_T) \mid S_t = S].$$

Notice that given $S_t = S$, we have

$$S_T = S \exp\left((r - \frac{\sigma^2}{2})(T-t) + \sigma\mathcal{N}(0, T-t)\right).$$

Therefore,

$$\frac{d}{dS}\Big(g(S_T)\Big) = \exp\left((r - \frac{\sigma^2}{2})(T-t) + \sigma\mathcal{N}(0, T-t)\right) g'(S_T) = \frac{S_T}{S} g'(S_T).$$

This implies that

$$\begin{aligned}
\Delta_t^{\mathrm{BS}}(S) &= e^{-r(T-t)} \hat{\mathbb{E}}\Big[\frac{S_T}{S} g'(S_T) \mid S_t = S\Big] = e^{-r(T-t)} \hat{\mathbb{E}}\Big[\frac{d}{dS}\Big(g(S_T)\Big) \mid S_t = S\Big] \\
&= \frac{d}{dS}\Big(e^{-r(T-t)} \hat{\mathbb{E}}[g(S_T) \mid S_t = S]\Big) = \partial_S V(t, S).
\end{aligned}$$

3.3.4 Completeness of the Black-Scholes model

Similar to the binomial model, the Black-Scholes market is complete; every contingent claim is replicable. For the moment, it is not our concern to show this rigorously. Instead, we accept this fact and would rather emphasize how to replicate a contingent claim. In order to replicate a Markovian contingent claim with payoff $g(S_T)$ in the Black-Scholes model, we start by recalling from Section 2.3.3 that the replicating portfolio in the binomial model is determined by $\Delta^{\mathrm{bi}}_{t_i}(S_{t_i})$ given by (2.3.4) and the replicating portfolio is written as

$$V^{\mathrm{bi}}(t,S_t) = V^{\mathrm{bi}}(0,S_0) + R\sum_{i=0}^{t-1}(V^{\mathrm{bi}}(i,S_i) - \Delta^{\mathrm{bi}}_i(S_i)S_i) + \sum_{i=0}^{t-1}\Delta^{\mathrm{bi}}_i(S_i)(S_{i+1}-S_i). \tag{3.3.13}$$

Notice that $R = r\delta + o(\delta)$. By taking the limit from (3.3.13), we obtain

$$V(T,S_T) = V(0,S_0) + r\int_0^T (V(t,S_t) - \Delta_t(S_t)S_t)dt + \int_0^T \Delta_t(S_t)dS_t. \tag{3.3.14}$$

In the above $V(t,S)$ is the Black-Scholes price of the contingent claim, and $\Delta_t(S_t)$ satisfies (3.3.12). The first integral in (3.3.14) is a simple Riemann integral. The second integral is a more complicated stochastic integral; the integrator dS_t is an Itô stochastic, which is presented in Section C.4. But for the moment, you can interpret the Itô integral in (3.3.14) as the limit of the discrete stochastic integral $\sum_{i=0}^{N-1}\Delta_{t_i}(S_{t_i})(S_{t_{i+1}} - S_{t_i})$.

Exercise 3.3.7. *Repeat the above calculation to show that the discounted value of the option $(1+R)^{-t}V^{bi}(t,S_t)$ converges to*

$$e^{-rT}V(T,S_T) = V(0,S_0) + \int_0^T e^{-rt}\Delta_t(S_t)dS_t.$$

3.3.5 Error of discrete hedging in the Black-Scholes model and Greeks

Equation (3.3.14) suggests adjusting the Delta continuously in time to replicate the contingent claim. On one hand, this is a useful formula, because in reality trading can happen with enormous speed which makes continuous time a fine approximation. However, in practice, the time is still discrete, and hedging is only a time lapse. Therefore, it is important to have some estimation of the error of discrete-time hedging in the Black-Scholes framework.

Let us consider that the Black-Scholes model is running continuously in time, but we only adjust our position on the approximately replicating portfolio at times $t_i := \delta i$ where $\delta = \frac{T}{N}$ and $i = 0,1,...,N$. By setting aside the accumulated error until time t_i, we can assume that our hedge has been perfect until time t_i, for some i, i.e.

$$V(t_i,S_{t_i}) = \Delta^{\mathrm{BS}}_{t_i}(S_{t_i})S_{t_i} + Y_{t_i},$$

where $V(t,S)$ is the Black-Scholes value of the contingent claim and Y_t is the position in cash. At time $t_{i+1} = t_i + \delta$, the value of the portfolio is

$$\Delta_{t_i}^{\mathrm{BS}}(S_{t_i})S_{t_i+\delta} + e^{r\delta}Y_{t_i}.$$

Since by (3.3.12), $\Delta_{t_i}^{\mathrm{BS}}(S_{t_i}) = \partial_S V(t_i, S_{t_i})$, the error is given by

$$\begin{aligned}
\mathrm{Err}_{t_i}(\delta) :=& \hat{\mathbb{E}}\Big[V(t_i+\delta, S_{t_i+\delta}) - \partial_S V(t_i, S_{t_i})S_{t_i+\delta} - e^{r\delta}Y_{t_i} \mid S_{t_i}\Big] \\
=& \hat{\mathbb{E}}\Big[V(t_i+\delta, S_{t_i+\delta}) - \partial_S V(t_i, S_{t_i})S_{t_i+\delta} - e^{r\delta}(V(t_i, S_{t_i}) - \partial_S V(t_i, S_{t_i})S_{t_i}) \mid S_{t_i}\Big] \\
=& \hat{\mathbb{E}}\Big[V(t_i+\delta, S_{t_i+\delta}) - V(t_i, S_{t_i}) - \partial_S V(t_i, S_{t_i})(S_{t_i+\delta} - S_{t_i}) \\
& + (e^{r\delta} - 1)(V(t_i, S_{t_i}) - \partial_S V(t_i, S_{t_i})S_{t_i}) \mid S_{t_i}\Big] \\
=& \hat{\mathbb{E}}\Big[V(t_i+\delta, S_{t_i+\delta}) - V(t_i, S_{t_i}) - \partial_S V(t_i, S_{t_i})(S_{t_i+\delta} - S_{t_i}) \\
& - \delta r(V(t_i, S_{t_i}) - \partial_S V(t_i, S_{t_i})S_{t_i})) \mid S_{t_i}\Big] + O(\delta^2)
\end{aligned}$$

By the Taylor formula, the price of the option is

$$\begin{aligned}
V(t_i+\delta, S_{t_i+\delta}) = & V(t_i, S_{t_i}) + \partial_t V(t_i, S_{t_i})\delta + \partial_S V(t_i, S_{t_i})(S_{t_i+\delta} - S_{t_i}) \\
& + \frac{1}{2}\partial_{SS} V(t_i, S_{t_i})(S_{t_i+\delta} - S_{t_i})^2 + \frac{1}{2}\partial_{tt} V(t_i, S_{t_i})\delta^2 \\
& + \partial_{St} V(t_i, S_{t_i})(S_{t_i+\delta} - S_{t_i})\delta + o(\delta).
\end{aligned}$$

Notice that since

$$S_{t_i+\delta} - S_{t_i} = \sigma S_{t_i}\sqrt{\delta}\mathcal{N}(0,1) + O(\delta), \quad \text{with } \mathcal{N}(0,1) \text{ independent of } S_{t_i}, \tag{3.3.15}$$

the conditional expectation $\hat{\mathbb{E}}[(S_{t_i+\delta} - S_{t_i})^2 \mid S_{t_i}] = \sigma^2 S_{t_i}^2\delta + O(\delta^{\frac{3}{2}})$. Therefore, by the properties of the conditional expectation, we have

$$\begin{aligned}
\hat{\mathbb{E}}[\partial_S V(t_i, S_{t_i})(S_{t_i+\delta} - S_{t_i}) \mid S_{t_i}] &= \partial_S V(t_i, S_{t_i})\hat{\mathbb{E}}[S_{t_i+\delta} - S_{t_i} \mid S_{t_i}] = 0 \\
\hat{\mathbb{E}}[\partial_{SS} V(t_i, S_{t_i})(S_{t_i+\delta} - S_{t_i})^2 \mid S_{t_i}] &= \partial_{SS} V(t_i, S_{t_i})\hat{\mathbb{E}}[(S_{t_i+\delta} - S_{t_i})^2 \mid S_{t_i}] \\
&= \sigma^2 S_{t_i}^2 \partial_{SS} V(t_i, S_{t_i})\delta + O(\delta^{\frac{3}{2}}),
\end{aligned}$$

and we can write the error term[10] by

$$\mathrm{Err}_{t_i}(\delta) := \delta(\partial_t V(t_i, S_{t_i}) + \frac{\sigma^2 S_{t_i}^2}{2}\partial_{SS}V(t_i, S_{t_i}) + r\partial_S V(t_i, S_{t_i})S_{t_i}) - rV(t_i, S_{t_i})) + O(\delta^{\frac{3}{2}}). \tag{3.3.16}$$

where the term $O(\delta^{\frac{3}{2}})$ depends on the higher derivative $\partial_{tS}V$, σ, S_{t_i}, and r.

Before finishing the error estimation, let us briefly explain some of the important terms which show up in (3.3.16).

1) The first derivative $\partial_S V$ of the option price, *Delta*, is denoted by $\Delta(t, S)$ and determines the sensitivity of the value of the option with respect to the price of the underlying.

2) The second derivative $\partial_{SS}V$ of the option price, which is called *Gamma* and denoted by $\Gamma(t, S)$ and determines the convexity of the option value on the price of underlying.

3) The time derivative $\partial_t V$, which is called *time decay factor* or *Theta* and is denoted by $\Theta(t, S)$, determines how the price of option evolves over time.

As a function of time-to-maturity $\tau = T - t$, by the abuse of notation, we define $V(\tau, S) := V(t, S)$ and therefore we have

$$\begin{aligned}
\Delta(\tau, S) &= \partial_S V(\tau, S)\\
\Gamma(\tau, S) &= \partial_{SS} V(\tau, S)\\
\Theta(\tau, S) &= -\partial_\tau V(\tau, S)
\end{aligned}$$

For example in the case of call option with strike K and maturity T, by taking derivatives ∂_{SS} and ∂_τ in (3.3.11), we have

$$\Delta(\tau, S) = \Phi(d_1),\ \Gamma(\tau, S) = \frac{1}{S\sigma\sqrt{\tau}}\Phi'(d_1) \text{ and } \Theta(\tau, S) = -\frac{S\sigma}{2\sqrt{\tau}}\Phi'(d_1) - rKe^{-r\tau}\Phi(d_2). \tag{3.3.17}$$

The Δ, Γ, and Θ of a call option in the Black-Scholes model is shown in Figure 3.3.1.

To continue with the error estimation, we need the following proposition.

Proposition 3.3.1. *For a European Markovian contingent claim, the Black-Scholes price satisfies*

$$\Theta(\tau, S) = -\frac{\sigma^2 S^2}{2}\Gamma(\tau, S) - rS\Delta(\tau, S) + rV(\tau, S).$$

Proof. Apple Proposition (C.1) to the martingale $e^{-rt}V(t, S_t)$. □

[10] Here we heuristically assumed $\mathcal{N}(0, 1) \sim 1$. A more rigorous treatment of this error is in by calculating the L^2 error by calculating $\hat{\mathbb{E}}[(\mathrm{Err}_{t_i}(\delta))^2]^{\frac{1}{2}}$.

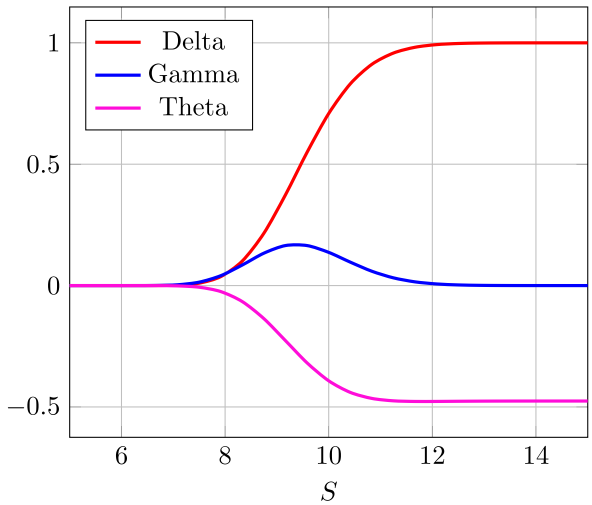


Figure 3.3.1: Greeks of a call option with $\tau = 1$, $\sigma = .1$, $r = .05$, and $K = 10$. As you see the significant sensitivity is near the ATM.

As a result of this proposition, the term of order δ in (3.3.16) vanishes, and thus one step error is of order $O(\delta^{\frac{3}{2}})$. Since $\delta = \frac{T}{N}$, we have

$$\mathrm{Err}(\delta) := \hat{\mathbb{E}}\left[\sum_{i=0}^{N-1} \mathrm{Err}_{t_i}(\delta)\right] = O(\sqrt{\delta}),$$

which converges to 0 as quickly as $\sqrt{\delta}$ when $\delta \to 0$.

Discrete hedging without a money market account

One reason to completely disregard the money market account is because the risk-free interest rate r is not exactly constant. The money market is also under several risks which is a different topic. One way to tackle the interest rate risks is to completely exclude the risk-free money market account from the hedging portfolio, and try to measure the hedging error in this case. To understand this better, let us first consider the issuer of an option that is long in $\Delta^{\mathrm{BS}}(t_i, S_{t_i})$ units of the underlying. Then the change in the portfolio from time t_i to time $t_i + \delta$ is

$$\begin{aligned}
&\hat{\mathbb{E}}[V(t_i+\delta, S_{t_i+\delta}) - V(t_i, S_{t_i}) - \Delta^{\mathrm{BS}}(t_i, S_{t_i})(S_{t_i+\delta} - S_{t_i})] \\
&\qquad = (\partial_t V(t_i, S_{t_i}) + \frac{\sigma^2 S_{t_i}^2}{2}\partial_{SS} V(t_i, S_{t_i}))\delta + o(\delta),
\end{aligned}$$

where we use (3.3.15) on the right-hand side. This error is related to the loss/profit of not perfectly hedging and is called a *slippage error*; see Figure 3.3.2.

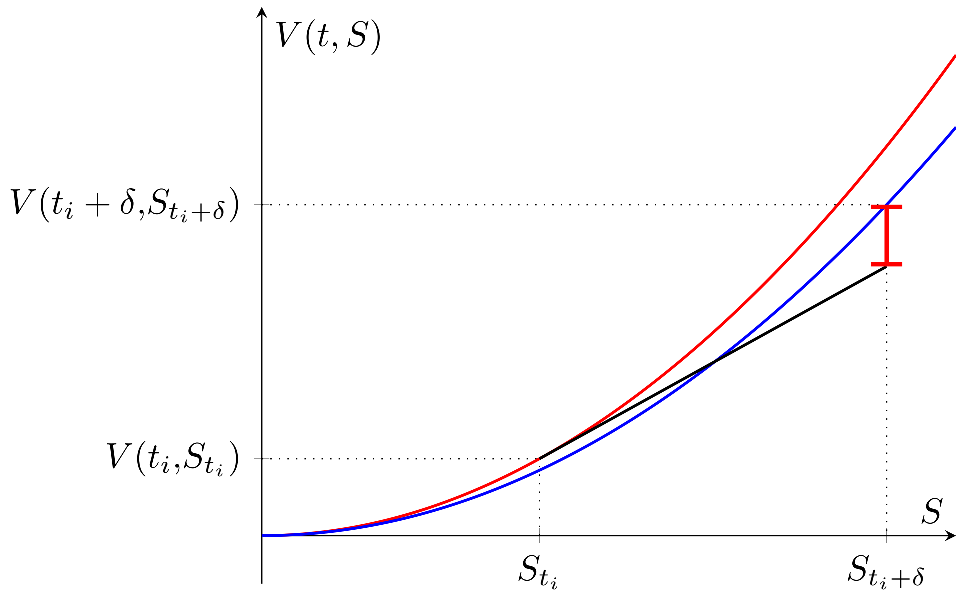


Figure 3.3.2: The red curve is the value of the option at time t_i, the blue curve is the value of the option at time t_{i+1}. The slippage error is shown in burgundy.

The loss/profit from slippage can be calculated by using the same technology as in the previous section; the slippage error during the time interval $[t_i, t_i+\delta]$ is given by

$$(\Theta(t_i, S_{t_i}) + \frac{\sigma^2 S_{t_i}^2}{2}\Gamma(t_i, S_{t_i}))\delta + o(\delta)$$

As illustrated in Figure 3.3.4, when, for instance, the time decay factor is negative and Gamma is positive, for small changes in the price of the asset, we lose, and for larger changes we gain. As seen in (3.3.17), it is a typical situation to have negative Θ and positive Γ for call options (or put options or any European Markovian option with a convex payoff function). See Figure 3.3.3.

Exercise 3.3.8. *Show that if the payoff function $g(S_T)$ is a convex function on S_T, then the Markovian European contingent claim with payoff $g(S_T)$ has nonnegative Γ; $V(\tau, S)$ is convex on S for all τ.*

Let $\tilde{S}_t = e^{-rt}S_t$ and $\tilde{V}(t, \tilde{S}_t) = e^{-rt}V(t, S_t) = e^{-rt}V(t, e^{rt}\tilde{S}_t)$ be respectively the discounted underlying price and discounted option price. Then, we can show that

$$\partial_t \tilde{V}(t, \tilde{S}) = -\frac{\sigma^2 \tilde{S}^2}{2}\partial_{\tilde{S}\tilde{S}}\tilde{V}(t, \tilde{S}).$$

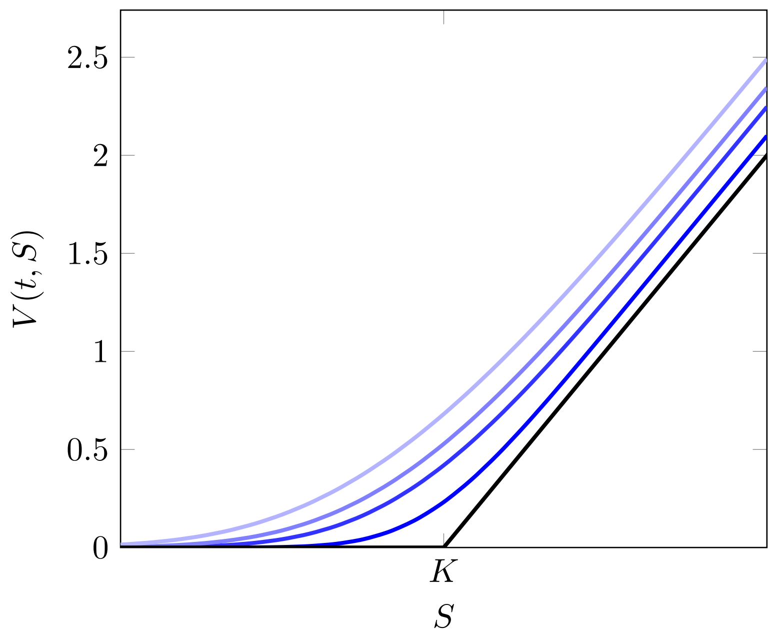


Figure 3.3.3: The time decay for the price of call option in the Black-Scholes model. As $\tau \to 0$, the color becomes darker.

Exercise 3.3.9. *Use Proposition 3.3.1 to show the above equality.*

This suggests that if the interest rate is nearly zero, then the lack of a money market in the replication does not impose any error. Otherwise, when the interest rate is large, the slippage error is significant and is equal to

$$(\Theta(t_i, S_{t_i}) + \frac{\sigma^2 S_{t_i}^2}{2}\Gamma(t_i, S_{t_i}))\delta + o(\delta) = r(V(t_i, S_{t_i}) - S_{t_i}^2\Delta(t_i, S_{t_i}))\delta + o(\delta),$$

which can accumulate to a large number.

Other Greeks

Two other Greeks are *Rho*, denoted by ρ, and *Vega*, denoted by $\mathcal{V}$, which respectively measure the sensitivity with respect to interest rate r and volatility σ, i.e.,

$$\rho(\tau, S) := \partial_r \left(e^{-r\tau}\hat{\mathbb{E}}[g(S_\tau) \mid S_0 = s]\right) \quad \text{and} \quad \mathcal{V}(\tau, S) := \partial_\sigma \left(e^{-r\tau}\hat{\mathbb{E}}[g(S_\tau) \mid S_0 = s]\right).$$

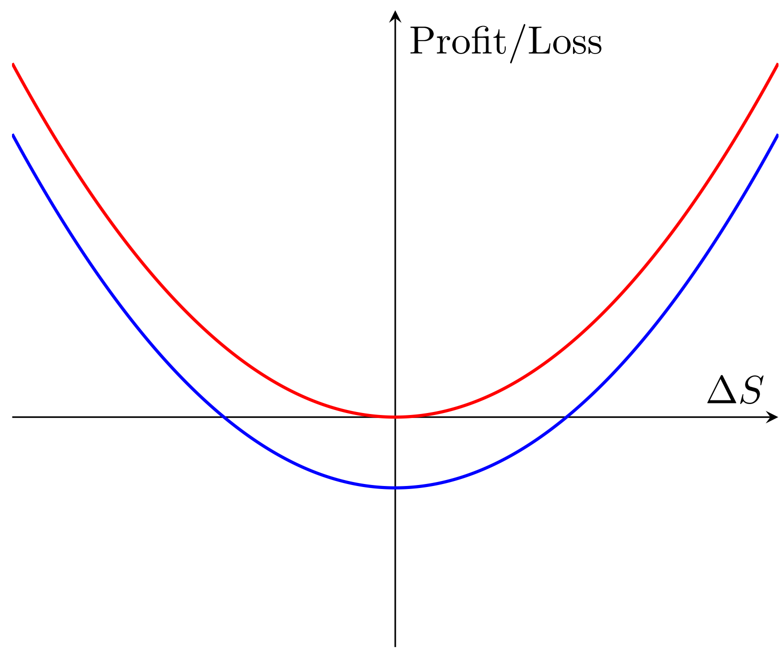


Figure 3.3.4: The loss/profit of the discrete hedging in the Black-Scholes model. The red graph shows the loss/profit without the time decay factor. The blue includes the time decay factor, too.

For a call option, these derivatives are given by

$$\rho(\tau, S) = e^{-r\tau} K\tau\Phi(d_2) \quad \text{and} \quad \mathcal{V}(\tau, S) = S\sqrt{\tau}\Phi'(d_1).$$

Figures (3.3.1) and 3.3.5 show the Greeks Δ, Γ, Θ, ρ and $\mathcal{V}$ for a call option as a function of S.

Exercise 3.3.10. *The third derivative of the Black-Scholes price with respect to S is called speed. Find a closed-form solution for speed.*

Example 3.3.1. *The payoff in Figure 3.3.6 can be written as $(S_T - K_1)_+ - (S_T - K_2)_+ - (S_T - K_3)_+ + (S_T - K_4)_+$. Therefore, the closed-form solution for the Black-Scholes price of the option is given by*

$$V(\tau, S) = C(\tau, K_1, S) - C(\tau, K_2, S) - C(\tau, K_3, S) + C(\tau, K_4, S).$$

All the Greeks of the option are also a linear combination of the Greeks of these call options. For instance,

$$\Delta(t, S) = \Phi(d_1(\tau, K_1, S)) - \Phi(d_1(\tau, K_2, S)) - \Phi(d_1(\tau, K_3, S)) + \Phi(d_1(\tau, K_4, S)).$$

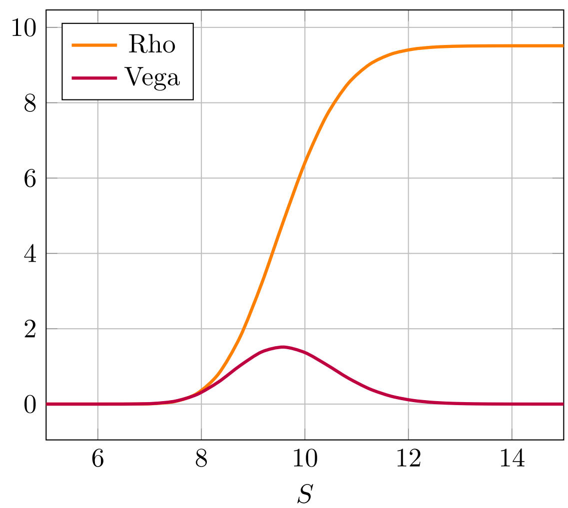


Figure 3.3.5: ρ and $\mathcal{V}$ of a call option with $\tau = 1$, $\sigma = .1$, $r = .05$, and $K = 10$.

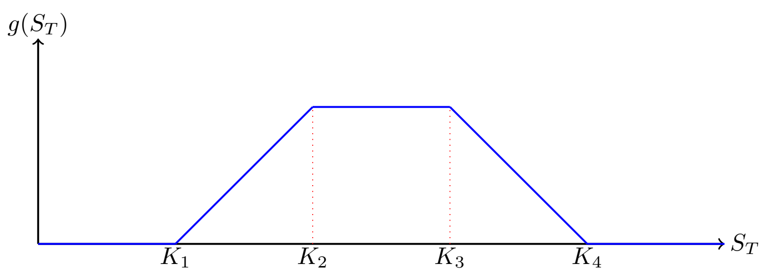


Figure 3.3.6: Payoff of Example 3.3.1 $g(S_T) = (S_T - K_1)_+ - (S_T - K_2)_+ - (S_T - K_3)_+ + (S_T - K_4)_+$.

Exercise 3.3.11. *Write the payoffs in Figure 3.3.7 as a linear combination of call options and derive a closed-form formula for the Black-Scholes price, the Delta, the Gamma, and the time decay of options with these payoffs.*

Exercise 3.3.12 (Bull and bear call spreads)**.** *Write the payoffs in Figure 3.3.8 as linear combination of call options with different strikes and possibly some cash and give the closed form formula for them.*

3.3.6 Time-varying Black-Scholes model

Recall from Section 2.4 that the binomial model can be calibrated to time-dependent parameters. Let $\delta = \frac{T}{N}$ and consider the discrete time instances $t_i = i\delta$. The time-varying

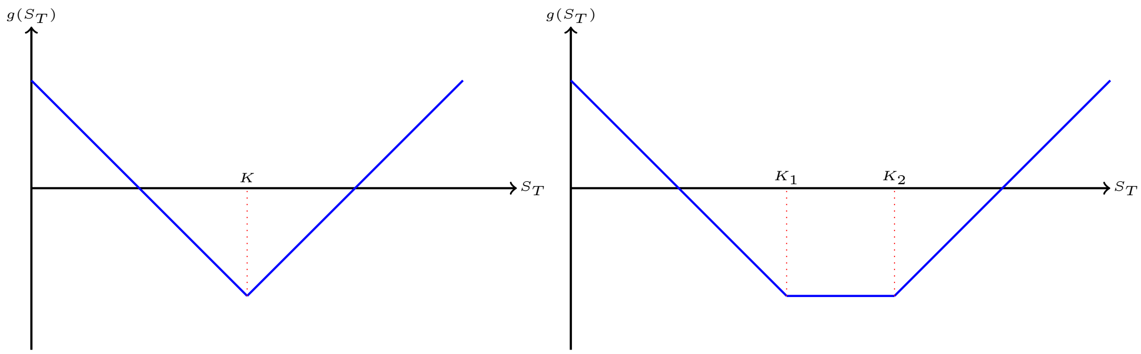


Figure 3.3.7: Left: payoff of straddle. Right: payoff of strangle.

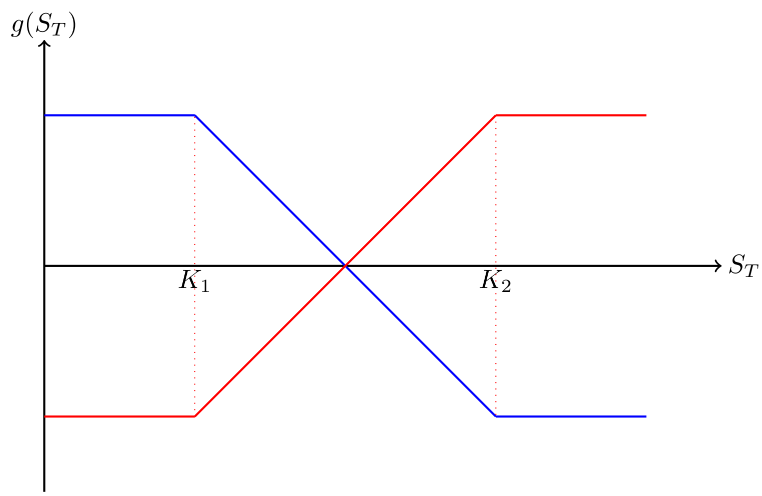


Figure 3.3.8: Red: Bull-spread call. Blue: Bear-spread call

binomial can be given by

$$\ln(S_{t_{k+1}}) = \ln(S_{k\delta}) + \ln(H_{k+1}),$$

where $\{H_k\}_{k=1}^N$ is a sequence of independent random variables with the following distribution under the risk-neutral probability

$$H_k = \begin{cases} 1 + \delta r_{t_k} + \sqrt{\delta}\sigma_{t_k}\alpha_k & \text{with probability } \hat{\pi}_k \\ 1 + \delta r_{t_k} - \sqrt{\delta}\sigma_{t_k}\beta_k & \text{with probability } 1 - \hat{\pi}_k \end{cases},$$

where α_k, β_k are given by (2.4.6) and (2.4.7) for time-dependent λ_t. Therefore, equation (3.3.4) takes the following time-dependent form:

$$\ln(S_{t_k}) = \ln(S_0) + \sum_{i=0}^{k-1}(r_{t_i} - \frac{\sigma_{t_i}^2}{2})\delta + \sqrt{\delta}\sum_{k=1}^{N}\sigma_{t_k}Z_k. \tag{3.3.18}$$

Analogous to (3.3.5), we have $\pi_k = \frac{\beta_k}{\alpha_k+\beta_k}$ and

$$\hat{\mathbb{E}}[Z_k] = 0 \quad \text{and} \quad \hat{\mathbb{E}}[Z_k^2] = 1.$$

Therefore,

$$\chi_{\sqrt{\delta}\sigma_{t_k}Z_k}(\theta) = \hat{\mathbb{E}}[e^{i\theta\sigma_{t_k}Z_k}] = 1 + i\theta\sigma_{t_k}\delta\hat{\mathbb{E}}[Z_1] - \frac{\theta^2\sigma_{t_k}^2\delta\hat{\mathbb{E}}[Z_1^2]}{2} + o(\delta).$$

This implies that

$$\chi_{\sqrt{\delta}\sum_{k=1}^N\sigma_{t_k}Z_k}(\theta) = \prod_{k=1}^{N}\left(1 - \frac{1}{2}\delta\theta^2\sigma_{t_k}^2 + o(\delta^2)\right) = \prod_{k=1}^{N}e^{-\frac{1}{2}\theta^2\sigma_{t_k}^2\delta}+o(\delta) = e^{-\frac{\theta^2}{2}\sum_{i=1}^N\sigma_{t_k}^2\delta}+o(\delta).$$

As $\delta \to 0$,

$$\chi_{\sqrt{\delta}\sum_{k=1}^N\sigma_{t_k}Z_k}(\theta) \to e^{-\frac{\theta^2}{2}\int_0^T\sigma_t^2dt},$$

which is the characteristic function of $\mathcal{N}(0, \int_0^T\sigma_t^2dt)$. Thus in the limit, we have

$$S_T = S_0\exp\left(\int_0^T(r_t - \frac{\sigma_t^2}{2})dt + \int_0^T\sigma_t^2dt\mathcal{N}(0,1)\right).$$

As a matter of fact, for any t we have

$$\begin{aligned} S_t &= S_0\exp\left(\int_0^t(r_u - \frac{\sigma_u^2}{2})du + \mathcal{N}\left(0, \int_0^t\sigma_u^2du\right)\right) \\ S_T &= S_t\exp\left(\int_t^T(r_u - \frac{\sigma_u^2}{2})du + \mathcal{N}\left(0, \int_t^T\sigma_u^2du\right)\right). \end{aligned}$$

and the random variables $\mathcal{N}\left(0, \int_t^T\sigma_u^2du\right)$ and $\mathcal{N}\left(0, \int_0^t\sigma_u^2du\right)$ are independent.

Usually the interest rate r_t and the volatility σ_t are not given and we have to estimate them from the data. In the next section, we discuss some estimation methods for these two parameters.

Using the variable r and σ, we can rewrite the Black-Scholes formula for a call option by

$$C(T,K,S,t) := e^{-\int_t^T r_u du}\hat{\mathbb{E}}[(S_T-K)_+ \mid S_t = S] = S\Phi(d_1) - e^{-\int_t^T r_u du}K\Phi(d_2),$$
$$d_1 = \frac{1}{\sqrt{\int_t^T \sigma_u^2 du}}\left(\ln(S/K) + \int_t^T \left(r_u + \frac{\sigma_u^2}{2}\right)du\right) \text{ and}$$
$$d_2 = \frac{1}{\sqrt{\int_t^T \sigma_u^2 du}}\left(\ln(S/K) + \int_t^T \left(r_u - \frac{\sigma_u^2}{2}\right)du\right).$$

For a general European payoff $g(S_T)$ we have the Black-Scholes price given by

$$V(t,S) := e^{-\int_t^T r_u du}\hat{\mathbb{E}}[g(S_\tau) \mid S_0 = S]$$
$$= \frac{e^{-\int_t^T r_u du}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g\left(S\big(e^{-\int_t^T (r_u - \frac{\sigma_u^2}{2})du + x\sqrt{\int_t^T \sigma_u^2 u du}}\big)\right)e^{-\frac{x^2}{2}}dx.$$

3.3.7 Black-Scholes with yield curve and forward interest rate

Recall from Section 1.1.5 that the yield curve $R_t(T)$ and forward curve $F_t(T)$ of a zero bond are defined by

$$B_t(T) = e^{-(T-t)R_t(T)} = e^{-\int_t^T F_t(u)du} \quad \text{or} \quad \text{or} \quad R_t(T) := -\frac{1}{T-t}\ln B_t(T) \quad F_t(T) := -\frac{d\ln B_t(T)}{dT}.$$

Since setting a model for the forward rate is equivalent to setting a model for the short rate r_t, in the Black-Scholes formula with a time-varying interest rate, one can equivalently use the forward rate or the yield curve. Assume that σ is constant. Then, the Black-Scholes pricing formula becomes

$$C(T,K,S,t) := e^{-\int_t^T F_t(u)du}\hat{\mathbb{E}}[(S_T-K)_+ \mid S_t = S] = S\Phi(d_1) - e^{-\int_t^T F_t(u)du}K\Phi(d_2),$$
$$d_1 = \frac{1}{\sqrt{\sigma^2(T-t)}}\left(\ln(S/K) + \int_t^T F_t(u)du + \frac{\sigma^2}{2}(T-t)\right) \text{ and}$$
$$d_2 = \frac{1}{\sqrt{\sigma^2(T-t)}}\left(\ln(S/K) + \int_t^T F_t(u)du - \frac{\sigma^2}{2}(T-t)\right).$$

and

$$\begin{aligned} C(T,K,S,t) &:= e^{-R_t(T)(T-t)}\hat{\mathbb{E}}[(S_T-K)_+ \mid S_t=S] = S\Phi(d_1) - e^{-R_t(T)(T-t)}K\Phi(d_2), \\ d_1 &= \frac{1}{\sqrt{\sigma^2(T-t)}}\left(\ln(S/K) + R_t(T)(T-t) + \frac{\sigma^2}{2}(T-t)\right) \text{ and} \\ d_2 &= \frac{1}{\sqrt{\sigma^2(T-t)}}\left(\ln(S/K) + R_t(T)(T-t) - \frac{\sigma^2}{2}(T-t)\right). \end{aligned} \tag{3.3.19}$$

3.3.8 Black-Scholes model and Brownian motion

In Section B.5, we show that the symmetric random walk converges to the Brownian motion B_t. The same principle shows that the linear interpolation of the summation $\sum_{k=1}^{N} Z_k$ in the logarithm of the asset price in the binomial model in (3.3.4) also converges to the Brownian motion. Therefore, one can write the Black-Scholes model (3.3.7) by using the Brownian motion B_t;

$$S_t = S_0 \exp\left((r - \frac{\sigma^2}{2})t + \sigma B_t\right). \tag{3.3.20}$$

Here B_t is a Gaussian random variable with mean zero and variance t. The above process is called a geometric Brownian motion (GBM for short)[11].

Markovian property of the Black-Scholes model

Since for Brownian motion the increment $B_{s+t} - B_t$ is independent of B_t,

$$S_t = S_0 \exp\left((r - \frac{\sigma^2}{2})t + \sigma B_t\right) \quad \text{and} \quad S_{t+s} = S_t \exp\left((r - \frac{\sigma^2}{2})s + \sigma(B_{t+s} - B_t)\right)$$

are independent. Verbally, this means that future price movements are independent of past movements. In other words, given the history of the movement of an asset's price until time t, i.e., S_u for all $u \in [0,t]$, the distribution of S_{t+s} only depends on S_t and that part of history during $[0,t)$ is irrelevant. As a result, for any function $g : \mathbb{R}_+ \to \mathbb{R}$, we have

$$\hat{\mathbb{E}}[g(S_{t+s}) \mid S_u :\ \forall u \in [0,t]] = \hat{\mathbb{E}}[g(S_{t+s}) \mid S_t] = \hat{\mathbb{E}}\left[g\left(S_t \exp\left((r - \frac{\sigma^2}{2})s + \sigma(B_{t+s} - B_t)\right)\right)\right].$$

The pricing formula (3.3.10) is precisely derived from the Markovian property of the Black-Scholes model.

[11]Paul Samuelson, in the 1950's, first came up with the idea of using GBM to model the risky asset price. His primary motivation is that GBM never generates negative prices, which overcomes one of the drawbacks of Bachelier model, negative prices for an asset.

As a result of the Markovian property of GBM, one can write

$$S_{t+dt} - S_t = S_t \left(\exp\left((r - \frac{\sigma^2}{2})dt + \sigma(B_{t+dt} - B_t) \right) - 1 \right)$$
$$= S_t\Big((r - \frac{\sigma^2}{2})dt + \sigma(B_{t+dt} - B_t) + \frac{1}{2}\sigma^2(B_{t+dt} - B_t)^2 \Big) + o(dt).$$

Therefore, the short-term return of an asset in the Black-Scholes model is given by

$$\frac{S_{t+dt} - S_t}{S_t} = (r - \frac{\sigma^2}{2})dt + \sigma(B_{t+dt} - B_t) + \frac{1}{2}\sigma^2(B_{t+dt} - B_t)^2 + o(dt).$$

This is, in particular, consistent with Assumption 2.4.1 and the definition of mean return and volatility, i.e.

$$\hat{\mathbb{E}}\left[\frac{S_{t+dt} - S_t}{S_t}\right] = (r - \frac{\sigma^2}{2})dt + \sigma\hat{\mathbb{E}}[B_{t+dt} - B_t] + \frac{1}{2}\sigma^2\hat{\mathbb{E}}[(B_{t+dt} - B_t)^2] + o(dt) = rdt + o(dt).$$
$$\text{var}\left(\frac{S_{t+dt} - S_t}{S_t}\right) = \sigma^2 dt + o(dt).$$

Inspired from the above formal calculation, We formally write the Black-Scholes model for the asset price as

$$\frac{dS_t}{S_t} = rdt + \sigma dB_t. \tag{3.3.21}$$

The above equation, which describes (3.3.20) in the differential form, is called the *Black-Scholes stochastic differential equation.*

Martingale property of the Black-Scholes model

Similar to the binomial model, in the Black-Scholes model the lack of arbitrage is equivalent to the martingale property of the discounted asset price. The discounted asset price in the Black-Scholes model is given by

$$\tilde{S}_t = e^{-rt}S_t = S_0 \exp\left(-\frac{\sigma^2}{2}t + \sigma B_t\right).$$

The conditional expectation of $\tilde{S}_{t+s}$ given $\tilde{S}_t$ is then given by

$$\begin{aligned}
&S_0\hat{\mathbb{E}}\left[\exp\left(-\frac{\sigma^2}{2}(t+s)+\sigma B_{t+s}\right)\Big|\tilde{S}_t\right]\\
&= S_0\exp\left(-\frac{\sigma^2}{2}t+\sigma B_t\right)\hat{\mathbb{E}}\left[\exp\left(-\frac{\sigma^2}{2}s+\sigma(B_{t+s}-B_t)\right)\Big|\tilde{S}_t\right]\\
&= \tilde{S}_t\hat{\mathbb{E}}\left[\exp\left(-\frac{\sigma^2}{2}s+\sigma(B_{t+s}-B_t)\right)\Big|\tilde{S}_t\right].
\end{aligned}$$

By the independence of the increments of Brownian motion, we have

$$\hat{\mathbb{E}}\left[\exp\left(-\frac{\sigma^2}{2}s+\sigma(B_{t+s}-B_t)\right)\Big|\tilde{S}_t\right]=\hat{\mathbb{E}}\left[\exp\left(-\frac{\sigma^2}{2}s+\sigma(B_{t+s}-B_t)\right)\right],$$

and therefore,

$$\hat{\mathbb{E}}\left[\tilde{S}_{t+s}|\tilde{S}_t\right]=\tilde{S}_t\hat{\mathbb{E}}\left[\exp\left(-\frac{\sigma^2}{2}s+\sigma(B_{t+s}-B_t)\right)\right].$$

On the other hand since $B_{t+s}-B_t\sim\mathcal{N}(0,s)$, we have

$$\hat{\mathbb{E}}\left[\exp\left(\sigma(B_{t+s}-B_t)\right)\right]=\exp\left(\frac{\sigma^2}{2}s\right),$$

and therefore, $\hat{\mathbb{E}}\left[\tilde{S}_{t+s}|\tilde{S}_t\right]=\tilde{S}_t$.

In addition to the asset price, the discounted price of a newly introduced contingent claim must be a martingale to preserve the no-arbitrage condition. Recall from formula (3.3.10) that the price of a contingent claim with payoff $g(S_T)$ is given by

$$V(t,S_t)=e^{-r(T-t)}\hat{\mathbb{E}}[g(S_T)\mid S_t].$$

If we define the discounted price by $\tilde{V}(t,S_t)=e^{-rt}V(t,S_t)$, then we can write the above as

$$\tilde{V}(t,S_t)=e^{-rT}\hat{\mathbb{E}}[g(S_T)\mid S_t].$$

By the tower property of conditional expectation, we have

$$\hat{\mathbb{E}}\left[\tilde{V}(t+s,S_{t+s})|S_t\right]=\hat{\mathbb{E}}\left[e^{-rT}\hat{\mathbb{E}}[\hat{\mathbb{E}}[g(S_T)\mid S_{t+s}]\Big|S_t\right]=e^{-rT}\hat{\mathbb{E}}\left[g(S_T)|S_t\right]=\tilde{V}(t,S_t).$$

Therefore, the price of the contingent claim is a martingale under risk-neutral probability.

3.3.9 Physical versus risk-neutral in the Black-Scholes model

Recall from (2.4.1) that the binomial model under physical probability is given by

$$S_{(k+1)\delta} = S_{k\delta} H_{k+1}, \quad \text{for} \quad k = 0, ..., N-1,$$

where the sequence of i.i.d. random variables $\{H_k\}_{k=1}^N$ is given by

$$H_k = \begin{cases} u & \text{with probability } \ p \\ \ell & \text{with probability } \ 1-p \end{cases}$$

Then, the dynamics of the asset price under physical probability is given by

$$\ln(S_T) = \ln(S_0) + \sum_{k=1}^{N} \ln(H_k).$$

$\ln(H_k)$ is the log return $\mathbf{R}_{k\delta}^{\log}$. Recall from (3.3.3) that the first two moments of log return are given by

$$\mathbb{E}[\ln(H_k)] = (\mu - \frac{1}{2}\sigma^2)\delta + O(\delta^2) \text{ and } \mathbb{E}[\ln(H_k)^2] = \sigma^2\delta + o(\delta).$$

If we define $Z_k' := \frac{\ln(H_k) - (\mu - \sigma^2/2)}{\sigma\sqrt{\delta}}$, one can write

$$\ln(S_T) = \ln(S_0) + \left(\mu - \frac{1}{2}\sigma^2\right)T + \sigma\sqrt{T} \cdot \frac{1}{\sqrt{N}} \sum_{k=1}^{N} Z_k',$$

where $\{Z_k'\}$ is a sequence of i.i.d. random variables satisfying

$$\mathbb{E}[Z_k'] = o(\delta), \quad \text{and} \quad \mathbb{E}[(Z_k')^2] = 1 + o(1).$$

If follows from (3.3.6) that $\frac{1}{\sqrt{N}} \sum_{k=1}^{N} Z_k'$ converges weakly to $\mathcal{N}(0,1)$, and under physical probability, the Black-Scholes model is described by

$$S_t = S_0 \exp\left((\mu - \frac{\sigma^2}{2})t + \sigma B_t\right).$$

In other words, by switching from risk-neutral probability to physical probability, we adjust the mean return of the asset from r to the risk-free interest rate μ.

Remark 3.3.2. *For the purpose of derivative pricing, physical probability is irrelevant. This is because by the fundamental theorem of asset pricing, Theorem 2.1.1, the price of any derivative is determined by the discounted expectation of payoff under risk-neutral* *probability. However, for portfolio management, physical probability is important, because it carries the long-term growth rate of the asset, μ. For example, an investor with x initial wealth wants to decide how to split his investment between a risk-free bond with interest rate r and a risky asset given by the Black-Scholes model*

$$S_t = S_0 \exp\left((\mu - \frac{\sigma^2}{2})t + \sigma B_t\right).$$

His objective is to maximize his expected wealth at time T subject to a constraint of the risk of portfolio measured by the variance of the wealth at time T, i.e.,

$$\max_\theta \{\mathbb{E}[X_T^\theta] - \lambda \text{var}(X_T^\theta)\}. \tag{3.3.22}$$

where X_T^θ is the wealth of the investor at time T if he chooses to invest $x_0 - \theta$ in the risk-free asset and θ in the risky asset. The wealth X_T^θ satisfies

$$X_T^\theta = e^{rT}(x_0 - \theta) + \theta e^{(\mu - \frac{\sigma^2}{2})T + \sigma B_T}.$$

Therefore,

$$\mathbb{E}[X_T^\theta] = e^{rT}(x_0 - \theta) + \theta e^{\mu} \quad \textit{and} \quad \text{var}(X_T^\theta) = \theta^2 e^{2\mu T}(e^{\sigma^2 T} - 1).$$

Therefore, the portfolio maximization problem (3.3.22) is given by

$$\max_\theta \left\{ e^{rT}(x_0 - \theta) + \theta e^{\mu T} - \lambda \theta^2 e^{2\mu T}(e^{\sigma^2 T} - 1) \right\}$$

and the solution is given by

$$\theta^* = \frac{e^{\mu T} - e^{rT}}{2\lambda e^{2\mu T}(e^{\sigma^2 T} - 1)}.$$

Volatility estimation

Notice that the log return of the Black-Scholes model satisfies

$$\ln\left(\frac{S_{t+\delta}}{S_t}\right) = (\mu - \frac{\sigma^2}{2})\delta + \sigma(B_{t+\delta} - B_t)$$

We consider the time discretization $t_0 = 0$, $t_{i+1} = t_i + \delta$, and $t_N = t$. Therefore, it follows from (B.15) (the quadratic variation of Brownian motion) that

$$\sum_{i=0}^{N-1} \ln\left(\frac{S_{t_{i+1}}}{S_{t_i}}\right)^2 = \sigma^2 \sum_{i=0}^{N-1} (B_{t_{i+1}} - B_{t_i})^2 + o(\delta) \to \sigma^2 t \tag{3.3.23}$$

One can use (3.3.23) to make an estimation of volatility. For any $t_i = i\delta$, evaluate

$$Y_{t_i} := \sum_{j=0}^{i-1} \ln\left(\frac{S_{t_{j+1}}}{S_{t_j}}\right)^2.$$

If δ is small enough, then Y_{t_i} should approximately be equal to $t_i\sigma^2$. This suggests that if we fit a line to the data $\{(t_i, Y_{t_i}) \ : i = 0, 1, ...\}$, the slope of line is σ^2. The schematic picture of this fitting is shown in figure 3.3.9.

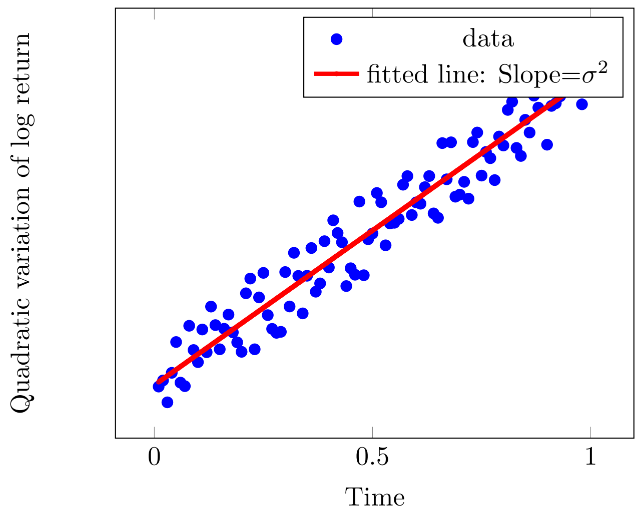


Figure 3.3.9: Quadratic variation estimation of volatility

Exercise 3.3.13 (Project)**.** *With the same set of data from Exercise 2.4.1, calculate the volatility by using the quadratic variation formula (3.3.23). Take $\delta = \frac{1}{250}$ year. Then, plot the estimated quadratic variation Y_{t_i} versus time t_i. Some examples of the generated plots are given in Figure 3.3.10. Then fit a line to the data by assuming $Y_t = \sigma^2 t +$ noise. To fit a line to the data points, you can use the least square method. The slope of the line gives you the volatility σ.*

3.3.10 Black-Scholes partial differential equation

Recall from Proposition 3.3.1 that the price $V(\tau, S)$ of the Markovian European contingent claim with payoff $g(S_T)$ at time $t = T - \tau$ (τ is time-to-maturity) when the asset price is S satisfies

$$-\partial_\tau V(\tau, S) = -\frac{\sigma^2 S^2}{2}\partial_{SS}V(\tau, S) - rS\partial_S(\tau, S) + rV(\tau, S).$$

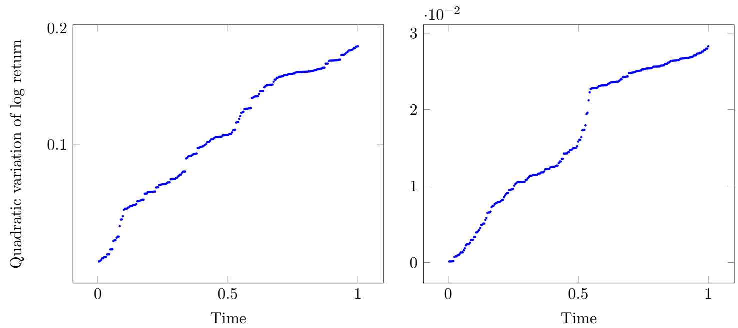


Figure 3.3.10: Quadratic variation estimation of volatility: Left: Tesla. Right: S&P500

The above equation is a partial differential equation called the *Black-Scholes equation*. One way to find the pricing formula for an option is to solve the Black-Scholes PDE. As with all PDEs, a boundary condition and an initial condition[12] are required to solve the PDE analytically or numerically. The initial condition for the Black-Scholes equation is the payoff of the derivative, i.e.

$$V(0, S) = g(S).$$

Notice that here when $\tau = 0$, we have $t = T$.

Notice that the above PDE holds in region $\tau \in [0, T]$ and $S > 0$. Therefore, we need a boundary condition at $S = 0$. This boundary condition is a little tricky to devise, because in the Black-Scholes model, the price of the asset S_t never hits zero; if the price of an asset is initially positive, then the price stays positive at all times. If the price of the asset is initially 0, then $S_T = 0$ and the price of the option with payoff $g(S_T)$ is given by

$$V(\tau, 0) = e^{-r\tau}\hat{\mathbb{E}}[g(S_T) \mid S_{T-\tau} = 0] = e^{-r\tau}g(0).$$

Therefore, the boundary condition for $S = 0$ is given by

$$V(\tau, 0) = e^{-r\tau}g(0).$$

To summarize, the Black-Scholes PDE for pricing a Markovian European contingent claim with payoff $g(S_T)$ is given by the following boundary value problem

$$\begin{cases} \partial_\tau V(\tau, S) & = \frac{\sigma^2 S^2}{2}\partial_{SS}V(\tau, S) + rS\partial_S V(\tau, S) - rV(\tau, S) \\ V(\tau, 0) & = e^{-r\tau}g(0) \\ V(0, S) & = g(S) \end{cases}. \tag{3.3.24}$$

Exercise 3.3.14. *Consider the option with payoff $g(S_T) = \frac{1}{\sqrt{S_T}}$. Find the Black-Scholes*

[12]For time-dependent PDEs.

price of this payoff at time $t = 0$ by solving PDE (3.3.24). Hint: Try to plug in $V(\tau, S) = e^{-r\tau}S^a$ into the Black-Scholes equation, for some constant a. Then, find the constant a. The boundary condition $V(\tau, 0) = e^{-r\tau}g(0)$ is unnecessary as $g(0) = \infty$.

Remark 3.3.3. *Somehow the boundary condition $V(\tau, 0) = e^{-r\tau}g(0)$ is redundant in equation (3.3.24), because in the Black-Scholes model, the price of the underlying asset never hits zero. But this condition is important for solving the Black-Scholes equation numerically.*

Heat equation and Black-Scholes model

Recall from Exercise 3.3.9 that the change of variable $\tilde{S} := e^{-r(T-\tau)}S$ and $\tilde{V}(\tau, \tilde{S}) := e^{-r(T-\tau)}V(\tau, S) = e^{-r(T-\tau)}V(\tau, e^{r(T-\tau)}\tilde{S})$ modifies the Black-Scholes equation to

$$\partial_\tau \tilde{V}(\tau, \tilde{S}) = \frac{\sigma^2 \tilde{S}^2}{2} \partial_{\tilde{S}\tilde{S}} \tilde{V}(\tau, \tilde{S}).$$

If we make a change of variables $x := \ln(\tilde{S})$ and $U(\tau, x) := \tilde{V}(\tau, e^x)$, then we obtain the heat equation

$$\partial_\tau U(\tau, x) = \frac{\sigma^2}{2} \partial_{xx} U(\tau, x).$$

Unlike (3.3.24), the heat equation holds on the whole space, i.e.

$$\begin{cases} \partial_\tau U(\tau, x) & = \frac{\sigma^2}{2} \partial_{xx} U(\tau, x) \\ U(0, x) & = e^{-rT} V(e^{rT} e^x) \end{cases}. \tag{3.3.25}$$

Notice that at time $t = 0$ (or $\tau = T$), the price of the Markovian European contingent claim with payoff $g(S_T)$ is equal to $V(T, S_0) = \tilde{V}(T, S_0) = U(T, \ln(S_0))$.

Transforming the Black-Scholes equation into the heat equation helps us to apply numerical techniques for the heat equation that are developed in Section 3.2.2 to obtain the Black-Scholes price of an option.

3.3.11 Numerical methods for the Black-Scholes price of a European option

Although the Black-Scholes price of call options, put options, digital options, or a linear combination of them has to a closed-form solution, there are many payoffs that does not. As a result, one need to develop numerical methods to evaluate the price of such options. This section, covers the finite-difference scheme for the Black-Scholes equation, the binomial model approximation of the Black-Scholes model, and the Monte Carlo approximation of the price of a European option in the Black-Scholes model. All the methods in Section 3.2.2 can also be applied after transferring the Black-Scholes into the heat equation, as discussed in Section 3.3.10.

Solving the Black-Scholes equation via finite-difference scheme

One can directly discretize the Black-Scholes equation (3.3.24) to apply the finite-difference method described in the previous section. See Figure 3.3.11. In this case, the computation domain has to be $[0, S_{\max}]$ for some $S_{\max} > 0$. The boundary condition at 0 is already assigned at (3.3.24) by

$$V(\tau, 0) = e^{-r\tau} g(0),$$

and the ABC at $S_{\max}$ is given by the growth of the payoff for large values of S. For a call option, for instance, the ABC is

$$V(t, S_{\max}) = S_{\max} \quad \text{for } S_{\max} \text{ sufficiently large.}$$

The rest of the approximation follows as was presented in Section 3.3.10. However, one should be cautious about applying explicit schemes, which need the CFL condition. Recall that the right-hand side of the CFL condition is always $\frac{1}{2}$ times the inverse of the coefficient of second derivative in the equation. Therefore, the CFL condition translates to

$$\frac{h}{k^2} \leqslant \frac{1}{\sigma^2 S^2}.$$

If $S_{\max}$ is chosen very large, for a fixed discretization k of variable S, one needs to set the discretization h of variable t very small. The downside of this method is that the time of the algorithm increases significantly. In such cases, transferring the Black-Scholes equation into the heat equation and implementing the explicit finite-difference scheme in Section 3.3.10 is more efficient.

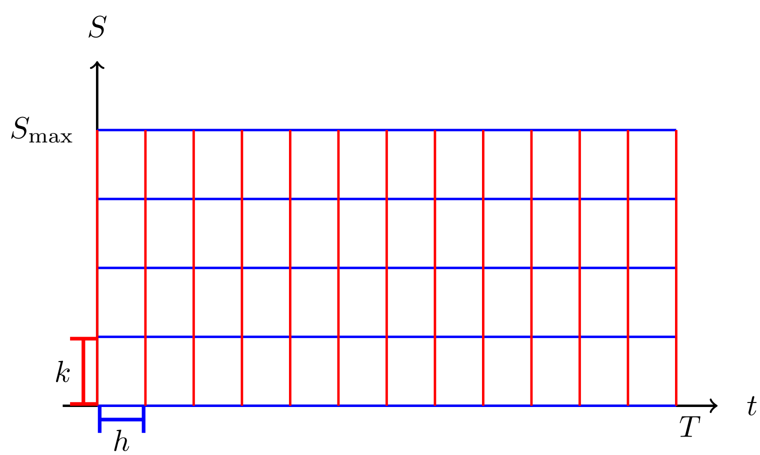


Figure 3.3.11: Finite-difference grid for the Black-Scholes equation. In the explicit scheme the CFL condition should be satisfied, i.e., $\frac{h}{k^2} \leqslant \frac{1}{\sigma S}$. This requires choosing a very small h. Artificial boundary conditions are necessary on both $S_{\max}$ and 0.

Binomial model scheme for the Black-Scholes equation

Recall from Section 3.3.1 that the Black-Scholes model is the limit of the binomial model under risk-neutral probability. Therefore, if necessary, one can use the binomial model to approximate the Black-Scholes price of the contingent claim. For implementation, one needs to choose a sufficiently large number of periods N or, equivalently, a small $\delta = \frac{T}{N}$. Suggested by (3.3.1), for a given interest rate r and volatility σ, we need to choose u and ℓ and the one-period interest rate R as follows.

$$u = e^{\delta(r-\frac{\sigma^2}{2})+\sqrt{\delta}\sigma\alpha}, \quad l = e^{\delta(r-\frac{\sigma^2}{2})-\sqrt{\delta}\sigma\beta}, \quad \text{and} \quad R = r\delta,$$

where α and β are given by (2.4.6). However, one can avoid the calculation of α and β by making some different choices. Notice that the binomial model has three parameters u, ℓ and R while the Black-Scholes parameters are only two. This degree of freedom provides us with some modifications of binomial model, which still converges to the Black-Scholes formula. This also simplifies the calibration process in Section 2.4 significantly. Here are some choices:

1) Symmetric probabilities:

$$u = e^{\delta(r-\frac{\sigma^2}{2})+\sqrt{\delta}\sigma}, \quad l = e^{\delta(r-\frac{\sigma^2}{2})-\sqrt{\delta}\sigma}, \quad \text{and} \quad R = r\delta,$$

Then

$$\hat{\pi}_u = \hat{\pi}_\ell = \frac{1}{2}.$$

2) Subjective return:

$$u = e^{\delta\nu+\sqrt{\delta}\sigma}, \quad l = e^{\delta\nu-\sqrt{\delta}\sigma}, \quad \text{and} \quad R = r\delta,$$

Then

$$\hat{\pi}_u = \frac{1}{2}\left(1+\sqrt{\delta}\frac{r-\nu-\frac{1}{2}\sigma^2}{\sigma}\right) \quad \text{and} \quad \hat{\pi}_\ell = \frac{1}{2}\left(1-\sqrt{\delta}\frac{r-\nu-\frac{1}{2}\sigma^2}{\sigma}\right).$$

To see why this choices work, recall (3.3.4) from Section 3.3.1. The only criteria for the convergence of the binomial model to the Black-Scholes model is that when we write the log of asset price in the binomial model as in (3.3.4),

$$\ln(S_T) = \ln(S_0) + (r - \frac{\sigma^2}{2})T + \sigma\sqrt{T}\cdot\frac{1}{\sqrt{N}}\sum_{k=1}^{N} Z_k,$$

the random variables Z_k, $k = 1, ..., N$ must satisfy

$$\hat{\mathbb{E}}[Z_1] = o(\delta) \quad \text{and} \quad \hat{\mathbb{E}}[Z_1^2] = 1 + o(1).$$

In Section 3.3.1, we made a perfect choice of $\hat{\mathbb{E}}[Z_1] = 0$ and $\hat{\mathbb{E}}[Z_1^2] = 1$.

Monte Carlo scheme for Black-Scholes

Recall from Section 3.3.2 that the price of a Markovian European contingent claim can be written as an expectation and/or a single integral.

$$V(\tau, S) := e^{-r\tau}\hat{\mathbb{E}}[g(S_\tau) \mid S_0 = S] = \frac{e^{-r\tau}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g\left(S\big(e^{(r-\frac{\sigma^2}{2})\tau+\sigma\sqrt{\tau}x}\big)\right) e^{-\frac{x^2}{2}}\, dx.$$

One way to estimate $V(T, S_0)$ is to generate a sample $x^{(1)}, ..., x^{(M)}$ of $\mathcal{N}(0, 1)$ and approximate the above expectation by

$$\frac{e^{-rT}}{M} \sum_{j=1}^{M} g\left(S_0\big(e^{(r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}x^{(j)}}\big)\right). \tag{3.3.26}$$

Another class of approximation methods, *quadrature methods*, directly targets the integral by choosing a large number $0 < x_{\max}$ and approximate the integral

$$\int_{-x_{\max}}^{x_{\max}} g\left(S\big(e^{(r-\frac{\sigma^2}{2})\tau+\sigma\sqrt{\tau}x}\big)\right) e^{-\frac{x^2}{2}}\, dx$$

by a finite Riemann sum. For instance, if we choose $\Delta x := \frac{x_{\max}}{L}$ and $x_{(j)} = j\Delta x$, the approximation goes as follows:

$$\begin{aligned} V(\tau, S) &\approx \frac{e^{-r\tau}}{\sqrt{2\pi}} \int_{-x_{\max}}^{x_{\max}} g\left(S\big(e^{(r-\frac{\sigma^2}{2})\tau+\sigma\sqrt{\tau}x}\big)\right) e^{-\frac{x^2}{2}}\, dx \\ &\approx e^{-rT} \sum_{j=-L}^{L-1} g\left(S_0\big(e^{(r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}x_{(j)}}\big)\right) e^{-\frac{x_{(j)}^2}{2}} \Delta x. \end{aligned}$$

Quadrature methods, if implemented carefully, can be more efficient than the naive Monte Carlo scheme (3.3.26).

Exercise 3.3.15 (Project)**.** *Consider the payoff of a bull-spread call in Figure 3.3.8 with $T = 1$, $K_1 = 10$, and $K_2 = 20$. Assume that the parameters of the underlying asset are given by $S_0 = 15$ and $\sigma = .02$, and that the interest rate is $r = .01$. Compare the following approximation schemes for the price of the bull-spread call. Record the time of the algorithms for each scheme to obtain four-digit accuracy.*

a) The closed-form solution for the value of this option from the Black-Scholes formula.

b) Implicit finite-difference with parameter $\theta = .5$.

c) Implicit finite-difference with parameter $\theta = 1$.

d) Explicit finite-difference (equivalently implicit with $\theta = 0$).

e) Symmetric binomial model.

f) Monte Carlo scheme.

g) Approximation by a Riemann sum.

3.3.12 Stock price with dividend in the Black-Scholes model

IN this section, we consider two types of dividend strategies. If the dividend is paid continuously, then there is a constant outflow of cash from the price of the asset. If the rate of dividend payment is D_t, the Black-Scholes model in (3.3.21) has to be modified to

$$dS_t = rS_t dt + \sigma dB_t - D_t dt.$$

Choosing $D_t := qS_t$ for $q \geqslant 0$, we obtain

$$dS_t = (r - q)S_t dt + \sigma S_t dB_t.$$

Remark 3.3.4. *For a continuous dividend rate q, the dividend yield in time period $[t, T]$ is given by $e^{-q(T-t)} \times 100$ percent.*

Especially, this choice guarantees that the dividend is always less than the asset price and that paying the dividend does not diminish the value of the asset. In exponential form, we have

$$S_t = S_0 \exp\left((r - q)t + \sigma B_t\right) = e^{-qt} S_0 \exp\left((r - \frac{\sigma^2}{2})t + \sigma B_t\right).$$

In this case, the Black-Scholes price of a European contingent claim with payoff $g(S_T)$ is given by

$$V(t, S) = e^{-r(T-t)}\hat{\mathbb{E}}[g(S_T) \mid S_t = S] = e^{-r(T-t)}\hat{\mathbb{E}}\Big[g\Big(e^{-q(T-t)}\tilde{S}_T\Big) \mid \tilde{S}_t = S\Big],$$

where $\tilde{S}_t$ satisfies the Black-Scholes equation without dividend, i.e.

$$\tilde{S}_t = S_0 \exp\left((r - \frac{\sigma^2}{2})t + \sigma B_t\right).$$

If the dividend strategy is time-varying q_t, then one can write the above pricing formula as

$$V(t,S) = e^{-r(T-t)}\hat{\mathbb{E}}\Big[g\Big(e^{-\int_t^T q_s ds}\tilde{S}_T\Big) \mid \tilde{S}_t = S\Big].$$

For continuous-rate dividend, the relation between the Greeks of the option in Proposition 3.3.1 is modified as in the following proposition.

Proposition 3.3.2. *Let S_t follow the Black-Scholes price with a continuous dividend rate q. For a European Markovian contingent claim, the Black-Scholes price satisfies*

$$\Theta(\tau,S) = -\frac{\sigma^2 S^2}{2}\Gamma(\tau,S) - (r-q)S\Delta(\tau,S) + rV(\tau,S).$$

The second type of dividend strategy is discrete. The discrete dividend is paid in times $0 \leqslant t_0 < t_1 < ... < t_N \leqslant T$ in amounts $D_0, ..., D_N$. Then, between the times of dividend payment, the asset price follows the Black-Scholes model, i.e.

$$S_t = S_{t_{n-1}} \exp\left((r - \frac{\sigma^2}{2})(t - t_{n-1}) + \sigma(B_t - B_{t_{n-1}})\right), \quad t \in [t_{n-1}, t_n).$$

Just a moment before the payment of dividend at time t_n-[13] the price of the asset is

$$S_{t_n-} := S_{t_{n-1}} \exp\left((r - \frac{\sigma^2}{2})(t_n - t_{n-1}) + \sigma(B_{t_n} - B_{t_{n-1}})\right).$$

But, after paying a dividend of D_n, this price drops to

$$S_{t_n} = S_{t_n-} - D_n = S_{t_{n-1}} \exp\left((r - \frac{\sigma^2}{2})(t_n - t_{n-1}) + \sigma(B_{t_n} - B_{t_{n-1}})\right) - D_n.$$

As mentioned in Section 2.3.4, the dividend is usually given as a percentage of the current asset price, i.e.

$$D_n = d_n S_{t_n-}, \quad \text{with} \quad d_n \in [0,1).$$

and we have

$$S_{t_n} = (1-d_n)S_{t_n-} = (1-d_n)S_{t_{n-1}} \exp\left((r - \frac{\sigma^2}{2})(t_n - t_{n-1}) + \sigma(B_{t_n} - B_{t_{n-1}})\right).$$

Remark 3.3.5. *For a discrete dividend, the dividend yield in time period $[t_{n-1}, t_n]$ is $d_n \times 100$ percent.*

[13] The moment before time t is denoted by $t-$.

Therefore,

$$S_T = \prod_{n=1}^{N}(1-d_n)S_0 \exp\left((r-\frac{\sigma^2}{2})T + \sigma B_T\right).$$

Remark 3.3.6. *If at time T there is a dividend payment, the payoff of a contingent claim $g(S_T)$ takes into account the price S_T after the payment of the dividend. In short, the price of the asset at the maturity is always ex-dividend.*

Proposition 3.3.3. *Let S_t follow the Black-Scholes price with the discrete dividend policy given by $d_0, ..., d_N \in [0,1)$ at times $0 \leqslant t_0 < t_1 < ... < t_N \leqslant T$. For a European Markovian contingent claim, the Black-Scholes price satisfies*

$$\Theta(\tau, S) = -\frac{\sigma^2 S^2}{2}\Gamma(\tau, S) - rS\Delta(\tau, S) + rV(\tau, S) \quad \textit{for} \quad \tau \in (T-t_n, T-t_{n-1}).$$

Then, the Black-Scholes price of a contingent claim with payoff $g(S_T)$ is given by

$$V(t,S) = e^{-r(T-t)}\hat{\mathbb{E}}[g(S_T) \mid S_t = S] = e^{-r(T-t)}\hat{\mathbb{E}}\Big[g\Big(\prod_{t<t_n\leqslant T}(1-d_n)\tilde{S}_T\Big) \,\Big|\tilde{S}_t = S\Big],$$

where $\tilde{S}_t$ satisfies the simple no-dividend Black-Scholes model.

Exercise 3.3.16. *Consider a portfolio of one straddle option with $K = 10$ and one strangle option with $K_1 = 8$ and $K_2 = 14$, both maturing at $T = 1$. See Figure 3.3.7. Assume that the underlying asset has parameters $S_0 = 2$ and $\sigma = .2$, and that it pays a 4% quarterly dividend. The interest rate is $r = .01$ (1%). Find the price of this portfolio and its Δ and Γ at time $t = 0$.*

Remark 3.3.7. *Dividend strategies are sometimes not known upfront and therefore should be modeled by random variables. If the dividend policy is a random policy that depends on the path of the stock price, then, the pricing formula will be more complicated even in the Black-Scholes model.*

Exercise 3.3.17 (Butterfly spread)**.** *Consider the payoff $g(S_T)$ shown in Figure 3.3.12.*

Consider the Black-Scholes model for the price of a risky asset with $T = 1$, $r = .04$, and $\sigma = .02$ with dividends paid quarterly with a dividend yield of 10%. Take $S_0 = 10$, $K_1 = 9$, and $K_2 = 11$. Find the Black-Scholes price, Δ, Γ, ρ, and $\mathcal{V}$ of this option at time $t = 0$. Find Θ at time $t = 0$ without taking derivatives with respect to S.

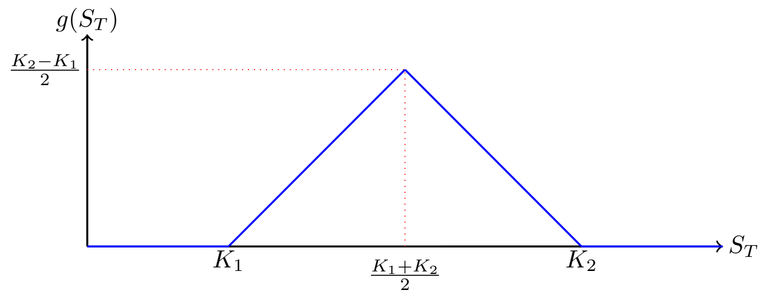


Figure 3.3.12: Butterfly spread payoff

The Black-Scholes equation with dividend

In the case of dividends, the Black-Scholes equation in (3.3.24) is given by

$$\begin{cases} \partial_\tau V(\tau, S) &= \frac{\sigma^2 S^2}{2}\partial_{SS}V(\tau, S) + (r-q)S\partial_S(\tau, S)V(\tau, S) - rV(\tau, S) \\ V(\tau, 0) &= e^{-r\tau}g(0) \\ V(0, S) &= g(S) \end{cases}.$$

This also allows us to solve the derivative pricing problem with more complicated dividend strategies. Let's assume that the dividend payment rate at time t is a function $q(t, S_t)$. Then, the Black-Scholes model with dividend is given by the SDE

$$dS_t = S_t(r - q(t, S_t))dt + \sigma S_t dB_t.$$

and Black-Scholes equation in (3.3.24) is given by

$$\begin{cases} \partial_\tau V(\tau, S) &= \frac{\sigma^2 S^2}{2}\partial_{SS}V(\tau, S) + (r-q(t, S))S\partial_S V(\tau, S) - rV(\tau, S) \\ V(\tau, 0) &= e^{-r\tau}g(0) \\ V(0, S) &= g(S) \end{cases}.$$

After the change of variables described in Section 3.3.10, we obtain

$$\begin{cases} \partial_\tau U(\tau, x) &= \frac{\sigma^2}{2}\partial_{xx}U(\tau, x) - q(T-\tau, e^{r(T-\tau)}e^x)\partial_x U(\tau, x), \\ U(0, x) &= e^{-rT}V(e^{rT}e^x) \end{cases}.$$

The above equation is a heat equation with a drift term given by $q(T-\tau, e^{r(T-\tau)}e^x)\partial_x U(\tau, x)$.

Exercise 3.3.18 (Project)**.** *Consider a European call option with $T = 1$ and $K = 2$. Assume that the parameters of the underlying asset are given by $S_0 = 2$ and $\sigma = .2$, and that the interest rate is $r = .01$ (1%). In addition, assume that the underlying asset pays* *dividends at continuous rate* $q(t, S_t) = .05e^{-.01t}S_t$*, i.e., 5% of the discounted asset price.*

a) *Write the Black-Scholes equation for this problem and convert it into a heat equation with a drift term.*

b) *Solve this problem numerically by using a finite-difference scheme.*

Example 3.3.2. *Consider a European option with payoff*

$$g(S_T) = e^{\alpha S_T}.$$

Assume that the interest rate is $r > 0$ *and that the volatility of the underlying asset is* $\sigma > 0$*, and at time* 0 *it has value* S_0*, and pays dividends at a continuous rate* $q(t, S_t) = qS_t$*, where* $q > 0$*. Then, the Black-Scholes equation is given by*

$$\begin{cases} \partial_\tau V(\tau, S) & = \frac{\sigma^2 S^2}{2}\partial_{SS}V(\tau, S) + (r - qS)S\partial_S V(\tau, S) - rV(\tau, S) \\ V(0, S) & = e^{\alpha S}. \end{cases}.$$

The boundary condition for the Black-Scholes equation at $S = 0$ *is given by*

$$V(\tau, 0) = 1.$$

Function $V(\tau, S) = e^{-r\tau}e^{\alpha S}$ *satisfies the Black-Scholes equation and the boundary conditions. Therefore, the price of the contingent claim with payoff* $e^{\alpha S_T}$ *at time-to-maturity* τ *(at time* $T - \tau$*) is given by* $V(\tau, S) = e^{-r\tau}e^{\alpha S}$ *if the asset price takes value* S *at that time.*

Exercise 3.3.19. *Consider a European option with payoff*

$$g(S_T) = S_T^{-5}e^{10S_T}.$$

Assume that the interest rate is $r = .1$ *and the underlying asset satisfies* $S_0 = 2$ *and* $\sigma = .2$*, and that pays dividends at a continuous rate of* $q(t, S_t) = .2S_t$*.*

a) *Write the Black-Scholes equation for this problem.*

b) *Solve this problem analytically by the method of separation of variables. Plug into the equation a solution candidate with the form* $e^{\alpha\tau}S^{-5}e^{10S}$ *and determine* α*.*

4

American options

Unlike European options, holder of an American options has the right but not the obligation to exercise any date before or at the maturity. When she chooses to exercise the option at time $t \in [0,T]$ (or in discrete-time models $t \in \{0,...,T\}$), she receives the payoff of the option $g(t, S_t)$; i.e., the payoff is a function $g(t, S)$ of the time of exercise t and the price of the underlying S at the time of exercise. For instance, the payoff of an American call option if exercises at time t is $(S_t - K)_+$.

To price an American option, we assume that the holder chooses to exercise the American option at the optimal time of exercise; the time that the holder receives the largest value possible. We will discuss the details of the optimal exercise time in the future.

We use $C_{\text{Am}}(T, K, S, t)$ and $P_{\text{Am}}(T, K, S, t)$ to denote the price of American call and American put at time t when the underlying asset price is S with maturity T and strike K, respectively. The following remarks are very important in our future study of American option.

Remark 4.0.1. *Notice that since exercising is always an option, the value of the American option never falls below the payoff. In fact, is can sometimes be strictly larger than the payoff.*

Remark 4.0.2. *Consider an American option with payoff $g(t, S)$ which can possibly take negative values. Since the holder has no obligation to exercise the option, she will not do so as long as $g(t, S_t) \leqslant 0$. In particular, if the American option has not been exercised before the maturity T, it is not optimal for her to exercise at T if $S(T, S_T) \leqslant 0$. Therefore, the actual payoff is $g_+(t, S) = \max\{0, g(t, S)\}$. On the other hand, if the payoff $g(T, S_T)$ is positive at maturity, it is always optimal to exercise it. Therefore, we always assume that the payoff of an American option is nonnegative.*

Remark 4.0.3. *The price p of an American option with payoff $g_1(t, S) + g_2(t, S)$ is not the same as the sum of the price p_1 of an American option with payoff $g_1(t, S)$ and price p_2 of an American option with payoff $g_2(t, S)$. In fact, the first value is always smaller than* *or equal to the sum of the other two. To see this, let $\tilde{t}$ be the optimal exercise time for the American option with payoff $g_1(t,S)+g_2(t,S)$, i.e., the present value of the option is*

$$p = B_0(\tilde{t})(g_1(\tilde{t}, S_{\tilde{t}}) + g_2(\tilde{t}, S_{\tilde{t}})).$$

Notice that time $\tilde{t}$ is not necessarily optimal for the other two American options with payoff $g_1(t,S)$ or $g_2(t,S)$. Therefore,

$$p_i \geqslant B_0(\tilde{t}) g_i(\tilde{t}, S_{\tilde{t}}), \quad \text{for} \quad i = 1, 2.$$

This implies that $p \leqslant p_1 + p_2$.

4.0.1 Pricing American option in the binomial model via examples

The key to the pricing of American options is to compare to values at each node of the model, i.e., *continuation value* and *exercise value*. We will properly define these values in this section and use them to price American option. We first present the pricing method in the naive case of one-period binomial model.

Example 4.0.1. *Consider a one-period binomial model with $S_0 = 1$, $u = 2.1$, $\ell = .6$ and $R = .1$ (for simplicity). We consider an American put option with strike K; the payoff is $(K-S)_+$. Similar to the European put, at the terminal time $T = 1$, the value of the option is known. However, at time $t = 0$ we are facing a different situation; we can choose to exercise and get the exercise value of $E := (K-1)_+$, or we can continue. If we continue, we will have a payoff of $(K-2.1)_+$ or $(K-.6)_+$ depending of the future events. The value of continuation is obtain via taking risk-neutral expectation:*

$$C := \frac{1}{1+R}\hat{\mathbb{E}}[(K-S_T)_+] = \frac{1}{1.1}\left(\frac{1}{3}(K-2.1)_+ + \frac{2}{3}(K-.6)_+\right).$$

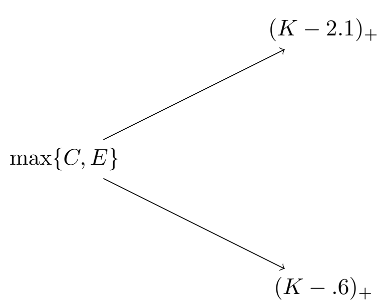


Figure 4.0.1: American put option in one-period binomial model

Next, we should compare continuation value C and exercise value E. If $C > E$, we should not exercise at time 0 and should wait until time 1. Otherwise when $C \leqslant E$, it is optimal *to exercise. For example when* $K = 3$*, we have* $C = 1.7273 < E = 2$ *and thus we exercise. However, when* $K = 1$*, we have* $C = .2424 > E = 0$ *and thus we do not exercise.*

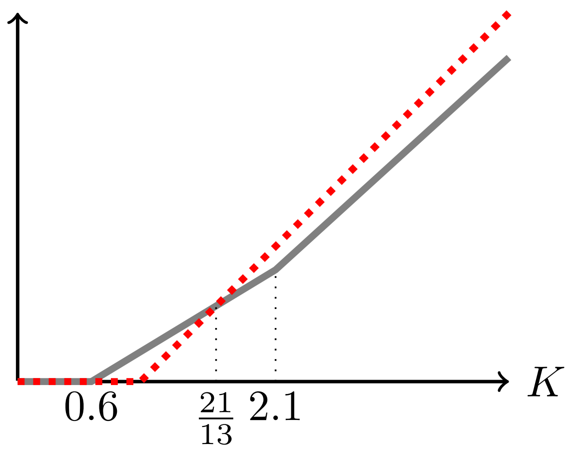


Figure 4.0.2: One-period binomial model: continuation value $C = \frac{1}{1.1}\left(\frac{1}{3}(K-2.1)_+ + \frac{2}{3}(K-.6)_+\right)$ (gray) and exercise value $E = (K-1)_+$ (red) of an American put option as a function of strike price K.

As seen in Figure 4.0.2, only American put options with strike K *in* $(.6, \frac{21}{13})$ *generate a larger continuation value and therefore should not be exercises at time* 0*.*

To illustrate more, we consider a two period binomial model in the following example.

Example 4.0.2. *Consider a two-period binomial model with the same parameters as in Exercise 4.0.1, i.e.,* $u = 2.1$*,* $\ell = .6$ *and* $R = .1$*. We consider an American put option with strike* K*; the payoff is* $(K-S)_+$*. The terminal time* $T = 2$*, the value of the option is known. At time* $t = 1$*, there are two nodes and at time* $t = 0$ *there is one node, at each of which we have to find exercise value and continuation value. The exercise values are given by the payoff as seen in Figure 4.0.3*

Here C_0*,* $C_{1,1}$*, and* $C_{1,2}$ *are the continuation values at nodes* $t = 0$*,* $t = 1$ *price-down, and* $t = 1$ *price-up, respectively. Next, we should compare continuation value and exercise value at each node in a backward manner. At time* $t = 1$*, the continuation value is*

$$C_{1,2} = \frac{1}{1+R}\hat{\mathbb{E}}[(K-S_T)_+ \mid S_1 = 2.1] = \frac{1}{1.1}\left(\frac{1}{3}(K-4.41)_+ + \frac{2}{3}(K-1.26)_+\right)$$
$$C_{1,1} = \frac{1}{1+R}\hat{\mathbb{E}}[(K-S_T)_+ \mid S_1 = .6] = \frac{1}{1.1}\left(\frac{1}{3}(K-1.26)_+ + \frac{2}{3}(K-.36)_+\right)$$

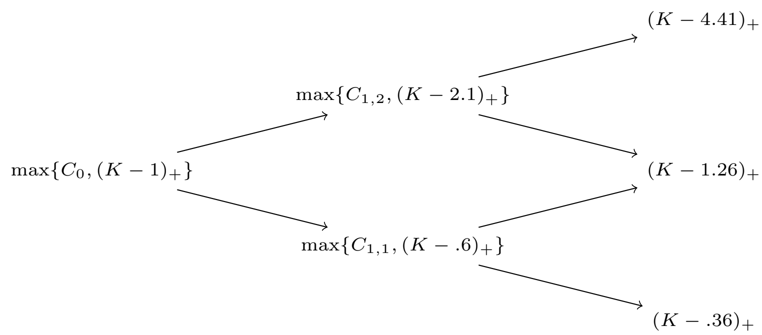


Figure 4.0.3: American put option in two-period binomial model

Therefore, the price of the option at these nodes are

$$V_{1,2} = \max\left\{\frac{1}{1.1}\left(\frac{1}{3}(K-4.41)_+ + \frac{2}{3}(K-1.26)_+\right), (K-2.1)_+\right\}$$
$$V_{1,1} = \max\left\{\frac{1}{1.1}\left(\frac{1}{3}(K-1.26)_+ + \frac{2}{3}(K-.36)_+\right), (K-.6)_+\right\}$$

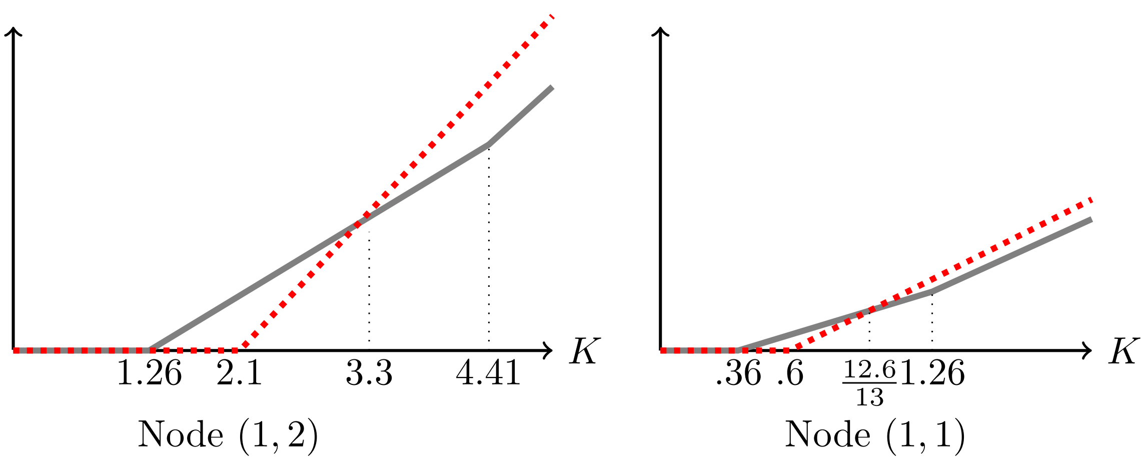


Figure 4.0.4: Two-period binomial model: at each nodes at time $t = 1$, continuation value (gray) is compared to exercise value (red) as a function of strike price K.

As you see from Figure 4.0.4, if we choose the strike of the option K in $(1.26, 3.3)$, then at node $(1,2)$ the continuation value is larger, otherwise we exercise at this node. If we choose K in $(.36, \frac{12.6}{13})$, at node $(1,1)$ the continuation value is larger.In this example, values of K that imposes the continuation in node $(1,1)$ is disjoint from those which imposed continuation at node $(1,2)$. Therefore, we exercise the American option on at least one of *these nodes.*

Now let fix our put option by choosing $K = 2$. *Then, we continue at node* $(1,2)$ *and exercise at node* $(1,1)$. *Therefore, at time* $t = 1$, *the option takes values*

$$V_{1,2} = C_{1,2} \approx 0.4485 \quad \text{and} \quad V_{1,1} = 1.4.$$

At time $t = 0$, *we need to see if it is optimal to exercise or if it is optimal to continue. The situation at node* $(0,0)$ *is explained in Figure 4.0.5. The exercise value is* $E = 1$ *but the continuation value is positive, i.e.,* $C = .9844$. *Therefore, it is optimal to exercise.*

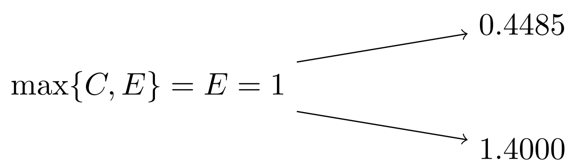


Figure 4.0.5: American put option at time $t = 0$ in two-period binomial model of Example 4.0.2.

Remark 4.0.4. *As a general rule, when it is optimal to exercise, i.e., the continuation value is less than or equal to exercise value, there is no point in continuing the option. This is because the continuation value remains equal to exercise value since then after. This can be observed from Example 4.0.2 by taking for example* $K = 1$. *The schematic pattern of exercise and continuation nodes is presented in Figure 4.0.6.*

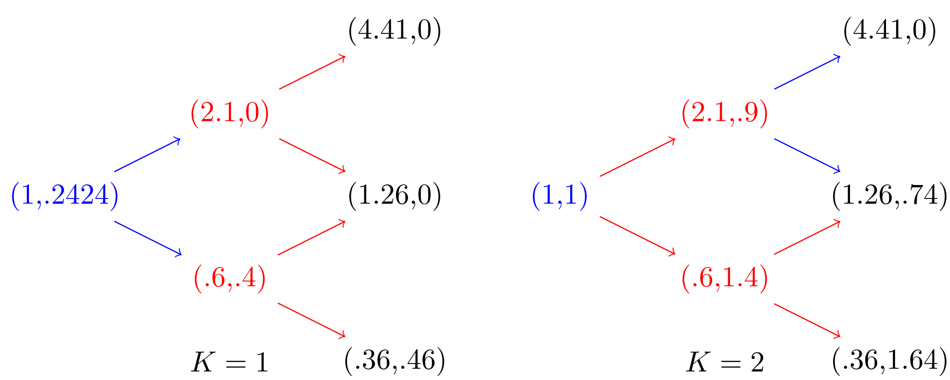


Figure 4.0.6: The pattern of continuation versus exercise in a two-period binomial model in Example 4.0.2. The red nodes are the exercise nodes and blue nodes are continuation nodes. In the pair (a, b), a is the asset price and b is the continuation value.

Exercise 4.0.1. *In Example 4.0.2, take the following values of* K.

1) $K = 1$.

2) $K = 3$.

4.1 Pricing American option in the binomial model; problem formulation

In this section, we formulate the problem of pricing an American option with (nonnegative) payoff $g(t, S)$ in T-period binomial model. Notice that to price an American option, it is important to know the best exercise time for the holder of the option. Since the price of the underlying is randomly changing over time, the best execution time can also be a random time. To formulate an exercise strategy as a random time, we need to define the notion of *stopping time*. Recalling from Definition B.19, we define an exercise strategy for the multi period binomial model.

Definition 4.1.1. *An exercise strategy for an American option is a stopping time with respect to price of underlying* $\{S_t\ :\ t = 0, ...\}$*, i.e., is a random variables* $\tau : \Omega \to \{0, ..., T\}$ *such that for any* $t = 0, ..., T$*, the even* $\tau \leqslant t$ *is known given the values of* S_u *for* $u \leqslant t$*.*

In other words, exercise strategy is a stopping time with respect to the information generated by the price process. Given the holder of an American option chooses a specific stopping time τ, the corresponding value of the option is obtained through

$$\hat{\mathbb{E}}\Big[\frac{1}{(1+R)^{\tau}}g(\tau, S_\tau)\Big].$$

Therefore, the (optimal) value of the option for the holder is given by the maximum value over all exercise strategies;

$$V_0 := \max_{\tau\in\mathcal{T}_0}\hat{\mathbb{E}}\Big[\frac{1}{(1+R)^{\tau}}g(\tau, S_\tau)\Big],$$

where the maximum is taken over the set $\mathcal{T}_0$ all stopping times τ with values in $\{0, ..., T\}$. Since the binomial model can be constructed on a finite sample space (the set of all price process paths is finite), then the set $\mathcal{T}_0$ is finite and therefore in the above e can show the value of the American option with a maximum. More generally, given the option has not been exercised until time t, then the value of the option is given

$$V_t := \max_{\tau\in\mathcal{T}_t}\hat{\mathbb{E}}[\frac{1}{(1+R)^{\tau-t}}g(\tau, S_\tau) \mid \mathcal{F}_t]. \tag{4.1.1}$$

Here the maximum is taken over the finite set $\mathcal{T}_t$ of all stopping times τ with values in $\{t, ..., T\}$. Since $\mathcal{T}_0$ is finite, there exists an optimal exercise strategy (stopping time) ϱ^*

with values in $\{0, ..., T\}$ such that

$$V_0 = \hat{\mathbb{E}}\Big[\frac{1}{(1+R)^{\varrho^*}} g(\varrho^*, S_{\varrho^*})\Big]$$

Similarly at time t, there exists an optimal a stopping time ϱ_t^* with values in $\{t, ..., T\}$ such that

$$V_t = \hat{\mathbb{E}}\Big[\frac{1}{(1+R)^{\varrho_t^* - t}} g(\varrho_t^*, S_{\varrho_t^*}) \mid \mathcal{F}_t\Big]$$

We explain the methodology for pricing American option in the simple one-period case by choosing $t = T-1$, i.e.

$$V_{T-1} := \max_{\tau \in \mathcal{T}_{T-1}} \hat{\mathbb{E}}[\frac{1}{(1+R)^{\tau - T+1}} g(\tau, S_\tau) \mid \mathcal{F}_{T-1}],$$

where the maximum is taken over all stopping times τ with two values $T-1$ and T; the former corresponds to decision to exercise at time $T-1$, while the latter is to continue untill maturity T. Given $\mathcal{F}_{T-1}$, for any stopping time τ with two values $\{T-1, T\}$, the events $\{\tau = T-1\} = \{\tau \leqslant T-1\}$ and $\{\tau = T\} = \{\tau \leqslant T-1\}^\complement$ are both known. Therefore, either $\tau = T-1$ a.s. or $\tau = T$ a.s.. When $\tau = T-1$ we obtain the exercise value $g(T-1, S_{T-1})$; otherwise $\tau = T$ yields the continuation value $\frac{1}{1+R}\hat{\mathbb{E}}[g(T, S_T) \mid \mathcal{F}_{T-1}]$. Therefore, the choice of exercise strategy boils down to choosing the maximum of the exercise value $g(T-1, S_{T-1})$ and the continuation value $\frac{1}{1+R}\hat{\mathbb{E}}[g(T, S_T) \mid \mathcal{F}_{T-1}]$, i.e.

$$V_{T-1} = \max\Big\{\frac{1}{1+R}\hat{\mathbb{E}}[g(T, S_T) \mid \mathcal{F}_{T-1}], g(T-1, S_{T-1})\Big\}.$$

By using the Markov property of the binomial model, we have $\hat{\mathbb{E}}[g(T, S_T) \mid \mathcal{F}_{T-1}] = \hat{\mathbb{E}}[g(T, S_T) \mid S_{T-1}]$ and thus the value of the American option is a function of time $T-1$ and asset price S_{T-1}, i.e.

$$V(T-1, S_{T-1}) = \max\Big\{\frac{1}{1+R}\hat{\mathbb{E}}[g(T, S_T) \mid S_{T-1}], g(T-1, S_{T-1})\Big\}.$$

In this one period situation, it is not hard to see that the optimal stopping time is obtained by the first time t (among $T-1$, and T) that the exercise value $g(t, S_t)$ is not smaller than continuation value $\frac{1}{1+R}\hat{\mathbb{E}}[g(T, S_T) \mid S_{T-1}]$. Notice that from Remark 4.0.2 the value of the American option at maturity time T is always equal $g(T, S_T)$ and

$$V(T-1, S_{T-1}) = \max\Big\{\frac{1}{1+R}\hat{\mathbb{E}}[V_T \mid S_{T-1}], g(T-1, S_{T-1})\Big\}.$$

In the sequel, we would like to extend the above implication to all t. More precisely, we seek to show the following theorem.

Theorem 4.1.1. *The value of the American option with payoff function $g(t, S)$ is a function $V(t, S)$ of time t and the asset price S which satisfies*

$$V(t, S_t) = \max\Big\{\frac{1}{1+R}\hat{\mathbb{E}}[V(t+1, S_{t+1}) \mid S_t], g(t, S_t)\Big\} \tag{4.1.2}$$

Here the supremum is over all stopping times with values in $\{t, ..., T\}$. In addition, we show that an optimal exercise strategy is given by the stopping time τ_t^* defined by the first time $u \geqslant t$ such that $V(u, S_u) = g(u, S_u)$, i.e.,

$$\tau_t^* = \inf\{u \,:\, u = t, ..., T, \quad \textit{such that } V(u, S_u) = g(u, S_u)\}. \tag{4.1.3}$$

Recall that from Remark 4.0.1, we already know that $V_t \geqslant g(t, S_t)$. Therefore (4.1.3) representation for an optimal exercise implies that $V(t, S_t) > g(t, S_t)$ if and only if $\tau_t^* > t$.

To prove Theorem 4.1.1, we define a new process $\tilde{V}$ by

$$\begin{cases} \tilde{V}_T := g(T, S_T) \\ \tilde{V}_t := \max\big\{\frac{1}{1+R}\hat{\mathbb{E}}[\tilde{V}_{t+1} \mid \mathcal{F}_t], g(t, S_t)\big\} \qquad t = 0, ..., T-1 \end{cases}. \tag{4.1.4}$$

The following lemma presents some key properties for the study of an American option.

Lemma 4.1.1. *The following properties hold for V_t and $\tilde{V}_t$ defined by 4.1.1 and 4.1.4, respectively.*

i) Both $\frac{V_t}{(1+R)^t}$ and $\frac{\tilde{V}_t}{(1+R)^t}$ are supermartingales with respect to filtration generated by the asset price.

ii) V_t and $\tilde{V}_t$ are greater than or equal to $g(t, S_t)$.

iii) V_t is the smallest process with property (i) and (ii).

iv) $\Big\{\frac{V_{u\wedge\tau_t^*}}{(1+R)^{u\wedge\tau_t^*}} \,:\, u = t, ..., T\Big\}$ is a martingale.

Property (i) asserts that the discounted price of an American option is a supermartingale and property (ii) simply restates that the value of the American option is never smaller than the payoff. Property (iii), which is a crucial property of the value of American option, indicates that the discounted values of the American option is the smallest supermartingale which is greater than or equal to the discounted payoff.

Proof of Lemma 4.1.1. Given $\mathcal{F}_t$, for any stopping time τ with values in $\{t, ..., T\}$, the occurrence of the event $\{\tau \leqslant t\} = \{\tau = t\}$ is known. Therefore,

- If $\tau = t$ has happened then strategy τ suggests to exercise and

$$\hat{\mathbb{E}}\Big[\frac{1}{(1+R)^{\tau-t}}g(\tau, S_\tau) \mid \mathcal{F}_t\Big] = g(t, S_t).$$

- If $\tau = t$ has not happened, then strategy τ suggests to continue.

Therefore,

$$V_t = \max_{\tau\in\mathcal{T}_t} \hat{\mathbb{E}}[\frac{1}{(1+R)^{\tau-t}} g(\tau, S_\tau) \mid \mathcal{F}_t] = \max\big\{\max_{\tau\in\mathcal{T}_{t+1}} \hat{\mathbb{E}}[\frac{1}{(1+R)^{\tau-t}} g(\tau, S_\tau) \mid \mathcal{F}_t], g(t, S_t)\big\}$$

$$\geqslant \frac{1}{1+R} \max_{\tau\in\mathcal{T}_{t+1}} \hat{\mathbb{E}}[\frac{1}{(1+R)^{\tau-t-1}} g(\tau, S_\tau) \mid \mathcal{F}_t]. \tag{4.1.5}$$

Let ϱ^*_{t+1} be an optimal exercise for V_{t+1}, i.e.

$$V_{t+1} = \hat{\mathbb{E}}\Big[\frac{1}{(1+R)^{\varrho^*_{t+1}-t-1}} g(\varrho^*_{t+1}, S_{\varrho^*_{t+1}})\Big].$$

Then, in particular we have

$$V_t \geqslant \hat{\mathbb{E}}[\frac{1}{(1+R)^{\varrho^*_{t+1}-t}} g(\varrho^*_{t+1}, S_{\varrho^*_{t+1}}) \mid \mathcal{F}_t] = \frac{1}{1+R}\hat{\mathbb{E}}[V_{t+1} \mid \mathcal{F}_t].$$

Therefore, $\frac{V_t}{(1+R)^t}$ is a supermartingale. The supermartingale property for $\frac{\tilde{V}_t}{(1+R)^t}$ is a direct consequence of its definition (4.1.4), i.e.

$$\tilde{V}_t = \max\big\{\frac{1}{1+R}\hat{\mathbb{E}}[\tilde{V}_{t+1} \mid \mathcal{F}_t], g(t, S_t)\big\} \geqslant \frac{1}{1+R}\hat{\mathbb{E}}[\tilde{V}_{t+1} \mid \mathcal{F}_t].$$

Hence, (i) is proven. (ii) is a straightforward consequence of choosing $\tau \equiv t$ in (4.1.1) and (4.1.4).

To show (iii), let Y_t be another process which satisfies (i)-(ii). In particular, $Y_T \geqslant g(T, S_T) = V_T$. Inductively, assume that $Y_u \geqslant V_u$ for $u > t$. If $V_t = g(t, S_t)$, then property (ii) for Y_t implies $V_t = g(t, S_t) \leqslant Y_t$. Otherwise, $V_t > g(t, S_t)$ and we deduce from (4.1.5) that

$$V_t = \max_{\tau\in\mathcal{T}_{t+1}} \hat{\mathbb{E}}[\frac{1}{(1+R)^{\tau-t}} g(\tau, S_\tau) \mid \mathcal{F}_t].$$

Since Y_t satisfies (ii), we have

$$V_t \leqslant \max_{\tau\in\mathcal{T}_{t+1}} \hat{\mathbb{E}}[\frac{1}{(1+R)^{\tau-t}} Y_\tau \mid \mathcal{F}_t].$$

By optional sampling theorem, Theorem B.6, we obtain

$$\hat{\mathbb{E}}[\frac{1}{(1+R)^{\tau-t}} Y_\tau \mid \mathcal{F}_t] \leqslant Y_t.$$

Hence, $V_t \leqslant Y_t$. □

Remark 4.1.1 (Snell envelope)**.** *For any given payoff process* $g := \{g_t\}_{t \geqslant 0}$*, for instance* $g_t = \frac{g(t,S)}{(1+R)^t}$*, the smallest supermartingale which is greater than or equal to* g *is called* ***Snell envelope*** *of* g*. Therefore, the discounted value of the American option is the Snell envelope of the discounted payoff.*

First, we use the properties (i)-(iii) in Lemma 4.1.1 to show that $V_t = \tilde{V}_t$ for all t.

Lemma 4.1.2. *For any* $t = 0, ..., T$*, we have* $V_t = \tilde{V}_t$*.*

Proof. Obviously, $V_T = g(T, S_T) = \tilde{V}_T$. If $V_{t+1} = \tilde{V}_{t+1}$, then

$$\tilde{V}_t = \max\Big\{\frac{1}{1+R}\hat{\mathbb{E}}[\tilde{V}_{t+1} \mid \mathcal{F}_t], g(t, S_t)\Big\} = \max\Big\{\frac{1}{1+R}\hat{\mathbb{E}}[V_{t+1} \mid \mathcal{F}_t], g(t, S_t)\Big\}.$$

By (i) (supermartingale property of discounted V) and (ii), we have

$$\tilde{V}_t \leqslant \max\big\{V_t, g(t, S_t)\big\} = V_t.$$

On the other hand, since V_t is the smallest process with properties (i) and (ii), we must have $V_t \leqslant \tilde{V}_t$, which verifies the equality. □

Next lemma is on the Markovian property of the price of American option.

Lemma 4.1.3. *The value of the American option* V_t *is a function of time* t *and asset price* S*, i.e.,* $V_t = V(t, S_t)$*.*

Proof. If $V_{t+1} = V(t+1, S_{t+1})$, then by Markovian property of binomial model, we obtain $\hat{\mathbb{E}}[V(t+1, S_{t+1}) \mid \mathcal{F}_t] = \hat{\mathbb{E}}[V(t+1, S_{t+1}) \mid S_t]$ and

$$V_t = \max\Big\{\frac{1}{1+R}\hat{\mathbb{E}}[V(t+1, S_{t+1}) \mid S_t], g(t, S_t)\Big\}$$

is a function $V(t, S_t)$ of time and the asset price. Hence, (4.1.2) follows. □

Finally, we show that (4.1.3) provides an optimal exercise.

Lemma 4.1.4. *The stopping time* τ_t^* *given by 4.1.3 satisfies*

$$V_t = \hat{\mathbb{E}}\Big[\frac{1}{(1+R)^{\tau_t^*-t}} g(\tau_t^*, S_{\tau_t^*}) \mid \mathcal{F}_t\Big]$$

Proof. Notice that by definition, $V_{\tau_t^*} = g(\tau_t^*, S_{\tau_t^*})$. Hence, by the optional sampling theorem, Theorem (B.6), and property (iv), we obtain that

$$V_t = V_{t \wedge \tau_t^*} = \hat{\mathbb{E}}\Big[\frac{1}{(1+R)^{\tau_t^*-t}} V(\tau_t^*, S_{\tau_t^*}) \mid \mathcal{F}_t\Big] = \hat{\mathbb{E}}\Big[\frac{1}{(1+R)^{\tau_t^*-t}} g(\tau_t^*, S_{\tau^*}) \mid \mathcal{F}_t\Big],$$

which shows that τ_t^* is optimal. □

The above discussion suggests the following algorithm in pricing American options.

Pricing American options in the binomial model

At time T, the value of the option is $g(T, S_T(j))$.
for each $t = T-1, ..., 0$ **do**
  for each $j = 1, ..., t+1$ **do**
Exercise value $= g(t, S_t(j))$.
Continuation value $= \frac{1}{1+R}\hat{\mathbb{E}}[V(t+1, S_{t+1}) \mid S_t = S_t(i)] = \frac{1}{1+R}\left(V_{t+1}^{i+1}\hat{\pi}_u + V_{t+1}^{i}\hat{\pi}_\ell\right)$.
The value of the option $V(t, S_t(i)) = \max\left\{g(t, S_t(i)), \frac{1}{1+R}\hat{\mathbb{E}}[V(t+1, S_{t+1}) \mid S_t = S_t(i)]\right\}$.
If $V(t, S_t(i)) = g(t, S_t(i))$, we exercise the option and stop.
If $V(t, S_t(i)) > g(t, S_t(i))$, we continue and the replicating portfolio is given by $\Delta_t(S_t(j))$ units of risky assets and $V(t, S_t(i)) - \Delta_t(S_t(j))S_t(j)$ is cash.
  end for
end for

Remark 4.1.2 (Path dependent American option)**.** *If the payoff of the American option depends on the path, one can adjust the algorithm by considering the path dependent continuation value and path dependent exercise value. Exercise 4.1.1 provides an example of such kind.*

Example 4.1.1 (American call option on a nondividend asset does not exist!)**.** *In this case the payoff is given by $g(t, S) = (S - K)_+$. Then, the price of the American call option is the same as the price of European call option! In fact, this is true in any model where the pricing of European claims is carried by the risk-neutral probability. See Proposition 4.1.1.*

This situation is not unique to American call option. The following proposition further elaborate on this matter.

Proposition 4.1.1. *Consider an American option with a convex payoff $g(S)$ such that $g(0) = 0$ on an asset which pays no dividend. Assume that the price of a European contingent claim is given by*

$$V^{Eu}(t, S) = B_t(T)\hat{\mathbb{E}}[g(S_T) \mid S_t = S],$$

where $\hat{\mathbb{E}}$ is expectation under risk-neutral probability, and the discounted asset price $\tilde{S}_t = \frac{S_t}{(1+R)^t}$ is a martingale under risk-neutral probability. Then, the price of American option $V^{Am}(t, S)$ with payoff $g(S)$ is the same as V^{Eu}, i.e.

$$V^{Eu}(t, S) = V^{Am}(t, S).$$

Remark 4.1.3. *Proposition 4.1.1 is only true when the underlying asset does not pay dividend. See Exercise 4.1.2.*

Exercise 4.1.1. *Consider a two-period binomial model for a risky asset with each period equal to a year and take $S_0 = \$1$, $u = 1.5$ and $\ell = 0.6$. The interest rate for both periods is $R = .1$.*

a) Price an American put option with strike $K = .8$.

b) Price an American call option with strike $K = .8$.

c) Price an American option with a path dependent payoff which pays the running maximum[1] of the path.

Remark 4.1.4. *A naive example of an American option is the case where the payoff is \$1. In this case, if the interest rate is positive, it is optimal to exercise the option right away. However, if interest rate is 0, the time of exercise can be any time. This naive example has an important implication. If interest rate is zero, one can remove condition $g(0) = 0$ from Proposition 4.1.1, simply by replacing payoff g by $\tilde{g}(S) = g(S) - g(0)$. Since cash value of $g(0)$ does not change value over time, the value of the American option with payoff $g(S)$ is $g(0)$ plus the value of American option with payoff $\tilde{g}(S)$.*
For negative interest rate, the exercise date will be postponed compared to the positive interest rate.
For example, if the interest rate is zero, the price of American put is equal to the price of European put.

Hedging American option in the binomial model

Hedging American option in the binomial model follows the same way as European option. The only difference is that the hedging may not continue until maturity because of the early exercise. Given that we know the price $V(t+1, S_{t+1}(i))$ of the American option at time $t+1$ at all states $i = 1, ..., t+2$, to hedge at time t and state j, we need to keep $\Delta_t(S_t(j))$ units of risky asset in the replicating portfolio and $V(t, S_t(i)) - \Delta_t(S_t(j))S_t(j)$ in cash, where $\Delta_t(S)$ is given by (2.3.4), i.e.

$$\Delta(t, S) := \frac{V(t+1, Su) - V(t+1, S\ell)}{S(u-\ell)} \quad \text{for } t < \tau^*.$$

Notice that hedging an American contingent claim is only matters before the time of the exercise. At the time of the exercise or thereafter, there is no need to hedge. However, if the holder of the American claim decides not to exercise at time τ^*, the issuer can continue hedging with no hassle. For example, in Example 4.0.2 with $K = 2$, it is optimal for the

[1] The running maximum at time t is the maximum of the price until or at time t.

holder to exercise the option at the beginning, where the price of asset is \$ 1. Therefore, the issuer need \$ 1 to replicate. However, if the holder continues, the replication problem in the next period leads to solving the following system of equation.

$$\begin{cases} 2.1a + 1.1b & = .4485 \\ .6a + 1.1b & = 1.4000 \end{cases}$$

which yields $a = -1.903$ and $b = 2.3107$. Therefore, the issuer needs \$ $a + b = .4077$ to replicate the option which is less than \$ 1 if the holder exercises the option.

Exercise 4.1.2. *Consider a two-period binomial model for a risky asset with each period equal to a year and take $S_0 = \$1$, $u = 1.2$ and $\ell = 0.8$. The interest rate for both periods is $R = .05$.*

a) If the asset pays 10% divided yield in the first period and 5% in the second period, find the price of an American and European call options with strike $K = .8$.

b) Construct the replicating portfolio for both American and European call option.

4.2 Pricing American option in the Black-Scholes model

In continuous time, including Black-Scholes model, the definition of exercise policy (stopping time) for American options is not easy to define. The definition needs to use filtration and σ-algebra from measure theory[2]. Here we avoid technical discussion of stopping times and only present the solution for American option in the Black-Scholes model.

First notice that Proposition 4.1.1 implies that the American call option is Black-Scholes model has the same price as European call if the underlying does not pay any dividend. Therefore, our focus here is on the derivatives such as American put or American call on a dividend-paying asset.

The key to solve the American option problem in the Black-Scholes model is to set up a *free boundary problem*. This type of problems are widely studied in physics. For example, in order to understand how an ice cube is melting over time, we need to solve a free boundary problem. Or if we push an elastic object to a certain shape, after releasing, the shape starts changing in a certain way which can be realized by solving a free boundary problem. See Figure 4.2.1.

Another simple example occurs if we add obstacle underneath a hanging elastic rubber. An elastic rubber fixed at two level points takes shape as a piece of parabola shown in left image in Figure 4.2.2. The equation satisfied by the free rubber is $u'' = c$ where the constant c depends on the physical properties of the rubber. If we position an obstacle underneath the rubber such that the elastic rubber is touched, then it changes the shape of the red

[2] For more information see for example [23, 31] or for a more advanced text see [17, Chapter 1].

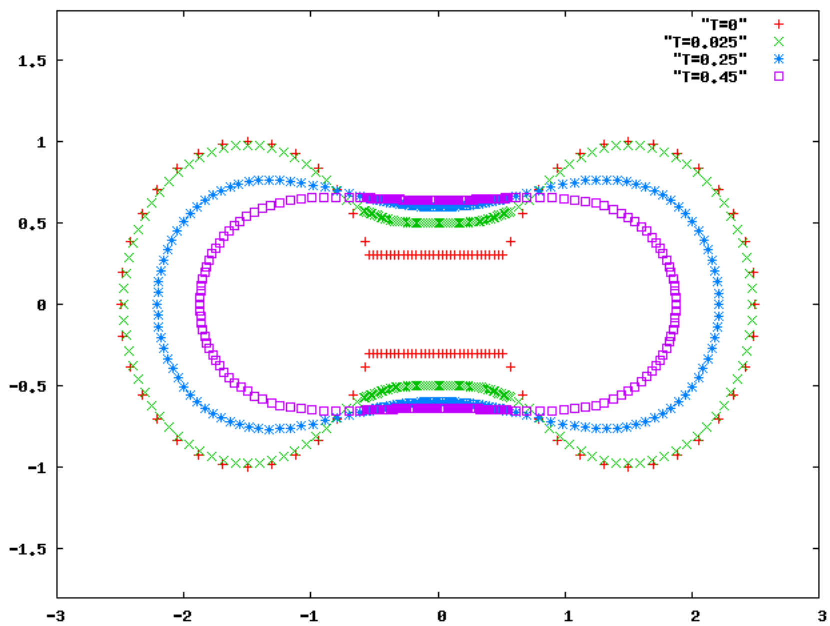


Figure 4.2.1: Exterior membrane of a tube (purple shape) is forcefully shaped into a dumbbell (red shape). Upon release of the forces, the surface of the shape starts moving; each point moves at a speed proportional to the curvature of the surface. Eventually, it returns back to the original shape. The picture is adopted from [12].

curve shown in right image in Figure 4.2.2. The position of the red curve satisfies the same equation $u'' = c$ but only at the points where the rubber in not touching the obstacle. Inside the touching region (x_0, x_1), the rubber takes the shape of the obstacle. At the two endpoints of touching region, the shape of the rubber is once continuously differentiable; i.e., if the $u(x_0) = u(x_1) = y_0$ the height of the obstacle and $u'(x_0) = u'(x_1) = 0$.

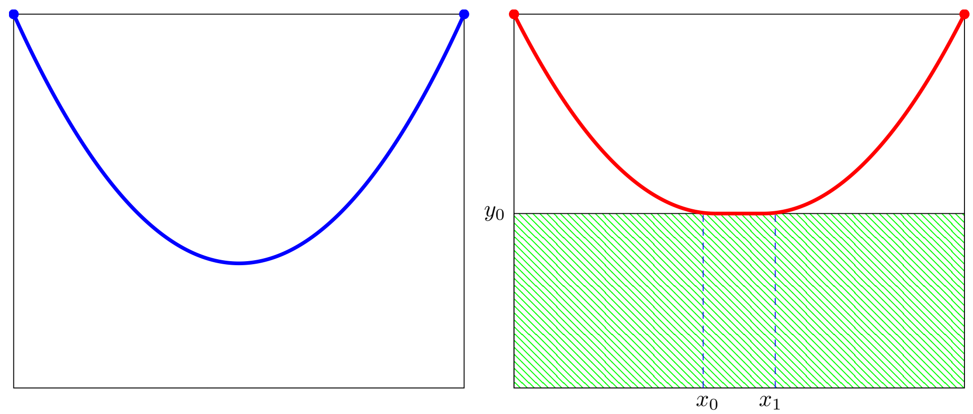


Figure 4.2.2: The blue curve on the left shows the position of a free elastic hanging at two points. The red curve shows the same elastic rubber hanging at the same points but bounded below by an obstacle.

After this short introduction, we come back to the problem of pricing and hedging American option. For the sake of simplicity, we only consider simple case where there is only one free boundary. This for example occurs when we have an American put or an American call on a dividend-paying asset. Below, we list the important facts that you need to know about the free boundary.

a) The domain for the problem $\{(t, S) : t \in [0, T], \quad \text{and} \quad S \in [0, \infty)\}$ is divided into to parts separated by a curve $C := \{(t, S^*(t)) : t \in [0, T]\}$. The curve C is called *free boundary.*

b) On one side of the boundary it is not optimal to exercise the option. This side is called *continuation region*, e.g. if the asset price $S_s > S^*(s)$ for all $s \leqslant t$, it has never been optimal to exercise the option before or at time t.

c) The other side of the boundary is called *exercise region*, e.g. if the asset price $S_t \leqslant S^*(t)$, then it is optimal to exercise the option at time t. Therefore, the optimal stopping time is the first time that the pair (t, S_t) hits the exercise region. More precisely, the asset price S_t hits the free boundary at time t at point $S^*(t)$.

$$\tau^* := \inf \{t \leqslant T : S_t = S^*(t)\}.$$

It is important to notice that the exercise boundary is an unknown in pricing American options in continuous time. The other unknown is the price of the American option. We next explain that finding the price of the American option also gives us the free boundary. The relation between these two is lied in the following representation of the stopping policy. For an American option with payoff $g(t, S_t)$, we have the optimal stopping τ^* given by

$$\tau^* = \inf \{t \leqslant T \,:\, V(t, S_t) = g(t, S_t)\}\,.$$

See [24] for more details. Therefore, finding the price of American option is the first priority here, which will be explained in the sequel.

Finding the edges of exercise boundary

So far, we learned that the domain of the problem is split into continuation boundary and exercise boundary; e.g. for American put at each time t, $(0, S^*(t))$ is the interior of exercise region and $(S^*(t), \infty)$ is the interior of the continuation region. determining continuation region is the matter of guess-and-check; we must look at the payoff of the American option to guess the topology of the continuation region. One general rule is that **it is not optimal to exercise in the out of money region or where the payoff takes its minimum value**. We consider the following examples to clarify this rule:

a) **American put.** An American put option with strike K is out-of-money if $S_t > K$, and since it is not optimal to exercise in the out-of-money region, the exercise boundary should be inside the in-the-money region, i.e., $S^*(t) \leqslant K$. See Figure 4.2.3 on the left. At the maturity $(t = T)$, $S^*(T) = K$.

b) **American call.** An American call option with strike K on a dividend-paying asset is out-of-money if $S_t < K$, and since it is not optimal to exercise in the out-of-money region, the exercise boundary should be inside the in-the-money region, i.e., $S^*(t) \geqslant K$. See Figure 4.2.3 on the left. At the maturity $(t = T)$, $S^*(T) = K$.

c) **Straddle.** Several other options including strangle, bull and bear spread, etc can also be argued in the similar fashion. However, we only explain it for straddle. First notice that one can shift the payoff of a straddle option by cash amount of $\frac{K}{2}$ so that the new payoff is positive. The key to analyze the straddle is that it is the least desirable to exercise the option at and around the minimum point. Therefore, we can guess that there are two free boundaries, $S_1^*(t)$ and $S_2^*(t)$, located symmetrically on two sides of the minimum point of the payoff K. At the maturity $(t = T)$, $S_1^*(T) = S^*2(T) = K$; the free boundaries collapse to K.

The position of the free boundary with respect time in the three examples is sketched in Figure 4.2.5.

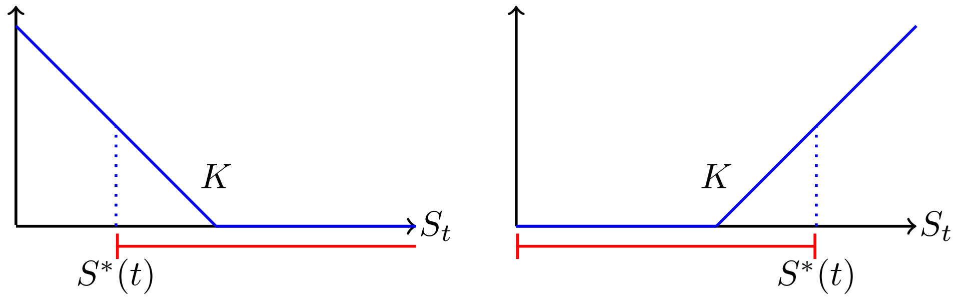


Figure 4.2.3: Left: Free boundary of American put at time t. Right: Free boundary of American call (on dividend-paying asset) at time t. The continuation region in marked with |—|.

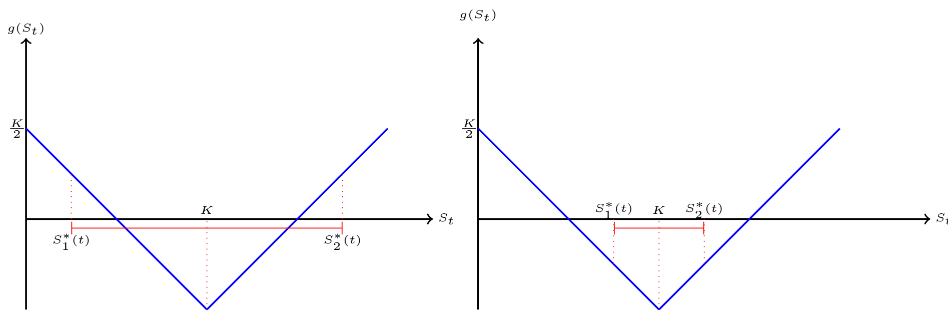


Figure 4.2.4: Two free boundaries of straddle.The continuation region in marked with |—|. The free boundaries can occur both on the positive regions of the payoff and on the negative region of the payoff. If the time-to-maturity is long, we guess that the free boundaries are wides apart and as time-to-maturity decreases, they get closer.

Remark 4.2.1. *The price of an American option with payoff equal to $g_1(S) + g_2(S)$ is not equal to, but only smaller than, the summation of prices of an American option with payoff $g_1(S)$ and an American option with payoff $g_2(S)$ Similar phenomenon is observed in Exercise 1.3.3.*

Smooth fit

The main tool in finding the free boundary and to evaluate American options in the principle of smooth fit.

Let's denote the price of the American option at time t when the asset price is equal to $S_t = S$ by $V(t, S)$. We provide the methodology for American put option. For other cases, the method can be adopted after necessary modifications. Before presenting this principle, we shall explain that in the continuation region, the price function $V(t, S)$ of the American

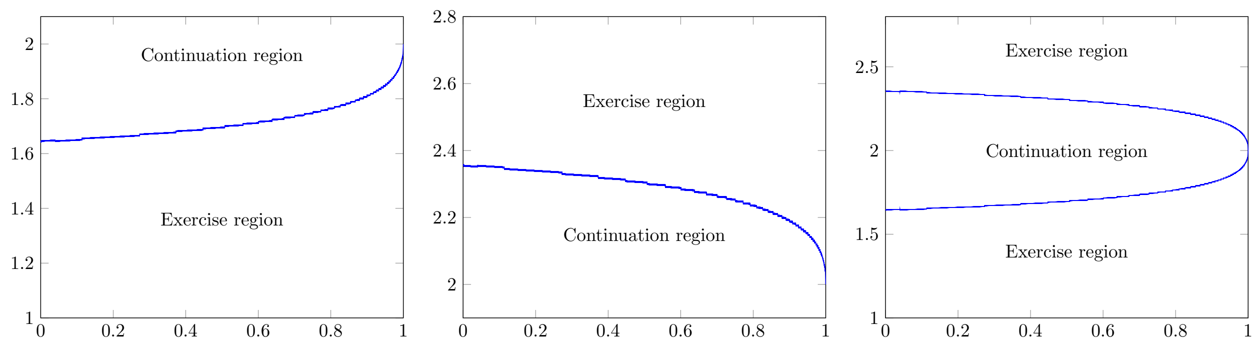


Figure 4.2.5: Sketch of the position of American option exercise boundary in time (horizontal axis). Left: American put. Middle: American call (dividend). Right: American Straddle.

put option satisfies Black-Scholes equation, i.e.

$$\begin{cases} \partial_t V(t,S) + \frac{\sigma^2 S^2}{2}\partial_{SS}V(t,S) + rS\partial_S V(t,S) - rV(t,S) = 0 & \text{for} \quad S > S^*(t) \\ V(t,S) = g(t,S) & \text{for} \quad S \leqslant S^*(t) \; . \\ V(T,S) = g(T,S); \end{cases} \tag{4.2.1}$$

Therefore, $S^*(t)$ serves as a boundary for the above equation. However, $S^*(t)$ itself is an unknown. With the Black-Scholes equation, we have two unknowns but only one equation. Principal of smooth fit provides a second equation.

Proposition 4.2.1 (Principal of smooth fit)**.** *Assume that the payoff $g(t,S)$ of the American option is twice continuously differentiable with respect to S and continuous in t. Then, at the free boundary $S^*(t)$, we have*

$$V(t, S^*(t)) = g(t, S^*(t)), \quad \textit{and} \quad \partial_S V(t, S^*(t)) = \partial_S g(t, S^*(t)),$$

for all $t < T$.

For example for American put we have $g(t,S) = (K-S)_+$. Therefore, principle of smooth fit implies that $V(t,S^*(t)) = (K - S^*(t))_+$ and $\partial_S V(t,S^*(t)) = -1$.

To see how smooth fit can be used in pricing American options, we provide some exactly solvable example in the following. These examples are called *perpetual American options* with maturity $T = \infty$. As a result, the price of the American option does not depend on time and the term $\partial_t V$ in PDE (4.2.1) vanishes. Therefore, the pricing function $V(S)$ satisfies the ODE

$$\frac{\sigma^2 S^2}{2} V''(S) + rSV'(S) - rV(S) = 0,$$

in the continuation region.

Example 4.2.1 (Perpetual American put option)**.** *A perpetual American option is an option with maturity $T = \infty$. In practice, there is no perpetual option. However, if the maturity is long (10 years), then one can approximate the price with the price of a perpetual option.*

The key observation is that since the Black-Scholes model is time-homogeneous, the free boundary of the perpetual American put option does not depend on time, i.e., $S^*(t) = S^*$ for some unknown constant $S^* < K$. On the other hand, since the time horizon is infinite, the price $V(t,S)$ of the American put does not depend on t, i.e., $\partial_t V(t,S) = 0$. Thus, we have

$$\frac{\sigma^2 S^2}{2} V'(S) + rSV'(S) - rV(S) = 0.$$

The general solution of the above equation is given by

$$V(S) = c_1 S + c_2 S^{-\frac{2r}{\sigma^2}}.$$

One can argue that c_1 must be equal to 0. Since as $S \to \infty$, the option goes deep out-of-money and becomes worthless. To find c_2, we use principle of smooth fit.

$$\begin{aligned} c_2 (S^*)^{-\frac{2r}{\sigma^2}} &= K - S^* \\ c_2 &= \frac{\sigma^2}{2r} (S^*)^{\frac{2r}{\sigma^2}+1}. \end{aligned}$$

Thus, $S^* = \frac{rK}{r+\frac{\sigma^2}{2}}$ and

$$V(S) = \begin{cases} K - S & S \geqslant \frac{rK}{r+\frac{\sigma^2}{2}} \\ \frac{\sigma^2}{2r}\left(\frac{rK}{r+\frac{\sigma^2}{2}}\right)^{\frac{2r}{\sigma^2}+1} S^{-\frac{2r}{\sigma^2}} & S < \frac{rK}{r+\frac{\sigma^2}{2}} \end{cases}.$$

Example 4.2.2 (Perpetual American call option on continuous dividend-paying asset)**.** *Consider a continuous constant dividend rate $q > 0$. The free boundary in this case is given by a constant S^* with $S^* > K$ and the price of American option satisfies*

$$\frac{\sigma^2 S^2}{2} V''(S) + (r - q)SV(S) - rV(S) = 0.$$

The general solution of the above equation is given by

$$V(S) = c_1 S^{\gamma_1} + c_2 S^{\gamma_2},$$

where γ_1 and γ_2 are roots of

$$\frac{\sigma^2}{2}\gamma^2 + \Big(r - q - \frac{\sigma^2}{2}\Big)\gamma - r = 0.$$

Notice that γ_1 and γ_2 have opposite sign and the positive one is strictly larger than 1. With out loss in generality, we assume that $\gamma_1 < 0 < 1 < \gamma_2$.

One can argue that c_1 must be equal to 0. Since as $S \to 0$, the option goes deep out-of-money and becomes worthless. To find c_2, we use principle of smooth fit.

$$\begin{aligned} c_2(S^*)^{\gamma_2} &= S^* - K \\ c_2 &= \frac{1}{\gamma_2(S^*)^{\gamma_2 - 1}}. \end{aligned}$$

Thus, $S^* = \frac{\gamma_2 K}{\gamma_2 - 1}$ and

$$V(S) = \begin{cases} S - K & S \leqslant \frac{\gamma_2 K}{\gamma_2 - 1} \\ \frac{1}{\gamma_2}(S^*)^{1-\gamma_2} S^{\gamma_2} & S > \frac{\gamma_2 K}{\gamma_2 - 1} \end{cases}.$$

Exercise 4.2.1. *Formulate and solve the free boundary problem for the perpetual American options with following payoffs.*

a) $(S - K)_+ + a$ *where* $a > 0$.

b) $(K - S)_+ + a$ *where* $a > 0$.

c) Straddle

d) Strangle

e) Bull call spread

f) Bear call spread

American option with finite maturity

Unlike perpetual American option, when $T < \infty$, there is no closed-form solution for the free boundary problem (4.2.1). Therefore, numerical methods should be used to approximate the solution.

The simplest among numerical methods is the binomial approximation. One needs to choose large number of periods N and apply the algorithm of "Pricing American options in the binomial model". The parameters of the binomial model u, ℓ , and R can be chosen according to symmetric probabilities, subjective return, or any other binomial which converges to the specific Black-Scholes model.

The finite-difference scheme can be employed to solve free boundary problem numerically. Similar to Section (3.3.10), one can apply the change of variables $U(\tau,x) = e^{-r\tau}V(T-\tau, e^{r\tau+x})$, to derive a heat equation with free boundary for U, i.e.

$$\begin{cases}\partial_\tau U(\tau,x) = \frac{\sigma^2}{2}\partial_{xx}U(\tau,x) + r\partial_x V(\tau,x) - rU(\tau,x) & \text{for} \quad x > x^*(\tau)\\ U(\tau,x) = g(T-\tau,e^x) & \text{for} \quad x \leqslant x^*(\tau)\\ U(0,x) = g(T,e^x) \end{cases},$$

where $x^*(\tau) = \ln(S^*(\tau))$.

For the simplicity, we only consider American put option, where we have $U(\tau,x) \geqslant g(T-\tau,e^x)$. Whether we want to apply explicit or implicit scheme to the above problem, because of the free boundary, we need to add an intermediate step between step i and step $i+1$ of the scheme. Suppose that the approximate solution $\hat{U}(\tau_i,x_j)$ at τ_i is known for all j. Therefore, finite-difference scheme (3.2.19) or (3.2.20) provides an approximate solution, denoted by $\hat{U}(\tau_{i+\frac{1}{2}},x_j)$ for the heat equation without including the free boundary[3]. Then, to find an approximate solution $\hat{U}(\tau_{i+1},x_j)$ for the free boundary problem at τ_{i+1}, one only needs to set

$$\hat{U}(\tau_{i+1},x_j) := \max\{\hat{U}(\tau_{i+\frac{1}{2}},x_j), g(T-\tau,e^{x_j})\}.$$

To summarize we have the following:

$$\begin{cases}\hat{U}(\tau_{i+\frac{1}{2}},\cdot) & := A\hat{U}(\tau_i,\cdot)\\ \hat{U}(\tau_{i+1},x_j) & := \max\{\hat{U}(\tau_{i+\frac{1}{2}},x_j), g(T-\tau,e^{x_j})\} \ \ \forall j\end{cases}.$$

Here A can be implicit, explicit or mixed scheme.

The above method is in the category of *splitting method*, where there is one or more intermediate steps in the numerical schemes to go from step i to step $i+1$.

The splitting method described in this section can also be applied directly to Black-Scholes equation with free boundary

$$\begin{cases}\partial_\tau V(\tau,S) = \frac{\sigma^2S^2}{2}\partial_{SS}V(\tau,S) + rS\partial_S V(\tau,S) - rV(\tau,S) & \text{for} \quad S > S^*(\tau)\\ V(\tau,S) = g(T-\tau,S) & \text{for} \quad S \leqslant S^*(\tau)\\ V(0,S) = g(T,S); \end{cases}.$$

The CFL condition is not different in the case of free boundary problems.

The Monte Carlo methods for American options are more complicated than for the European options and in beyond the scope of this lecture notes. For more information of the Monte Carlo methods for American options, see [19], [7], or the textbook [15].

[3]The subscript in $\tau_{i+\frac{1}{2}}$ indicate that finite-difference evaluation in each step in an intermediate step.

American call option on discrete dividend-paying asset

Unlike continuous dividend problem, the discrete dividend cannot be solved as a single free boundary problem. Consider an asset which pays dividend yield of $d \in (0, 1)$ at times $t_1 < t_2 < ... < t_n = T$. Proposition 4.1.1 suggests that at any time $t \in [t_i, t_{i+1})$ between the times of dividend payments it is better to wait and not to exercise. However, at time t_i of the dividend payment, the price of the asset decreases by the dividend and so does the price of the call option. Therefore, if the continuation is not optimal, the option should be exercises at the moment just before the time of a dividend payment.

Appendices

A Convex optimization

In this section, we briefly review the important result that we need from linear and convex optimization for this course. We start with reminding the notion of convex set and convex function.

A.1 Convex functions

Definition A.1. *A set $A \subseteq \mathbb{R}^d$ is called convex if for any $\lambda \in (0,1)$ and $x, y \in A$, we have*

$$\lambda x + (1-\lambda)y \in A.$$

In other words, a convex set is a set which contains all the segments with endpoints inside the set.

Example A.1. *The unit disk $\{x \in \mathbb{R}^d : |x| \leqslant 1\}$ and the unit cube $[0,1]^d$ are convex sets. The volume given by inequalities $a_i \cdot x \leqslant b_i$ for $i = 1, ..., n$, $a_i \in \mathbb{R}^d$ and $b_i \in \mathbb{R}$ (enclosed within n hyperplanes) is a convex set. Euclidean space R^d and the empty set are also convex.*

Definition A.2. *Let A be a convex set. A real function $f : A \subseteq \mathbb{R}^d \to \mathbb{R}$ is called convex if for any $\lambda \in (0,1)$ and $x, y \in \mathbb{R}^d$, we have*

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda)f(y).$$

A function f is called concave if $-f$ is convex. The convexity (concavity) is called strict if the inequality above is strict when $x \neq y$.

Proposition A.1. *A set $A \subseteq \mathbb{R}^d$ is called convex. Then, a function $f : A \subseteq \mathbb{R}^d \to \mathbb{R}$ is convex if and only if*

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leqslant \sum_{i=1}^{n} \lambda_i f(x_i),$$

for all $x_1, ..., x_n \in A$ and $\lambda_1, ..., \lambda_n \in \mathbb{R}_+$ with $\sum_{i=1}^{n} \lambda_i = 1$.

If a convex function f is twice differentiable, then the Hessian matrix of second derivatives of f, $\nabla^2 f$, has all eigenvalues nonnegative. In one dimensional case, this is equivalent to $f'' \geqslant 0$.

However, not all convex function are twice differentiable or even differentiable. We actually know that all convex functions are continuous. In addition, we can show that the one-sided directional derivatives of a continuous function exits.

Proposition A.2. *Let A be a convex set and $f : A \subseteq \mathbb{R}^d \to \mathbb{R}$ be a convex function. Then, for all x in the interior of A and all vectors $v \in \mathbb{R}^d$, the directional derivative of f at x is* *the direction of* v,

$$\nabla_v f(x) := \lim_{\varepsilon \to 0^+} \frac{f(x+\varepsilon v) - f(x)}{\varepsilon}$$

exits and satisfies $f(x+tv) \geqslant f(x) + t\nabla_v f(x)$ *for all* $t \geqslant 0$. *In particular,* f *is continuous at all points of* A.

In one dimensional case, for a convex functions f the above Proposition implies that the right and left derivatives, $f'(\cdot+)$ and $f'(\cdot-)$, exist at all points, and in particular a convex function is always continuous.

The above proposition has an important implication about the tangent hyperplane to a convex function.

Corollary A.1. *Consider a convex function* f *and* $x_0 \in \mathbb{R}^d$. *Then, there exists some* $u \in \mathbb{R}^d$ *such that the hyperplane* $y = f(x_0) + (x - x_0) \cdot u$ *always lies underneath the surface, i.e.*

$$f(x) \geqslant f(x_0) + (x - x_0) \cdot u \quad \textit{for all } x \in \mathbb{R}^d.$$

When f is differentiable, the linear approximation of f is always *under-approximating* the function, i.e.

$$f(x) \geqslant f(x_0) + (x - x_0) \cdot \nabla f(x_0) \quad \text{for all } v \in \mathbb{R}^d.$$

If a function f is twice differentiable, the Hessian $\nabla^2 f$ is the matrix which contains all second derivatives of the function f, i.e. $\left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right]_{d\times d}$. For twice differentiable function one can provide a criteria for convexity based on the eigenvalues of the Hessian matrix.

Corollary A.2. *If all eigenvalues of the Hessian matrix* $\nabla^2 f$ *are nonnegative (positive) at all points, then the function* f *is convex (resp. strictly convex).*

One can also use the Hessian matrix to find the local minimum and maximums of a function by checking its local convexity and concavity which is given in the following result. Recall that a point x_0 for a differentiable function is called *critical* if $\nabla f(x_0) = 0$.

Proposition A.3. *A critical point* x_0 *for a differentiable function* f *is a local minimum (resp. maximum) if and only if* f *is convex (resp. concave) in a neighborhood of* x_0.

As a result of the above proposition we have the second order derivative test in multivariate calculus.

Proposition A.4. *A critical point* x_0 *for a second order differentiable function* f *is*

- ***i)*** *a local minimum if the Hessian* $\nabla^2 f(x_0)$ *is positive-definite.*
- ***ii)*** *a local maximum if the Hessian* $\nabla^2 f(x_0)$ *is negative-definite.*
- ***iii)*** *a saddle point if the Hessian* $\nabla^2 f(x_0)$ *has both negative and positive eigenvalues.*

Exercise A.1. *Find and categorize all the critical point of the function* $f(x,y) = xe^{-\frac{x^2+y^2}{2}}$.

Exercise A.2. *Find and categorize all the critical point of the function* $f(x,y) = xye^{-\frac{x^2+y^2}{2}}$.

A.2 Convex constrained optimization

Consider a convex function $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ where the set D is given by

$$K = \{x \in \mathbb{R}^n \; : \; g_i(x) \leqslant 0 \;\text{ for }\; i = 1, ..., m\}.$$

We assume that all functions g_i for $i = 1, ..., m$ are convex and therefore the set K is closed and convex. The primal problem is to minimize f in D, i.e.

$$\min f(x) \quad \text{subject to constraints} \;\; g_i(x) \leqslant 0 \quad \text{for} \quad i = 1, ..., M.$$

If f is strictly concave, then the optimizer is unique if it exists. Notice that the existence of the optimizer is subject to nonemptiness of the set D.

An equality constraint can also be described in the above form if the function g_i is affine[4] and both $g_i(x) \leqslant 0$ and $-g_i(x) \leqslant 0$. For later analysis, we specify constraint equalities in the problem separately and define the *feasibility* set by

$$K = \{x \in \mathbb{R}^n \; : \; g_i(x) \leqslant 0 \;\text{ for }\; i = 1, ..., m \;\text{ and }\; h_j(x) = 0 \;\text{ for }\; j =, ..., k\}.$$

Notice that since g_is are convex and h_js are affine, K is convex, for any number $\lambda \in (0, 1)$ and any $x, y \in K$, $\lambda x + (1 - \lambda)y \in K$. Given the convex functions g_i for $i = 1, ..., m$ and affine functions h_j for $j = 1, ..., k$, the convex constrained optimization problem is given by

$$P := \min f(x) \quad \text{subject to constraints} \;\; \begin{cases} g_i(x) \leqslant 0 & \text{for} \quad i = 1, ..., m \\ h_j(x) = 0 & \text{for} \quad j = 1, ..., k \end{cases}. \tag{A.1}$$

Function f is called the *objective* function and the inequalities $g_i \leqslant 0$ and equations $h_j = 0$ are called constraints. If $K \neq \emptyset$, the problem is called feasible. if $f(x)$ is strictly convex, then the minimizer is unique; for any two distinguished minimizers x_1 and x_2 with minimum value $P = f(x_1) = f(x_2)$, we have $f\left(\frac{x_1+x_2}{2}\right) < \frac{1}{2}(f(x_1) + f(x_2)) = P$ which contradicts with that x_1 and x_2 are minimizers.

Duality method is one of the useful approaches to solve the convex optimization problems. To present the dual problem, we first introduce the Lagrangian

$$L(\mu, \lambda) = \min_{x \in \mathbb{R}^n} \Big\{ f(x) + \mu \cdot g(x) + \lambda \cdot h(x) \Big\}.$$

Here $\mu \in \mathbb{R}^m$ and $g(x) = (g_1(x), ..., g_m(x)))$, $\lambda \in \mathbb{R}^k$ and $h(x) = (h_1(x), ..., h_m(x)))$ and $\cdot$ is the dot product in the proper Euclidean space. Then, the dual problem is given by

$$D := \max_{\mu \in \mathbb{R}^m_+, \; \lambda \in \mathbb{R}^k} L(\mu, \lambda).$$

[4] An affine function is a linear function plus a constant; $h(x) = a \cdot x + b$, where $a \in \mathbb{R}^m$ and $b \in \mathbb{R}$.

Notice that for $x \in K$, $L(\mu, \lambda) \leqslant f(x) + \mu \cdot g(x) + \lambda \cdot h(x) \leqslant f(x)$. Therefore, $L(\mu, \lambda) \leqslant \min_{x \in K} f(x)$ and thus, $D \leqslant P$ which is called the *weak duality*. We can show that under the following condition that the *strong duality* holds.

Assumption A.1 (Slater condition). *There exist a point x_0 such that $g_i(x_0) < 0$ for all $i = 1, ..., m$ and $h_j(x_0) = 0$ for all $j = 1, ..., k$.*

Theorem A.1 (Duality). *Let f, g_i for $i = 1, ..., n$ and h_j for $j = 1, ..., m$ are convex and Slater condition A.1 holds. Then $D = P$. In addition, the dual maximizer $(\mu^*, \lambda^*) \in \mathbb{R}^m_+ \times \mathbb{R}^k$ exists whenever $D > -\infty$, i.e.*

$$\max_{\mu \in \mathbb{R}^m_+,\ \lambda \in \mathbb{R}^k} L(\mu, \lambda) = L(\mu^*, \lambda^*).$$

A proof of this theorem can be found in [8, Section 5.3.2]. The following example shows if Slater condition fail, the strong duality does not necessarily hold.

Example A.2. *Take $d = m = 2$, $n = 1$ with $f(x_1, x_2) = x_1^2 + x_2$, $g_1(x) = x_2$, $h_1(x) = x_1 + x_2$ and $h_2(x) = x_1 - x_2$. Since feasibility set is a singleton; $K = \{(0,0)\}$, the Slater condition does not hold. Therefore, $P = f(0,0) = 0$. On the other hand, $L(\mu, \lambda) = -\infty$ unless $\lambda = 0$. Specifically, $L(\mu, 0) = \min_{(x_1, x_2)} f(x_1, x_2) + \mu x_2$. Then, we also have*

$$L(\mu, 0) = \min_{x_1} x_1^2 + \min_{x_2} x_2(1 + \mu) = -\infty,$$

since $\mu \geqslant 0$. Therefore, $D = -\infty < P = f(0,0) = 0$.

Remark A.1. *Notice that the duality can sometimes hold when the Slater condition does not hold or when the problem is not even feasible. For instance, if $K = \varnothing$, since $\min_{\varnothing} = \infty$, $P = \infty$. On the other hand, for and arbitrary point x_0, let I and J respectively be the set of all indices i and j such that $g_i(x_0) > 0$ and $h_j(x_0) \neq 0$. Since $K = \varnothing$, at least one of I or J is nonempty. Then, choose μ and λ such that $\mu_i = 0$ if and only if $i \notin I$ and $\lambda_j = 0$ if and only if $j \notin J$. Therefore,*

$$L(\mu, \lambda) \leqslant f(x_0) + \mu \cdot g(x_0) + \lambda \cdot h(x_0).$$

By sending $\mu_i \to +\infty$ for $i \in I$ and $\lambda_j \to \pm\infty$ whether $h_j(x_0)$ is negative/positive, we obtain that $\max_{\mu \in \mathbb{R}^m_+,\ \lambda \in \mathbb{R}^k} L(\mu, \lambda) = +\infty$.

One of the practical methods of finding the optimal points for primal and dual problem is through the KKT[5], which provided a necessary condition of optimality. Additionally with Slater condition, KKT is also sufficient.

[5] Karush-Kuhn-Tucker

Theorem A.2 (KKT optimality condition)**.** *Assume that f, g and h are differentiable and $x^* \in \mathbb{R}^d$ and $(\mu^*, \lambda^*) \in \mathbb{R}_+^m \times \mathbb{R}^k$ are respectively the optimizer of the primal and dual problems. Then,*

$$\begin{aligned} \partial_{x_\ell} f(x^*) + \mu \cdot \partial_{x_\ell} g(x^*) + \lambda \cdot \partial_{x_\ell} h(x^*) &= 0 \;\; \text{for} \;\; \ell = 1, ..., n \\ \mu_i^* g_i(x^*) &= 0 \;\; \text{for} \;\; i = 1, ..., m. \\ h_j(x^*) &= 0 \;\; \text{for} \;\; j = 1, ..., k \end{aligned} \tag{A.2}$$

Condition $\mu_i^* g_i(x^*) = 0$ is called the complementary slackness condition. In addition, if the Slater condition A.1 holds and (A.2) is satisfied, then $x^* \in \mathbb{R}^d$ and $(\mu^*, \lambda^*) \in \mathbb{R}_+^m \times \mathbb{R}^k$ are respectively the optimizer of the primal and dual problems.

Example A.3. *We want to find the maximum volume of an open lid box with a fixed surface area s by solving the constrained maximization problem*

$$\max xyz \quad \text{subject to constraints} \;\; xy + 2xz + 2yz = s.$$

Notice that since the Slater condition holds for $x_0 = y_0 = \sqrt{s}$ and $Z_0 = 0$, the dual problem and primal problem have the same value. The KKT condition suggest to solve the following to find candidates for the primal and dual problem.

$$\begin{aligned} xy + 2\mu(x + y) &= 0 \\ xz + \mu(x + 2z) &= 0 \\ yz + \mu(y + 2z) &= 0 \\ xy + 2xz + 2yz &= s \end{aligned}$$

Solving the above system of equation yields

$$x^* = y^* = 2z^* = \frac{\sqrt{3s}}{3} \;\; \text{and} \;\; \mu^* = -\frac{\sqrt{3s}}{12},$$

and the maximum volume is $\frac{s\sqrt{3s}}{16}$.

Exercise A.3. *Find the volume of the largest box under the constraint that the sum of the diagonals of the three sides sharing a corner equal is to s.*

The duality method, while can be used as a computational tool, it can also provide us with some qualitative results about the optimization problem that we are studying. For instance, in Section 2.1.6, the dual problem for evaluation of model risk is the problem of superreplication. In the next sections, we formulate the dual problem for the linear and quadratic programming.

Linear programming

Linear optimization is when the objective function f and the constraints are all linear; for some constants $p \in \mathbb{R}^d$, $a_i \in \mathbb{R}^d$, $b_i \in \mathbb{R}$ for $i = 1, .., m$, and $c_j \in \mathbb{R}^d$ and $d_j \in \mathbb{R}$ for $j = 1, ..., k$.

$$f(x) = p \cdot x, \quad g_i(x) = a_i \cdot x + b_i, \quad \text{and} \quad h_j(x) = c_j \cdot x + d_j.$$

Therefore,

$$P := \min p \cdot x \quad \text{subject to constraints} \quad \begin{cases} a_i \cdot x + b_i \leqslant 0 & \text{for} \quad i = 1, ..., m \\ c_j \cdot x + d_j = 0 & \text{for} \quad j = 1, ..., k \end{cases}. \tag{A.3}$$

Let matrices A and C consist of rows $a_1, ..., a_m$ and $c_1, ..., c_k$, respectively and set $d := (d_1, ..., d_k)$ and $b := (b_1, ..., b_m)$ as column vectors. Then, the linear programming can be written in the following compact form.

$$\min p \cdot x \quad \text{subject to constraints} \quad \begin{cases} Ax + b & \leqslant 0 \\ Cx + d & = 0 \end{cases}.$$

Each equation $c_j \cdot x + d_j = 0$ is a hyperplane and each inequality $a_i \cdot x + b_i \leqslant 0$ is a half-space in the d-dimensional Euclidean space. The feasibility set K can easily be empty, if the constraints are made by parallel hyperplanes; for example $a_i x + b_i \leqslant 0$ and $a_{i'} x + b_{i'} \leqslant 0$ with $a_i = -a_{i'}$ and $b_i = -b_{i'} - 1$. To avoid this situation, we assume that the vectors $a_1, ..., a_m$ and $c_1, ..., c_k$ are linearly independent in $\mathbb{R}^d$. As a result, $m + k \leqslant d$ the matrices A and C are full rank, where A and C are matrices with rows $a_1, ..., a_m$ and $c_1, ..., c_k$, respectively. In addition, we need to assume that the column vector $d = (d_1, ..., d_k)$ is in the range of $-C$ and the orthant $\{y \in \mathbb{R}^m : y \geqslant b_i \ \text{ for } \ i = 1, ..., m\}$ intersects with the range of $-A$, $Cx + d = 0$ and $Ax + b \leqslant 0$ together have a solution for x. and are column vectors.

The Lagrangian for the linear programming is given by

$$L(\mu, \lambda) = \min_{x \in \mathbb{R}^d} \{p \cdot x + \mu \cdot (Ax + b) + \lambda \cdot (Cx + d)\}.$$

Notice that if (μ, λ) are such that $p^\top + (\mu^*)^\top A + (\lambda^*)^\top C \neq 0$ then,

$$\min_{x \in \mathbb{R}^d} \left\{ (p^\top + \mu^\top A + \lambda^\top C) \cdot x \right\} + \mu^\top b + \lambda^\top d = -\infty.$$

Therefore,

$$L(\mu, \lambda) = \begin{cases} -\infty & p^\top + (\mu^*)^\top A + (\lambda^*)^\top C \neq 0 \\ \mu^\top b + \lambda^\top d & p^\top + (\mu^*)^\top A + (\lambda^*)^\top C = 0 \end{cases}.$$

Therefore, the dual problem can be written as

$$D := \max_{\mu \geqslant 0, \lambda} \mu^\mathsf{T} b + \lambda^\mathsf{T} d \quad \text{subject to constraints} \;\; p^\mathsf{T} + (\mu^*)^\mathsf{T} A + (\lambda^*)^\mathsf{T} C = 0 \tag{A.4}$$

Theorem A.3 (Linear programming duality). *Consider the following two linear programming problems* (A.3) *and* (A.4)*. If either of the problems has an optimal solution, so does the other one and both problems have the same value, $D = P$.*

Exercise A.4. *Write the dual problem and KKT condition for the standard linear programming equation:*

$$\min p \cdot x \quad \textit{subject to constraints} \quad \begin{cases} x_i \leqslant 0 & \textit{for} \quad i = 1, ..., d \\ c_j \cdot x + d_j = 0 & \textit{for} \quad j = 1, ..., k \end{cases}.$$

Exercise A.5 (Project). *Study the following algorithms for the linear programming problem in Exercise 2.1.10. Find a package that has both methods and compare the running time of each method on the same problem.*

a) *Simplex method*

b) *Interior points*

Quadratic programming

Consider a positive-definite symmetric[6] d-by-d matrix M, $a_i \in \mathbb{R}^d$, $b_i \in \mathbb{R}$ for $i = 1, .., m$, and $c_j \in \mathbb{R}^d$ and $d_j \in \mathbb{R}$ for $j = 1, ..., k$.

$$f(x) = \frac{1}{2} x \cdot Mx, \quad g_i(x) = a_i \cdot x + b_i, \quad \text{and} \quad h_j(x) = c_j \cdot x + d_j.$$

Therefore, the quadratic optimization problem is given by

$$\min \frac{1}{2} x \cdot Mx \quad \text{subject to constraints} \quad \begin{cases} a_i \cdot x + b_i \leqslant 0 & \text{for} \quad i = 1, ..., m \\ c_j \cdot x + d_j = 0 & \text{for} \quad j = 1, ..., k \end{cases}. \tag{A.5}$$

The Lagrangian for the linear programming is given by

$$L(\mu, \lambda) = \min_{x \in \mathbb{R}^d} \left\{ \frac{1}{2} x \cdot Mx + \mu \cdot (Ax + b) + \lambda \cdot (Cx + d) \right\}.$$

[6] A matrix A is symmetric if $A^\mathsf{T} = A$ and is positive-definite if all its eigenvalues are

The dual problem for (A.5) is also a quadratic problem; the minimizer for x in the Lagrangian function $L(\mu, \lambda)$ is given by $x(\mu, \lambda) = -M^{-1}(A^\top \mu + C^\top \lambda)$ and therefore Lagrangian $L(\mu, \lambda)$ is a quadratic function of μ and λ. Therefore,

$$L(\mu, \lambda) = -\frac{1}{2}\mu \cdot \tilde{A}\mu + \mu \cdot b - \frac{1}{2}\lambda \cdot \tilde{C}\lambda + \lambda \cdot d,$$

where $\tilde{A} = AM^{-1}A^\top$ and $\tilde{C} = CM^{-1}C^\top$ are indeed nonnegative definite symmetric matrices. The dual problem can now be decoupled into two problems;

$$\max_{\mu \geqslant 0} \left\{ -\frac{1}{2}\mu \cdot \tilde{A}\mu + \mu \cdot b \right\} \quad \text{and} \quad \max_{\lambda} \left\{ -\frac{1}{2}\lambda \cdot \tilde{C}\lambda + \lambda \cdot d \right\}$$

The maximization problem on λ is an unconstrained problem which leads to $\lambda^* = \tilde{C}^{-1}d$ when $\tilde{C}$ is invertible. The maximization problem on μ is, however, a constrained problem. One way to find the primal and the dual optimal variables X^*, μ^* and λ^* is through KKT condition in Theorem (A.2), which is written as

$$\begin{aligned} Mx^* + A^\top \mu^* + C^\top \lambda^* &= 0 \\ \mu_i^*(a_i x^* + b_i) &= 0 \;\; \text{for} \;\; i = 1, ..., m \\ c_j x^* + d_j &= 0 \;\; \text{for} \;\; j = 1, ..., k \end{aligned}$$

When the dual maximizer $\mu_i^* > 0$, then by KKT condition the ith constraint must hold with equality, $a_i x^* + b_i = 0$. In this case, we call the constraint an *active* constraint.

Exercise A.6 (Project)**.** *Study the following algorithms for the linear programming problem in Section 1.2. Find a package that has both methods and compare the running time of each method on the same problem.*

a) *Active set*

b) *Interior points*

Computational tools for convex optimization problems

The computational methods for convex optimization problems are vast and we do not intend to study them in this book. Instead, we briefly introduce some of the tools which you can use to solve optimization problems in finance.

A well-developed convex optimization tool is `CVX` under `MatLab` created by Michael Grant and Stephen Boyd. The academic (noncommercial) version of the toolbox is free and under a GNU General Public License; it can be used or redistributed but not altered. However, it works under commercial software `MatLab`. The home of `CVX` is `http://cvxr.com`.

Another convex optimization tool is `CVXOPT` which is a `Python` Program created by Martin Andersen, Joachim Dahl, and Lieven Vandenberghe. The license is also a GNU General Public License. The home of `CVXOPT` is `https://cvxopt.org`.

As a fun fact, Stephen Boyd from Stanford University and Lieven Vandenberghe from UCLA are the authors of the convex optimization book [8] which has more than 40,000 citations and is widely used as textbook for optimization courses.

B A review of probability theory

In financial modeling, most uncertainties can effectively bemodeled by probability. Probability was not considered part of mathematics until approximately four hundred years ago by a series of correspondence between Blaise Pascal (1623–1662) and Pierre de Fermat (1607–1665). Most of the problems were motivated by observation in gambling and games of chance.

Example B.1 (Empirical observation)**.** *One wins a game of rolling one die if he achieves at least one six in four trials. In a different game, one wins in rolling two dice if he achieves at least one double six in twenty-four trials. Chevalier de Méré a.k.a. Antoine Gombaud (1607–1684) wrote to Pascal that these two games must have the same probability of winning; in the latter, the chance of getting a favorable outcome is a round in six times less than the former, while the number of trials is six times more. Chevalier de Méré, however, discovered that the two games are not empirically the same. More specifically, he observed that the first game has a winning chance of slightly more than 50% while the second game's chance is slightly greater than that. Pascal and Fermat discussed the problem until Fermat eventually solved it. The solution is as follows.*

The chance of losing the first game is $\left(\frac{5}{6}\right)^4$. *Therefore, the change of winning is* $1-\left(\frac{5}{6}\right)^4 \approx 0.5177$.*The second problem has the change of winning equal to* $1-\left(\frac{35}{36}\right)^{24} \approx 0.4914$.

It is astonishing that Chevalier de Méré could empirically distinguish a difference of 0.0263 *between the two chances of winning.*

Example B.2 (Problem of points)**.** *Three players are playing over a stake. The condition of winning is that whoever wins a certain number of rounds takes the whole stake. Player 1 needs one more round to win the stake, while players 2 and 3 need two rounds. However, outside circumstances dictates that they have to suddenly stop playing and agree to divide the stake according to current situation of the game. Obviously, player 1 deserves a greater share of the stake than the other two players, who should receive the same share. What is the fair share of the stake for each player?*

> *The game is over after at most three rounds. Therefore, for each possible outcome of the next three rounds a winner can be decided. For example, if the outcome of the next three rounds is* $(2, 3, 1)$ *(player 2 wins, then player 3 wins, and finally player 1 wins), then player 1 is the winner. There are twenty-seven possible outcomes in the next three rounds in seventeen of which player 1 wins. Players 2 and 3 each win in exactly five of the twenty-seven outcomes. Therefore, if all players have the same chance of winning each round of the game, the stake should be divided by* 17*,* 5 *and* 5 *between players 1, 2, and 3, respectively.*

For about a hundred more years, probability theory continued to be used primarily to address gambling-related problems, until Jacob Bernoulli (1655–1705), Abraham de Moivre (1667–1754), and Thomas Bayes (1702–1761) introduced the first limit theorems. Bernoulli proved that if the probability of heads in tossing a coin is p, then the frequency of heads in a sequence of n trials converges to p as $n \to \infty$. De Moivre showed that the empirical distribution of the number of heads in n trials converges to a normal distribution.

> **Law of large numbers.** If r is the number of heads in n trials of tossing a coin with heads probability p, then the probability of that $|p - \frac{r}{n}| \geqslant \epsilon$ converges to zero as $n \to \infty$.

In n trials of tossing a coin with heads probability p, the probability of r heads is given by $\binom{n}{r} p^r (1-p)^{n-r}$ for $r = 0, ..., n$. This probability is called the binomial distribution.

> **De Moivre-Laplace.** In the binomial distribution, $\binom{n}{r} p^r (1-p)^{n-r} \sim \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ holds when as $n \to \infty$ and $\frac{r-np}{\sqrt{np(1-p)}} \to x$.

The above theorem is an early version of the *central limit theorem*, which will be presented later in Theorem B.7 in Section B.4.

Bayes also contributed the concept of conditional probability, which deals with how the occurrence of a specific event can affect the probabilities of other events. This gave birth to the concept of conditional probability. Consider the Venn diagram in Figure B.1. Assume that events A, B, and C have their ex ante probabilities. If we discover that event A has happened, the ex post probability of A is now equal to 1. Another event B is now restricted to whatever remains of it inside A, i.e., $A \cap B$, and their ex post probabilities are obtained by rescaling the ex ante probability of $A \cap B$ by the ex ante probability of A. For example, the probability of event B given by

$$\text{ex post probability of } A \text{ given } B = \frac{\text{ex ante probability of } A \cap B}{\text{ex ante probability of } A}.$$

In Figure B.1, the ex post probability of C is zero, since the ex ante probability of $A \cap C$ is zero.

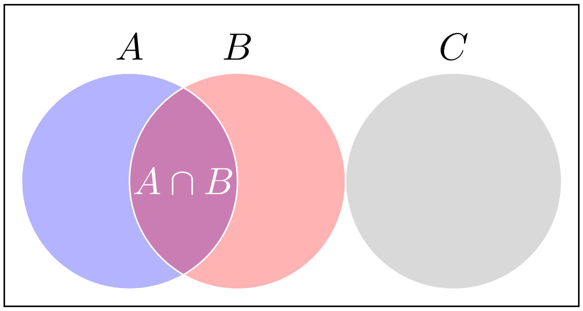


Figure B.1: Given A, the ex post probability of event B is the probability of event B inside A ($A \cap B$) divided by the probability of A.

In the above, we explained how to find the ex post probability in terms of the ex ante probability. The Bayes formula explains how to obtain the ex ante probability in terms of the ex post probability. Bayes formula, one of the most influential results in probability, constructs the foundation of Bayesian statistics. To explain the formula, first we need to introduce some notations. By $\mathbb{P}(A)$ we denote the ex ante probability of event A, and by $\mathbb{P}(A|B)$ we denote the ex post probability of A given B.

Bayes formula 1.

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^{\complement})\mathbb{P}(B^{\complement})$$

If $B_1, B_2, ...$ are mutually exclusive and

$$\mathbb{P}(B_1 \cup B_2 \cup \cdots) = 1,$$

then

$$\mathbb{P}(A) = \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \cdots .$$

When ex post probabilities of an event A conditional on a certain event B and the ex ante probability of B are known, one can use Bayes formula to find the ex post probability of B conditional on A.

Bayes formula 2.

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^{\complement})\mathbb{P}(B^{\complement})}$$

We elaborate on the notion of conditional probability and the Bayes theorem via a famous example in conditional probability, namely the three prisoners problem.

Example B.3 (Three prisoners problem). *Three suspects (Bob, Kevin, and Stuart) are equally likely to be convicted. However, the judge has made up his mind and is going to pronounce one and only one of them guilty the following morning. None of the three prisoners knows who is going to be announced; however, the warden has been told the name of the guilty by the judge and has been given strict orders not to pass the information to the prisoners. Stuart argues with the warden that at least one of the other two, Kevin or Bob, is not guilty. If the warden names one of the other two, Stuart still cannot know if he himself is guilty or not. The warden counterargues that if he reveals the name of a nonguilty person, Stuart's chance of conviction increases from $\frac{1}{3}$ to $\frac{1}{2}$, because now there are two prisoners one of whom is going to be guilty. Is the warden correct?*

To answer the question, we formulate the problem using conditional probabilities. In this case, if the warden names Bob as not guilty, then the ex ante probability of Stuart's conviction is given by

$$\frac{\textit{ex ante probability of conviction of Stuart} \bigcap \textit{Bob named not guilty by the warden}}{\textit{ex ante probability of Bob named not guilty by the warden}}.$$

The ex ante probability of Bob named not guilty by the warden can be calculated using the Bayes formula. Let's denote this event by $\tilde{B}$, and let S, K, and B be the events that either Stuart, Kevin, or Bob, respectively, is announced guilty. By conditional probability, one needs to calculate

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

Since $\mathbb{P}(S \cup K \cup B) = 1$, it follows from the first Bayes formula that

$$\mathbb{P}(\tilde{B}) = \mathbb{P}(\tilde{B}|S)\mathbb{P}(S) + \mathbb{P}(\tilde{B}|K)\mathbb{P}(K) + \mathbb{P}(\tilde{B}|B)\mathbb{P}(B).$$

Notice that $\mathbb{P}(K) = \mathbb{P}(B) = \mathbb{P}(S) = \frac{1}{3}$[7], $\mathbb{P}(\tilde{B}|S) = \frac{1}{2}$, $\mathbb{P}(\tilde{B}|K) = 1$, and $\mathbb{P}(\tilde{B}|B) = 0$. Therefore,

$$\mathbb{P}(\tilde{B}) = \frac{1}{6} + \frac{1}{3} + 0 = \frac{1}{2}.$$

On the other hand,

$$\mathbb{P}(S \cap \tilde{B}) = \mathbb{P}(\tilde{B}|S)\mathbb{P}(S) = \frac{1}{6}.$$

Therefore, $\mathbb{P}(S|\tilde{B}) = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$.

In Example B.3, the revealed information does not change the ex ante probability. This is, however, a coincidence and not necessarily true. The following exercise addresses this issue.

[7] Here we assume that they are equally likely to be pronounced guilty. One can adjust these probabilities and solve the problem accordingly.

Exercise B.1. *Repeat the solution to Example B.3 with* $\mathbb{P}(K) = \mathbb{P}(S) = \frac{1}{6}$ *and* $\mathbb{P}(B) = \frac{2}{3}$*. Beware as here is some ambiguity about how to choose* $\mathbb{P}(\tilde{B}|S)$*!*

Example B.4. *In a burglary hearing, there is a probability* p_G *that the suspect is guilty. In actual fact, the suspect is left-handed. If it is brought before the court that the burglar is right-handed, the ex post probability that the suspect is guilty vanishes to zero. However, if the evidence shows that the burglar is left-handed, we need to use the Bayes formula to find the ex post probability that the suspect is guilty. In this case, let* G *be the event that the suspect is guilty and* L *be the event that the burglar is left-handed that, we assume, has the same ex ante probability of any individual is left-handed*[8]*. Then,*

$$\mathbb{P}(G|L) = \frac{\mathbb{P}(L|G)\mathbb{P}(G)}{\mathbb{P}(L)} = \frac{\mathbb{P}(G)}{\mathbb{P}(L)}.$$

Notice that $\mathbb{P}(L|G) = 1$*, because the suspect is left-handed. On the other hand,*

$$\mathbb{P}(L) = \mathbb{P}(L|G)\mathbb{P}(G) + \mathbb{P}(L|G^{\complement})\mathbb{P}(G^{\complement}) = \mathbb{P}(G) + \mathbb{P}(L|G^{\complement})\mathbb{P}(G^{\complement})$$

Notice that $\mathbb{P}(G) = p_G$*,* $\mathbb{P}(G^{\complement}) = 1 - p_G$*, and* $\mathbb{P}(L|G^{\complement})$ *is roughly equal to the percentage of the left-handed population, denoted by* p_L*. Therefore,*

$$\mathbb{P}(G|L) = \frac{p_G}{p_G + (1 - p_G)p_L}.$$

Notice that when $P_L < 1$*, then* $\mathbb{P}(G|L) > \mathbb{P}(G) = p_G$*, because* $p_G + (1 - p_G)p_L < 1$*. For instance, when* $p_L = .1$ *and* $p_G = .15$*, then the ex post probability* $\mathbb{P}(G|L)$ *is* $\frac{15}{23.5}$*. If* $p_L = 0$*, then no one in the population is left-handed except for the suspect, which makes him guilty. If* $p_L = 1$*, then everyone in the population is left-handed,* $\mathbb{P}(G|L) = p_G$*, and the event* L *does not add to the information we already have.*

Exercise B.2. *Show that if the ex ante probability of event* A *is 1, then* $\mathbb{P}(B|A) = \mathbb{P}(B)$*.*

The notion of conditional probability (the handling of new pieces of information in probability) is one of the most influential tools in applications of probability theory. Most important applications appear in Bayesian statistics, through which they contribute to all other areas of science.

We conclude the discussion of early advances in probability theory by mentioning that these early developments are gathered in *Essai philosophique sur les probabilités* by Pierre-Simone Laplace (1749–1827) ([18]). However, the subject still lacked mathematical rigor compared to other areas of mathematics until a breakthrough happened in the twentieth century. Andrey Nikolaevich Kolmogorov (1903–1987) initiated the foundation of probability theory through *measure theory*, a topic in mathematical analysis. We provide a more

[8] It is logical to assume that the probability that a burglar is left-handed is the same as any other individual.

rigorous treatment of probability theory in Section B.2. In the next section, we present the discrete theory of probability, which does not need measure theory.

B.1 Basic concepts and definitions of discrete probability

A (finite, countably infinite) *sample space* is a collection of all possible outcomes of a random experiment. Any subset A of the sample space Ω is called an *event*.

Example B.5 (Flipping a coin). *The sample space is* $\{head, tail\}$.

Example B.6 (Arrow-Debreu market model in Section 2.1). *The sample space can be chosen to be* $\{1, ..., M\}$*, i.e., the collection of all the possible states of the system.*

Example B.7 (T-period binomial model). *The sample space of the* T*-period binomial model in Section 2.3 can be chosen to be the collection of all* T*-sequences of the form* $(a_1, ..., a_T)$ *where each* a_i *is either* u *or* ℓ*. Each sample addresses the complete movements of the asset price over time.*

A *probability* over a finite (countably infinite) sample space $\Omega = \{\omega_1, ..., \omega_M\}$ ($\Omega = \{\omega_1, \omega_2, ...\}$) is a vector $\pi = (\pi_1, ..., \pi_M)$ ($\pi = (\pi_1, \pi_2, ...)$) of nonnegative values such that $\sum_{\omega\in\Omega} \pi_i = 1$. For simplicity, we write $\mathbb{P}^\pi(\omega_j) = \pi_j$ if no confusion occurs. The collection of all subsets of Ω determines the set of all *events*. The probability of an event $A \subseteq \Omega$ is then defined by

$$\mathbb{P}(A) := \sum_{\omega\in A} \pi_{\omega_i}.$$

Evidently, $\mathbb{P}$ satisfies

1) $\mathbb{P}(\emptyset) = 0$,

2) $\mathbb{P}(A^\complement) = 1 - \mathbb{P}(A)$ for all $A \subseteq \Omega$, and,

3) If $\{A_n\}_{n\geqslant 1}$ is a sequence of **disjoint** events, then

$$\mathbb{P}\left(\bigcup_{n\geqslant 1} A_n\right) = \sum_{n\geqslant 1} \mathbb{P}(A_n).$$

Example B.8 (Flipping a coin). *In flipping a fair coin, the probability of either heads or tails is* $p = \frac{1}{2}$*. In flipping a unfair coin, the probability of heads is* $p \neq \frac{1}{2}$*, and therefore the probability of tails is* $1 - p$*. In two consecutive flips of a fair coin, the probability of having* (H, H)*,* (H, T)*,* (T, H)*, or* (T, T) *is equally* $1/4$*. If the coin is not fair, then the sample space is not changed. But the probabilities of these outcomes change to* $\pi_H = p$ *and* $\pi_T = 1 - p$ *where* $p \in [0, 1]$*. In two consecutive flips, we have* $\pi_{H,H} = p^2$*,* $\pi_{H,T} = \pi_{T,H} = p(1 - p)$*, and* $\pi_{T,T} = (1 - p)^2$*.*

Example B.9 (Single asset T-period binomial model)**.** *In a binomial model with T-periods, a risk-neutral probability assigns the probability $\pi_u^k \pi_\ell^{T-k}$ to an outcome $(a_1, ..., a_T)$ in which k of the entities are u and the $T - k$ remaining are ℓ.*

A *random variable* X is a function from sample space to $\mathbb{R}^d$, $X : \omega \in \Omega \mapsto X(\omega) \in \mathbb{R}^d$. The values that X takes with positive probability are called the values of the random variable, $x \in \mathbb{R}^d$ such that $\mathbb{P}(\omega : X(\omega) = x) > 0$. To simplify the notation, we often write $\mathbb{P}(X = x)$ for $\mathbb{P}(\omega : X(\omega) = x)$. When the sample space is finite or countably infinite, the random variables can only take finitely of countably infinitely many values. Random variables with at most countably infinitely many values are called *discrete random variables.*

Remark B.1. *Notice that the values of a random variable are relative to the choice of probability measure. For example, a random variable $X : \{0, 1, 2\} \to \mathbb{R}$ defined by $X(x) = x$ has values $\{0, 1, 2\}$ relative to probability $\mathbb{P}(0) = \mathbb{P}(1) = \mathbb{P}(2) = 1/3$. However, relative to a new probability $\mathbb{Q}(0) = \mathbb{Q}(1) = 1/2$ and $\mathbb{Q}(2) = 0$, the set of values is given by $\{0, 1\}$.*

Example B.10. *Recall from Example B.6 that the sample space for the Arrow-Debreu market model is the set of states $\Omega = \{1, ..., M\}$. Therefore, a random variable is given by a function $X : \{1, ..., M\} \to \mathbb{R}$. In particular, the payoff of an asset (the price of an asset at time 1 for each state of the market) in the Arrow-Debreu market model is a random variable. For instance, the payoff of asset i given in diagram 2.1.1 is a random variable $\mathbf{P}_{i,\cdot} : \Omega \to \mathbb{R}$ such that*

$$\mathbf{P}_{i,\cdot}(j) = P_{i,j}.$$

Example B.11 (Bernoulli random variable)**.** *Flipping a coin creates a Bernoulli random variable by assigning values to the outcomes heads and tails. The Bernoulli random variable X takes the value 1 if the coin turns heads and 0 otherwise. If the coin has a probability of tails equal to p, then $X = 1$ has probability p and $X = 0$ has probability $1 - p$.*

Definition B.1. *For an event A, the indicator of A is a random variable that takes value 1 if A occurs and 0 otherwise. The indicator of A is denoted by 1_A[9]. The indicator random variable 1_A is a Bernoulli random variable that takes value 1 with probability $\mathbb{P}(A)$ and value 0 with probability $1 - \mathbb{P}(A)$.*

Example B.12 (Binomial random variable)**.** *In flipping a coin n times, the binomial random variable X takes the value of the number of heads. The set of values of X is $\{0, ..., n\}$. If the coin has a probability of heads equal to p, then for x in the set of values, the probability $X = x$ is given by*

$$\binom{n}{x} p^x (1 - p)^{n-x}.$$

[9] the indicator is also denoted by χ_A in some literature.

Example B.13 (Random walk)**.** *In a game of chance, in each round a player flips a coin. If it turns tails, he gains \$1; otherwise, he loses \$1. Technically, each round has an outcome given by* $2X - 1$ *where* X *is a brand new Bernoulli random variable with outcomes* 1 *and* 0*. The player's accumulated reward after two rounds is a random variable* W_2 *that take values* -2*,* 0*, and* 2 *with probabilities* 1/4*,* 1/2*, and* 1/4*, respectively. If the coin has a probability of heads equal* p*, then* $\mathbb{P}(W_2 = 2) = p^2$*,* $\mathbb{P}(W_2 = -2) = (1-p)^2$*, and* $\mathbb{P}(W_2 = 0) = 2(1-p)p$*. Here,* 1 *is not among the values of* W_2*, because* $\mathbb{P}(W_2 = 1) = 0$*.*

If the player continues this game, the sequence of his accumulated wealth at all times, $W_0, W_1, W_2, ...$*, is called a random walk. At time* t*,* W_t *takes values* $\{-t, -t+2, ..., t\}$ *for* $t = 1, ..., T$*. The probability of* $W_t = x$ *is given by*

$$\binom{t}{k} p^k (1-p)^{t+1-k} \quad \textit{with} \quad k = \frac{t+x}{2}.$$

In Example B.13, a proper sample space can be given by

$$\Omega := \{(a_1, a_2, ...) \;\; a_i = H \;\; \textit{or} \;\; T \textit{ for } \; i = 1, 2, ...\}, \tag{B.1}$$

when the game is infinite.

Exercise B.3. *In Example B.13, calculate* $\mathbb{P}(W_2 = 2 \mid W_1 = 1)$ *and* $\mathbb{P}(W_3 = -1 \mid W_1 = 1)$*.*

Example B.14 (Negative binomial random variable)**.** *In flipping a coin, the negative binomial random variable* X *counts the number of heads before* r *number of tails appear. The set of values of* X *is* $\{r, r+1, ...\}$*. If the coin has a probability of heads equal to* p*, then for* x *in the set of values, the probability* $X = x$ *is given by*

$$\binom{x+r-1}{x} p^x (1-p)^r.$$

Exercise B.4. *In Example B.14, find an appropriate sample space and an accurate probability on this sample space.*

Example B.15 (T-period binomial model)**.** *For the binomial model with* T *periods, consider the sample space described in Example B.7. Recall from Section 2.3 that in the binomial model with* T *periods, the price* S_t *of the asset at time* t *is a random variable that takes values* $\{S_0 u^k \ell^{t-k} : k = 0, ..., t\}$*. Thus, one can say that under risk-neutral probability, we have* $\hat{\mathbb{P}}^\pi(S_t = S_0 u^k \ell^{t-k}) = \binom{t}{k}(\hat{\pi}_u)^k (\hat{\pi}_\ell)^{t-k}$*. This is because exactly* $\binom{t}{k}$ *of outcomes in the sample space lead to* $S_t = S_0 u^k \ell^{t-k}$*, and each outcome in the sample space has probability* $(\hat{\pi}_u)^k (\hat{\pi}_\ell)^{t-k}$*. Under physical probability, see 2.4.1; the probability of* $S_t = S_0 u^k \ell^{t-k}$ *changes to* $\binom{t}{k} p^k (1-p)^{t-k}$*.*

Remark B.2 (Random walk as a corner stone of financial models)**.** *A binomial model is related to the random walk in Example B.13 through taking logarithm. If* $V_t = \ln(S_t)$ *then*

V_t takes values $\ln(S_0) + k\ln(u) + (t-k)\ln(\ell)$ for $k = 0, ..., t$. In other words, V_t is the position of a generalized random walk after t rounds, starting at $\ln(S_0)$ which moves to $\ln(u)$ or $\ln(\ell)$ in each round, respective to the outcomes of a coin. If $S_0 = 1$, $u = e$ and $\ell = e^{-1}$, then the random walk V_t is the standard random walk W_t described in Example B.13. Otherwise, $V_t = \ln(S_0) + \mu t + \sigma W_t$ where $\mu = \frac{\ln(u)+\ln(\ell)}{2}$ and $\sigma = \frac{\ln(u)-\ln(\ell)}{2}$. In other words,

$$S_t = S_0 \exp\left(\mu t + \sigma W_t\right).$$

Example B.16. *Two random variables representing the payoff of two risky assets in the Arrow-Debreu market model that also includes a zero bond with yield $R = 0$, is shown in Figure B.2. From Exercise 2.1.4, we know that all risk-neutral probabilities are given by $\hat{\pi} = (t/2, (1-t)/2, (1-t)/2, t/2)^{\mathsf{T}}$ with $t \in (0, 1)$. The conditional probability of $X = y$*

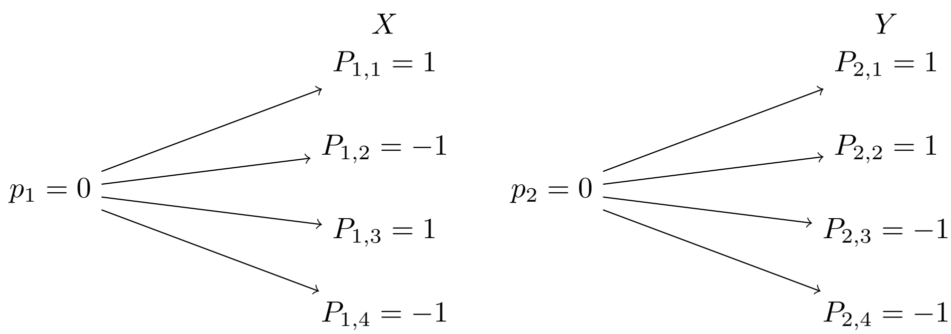


Figure B.2: Example B.16

given $Y = x$, for $x = \pm 1$ and $y = \pm 1$, is found below.

$$\begin{aligned}
\mathbb{P}(X = 1 | Y = 1) &= \frac{\mathbb{P}(X = 1 \& Y = 1)}{\mathbb{P}(Y = 1)} = \frac{\hat{\pi}_1}{\hat{\pi}_1 + \hat{\pi}_2} = t \\
\mathbb{P}(X = -1 | Y = 1) &= \frac{\mathbb{P}(X = -1 \& Y = 1)}{\mathbb{P}(Y = 1)} = \frac{\hat{\pi}_2}{\hat{\pi}_1 + \hat{\pi}_2} = 1 - t \\
\mathbb{P}(X = 1 | Y = -1) &= \frac{\mathbb{P}(X = 1 \& Y = -1)}{\mathbb{P}(Y = -1)} = \frac{\hat{\pi}_3}{\hat{\pi}_3 + \hat{\pi}_4} = 1 - t \\
\mathbb{P}(X = -1 | Y = -1) &= \frac{\mathbb{P}(X = -1 \& Y = -1)}{\mathbb{P}(Y = -1)} = \frac{\hat{\pi}_4}{\hat{\pi}_3 + \hat{\pi}_4} = t
\end{aligned} \tag{B.2}$$

Independence

Recall that the conditional probability of event B given that event A has occurred is defined by

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

The conditional probability gives birth to the important notion of independence. Two events A and B are called *independent* if

$$\mathbb{P}(B \mid A) = \mathbb{P}(B),$$

or equivalently

$$\mathbb{P}(A \mid B) = \mathbb{P}(A).$$

It is easier to write independence as

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

Two random variables X and Y are called independent if each event related to X is independent of each event related to Y, i.e.

For all x in values of X and all y is values of Y, we have $$\mathbb{P}(X = x,\ Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

Exercise B.5. *Show that two events A and B are independent if and only if the indicator random variables 1_A and 1_B are independent.*

In modeling random experiments, independence is a common-sense knowledge or an assumption inside a model. For example in a random walk, the outcomes of two different rounds are assumed independent, because of the belief that two flips of a coin are independent trials. As a result, two random variables W_5 and $W_8 - W_5$ are independent.

Exercise B.6. *Show that in a random walk, W_5 and $W_8 - W_5$ are independent provided that the outcome of each round is independent of other rounds.*

Defining independence for more than two events (equivalently random variables) is a little tricky. We call X independent of the sequence of random variables $X_1, X_2, ..., X_n$ if

$$\mathbb{P}(X = x,\ X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = \mathbb{P}(X = x)\mathbb{P}(X_1 = x_1, X_2 = x_2, ..., X_n = x_n).$$

This indicates that any event related to the values of X is independent of any event related to the values of $X_1, X_2, ..., X_n$, i.e.

$$\mathbb{P}(X \in A,\ (X_1, X_2, ..., X_n) \in B) = \mathbb{P}(X \in A)\mathbb{P}((X_1, X_2, ..., X_n) \in B).$$

A finite sequence $X_1, X_2, ..., X_n$ is called independent sequence of random variables if for each i, X_i and $\{X_j : j \neq i\}$ are independent.

As observed in the following exercise, to show the independence of a sequence of random variables, it is not enough to check that each pair of random variables X_i and X_j are independent.

Exercise B.7. *In two consecutive flips of a fair coin, let A be the event that the first flip turns heads, B be the event that the second flip turns heads and C be the event that only one of the flips turns heads. Show that A, B, and C are not an independent sequence of random variables, but they are pairwise independent.*

An equivalent definition of independence of a sequence of random variables is as follows. A finite sequence $X_1, X_2, ..., X_n$ is called independent if for all subsets $A_1, ..., A_n$ of values of $X_1, ..., X_n$, respectively, we have

$$\mathbb{P}(X_1 \in A_1, ..., X_n \in A_n) = \mathbb{P}(X_1 \in A_1)...\mathbb{P}(X_n \in A_n). \tag{B.3}$$

This can also be extended to an infinite sequence of random variables. An infinite sequence $X_1, X_2, ...$ is called independent if each finite subset $\{X_{i_1}, ..., X_{i_n}\}$ makes an independent sequence. Having defined the notion of independence for a sequence of random variables, we can now properly define a random walk, as the previous definition in Example B.13 is more heuristic than rigorous.

Definition B.2. *Let $\xi_1, \xi_2, ...$ be a sequence of independent and identically distributed (i.i.d.) random variables such that $\mathbb{P}(\xi_i = 1) = p$ and $\mathbb{P}(\xi_i = -1) = 1 - p$ for some $p \in (0, 1)$. For $x \in \mathbb{Z}$, the sequence $W_0 = x, W_1, W_2, ...$ with*

$$W_n := x + \sum_{i=1}^{n} \xi_i \quad \text{for } \ n \geqslant 1,$$

is called a random walk. When $p = \frac{1}{2}$, we call it a symmetric random walk; otherwise, it is called a biased random walk.

Let X be a random variable on a discrete sample space Ω with probability vector $\pi(\omega)$ for each $\omega \in \Omega$. Then, the *expectation* or *expected value* of X is defined by

$$\mathbb{E}[X] := \sum_{\omega \in \Omega} X(\omega)\pi(\omega). \tag{B.4}$$

If a random variable X takes values $x_1, x_2, ...$ with probabilities $p_1, p_2, ...$, respectively, then the expectation of X can equivalently be given by

$$\mathbb{E}[X] := \sum_{i=1}^{\infty} x_i p_i.$$

In particular, if values of X are finitely many $x_1, x_2, ..., x_n$ with probabilities $p_1, p_2, ..., p_n$, respectively, then the expectation of X is equivalently given by

$$\mathbb{E}[X] := \sum_{i=1}^{n} x_i p_i. \tag{B.5}$$

The advantage of (B.5) over (B.4) is that a sample space can be very large while the random variable only takes small number of values. For example, in Example B.9, a single-asset T-period binomial model generates a sample space of all paths of the asset price of size 2^T, while the values of random variable S_t are only $T+1$.

By straightforward calculations, the expectation of a function $f(x)$ of X is given by

$$\mathbb{E}[f(X)] = \sum_{i=1}^{\infty} f(x_i)p_i.$$

The variance of a random variable is defined by

$$\text{var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

One of the important result in the expectation of random variables is the Jensen inequality for convex functions.

Corollary B.1. *Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that hosts a $\mathbb{R}^d$-valued random variable X and let $f : \mathbb{R}^d \to \mathbb{R}$ be convex function f. Then, we have*

$$f(\mathbb{E}[X]) \leqslant \mathbb{E}[f(X)],$$

provided that both of the expectations exist.

Proof. By Corollary A.1, there exists a vector $u \in \mathbb{R}^d$ such that $f(\mathbb{E}[X]) + (X - \mathbb{E}[X]) \cdot u \leqslant f(X)$. By taking the expected value from both sides, we obtain the desired result. □

The Jensen inequality is reduced to the definition of convexity (A.2) when X is a random variable with two values x_1 and x_2. More precisely, if $\mathbb{P}(X = x_1) = \lambda$ and $\mathbb{P}(X = x_2) = 1-\lambda$, we have

$$f(\mathbb{E}[X]) = f(\lambda x_1 + (1 - \lambda x_2)) \leqslant \lambda f(x_1) + (1 - \lambda) f(x_2) = \mathbb{E}[f(X)].$$

Let Y be another random variable with values $y_1, ..., y_m$. To write the expectation of a function $f(x, y)$ of two random variables X and Y, we need to know the joint (mutual) probabilities of the pair (X, Y), i.e.,

$$p_{i,j} := \mathbb{P}(X = x_i,\ Y = y_j) \quad \text{for} \quad i = 1, ..., n, \quad \text{and} \quad j = 1, ..., m.$$

Notice that although $\mathbb{P}(X = x_i)$ and $\mathbb{P}(Y = y_j)$ are positive, $p_{i,j}$ can be zero, which means that if a value x_i is realized for the random variable X, then y_j cannot be realized for Y, and vice versa. Then, we define the expected value of $f(X, Y)$ by

$$\mathbb{E}[f(X, Y)] := \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f(x_i, y_j) p_{i,j}.$$

With the above definition, one show the Cauchy-Schwartz inequality.

Theorem B.1. *For two random variables X and Y we have*

$$\mathbb{E}[XY] \leqslant \sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}.$$

The equality holds if and only if X and Y are linearly dependent, i.e., if $aX + bY + c = 0$ for some constants $a, b, c \in \mathbb{R}$ such that at least one of them is nonzero.

The covariance between two random variables is defined by

$$\operatorname{cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

By applying the Cauchy-Schwartz inequality, we obtain

$$\operatorname{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \leqslant \sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2]}\sqrt{\mathbb{E}[(Y - \mathbb{E}[Y])^2]}.$$

In the above, the equality holds if and only if X and Y are linearly dependent. The correlation between two random variables is defined by

$$\operatorname{cor}(X, Y) := \frac{\operatorname{cov}(X, Y)}{\sqrt{\operatorname{var}(x)\operatorname{var}(Y)}}.$$

The Cauchy-Schwartz inequality shows that $-1 \leqslant \operatorname{cor}(X, Y) \leqslant 1$, and either of the equalities holds if and only if if $aX + bY + c = 0$ for some constants $a, b, c \in \mathbb{R}$ such that at least one of them is nonzero. In the case of equality, when $ab < 0$ (respectively $ab > 0$), then $\operatorname{cor}(X, Y) = 1$ (respectively $\operatorname{cor}(X, Y) = -1$).

We can define conditional expectation similarly by replacing the probabilities $p_{i,j}$ with the conditional probabilities

$$p_{i|j} := \mathbb{P}(X = x_i \mid Y = y_j) = \frac{p_{i,j}}{p_j^Y},$$

with $p_j^Y = \mathbb{P}(Y = y_j)$. More precisely,

$$\mathbb{E}[f(X, Y) \mid Y = y_j] := \sum_{i=1}^{\infty} f(x_i, y_j) p_{i|j} = \frac{1}{p_j^Y} \sum_{i=1}^{\infty} f(x_i, y_j) p_{i,j}. \tag{B.6}$$

Notice that if $\mathbb{P}(Y = y) = 0$, then $\mathbb{E}[f(X, Y) \mid Y = y]$ in (B.6) is not defined. However, we can define function $h : y \mapsto \mathbb{E}[f(X, Y) \mid Y = y]$ on the set of values of the random variable Y. This in particular helps us to define the random variable

$$\mathbb{E}[f(X, Y) \mid Y] := h(Y).$$

Remark B.3. *Notice the difference between* $\mathbb{E}[f(X,Y) \mid Y = y_j]$, $\mathbb{E}[f(X,Y) \mid Y = y]$, *and* $\mathbb{E}[f(X,Y) \mid Y]$. $\mathbb{E}[f(X,Y) \mid Y = y_j]$ *is a real number,* $\mathbb{E}[f(X,Y) \mid Y = y]$ *is a real function on variable* y*, and finally* $\mathbb{E}[f(X,Y) \mid Y]$ *is a random variable.*

Example B.17. *With regard to the above remark, we find* $\mathbb{P}(X = 1|Y)$, $\mathbb{P}(X = -1|Y)$, *and* $\mathbb{E}[X|Y]$ *in Example B.16. It follows from (B.2) that* $h(y) = \mathbb{P}(X = 1|Y = y)$ *is given by*

$$h(y) = \begin{cases} t & \text{when } \; y = 1 \\ 1-t & \text{when } \; y = -1 \end{cases}$$

Thus, $\mathbb{P}(X = 1|Y) = t\delta_1(Y) + (1-t)\delta_{-1}(Y)$. *Here,* $\delta_x(y)$ *is* 1 *when* $y = x$ *and* 0 *otherwise. Similarly,* $\mathbb{P}(X = -1|Y) = (1-t)\delta_1(Y) + t\delta_{-1}(Y)$. *Finally,*

$$\begin{aligned} \mathbb{E}[X|Y] &= \mathbb{P}(X = 1|Y) - \mathbb{P}(X = -1|Y) = t\delta_1(Y) + (1-t)\delta_{-1}(Y) - (1-t)\delta_1(Y) - t\delta_{-1}(Y) \\ &= (2t-1)\delta_1(Y) + (1-2t)\delta_{-1}(Y). \end{aligned}$$

As a particular case, when $f(x,y) = x$, we have

$$h(y) = \mathbb{E}[X \mid Y = y] \quad \text{and} \quad \mathbb{E}[X \mid Y] = h(Y).$$

Corollary B.2. *If* X *and* Y *are random variables and* f *is a real function, then we have*

$$\mathbb{E}[f(Y)X \mid Y] = f(Y)\mathbb{E}[X \mid Y].$$

The following proposition, which is a direct result of (B.6), explains a very important property of independent random variables.

Proposition B.1. X *and* Y *are independent if and only if, for any real function* $f(x,y)$ *of* X *and* Y*, we have*

$$\mathbb{E}[f(X,Y) \mid Y = y] = \mathbb{E}[f(X,y)] \; \text{ for all } y \text{ in the set of values of } Y.$$

Corollary B.3. *If* X *and* Y *are independent and* f *is a real function, then we have*

$$\mathbb{E}[f(X) \mid Y] = \mathbb{E}[f(X)].$$

One of the important properties of conditional expectation is the *tower property*, which is presented in the next proposition.

Proposition B.2 (Tower property of conditional expectation)**.** *Let* X*,* Y *and* Z *be random variables. Then,*

$$\mathbb{E}[\mathbb{E}[X \mid Y, Z] \mid Y] = \mathbb{E}[X \mid Y].$$

In particular,

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X].$$

The following proposition provides an equivalent representation for independence through the conditional expectation.

B.2 General probability spaces

Some random experiments can generate uncountable number of outcomes, e.g., choosing a point in the unit interval $[0, 1]$ or choosing a chord of a unit circle. In such cases, the definitions in Section B.1 don't make sense; e.g., the summation in (B.4) relies on the countability of the sample space. The complexity of uncountable sample spaces is twofold. First, the basic definitions, such as the expected value in (B.4), use summation, which relies on countability of values. In Section B.2, we will consider the continuous random variables in which the summations can be replaced by an integral, such as Example B.18. Secondly, there is mathematical challenge in defining an event in an uncountable sample space, which we will describe later in this section.

Example B.18. *In choosing a random number uniformly inside* $[0, 1]$*, the probability of the following events can be calculated by integration.*

i) The probability that the random number is in $(a, b) \subseteq [0, 1]$ *equals* $\int_a^b f(x)dx = b - a$*. Here* $f(x) = 1_{[0,1]}(x)$ *is called the uniform probability density function. This probability is the area enclosed by* $f(x)$ *and* x *axis above interval* (a, b)

ii) The probability that the random number is $.5$ *equals 0 as the area described in (i) is zero.*

iii) The probability that the random number is a rational number is 0. Rational numbers are countable so the probability of this event is sum of the probability of each rational number, which is summation of countable number of zeros.

iv) The probability that the random number is an irrational number is 1*. This probability is the complement of the probability of rationals, which is* 0*.*

The random experiment in the above example is describing a continuous random variable. In such random variables, the probability of events and expected values of random variables can be calculated by integration. However, in general, there can be random variables that are neither continuous nor discrete. In that case, concepts such as expectation and conditional expectation should be defined differently; they requires advanced techniques from measure theory. Measure theory was gradually developed as a theory for integration by several mathematicians such as Émile Borel (1871–1956), Henri Lebesgue (1875–1941), Johann Radon (1887–1956), and Maurice Fréchet (1878–1973). However, Andrey Kolmogorov (1903–1987) was the first who noticed that this theory can be used as a foundation for probability theory; a probability is a *nonnegative finite measure* (normalized to mass one) and the expectation is an *integral with respect to that measure.*

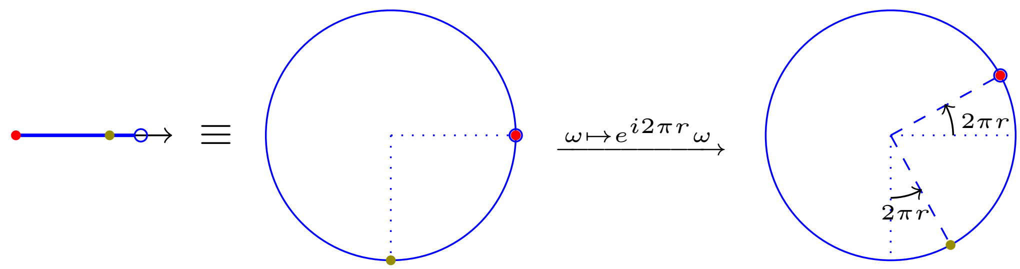


Figure B.3: Description of the transformations in Example B.19

Another difficulty that arises in uncountable sample spaces is the meaning of an event. In discrete space, any subset of the sample space is an event and any function from sample space to $\mathbb{R}^d$ is a random variable. However, in uncountable sample spaces, there are some subsets of the sample space for which no value as a probability can be assigned. Therefore, an event and a random variable should be defined in a proper way by using the concept of *measurability*; only *measurable* sets are events, and measurable real functions make random variables. For more extensive study on the concept of probability measures and measurable functions see [1] or [4]. Here, we address the issue briefly. First, we provide an example to show how a *nonmeasurable* set looks like.

Example B.19 (A nonmeasurable set exists)**.** *Consider $\Omega = [0,1)$, identified by a unit circle through the transformation $\omega \mapsto e^{i2\pi\omega}$ shown in Figure B.3. Here, $i = \sqrt{-1}$. Now, for any rational number r in $[0,1)$, consider the rotational transformation on the circle given by $e^{i2\pi r} : x = e^{i2\pi\omega} \mapsto e^{i2\pi r}x$. If we consider the uniform probability on $\Omega = [0,1)$, the induced probability on the circle is also uniform. The uniform probability on the circle is invariant under rotations;*

$$\mathbb{P}(B) = \mathbb{P}(e^{i2\pi r}B) \quad \textit{for any subset } B \textit{ on the circle.}$$

*For $x = e^{i2\pi\omega}$, define the orbit of x by $O(x) := \{xe^{i2\pi r} : r \in \mathbb{Q} \cap [0,1)\}$. Since $O(x)$ is countable and the unit circle is uncountable, there are uncountably many disjoint orbits with a union equal to the unit circle. Let A be a set that has **exactly** one point from each distinct orbit. Then, one can see that the countable **disjoint** union*

$$\bigcup_{r \in \mathbb{Q} \cap [0,1)} e^{i2\pi r}(A)$$

covers the unit circle. Here, $e^{i2\pi r}(A)$ means the image of A under the rotational transformation $e^{i2\pi r}$. This is because: (1) If $x \in e^{i2\pi r}(A) \cap e^{i2\pi r'}(A)$ for $0 \leqslant r < r' < 1$ rationals , we have then both $e^{i2\pi(1-r)}x$[10] and $e^{i2\pi(1-r')}x$ belonging to A and being members of $O(x)$,

[10] If $y = e^{i2\pi r}x$, then $x = ye^{-i2\pi r}$. But, since $-r$ is not in $[0,1)$ and $1 = e^{i2\pi}$, we can write $x = ye^{i2\pi(1-r)}$.

which contradicts the choice of A; and (2) For each x, there is a member $y \in O(x) \in A$; there exists a rational r such that $y = e^{i2\pi r}x$. Thus, $x = e^{i2\pi(1-r)}y \in e^{i2\pi(1-r)}(A)$. Therefore, we have

$$\mathbb{P}\left(\bigcup_{r\in\mathbb{Q}\cap[0,1)} e^{i2\pi r}(A)\right) = \sum_{r\in\mathbb{Q}\cap[0,1)} \mathbb{P}\left(e^{i2\pi r}(A)\right).$$

It follows from the invariance of uniform probability under rotation that the probability of A is the same as the probability of $e^{i2\pi r}(A)$ for all $r \in \mathbb{Q} \cap [0,1)$, if such a probability exists.

$$\mathbb{P}\left(\bigcup_{r\in\mathbb{Q}\cap[0,1)} e^{i2\pi r}(A)\right) = \sum_{r\in\mathbb{Q}\cap[0,1)} \mathbb{P}(A).$$

Now if A has probability 0 under uniform measure, then so does the countable union $\bigcup_{r\in\mathbb{Q}\cap[0,1)} e^{i2\pi r}(A)$, which is a contradiction, because the union $\bigcup_{r\in\mathbb{Q}\cap[0,1)} e^{i2\pi r}(A)$ makes up the whole circle, which has probability 1. Otherwise, if A has a nonzero probability, then the sum on the left-hand side is infinite. Either way, it is impossible to assign a probability to A; A is nonmeasurable.

Uncountable sample spaces appear in probability even when the experiment is a discrete one. For instance, consider the following version of the law of large numbers:

Theorem B.2 (Law of Large Numbers)**.** *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of independent random variables that take values 0 or 1 with equal likelihood. Then,*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{\infty} X_n = \frac{1}{2}$$

except on a set of outcomes with probability zero.

In this theorem, we consider a sample space Ωthat consists of all sequences $\omega = (\omega_1, ...)$ such that $\omega_n = H$ or T based on the outcome of the nth coin flip. The random variable X_n is defined by 1 if $\omega_n = H$ and 0 if $\omega_n = T$. This sample space is uncountable! To see this, we construct a map which takes it into the interval $[0,1)$:

$$\omega \mapsto \sum_{n=1}^{\infty} \frac{X_n}{2^n}.$$

The right-hand-side above is the binary representation of a number in $[0,1)$. Therefore, the sample space Ω is uncountable and not significantly different from $[0,1)$.

Remark B.4. *Notice that some numbers can have two different binary representations. For example, $.5 = \frac{1}{2} = \sum_{n=2}^{\infty} \frac{1}{2^n}$. However, the even in Ω that corresponds to all these numbers has probability zero.*

σ-field of events

One of the important notions in the probability theory is dealing with the definition of an event, a random variable, and the information related to them, which is presented in the following definition.

Definition B.3 (σ-field)**.** *For a (possibly uncountable) sample space Ω, the set of events should use the notion of σ-field from measure theory. A σ-field $\mathcal{F}$ is a collection of subsets of Ω that satisfies*

a) $\varnothing$ and $\Omega \in \mathcal{F}$.

b) If $A \in \mathcal{F}$, then $A^{\complement} \in \mathcal{F}$.

c) For a sequence $\{A_n\}_{n=0}^{\infty} \subseteq \mathcal{F}$, $\bigcup_{n=0}^{\infty} A_n \in \mathcal{F}$.

A sample space Ω along with a σ-field is called a measurable space.

Usually, a σ-field can be specified by defining a set of elementary events and expanding this set to a whole σ-field, which contains all events. For instance, in a discrete sample space $\Omega = \{\omega_1, \omega_2, ...\}$, one can define the elementary events to be $E_n = \{\omega_n\}$. Then, any event is a countable union of elementary events.

In uncountable sample spaces, the methodology of working with elementary events is inevitable. We consider a set of elementary events $\{E_n : n \geqslant 1\}$ and construct the **smallest** σ-field that contains all these elementary events, namely the *σ-field generated by elementary sets* $\{E_n : n \geqslant 1\}$ denoted by $\sigma(E_n : n \geqslant 1)$.

Example B.20. *Consider $\Omega = \mathbb{R}$ and the elementary events given by the singleton sets $E_n := \{x_n\}$ for $n = 1, 2, ...$ Then, by property (b) the complement of singleton elementary events, $E_n^{\complement} = \mathbb{R} \backslash \{x_n\}$ is in the σ-field generated by E_ns. Also, any finite or countable number of x_ns and the complement of it make events in this σ-field. For instance, the σ-field generated by $\{1\}$ is*

$$\Big\{\varnothing, \mathbb{R}, \{1\}, \mathbb{R}\backslash\{1\}\Big\}.$$

The σ-field generated by $\{1\}$ and $\{2\}$ is

$$\Big\{\varnothing, \mathbb{R}, \{1\}, \{2\}, \{1,2\}, \mathbb{R}\backslash\{1\}, \mathbb{R}\backslash\{2\}, \mathbb{R}\backslash\{1,2\}\Big\}.$$

Example B.21. *The σ-field generated by a single event, A has four distinct events: $\varnothing$, Ω, A, and $A^{\complement}$. For example, the σ-field generated by the set of all rational numbers in $\mathbb{R}$ consist of $\varnothing$, $\mathbb{R}$, the set of rational numbers, and the set of irrational numbers.*

In the discrete example from the last paragraph, the σ-field is the same as the set of all subsets of Ω. In Chapter 2, the sample space is $\Omega = (\mathbb{R}^{d+1})^T$. We consider each elementary event given by an open set in $(\mathbb{R}^{d+1})^T$. Then, we work with the *Borel σ-field* $\mathcal{B}(\Omega)$, the σ-field generated by all open subsets of Ω. This method can be generalized to all samples spaces that are topological spaces. The Borel σ-field in $\mathbb{R}$, $\mathcal{B}(\mathbb{R})$, is generated by the set of all open intervals (a, b). It is not hard to see that all intervals, closed or semi-closed, are also included in the Borel σ-field. For instance,

$$(a, b] = \bigcap_{n \geqslant 1} \Big(a, b + \frac{1}{n}\Big).$$

Also, all single points are in the Borel σ-field; $\{b\} = \bigcup_{n \geqslant 1} [b, b + \frac{1}{n})$. Also, $\mathcal{B}(\mathbb{R})$ can be generated by half-open intervals of the form $(a, b]$, because

$$(a, b) = \bigcup_{n=1}^{\infty} \Big(a, b - \frac{1}{n}\Big].$$

Exercise B.8. *Show that $\mathcal{B}(\mathbb{R})$ can be generated by half-open intervals of the form $[a, b)$.*

Exercise B.9. *Check whether the following sets make a σ-field. In case they are not a σ-field, find the missing events that makes them a σ-field.*

i) $\big\{\mathbb{R}, \{a\}, \{b\}, \{a, b\}, \mathbb{R}\backslash\{b\}, \mathbb{R}\backslash\{a\}\big\}$ *where a and b are two distinct real numbers.*

ii) $\big\{\varnothing, [0, 1], [0, \frac{2}{3}), (\frac{1}{3}, 1]\big\}$.

iii) $\big\{\varnothing, \mathbb{R}\big\}$.

The Borel σ-field on $[0, 1]$, $\mathcal{B}([0, 1])$, is the σ-field generated by the intervals of the form (a, b), $[0, b)$, or $(a, 1]$, where $0 \leqslant a < b \leqslant 1$. Equivalently, $\mathcal{B}([0, 1])$ is generated by the half-open intervals of the form $[a, b)$ or $(a, b]$.

Example B.22. *The Borel σfield $\mathcal{B}([0, 1])$ can be generated by either of the following elementary events:*

i) The set of intervals of the form (a, b), $[0, b)$, or $(a, 1]$, where $0 \leqslant a < b \leqslant 1$,

ii) The set of intervals of the form $[a, b)$ or $(a, b]$, where $0 \leqslant a < b \leqslant 1$, and

iii) The set of intervals of the form $[a, b]$, where $0 \leqslant a < b \leqslant 1$.

To show this, we need to demonstrate that each elementary event of one set can be generated by the elementary events from the other sets. In the set of elementary events in (i), the interval (a, b) can be written as

$$(a, b) = \bigcup_{n=1}^{\infty} \Big(a, b - \frac{1}{n}\Big].$$

Therefore, (a,b) can be generated by the events that are described in (ii). $[0,b)$ and $(a,1]$ already belong to the set that are described in (ii). Therefore, all elementary events in (i) can be generated by elementary events in (ii).

In addition, we have

$$[a,b)=\bigcup_{n=1}^{\infty}\left[a,b-\frac{1}{n}\right] \text{ and } (a,b]=\bigcup_{n=1}^{\infty}\left[a+\frac{1}{n},b\right],$$

which implies that all elementary events in (ii) can be generated by elementary events in (iii).

Finally, we have

$$[a,b]=\bigcap_{n=1}^{\infty}\left(a-\frac{1}{n},b+\frac{1}{n}\right),\ (a,1]=\bigcap_{n=1}^{\infty}\left(a-\frac{1}{n},1\right],\ \text{and } [0,b)=\bigcap_{n=1}^{\infty}\left[0,b+\frac{1}{n}\right).$$

Therefore, all elementary events in (iii) can be generated by elementary events in (i).

The Borel σ-field on the half-line $[0,\infty)$, $\mathcal{B}([0,\infty))$, is the σ-field generated by the sets of the form (a,b) or $[0,b)$, where $0 \leqslant a < b$.

Exercise B.10. *Show that $\mathcal{B}([0,\infty))$ can be generated by either of the following elementary events:*

- *i) The set of intervals of the form (a,b) or $[0,b)$, where $0 \leqslant a < b < \infty$, and*
- *ii) The set of intervals of the form $[a,b)$, where $0 \leqslant a < b < \infty$.*

Hint: Show that each elementary event of one set can be generated by the elementary events from the other set.

The following example is a useful use of elementary events in analyzing an event of interest and a method of finding the probability of such an event.

Example B.23. *In Example B.13, let the event of ruin A be the collection of all outcomes for which the gambler's wealth eventually hits zero. For instance, if $W_0 = 10$, any outcome with ten loses in a row since the start of the game is in this set. Also, any outcome with the number of losses at some time larger than the number of wins plus ten is in the ruin event A. The ruin event is uncountable. We attempt to find the probability of this event later in Example B.25, namely the ruin probability. Before that, we need to write this event as a union of disjoint simpler events for which the probability can easily be evaluated.*

$$A=\dot{\bigcup_{n=1}^{\infty}} A_n,$$

where A_n is the event that the ruin happens exactly at the nth round. Notice that A_n is made of all outcomes such that (1) the nth round is a lost round, (2) in the first $n-1$st rounds there are $x-1$ lost rounds more than the number of wins. Therefore, if m is the number of won rounds in the $n-1$st rounds, we must have $2m+x-1=n-1$, or $m=\frac{n-x}{2}$. As a result, A_n is a nonempty event if and only if $n-x$ is an even number. Therefore, one can write

$$A = \dot{\bigcup_{m=1}^{\infty}} A_{x+2m}.$$

Exercise B.11. *Recall the sample space given by (B.1) for a random walk. Show that in Example B.23 the event of ruin is uncountable.*

Many complicated sets can be found in the Borel σ-field. For more discussion of Borel sets, see [3, Chapter 7]. However, in most cases, it is enough to work with the elementary events that build the σ-field.

Probability via measure theory

Given a σ-field of events, we can define a probability measure in the most general form.

Definition B.4. *A measure space is a $(\Omega, \mathcal{F})$ such that Ω is a sample space and $\mathcal{F}$ is a σ-field on Ω. A measure on a measure space $(\Omega, \mathcal{F})$ is a function from $\mathbb{P}: \mathcal{F} \to \mathbb{R} \cup \{\infty\}$ such that for any sequence $\{A_n\}_{n=0}^{\infty} \subseteq \mathcal{F}$ of **disjoint** events,*

$$\mathbb{P}\left(\bigcup_{n=0}^{\infty} A_n\right) = \sum_{n=0}^{\infty} \mathbb{P}(A_n).$$

A probability measure is a measure with the following properties.

a) For any $A \in \mathcal{F}$ then $\mathbb{P}(A) \geqslant 0$.

b) $\mathbb{P}(\Omega) = 1$.

Notice that the above definition leads to many important properties of the probability:

a) $\mathbb{P}(A^{\complement}) + \mathbb{P}(A) = \mathbb{P}(A^{\complement} \cup A) = \mathbb{P}(\Omega) = 1$. Therefore, $\mathbb{P}(A^{\complement}) = 1 - \mathbb{P}(A)$.

b) $\mathbb{P}(\varnothing) = 1 - \mathbb{P}(\Omega) = 0$.

c) **Continuity of probability.** For any shrinking sequence of events $A_1 \supseteq A_2 \supseteq \cdots$, we have

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \left(\bigcap_{n=0}^{\infty} A_n\right)$$

Exercise B.12. *Show (c) in the above.*

Triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space. To define any other measure $\mathbb{P}$ on a measurable space $(\Omega, \mathcal{F})$, it is sufficient to define it on the elementary sets that generate the σ-field $\mathcal{F}$. For example, to define a uniform probability measure on $[0, 1]$ on the Borel σ-field $\mathcal{B}([0, 1])$, one only needs to define

$$\mathbb{P}((a, b)) = b - a, \quad \text{for} \quad (a, b) \in [0, 1].$$

Since probability of singleton is zero, we have

$$\mathbb{P}([0, b)) = \mathbb{P}(\{0\}) + \mathbb{P}((0, b)) = 0 + b.$$

Similarly, $\mathbb{P}((a, 1]) = 1 - a$.

Here is a more complicated example. The random variable in this example is not discrete or continuous.

Example B.24. *We choose a random number in $[0, 1]$ in the following way. We flip a fair coin. If the coin lands heads, we choose a number uniformly in $[0, 1]$. If the coin lands tails, we choose $\sqrt{2}/2$. We define the corresponding probability measure on $\mathcal{B}([0, 1])$ by defining it only on the elementary events:*

$$\mathbb{P}((a, b)) = \begin{cases} \frac{1}{2}(b - a) & \text{if} \quad \sqrt{2}/2 \notin (a, b) \in [0, 1] \\ \frac{1}{2}(b - a) + \frac{1}{2} & \text{if} \quad \sqrt{2}/2 \in (a, b) \in [0, 1] \end{cases}.$$

Notice that although the above probability is only defined on open intervals, one can find probability of other events. For instance, $\mathbb{P}(\{\sqrt{2}/2\}) = \lim_{n \to \infty} \left(\sqrt{2}/2 - \frac{1}{n}, \sqrt{2}/2 + \frac{1}{n}\right) = \frac{2}{n} + \frac{1}{2} = \frac{1}{2}$.

Example B.25. *In this example, we follow up on the probability of ruin in Example B.23. Recall that the event of ruin is represented as the union of disjoint events:*

$$A = \dot{\bigcup_{m=0}^{\infty}} A_{x+2m}.$$

Therefore, by the definition of probability

$$\mathbb{P}(A) = \sum_{m=0}^{\infty} \mathbb{P}(A_{x+2m}).$$

If the chance of winning any round is p, then

$$\mathbb{P}(A_{x+2m}) = \binom{x + 2m - 1}{m} p^m (1 - p)^{x+m}.$$

Therefore,

$$\mathbb{P}(A) = \sum_{m=0}^{\infty} \binom{x+2m-1}{m} p^m (1-p)^{x+m}.$$

The above probability is not easy to calculate in this series form, and it is easier to find it by using a conditioning technique.

Example B.26. *In an infinite sequence of coin flips, we consider Ω to be the set of all sequences $(\omega_1, \omega_2, \cdots)$ such that $\omega_i = H$ or T for $n \geqslant 1$. A σ-field can be defined by specifying some elementary events. For a finite sequence $n_1, ...n_m$ of positive integers, an elementary event $E_{n_1,...n_m}$ is defined to be the set of all $(\omega_1, \omega_2, \cdots)$ such that $\omega_{n_j} = H$ for $j = 1, ..., m$. Then, on the σ-field generated by a fair coin, we have the corresponding probability measure given by $\mathbb{P}(E_{n_1,...n_m}) = \left(\frac{1}{2}\right)^m$. For an unfair coin with the probability of heads given by p, we have a different probability measure $\mathbb{Q}(E_{n_1,...n_m}) = p^m$. Based on the assignment of probability , we can determine the probability of all events in the σ-field generated by the elementary events. For instance, let $E_{m,-n}$ be the event that the nth flip is heads and the mth flip is tails. Then,*

$$E_m = E_{m,n} \cup E_{m,-n}.$$

Notice that $E_{m,n}$ and $E_{m,-n}$ are disjoint events. Thus, $\mathbb{P}(E_{m,n'}) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$. For an unfair coin, we have $\mathbb{Q}(E_{m,n'}) = p - p^2 = p(1-p)$. Recalling the notion of independence from (B.3), one can see that for $n \neq m$, E_n and E_m are independent events, under both measures $\mathbb{P}$ and $\mathbb{Q}$. More precisely, since $E_n \cap E_m = E_{m,n}$, we have

$$\mathbb{P}(E_{m,n}) = \frac{1}{4} = \mathbb{P}(E_m)\mathbb{P}(E_n) \quad \textit{and} \quad \mathbb{Q}(E_{m,n}) = p^2 = \mathbb{Q}(E_m)\mathbb{Q}(E_n).$$

Generally speaking, $E_{n_1} \cap \cdots \cap E_{n_m} = E_{n_1,...,n_m}$, and therefore the σ-field generated by the sequence $\{E_n : n \geqslant 1\}$ is the same as the σ-field generated by the elementary sets of the form $E_{n_1,...,n_m}$.

Recall that mapping

$$\omega = (\omega_1, \omega_2, \cdots) \mapsto \sum_{n=1}^{\infty} \frac{a_n}{2^n}$$

with $a_n = 1$ when $\omega_n = H$ and $a_n = 0$ when $\omega_n = T$ maps Ω to $[0, 1]$. Under this mapping, the elementary event E_n is mapped to

$$\bigcup_{i=0}^{2^{n-1}} \Big[\frac{2i+1}{2^n}, \frac{2i+2}{2^n}\Big).$$

Although tedious, one can find the exact intervals that make the image of $E_{n_1,...n_m}$ under this mapping and then use it to verify that the σ-field generated by these events is exactly *the Borel σ-field $\mathcal{B}([0,1])$. The following theorem is a helpful tool in that regard.*

Exercise B.13. *Show that any interval of the form $\left[\frac{k_1}{2^n}, \frac{k_2}{2^n}\right)$ with $0 \leqslant k_1 < k_2 \leqslant 2^n$ can be written by a combination of a finite union, intersection and complement of the sets that are image of the elementary events of the form E_n under the mapping described in Example B.26. Use this to show that the σ-field generated by the elementary events of the form E_n correspond to the Borel σ-field $\mathcal{B}([0,1])$.*

Given a probability measure $\mathbb{P}$, the events that have a probability of 0 under $\mathbb{P}$ are called *null events*. For instance, in choosing a point uniformly in $[0,1)$, the probability of any single point is 0; therefore, the singleton $\{x\}$ is a null event. In Example B.24, the only singleton that is not a null event is $\left\{\frac{\sqrt{2}}{2}\right\}$. The set of all rational numbers is a null event in both cases. In Example B.26, one can show that an event of a particular pattern appearing periodically along a sequence of coin flips is a null event. For instance, the event of the pattern THH appearing periodically along a sequence of coin flips is a null event.

Exercise B.14. *Show that the event of all outcomes that have the pattern THH (or any other particular pattern) appearing periodically along a sequence of coin flips is a null event. Use the same idea to extend the result: the event of all outcomes that have a periodic pattern is a null event. (Hint: the set of all patterns is a countable set.)*

Random variables

In this section, we explain the definition of a random variable. Random variables represent random quantities that are related to a random experiment. For example, in Example B.13, the value W_n of a random walk at time n is a random variable.

Definition B.5. *A random variable on a measure space $(\Omega, \mathcal{F})$ is a function $X : \Omega \to \mathbb{R}^d$ such that for any Borel set $A \in \mathcal{B}(\mathbb{R}^d)$, the inverse image of A under X, $X^{-1}(A)$, belongs to the set of events $\mathcal{F}$.*

In the sequel, we denote the inverse image of A under X, $X^{-1}(A)$, by $\{X \in A\}$. Since $\mathcal{B}(\mathbb{R}^d)$ is generated by open sets in $\mathbb{R}^d$, the condition "$\forall A \in \mathcal{B}(\mathbb{R}^d)$, $\{X \in A\} \in \mathcal{F}$" in Definition B.5 needs to hold only for all open sets instead of all Borel sets; $A \in \mathcal{B}(\mathbb{R}^d)$ for all open sets $A \subseteq \mathbb{R}^d$. If $\Omega = \mathbb{R}^d$ and $\mathcal{F} = \mathcal{B}(\mathbb{R}^d)$, a random variable is called a *Borel-measurable* function.

In practice, important random quantities are described by random variables. See the example below.

Example B.27. *In an infinite sequence of coin flips, Example B.26, we define the random variable X as follows. For $\omega = (\omega_1, \omega_2, \cdots) \in \Omega$*

$$X(\omega) = \sum_{n=1}^{\infty} \frac{a_n}{3^n}$$

such that $a_n = 0$ *if* $\omega_n = H$*, and* $a_n = 2$ *if* $\omega_n = T$*. Here,* $\sum_{n=1}^{\infty} \frac{a_n}{3^n} = 0.a_1a_2\cdots$ *is the ternary representation of a number in* $[0,1]$*.* X *maps* Ω *to* $[0,1]$*. To check if* X *is a random variable, we need to check that the inverse of an open interval is in the* σ*-field generated by the elementary sets* $A_{n_1,\ldots n_m}$ *described in Example B.26. We leave this task to the following exercise.*

Remark B.5. *The image of* X *in Example B.27 is called a Cantor set.*

Exercise B.15. *Show that the mapping* X *defined in Example B.27 is a random variable. Hint: Assume a ternary representation for the endpoints of the interval* (a,b)*.*

Let X be a random variable. The σ-field generated by X is the **smallest** σ-field that contains all events $\{X \in A\}$ for all Borel sets A, and is denoted by $\sigma(X)$. In other words, it is the $\sigma(X)$ that contains all events related to X. $\sigma(X)$ can equivalently be expressed as the smallest σ-field that contains all events $\{X \in A\}$ for all open sets A.

Independence

It is possible to define the notion of independence for two random variables X and Y by using the inverse images $\{X \in A\}$ and $\{Y \in B\}$.

Definition B.6. *A sequence of random variables* $\{X_n : n \geqslant 1\}$ *are called independent if for any sequence of Borel sets* $\{A_n : n \geqslant 1\} \subseteq \mathcal{B}(\mathbb{R})$*, the sequence of events* $\{X_n^{-1}(A_n) : n \geqslant 1\}$ *are independent.*

In particular, two random variables X *and* Y *are called independent if for any two Borel sets* A *and* B*,* $\{X \in A\}$ *and* $\{Y \in B\}$ *are independent,*

$$\mathbb{P}(X \in A \,\&\, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

The following proposition can be used as an alternative definition of independence of two random variables.

Proposition B.3. *Two random variables* X *and* Y *are independent if and only if for any two bounded Borel-measurable functions* $h_1(x)$ *and* $h_2(y)$ *we have*

$$\mathbb{E}[h_1(X)h_2(Y)] = \mathbb{E}[h_1(X)]\mathbb{E}[h_2(Y)]. \tag{B.7}$$

As a consequence of the above proposition, if X and Y are independent, then for any two bounded Borel-measurable functions $h_1(x)$ and $h_2(y)$, $h_1(X)$ and $h_2(Y)$ are also independent.

Example B.28. *In Example B.13, consider random variables* W_n *and* $W_m - W_n$ *with* $m > n$*. These random variables are independent. This is because,* $W_n = W_0 + \sum_{i=1}^{n} \xi_i$ *and* $W_m - W_n = \sum_{i=n+1}^{m} \xi_i$*, and the vectors* $(\xi_1, \ldots, \xi_n)$ *and* $(\xi_{n+1}, \ldots, \xi_m)$ *are independent. As a result* $h_1(\xi_1, \ldots, \xi_n) := W_0 + \sum_{i=1}^{n} \xi_i$ *and* $h_2(\xi_{n+1}, \ldots, \xi_m) = \sum_{i=n+1}^{m} \xi_i$ *are independent.*

One can assert the independence of a sequence of random variables in terms of the sequence of σ-fields generated by them.

Proposition B.4. *A sequence of random variables $\{X_n : n \geqslant 1\}$ are independent if any sequence of events $\{E_n : n \geqslant 1\}$ such that $E_n \in \sigma(X_n)$ for all $n > 1$ is an independent sequence.*

In particular, two random variables X and Y are independent if for any two events $F_1 \in \sigma(X)$ and $F_2 \in \sigma(Y)$, F_1 and F_2 are independent,

$$\mathbb{P}(F_1 \cap F_2) = \mathbb{P}(F_1)\mathbb{P}(F_2).$$

The above proposition gives rise to the notion of independent σ-fields of events.

Definition B.7. *A sequence of random variables σ-fields $\{\mathcal{F}_n : n \geqslant 1\}$ are called independent if any sequence of events $\{F_n : n \geqslant 1\}$ such that $F_n \in \mathcal{F}_n$ is an independent sequence.*

In particular, two σ-fields $\mathcal{F}$ and $\mathcal{G}$ are called independent if for any two events $F \in \mathcal{F}$ and $G \in \mathcal{G}$, we have

$$\mathbb{P}(F \cap G) = \mathbb{P}(F)\mathbb{P}(G).$$

Example B.29. *In Example B.13, consider the σ-field generated by the outcomes of the even rounds and the σ-field generated by the outcomes of the odd rounds; denote them by $\mathcal{F}$ and $\mathcal{G}$, respectively. Since, the outcomes of each round in independent others, $\mathcal{F}$ and $\mathcal{G}$ are independent σ-fields.*

Expected value and integration

Before Andrey Kolmogorov used the concept of a measure to define probability, measures were used to extend the notion of integration. Integration with respect to measures is also important in probability theory to define the expected value of random variables. Here, we define the expected value of a random variable by the integral of that random variable with respect to a probability measure in the most general form, which included uncountable sample spaces.

If a random variable X is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the expected value of X is defined as an integral of X with respect to measure $\mathbb{P}$, denoted by

$$\mathbb{E}[X] := \int X d\mathbb{P}.$$

The definition of the integral is a little cumbersome. Thus, we only outline the steps.

Step 1) For an indicator random variable $X = 1_A$ where A is an event in $\mathcal{F}$, we have

$$\mathbb{E}[X] := \int_A d\mathbb{P} = \mathbb{P}(A).$$

Step 2) For step random variable X, a finite linear combination of indicator random variables $X = \sum_{j=1}^{N} a_j 1_{A_j}$, we have

$$\mathbb{E}[X] := \sum_{j=1}^{N} a_j \mathbb{P}(A_j).$$

Step 3) If the random variable X is nonnegative,$X \geqslant 0$, then X can be approximated from below by an increasing sequence of step random variables; $X_n \uparrow X$. Then,

$$\mathbb{E}[X] := \sup_n \mathbb{E}[X_n].$$

Notice that one can show that the value $\mathbb{E}[X]$ is independent of the choice of the increasing sequence of step random variables and thus is well defined.

Step 4) For a general random variable X, we decompose X into positive and negative parts; $X = X_+ - X_-$ with $X_+ = \max 0, X$ and $X_- = \max 0, -X$. Then,

$$\mathbb{E}[X] := \mathbb{E}[X_+] - \mathbb{E}[X_-].$$

In Step (4) in the above, the expected value $\mathbb{E}[X]$ is ∞ when $\mathbb{E}[X_+] = \infty$ and $\mathbb{E}[X_-] < \infty$, and it is $-\infty$ when $\mathbb{E}[X_+] < \infty$ and $\mathbb{E}[X_-] = \infty$. However, $E[X]$ cannot be defined when we have both $\mathbb{E}[X_+] = \infty$ and $\mathbb{E}[X_-] = \infty$. This motivates the following definition.

Definition B.8. *A random variable is called integrable if* $\mathbb{E}[|X|] = \mathbb{E}[X_+] + \mathbb{E}[X_-] < \infty$.

The following example shows why we need to separate the positive part and the negative part of a random variable in Step 4 above.

Example B.30. *Consider the uniform probability measure on probability space given by* $\Omega = [-1, 1]^2$*, the two dimensional square, and a random variable* X *given by* $X(x, y) = \frac{xy}{(x^2+y^2)^2}$[11]*. The positive part of* X *is* $[0, 1]^2 \cup [-1, 0]^2$*. It is easy (by switching to polar coordinates) to see that*

$$\int_{[0,1]^2} X_+ d\mathbb{P} = \int_0^1 \int_0^1 \frac{xy}{(x^2 + y^2)^2} dxdy + \int_{-1}^0 \int_{-1}^0 \frac{xy}{(x^2 + y^2)^2} dxdy = \infty,$$

and therefore, X *is not integrable. We conclude that* $\mathbb{E}[X]$ *does not exist. However, the iterated integral below exists and is equal to zero, i.e.*

$$\int_{-1}^1 \int_{-1}^1 X(x, y) dxdy = 0.$$

[11] X is undefined at the single point $(0, 0)$. But, the probability of a single point is 0. Therefore, we can define $X(0, 0)$ by any value of choice; for example, $X(0, 0) = 0$.

Therefore, for random variable X, the iterated integral is not a well defined notion of expectation.

There are three main convergence theorems for integrals with respect to a measure, which are very useful in the probability theory. We conclude this section by presenting them.

Theorem B.3 (Monotone convergence theorem). *Let $\{X_n\}_{n\geqslant 1}$ be a sequence of nonnegative and increasing random variables. Then, $X := \lim_{n\to\infty} X_n$ is a random variable and $\mathbb{E}[X] = \lim_{n\to\infty} \mathbb{E}[X_n]$.*

Example B.31. *In Example B.27, we assume the coin is a fair coin. We shall find the expected value of the introduced random variable by introducing the sequence of random variables $\{X_n\}_{n=1}^{\infty}$ defined by*

$$X_n(\omega) := \sum_{j=1}^{n} \frac{a_j}{3^j},$$

such that $a_j = 0$ if $\omega_j = H$, and $a_j = 2$ if $\omega_j = T$. Notice that by the monotone convergence theorem, $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$. Observe that $X_n = X_{n-1} + \xi_n$, where ξ_n takes two values 0 and $\frac{2}{3^n}$ with equal probabilities $\frac{1}{2}$. Therefore, $\mathbb{E}[X_n] = \mathbb{E}[X_{n-1}] + \mathbb{E}[\xi_n]$. Since $\mathbb{E}[\xi_n] = \frac{1}{3^n}$, we have $\mathbb{E}[X_n] = \mathbb{E}[X_{n-1}] + \frac{1}{3^n}$. By solving $\mathbb{E}[X_n] = \mathbb{E}[X_{n-1}] + \frac{1}{3^n}$ recursively, we obtain $\mathbb{E}[X_n] = \sum_{j=1}^{n} \frac{1}{3^j}$. As $n \to \infty$, if follows from the monotone convergence theorem that $\mathbb{E}[X] = \sum_{j=1}^{\infty} \frac{1}{3^j} = \frac{1}{2}$.

Exercise B.16. *Repeat the calculation in Example B.31 for an unfair coin with a probability p of heads.*

Theorem B.4 (Fatou's lemma). *Let $\{X_n\}_{n\geqslant 1}$ be a nonnegative sequence of random variables. Then, $X = \liminf_{n\to\infty} X_n$ is a random variable, and $\mathbb{E}[X] \leqslant \liminf_{n\to\infty} \mathbb{E}[X_n]$.*

Example B.32. *On the sample space $\Omega = [0, 1)$ equipped with uniform probability measure, define the sequence of random variables $\{Y_n\}_{n=1}^{\infty}$ by*

$$Y_n(\omega) = \begin{cases} 1 & \text{when for } j = 1, ..., n,\ a_j = 0 \text{ or } 2 \\ 0 & \text{otherwise} \end{cases},$$

where $\omega = \sum_{j=1}^{\infty} \frac{a_j}{3^n}$ is the ternary expansion of $\omega \in \Omega$. Define

$$Y(\omega) := \lim_{n\to\infty} Y_n(\omega).$$

It is straightforward to see that

$$Y(\omega) = \begin{cases} 1 & \text{when } a_j = 0 \text{ or } 2 \text{ for all } j \geqslant 1 \\ 0 & \text{otherwise} \end{cases}.$$

In other words, Y is the indicator random variable of the event that contains all numbers in $[0,1)$ whose ternary representation has only digits 0 or 2. This event is the same as the Cantor set in Remark B.5 and is denoted by C. Therefore, $Y = 1_C$ and $\mathbb{E}[Y] = \mathbb{P}(C)$. To find the $\mathbb{P}(C)$, we use Fatou's lemma, Theorem B.4, and write

$$\mathbb{E}[Y] = \mathbb{P}(C) \leqslant \liminf_{n\to\infty} \mathbb{E}[Y_n].$$

Next, we will show that $\liminf_{n\to\infty} \mathbb{E}[Y_n] = 0$. To see this, first observe that Y_n is also an indicator random variable. For example for $n = 1$, $Y_1 = 1_{[0,1/3)\cup[2/3,1)}$ and $Y_2 = 1_{[0,1/9)\cup[2/9,1/3)\cup[2/3,7/9)\cup[8/9,1)}$, and so on. Generally speaking,

$$Y_n = 1_{A_n},$$

where A_n is the union of 2^n intervals each of size $\frac{1}{3^n}$. Therefore, under uniform probability,

$$\mathbb{E}[Y_n] = \mathbb{P}(A_n) = \frac{2^n}{3^n} \to 0, \text{ as } n \to \infty.$$

This means that the Cantor set has a probability of 0 under uniform probability.

Theorem B.5 (Lebesgue convergence theorem). *Let $\{X_n\}_{n\geqslant 1}$ be a convergent sequence of random variables and ξ be an integrable random variable such that $|X_n| \leqslant |\xi|$ for all $n \geqslant 1$. Then, $X := \lim_{n\to\infty} X_n$ is a random variable, and $\mathbb{E}[X] = \lim_{n\to\infty} \mathbb{E}[X_n]$.*

Remark B.6. *In Example B.32, we could use Lebesgue convergence theorem, Therem B.5, to show that $\mathbb{E}[Y] = 0$. This is because all random variables Y_n are indicators, and therefore, their absolute values are bounded by the constant random variable $\xi = 1$.*

Equality of random variables

Recall the simple version of the law of large numbers in Theorem B.2. Here, we put the emphasis on the exception set, where the limit inside the theorem does not converge to $\frac{1}{2}$. One possible outcome inside this set is given by $X_n = 0$ for $n \equiv 0 \pmod 3$, and $X_n = 1$ for $n \equiv 1$ or $2 \pmod 3$. Then, the limit on the left-hand side converges to $\frac{2}{3}$, not $\frac{1}{2}$. The law of large numbers in Theorem B.2 asserts that the collection of all outcomes such that $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} X_n$ exists and is equal to $\frac{1}{2}$ is an event with a probability of 1. In other words, the random variable $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} X_n$ is defined to be the constant $\frac{1}{2}$ except on an event with probability 0.

Definition B.9. *Two random variables X and Y are considered equal if the event*

$$\{\omega \in \Omega \,:\, X(\omega) \neq Y(\omega)\}$$

has a probability of zero. Then, we write $X = Y$ $\mathbb{P}$ almost surely; $X = Y$ $\mathbb{P}$-a.s. for short; or just a.s. whenever the probability measure is assumed.

For instance, Theorem B.2 indicates that the random variable on the left-hand side, $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{\infty} X_n$, is equal to the random variable on the right-hand side, $\frac{1}{2}$, a.s. In addition, it also implies that the limit $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{\infty} X_n$ exists a.s., which is not an obvious fact.

As a result of the following Lemma, Theorem B.2 shows that the random variable $X(\omega) = \lim_{N\to\infty} \sum_{n=1}^{N} X_n - \frac{1}{2}$ is zero a.s.: a very fancy representation for zero!

Lemma B.1. *For a nonnegative random variable X, $\mathbb{E}[X] = 0$ if and only if $X = 0$ a.s.*

As a consequence, we can define the convergence of a sequence of random variables.

Definition B.10. *A sequence $\{X_n\}_{n=1}^{\infty}$ of random variables is called convergent a.s. if the event*

$$\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) \ \text{does not exist}\}$$

has a probability of 0. Then, the limit random variable $X := \lim_{n\to\infty} X_n$ exists a.s. and is called the limit of $\{X_n\}_{n=1}^{\infty}$.

Law of a random variable

In the remainder of this section, we show how to define expectation for general random variables by means of the distribution function, without appealing to the notion of integration with respect to a measure.

Definition B.11. *For a random variable X with real values, the cumulative distribution function (or cdf or simply "distribution function") $F_X(x)$ is defined by*

$$F_X(x) := \mathbb{P}(X \leqslant x).$$

The definition of a distribution function is related to the probability measure $\mathbb{P}$ by $\mathbb{P}(X \in (a, b]) = F_X(b) - F_X(a)$. In fact, the distribution function induces a probability measure on $\mathbb{R}$, denoted by $\mathbb{P}_X$, which is defined on open intervals by

$$\mathbb{P}_X((a, b]) = F_X(b) - F_X(a).$$

recall that to define a probability measure on $\mathcal{B}(\mathbb{R})$, we only need to define it on the open intervals, or equivalently on the half-open intervals of the form $(a, b]$.

Example B.33. *Let $X = 1_A$. Then, $F_X(x) = 0$ if $x < 1$ and $F_X(x) = 1$ if $x \geqslant 1$; see Figure B.4.*

Example B.34. *In Example B.24, define a random variable that is equal to the value of the picked random number. Then, the law of X is shown in Figure B.5.*

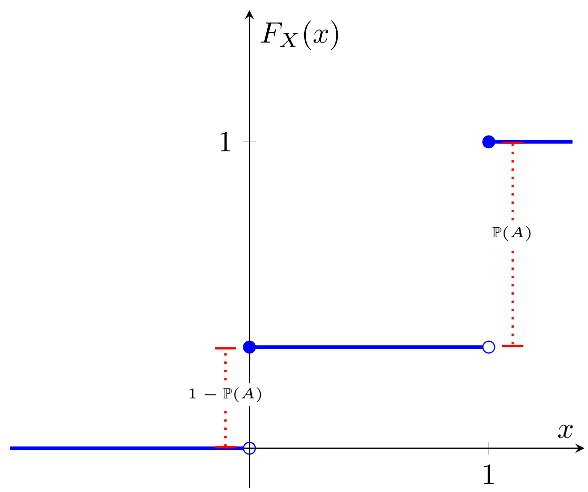


Figure B.4: The distribution function for an indicator.

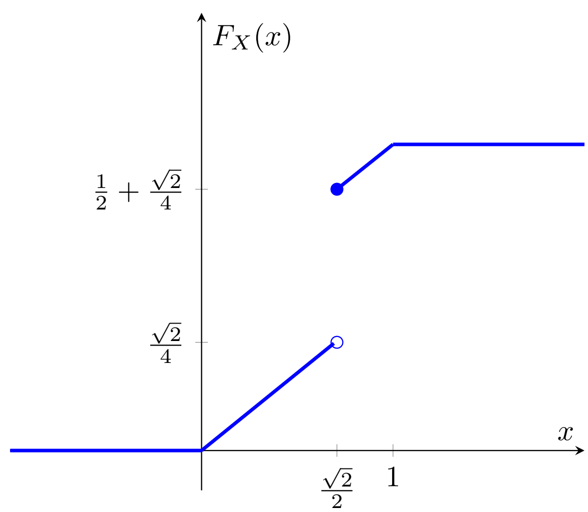


Figure B.5: The distribution function for the random number chosen in Example B.24.

For a discrete random, F_X is a step function. For nondiscrete random variables $F_X(x)$ is not a step function. The distribution function of random variable X in Example B.27 is illustrated in Figure B.6[12].

In general, the distribution function of a random variable X satisfies the following properties.

[12] To be precise, the distribution function of random variable X in Example B.27 is approximated by the distribution function of random variable X_n with a large n.

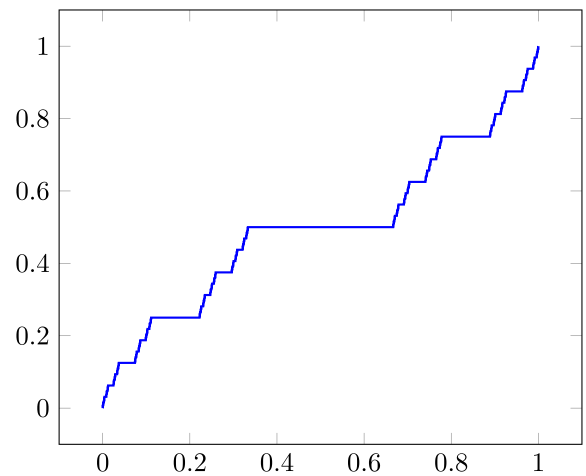


Figure B.6: Devil's Staircase: the distribution function of the random variable X in Example B.27.

a) $F_X(-\infty) = 0$ and $F_X(\infty) = 1$.

b) F_X is an increasing right-continuous function with a left-limit at all points.

Exercise B.17. *Show that* $\mathbb{P}(X \in [a, b]) = F_X(b) - F_X(a-)$, *where* $F_X(a-)$ *is the left limit of* F_X *at* a.

It is also well known that if a function F satisfies the properties (a) and (b) of a distribution function, then there is always a random variable X in a proper sample space such that $F_X = F$.

The probability P_X is called the *law of* X. Notice that knowing F_X and knowing the law of X are equivalent, but they are not equivalent to knowing $\mathbb{P}$. In fact, we shall see that in many important calculations about random variables, we do not exactly need to know $\mathbb{P}$; in most applications, the law of a random variable is all we need. Therefore, we will now introduce expected value in terms of the law of the random variable.

Notice that F_X is an increasing function, and therefore one can define the Riemann-Stieltjes integral with respect to F_X as an alternative to the more complicated notion of integration with respect to measure $\mathbb{P}$. In this case, the integral with respect to measure $\mathbb{P}_X$ is also simplified to a Riemann-Stieltjes[13] integral.

Proposition B.5. *For an integrable random variable* X *the expectation of* X *satisfies*

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x dF_X(x).$$

[13] Read"Steel-chess".

In particular, for any function g,

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)F_X(dx),$$

given $\mathbb{E}[|g(X)|] < \infty$.

Similarly, for a random vector $X = (X_1, ..., X_d)$, the distribution function is defined by

$$F_X(x_1, ..., x_d) := \mathbb{P}(X_1 \leqslant x_1, ..., X_d \leqslant x_d).$$

The law of X is a probability measure $\mathbb{P}_X$ on $\mathbb{R}^d$ such that

$$\mathbb{P}_X((-\infty, x_1] \times \cdots \times (-\infty, x_d]) = F_X(x_1, ..., x_d).$$

Then, the expectation of $g(X_1, ..., X_d)$ is given by

$$\mathbb{E}[g(X_1, ..., X_d)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, ..., x_d)F_X(dx_1, ..., dx_d),$$

given $\mathbb{E}[|g(X_1, ..., X_d)|] < \infty$.

Exercise B.18. *If* $\{X_i\}_{i=1}^{\infty}$ *is a sequence of independent Bernoulli random variables with equally likely values* 0 *and* 1*, show that* $U = \sum_{i=1}^{\infty} \frac{X_i}{2^i}$ *is uniformly distributed on* $[0, 1]$.

Conditional expectation

Conditional expectation cannot simply be defined in terms of the distribution function; more advanced methods are needed. Unlike discrete probability setting, (B.6) in Section B, we need to define $\mathbb{E}[X \mid Y]$ in general form by using a powerful tool in analysis, namely the Radon-Nikodym theorem. To give you a glimpse of the definition, we recall from Remark B.3 in the discrete setting that $\mathbb{E}[X \mid Y = y]$ is a function of the variable y, denoted by $h(y)$. Then, we use this function to define $\mathbb{E}[X \mid Y]$ by $h(Y)$. Similarly, we try to find $\mathbb{E}[X \mid Y]$ among the random variables of the form $h(Y)$ for some (Borel measurable) function $h : \mathbb{R} \to \mathbb{R}$.

Definition B.12 (Conditional expectation)**.** *Let* X *be an integrable random variable.* $\mathbb{E}[X \mid Y]$ *is the* ***unique*** *random variable of the form* $h(Y)$*, where* $h : \mathbb{R} \to \mathbb{R}$ *is a Borel measurable function that satisfies*

$$\mathbb{E}[h(Y)g(Y)] = \mathbb{E}[Xg(Y)], \tag{B.8}$$

for all bounded real functions g with the domain containing the set of values of Y.

The Radon-Nikodym theorem from measure theory guarantees the existence and the uniqueness of an integrable random variable $\mathbb{E}[X \mid Y]$ in the a.s. sense. Given Definition B.12 for conditional expectation, all the results of Corollary B.2, Proposition B.1, and Proposition B.3 hold for general random variables.

Corollary B.4. *Let X and Y be random variables and f be a real function, such that $f(Y)X$ is integrable. Then,*

$$\mathbb{E}[f(Y)X \mid Y] = f(Y)\mathbb{E}[X \mid Y] \text{ a.s..}$$

We only provide a proof for the above corollary and leave the rest of the result of this section to the reader.

Proof. By uniqueness of conditional expectation, it is sufficient to show that both $\mathbb{E}[f(Y)X \mid Y]$ and $f(Y)\mathbb{E}[X \mid Y]$ satisfy the Definition B.12. In other words, for any bounded real functions g with the domain containing the set of values of Y, we have

$$\mathbb{E}[h(Y)g(Y)] = \mathbb{E}[Xg(Y)],$$

with $h(Y) = \mathbb{E}[f(Y)X \mid Y]$ or $h(Y) = f(Y)\mathbb{E}[X \mid Y]$. Without lack of generality, we assume that f is a bounded function. Then, $f(Y)g(Y)$ is also a bounded function. Therefore,

$$\mathbb{E}[\mathbb{E}[X \mid Y]f(Y)g(Y)] = \mathbb{E}[Xf(Y)g(Y)].$$

On the other hand,

$$\mathbb{E}[\mathbb{E}[f(Y)X \mid Y]g(Y)] = \mathbb{E}[Xf(Y)g(Y)].$$

Therefore, the uniqueness of conditional expectation implies the desired result. □

Proposition B.6. *X and Y are independent if and only if for any real function $f(x, y)$ of X and Y, we have*

$$\mathbb{E}[f(X, Y) \mid Y = y] = \mathbb{E}[f(X, y)] \text{ for all } y \text{ in the set of values of } Y,$$

provided that both $f(X, Y)$ and $f(X, y)$ are integrable for all y in the set of values of Y.

Corollary B.5. *Two random variables X and Y are independent if and only if for any a real function f such that $f(X)$ is integrable, we have*

$$\mathbb{E}[f(X) \mid Y] = \mathbb{E}[f(X)] \text{ a.s..}$$

Proposition B.7 (Tower property of conditional expectation)**.** *Let X, Y, and Z be integrable random variables. Then,*

$$\mathbb{E}[\mathbb{E}[X \mid Y, Z] \mid Y] = \mathbb{E}[X \mid Y] \text{ a.s..}$$

In particular,

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X].$$

Remark B.7. *The two following comments often help in finding conditional expectation in general form.*

1) *One can verify (B.12) over a smaller set of functions to guarantee $\mathbb{E}[X \mid Y] = h(Y)$. For instance, if for any bounded function g we have $\mathbb{E}[h(Y)g(Y)] = \mathbb{E}[Xg(Y)]$, then $\mathbb{E}[X \mid Y] = h(Y)$ holds true. Equivalently, if for any Borel set (or open set) A we have $\mathbb{E}[h(Y)1_{Y\in A}] = \mathbb{E}[X1_{Y\in A}]$, then $\mathbb{E}[X \mid Y] = h(Y)$.*

2) *If for a constant a, $A_a = \{Y = a\}$ is an event with positive probability, then $\mathbb{E}[X \mid Y = a] = h(a)$ and is also constant. Then, one can find constant $h(a)$ by using $\mathbb{E}[h(Y)1_{Y\in A_a}] = \mathbb{E}[X1_{Y\in A_a}]$ in the following.*

$$\mathbb{E}[h(Y)1_{\{Y\in A_a\}}] = h(a)\mathbb{P}(Y = a) = \mathbb{E}[X1_{\{Y\in A_a\}}] \implies h(a) = \frac{\mathbb{E}[X1_{\{Y\in A_a\}}]}{\mathbb{P}(Y = a)}.$$

This means that the conditional expectation $\mathbb{E}[X \mid Y]$ on the event $\{Y{=}a\}$ is constant and is equal to the average of X over the event $\{Y = a\}$.

Exercise B.19. *Use Remark B.7-(1) to show Corollary B.4, Proposition B.6 and Proposition B.5.*

Recall that X be a random variable. The σ-field generated by X is the **smallest** σ-field that contains all events $\{X \in A\}$ for all Borel sets A, and is denoted by $\sigma(X)$. In other words, it is the $\sigma(X)$ that contains all events related to X. $\sigma(X)$ can equivalently be expressed as the smallest σ-field that contains all events $\{X \in A\}$ for all open sets A. We say that a random variable Z is $\sigma(X)$-measurable, or measurable with respect to $\sigma(X)$, if for any Borel set A, the event $\{Z \in A\}$ is in $\sigma(X)$. It is known that if Z is $\sigma(X)$-measurable, then there exists a Borel function $h : \mathbb{R} \to \mathbb{R}$ such that $Z = h(X)$. In particular, since $\mathbb{E}[X \mid Y]$ is $\sigma(Y)$-measurable, $\mathbb{E}[X \mid Y] = h(Y)$.

More generally, for a σ-field $\mathcal{G} \subseteq \mathcal{F}$, we say that a random variable Z is $\mathcal{G}$-measurable, or measurable with respect to $\mathcal{G}$, if for any Borel set A, the event $\{Z \in A\}$ is in $\mathcal{G}$. One can define $\mathbb{E}[X \mid \mathcal{G}]$, the conditional expectation of X given the events in $\mathcal{G}$, in a similar fashion as Definition B.12, for which existence and uniqueness are guaranteed by the Radon-Nykodim theorem.

Definition B.13 (Conditional expectation with respect to σ-field)**.** *Let X be an integrable random variable and $\mathcal{G}$ be a sub σ-field of $\mathcal{F}$. $\mathbb{E}[X \mid \mathcal{G}]$ is the **unique** random variable Z that satisfies*

$$\mathbb{E}[ZY] = \mathbb{E}[XY], \tag{B.9}$$

for all $\mathcal{G}$-measurable random variables Y such that XY is integrable.

Remark B.7 holds also for (B.13): one only needs to verify $\mathbb{E}[Z1_G] = \mathbb{E}[X1_G]$, for all $G \in \mathcal{G}$, to show that $Z = \mathbb{E}[X \mid \mathcal{G}]$. If $\mathcal{G}$ is generated by a set of elementary events $\mathcal{E}$, then one only needs to verify that $\mathbb{E}[Z1_E] = \mathbb{E}[X1_E]$ for all $E \in \mathcal{E}$. In addition, let A is an event in $\mathcal{G}$ such that $\mathbb{P}(A) > 0$ and for any events $B \in \mathcal{G}$ such that $B \subseteq A$, we have $\mathbb{P}(B) = \mathbb{P}(A)$ or $\mathbb{P}(B) = 0$. Then, we have

$$\mathbb{E}[X \mid \mathcal{G}] = \frac{\mathbb{E}[X1_{\{Y \in A\}}]}{\mathbb{P}(A)} \quad \text{on } A.$$

Example B.35. *Consider the probability space $((0,1), \mathcal{B}((0,1)), \mathbb{P})$ where $\mathbb{P}$ is uniform probability measure, and let $\mathcal{G}$ be the σ-field generated by the intervals $(\frac{i}{4}, \frac{i+1}{4})$, for $i = 0,1,2,3$. Define the random variable $X : \omega \in (0,1) \mapsto \mathbb{R}$ to be $X(\omega) = 1_{(1/3,2/3)} - 1_{(0,1/3)} - 1_{(2/3,1)}$. We would like to find $\mathbb{E}[X|\mathcal{G}]$. The **key observation** here is that the σ-field $\mathcal{G}$ can be generated by a random variable that takes a constant value in interval $(\frac{i}{4}, \frac{i+1}{4})$, for $i = 0,1,2,3$. For instance, one can choose (among all other valid choices)*

$$Y = \sum_{i=0}^{3} i 1_{(\frac{i}{4}, \frac{i+1}{4})}$$

and $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X|Y]$. Therefore, since Y is constant in each interval $(\frac{i}{4}, \frac{i+1}{4})$, so is $\mathbb{E}[X|\mathcal{G}]$, and we can write

$$\mathbb{E}[X|\mathcal{G}] = \sum_{i=0}^{3} a_i 1_{(\frac{i}{4}, \frac{i+1}{4})}.$$

In Definition B.13, one can take $Y = 1_{(\frac{i}{4}, \frac{i+1}{4})}$ to obtain

$$a_i \mathbb{E}[1_{(\frac{i}{4}, \frac{i+1}{4})}] = \mathbb{E}[X1_{(\frac{i}{4}, \frac{i+1}{4})}] = \int_{\frac{i}{4}}^{\frac{i+1}{4}} X(\omega) d\omega = \begin{cases} -\frac{1}{4} & i = 0 \\ \frac{1}{12} & i = 1 \\ \frac{1}{12} & i = 2 \\ -\frac{1}{4} & i = 3 \end{cases}$$

Therefore,

$$a_i = \begin{cases} -1 & i = 0 \\ \frac{1}{3} & i = 1 \\ \frac{1}{3} & i = 2 \\ -1 & i = 3 \end{cases},$$

and

$$\mathbb{E}[X|\mathcal{G}] = -1_{(0,\frac{1}{4})} + \frac{1}{3} 1_{(\frac{1}{4},\frac{1}{2})} + \frac{1}{3} 1_{(\frac{1}{2},\frac{3}{4})} - 1_{(\frac{3}{4},1)}.$$

The conditional expectation $\mathbb{E}[X \mid \mathcal{G}]$ can be understood by using the concept of regression.

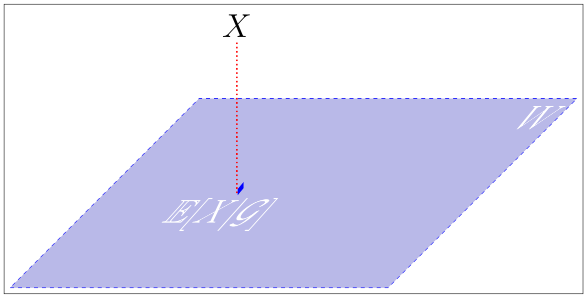


Figure B.7: Geometric interpretation of conditional expectation with respect to a σ-field.

Consider the vector space of square integrable random variables; $V = \{\xi : \mathbb{E}[\xi^2] < \infty\}$ and a linear subspace W of V including all $\mathcal{G}$-measurable random variables. Then, for a random variable $X \in V$, the conditional expectation $Z := \mathbb{E}[X|\mathcal{G}]$ is the closest point in W to X; $\mathbb{E}[X|\mathcal{G}]$ minimizes

$$\mathbb{E}[|X - Y|^2]$$

over all $Y \in W$ ($\mathcal{G}$-measurable square-integrable random variables). See Figure B.7.

Remark B.8. *Any $\sigma(Y)$-measurable random variable $\hat{Y}$ can be written as $h(Y)$ such that $h : \mathbb{R} \to \mathbb{R}$ is a Borel function. Therefore, we have $\mathbb{E}[X \mid Y]$ by $\mathbb{E}[X \mid \sigma(Y)]$.*

Having the conditional expectation defined, one can expand the notion of independence to a general case. One can show, as in Proposition B.9, that two random variables X and Y are independent if and only if $\mathbb{E}[g(X) \mid \sigma(Y)] = \mathbb{E}[g(X)]$ for any Borel function g (or, equivalently, $\mathbb{E}[h(Y) \mid \sigma(X)] = \mathbb{E}[h(Y)]$ for any Borel function h), provided that the expectations exist.

Now, one can define the independence between two σ-fields: two σ-fields $\mathcal{G}$ and $\mathcal{H}$ are independent if for any $\mathcal{G}$-measurable random variable Z, $\mathbb{E}[Z|\mathcal{H}] = \mathbb{E}[Z]$ (or, equivalently, for any $\mathcal{H}$-measurable random variable Z, $\mathbb{E}[Z|\mathcal{G}] = \mathbb{E}[Z]$). Given Definition B.13 for conditional expectation, all the results of Corollary B.4, Proposition B.5 and Proposition B.7 hold for general random variables.

Corollary B.6. *Let X and Y be random variables. Then, provided that Y is $\mathcal{G}$-measurable, we have*

$$\mathbb{E}[YX \mid \mathcal{G}] = Y\mathbb{E}[X \mid \mathcal{G}] \text{ a.s.},$$

given that X and XY are integrable.

Proposition B.8 (Tower property of conditional expectation)**.** *Let X be an integrable random variable and $\mathcal{F}$ and $\mathcal{G}$ be two σ-fields such that $\mathcal{G} \subseteq \mathcal{F}$. Then,*

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{F}] \mid \mathcal{G}] = \mathbb{E}[X \mid \mathcal{G}] \text{ a.s.}$$

In particular,

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{F}]] = \mathbb{E}[X].$$

Proposition B.9. *X is independent of all $\mathcal{G}$-measurable random variables if and only if for any bounded real function f such that $f(X)$ is integrable, we have*

$$\mathbb{E}[f(X) \mid \mathcal{G}] = \mathbb{E}[f(X)] \ a.s.$$

In particular, X is independent of Y if and only if $\mathbb{E}[f(X) \mid \sigma(Y)] = \mathbb{E}[f(X)]$.

Corollary B.7. *Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that hosts an integrable random variable X, and let $\mathcal{G}$ be a σ-field such that $\mathcal{G} \subseteq \mathcal{F}$. Then, for any convex function f, we have*

$$f(\mathbb{E}[X \mid \mathcal{G}]) \leqslant \mathbb{E}[f(X) \mid \mathcal{G}],$$

provided that $f(X)$ is integrable.

The proof of this corollary is exactly the same line of argument as in Corollary B.1.

Continuous random variables

In this section, we review the basic concepts of continuous random variables without referring to measurability issues.

Definition B.14. *A random variable X is called continuous if there exits a nonnegative measurable function $f_X : \mathbb{R}^d \to \mathbb{R}$ such that*

$$F_X(x) := \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f_X(y) dy.$$

In this case, the function f_X is called the probability density function (pdf) of the continuous random variable X.

Working with continuous random variables is often computationally convenient; one can accurately approximate the integrals to estimate relevant quantities such as probability of certain events and expected value of certain random variables. For example, when X is a univariate continuous random variable with pdf f_X, the expected value of X is given by

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx.$$

More generally, for a function $h : \mathbb{R} \to \mathbb{R}$, we have

$$\mathbb{E}[h(X)] = \int_{-\infty}^{\infty} h(x) f_X(x) dx.$$

If $I \subseteq \mathbb{R}$, then

$$\mathbb{P}(X \in I) = \int_I f_X(x)dx.$$

Example B.36 (Normal distribution)**.** *A continuous random variable X with density*

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \text{for all} \quad x \in \mathbb{R}$$

is called a standard Gaussian random variable. By using integration techniques, one can see that $\mathbb{E}[X] = 0$ and $\mathrm{var}(X) := \mathbb{E}[X^2] = 1$.

If $Y = \sigma X + \mu$ for $\sigma > 0$ and $\mu \in \mathbb{R}$, then Y is also a continuous random variable with density

$$f_Y(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{for all} \quad x \in \mathbb{R}.$$

Then, Y is called a normal random variable with mean μ and variance σ^2 and is denoted by $Y \sim \mathcal{N}(\mu, \sigma^2)$.

Exercise B.20. *Show that when $X \sim \mathcal{N}(0,1)$, we have $\mathbb{E}[X] = 0$ and $\mathrm{var}(X) := \mathbb{E}[X^2] = 1$.*

When $X = (X_1, ..., X_d)$ is a *jointly* continuous random vector, we refer to its pdf f_X as the *joint probability density function* of $X_1, ..., X_d$ to emphasize its multi-dimensionality. For simplicity, let's focus on two random variables. Let the joint pdf of (X, Y) be $f_{X,Y}(x, y)$. Then, it is easy to see that

a) The pdf of X, f_X, satisfies

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy.$$

A similar formula holds for the pdf of Y.

b) For a function $h : \mathbb{R}^2 \to \mathbb{R}$, we have

$$\mathbb{E}[h(X)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) f_{X,Y}(x, y)dxdy$$

c) If $K \subseteq \mathbb{R}^2$, then

$$\mathbb{P}((X, Y) \in K) = \iint_K f_{X,Y}(x, y)dA.$$

Example B.37 (Bivariate normal distribution). *Let C be a symmetric positive-definite matrix and $\mu = (\mu_1, ..., \mu_d) \in \mathbb{R}^d$. A jointly continuous random vector $X = (X_1, ..., X_d)$ with density*

$$f_X(x) = \frac{1}{(2\pi\, det(C))^{d/2}} e^{-\frac{(x-\mu)C^{-1}(x-\mu)^{\mathsf{T}}}{2}} \quad \text{for all} \quad x \in \mathbb{R}^d$$

is called a multi-variate normal random vector. For each $j = 1, ..., d$, $\mathbb{E}[X_j] = \mu_j$. The matrix C is called the covariance matrix of X, because its entities correspond to the covariance of components of X, i.e.,

$$C_{i,j} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)].$$

In particular, for $d = 2$, for a positive-definite matrix[14]

$$C = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

and $\mu = 0$, we have

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}} e^{-\frac{\sigma_2^2 x^2 - 2\sigma_{12}xy + \sigma_1^2 y^2}{2(\sigma_1^2\sigma_2^2 - \sigma_{12}^2)}} \quad \text{for all} \quad x \in \mathbb{R}.$$

Here by evaluating double integrals, we can see that σ_1^2 and σ_2^2 are variances of X and Y, respectively, and that σ_{12} is the covariance of X and Y.

Exercise B.21. *In Example B.37, show that $cov(X, Y) := \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] = C_{i,j}$.*

Exercise B.22. *In Example B.37, show that σ_1^2 and σ_2^2 are variance of X and Y respectively, and σ_{12} is the covariance of X and Y.*

Defining and calculating conditional probability and conditional expectation is also done through integral definition for continuous random variables. Let (X, Y) be a jointly continuous random variable with density $f_{X,Y}(x, y)$. Then, the *conditional density of X given $Y = y$* is defined by

$$f_{X|Y}(x \mid y) := \frac{f_{X,Y}(x, y)}{f_Y(y)},$$

provided that y is in the set of values of Y, $\{y : f(y) \neq 0\}$. Using the above definition, the conditional probability of $X \in I$ given $Y = y$ is given by

$$\mathbb{P}(X \in I \mid Y = y) := \frac{\int_I f_{X,Y}(x, y)dx}{f(y)} \tag{B.10}$$

Similarly, if $h : \mathbb{R} \to \mathbb{R}$ is a function, then the conditional expectation of $h(X)$ given $Y = y$

[14] In order for C to be positive-definite, is necessary and sufficient to have $\sigma_1, \sigma_2 \neq 0$, and $\sigma_1^2\sigma_2^2 - \sigma_{12}^2 > 0$.

is given by

$$\mathbb{E}[h(X) \mid Y = y] := \frac{\int_{-\infty}^{\infty} h(x) f_{X,Y}(x,y)dx}{f(y)}. \tag{B.11}$$

In particular, the conditional expectation of X given $Y = y$ is given by

$$\mathbb{E}[X \mid Y = y] := \frac{\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dx}{f_Y(y)}. \tag{B.12}$$

Notice that conditional expectation and conditional probability in (B.10), (B.11), and(B.12) are functions of the variable y. The domain of all these is the set of values of Y, i.e., $\{y : f_Y(y) \neq 0\}$. This, in particular, can be useful in defining *conditional distribution* and *conditional probability given* Y. Let's first make the definition for *conditional probability of* $X \in I$ *given* Y. Consider a function $\mathcal{Y}$ that maps y onto $\mathcal{Y}(y) := \mathbb{P}(X \in I \mid Y = y) = \frac{\int_I f_{X,Y}(x,y)dx}{f_Y(y)}$. Then, one can define

$$\mathbb{P}(X \in I \mid Y) := \mathcal{Y}(Y) = \frac{\int_I f_{X,Y}(x,Y)dx}{f_Y(Y)}.$$

Notice that, unlike $\mathbb{P}(X \in I \mid Y = y)$ which is a real function, $\mathbb{P}(X \in I \mid Y)$ is a random variable that is completely dependent on random variable Y. Similarly, we have

$$\mathbb{E}[h(X) \mid Y] := \frac{\int_{-\infty}^{\infty} h(x) f_{X,Y}(x,Y)dx}{f_Y(Y)} \quad \text{and} \quad \mathbb{E}[X \mid Y] := \frac{\int_{-\infty}^{\infty} x f_{X,Y}(x,Y)dx}{f_Y(Y)}.$$

For continuous random variables independence can be defined in terms of the joint pdf $f(x,y)$; let X and Y be jointly continuous. Then, X and Y are called independent if f is a separable function, i.e.

$$f_{X,Y}(x,y) = g(x)h(y).$$

Notice that the choice of h and g is not unique and varies by multiplying or dividing constants. In this case, one can write the separation in a standard form

$$f_{X,Y}(x,y) = f_X(x) f_Y(y), \tag{B.13}$$

where $f_X(x)$ and $f_Y(y)$ are, respectively, the pdf of x and the pdf of Y.

Example B.38. *From Example B.37 and (B.13), one can see that bivariate normal random variables are independent if and only if they are uncorrelated, i.e., they have zero correlation or simply* $\sigma_{12} = 0$.

Exercise B.23. *Show that (B.7) and (B.13) are equivalent.*

B.3 Martingales

The conditional expectation defined in the previous section is required in the definition of a martingale. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, i.e., a sample space Ω, a σ-field $\mathcal{F}$, and a probability measure $\mathbb{P}$ and let $X := \{X_t\}_{t=0}^T$ (possibly $T = \infty$) be a *discrete-time stochastic process*. Here, a stochastic process is simply a sequence of random variables indexed by time; for any $t = 0, 1, ...,$ X_t is a random variable.

Definition B.15. *A discrete-time stochastic process* $\{M_t\}_{t=0}^T$ *on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *is called a martingale with respect to* X *if*

a) M_t *is integrable for all* $t = 0, ..., T$, $\mathbb{E}[|M_t|] < \infty$.

b) The conditional expectation M_t *given* $X_s, X_{s-1}, ..., X_0$ *is equal to* M_s,

$$\mathbb{E}[M_t|X_s, X_{s-1}, ..., X_0] = M_s, \text{ for } s < t.$$

Condition (a) in the definition of martingale is technical and guarantees the existence of the conditional expectation in condition (b). Condition (b) in the definition of martingale implies that M_s is a $\sigma(X_0, ..., X_s)$-measurable random variable, for all $s \geqslant 0$. This is because the conditional expectation $\mathbb{E}[M_t|X_s, X_{s-1}, ..., X_0]$ is a $\sigma(X_0, ..., X_s)$-measurable random variable.

Condition (b) can also be given equivalently by

b′) $\mathbb{E}[M_t|X_{t-1}, ..., X_0] = M_{t-1}$, for $t \geqslant 1$.

Notice that, as a result of the tower property for conditional expectation, if (b′) holds, we can write

$$\mathbb{E}[\mathbb{E}[M_t|X_{t-1}, ..., X_0]|X_s, ..., X_0] = \mathbb{E}[M_{t-1}|X_s, ..., X_0].$$

By applying the tower property inductively, we obtain

$$\mathbb{E}[M_t|X_s, ..., X_0] = \mathbb{E}[M_{t-1}|X_s, ..., X_0] = \cdots = \mathbb{E}[M_{s+1}|X_s, ..., X_0] = M_s.$$

Example B.39. *Let* Y *be an arbitrary integrable random variable. Then,* $M_t := \mathbb{E}[Y \mid X_t, ..., X_0]$ *is a martingale with respect to* X.

Example B.40 (Symmetric random walk)**.** *The symmetric random walk in Definition B.2 is a martingale. Since* $W_i - W_{i-1} = \xi_i$, *we have*

$$\begin{aligned}\mathbb{E}[W_{t+1}|W_t, ..., W_0] &= \mathbb{E}[W_t + \xi_{t+1}|\xi_t, ..., \xi_1, W_0] \\ &= \mathbb{E}[W_t|\xi_t, ..., \xi_1, W_0] + \mathbb{E}[\xi_{t+1}|\xi_t, ..., \xi_1, W_0].\end{aligned}$$

$\mathbb{E}[W_t|\xi_t, ..., \xi_1, W_0]$ *is simply* $\mathbb{E}[W_t|W_t, ..., W_0] = W_t$. *On the other hand, since* $\{\xi_n\}_{n=1}^\infty$ *is an independent sequence of random variables, it follows from Corollary B.3 that*

$$\mathbb{E}[\xi_{t+1}|\xi_t, ..., \xi_1, W_0] = \mathbb{E}[\xi_{t+1}] = 0,$$

and therefore, $\mathbb{E}[W_{t+1}|W_t, ..., W_0] = W_t$.

Example B.41 (Multiperiod binomial model)**.** *Under the risk-neutral probability, the discounted price of the asset,* $\hat{S}_t := \frac{S_t}{(1+R)^t}$ *is a martingale. Recall from Section II.2.3 that*

$$S_t = S_{t-1}H_t,$$

where H_i *is a sequence of i.i.d. random variables under the risk-neutral probability that is given by*

$$\mathbb{P}^\pi(H_i = u) = \frac{1 + R - \ell}{u - \ell} \quad \text{and} \quad \mathbb{P}^\pi(H_i = l) = \frac{u - 1 - R}{u - \ell}.$$

Thus, H_t *is independent of* $S_{t-1} = S_0 H_{t-1} \cdots H_1$ *and*

$$\begin{aligned}\mathbb{E}[S_t|S_{t-1}, ..., S_0] &= \mathbb{E}[S_{t-1}H_t|H_{t-1}, ..., H_1, S_0] \\ &= S_{t-1}\mathbb{E}[H_t|H_{t-1}, ..., H_1, S_0] = S_{t-1}\mathbb{E}[H_t] = (1 + R)S_{t-1}.\end{aligned}$$

In the second equality above, we used Corollary B.2 and the third equality is the result of Corollary B.3.

Remark B.9. *As a result of tower property of conditional expectation, the expectation of a martingale remain constant with time, i.e.,* $\mathbb{E}[M_t] = \mathbb{E}[M_0]$.

Example B.42. *Let* $\{W_t\}_{t=0}^\infty$ *be a symmetric random walk from Example B.2 and define* $M_t := W_t^2 - t$. *Then,* W_t *is a martingale with respect to* $\{\xi_t\}_{t=1}^\infty$. *To see this, we need to show*

$$\mathbb{E}[W_{t+1}^2 - (t + 1)|\xi_t, ..., \xi_1] = W_t^2 - t.$$

Recall that $W_{t+1} = W_t + \xi_{t+1}$. *Thus,*

$$\begin{aligned}\mathbb{E}[W_{t+1}^2|\xi_t, ..., \xi_0] &= \mathbb{E}[(W_t + \xi_{t+1})^2|\xi_t, ..., \xi_1] \\ &= \mathbb{E}[W_t^2|\xi_t, ..., \xi_1] + \mathbb{E}[\xi_{t+1}^2|\xi_t, ..., \xi_1] + 2\mathbb{E}[W_t\xi_{t+1}|\xi_t, ..., \xi_1].\end{aligned}$$

It follows from Corollary B.2 that

$$\begin{aligned}\mathbb{E}[W_t^2|\xi_t, ..., \xi_1] &= W_t^2 \\ \mathbb{E}[W_t\xi_{t+1}|\xi_t, ..., \xi_1] &= W_t\mathbb{E}[\xi_{t+1}|\xi_t, ..., \xi_1].\end{aligned}$$

On the other hand, by Corollary B.3, we have

$$\begin{aligned}\mathbb{E}[\xi_{t+1}^2|\xi_t, ..., \xi_1] &= \mathbb{E}[\xi_{t+1}^2] = 1 \\ \mathbb{E}[\xi_{t+1}|\xi_t, ..., \xi_1] &= \mathbb{E}[\xi_{t+1}] = 0\end{aligned}$$

Thus, $\mathbb{E}[X_{t+1}^2|\xi_t, ..., \xi_0] = W_t^2 + 1$, *and therefore,*

$$\mathbb{E}[M_{t+1}|\xi_t, ..., \xi_1] = W_t^2 + 1 - (t + 1) = M_t.$$

In the definition of a martingale, Definition B.15, process X models the dynamics of the information as time passes. At time t, the occurrence or absence of all events related to $X_t, ..., X_0$ are known. The conditional expectation $\mathbb{E}[M_t|X_s, X_{s-1}, ..., X_0]$ should be read as expectation of M_t given the information gathered from the realization of process X until time s. It follows from Remark B.8 that if we denote $\mathcal{F}_s^X := \sigma(X_s, ..., X_0)$, then part (b) in Definition B.15 of martingale can be written as

$$\mathbb{E}[M_t|\mathcal{F}_s^X] = M_s.$$

This motivates the definition of a *filtration*.

Definition B.16. *A filtration is a sequence of σ-fields $\mathbb{F} := \{\mathcal{F}_s\}_{s=0}^{\infty}$ such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $t \geqslant s$.*

For example, if we set $\mathcal{F}_s^X := \sigma(X_s, ..., X_0)$, then

$$\mathcal{F}_s^X = \sigma(X_s, X_{s-1}, ..., X_0) \subseteq \mathcal{F}_t^X = \sigma(X_t, ..., X_s, X_{s-1}, ..., X_0).$$

In this case, we call $\mathbb{F}^X$ the filtration generated by X. $\mathbb{F}^X$ represents the accumulated information that are revealed by the process X as time passes.

We a filtration $\{\mathcal{F}_t\}_{t=0}^{\infty}$ is given, a stochastic process $\{Y_t\}_{t=0}^{\infty}$ is called *adapted* with respect to the filtration if for all t, Y_t is an $\mathcal{F}_t$-measurable random variables. For instance, in Example B.13, the random walk $\{W_t\}_{t=0}^{\infty}$ is adapted with respect to $\{\mathcal{F}_t^{\xi}\}_{t=0}^{\infty}$. Here, W_0 is a constant and $\mathcal{F}_t^{\xi} = \sigma(\xi_1, ..., \xi_t)$.

Given Definition B.13 of conditional expectation with respect to a σ-field, one can now define a martingale with respect to a given filtration without appealing to a process X. A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $\mathbb{F} := \{\mathcal{F}_s\}_{s=0}^{\infty}$ is called a *filtered* probability space.

Definition B.17. *Consider a filtered probability space, i.e., $(\Omega, \mathcal{F}, \mathbb{F} := \{\mathcal{F}_s\}_{s=0}^{\infty}, \mathbb{P})$. A discrete-time stochastic process $\{M_t\}_{t=0}^{T}$ is called a martingale with respect to filtration $\mathbb{F}$ if*

a) The expected value of $|M_t|$ is finite for all $t = 0, ..., T$, i.e., $\mathbb{E}[|M_t|] < \infty$.

b) The conditional expectation M_t given $\mathcal{F}_s$ is equal to M_s, $\mathbb{E}[M_t \mid \mathcal{F}_s] = M_s$ for $s \leqslant t$.

We can equivalently present (b) as

$$\mathbb{E}[M_t \mid \mathcal{F}_{t-1}] = M_{t-1}.$$

By applying the tower property in Proposition B.8 inductively, we obtain

$$\mathbb{E}[\mathbb{E}[M_t \mid \mathcal{F}_{t-1}] \mid \mathcal{F}_s] = \cdots = \mathbb{E}[M_{s+1} \mid \mathcal{F}_s] = M_s \quad \text{and} \quad \mathbb{E}[M_t] = M_0.$$

Super/submartingales

Motivated by American option pricing and many other applications, we define a supermartingale and a submartingales

Definition B.18. *A discrete-time stochastic process* $\{M_t\}_{t=0}^T$ *on a filtered probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *is called a supermartingale (or, respectively, a submartingale) with respect to a filtration* $\mathbb{F}$ *if*

a) M_t *is integrable for all* $t = 0, ..., T$, *i.e.,* $\mathbb{E}[|M_t|] < \infty$.

b) $\mathbb{E}[M_t \mid \mathcal{F}_s] \leqslant M_s$, *for* $s < t$ *(or, respectively,* $\mathbb{E}[M_t \mid \mathcal{F}_s] \geqslant M_s$, *for* $s < t$*).*

M_t is a supermartingale if and only if $-M_t$ is a submartingale. So, we can only focus our study of supermartingales. A martingale is simultaneously a submartingale and a supermartingale. The supermartingale property can equivalently be given by

$$\mathbb{E}[M_t \mid \mathcal{F}_{t-1}] \leqslant M_{t-1}.$$

Example B.43 (asymmetric random walk)**.** *An asymmetric random walk with* $\mathbb{P}(\xi_n = 1) = p < \frac{1}{2}$ *is a* ***strict*** *supermartingale.*

$$\begin{aligned}\mathbb{E}[W_{t+1} \mid \xi_t, ..., \xi_1] &= \mathbb{E}[W_t + \xi_{t+1} \mid \xi_t, ..., \xi_1] \\ &= \mathbb{E}[W_t \mid \xi_t, ..., \xi_1] + \mathbb{E}[\xi_{t+1} \mid \xi_t, ..., \xi_1].\end{aligned}$$

$\mathbb{E}[W_t \mid \xi_t, ..., \xi_1]$ *is simply* W_t. *On the other hand, since* $\{\xi_i\}_{i=1}$ *is an independent sequence of random variables, it follows from Corollary B.3 that*

$$\mathbb{E}[\xi_{t+1} \mid \xi_t, ..., \xi_1] = \mathbb{E}[\xi_{t+1}] = 2p - 1 < 0.$$

Thus, $\mathbb{E}[W_{t+1} \mid \xi_t, ..., \xi_1] < W_t$.

The following corollary shows how submartingales naturally show up from convex transformation of martingales.

Lemma B.2. *Let* $\{M_t\}_{t\geqslant 0}$ *be a martingale with respect to filtration* $\mathbb{F}$ *and* f *be a convex function, and define* $Y_t := f(M_t)$. *Then,* $\{Y_t\}_{t\geqslant 0}$ *is a submartingale with respect to filtration* $\mathbb{F}$ *if* $\mathbb{E}[|f(M_t)|] < \infty$ *for all* $t \geqslant 0$.

Proof. By corollary B.7, we know that

$$\mathbb{E}[f(M_t) \mid \mathcal{F}_s] \geqslant f(\mathbb{E}[M_t \mid \mathcal{F}_s]) = f(M_s).$$

In the above, the equality comes from the martingale property of $\{M\}_{t=0}^{\infty}$. □

If $\{M\}_{t=0}^{\infty}$ is a submartingale, we can still have a slightly weaker version of the above corollary.

Lemma B.3. *Let $\{M_t\}_{t\geqslant 0}$ be a submartingale with respect to filtration $\mathbb{F}$ and f be a convex nondecreasing function such that $\mathbb{E}[|f(M_t)|] < \infty$ for all $t \geqslant 0$. Define $Y_t := f(M_t)$. Then, $\{Y_t\}_{t\geqslant 0}$ is a submartingale with respect to filtration $\mathbb{F}$.*

Proof. By corollary B.7, we know that

$$\mathbb{E}[f(M_t) \mid \mathcal{F}_s] \geqslant f(\mathbb{E}[M_t \mid \mathcal{F}_s]) \geqslant f(M_s).$$

In the above, the second inequality comes from the submartingale property of $\{M\}_{t=0}^{\infty}$, i.e., $\mathbb{E}[M_t \mid \mathcal{F}_s] \geqslant M_s$ and nondecreasingness property of f. □

The following corollaries are the result of Lemmas B.2 and B.3

Corollary B.8. *If $M = \{M_t\}_{t\geqslant 0}$ is a martingale, then $M_+ = \{\max\{0, M_t\}\}_{t\geqslant 0}$, $|M| := \{|M_t|\}_{t\geqslant 0}$, and $|M|^p := \{|M_t|^p\}_{t\geqslant 0}$ for $p > 1$ are submartingales; for the last one we need the assumption that $\mathbb{E}[|M_t|^p] < \infty$ for all $t \geqslant 0$.*

Stopping time and optional sampling theorem

Definition B.19. *A stopping time with respect to filtration $\{\mathcal{F}_t\}_{t=0}^{\infty}$ is a random variables $\tau : \Omega \to \{0, 1, ...\}$ such that for any t, the even $\{\tau \leqslant t\}$ belongs to $\mathcal{F}_t$.*

A stopping time is a random time that does not anticipate the release of information that is defined by a filtration. For instance, recall that the filtration generated by the process X is given by $\{\mathcal{F}_t^X\}_{t=0}^{\infty}$, where $\mathcal{F}_t^X = \sigma(X_0, ..., X_t)$ increases as time t passes. Then, the event that τ happened before or at t should belong to $\sigma(X_0, ..., X_t)$ to make *tau* a stopping time. Some examples of a stopping time are presented below.

i) A deterministic time $\tau \equiv t_0$ is a stopping time. Then, the event $\{t_0 \leqslant t\}$ is either Ω or $\varnothing$, if $t_0 \leqslant t$ or $t_0 > t$, respectively.

ii) Let τ_a be the first time that a stochastic process $\{X\}_{t=0}^{\infty}$ such as the price of an asset is greater than or equal to a. Then, the event $\{\tau \leqslant t\}$ can also be represented by al the outcomes such that $X_u \geqslant a$ for some $u \leqslant t$. Therefore, it belongs to $\sigma(X_0, ..., X_t)$.

iii) Assume that $\{Y_t\}_{t\geqslant 0})$ be a stochastic process which is measurable with respect to $\sigma(X_0, ..., X_t)$. Let τ be the first time that a stochastic process $\{Y_t\}_{t\geqslant 0})$ enters the interval (a, b). Then, τ is a stopping time.

iv) If $\{Y_t\}_{t\geqslant 0})$ is *predictable*, i.e., measurable with respect to $\sigma(X_0, ..., X_{t-1})$, then the stopping time in (iii) is called a *predictable stopping time.* In other words, we know that the event $\{\tau \leqslant t\}$ is going to happen one period ahead of time t.

v) If Y_t is not measurable with respect to $\sigma(X_0, ..., X_t)$; but it is measurable with respect to $\sigma(X_0, ..., X_{t+1})$, then τ is not a stopping time. Because we know about the occurrence of $\{\tau \leqslant t\}$ not any time sooner than $t+1$.

Lemma B.4. *If τ and ϱ are two stopping times with respect to a filtration $\{\mathcal{F}_t\}_{t=0}^{\infty}$, then $\tau \vee \varrho := \max\{\tau, \varrho\}$, and $\tau \wedge \varrho := \min\{\tau, \varrho\}$ are stopping times with respect to filtration $\{\mathcal{F}_t\}_{t=0}^{\infty}$. In particular, for a deterministic time t_0, $\tau \wedge t_0$ is a stopping time bounded by t_0 with respect to filtration $\{\mathcal{F}_t\}_{t=0}^{\infty}$.*

Proof. We only present the proof for $\tau \vee \varrho$ as the rest of the proof follows through a similar line of arguments. The event $\{\tau \vee \varrho \leqslant t\}$ is equal to $\{\tau \leqslant t\} \cap \{\varrho \leqslant t\}$. Since, we have $\{\tau \leqslant t\} \in \mathcal{F}_t$ and $\{\varrho \leqslant t\} \in \mathcal{F}_t$, $\{\tau \vee \varrho \leqslant t\} = \{\tau \leqslant t\} \cap \{\varrho \leqslant t\} \in \mathcal{F}_t$, and therefore, $\tau \vee \varrho$ is a stopping time with respect to a filtration $\{\mathcal{F}_t\}_{t=0}^{\infty}$. □

For the simplicity, a stopping time with respect to a known filtration is called a stopping time, whenever there is no confusion.

Definition B.20 (Stopped process)**.** *For a process $\{X_t\}_{t=0}^{\infty}$ that is adapted to a filtration and a stopping time τ with respect to the same filtration, the random variable X_τ is defined to be equal X_t on the event $\tau = t$.*

Example B.44. *Consider a random walk W, and let the stopping time τ_a be the first time the random walk is greater than or equal to a. Since the random walk moves one step at a time, τ_a is the first time the random walk hits a. One can see this as the first time the wealth of a gambler who bets only one dollar at each round becomes equal to a. Therefore, $W_{\tau_a} = a$.*

Let $\{M\}_{t=0}^{\infty}$ is a martingale and τ be a stopping time with respect to filtration $\mathbb{F}$ such that $\tau \geqslant t_0$ for some deterministic time t_0, it is natural to ask whether $\mathbb{E}[M_\tau \mid \mathcal{F}_{t_0}] = M_{t_0}$. In particular when $t_0 = 0$, we want to see whether $\mathbb{E}[M_\tau] = M_0$. The answer in general case is no according to the next example.

Example B.45. *Recall the Saint Petersburg paradox in Example 2.2.4. Let τ be the first time that the gambler wins a round, i.e.,*

$$\tau := \inf\{t \,:\, \xi_t = 1\}.$$

Notice that the doubling strategy always generates exactly one dollar more than the initial wealth, i.e., $\mathcal{W}_\tau = \mathcal{W}_0 + 1$. Let $t_0 = 0$. If the game is fair, i.e., $\mathbb{P}(\xi = 1) = \mathbb{P}(\xi = -1) = \frac{1}{2}$, the random walk, and therefore, the doubling strategy make the wealth process a martingale. However, $\mathbb{E}[\mathcal{W}_\tau] = \mathcal{W}_0 + 1 > \mathcal{W}_0$.

Under some additional condition, we may hope for $\mathbb{E}[M_\tau \mid \mathcal{F}_{t_0}]$ and $\mathbb{E}[M_\tau] = M_0$ to hold true. The following Theorem is providing a set of sufficient conditions.

Theorem B.6 (Optional sampling)**.** *Let τ be stopping time bounded by a deterministic time T, i.e., $\tau \leqslant T$ and $\{M\}_{t=0}^{\infty}$ be a supermartingale. Then for any $t \leqslant T$,*

$$\mathbb{E}[M_\tau \mid \mathcal{F}_t] \leqslant M_{\tau \wedge t}.$$

In particular, $\mathbb{E}[M_\tau] \leqslant M_0$.

It is obvious that for a martingale, the inequalities in the assertion of Theorem B.6 turns to equalities, i.e., $\mathbb{E}[M_\tau \mid \mathcal{F}_t] = M_{\tau \wedge t}$ and $\mathbb{E}[M_\tau] \leqslant M_0$.
To show Theorem B.6, we need the following lemma.

Lemma B.5. *Let $\{M\}_{t=0}^{\infty}$ be a supermartingale. Then, the stopped process*

$$M_{\tau \wedge t} = \begin{cases} M_t & \tau > t \\ M_\tau & \tau \leqslant t \end{cases}$$

is also a supermartingale.

Proof of Theorem B.6. Notice that because $\tau \leqslant T$, $\tau \wedge T = \tau$ and by Lemma B.5, one can write

$$\mathbb{E}[M_\tau \mid F_t] = \mathbb{E}[M_{\tau \wedge T} \mid F_t] = M_{\tau \wedge t}.$$

□

Proof of Lemma B.5. It suffices to show $\mathbb{E}[|M_{\tau \wedge t}|] < \infty$ and

$$\mathbb{E}[M_{\tau \wedge t} \mid \mathcal{F}_{t-1}] = M_{\tau \wedge (t-1)}.$$

Notice that $M_{\tau \wedge t} = M_\tau 1_{\{\tau < t\}} + M_t 1_{\{\tau \geqslant t\}}$. Therefore,

$$\mathbb{E}[M_{\tau \wedge t} \mid \mathcal{F}_{t-1}] = \mathbb{E}[M_\tau \mid 1_{\{\tau < t\}} \mid \mathcal{F}_{t-1}] + \mathbb{E}[M_t 1_{\{\tau \geqslant t\}} \mid \mathcal{F}_{t-1}]$$

Since $\{\tau < t\} = \{\tau \leqslant t-1\} \in \mathcal{F}_{t-1}$ and $\{\tau \geqslant t\} = \{\tau < t\}^{\complement} \in \mathcal{F}_{t-1}$, we have

$$\mathbb{E}[M_{\tau \wedge t} \mid \mathcal{F}_{t-1}] = \mathbb{E}[M_\tau 1_{\{\tau < t\}} \mid \mathcal{F}_{t-1}] + 1_{\{\tau \geqslant t\}} \mathbb{E}[M_t \mid \mathcal{F}_{t-1}].$$

By supermartingale property of $\{M\}_{t=0}^{\infty}$, we have $\mathbb{E}[M_t \mid \mathcal{F}_{t-1}] \leqslant M_{t-1}$ and

$$\mathbb{E}[M_{\tau \wedge t} \mid \mathcal{F}_{t-1}] \leqslant \mathbb{E}[M_\tau 1_{\{\tau < t\}} \mid \mathcal{F}_{t-1}] + M_{t-1} 1_{\{\tau \geqslant t\}}.$$

On the other hand,

$$1_{\{\tau < t\}} \mathbb{E}[M_\tau \mid \mathcal{F}_{t-1}] = \sum_{i=0}^{t-1} \mathbb{E}[M_\tau 1_{\{\tau = i\}} \mid \mathcal{F}_{t-1}] = \sum_{i=0}^{t-1} \mathbb{E}[M_i 1_{\{\tau = i\}} \mid \mathcal{F}_{t-1}].$$

Since $1_{\{\tau=i\}}M_i$ is$\mathcal{F}_{t-1}$-measurable for $i \leqslant t-1$, we have

$$\mathbb{E}[M_i 1_{\{\tau=i\}} \mid \mathcal{F}_{t-1}] = 1_{\{\tau=i\}} M_i,$$

and

$$1_{\{\tau<t\}}\mathbb{E}[M_\tau \mid \mathcal{F}_{t-1}] = \sum_{i=0}^{t-1} M_i 1_{\{\tau=i\}} = M_\tau 1_{\{\tau\leqslant t-1\}}.$$

Therefore,

$$\mathbb{E}[M_{\tau\wedge t} \mid \mathcal{F}_{t-1}] \leqslant M_\tau 1_{\{\tau\leqslant t-1\}} + M_{t-1} 1_{\{\tau\geqslant t\}} = M_{\tau\wedge(t-1)}.$$

□

B.4 Characteristic functions and weak convergence

The characteristic function of a (univariate) random variable is a complex function defined by

$$\chi_X(\theta) := \mathbb{E}[e^{i\theta X}].$$

When X is a discrete random variable, then $\chi_X(\theta) = \sum_{i=1}^{\infty} e^{i\theta x_i} p_1$, and when X is continuous with pdf f, then $\chi_X(\theta) = \int_{-\infty}^{\infty} e^{i\theta x} f(x)dx$. In the latter case, the characteristic function is the Fourier transform of the pdf f.

If F is the distribution function of X, the characteristic function can be equivalently given by

$$\chi_X(\theta) = \int_{\mathbb{R}} e^{i\theta x} dF(x),$$

where the above integral should be interpreted as Reimann-Stieltjes integral. Especially for continuous random variables we have

$$\chi_X(\theta) = \int_{\mathbb{R}} e^{i\theta x} f(x)dx.$$

Therefore, for a continuous random variable, the characteristic function is the *Fourier transform* of the pdf. Therefore, if we know the characteristic function of a distribution, then one can find the distribution function by inverse Fourier transform.

Example B.46. *We want to find the characteristic function of $Y \sim \mathcal{N}(\mu, \sigma)$. First tale the standard case of $X \sim \mathcal{N}(0, 1)$.*

$$\chi_X(\theta) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\theta x} e^{-\frac{x^2}{2}} dx = \frac{e^{-\frac{\theta^2}{2}}}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{(x-i\theta)^2}{2}} dx = e^{-\frac{\theta^2}{2}}.$$

Now, since $Y = \mu + \sigma X$, we have

$$\chi_Y(\theta) = \mathbb{E}[e^{i\theta(\mu+\sigma X)}] = e^{i\mu\theta}\chi_X(\sigma\theta) = e^{i\mu\theta - \frac{\sigma^2\theta^2}{2}}.$$

For the *inversion theorem*, see [10, Theorem 3.3.4], the characteristic function **uniquely** determines the distribution of the random variable. Therefore, all the information

For a random vector (X, Y), the characteristic function is defined as

$$\chi(\theta_1, \theta_2) := \mathbb{E}[e^{i(\theta_1 X + \theta_2 Y)}].$$

One of the important implications of definition of independent random variable in Remark B.3 is that if X and Y are independent, then for any (θ_1, θ_2) we have

$$\chi(\theta_1, \theta_2) = M_X(\theta_1)M_Y(\theta_2). \tag{B.14}$$

The inverse is also true; see [10, Theorem 3.3.2]. This provides an easy way to formulate and verify the independence of random variable in theory.

Proposition B.10. *X and Y are independent if and only if (B.14) holds true.*

Weak convergence

The most well-known place where the weak convergence comes to play is the *central limit theorem* (CLT).

Theorem B.7. *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with expectation $\mu = \mathbb{E}[X_1]$ and standard deviation $\sigma := \sqrt{\mathbb{E}[X_1^2] - \mu^2}$ and define $W_n := \frac{\sum_{i=1}^n (X_i - \mu)}{\sigma\sqrt{n}}$. Then,*

$$\mathbb{P}(W_n \leqslant x) \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy,$$

as $n \to \infty$.

The appellation "weak" is originated from the fact that this convergence is weaker that the concept of *pointwise* or *almost sure* (shortly a.s.) convergence. The pointwise convergence indicates that the sequence of random variables$\{W_n(\omega)\}_n$ converges for all $\omega \in \Omega$; while almost sure convergence means the probability of the event

$$A_n := \{\omega \,:\, W_n(\omega)\}_n \text{ does not converge}\}$$

converges to zero as $n \to \infty$. Almost sure convergence, for instance, appears in the *law of large numbers* (LLN).

Theorem B.8. *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with expectation $\mu = \mathbb{E}[X_1]$ and standard deviation $\sigma := \sqrt{\mathbb{E}[X_1^2] - \mu^2}$. Then,*

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{n} \to 0 \quad \textit{almost surely,}$$

as $n \to \infty$. Here almost surely means that the probability that this convergence does not happen is zero.

In other words, the statistical average $\frac{\sum_{i=1}^{n} X_i}{n}$ converges to the expectation (mean), when $n \to \infty$, except for a set of outcomes with zero total probability. For example in the context of flipping a fair coin, the fraction of flips that the coin turns heads converges to $\frac{1}{2}$ exclusively. However, one can simply construct infinite sequences of heads and tails such as $H, T, T, H, T, T, \ldots$ with statistical average converging to some value other than $\frac{1}{2}$.

In the weak sense of convergence in CLT, the sequence of random variables$\{W_n(\omega)\}_n$ do not actually converge to a normal random variable over a significantly large part of the sample space, as a result of *law of iterated logarithms*, i.e.

$$\limsup_{n\to\infty} \frac{W_n}{\sqrt{2\log(\log(n))}} = 1 \quad \text{almost surely.}$$

Instead, the probability distribution function of W_n can be approximated by normal distribution function for large n. In this case, we say that $\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$ converges *weakly* or *in distribution* to standard Gaussian.

Definition B.21. *We say the sequence $\{Y_n\}_{n=1}^{\infty}$ of random variables converges weakly (or in distribution) to random variable Y if for the distribution functions we have*

$$F_{Y_n}(y) \to F_Y(y) \quad \textit{for any } y \textit{ such that } F_Y \textit{ is continuous at } y.$$

We denote the weak convergence by $Y_n \Rightarrow Y$.

The following example reveals a different aspect of weak convergence in regard to comparison with a.s. convergence.

Example B.47. *On a probability space $(\Omega, \mathbb{P})$, let the random variable $Y_n(\omega) = y_n$ for all ω (a.s.), i.e., Y_n is a constant random variable equal to y_n. If $y_n \to y$, then $Y_n \to Y \equiv y$ pointwise (a.s., respectively).*

Now, consider a sequence of possibly different probability space $(\Omega = \{0,1\}, \mathbb{P}_n)$ such that $\mathbb{P}_n(\{0\}) = 1$ if n is odd and $\mathbb{P}_n(\{0\}) = 0$ if n is even. For each n, let the random variable $Z_n : \Omega \to \mathbb{R}$ be defined by

$$Z_n(\omega) = \begin{cases} \frac{1}{n} & (\omega = 0 \quad \textit{and } n \textit{ is odd}) \quad \textit{or} \quad (\omega = 1 \quad \textit{and } n \textit{ is even}) \\ 1 & \textit{otherwise.} \end{cases}$$

In particular, the distribution of Z_n is a Dirac distribution located at $\frac{1}{n}$;

$$F_n(x) := \mathbb{P}_n(Z_n \leqslant x) = \begin{cases} 1 & x \geqslant \frac{1}{n} \\ 0 & \text{otherwise.} \end{cases}$$

Z_n does not converge pointwise, since $Z_n(0)$ alternates between $\frac{1}{n}$ and 1 as n increases successively. However, the distribution function $F_n(x)$ of Z_n converges to the distribution function

$$F(x) := \begin{cases} 1 & x \geqslant 0 \\ 0 & \text{otherwise.} \end{cases}$$

The following proposition is one of the equivalent conditions of weak convergence.

Proposition B.11. *$Y_n \Rightarrow Y$ if and only if for any **bounded continuous** function $f : \mathbb{R} \to \mathbb{R}$ we have*

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

Notice that the expectation in $\mathbb{E}[Y_n]$ and $\mathbb{E}[Y]$ are to be interpreted in different sample spaces with different probabilities.

One of the ways to establish weak convergence results is to use characteristic functions; see [10, Theorem 3.3.6]

Theorem B.9. *Let $\{X_n\}$ be a sequence of random variables such that for any θ, $\chi_{X_n}(\theta)$ converges to a function $\chi(\theta)$ which is continuous at $\theta = 0$. Then, X_n converges weakly to a random variable X with characteristic function χ.*

As a consequence of the this Theorem, one can easily provide a formal derivation for central limit theorem. Let $S_n := \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}$. Thanks to the properties of characteristic function and (B.14), the characteristic function of S_n is given by

$$\chi_{S_n}(\theta) = \prod_{i=1}^{n} \chi_{X_i-\mu}(\theta/(\sigma\sqrt{n})) = \left(\chi_{X_1-\mu}(\theta/(\sigma\sqrt{n}))\right)^n.$$

Here we used identical distribution of sequence $\{X_n\}$ to write the last equality. Since $e^{ix} = 1 + ix - \frac{x^2}{2} + o(x^2)$, we can write

$$\chi_{X_i-\mu}(\theta) = 1 - \frac{\sigma^2\theta^2}{2} + o(\theta^2).$$

Therefore,

$$\chi_{S_n}(\theta) = \left(\chi_{(X_i-\mu)}(\theta/(\sigma\sqrt{n}))\right)^n = \left(1 - \frac{\theta^2}{2n} + o(n^{-1})\right)^n \to e^{-\frac{\theta^2}{2}},$$

as $n \to \infty$. This finished the argument since $e^{-\frac{\theta^2}{2}}$ is the characteristic function of standard Gaussian.

Remark B.10. *In the Definition B.21 of weak convergence, only the distribution of random variables matters, and not the sample space of each random variable. Therefore, in weak convergence, the random variables in sequence $\{Y_n\}_{n=1}^{\infty}$ can live in different sample spaces. However, one can make one universal sample space for all Y_n's and Y; more precisely, if Y_n and Y are defined on probability space $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ and $(\Omega, \mathcal{F}, \mathbb{P})$, then $\tilde{\Omega} = \Omega \times \prod_n \Omega_n$ is a universal sample space. The random variables Y_n and Y are redefined on $\tilde{\Omega}$ by*

$$\tilde{Y}(\omega, \omega_1, \omega_2, ...) = Y(\omega), \quad \text{and} \quad \tilde{Y}_n(\omega, \omega_1, \omega_2, ...) = Y_n(\omega_n).$$

The distribution of random variables on $\tilde{\Omega}$ is determined by the probability $\tilde{\mathbb{P}} := \mathbb{P} \otimes \prod_n \mathbb{P}_n$. Therefore, the weak convergence of random variables can be reduced to weak convergence of probabilities on a single sample space.

Weak convergence of probabilities

If the sample space is a Polish space (complete metrizable topological space), then one can define weak convergence of probabilities (or even measures). Sample spaces with a topology contribute to the richness of the probabilistic structure; the concept of convergence of probabilities can be defined.

Definition B.22. *Consider a sequence of probabilities $\{\mathbb{P}_n\}_n$ on a Polish probability space $(\Omega, \mathcal{F})$. We say $\mathbb{P}_n$ converges weakly to a probability $\mathbb{P}$ on $(\Omega, \mathcal{F})$, denoted by $\mathbb{P}_n \Rightarrow \mathbb{P}$, if for any bounded continuous function $f : \Omega \to \mathbb{R}$ we have*

$$\lim_{n\to\infty} \mathbb{E}^n[f] = \mathbb{E}[f]$$

Notice that in the above definition topology of Ω has been used in the continuity of function f.

One can always reduce the weak convergence of random variables to weak convergence of probabilities in the Polish space $\mathbb{R}^d$. Let $\{Y_n\}_{n=1}^{\infty}$ be a sequence of random variables and Y be a random variable all with values in $\mathbb{R}^d$. Then, the distributions of Y_n's and Y defined probability measures in R^d as follows.

$$\tilde{\mathbb{P}}_n(A) = \mathbb{P}_n(Y \in A) \text{ and } \tilde{\mathbb{P}}(A) = \mathbb{P}(Y \in A).$$

Notice that, as emphasized in Remark B.10, Y_n and Y live in different probability spaces $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ and $(\Omega, \mathcal{F}, \mathbb{P})$, respectively. However, the probabilities $\tilde{\mathbb{P}}_n$ and $\tilde{\mathbb{P}}$ are defined on the same sample space $\mathbb{R}^d$.

Corollary B.9. *$Y_n \Rightarrow Y$ if and only if $\tilde{\mathbb{P}}_n \Rightarrow \tilde{\mathbb{P}}$.*

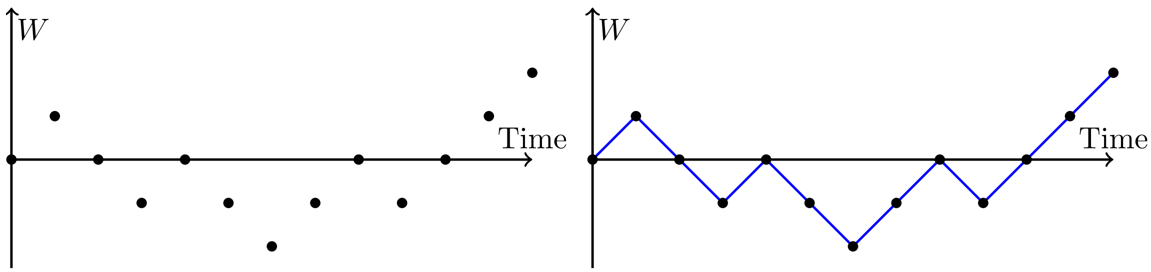


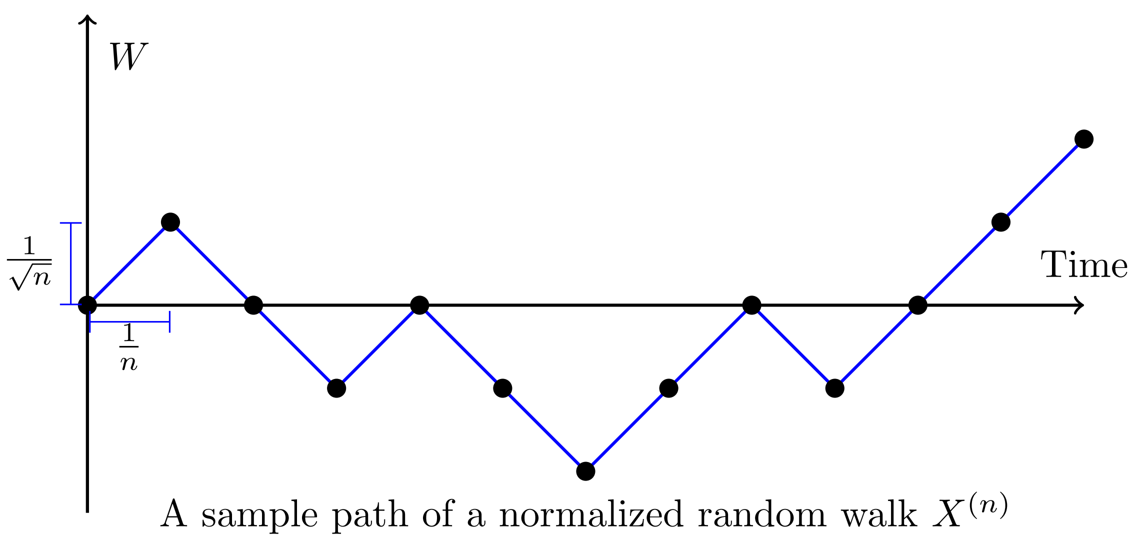
Figure B.8: Left: Sample path of a random walk. Right: Interpolated sample path of a random walk

B.5 Donsker Invariance Principle and Brownian motion

In this section, we heuristically construct Brownian motion (or Wiener process) as the weak limit of symmetric discrete-time random walk in Definition B.2. First, we make the *sample paths* of random walk continuous by linear interpolation. For $t \in [0, \infty)$, we define the interpolated random walk by $W_t := W_{[t]} + (t - [t])W_{[t]+1}$; see Figure B.8. Then, $\{W_t : t \geqslant 0\}$ becomes a *continuous-time stochastic process* with continuous sample paths, i.e., for each $t \geqslant 0$, W_t is a random variable and for any *realization* ω of random walk, $W_t(\omega)$ is continuous in t.

Motivated by central limit theorem, we defined normalized random walk by

$$X_t^{(n)} := \frac{1}{\sqrt{n}} W_{nt} \quad \text{for all} \quad t \in [0, \infty).$$



Then, we define *Brownian motion* is the weak limit of $X_t^{(n)}$ as $n \to \infty$. Indeed, a rigorous definition of the Brownian motion is way more technical and requires advanced techniques from analysis and measure theory. Here we only need the properties which characterize a Brownian motion.

Remark B.11. *In the above construction, one can take any sequence of i.i.d. random variables* $\{\xi_n\}_{n\geqslant 0}$ *with finite variance* σ^2 *and define* $X_t^{(n)} := \frac{1}{\sigma\sqrt{n}} W_{nt}$*. The rest of the arguments in this section can be easily modified for this case.*

Because $\{\xi_i\}_{i=1}^{\infty}$ are i.i.d., for any $t = \frac{k}{n}$ and $s = \frac{\ell}{n}$ with $t > s$, $X_t^{(n)} - X_s^{(n)}$ has mean zero and variance $t-s$. If $t > s$ are real numbers, then the mean is still zero but the variance is $\frac{1}{n}([tn]-[sn])$ which converges to $t-s$. In addition, $X_t^{(n)} - X_s^{(n)}$ is independent of $X_u^{(n)}$ when $u \leqslant s$. By central limit theorem, $X_t^{(n)} - X_s^{(n)} \Rightarrow \mathcal{N}(0, t-s)$, a normal distribution with mean zero and variance $t-s$. This suggests that Brownian motion inherits the following properties in the limit from $X_t^{(n)}$:

a) B has continuous sample paths,

b) $B_0 = 0$,

c) when $s < t$, the increment $B_t - B_s$ is a normally distributed random variables with mean 0 and variance $t-s$ and is independent of B_u; for all $u \leqslant s$.

The properties above fully characterize the Brownian motion.

Definition B.23. *A stochastic process is a Brownian motion if it satisfies the properties (a)-(c) above.*

Property (c) in the definition of Brownian motion also implies some new properties for the Brownian motion which will be useful in modeling financial asset prices.

- **Time-homogeneity.** Brownian motion is time homogeneous, i.e., $B_t - B_s$ has the same distribution as B_{t-s}.

- **Markovian.** The distribution of B_t given $\{B_u : u \leqslant s\}$ has the same distribution as B_t given B_s, i.e., the most recent past is the only relevant information. Notice that $B_t = B_s + B_t - B_s$. Since by property (c) $B_t - B_s$ is independent of $\mathcal{F}_s := \{B_u : u \leqslant s\}$, the distribution of $B_t = B_s + B_t - B_s$ given $\mathcal{F}_s = \{B_u : u \leqslant s\}$ is normal with mean B_s and variance $t-s$, which only depends on the most recent past B_s. In other word, by Proposition B.6

$$\mathbb{E}[f(B_t) \mid \mathcal{F}_s] = \mathbb{E}[f(B_t - B_s + B_s) \mid \mathcal{F}_s] = \mathbb{E}[f(B_t - B_s + B_s) \mid B_s]$$

- **Martingale.** Finally, the conditional expectation $\mathbb{E}[B_t \mid B_u : u \leqslant s]$ can be shown to be equal to B_s.

$$\mathbb{E}[B_t \mid B_u : u \leqslant s] = \mathbb{E}[B_t \mid B_s] = \mathbb{E}[B_s \mid B_s] + \mathbb{E}[B_t - B_s \mid B_s] = B_s.$$

  In the above we used property (c) to conclude that $\mathbb{E}[B_t - B_s \mid B_s] = 0$.

A typical sample paths of Brownian motion is shown in Figure B.9.

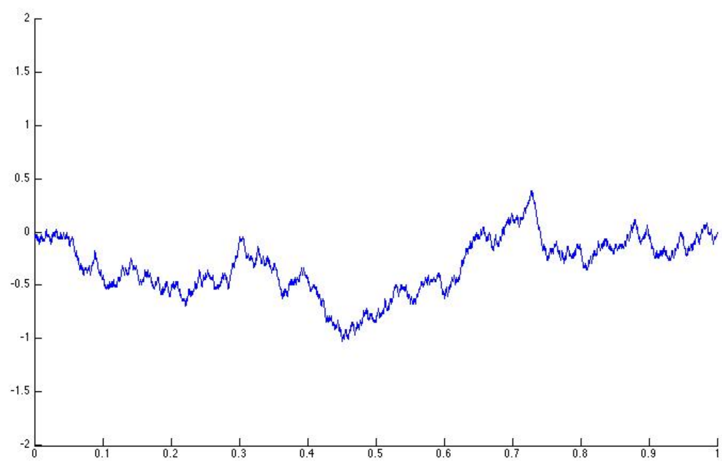


Figure B.9: A sample paths of Brownian motion

Sample space for Brownian motion

In order to construct a Brownian motion, we need to specify the sample space. In the early work of Kolmogorov, we choose the sample space to be $(\mathbb{R}^d)^{[0,\infty)}$, i.e., space of all functions from $[0,\infty)$ to $\mathbb{R}^d$. This is motivated by the fact that for any $\omega \in \Omega$, the *sample path* of Brownian motion associated with sample ω is given by the function $B_t(\omega) : [0,\infty) \to \mathbb{R}^d$. See Figure B.9. Kolmogorov made a theory which in particular resulted in the existence of Brownian motion. While it is not hard to show property (c), in his theory it is not easy to show property (a) of Brownian motion, i.e., the sample paths are continuous. By using the weak convergence result of Yuri Prokhorov, Norbert Wiener take the construction of Brownian motion to a new level by taking the sample space $\Omega := C([0,\infty); \mathbb{R}^d)$, the space of all continuous functions. This way property (a) becomes trivial, while, property (c) is more challenging.

Among all the continuous functions in $C([0,\infty); \mathbb{R}^d)$ only a small set can be a sample paths of a Brownian motion. In the following, we present some of the characteristics of the paths of Brownian motion.

i) Sample paths of Brownian motion are nowhere differentiable. In addition,

$$\limsup_{\delta\to 0}\frac{B_{t+\delta}-B_t}{\delta}=\infty,\quad\text{and}\quad \liminf_{\delta\to 0}\frac{B_{t+\delta}-B_t}{\delta}=-\infty.$$

ii) Sample paths of Brownian motion are of bounded quadratic variation variations. More precisely, the quadratic variation of the path of Brownian motion until time t, equals t, i.e.,

$$\lim_{\|\Pi\|\to 0}\sum_{i=0}^{N-1}(B_{t_{i+1}}-B_{t_i})^2=t,\quad\text{a.s.,}\tag{B.15}$$

where for the partition $\Pi:=\{t_0=0<t_1<\cdots<t_N=t\}$, $\|\Pi\|=\max_{i=0,\ldots,N-1}(t_{i+1}-t_i)$

iii) Sample paths of Brownian motion are not of bounded variations almost surely, i.e., for

$$\sup_{\Pi}\sum_{i=0}^{N-1}|B_{t_{i+1}}-B_{t_i}|=\infty,\quad\text{a.s.,}$$

where the supremum is over all partitions $\Pi:=\{t_0=0<t_1<\cdots<t_N=t\}$.

In the above, property (iii) is a result of (ii). More precisely, for a a continuous function $g:[a,b]\to\mathbb{R}$, a nonzero quadratic variation implies infinite bounded variation. To show this, let's assume that g is continuous and bounded variation B. Then, g is uniformly continuous on $[a,b]$, i.e., for any $\varepsilon>0$, there exists a $\delta>0$ such that if $\|\Pi\|\leqslant\delta$, then

$$\max_{i=0,\ldots,N-1}\{|g(t_{i+1})-g(t_i)|\}\leqslant\varepsilon.$$

Thus,

$$\sum_{i=0}^{N-1}(g(t_{i+1})-g(t_i))^2\leqslant\max_{i=0,\ldots,N-1}\{|g(t_{i+1})-g(t_i)|\}\sum_{i=0}^{N-1}|g(t_{i+1})-g(t_i)|\leqslant\varepsilon B.$$

By sending $\varepsilon\to 0$, we obtain that the quadratic variation vanishes.

The following exercise shows the relation between the quadratic variation is martingale properties of Brownian motion.

Exercise B.24. *Show that B_t^2-t is a martingale, i.e.*

$$\mathbb{E}[B_{t+s}^2-(t+s)\mid B_s]=B_s^2-s.$$

C Stochastic analysis

Calculus is the study of derivative (not financial) and integral. Stochastic calculus is therefore the study of integrals and differentials of stochastic objects such as Brownian motion. In this section, we provide a brief overview of the stochastic integral and Itô formula (stochastic chain rule). The application to finance is provided in Part 3.

In calculus, the Riemann integral $\int_a^b f(t)dt$ is defined by the limit of Riemann sums:

$$\lim_{\delta \to 0} \delta \sum_{i=1}^{N} f(t_i^*),$$

where $\delta = \frac{b-a}{N}$, $t_0 = a$, $t_i = t_0 + i\delta$, and t_i^* is an arbitrary point in interval $[t_{i-1}, t_i]$. The Reiman integral can be defined for a limited class of integrands, i.e., the real functions f is Riemann integrable on $[a, b]$ if and only if it is bounded and continuous almost everywhere. A natural extension of Riemann integral is Lebesgue integral, which can be defined on a large class of real functions, i.e., bounded measurable function on $[a, b]$.

A more general form of Riemann integral, Riemann-Stieltjes integral is defined in a similar fashion. For two real functions $f, g : [a, b] \to \mathbb{R}$, the integral of the *integrand* f with respect to *integrator* g is defined by

$$\int_a^b f(t)dg(t) \lim_{\delta \to 0} \sum_{i=1}^{N} f(t_i^*)(g(t_i) - g(t_{i-1}).$$

For Riemann-Stieltjes integral, and its extension, Lebegue-Stieltjes integral, to be well-defined, we need some conditions on f and g.

The condition on the integrand f is similar to the those in Riemann and Lebesgue integrals. For example if f is continuous almost everywhere and at the points of discontinuity of g, then no further condition needs to be imposed on f in the Riemann-Stieltjes integral. For the Lebesgue-Stieltjes integral, f only needs to be measurable. However, for g a very crucial condition is needed to make the integral well-defined. g must be of bounded variation, i.e.

$$\lim_{\delta \to 0} \sum_{i=0}^{N} |g(t_{i+1}) - g(t_i)| < \infty.$$

No matter how nice the function f is, if g is unbounded variation, Riemann-Stieltjes or Lebesgue-Stieltjes integral cannot be defined. As seen in Section B.5, the sample paths of Brownian motion are of unbounded variation, which makes it impossible to use them as the integrator. Therefore, the integral $\int_a^b f(t)dB_t$ with respect to Brownian motion cannot be defined pathwise in the sense of the Riemann-Stitljes or Lebesgue-Stieltjes integrals.

In this section, we define a new notion of integral, Itô integral[15], which makes sense of $\int_a^b f(t)dB_t$ in a useful way for some applications, including finance.

One of the major tools in stochastic analysis is the stochastic chain rule. Recall that the chain rule in the differential form is written as $dh(g(t)) = g'(t)h'(g(t))dt$, which can be used in the change of variable in integral. If $v(t) = h(g(t))$ and h and g are differentiable functions, then

$$\int_a^b f(t)dv(t) = \int_a^b f(t)g'(t)h'(g(t))dt.$$

The right-hand side above is a Riemann (Lebesgue) integral. As a matter of fact, change of variable formula for Reimann (Lebesgue) integral is the integral format of the chain rule. For bounded variation but not necessarily differentiable function g, the chain rule in the change of variable for Riemann-Stieltjes (Lebesgue-Stieltjes) integral can be written in a slightly different way. More precisely, if $v(t) = h(g(t))$ and h is a differentiable function and g is of bounded variation, then

$$\int_a^b f(t)dv(t) = \int_a^b f(t)h'(g(t))dg(t).$$

In the chain rule for Itô stochastic integral, an extra term appears. If $v(t) = h(B_t)$ and h is a **twice differentiable** function, then

$$\int_a^b f(t)dv(t) = \int_a^b f(t)h'(B_t)dB_t + \frac{1}{2}\int_a^b f(t)h''(B_t)dt.$$

As expected the term $\int_a^b f(t)h'(B_t)dB_t$ shows up like in chain rule. However, the term $\frac{1}{2}\int_a^b f(t)h''(B_t)dt$, which is a simple Riemann integral, is unprecedented. In the remaining of this section, we provide a heuristic argument why this term should be in the chain rule for Itô stochastic integral.

C.1 Stochastic integral with respect to Brownian motion and Itô formula

We first introduce a special case of Itô integral, called Wiener integral[16]. In Wiener integral, we assume that the integrand f is simply a real function and is not stochastic. The the partial sums

$$S_\delta := \sum_{i=1}^{N} f(t_{i-1})(B_{t_i} - B_{t_{i-1}}),$$

[15]Named after Japanese mathematician Kiyosi Itô, 1915-2008.
[16]Named after American mathematician Norbert Wiener, 1894-1964.

is a Gaussian random variable with mean zero and variance

$$\delta \sum_{i=1}^{N} f^2(t_{i-1}).$$

Exercise C.1. *Show that S_δ is a Gaussian random variable with mean zero and variance $\delta \sum_{i=1}^{N} f^2(t_{i-1})$.*

Therefore, it follows form

$$\lim_{\delta \to 0} \delta \sum_{i=1}^{N} f^2(t_{i-1}) = \int_a^b f^2(t)dt,$$

that $S_\delta \Rightarrow X$ where X is a normal random variable with mean zero and variance $\int_a^b f^2(t)dt$.
Notice that in the partial sum for S_δ, we choose t_{i-1}, i.e., the left endpoint on the interval $[t_{i-1}, t_i]$. This choice is not crucial to achieve the limit. If we would choose different points on the interval, we still obtain the same limiting distributions. See Exercise C.2.

Exercise C.2. *Calculate that the mean and the variance of partial sums below:*

a) $\sum_{i=1}^{N} f(t_i)(B_{t_i} - B_{t_{i-1}})$.

b) $\sum_{i=1}^{N} f\left(\frac{t_i + t_{i-1}}{2}\right)(B_{t_i} - B_{t_{i-1}})$.

c) $\sum_{i=1}^{N} f(t_{i-1})(B_{t_i} - B_{t_{i-1}})$.

Then, show that in each case the limit of the calculated quantities as $\delta \to 0$ is the same.

Itô integral extends Wiener integral to stochastic integrands. The integrand f is now a function of time $t \in [a, b]$ and ω in sample space Ω. For our analysis in this notes, we only need to define Itô integral on the integrands of the form $f(t, B_t)$ where $f : [a, b] \times \mathbb{R} \to \mathbb{R}$ is a measurable function. Similar to the Winer integral we start with the partial sum

$$\sum_{i=0}^{N-1} f(t_i, B_{t_i})(B_{t_{i+1}} - B_{t_i}).$$

Exercise C.3. *Calculate that the mean and the variance of partial sums below:*

a) $\sum_{i=1}^{N} B_{t_i}(B_{t_i} - B_{t_{i-1}})$.

b) $\sum_{i=1}^{N} B_{\frac{t_i + t_{i-1}}{2}}(B_{t_i} - B_{t_{i-1}})$.

c) $\sum_{i=1}^{N} B_{t_{i-1}}(B_{t_i} - B_{t_{i-1}})$.

Then, show that in all cases the limits of the calculated quantities as $\delta \to 0$ *are different.*

$$\int_0^T f(u, B_u) dB_u := \mathbb{P} - \lim_{\delta \to 0} \sum_{i=0}^{N-1} f(t_i, B_{t_i})(B_{t_{i+1}} - B_{t_i}). \tag{C.1}$$

The notation $\mathbb{P} - \lim$ means the limit is in probability, i.e., for any $\varepsilon > 0$,

$$\mathbb{P}\left(\Big| \sum_{i=0}^{N-1} f(t_i, B_{t_i})(B_{t_{i+1}} - B_{t_i}) - \int_0^T f(u, B_u) dB_u \Big| > \varepsilon \right) \to 0 \quad \text{as} \quad \delta \to 0.$$

The choice of starting point t_i in the interval $[t_i, t_{i+1}]$ in $f(t_i, B_{t_i})$ is crucial. This is because choosing other point in the interval $[t_i, t_{i+1}]$ leads to different limits. Fo instance,

$$\sum_{i=0}^{N-1} f\Big(\frac{t_i + t_{i+1}}{2}, \frac{B_{t_i + t_{i+1}}}{2}\Big)(B_{t_{i+1}} - B_{t_i})$$

converges to

$$\int_0^T f(u, B_u) dB_u + \int_0^T \partial_x f(u, B_u) du.$$

Martingale property of stochastic integral

Consider the discrete sum which converges to the stochastic integral, i.e.

$$M_T := \sum_{i=0}^{N-1} f(t_i, B_{t_i})(B_{t_{i+1}} - B_{t_i})$$

Assume that the values of $B_0, ..., B_{t_j}$ are given. We want to evaluate the conditional expectation of the stochastic sum M_T, i.e.

$$\mathbb{E}[M_T \mid B_0, ..., B_{t_j}].$$

Then, we split the stochastic sum into to parts

$$M_T := \sum_{i=0}^{j-1} f(t_i, B_{t_i})(B_{t_{i+1}} - B_{t_i}) + \sum_{i=j}^{N-1} f(t_i, B_{t_i})(B_{t_{i+1}} - B_{t_i}).$$

The first summation of the right-hand side above is known given $B_0, ..., B_{t_j}$. Thus,

$$\begin{aligned}\mathbb{E}[M_T \mid B_0, ..., B_{t_j}] &= \sum_{i=0}^{j-1} f(t_i, B_{t_i})(B_{t_{i+1}} - B_{t_i}) + \mathbb{E}\left[\sum_{i=j}^{N-1} f(t_i, B_{t_i})(B_{t_{i+1}} - B_{t_i}) \mid B_0, ..., B_{t_j}\right] \\ &= \sum_{i=0}^{j-1} f(t_i, B_{t_i})(B_{t_{i+1}} - B_{t_i}) + \sum_{i=j}^{N-1} \mathbb{E}\left[f(t_i, B_{t_i})(B_{t_{i+1}} - B_{t_i}) \mid B_0, ..., B_{t_j}\right]\end{aligned}$$

Each term in the second summation of right-hand side above can be calculated by tower property of conditional expectation

$$\mathbb{E}\left[f(t_i, B_{t_i})(B_{t_{i+1}} - B_{t_i}) \mid B_0, ..., B_{t_j}\right] = \mathbb{E}\left[f(t_i, B_{t_i})\mathbb{E}[B_{t_{i+1}} - B_{t_i} \mid B_0, ..., B_{t_i}] \mid B_0, ..., B_{t_j}\right].$$

Since $B_{t_{i+1}} - B_{t_i}$ is independent of $B_0, ..., B_{t_i}$,

$$\mathbb{E}[B_{t_{i+1}} - B_{t_i} \mid B_0, ..., B_{t_i}] = \mathbb{E}[B_{t_{i+1}} - B_{t_i}] = 0.$$

This implies that the second summation vanishes and we have

$$\mathbb{E}[M_T \mid B_0, ..., B_{t_j}] = \sum_{i=0}^{j-1} f(t_i, B_{t_i})(B_{t_{i+1}} - B_{t_i}) =: M_{t_j}.$$

In other words, given the Brownian motion up to time t_j, the expected values of M_T is equal to M_{t_j}. In probability terms, we call this a *martingale*. By some more technical tools, one can show that given the path of a Brownian motion until time $t < T$, the expected value of the stochastic integral $\int_0^T f(s, B_s)dB_s$ is equal to $\int_0^t f(s, B_s)dB_s$, i.e.

$$\mathbb{E}\left[\int_0^T f(s, B_s)dB_s \middle| B_s \quad \text{for} \quad s \in [0, t]\right] = \int_0^t f(s, B_s)dB_s.$$

One of the consequence of martingale property is that the expectation of stochastic integral is zero, i.e.

$$\mathbb{E}\left[\int_0^T f(s, B_s)dB_s\right] = \int_0^0 f(s, B_s)dB_s = 0.$$

Remark C.1. *The martingale property of the stochastic integral with respect to Brownian motion is basically a result of martingale property of Brownian motion. Riemann integral* $\int_0^t f(s, B_s)ds$ *is a martingale if and only if* $f \equiv 0$. *Intuitively, if we assume* $\int_0^t f(s, B_s)ds$ *is a martingale, we have*

$$\mathbb{E}\Big[\int_t^{t+\delta} f(s, B_s)ds \Big| \mathcal{F}_t\Big] = 0.$$

By dividing both sides by δ and then sending $\delta \to 0$, we obtain

$$\lim_{\delta\to 0}\mathbb{E}\Big[\frac{1}{\delta}\int_t^{t+\delta} f(s,B_s)ds\Big|\mathcal{F}_t\Big] = \mathbb{E}[f(t,B_t) \mid \mathcal{F}_t] = f(t,B_t).$$

C.2 Itô formula

One of the important implications of Itô integral is a very powerful tool called *Itô formula*. Itô formula is the stochastic version of Taylor expansion. To understand this better let's try to write Taylor expansion for $V(t+\delta, B_{t+\delta})$ about the point $V(t, B_t)$.

$$V(t+\delta,B_{t+\delta}) = V(t,B_t) + \delta\partial_t V(t,B_t) + \partial_x V(t,B_t)(B_{t+\delta}-B_t) + \frac{1}{2}\partial_{xx}V(t,B_t)(B_{t+\delta}-B_t)^2 + o(\delta). \tag{C.2}$$

The remaining term in of order $o(\delta)$ since $B_{t+\delta} - B_t \sim O(\sqrt{\delta})$. Also, we know that

$$(B_{t+\delta}-B_t)^2 - \delta \sim o(\delta), \ \ (B_{t+\delta}-B_t)\delta \sim o(\delta), \quad \text{and trivially } \ \delta^2 \sim o(\delta). \tag{C.3}$$

If we take conditional expectation with respect to B_t, we obtain

$$\begin{aligned}\mathbb{E}[V(t+\delta,B_{t+\delta}) \mid B_t] =& V(t,B_t) + \delta\partial_t V(t,B_t) + \partial_x V(t,B_t)\mathbb{E}[(B_{t+\delta}-B_t) \mid B_t] \\ &+ \frac{1}{2}\partial_{xx}V(t,B_t)\delta + o(\delta). \\ =& V(t,B_t) + (\partial_t V(t,B_t) + \frac{1}{2}\partial_{xx}V(t,B_t))\delta + o(\delta)\end{aligned}$$

Notice that here, by the martingale property of Brownian motion, we have $\mathbb{E}[(B_{t+\delta}-B_t) \mid B_t] = 0$. Then, we obtain

$$\mathbb{E}[V(T,B_T)] = V(0,B_0) + \mathbb{E}\left[\sum_{i=0}^{N-1}(\partial_t V(t_i,B_{t_i}) + \frac{1}{2}\partial_{xx}V(t_i,B_{t_i}))\delta\right] + o(1),$$

which in the limit converges to

$$\mathbb{E}[V(T,B_T)] = V(0,B_0) + \mathbb{E}\left[\int_0^T (\partial_t V(t_i,B_t) + \frac{1}{2}\partial_{xx}V(t,B_t))dt\right].$$

The above formula is called Dynkin formula. If we don't take conditional expectation, we can write

$$V(t+\delta,B_{t+\delta}) = V(t,B_t) + (\partial_t V(t,B_t) + \frac{1}{2}\partial_{xx}V(t,B_t))\delta + \partial_x V(t,B_t)(B_{t+\delta}-B_t) + o(\delta).$$

Then, we obtain

$$V(T,B_T) = V(0,B_0)+\sum_{i=0}^{N-1}(\partial_t V(t_i,B_{t_i})+\frac{1}{2}\partial_{xx}V(t_i,B_{t_i}))\delta+\sum_{i=0}^{N-1}\partial_x V(t_i,B_{t_i})(B_{t_{i+1}}-B_{t_i})+o(1).$$

which in the limit converges to

$$V(T,B_T) = V(0,B_0) + \int_0^T (\partial_t V(t,B_t) + \frac{1}{2}\partial_{xx}V(t,B_t))dt + \int_0^T \partial_x V(t,B_t)dB_t. \quad \text{(C.4)}$$

In the above, (C.4) is referred to as Itô formula.
In a less formal way, Utô formula is given by

$$dV(t,B_t) = (\partial_t V(t,B_t) + \frac{1}{2}\partial_{xx}V(t,B_t))dt + \partial_x V(t,B_t)dB_t.$$

However, it has to be interpreted as (C.4).

Exercise C.4. *Use Itô formula to calculate $dV(t,B_t)$ in the following cases.*

a. $V(t,x) = e^{ax}$

b. $V(t,x) = e^{-t}e^{ax}$

c. $V(t,x) = e^{-t}\cos(x)$

d. $V(t,x) = e^{-t}x^a$

where a is a given constant.

C.3 Martingale property of stochastic integral and partial differential equations

Why martingale property of stochastic integral is important? Recall the Itô formula

$$dV(t,B_t) = (\partial_t V(t,B_t) + \frac{1}{2}\partial_{xx}V(t,B_t))dt + \partial_x V(t,B_t)dB_t.$$

Assume that $V(t,x)$ satisfies the PDE

$$\partial_t V(t,x) + \frac{1}{2}\partial_{xx}V(t,x) = 0. \quad \text{(C.5)}$$

Then,

$$dV(t,B_t) = \partial_x V(t,B_t)dB_t.$$

and thus $V(t,B_t)$ is a martingale, i.e.

$$V(t,B_t) = \mathbb{E}[V(T,B_T) \mid B_s \quad \text{for} \quad s \in [0,t]].$$

Conversely, if $V(t,B_t)$ is a martingale, then $V(t,x)$ must satisfy the PDE (C.5).

C.4 Stochastic integral and Stochastic differential equation

Riemann integral allows us to write a differential equation $\frac{dx}{dt} = f(t,x(t)),\ x(0) = x_0$ as an integral equation

$$x(t) = x_0 + \int_0^t f(s,x(s))ds. \tag{C.6}$$

Integral equations are more general because the solution does not necessarily need to be differentiable. For instance if

$$f(t,x) = \begin{cases} 1 & t \geqslant 1 \\ 0 & t < 1 \end{cases}.$$

The solution to the integral equation is $x(t) = x_0 + (t-1)_+$, which is not differentiable at $t = 1$. While $dx = f(t,x(t))dt$ should be interpreted as (C.6).

Itô integral allows us to define *stochastic differential equations* (SDE for short) in integral form. For example, the Black-Scholes differential equation is given by

$$\frac{dS_t}{S_t} = rdt + \sigma dB_t.$$

The true meaning of this term is

$$S_t = S_0 + r\int_0^t S_u du + \sigma \int_0^t S_u dB_u.$$

The solution is a stochastic process S_t which satisfies the SDE. In the above case, it is easy to verify, by means of Itô formula, that the geometric Brownian motion

$$S_t = S_0 \exp\left((r - \frac{1}{2}\sigma^2)t + \sigma B_t\right).$$

Take $V(t,x) = S_0 \exp\left((r-\frac{1}{2}\sigma^2)t + \sigma x\right)$. It follows from Itô formula that

$$V(t,B_t) = V(0,B_0) + \int_0^t (\partial_t V(s,B_s) + \frac{1}{2}\partial_{xx}V(s,B_s))dt + \int_0^t \partial_x V(s,B_s)dB_s.$$

Since $V(0,x) = S_0$, $\partial_t V(s,x) = (r - \frac{1}{2}\sigma^2)V(s,x)$, $\partial_x V(s,x) = \sigma V(s,x)$, and $\partial_{xx}V(s,x) =$

$\sigma^2 V(s,x)$, we have

$$V(t,B_t) = S_0 + r\int_0^t V(s,B_s)ds + \sigma\int_0^t V(s,B_s)dB_s.$$

In general for a pair of given functions $\mu(t,x)$ and $\sigma(t,x)$, an equation of the form

$$dS_t = \mu(t,S_t)dt + \sigma(t,S_t)dB_t \tag{C.7}$$

is called a *stochastic differential equation* (SDE). A solution S_t is a process such that the Lebesgue integral

$$\int_0^t \mu(s,S_s)ds < \infty \ \ \mathbb{P}\text{-a.s.},$$

the Itô integral

$$\int_0^t \sigma(s,S_s)dB_s$$

is well-defined, and the following is satisfied:

$$S_t = S_0 + \int_0^t \mu(s,S_s)ds + \int_0^t \sigma(s,S_s)dB_s.$$

For CEV model in Section **??**, SDE is written as

$$\frac{dS_t}{S_t} = rdt + \sigma S_t^\beta dB_t$$

or in integral form

$$S_t = S_0 + r\int_0^t S_u du + \sigma\int_0^t S_u^{1+\beta}dB_u.$$

One of the most important applications of the Itôs formula is the chain rule in the stochastic form. Consider SDE

$$dS_t = \mu(t,S_t)dt + \sigma(t,S_t)dB_t \tag{C.8}$$

and $V(t,S_t)$ where $V(t,x)$ is a function one time continuously differentiable in t and twice continuously differentiable on x. Then we can write

$$V(t+\delta,S_{t+\delta}) = V(t,S_t) + \partial_t V(t,S_t)\delta + \partial_x V(t,S_t)(S_{t+\delta}-S_t) + \frac{1}{2}\partial_{xx}V(t,S_t)(S_{t+\delta}-S_t)^2 + o(\delta).$$

Notice that by (C.8), we have

$$S_{t+\delta} - S_t \approx \mu(t,S_t)\delta + \sigma(t,S_t)(B_{t+\delta}-B_t).$$

Therefore, we have

$$(S_{t+\delta} - S_t)^2 = \sigma^2(t, S_t)(B_{t+\delta} - B_t)^2 + o(\delta).$$

and

$$\begin{aligned}
V(t+\delta, S_{t+\delta}) &= V(t, S_t) + \partial_t V(t, S_t)\delta + \partial_x V(t, S_t)(\mu(t, S_t)\delta + \sigma(t, S_t)(B_{t+\delta} - B_t)) \\
&\quad + \frac{1}{2}\partial_{xx} V(t, S_t)(\mu(t, S_t)\delta + \sigma(t, S_t)(B_{t+\delta} - B_t))^2 + o(\delta) \\
&= V(t, S_t) + \partial_t V(t, S_t)\delta + \partial_x V(t, S_t)(\mu(t, S_t)\delta + \sigma(t, S_t)(B_{t+\delta} - B_t)) \\
&\quad + \frac{1}{2}\partial_{xx} V(t, S_t)\sigma^2(t, S_t)\delta + o(\delta) \\
&= V(t, S_t) + (\partial_t V(t, S_t) + \partial_x V(t, S_t)\mu(t, S_t) + \frac{1}{2}\partial_{xx} V(t, S_t)\sigma^2(t, S_t))\delta \\
&\quad + \sigma(t, S_t)\partial_x V(t, S_t)(B_{t+\delta} - B_t)) + o(\delta).
\end{aligned}$$

Or in the integral form we have

$$\begin{aligned}
V(t, S_t) =& V(0, S_0) + \int_0^t \Big(\partial_t V(u, S_u) + \partial_x V(u, S_u)\mu(u, S_u) + \frac{1}{2}\partial_{xx} V(u, S_u)\sigma^2(u, S_u)\Big) du \\
&+ \int_0^T \sigma(u, S_u)\partial_x V(u, S_u) dB_u.
\end{aligned} \tag{C.9}$$

Proposition C.1. *If the function $V(t, S)$ is twice continuously differentiable and $\{V(t, S_t)\}_{t=0}^T$ is a martingale, then*

$$\partial_t V(t, S) + \partial_x V(t, S)\mu(t, S) + \frac{1}{2}\partial_{xx} V(t, S)\sigma^2(t, S) = 0$$

for all $t \in [0, T]$ and S.

Proof. By the Itô formula (C.9), we have

$$\begin{aligned}
&\int_0^t \Big(\partial_t V(u, S_u) + \partial_x V(u, S_u)\mu(u, S_u) + \frac{1}{2}\partial_{xx} V(u, S_u)\sigma^2(u, S_u)\Big) du \\
&\quad = V(t, S_t) - V(0, S_0) - \int_0^T \sigma(u, S_u)\partial_x V(u, S_u) dB_u.
\end{aligned}$$

Since the right-hand side is a martingale, so is the left-hand side. However, a Riemann integral on the left cannot be a martingale unless

$$\partial_t V(t, S_t) + \partial_x V(t, S_t)\mu(t, S_t) + \frac{1}{2}\partial_{xx} V(t, S_t)\sigma^2(t, S_t) = 0.$$

This is because if the above term is not zero at some time t, e.g. positive, the Riemann integral becomes increasing for a short interval $[t, t+\delta]$, which contradicts the martingale property. □

C.5 Itô calculus

The calculations in the previous section can be obtained from a formal calculus. First, we formally write (C.2) as

$$dV(t, B_t) = \partial_t V(t, B_t)dt + \partial_x V(t, B_t)dB_t + \frac{1}{2}\partial_{xx} V(t, B_t)(dB_t)^2.$$

Then, we present (C.3) in formal form of

$$(dB_t)^2 = dt, \;\; dB_t dt = dt dB_t = 0 \;\; , \quad \text{and} \quad (dt)^2 = 0. \tag{C.10}$$

which implies the Itô formula for Brownian motion.

$$dV(t, B_t) = (\partial_t V(t, B_t) + \frac{1}{2}\partial_{xx} V(t, B_t))dt + \partial_x V(t, B_t)dB_t.$$

For the Itô formula for process S_t in (C.8), we can formally write

$$dV(t, S_t) = \partial_t V(t, S_t)dt + \partial_x V(t, S_t)dS_t + \frac{1}{2}\partial_{xx} V(t, S_t)(dS_t)^2.$$

Then, we use (C.10) to obtain

$$(dS_t)^2 = \mu^2(dt)^2 + 2\mu\sigma dt dB_t + \sigma^2(dB_t)^2 = \sigma^2 dt.$$

Thus,

$$dV(t, S_t) = (\partial_t V(t, S_t) + \frac{\sigma^2(t, S_t)}{2}\partial_{xx} V(t, S_t))dt + \partial_x V(t, S_t)dS_t.$$

Theorem C.1 (Itô formula)**.** *Consider S_t given by (C.8) and assume that function V is once continuously differentiable in t and twice continuously differentiable in x. Then, we have*

$$dV(t, S_t) = (\partial_t V(t, S_t) + \mu(t, S_t)\partial_x V(t, S_t) + \frac{\sigma^2(t, S_t)}{2}\partial_{xx} V(t, S_t))dt + \sigma(t, S_t)\partial_x V(t, S_t)dB_t.$$

Similar to Section C.3, we can use martingale property of stochastic integral to obtain a PDE. More precisely, $V(t, S_t)$ is a martingale if and only if $V(t, x)$ satisfies

$$\partial_t V(t, x) + \mu(t, x)\partial_x V(t, x) + \sigma^2(t, x)\frac{1}{2}\partial_{xx} V(t, x) = 0.$$

Exercise C.5. *Let*

$$dS_t = S_t dt + 2S_t dB_t.$$

Calculate $dV(t, S_t)$ in the following cases.

a. $V(t,x) = e^{ax}$

b. $V(t,x) = e^{-t}e^{ax}$

c. $V(t,x) = e^{-t}\cos(x)$

d. $V(t,x) = e^{-t}x^a$

where a is a constant.

Exercise C.6. *In each of the following SDE, find the PDE for the function $V(t,x)$ such that $V(t, S_t)$ is a martingale.*

a. $dS_t = \sigma dB_t$ *where σ is constants.*

b. $dS_t = \kappa(m - S_t)dt + \sigma dB_t$ *where κ, m and σ are constants.*

c. $dS_t = \kappa(m - S_t)dt + \sigma\sqrt{S_t}dB_t$ *where κ, m and σ are constants.*

d. $dS_t = rS_t dt + \sigma S_t^2 dB_t$ *where r and σ are constants.*

Exercise C.7. *Consider the SDE*

$$dS_t = rS_t dt + \sigma S_t dB_t \quad \text{where } r \text{ and } \sigma \text{ are constants.}$$

a. *Find the ODE for the function $V(x)$ such that $e^{-rt}V(S_t)$ is a martingale.*

b. *Find all the solutions to the ODE in (a).*

Bibliography

[1] M. R. Adams and V. Guillemin, Measure theory and probability, Springer, 1996.

[2] L. Bachelier, Théorie de la spéculation, Gauthier-Villars, 1900.

[3] D. P. Bertsekas and S. E. Shreve, Stochastic optimal control: the discrete time case, Mathematics in Science and Engineering, Elsevier, Burlington, MA, 1978.

[4] P. Billingsley, Probability and measure, John Wiley & Sons, 2008.

[5] F. Black and M. Scholes, The pricing of options and corporate liabilities, The journal of political economy, (1973), pp. 637–654.

[6] B. Bouchard, M. Nutz, et al., Arbitrage and duality in nondominated discrete-time models, The Annals of Applied Probability, 25 (2015), pp. 823–859.

[7] B. Bouchard and X. Warin, Monte-carlo valuation of american options: facts and new algorithms to improve existing methods, in Numerical methods in finance, Springer, 2012, pp. 215–255.

[8] S. P. Boyd and L. Vandenberghe, Convex optimization (pdf), Np: Cambridge UP, (2004).

[9] F. Delbaen and W. Schachermayer, A general version of the fundamental theorem of asset pricing, Mathematische Annalen, 300 (1994), pp. 463–520.

[10] R. Durrett, Probability. theory and examples. the wadsworth & brooks/cole statistics/probability series, Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, (1991).

[11] L. Eisenberg and T. H. Noe, Systemic risk in financial systems, Management Science, 47 (2001), pp. 236–249.

[12] A. Fahim, N. Touzi, and X. Warin, A probabilistic numerical method for fully nonlinear parabolic pdes, The Annals of Applied Probability, (2011), pp. 1322–1364.

[13] H. Föllmer and A. Schied, Stochastic finance: an introduction in discrete time, Walter de Gruyter, 2011.

[14] J. Franke, W. K. Härdle, and C. M. Hafner, Statistics of financial markets, vol. 2, Springer, 2004.

[15] P. Glasserman, Monte Carlo methods in financial engineering, vol. 53, Springer Science & Business Media, 2003.

[16] J. Hull, Options, Futures and Other Derivatives, Options, Futures and Other Derivatives, Pearson/Prentice Hall, 2015.

[17] I. Karatzas and S. Shreve, Brownian motion and stochastic calculus, vol. 113, Springer Science & Business Media, 2012.

[18] P.-S. Laplace, Essai philosophique sur les probabilités, H. Remy, 1829.

[19] F. A. Longstaff and E. S. Schwartz, Valuing american options by simulation: a simple least-squares approach, Review of Financial studies, 14 (2001), pp. 113–147.

[20] H. Markowitz, Portfolio selection, The journal of finance, 7 (1952), pp. 77–91.

[21] R. C. Merton, An analytic derivation of the efficient portfolio frontier, Journal of financial and quantitative analysis, 7 (1972), pp. 1851–1872.

[22] ——, Theory of rational option pricing, The Bell Journal of economics and management science, (1973), pp. 141–183.

[23] B. Oksendal, Stochastic differential equations: an introduction with applications, Springer Science & Business Media, 2013.

[24] G. Peskir and A. Shiryaev, Optimal stopping and free-boundary problems, Springer, 2006.

[25] P. Protter, A partial introduction to financial asset pricing theory, Stochastic processes and their applications, 91 (2001), pp. 169–203.

[26] W. Schachermayer and J. Teichmann, How close are the option pricing formulas of bachelier and black–merton–scholes?, Mathematical Finance, 18 (2008), pp. 155–170.

[27] S. Shreve, Stochastic calculus for finance I: the binomial asset pricing model, Springer Science & Business Media, 2012.

[28] S. E. Shreve, Stochastic calculus for finance II: Continuous-time models, vol. 11, Springer Science & Business Media, 2004.

[29] M. S. Taqqu, Bachelier and his times: a conversation with bernard bru, Finance and Stochastics, 5 (2001), pp. 3–32.

[30] N. Touzi and P. Tankov, No-arbitrage theory for derivative pricing, Cours de l'Ecole polytechnique, (2008).

[31] D. Williams, Probability with martingales, Cambridge university press, 1991.

Index