Introduction to Options Pricing Theory

Simone Calogero

March 15, 2016

Preface

This text presents a self-contained introduction to the options pricing theory based on the binomial model and on the Black-Scholes model. It is the main literature for the course "Options and Mathematics" at Chalmers, which provides the students with a first rudimentary knowledge in Mathematical Finance (in particular, without using Stochastic Calculus). The pre-requisites to follow this text are the standard basic courses in mathematics, such as Calculus and Linear Algebra. No previous knowledge on probability theory and finance are required. Each chapter is complemented with a number of exercises and Matlab codes. The exercises marked with the symbol (?) aim to critical thinking and do not necessarily have a well-defined unique solution. All other exercises will be solved in the class and the solution of the more involved ones will be uploaded thereafter on the course homepage. The latter can be found at

`http://www.math.chalmers.se/Math/Grundutb/CTH/mve095/1516/`

Remark: The Matlab codes presented in this text are not optimized. Moreover the powerful vectorization tools of Matlab are not employed, so as to make the codes easily adaptable to other computer softwares and languages. The task to improve the codes presented in this text is left to the interested reader.

Contents

Chapter 1

Warm-up

The purpose of this introductory chapter is threefold: (1) present a few basic financial concepts, (2) formulate the main assumptions for the mathematical models to be studied in the following chapters, (3) derive some fundamental qualitative properties of option prices.

1.1 Basic financial concepts

For a more detailed discussion on the concepts introduced in this section, see [2].

Financial assets

The term **financial asset** may be used to identify any object which, under specific regulations, can be bought and sold. This "definition" includes material assets (such as gold, oil, corn, etc.) and immaterial assets (e.g., stocks). A generic asset will be denoted by $\mathcal{U}$. The **price** of the asset is agreed by the **buyer** and the **seller** as a result of some kind of "negotiation". More precisely, the **ask price** is the minimum price at which the seller is willing to sell the asset, while the **bid price** is the maximum price that the buyer is willing to pay for the asset. When the difference between these two values, called **bid-ask spread**, becomes zero, the exchange of the asset takes place at the corresponding price. The price at time t of the asset $\mathcal{U}$ is denoted by $\Pi^{\mathcal{U}}(t)$. The asset price refers to the price of a **share** of the asset, where "share" stands for the minimum amount of an asset which can be traded (e.g., 1 ounce of gold).

Markets

Assets can be traded in **official** markets or in **over the counter (OTC)** markets. In the former case all trades are subject to a common legislation, while in the latter the exchange conditions are agreed upon by the individual traders. The most active regularized official markets are the stock markets, the option markets and the futures markets. Since official markets are more easily monitored, it is convenient to define the **market price** of an asset as the price (per share) of the asset in the official market where it is traded. In the following the market price of an asset will be referred to simply as the price (or value) of the asset. Buyers and sellers of assets in a market will be called **investors** or **agents**.

Trading

Besides the usual operations of "buying" and "selling" the asset, we need to consider an additional type of transaction, which is called **short-selling**. Short-selling an asset (typically a stock) is the practice of selling the asset without actually owning it. Concretely, an investor is short-selling N shares of an asset if the investor borrows the shares from a third party and then sell them immediately on the market. The reason for short-selling an asset is the expectation that the price of the asset will decrease in the future. More precisely, assume that N shares of an asset $\mathcal{U}$ are short-sold at time $t = 0$ for the price $\Pi^{\mathcal{U}}(0)$ and let $T > 0$ be the time at which the shares must be returned to their original owner. If $\Pi^{\mathcal{U}}(T) < \Pi^{\mathcal{U}}(0)$, then upon re-purchasing the N shares at time T, and returning them to the lender, the short-seller will make the profit $N(\Pi^{\mathcal{U}}(0) - \Pi^{\mathcal{U}}(T))$.

An investor is said to have a **long position** on an asset if the investor owns the asset and will therefore profit from an increase of the price of the asset. Conversely, the investor is said to have a **short position** on the asset if the investor will profit from a decrease of its value, as it happens for instance when the investor is short-selling the asset.

Finally we remark that any transaction in the market is subject to **transaction costs** (e.g., broker's commissions and lending fees for short-selling) and **transaction delays** (trading in real markets is not instantaneous).

Stocks

The **capital stock** of a company is the part of the company capital that is made publicly available for trading. Holding shares of a stock is equivalent to own a fraction of the company. Stocks are traded in the **stock markets**. For instance, over 300 company stocks are traded in the Stockholm stock exchange. The **index** of a stock market is a weighted average of the value of the stocks traded in this market. For instance, S&P500 (Standard and Poor 500) refers to 500 stock companies traded at the New York stock exchange (NYSE) and at the NASDAQ market. The price per share at time $t > 0$ of a generic stock will be denoted by $S(t)$.

Dividends

An asset (typically a stock) may occasionally pay a **dividend** to the shareholders. This means that a fraction of the asset price is deposited to the bank account of the shareholders. The day at which the dividend is paid, as well as its amount in percentage of the opening asset price at that day, are known in advance. After the dividend has been paid, the price of the asset diminishes of exactly the amount paid by the dividend.

Portfolio Position and Portfolio Process

Consider an agent that invests on N assets $\mathcal{U}_1,\dots,\mathcal{U}_N$. Assume that the agent trades on a_1 shares of the asset $\mathcal{U}_1$, a_2 shares of the asset $\mathcal{U}_2$,..., a_N shares of the asset $\mathcal{U}_N$. Here $a_i \in \mathbb{Z}$, where $a_i < 0$ means that the investor has a short position in the asset $\mathcal{U}_i$, while $a_i > 0$ means that the investor has a long position in the asset $\mathcal{U}_i$ (the reason for this interpretation will become soon clear). The vector $\mathcal{A} = (a_1, a_2, \dots, a_N) \in \mathbb{Z}^N$ is called a **portfolio position**, or simply a portfolio. The **value** of the portfolio at time t is given by

$$V_{\mathcal{A}}(t) = \sum_{i=1}^{N} a_i \Pi^{\mathcal{U}_i}(t), \tag{1.1}$$

where $\Pi^{\mathcal{U}_i}(t)$ denotes the price of the asset $\mathcal{U}_i$ at time t. The value of the portfolio measures the wealth of the investor: the higher is $V(t)$, the "richer" is the investor at time t. Now we see that when the price of the asset $\mathcal{U}_i$ increases, the value of the portfolio increases if $a_i > 0$ and decreases if $a_i < 0$, which explains why $a_i > 0$ corresponds to a long position on the asset $\mathcal{U}_i$ and $a_i < 0$ to a short position. We also remark that portfolios can be added by using the linear structure on $\mathbb{Z}^N$, namely if $\mathcal{A}, \mathcal{B} \in \mathbb{Z}^N$, $\mathcal{A} = (a_1, \dots, a_N)$, $\mathcal{B} = (b_1, \dots, b_N)$ are two portfolios and $\alpha, \beta \in \mathbb{Z}$, then $\mathcal{C} = \alpha\mathcal{A} + \beta\mathcal{B}$ is the portfolio $\mathcal{C} = (\alpha a_1 + \beta b_1, \dots, \alpha a_N + \beta b_N)$.

Suppose now that the investor changes the position on the assets during the interval $[0, T]$. Due to transaction delays, this can be done only at a finite number of times $0 = t_0 < t_1 < t_2 < \dots t_M = T$; for simplicity we assume that at each time $t_1, \dots t_M$ the change in the portfolio position occurs instantaneously. Let $\mathcal{A}_0$ denotes the initial (at time $t = t_0 = 0$) portfolio position of the investor and $\mathcal{A}_j$ denote the portfolio position of the investor in the interval of time $(t_{j-1}, t_j]$, $j = 1, \dots M$. The vector $(\mathcal{A}_1, \dots \mathcal{A}_M)$ is called a **portfolio process**. If we denote by a_{ij} the number of shares of the asset i in the portfolio $\mathcal{A}_j$, then we see that a portfolio process is in fact equivalent to the $N \times M$ matrix $A = (a_{ij})$, $i = 1, \dots, N$, $j = 1, \dots M$. The value $V(t)$ of the portfolio process at the time t is given by the value of the corresponding portfolio position at time t as defined by (1.1). Hence for $t \in (t_{j-1}, t_j]$ and $j = 1, \dots M$ the value of the portfolio process is given by

$$V(t) = V_{\mathcal{A}_j}(t) = \sum_{i=1}^{N} a_{ij} \Pi^{\mathcal{U}_i}(t),$$

while for $t = 0$ the value of the portfolio process is $V(0) = V_{\mathcal{A}_0}$. The initial value of a portfolio, when it is positive, is called the **initial wealth** of the investor.

A portfolio process is said to be **self-financing** if no cash is ever withdrawn or infused in the portfolio. Let us give an example. Suppose that at time $t_0 = 0$ the investor is short 400 shares on the asset $\mathcal{U}_1$, long 200 shares on the asset $\mathcal{U}_2$ and long 100 shares on the asset $\mathcal{U}_3$. This corresponds to the portfolio

$$\mathcal{A}_0 = (-400, 200, 100),$$

whose value is

$$V_{\mathcal{A}_0} = -400\, \Pi^{\mathcal{U}_1}(0) + 200\, \Pi^{\mathcal{U}_2}(0) + 100\, \Pi^{\mathcal{U}_3}(0).$$

If this value is positive, the investor needs an initial wealth to set up this portfolio position: the income deriving from short selling the asset $\mathcal{U}_1$ does not suffice to open the desired long position on the other two assets.

Assume that the investor keeps the same position in the interval $(0, 1]$, i.e., $\mathcal{A}_1 = \mathcal{A}_0$. The value of this portfolio at time $t = 1$ is

$$V_{\mathcal{A}_1}(1) = -400\,\Pi^{\mathcal{U}_1}(1) + 200\,\Pi^{\mathcal{U}_2}(1) + 100\,\Pi^{\mathcal{U}_3}(1).$$

Now suppose that at time $t = 1$ the investor buys 500 shares of $\mathcal{U}_1$, sells x shares of $\mathcal{U}_2$, and sells all the shares of $\mathcal{U}_3$. Then in the interval $(1, 2]$ the investor has a new portfolio which is given by

$$\mathcal{A}_2 = (100, 200 - x, 0),$$

whose value at time $t = 1$ is

$$V_{\mathcal{A}_2}(1) = 100\,\Pi^{\mathcal{U}_1}(1) + (200 - x)\,\Pi^{\mathcal{U}_2}(1).$$

The portfolio process is the pair $(\mathcal{A}_1, \mathcal{A}_2)$, or, equivalently, the matrix

$$A = \begin{pmatrix} -400 & 100 \\ 200 & 200 - x \\ 100 & 0 \end{pmatrix}.$$

We recall that changing the portfolio from $\mathcal{A}_1$ to $\mathcal{A}_2$ is not instantaneous (due to transaction delays) but we shall ignore this. The difference between the value of the two portfolios immediately after and immediately before the transaction is then

$$\begin{aligned} V_{\mathcal{A}_2}(1) - V_{\mathcal{A}_1}(1) &= 100\,\Pi^{\mathcal{U}_1}(1) + (200 - x)\,\Pi^{\mathcal{U}_2}(1) \\ &\quad - (-400\,\Pi^{\mathcal{U}_1}(1) + 200\,\Pi^{\mathcal{U}_2}(1) + 100\,\Pi^{\mathcal{U}_3}(1)) \\ &= 500\,\Pi^{\mathcal{U}_1}(1) - x\,\Pi^{\mathcal{U}_2}(1) - 100\,\Pi^{\mathcal{U}_3}(1). \end{aligned}$$

If this difference is positive, then the new portfolio cannot be created from the old one without extra cash. Viceversa, if this difference is negative, then the new portfolio is less valuable than the old one, the difference being equivalent to cash withdrawn from the portfolio. Hence for self-financing portfolio processes, we must have $V_{\mathcal{A}_2}(1) - V_{\mathcal{A}_1}(1) = 0$, i.e.,

$$x = \frac{500\Pi^{\mathcal{U}_1}(1) - 100\Pi^{\mathcal{U}_3}(1)}{\Pi^{\mathcal{U}_2}(1)}.$$

Of course, x will be an integer only in exceptional cases, which means that perfect self-financing strategies in real markets are almost impossible.

The **return** of a portfolio process in the interval $[0, T]$ is defined as

$$R(T) = V(T) - V(0), \tag{1.2}$$

where $V(t)$ denotes the value of the portfolio at time t, defined as above. If the return is positive, the investor makes a profit in the interval $[0, T]$, if it is negative the investor incurs in a loss. When $V(0) > 0$ we may introduce also the **relative return** of the portfolio given by

$$R_*(T) = \frac{V(T) - V(0)}{V(0)}. \tag{1.3}$$

Finally we remark that investment returns are commonly "annualized" by dividing the return $R(T)$ by the time T expressed in fraction of years (e.g., $T = 1$ week $= 1/52$ years).

Historical volatility

The historical volatility is a measure of how "wildly" an asset price changes in time, thereby giving information on its level of uncertainty. It is computed as the standard deviation of the log-returns of the asset based on historical data. More precisely, let $[t_0, t]$ be some interval of time in the past, with t denoting possibly the present time, and let $T = t - t_0 > 0$ be the length of this interval. Let us divide $[t_0, t]$ into n equally long periods, say

$$t_0 < t_1 < t_2 < \dots t_n = t, \quad t_i - t_{i-1} = h, \quad \text{for all } i = 1, \dots n.$$

Assume for instance that the asset is a stock. The **log-return** of the stock price in the interval $[t_{i-1}, t_i]$ is given by

$$R_i = \log S(t_i) - \log S(t_{i-1}) = \log\left(\frac{S(t_i)}{S(t_{i-1})}\right), \quad i = 1, \dots n. \tag{1.4}$$

The (corrected) sample variance of the log-returns is then

$$\Delta_T(t) = \frac{1}{n-1}\sum_{i=1}^{n}(R_i - \bar{R})^2,$$

where

$$\bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i = \frac{1}{n}\log\left(\frac{S(t)}{S(t_0)}\right) \tag{1.5}$$

is the mean log-return in the interval $[t_0, t]$. Note that the sample variance of log returns is dimensionless and it is expressed in percentage (e.g., $\Delta_T(t) = 0.01$ means 1% variance) . The **T-historical variance** of the asset is obtained by "annualizing" the sample variance, i.e., by dividing $\Delta_T(t)$ by h measured in fraction of years, so that the result is expressed in yearly percentage

$$\widehat{\sigma}_T^2(t) = \frac{1}{h}\frac{1}{n-1}\sum_{i=1}^{n}(R_i - \bar{R})^2. \tag{1.6}$$

(in other words, the historical variance is the one-year projection of the sample variance). The square root of the $T-$historical variance is the **T-historical volatility** at time t, that is

$$\widehat{\sigma}_T(t) = \frac{1}{\sqrt{h}}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(R_i - \bar{R})^2}. \tag{1.7}$$

Note carefully that the historical volatility depends on the partition being used to compute it.

Suppose for example that $t-t_0 = T = 20$ days, which is quite common in the applications, and let $t_1, \dots t_{20}$ be the market closing times at these days. Let $h = 1$ day $= 1/365$ years. Then

$$\widehat{\sigma}_{20}(t) = \sqrt{365}\sqrt{\frac{1}{19}\sum_{i=1}^{n}(R_i - \bar{R})^2}$$

is called the 20-days historical volatility. We remark that $h = 1/252$ is also commonly used as normalization factor, since there are 252 trading days in one year.

As a way of example, Figure 1.1 shows the 20-days volatility of four stocks in the Stockholm exchange from January 1$^{\text{st}}$, 2014 until May 2$^{\text{nd}}$, 2014 (88 trading days). These data have been obtained with MATHEMATICA, by running the following command on May 3$^{\text{rd}}$, 2014:

`FinancialData["`*ticker*`", "Volatility20Day", {2014, 1, 1}]`

Upon running this command, the software connects to *Yahoo Finance* and collects the 20-days volatility data for the stock identified by the ticker symbol *"ticker"*, starting from the date $\{2014, 1, 1\}$ (year, month, day) until the present day. Note that in a few cases the historical volatility remains approximatively constant within periods of about 20 days.

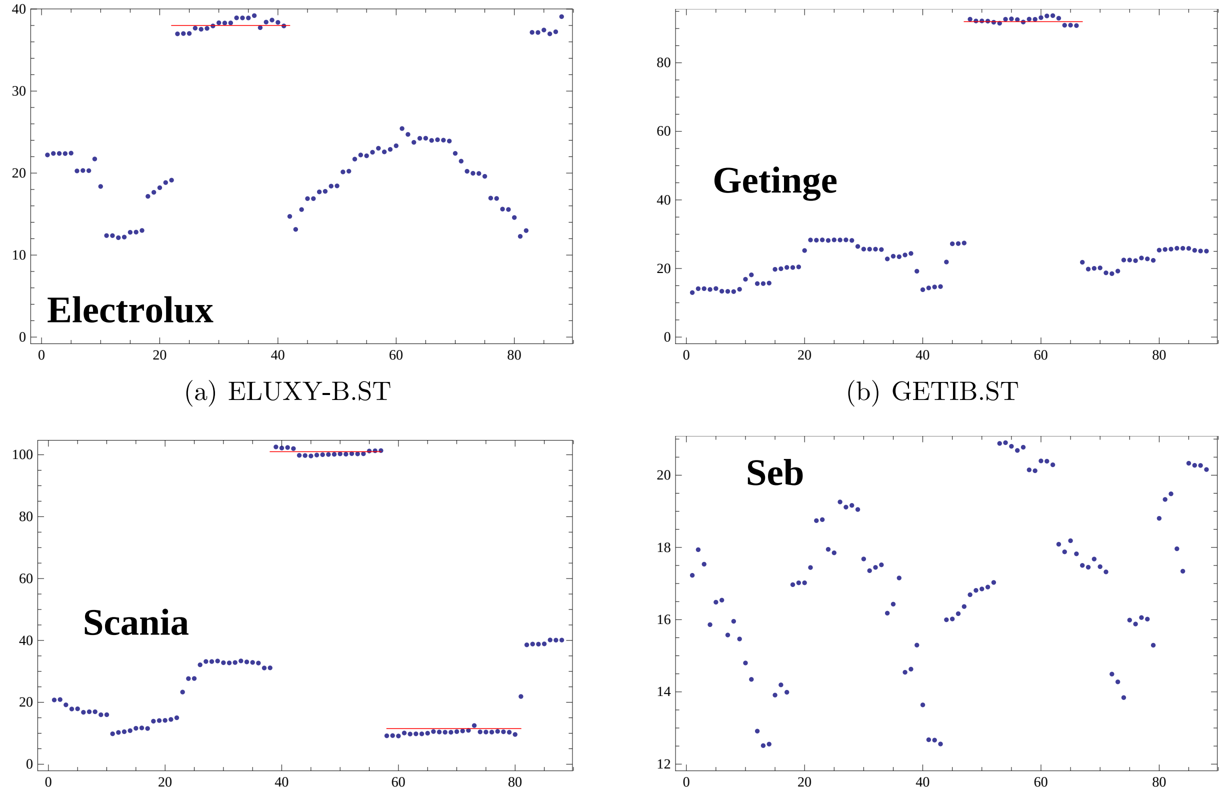
Financial derivatives. Options

A **financial derivative** (or derivative security) is an asset whose value depends on the performance of one (or more) other asset(s), which is called the **underlying asset**. There exist various types of financial derivatives, the most common being options, futures, forwards and swaps. In this section we discuss option derivatives on a single asset (typically a stock).

A **call option** is a contract between two parties: the buyer, or **owner**, of the call and the seller, or **writer**, of the call. The contract grants to the owner the right, but *not* the obligation, to buy the underlying asset for a given price, which is fixed at the time when the contract is stipulated and which is called **strike price** of the call. If the buyer can exercise this right only at some given time T in the future then the call option is called **European**, while if the option can be exercised at any time earlier than or equal to T, then the option is called **American**. The time T is called **maturity time**, or **expiration date** of the call. The writer of the call is obliged to sell the asset to the buyer if the latter decides to exercise the option. If the option to buy in the definition of a call is replaced by the option to sell, then the option is called a **put option**.

In exchange for the option, the buyer must pay a **premium** to the seller (options are not free). Suppose that the option is a European option with strike price K and maturity time T. Assume that the underlying is a stock with price $S(t)$ at time $t \leq T$ and let Π_0 be the premium paid by the buyer to the seller. In which case is it then convenient for the buyer to exercise the call at maturity? Let us define the **pay-off** of a European call as

$$Y = (S(T) - K)_+ := \max(0, S(T) - K) \quad \text{(call)},$$



(a) ELUXY-B.ST (b) GETIB.ST

(c) SCV-B.ST (d) SEB-C.ST

Figure 1.1: 20-days volatility of 4 stocks in the Stockholm Exchange Market on May 2^{nd}, 2014. The caption in each graph shows the ticker of the stock.

i.e., $Y > 0$ if the stock price at the expiration date is greater than the strike price of the call, while $Y = 0$ otherwise; similarly, for a European put we set

$$Y = (K - S(T))_+ \quad \text{(put)}.$$

Clearly, when $Y > 0$ it is more convenient for the buyer to exercise the option rather than buying/selling the stock on the market. Note however that the real **profit** for the buyer is given by $N(Y - \Pi_0)$, where N is the number of option contracts owned by the buyer. Typically, options are sold in multiples of 100 shares, that is to say, the minimum amount of options that one can buy is 100, which cover 100 shares of the underlying asset.

Let us introduce some further terminology. A European call (resp. put) is said to be **in the money** at time t if $S(t) > K$ (resp. $S(t) < K$). The call (resp. put) is said to be **out of the money** if $S(t) < K$ (resp. $S(t) > K$). If $S(t) = K$, the (call or put) option is said to be **at the money** at time t. The meaning of this terminology is self-explanatory, see Figure 1.2.

The pay-off of an American call exercised at time t is $Y(t) = (S(t) - K)_+$, while for an American put we have $Y(t) = (K - S(t))_+$. The quantity $Y(t)$ is also called **intrinsic value**

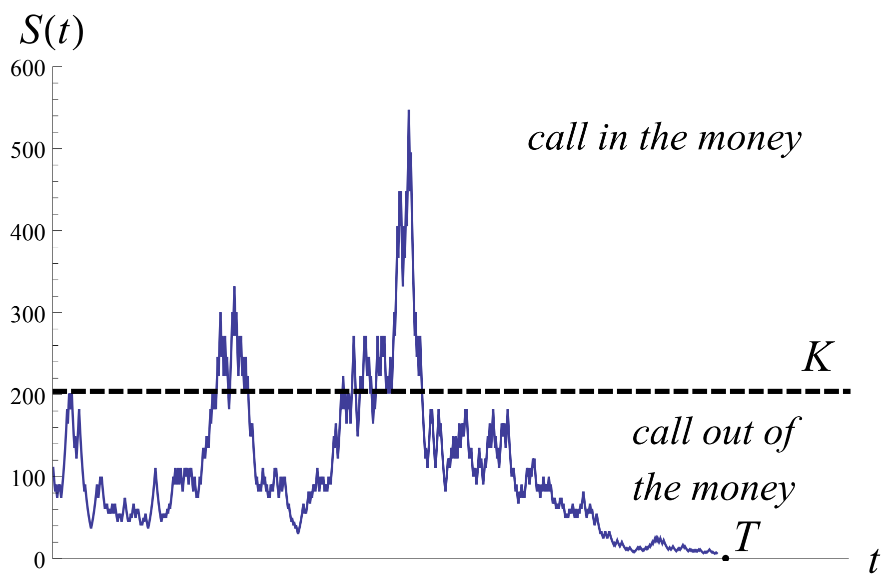


Figure 1.2: A call option with strike $K = 200$ is in the money in the upper region and out of the money in the lower region. A put option with the same strike is in the money in the lower region and out of the money in the upper region.

of the American option. In particular, the intrinsic value of an out-of-the-money American option is zero.

When dealing with options on a stock with price $S(t)$, we denote by $C(t, S(T), K, T)$, the price at time $t \in [0, T]$ of a European call option with strike K and maturity $T > 0$. The price of the European put option with the same parameters will be denoted by $P(t, S(t), K, T)$. Finally, we denote by $\widehat{C}(t, S(t), K, T)$ and $\widehat{P}(t, S(t), K, T)$ the values of the corresponding American call and put option.

Option markets

Option markets are relatively new compared to stock markets. The first one has been established in Chicago in 1974 (the Chicago Board Options Exchange, CBOE). In an option market anyone (after a proper authorization) can be the buyer or the seller of an option. Moreover options in the market are all American and thus can be exercised at any time prior or including the expiration date. Options are available on different assets (stocks, debts, indexes,...) and with different maturity times that can vary between one week and several months. Clearly, the deeper in the money is the option, the higher will be the price of the option in the market, while the price of an option deeply out of the money is usually quite low (but never zero!). It is also clear that the buyer of the option is the party holding the long position on the option, since the buyer owns the option and thus hopes for an increase of its value, while the writer is the holder of the short position.

One reason why investors buy call options is to protect a short position on the underlying asset. In fact, suppose that an investor is short-selling 100 shares of a stock at time $t = 0$ for the price $S(0)$ and let $t_0 > 0$ be the time at which the shares must be returned to the lender. At time $t = 0$ the investor buys 100 shares of an American call option on the stock with strike $K \approx S(0)$ and maturity later than t_0. If at time t_0 the price of the stock is no lower than $S(0)$, the investor will exercise the call, thereby obtaining 100 shares of the stock for the price $K \approx S(0)$, which the investor will then return to the lender with minimal losses*. At the same fashion, investors buy put options to protect a long position on the underlying asset.

Exercise 1.1 (?). *Can you think of a reason why investors sell options?*

Of course, speculation is also an important factor in option markets. However the standard theory of options pricing is firmly based on the interpretation of options as derivative securities and does not take speculation into account.

European, American and Asian derivatives

European call and put options are examples of more general contracts called **European derivatives**. Given a function $g : (0, \infty) \to \mathbb{R}$, a **standard** European derivative with pay-off $Y = g(S(T))$ and maturity time $T > 0$ is a contract that pays to its owner the amount Y at time $T > 0$. Here $S(T)$ is the price of the underlying stock at time T, while g is called the **pay-off function** of the derivative (e.g., $g(x) = (x - K)_+$ for European call options, while $g(x) = (K - x)_+$ for European put options). Hence, for a standard European derivative the pay-off depends only on the price of the stock at maturity and not on the earlier history of the price. An example of standard European derivative which is actually traded in the market (other than call and put options) is the **digital option**. Denoting by $H(x)$ the Heaviside function,

$$H(x) = \begin{cases} 1, & \text{for } x > 0 \\ 0, & \text{for } x \leq 0 \end{cases},$$

and letting $K, L > 0$ be constants expressed in units of currency (e.g., dollars), a standard European derivative with pay-off function $g(x) = LH(x - K)$ is called a **cash-settled digital option**; this derivative pays the amount L if $S(T) > K$, and nothing otherwise. A **physically-settled digital option** has the pay-off function $g(x) = xH(x - K)$, which means that, when the derivative expires in the money, i.e., when $S(T) > K$, the owner is paid-off with the stock. Digital options are also called **binary** options. Figure 1.3 shows the pay-off function of call, put and digital options with strike $K = 10$. Drawing the pay-off function of a derivative helps to get a first insight onto its properties.

Exercise 1.2. *Given $\Delta K > 0$, consider a standard European derivative with pay-off function*

$$g(x) = (x - K + \Delta K)_+ - 2(x - K)_+ + (x - K - \Delta K)_+.$$

*A short-selling strategy that is not covered by a suitable security is said to be naked.

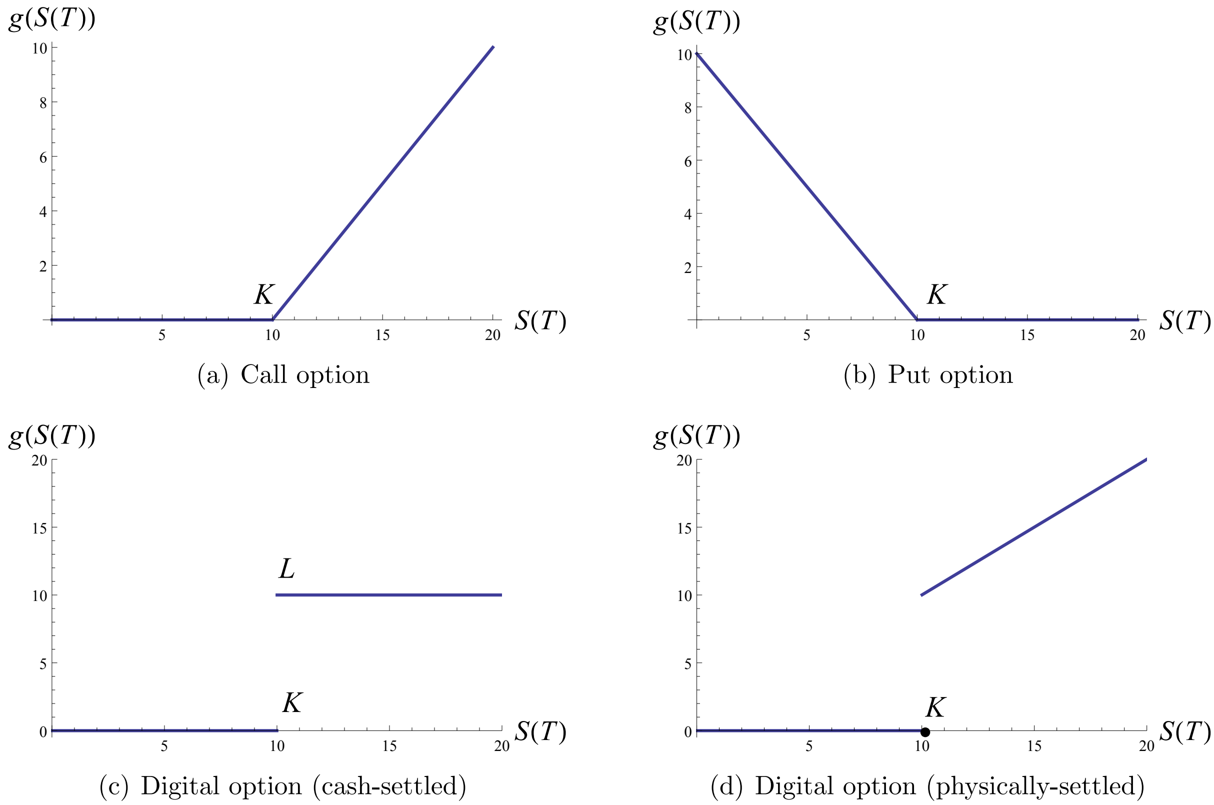


Figure 1.3: Pay-off function of some standard European derivatives.

Draw the graph of g and derive the range of $S(T)$ for which the derivative expires in the money.

If the pay-off depends on the history of the stock price during the interval $[0, T]$, and not just on $S(T)$, the European derivative is said to be **non-standard** (or **exotic**). An example of non-standard European derivative that is actually traded is the so-called **Asian call option**, the pay-off of which is given by $Y = (\frac{1}{T}\int_0^T S(t)\,dt - K)_+$.

The value at time t of a European derivative with pay-off Y and expiration date T will be denoted $\Pi_Y(t, T)$, or simply by $\Pi_Y(t)$ if there is no need to emphasize the expiration date.

The term "European" refers to the fact that the contract cannot be exercised before time T. For a **standard American derivative** the buyer can exercise the contract at any time $t \in (0, T]$ and so doing the buyer will receive the amount $Y(t) = g(S(t))$, where g is the pay-off function of the American derivative. Non-standard American derivatives can be defined similarly to the European ones.

Money market

A **money market** is a (OTC) market consisting of **risk-free** assets, i.e., assets whose value is always increasing in time. Like options, assets in the money market have finite maturity, which varies between one day and one year. The price of a generic risk-free asset in the money market at time t will be denoted by $B(t)$; the fact that the asset is risk-free means that $B(t_2) > B(t_1)$, for all $t_2 > t_1$, the difference $B(t_2) - B(t_1)$ being determined by the **interest rate** of the asset in the interval $[t_1, t_2]$. In particular, we say that a risk-free asset has **instantaneous interest rate** $r > 0$ in the interval $[t_1, t_2]$ if

$$B(t) = B(t_1) \exp(r(t - t_1)), \quad \text{for } t_1 \leq t \leq t_2.$$

The interest rate is always expressed in yearly percentage and typically varies on time.

As a way of example, suppose that an investor buys a risk-free asset at time $t = 0$ which expires at time $T > 0$. The seller will receive the quantity B_0. As part of the agreement, the seller promises to re-purchase the risk-free asset at time T for $B(T) > B_0$. Hence buying an asset in the money market is equivalent to lend money to the seller, while selling an asset in the money market is equivalent to borrow money from the buyer. The seller of the risk-free asset is said to hold the short position and the buyer the long position on the asset, although strictly speaking the long/short position refers to the interest rate (since the value of risk-free assets cannot decrease). Examples of risk-free assets in the money market are saving accounts, commercial papers and treasure bills.

Note that so far we have introduced three strategies that investors can undertake to obtain cash: short-selling an asset, writing an option or selling a risk-free asset.

Bonds

Risk-free assets with maturity longer than 1 year are called **bonds**. The simplest one is the **zero-coupon bond**. A zero-coupon bond with face value K and maturity T is a contract that promises to pay to its owner the amount K at time T. Hence a zero-coupon bond is a European style derivative with pay-off K and maturity T; the underlying asset is the interest rate of the bond. We denote by $B_K(t, T)$ the value of this contract at time $t < T$. Obviously $B_K(t, T) < K$, for $t < T$.

Zero-coupon bonds are often used to measure the performance of the underlying interest rate. This performance is expressed in terms of the **yield** $Y_t(T)$ of the interest rate, which is defined as the relative return of one share of a zero coupon bond with face value 1 purchased at time t and expiring at time $t + T$, i.e.,

$$Y_t(T) = \frac{1 - B_1(t, t + T)}{B_1(t, t + T)}.$$

The curve $T \to Y_t(T)$ is called **yield curve**. As a way of example, Figure 1.4 shows the yield curve of UK government bonds at $t =$January 19^{th}, 2016.

Exercise 1.3 (?). *Think about the fact that the yield curve showed in Figure 1.4 is not increasing. What does it mean?*

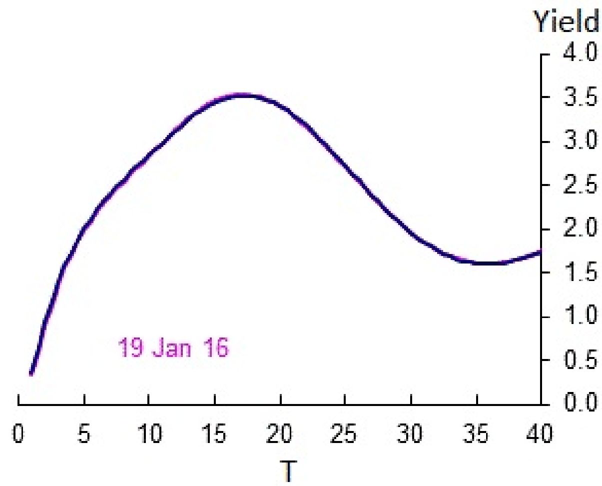


Figure 1.4: Yield curve $T \to Y_t(T)$ of UK government bonds for $t =$January 19$^{\text{th}}$, 2016. This image was extracted from the home page of Bank of England.

Exercise 1.4. *Compute the yield of a constant interest rate $r > 0$ and depict the yield curve with Matlab.*

Forwards and swaps

A **forward contract** with **delivery price** K and maturity (or delivery) time T on an asset $\mathcal{U}$ is a type of financial derivative stipulated by two parties in which one promises to the other to sell (and possibly deliver) the asset $\mathcal{U}$ at time T in exchange for the cash K. As opposed to option contracts, both parties in a forward contract are *obliged* to fulfill their part of the agreement. In particular, as they both have the same right/obligation, neither of the two parties has to pay a premium to the other when the contract is stipulated, that is to say, *forward contracts are free*†. The party who must sell the asset at maturity holds the short position, while the party who must buy the asset is the holder of the long position. Forward contracts are traded OTC (over the counter) and most commonly on commodities or market indexes, such as currency exchange rates, interest rates and volatilities. In the case that the underlying asset is an index, forward contracts are also called **swaps** (e.g., currency swaps, interest rate swaps, volatility swaps, etc.). Let us give two examples.

Example of forward contract on a commodity. Consider a farmer who grows wheat and a miller who needs wheat to produce flour. Clearly, the farmer interest is to sell the wheat for the highest possible price, while the miller interest is to pay the least possible for the wheat. The price of the wheat depends on many economical and non-economical factors (such as whether conditions, which affect the quality and quantity of harvests) and it is therefore

†In fact, the terminology used for forward contracts is "to enter a forward contract" and not "to buy/sell a forward contract".

quite volatile. As a form to reduce risks, the farmer and the miller stipulate a forward contract on the wheat in the winter (before the plantation, which occurs in the spring) with expiration date in the end of the summer (when harvest occurs) in order to lock the future price of the wheat at a value which is convenient for both of them.

Example of a currency swap. Suppose that a car company in Sweden promises to deliver a stock of 100 cars to another company in the United States in exactly one month. Suppose that the price of each car is fixed in Swedish crowns, say 100.000 crowns. Clearly the American company will benefit by an increase of the exchange rate crown/dollars and will be damaged in the opposite case. To avoid possible high losses, the American company stipulates a forward contract (currency swap) on 100×100.000=ten millions Swedish crowns expiring in one month and which gives the company the right *and* the obligation to buy ten millions crowns for a price in dollars agreed upon today. The other party of the forward contract could be a company exposed to the opposite risk, i.e., to an increase of the exchange rate crown/dollars.

As it is clear from these examples, the main purpose of forward contracts is to share risks. Irrespective of the movement of the underlying asset in the market, its price at time T for the holders of the forward contract will be K. We may define the "pay-off" for a long position in a forward contract as

$$Y = (\Pi^{\mathcal{U}}(T) - K),$$

while for the holder of the short position the "pay off" is $(K - \Pi^{\mathcal{U}}(T))$. Note however that the term "pay-off" in this case is a little inappropriate, because Y can be negative. In fact, it is clear that one of the two parties in a forward contract is always going to incur in a loss. If the loss is very large, this party could become insolvent, i.e., unable to fulfill the contract, and then both parties will end up losing. As explained below, the risk of insolvency is greatly reduced by trading futures contracts instead of forward contracts.

The delivery price agreed by the two parties in a forward contract represents a pondered estimation of what the price of the underlying asset will be in the future. In this respect, K is also called the **forward price** of the asset. More precisely, the T-forward price $\mathrm{For}_{\mathcal{U}}(t,T)$ of an asset $\mathcal{U}$ at time $t < T$ is the strike price of a forward contract on $\mathcal{U}$ stipulated at time t and with maturity T, while the current, actual price $\Pi^{\mathcal{U}}(t)$ of the asset is sometimes called the **spot** price[‡]. Clearly the forward price is unlikely to be a good estimation of the price of the asset at time T, since the consensus on this value is limited to the participants of the forward contract and different parties may agree to different delivery prices. The delivery price of futures contracts on the asset, which we define below, gives a better and more commonly accepted estimation for the future value of an asset.

Futures

Futures contracts are standardized forward contracts, i.e., rather than being OTC, they are negotiated in regularized markets. Perhaps the most interesting role of futures contracts

[‡]Note however that many assets, such as commodities, do not have an official spot price, but only a future price, as defined below.

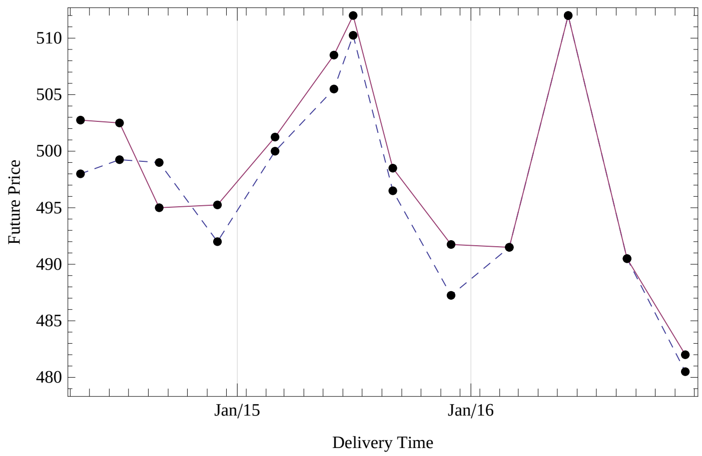


Figure 1.5: Futures price of corn on May 12, 2014 (dashed line) and on May 13, 2014 (continuous line) for different delivery times

is that they make trading on commodities possible for anyone. To this regard we remark that commodities, e.g. crude oil, wheat, etc, are most often sold through long term contracts, such as forward and futures contracts, and therefore they do not usually have an "official spot price", but only a future delivery price (commodities "spot markets" exist, but their role is marginal for the discussion in this section).

Futures markets are markets in which the object of trading are futures contracts. Unlike forward contracts, all futures contracts in a futures market are subject to the same regulation, and so in particular all contracts on the same asset $\mathcal{U}$ with the same time of maturity T have the same delivery price, which is called T-**future price** of the asset and denoted by $\mathrm{Fut}_{\mathcal{U}}(t,T)$. Thus $\mathrm{Fut}_{\mathcal{U}}(t,T)$ is the delivery price in a futures contract on the asset $\mathcal{U}$ with time of maturity T which is stipulated at time $t < T$. Futures markets have been existing for more than 300 years and nowadays the most important ones are the Chicago Mercantile Exchange (CME), the New York Mercantile Exchange (NYMEX), the Chicago Board of Trade (CBOT) and the International Exchange Group (ICE).

In a futures market, anyone (after a proper authorization) can stipulate a futures contract. More precisely, holding a long position in a futures contract in the futures market consists in the agreement to receive as a cash flow the change in the future price of the underlying asset during the time in which the position is held, while a short position on the same contract receives the opposite cash flow. Notice that the cash flow may be positive or negative. In a long position the cash flow is positive when the future price goes up and it is negative when the future price goes down. Moreover, in order to eliminate the risk of insolvency, the cash flow is distributed in time through the mechanism of the **margin account**. For example, assume that at $t = 0$ we open a long position in a futures contract expiring at time T. At the same time, we need to open a margin account which contains a certain amount of cash (usually, 10 % of the current value of the T-future price for each contract opened).

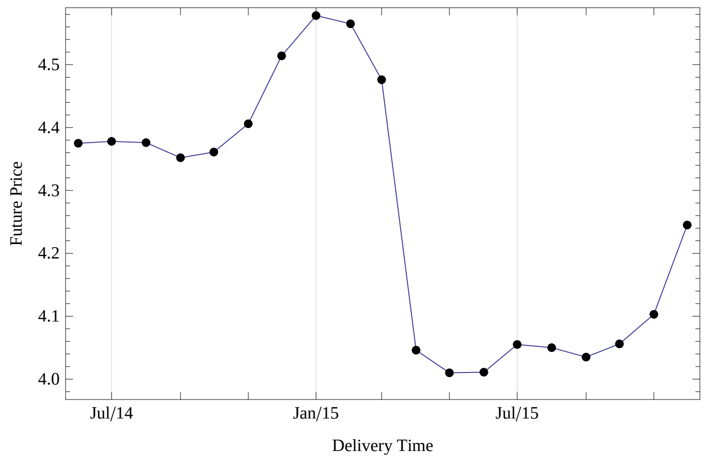


Figure 1.6: Futures price of natural gas on May 13, 2014 for different delivery times

At $t = 1$ day, the amount $\mathrm{Fut}_{\mathcal{U}}(1, T) - \mathrm{Fut}(0, T)$ will be added to the account, if it positive, or withdrawn, if it is negative. The position can be closed at any time $t < T$ (multiple of days), in which case the total amount of cash flown in the margin account is

$$(\mathrm{Fut}_{\mathcal{U}}(t, T) - \mathrm{Fut}_{\mathcal{U}}(t-1, T)) + (\mathrm{Fut}_{\mathcal{U}}(t-1, T) - \mathrm{Fut}_{\mathcal{U}}(t-2, T)) + \\ \cdots + (\mathrm{Fut}_{\mathcal{U}}(1, T) - \mathrm{Fut}_{\mathcal{U}}(0, T)) = (\mathrm{Fut}_{\mathcal{U}}(t, T) - \mathrm{Fut}_{\mathcal{U}}(0, T)).$$

(In fact, if the margin account becomes too low, and the investor does not add new cash to it, the position will be automatically closed by the exchange market).

If a long position is held up to the time of maturity, then the holder of the long position should buy the underlying asset. However futures contracts are often **cash settled** and not **physically settled**, which means that the delivery of the underlying asset does not occur, and the equivalent value in cash is paid instead.

Frictionless markets

As all mathematical models, also those in mathematical finance are based on a number of assumptions. Some of these assumptions are introduced only with the purpose of simplifying the analysis of the models and often correspond to facts that do not occur in reality. Among these "simplifying" assumptions it is common to impose that

1. There is no bid/ask spread
2. There are no transaction costs and trades occur instantaneously
3. An investor can trade any fraction of shares
4. No lack of liquidity: there is no limit to the amount of cash that can be borrowed from the money market

We have seen in the previous sections that real markets do not satisfy exactly these assumptions, although in some case they do it with excellent approximation. For instance if the investor is an agent working for a very large financial institution, then the above assumptions do reflect reality pretty well. However they work very badly for private investors. We summarize the validity of these assumptions by saying that the market has **no friction**. The idea is that when the above assumptions hold, trading proceeds "smoothly without resistance".

In a frictionless market we may define the portfolio process of an agent who is investing on N assets during the time interval $[0,T]$ as a function

$$\mathcal{A} : [0,T] \to \mathbb{R}^N, \quad \mathcal{A}(t) = (a_1(t), \dots, a_N(t)),$$

i.e., by assumptions 2 and 3, the number of shares $a_i(t)$ of each single asset at time t is now allowed to be any real number and to change at any arbitrary time in the interval $[0,T]$. Portfolio processes can be added using the linear structure in $\mathbb{R}^N$, namely if $\mathcal{B} = (b_1(t), \dots, b_N(t))$, then $\mathcal{A} + \mathcal{B}$ is the portfolio

$$\mathcal{A} + \mathcal{B} = (a_1(t) + b_1(t), \dots, a_N(t) + b_N(t)).$$

The value at time t of the portfolio process $\mathcal{A}$ is

$$V_{\mathcal{A}}(t) = \sum_{i=1}^{N} a(t)\Pi^{\mathcal{U}_i}(t),$$

and clearly

$$V_{\mathcal{A}}(t) + V_{\mathcal{B}}(t) = V_{\mathcal{A}+\mathcal{B}}(t).$$

Moreover it is clear that, thanks to assumption 3, perfect self-financial portfolio processes in frictionless markets always exist.

1.2 Qualitative properties of option prices

The purpose of this section is to derive some qualitative properties of option prices using only basic principles, without invoking any specific mathematical model for the market dynamics. Probably the most self-evident of all principles is the following, which we call the **no-dummy investor principle**[§]:

> *Investors prefer more to less and do not undertake trading strategies which result in a sure loss.*

This principle has a number of straightforward consequences, for instance:

(i) The price of stocks is always positive, i.e., $S(t) > 0$ for all t;

[§]More commonly (and repectfully) known as **rational investor principle**.

(ii) The price of options is always non-negative, e.g.,

$$C(t, S(t), K, T) \geq 0, \quad P(t, S(t), K, T) \geq 0,$$

for $t \leq T$ and similarly for American options;

(iii) The price of an option tends to the pay-off as maturity is approached, e.g.,

$$C(t, S(t), K, T) \to (S(T) - K)_+, \quad P(t, S(t), K, T) \to (K - S(T))_+,$$

as $t \to T^-$ and similarly for American options;

(iv) An American call or put option is at least as valuable as its European counterpart, i.e.,

$$\widehat{C}(t, S(t), K, T) \geq C(t, S(t), K, T), \quad \widehat{P}(t, S(t), K, T) \geq P(t, S(t), K, T)$$

(v) The price of an American call or put option is always larger or equal to its intrinsic value, i.e.,

$$\widehat{C}(t, S(t), K, T) \geq (S(t) - K)_+, \quad \widehat{P}(t, S(t), K, T) \geq (K - S(t))_+.$$

Exercise 1.5 (?). *Use the no-dummy investor principle to justify the properties (i)-(v).*

Any reasonable mathematical model for the price of stocks and options must be consistent with the properties (i)-(v) above. In the rest of this section they are assumed to hold without any further comment.

To derive more accurate properties of option prices we need to introduce the fundamental concept of arbitrage. Roughly speaking, an **arbitrage opportunity** is an investment strategy which entails no risk and a positive probability to give profit. For example, suppose that an investor is the seller of a call option $\mathcal{U}_1$ on a stock and the buyer of another call option $\mathcal{U}_2$ on the same stock with exactly the same parameters (time of maturity and strike). The value of this portfolio at the expiration date is zero. Now, if the investor were able to sell $\mathcal{U}_1$ for a price $\Pi^{\mathcal{U}_1} > \Pi^{\mathcal{U}_2}$, then this portfolio is clearly an arbitrage investment: By investing the difference $\Pi^{\mathcal{U}_1} - \Pi^{\mathcal{U}_2}$ in the money market, the agent has found a riskless investment which requires no initial wealth and which ensures a positive profit. However, why should the investor be able to find someone willing to pay $\Pi^{\mathcal{U}_1}$ for the call when the same call is available in the market for the lower price $\Pi^{\mathcal{U}_2}$? Purchasing the option at the price $\Pi^{\mathcal{U}_1}$ would be of course a dummy investment for the buyer. Nevertheless, due to the complexity of modern markets, arbitrage opportunities may actually exist, but only for a very short time, as they are quickly exploited and "traded away" by investors.

The previous discussion leads us to assume the validity of the **arbitrage-free principle:**

Asset prices in a market are such that no arbitrage opportunities can be found.

Asset prices that are consistent with this principle are said to be arbitrage free or **fair**.

The arbitrage-free principle can be used to derive a number of qualitative properties of option prices, which, unlike (i)-(v) above, do not follow trivially from the no-dummy investor principle. To this purpose it is convenient to express the arbitrage-free principle in a more quantitative form. There are several ways to do this, e.g. by requiring the absence of arbitrage portfolios in the market (see next chapter), or by imposing the so-called **dominance principle**:

Dominance Principle: *Suppose that $t < T$ is the present time and consider a portfolio which does not contain dividend-paying assets. If the holder of the portfolio can ensure that the portfolio exists up to time T and that its value is non-negative at time T, i.e., $V(T) \geq 0$, then $V(t) \geq 0$.*

The fact that the dominance principle must hold as a consequence of the arbitrage-free principle is clear. In fact, if $V(t) < 0$, then the investor needs no initial wealth to open the portfolio, while on the other hand the portfolio return is positive, since $V(T) - V(t) \geq -V(t) > 0$. Hence the given portfolio ensures a positive profit without taking any risk, which violates the arbitrage-free principle.

Let us comment further on the formulation of the dominance principle. First of all, the requirement that the investor must ensure the existence of the portfolio up to time T is not trivial and implies in particular that the portfolio cannot contain short positions on American derivatives (as the buyer may exercise the derivative prior to T). The reason to require that the assets pay no dividend is the following. Suppose that a stock pays a dividend of 2 % at time T. Just before that the investor open a short position on the stock and invest 99% of the income on a risk-free asset. Hence the value of this portfolio is negative, but it becomes instantaneously positive when the dividend is paid[¶].

The next theorem collects a number of properties that must be satisfied by option prices as a consequence of the dominance principle. We denote these properties by (vi)-(ix) in order to continue the list (i)-(v) given above.

Theorem 1.1. *Let $C(t, S(t), K, T)$ denote the price of a European call with strike K and maturity T on a stock with price $S(t)$ and let $P(t, S(t), K, T)$ be the price of the corresponding European put. Assume that there exists a risk-free asset in the money market with constant interest rate r. If the dominance principle holds, then, for all $t < T$,*

(vi) *The* **put-call parity** *holds*

$$S(t) - C(t, S(t), K, T) = Ke^{-r(T-t)} - P(t, S(t), K, T). \tag{1.8}$$

(vii) *If $r \geq 0$, then $C(t, S(t), K, T) \geq (S(t) - K)_+$; the strict inequality $C(t, S(t), K, T) > (S(t) - K)_+$ holds when $r > 0$.*

[¶]In practice this is not a feasible strategy as the profit is extremely small and highly surpassed by transaction costs.

(viii) If $r \geq 0$, the map $T \to C(t, S(t), K, T)$ is non-decreasing.

(ix) The maps $K \to C(t, S(t), K, T)$ and $K \to P(t, S(t), K, T)$ are convex[‖].

Proof. We reproduce here the arguments in [3, Ch.1].

(vi) Consider a constant portfolio $\mathcal{A}$ which is long one share of the stock and one share of the put option, and is short one share of the call and $K/B(T)$ shares of the risk-free asset. The value of this portfolio at maturity is

$$V_{\mathcal{A}}(T) = S(T) + (K - S(T))_+ - (S(T) - K)_+ - \frac{K}{B(T)} B(T) = 0.$$

Hence, by the dominance principle, $V_{\mathcal{A}}(t) \geq 0$, for $t < T$, that is

$$S(t) + P(t, S(t), K, T) - C(t, S(t), K, T) - K e^{-r(T-t)} \geq 0.$$

Now consider the portfolio $-\mathcal{A}$ with the opposite position on each asset. Again we have $V_{-\mathcal{A}}(T) = 0$ and thus $V_{-\mathcal{A}}(t) = -\mathcal{V}_{\mathcal{A}}(t) \geq 0$, for $t < T$. Hence

$$S(t) + P(t, S(t), K, T) - C(t, S(t), K, T) - K e^{-r(T-t)} \leq 0.$$

Thus the left hand side in the previous two inequalities must be zero, which gives the put-call parity.

(vii) We can assume $S(t) \geq K$, otherwise the claim is obvious (the price of a call cannot be negative). By the put-call parity, using that $P(t, S(t), K, T) \geq 0$,

$$C(t, S(t), K, T) = S(t) - K e^{-r(T-t)} + P(t, S(t), K, T) \geq S(t) - K e^{-r(T-t)};$$

the right hand side equals $S(t) - K$ for $r = 0$ and is strictly greater than this quantity for $r > 0$. As $S(t) - K = (S(t) - K)_+$ for $S(t) \geq K$, the claim follows.

(viii) Consider a portfolio $\mathcal{A}$ which is long one call with maturity T_2 and strike K, and short one call with maturity T_1 and strike K, where $T_2 > T_1 \geq t$. By the claim 2 we have

$$C(t_1, S(T_1), K, T_2) \geq (S(T_1) - K)_+ = C(t_1, S(T_1), K, T_1),$$

i.e., $V_{\mathcal{A}}(T_1) \geq 0$, for $t < T_1$. Hence $V_{\mathcal{A}}(t) \geq 0$, i.e., $C(t, S(t), K, T_2) \geq C(t, S(t), K, T_1)$, which is the claim.

(ix) We prove the statement for call options, the argument for put options being the same. Let $K_0, K_1 > 0$ and $0 < \theta < 1$ be given. Consider a portfolio $\mathcal{A}$ which is long one share of a call with strike $\theta K_1 + (1-\theta) K_0$ and maturity T, short θ shares of a call with strike

[‖] Recall that a real-valued function f on an interval I is convex if $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$, for all $x, y \in I$ and $\theta \in (0, 1)$.

K_1 and maturity T, short $(1-\theta)$ shares of a call with strike K_0 and maturity T. The value of this portfolio at maturity is

$$V_{\mathcal{A}}(T) = (S(T) - (\theta K_1 + (1-\theta)K_0))_+ - \theta(S(T) - K_1)_+ - (1-\theta)(S(T) - K_0)_+.$$

The convexity of the function $f(x) = (S(T) - x)_+$ gives $\mathcal{V}_{\mathcal{A}}(T) \geq 0$ and so $V_{\mathcal{A}}(t) \geq 0$ by the dominance principle. The latter inequality is

$$C(t, S(t), \theta K_1 + (1-\theta)K_0, T) \leq \theta C(t, S(t), K_1, T) + (1-\theta)C(t, S(t), K_0, T),$$

which is the claim for call options.

$\square$

Remark 1.1. In the applications of the put-call parity, one typically sets $e^{-r(T-t)} \approx 1$, which gives the simple relation $S(t) \approx C(t, S(t), K, T) - P(t, S(t), K, T)$ between the prices of a European call, the corresponding European put and the underlying. This relation is quite well-represented in the market, thereby giving indirect support to the arbitrage-free principle.

Exercise 1.6. *Consider the following table of European options prices at time $t = 0$:*

	CALL	
Maturity	**Strike**	**Price**
1 month	*104 Kr*	*20 Kr*
1 month	*110 Kr*	*16 Kr*
1 month	*116 Kr*	*10 Kr*
3 month	*110 Kr*	*15 Kr*

	PUT	
Maturity	**Strike**	**Price**
1 month	*96 Kr*	*14 Kr*
1 month	*100 Kr*	*16 Kr*
1 month	*104 Kr*	*18 Kr*
3 month	*110 Kr*	*18 Kr*

Assume that there exists risk-free asset with interest rate $r = 0$ and that the price of the underlying asset at time $t = 0$ is $S(0) = 100$ Kr. Explain why these prices are incompatible with the dominance principle. Find a constant portfolio position which violates the dominance principle.

Exercise 1.7. *Assume that the dominance principle holds. Consider a European derivative $\mathcal{U}$ with maturity time T and pay-off Y given by $Y = \min[(S(T)-K_1)_+, (K_2-S(T))_+]$, where $K_2 > K_1$ and $(x)_+ = max(0, x)$. Find a constant portfolio consisting of European calls and puts expiring at time T whose value at any time $t < T$ equals the value of $\mathcal{U}$ (i.e., which replicates the value of $\mathcal{U}$).*

Exercise 1.8. *Assume that there exists a risk-free asset with constant interest rate r and that the dominance principle holds. Show that the forward price of an asset $\mathcal{U}$ with spot price $\Pi^{\mathcal{U}}(t)$ is given by*

$$\mathrm{For}_{\mathcal{U}}(t, T) = e^{r(T-t)}\Pi^{\mathcal{U}}(t).$$

Interpret the result.

Consider now a no-dummy investor owning an American put option. When should the investor exercise the option? At any time $t < T$ we have, by (v),

$$\text{either}\quad \widehat{P}(t,S(t),K,T) > (K-S(t))_+ \quad \text{or} \quad \widehat{P}(t,S(t),K,T) = (K-S(t))_+.$$

Exercising the American put at a time t when the strict inequality $\widehat{P}(t,S(t),K,T) > (K-S(t))_+$ holds is a dummy decision, because the resulting pay-off is lower than the value of the derivative. The optimal strategy for the investor in this case would be to sell** an equivalent put option for its current price $\widehat{P}(t,S(t),K,T)$. This investment is completely risk-less as the investor has a long and a short position on the same derivative. On the other hand, if the equality $\widehat{P}(t,S(t),K,T) = (K-S(t))_+$ holds at time t, then the optimal strategy for the investor is to exercise the American put, as in this case the pay-off equals the value of the derivative, i.e., the investor takes full advantage of the American put. This leads us to introduce the following definition.

Definition 1.1. *At time $t < T$ is called an* **optimal exercise time** *for the American put with value $\widehat{P}(t,S(t),K,T)$ if*

$$\widehat{P}(t,S(t),K,T) = (K-S(t))_+.$$

A similar definition can be justified for American call options, i.e., the optimal exercise time of an American call is a time t at which $\widehat{C}(t,S(t),K,T) = (S(t)-k)_+$. However, assuming that the dominance principle holds and that there exists a risk-free asset with interest rate $r > 0$, we have $\widehat{C}(t,S(t),K,T) \geq C(t,S(t),K,T) > (S(t)-K)_+$, for $t < T$, see (vii). It follows that, in an arbitrage-free market, and provided the dominance principle holds, *it is never optimal to exercise an American call prior to maturity.* Note that the validity of the dominance principle requires in particular that the underlying pays no dividend. As opposed to this, it can be shown that if the underlying stock pays a dividend prior to the expiration of the American call, then it is optimal to exercise the American call just before the dividend is paid, provided the price of the stock is sufficiently large, see [3, Ch. 7] .

As in the absence of dividends the optimal strategy is to hold the American call until maturity, the dominance principle leads us to the following, last property on the fair price of options:

(x) When the underlying stock pays no dividend, the fair price of a European call and the fair price of an American call with same parameters are equal, i.e.,

$$\widehat{C}(t,S(t),K,T) = C(t,S(t),K,T).$$

Remark 1.2. The properties (i)-(x) established in this section depend only on the validity of the no-dummy investor principle and the arbitrage-free principle (in the form of the dominance principle) and *not* on the specific dynamics for the price of the stock and the options. In the following chapters we shall introduce models for the fair price of options in the market. These models also predict the validity of (i)-(x), whence they are consistent with the assumed principles.

**In a frictionless market a buyer can always be found.

Chapter 2

Binomial markets

In this and the following two chapters we present a time-discrete model for the fair price of options first proposed in [4] and which is known under the name of **binomial options pricing model**. The model is very popular among practitioners due to its implementation simplicity. The present chapter deals with the dynamics of the underlying asset, which we assume to be a stock. The following two chapters are concerned with European and American derivatives on the stock. The time-continuum analogue of the binomial model is the Black-Scholes model, which will be studied in Chapter 6.

2.1 The binomial stock price

The **binomial asset price** is a model for the evolution in time of the price of financial assets. It is often applied to stocks, hence we denote by $S(t)$ the price of the asset at time t. We are interested in monitoring the stock price in some finite time interval $[0, T]$, where $T > 0$ could be for instance the expiration date of an option on the stock. The price of the stock at time $t = 0$ is denoted by S_0 and is assumed to be known.

The binomial stock price can only change at some given pre-defined times $0 = t_0 < t_1 < t_2 < \cdots < t_N = T$; moreover the price at time t_{i+1} depends only on the price at time t_i and the result of "tossing a coin". Precisely, letting $u, d \in \mathbb{R}$, $u > d$, and $p \in (0, 1)$, we assume

$$S(t_{i+1}) = \begin{cases} S(t_i)e^u, & \text{with probability } p, \\ S(t_i)e^d, & \text{with probability } 1-p, \end{cases}$$

for all $i = 0, \ldots, N-1$. Here we may interpret p as the probability to get a head in a coin toss ($p = 1/2$ for a **fair coin**). We restrict to the standard binomial model, which assumes that the parameters u, d, p are time-independent and that the stock pays no dividend in the interval $[0, T]$. In the applications one typically chooses $u > 0$ and $d < 0$ (e.g., $d = -u$ is quite common), hence u stands for "up", since $S(t_{i+1}) = S(t_i)e^u > S(t_i)$, while d stands for "down", for $S(t_{i+1}) = S(t_i)e^d < S(t_i)$. In the first case we say that the stock price goes up at time t_{i+1}, in the second case that it goes down at time t_{i+1}.

Next we introduce a number of assumptions which simplify the analysis of the model without compromising its generality*. Firstly we assume that the times $t_0, t_1, \dots t_N$ are equidistant, that is

$$t_{i+1} - t_i = h > 0.$$

In the applications the value of h must be chosen much smaller than T. Without loss of generality we can pick $h = 1$, and so

$$t_1 = 1, \quad t_2 = 2, \quad \dots t_N = T = N,$$

with $N >> 1$. For instance, if $N = 67$ (the number of trading days in a period of 3 months), then $h = 1$ day and $S(t)$, for $t \in \{1, \dots N\}$, may refer to the closing price of the stock at each day. It is convenient to denote

$$\mathcal{I} = \{1, \dots, N\}.$$

Hence, from now on, we assume that the binomial stock price is determined by the rule $S(0) = S_0$ and

$$S(t) = \begin{cases} S(t-1)e^u, & \text{with probability } p \\ S(t-1)e^d, & \text{with probability } 1-p \end{cases}, \quad t \in \mathcal{I}. \tag{2.1}$$

Remark 2.1 (Notation)**.** The notation used in the present notes is the same as in [3], although it is slightly different from the one used in the standard literature on the binomial model, see e.g., [5]. In fact the binomial stock price is more commonly written as

$$S(t) = \begin{cases} S(t-1)u, & \text{with probability } p \\ S(t-1)d, & \text{with probability } 1-p \end{cases},$$

with $0 < d < u$. All the results in the present text can be translated into the standard notation by the substitutions $e^u \to u$, $e^d \to d$. In our notation the log-returns of the stock take a somehow simpler form, which is useful when passing to the time-continuum limit (see Section 5.6).

Each possible sequence $(S(1), \dots, S(N))$ of the future stock prices determined by the binomial model is called a **path** of the stock price. Clearly, there exists 2^N possible paths of the stock price in a N-period model. Letting

$$\{u, d\}^N = \{x = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N : x_t = u \text{ or } x_t = d, \ t \in \mathcal{I}\}$$

be the space of all possible N-sequences of "ups" and "downs", we obtain a unique path of the stock price $(S(1), \dots, S(N))$ for each $x \in \{u, d\}^N$. For instance, for $N = 3$ and $x = (u, u, u)$ the corresponding stock price path is given by

$$S_0 \to S(1) = S_0 e^u \to S(2) = S(1)e^u = S_0 e^{2u} \to S(3) = S(2)e^u = S_0 e^{3u},$$

*We come back to the general model in Section 2.4, where it is implemented with Matlab.

while for $x = (u, d, u)$ we obtain the path

$$S_0 \to S(1) = S_0 e^u \to S(2) = S(1)e^d = S_0 e^{u+d} \to S(3) = S(2)e^u = S_0 e^{2u+d}.$$

In general the possible paths of the stock price for the 3-period model can be represented as

$$\begin{tikzcd}
 & & & S(3) = S_0 e^{3u} \\
 & & S(2) = S_0 e^{2u} \arrow[ur, "u"] \arrow[dr, "d"] & \\
 & S(1) = S_0 e^{u} \arrow[ur, "u"] \arrow[dr, "d"] & & S(3) = S_0 e^{2u+d} \\
S(0) = S_0 \arrow[ur, "u"] \arrow[dr, "d"] & & S(2) = S_0 e^{u+d} \arrow[ur, "u"] \arrow[dr, "d"] & \\
 & S(1) = S_0 e^{d} \arrow[ur, "u"] \arrow[dr, "d"] & & S(3) = S_0 e^{u+2d} \\
 & & S(2) = S_0 e^{2d} \arrow[ur, "u"] \arrow[dr, "d"] & \\
 & & & S(3) = S_0 e^{3d}
\end{tikzcd}$$

which is an example of **binomial tree**. Note that the admissible values for the binomial price $S(t)$ at time t are given by

$$S(t) \in \{S_0 e^{ku+(t-k)d},\ k = 0, \dots, t\},$$

for all $t \in \mathcal{I}$, where k is the number of times that the price goes up up to the time t.

Definition 2.1. *Given $x = (x_1, \dots x_N) \in \{u, d\}^N$, the binomial stock price $S(t, x)$ at time $t \in \mathcal{I}$ corresponding to x is given by*

$$S(t, x) = S_0 \exp(x_1 + x_2 + \dots x_t).$$

The vector $S^x = (S(1, x), \dots S(N, x))$ is called the x-path of the binomial stock price. Moreover we define the probability[†] of the path S^x as

$$\mathbb{P}(S^x) = p^{N_u(x)}(1-p)^{N_d(x)},$$

where $N_u(x)$ is the number of u's in the sequence x and $N_d(x) = N - N_u(x)$ is the number of d's. The probability that the binomial stock price follows one of the two paths S^x, S^y is given by $\mathbb{P}(S^x) + \mathbb{P}(S^y)$ (and similarly for any number of paths).

[†]We shall say more about the probabilistic interpretation of the binomial model in Chapter 5.

Exercise 2.1. *Prove that*

$$\sum_{x\in\{u,d\}^N} \mathbb{P}(S^x) = 1.$$

2.2 Binomial markets

A 1+1 dimensional binomial market is a market that consists of a risky asset, say a stock, and a risk-free asset, such that the price of the risky asset is given by the binomial model (2.1). As to the risk-free asset, we assume that its value $B(t)$ at time $t \in \mathcal{I}$ is given by

$$B(t) = B_0 e^{rt}, \quad t \in \mathcal{I}, \tag{2.2}$$

where B_0 is the present (at time $t = 0$) value of the asset. In particular we assume that the interest rate of the risk-free asset r is constant (not necessarily positive). As we shall only consider markets with one stock and one risk-free asset, we refer to 1+1 dimensional binomial markets simply as binomial markets.

Remark 2.2 (Notation)**.** As a follow-up to Remark 2.1, we mention that our notation for the value of the risk-free asset is also slightly different from the standard one. In fact, most of the literature on the binomial model, e.g. [5], denotes the value of the risk-free asset at time t by $B(t) = B_0(1 + r)^t$. To translate our results into the standard notation one just needs to replace e^r with $(1 + r)$. As already reported in Remark 2.1, we follow the notation used in [3].

Remark 2.3 (Discounted price)**.** The **discounted price** of the stock in a binomial market is defined by $\widehat{S}(t) = e^{-rt}S(t)$ and has the following meaning: $\widehat{S}(t)$ is the amount that should be invested on the risk-free asset at time $t = 0$ in order that the value of this investment at time t replicates the value of the stock at time t. Note that, whenever $r > 0$, i.e., as long as investing in the risk-free asset ensures a positive return, we have $\widehat{S}(t) < S(t)$. The discounted price of the stock measures, roughly speaking, the loss in the stock value due to the "time-devaluation" of money expressed by the ratio $B_0/B(t) = e^{-rt}$. Equivalently, we say that the discounted price of a stock expresses the future stock price in terms of the "current value of money".

A **portfolio process** invested in a binomial market is a finite sequence

$$\{(h_S(t), h_B(t))\}_{t\in\mathcal{I}},$$

where $h_S(t)$ is the (real‡) number of shares invested in the stock and $h_B(t)$ is the number of shares invested in the risk-free asset during the time interval $(t - 1, t]$, $t \in \mathcal{I}$. The initial position of the investor is given by $(h_S(0), h_B(0))$. Without loss of generality we assume that

$$h_S(0) = h_S(1), \quad h_B(0) = h_B(1), \tag{2.3}$$

‡Recall that we assume that the agent can invest in any fractional number of shares of an asset. Of course, in the applications this has to be rounded to an integer number.

i.e, $(h_S(1), h_B(1))$ is the investor position on the closed interval $[0, 1]$ (and not just in the semi-open interval $(0, 1]$). We recall that $h_S(t) > 0$ means that the investor has a long position on the stock in the interval $(t-1, t]$, while $h_S(t) < 0$ corresponds to a short position.

Remark 2.4. It is clear that the investor will change the position on the stock and the risk-free asset according to the path followed by the stock price, and so $(h_S(t), h_B(t))$ is in general path-dependent. When we want to emphasize the dependence of a portfolio position on the path of the stock price we shall write

$$h_S(t) = h_S(t, x), \quad h_B(t) = h_B(t, x).$$

The value of the portfolio process at time t is given by

$$V(t) = h_S(t)S(t) + h_B(t)B(t). \tag{2.4}$$

We write $V(t) = V(t, x)$ if we want to emphasize the dependence of the portfolio value on the path of the stock price. Clearly,

$$V(t, x) = h_S(t, x)S(t, x) + h_B(t, x)B(t).$$

We say that the portfolio process is self-financing if purchasing more shares of one asset is possible only by selling shares of the other asset for an equivalent value (and not by infusing new cash into the portfolio), and, conversely, if any cash obtained by selling one asset is immediately re-invested to buy shares of the other asset (and not withdrawn from the portfolio). To translate this condition into a mathematical formula, recall that $(h_S(t), h_B(t))$ is the investor position on the stock and the risk-free asset during the time interval $(t-1, t]$. Let $V(t)$ be the value of this portfolio at the time t, given by (2.4). At the time t, the investor sells/buys shares of the assets. Let $(h_S(t+1), h_B(t+1))$ be the new position on the stock and the risk-free asset in the interval $(t, t+1]$. Then the value of the new portfolio at time t is given by

$$V'(t) = h_S(t+1)S(t) + h_B(t+1)B(t).$$

The difference $V'(t) - V(t)$, if not zero, corresponds to cash withdrawn or added to the portfolio as a result of the change in the position on the assets. In a self-financing portfolio, however, this difference must be zero. We thus must have $V'(t) - V(t) = 0$, which leads to the following definition.

Definition 2.2. *A portfolio process $\{(h_S(t), h_B(t))\}_{t\in\mathcal{I}}$ invested in a binomial market is said to be* **self-financing** *if*

$$h_S(t)S(t-1) + h_B(t)B(t-1) = h_S(t-1)S(t-1) + h_B(t-1)B(t-1) \tag{2.5}$$

holds for all $t \in \mathcal{I}$.

Remark 2.5. Constant (i.e., time-independent) portfolio processes are clearly self-financing. In fact, (2.5) holds if $h_S(t) = h_S(t-1)$ and $h_B(t) = h_B(t-1)$, for all $t \in \mathcal{I}$.

In the next theorem we show that the self-financing property of a portfolio process is very restrictive: the value of a self-financing portfolio at time $t = N$ determines the value of the portfolio at any earlier time $t = 0, \dots, N-1$. This result is crucial to justify our definition of binomial fair price of European derivatives in the next chapter.

Theorem 2.1. *Let $\{(h_S(t), h_B(t))\}_{t\in\mathcal{I}}$ be a self-financing portfolio process with value $V(N)$ at time $t = N$. Define*

$$q_u = \frac{e^r - e^d}{e^u - e^d}, \quad q_d = 1 - q_u = \frac{e^u - e^r}{e^u - e^d}. \tag{2.6}$$

Then for $t = 0, \dots N-1$, $V(t)$ is given by

$$V(t) = e^{-r(N-t)} \sum_{(x_{t+1},\dots,x_N)\in\{u,d\}^{N-t}} q_{x_{t+1}} \cdots q_{x_N} V(N, x). \tag{2.7}$$

In particular, at time $t = 0$ we have

$$V(0) = e^{-rN} \sum_{x\in\{u,d\}^N} (q_u)^{N_u(x)} (q_d)^{N_d(x)} V(N, x). \tag{2.8}$$

Moreover the portfolio value satisfies the following recurrence formula:

$$V(t-1) = e^{-r}[q_u V^u(t) + q_d V^d(t)], \quad \text{for } t \in \mathcal{I}, \tag{2.9}$$

where

$$V^u(t) = h_S(t)S(t-1)e^u + h_B(t)B(t-1)e^r$$

is the value of the portfolio at time t, assuming that the stock price goes up at this time, and

$$V^d(t) = h_S(t)S(t-1)e^d + h_B(t)B(t-1)e^r$$

is the value of the portfolio at time t, assuming that the stock price goes down at this time.

Proof. We prove the statement by induction on $t = 0, \dots N-1$.

Step 1. We first prove it for $t = N-1$. In this case the sum in (2.7) is over two terms, one for which $x_N = u$ and one for which $x_N = d$. Hence the claim becomes

$$V(N-1) = e^{-r}[q_u V(N, (x_1, \dots, x_{N-1}, u)) + q_d V(N, (x_1, \dots, x_{N-1}, d))]. \tag{2.10}$$

In the right hand side of (2.10) we replace

$$V(N, (x_1, \dots, x_{N-1}, u)) = h_S(N)S(N-1)e^u + h_B(N)B(N-1)e^r$$

and

$$V(N, (x_1, \dots, x_{N-1}, d)) = h_S(N)S(N-1)e^d + h_B(N)B(N-1)e^r,$$

which follow by the definition of portfolio value. So doing we obtain

$$\begin{aligned}
\text{r.h.s (2.10)} =& e^{-r}[q_u(h_S(N)S(N-1)e^u + h_B(N)B(N-1)e^r) \\
& + q_d(h_S(N)S(N-1)e^d + h_B(N)B(N-1)e^r)] \\
&= e^{-r}[h_S(N)S(N-1)(e^u q_u + e^d q_d) + h_B(N)B(N-1)e^r(q_u+q_d)] \\
&= e^{-r}[h_S(N)S(N-1)e^r + h_B(N)B(N-1)e^r] \\
&= h_S(N)S(N-1) + h_B(N)B(N-1),
\end{aligned} \tag{2.11}$$

where we used that $e^u q_u + e^d q_d = e^r$ and $q_u + q_d = 1$. By the definition of self-financing portfolio process, the expression in the last line of (3.12) is $V(N-1)$, which proves the claim for $t = N-1$.

Step 2. Now assume that the statement is true at time $t+1$, i.e.,

$$V(t+1) = e^{-r(N-t-1)} \sum_{(x_{t+2},\dots x_N)\in\{u,d\}^{N-t-1}} q_{x_{t+2}} \cdots q_{x_N} V(N,x).$$

Step 3. We now prove it at time t. Let

$$\begin{aligned}
V^u(t+1) &\equiv V(t+1) \text{ assuming } x_{t+1} = u \\
&= h_S(t+1)S(t)e^u + h_B(t+1)B(t)e^r
\end{aligned}$$

and similarly we define $V^d(t+1)$. Proceeding exactly as in (2.11) above gives the identity

$$e^{-r}[q_u V^u(t+1) + q_d V^d(t+1)] = h_S(t+1)S(t) + h_B(t+1)B(t).$$

By the self-financing property, the right hand side is $V(t)$. Thus

$$V(t) = e^{-r}[q_u V^u(t+1) + q_d V^d(t+1)], \tag{2.12}$$

which proves (2.9). On the other hand, by the induction hypothesis of step 2 we have

$$V^u(t+1) = e^{-r(N-t-1)} \sum_{(x_{t+2},\dots,x_N)\in\{u,d\}^{N-t-1}} q_{x_{t+2}} \cdots q_{x_N} V(N, x_1, \dots, x_t, u, x_{t+2}, \dots, x_N),$$

$$V^d(t+1) = e^{-r(N-t-1)} \sum_{(x_{t+2},\dots,x_N)\in\{u,d\}^{N-t-1}} q_{x_{t+2}} \cdots q_{x_N} V(N, x_1, \dots, x_t, d, x_{t+2}, \dots, x_N).$$

Replacing in (2.12) we obtain

$$\begin{aligned}
V(t) = e^{-r(N-t)} &\Bigg[q_u \sum_{(x_{t+2},\dots,x_N)\in\{u,d\}^{N-t-1}} q_{x_{t+2}} \cdots q_{x_N} V(N, x_1, \dots, x_t, u, x_{t+2}, \dots, x_N) \\
&+ q_d \sum_{(x_{t+1},\dots,x_N)\in\{u,d\}^{N-t-1}} q_{x_{t+2}} \cdots q_{x_N} V(N, x_1, \dots, x_t, d, x_{t+2}, \dots, x_N) \Bigg] \\
= e^{-r(N-t)} & \sum_{(x_{t+1},\dots x_N)\in\{u,d\}^{N-t}} q_{x_{t+1}} \cdots q_{x_N} V(N,x),
\end{aligned}$$

which completes the proof. $\square$

Remark 2.6. As the sum in (2.7) is over $(x_{t+1}, \dots, x_N)$, while $V(N)$ depends on the full path $x \in \{u, d\}^N$, then $V(t)$ depends on the path of the stock price up to time t, i.e.,

$$V(t) = V(t, x_1, x_2, \dots, x_t).$$

Remark 2.7. The proof makes use of the important identities

$$q_u + q_d = 1, \quad q_u e^u + q_d e^d = e^r, \tag{2.13}$$

which will be used several times in the sequel. Note that (2.13) uniquely determines the coefficients q_u, q_d as in (2.6).

Example. We conclude this section with an example of application of (2.7) and (2.9). Assume $N = 2$ and consider the constant (and then self-financing) portfolio $h_S(t) = 1$, $h_B(t) = -1$, $t = 0, 1, 2$. Then

$$V(2) = S(2) - B(2), \quad B(2) = B_0 e^{2r}.$$

The possible paths are $x \in \{u, d\}^2$. Hence $S(2)$ can take 3 possible values:

1. $S(2) = S(2, (u, u)) = S_0 e^{2u}$, for $x = (u, u)$;
2. $S(2) = S(2, (u, d)) = S(2, (d, u)) = S_0 e^{u+d}$, for $x = (u, d)$ or $x = (d, u)$;
3. $S(2) = S(2, (d, d)) = S_0 e^{2d}$, for $x = (d, d)$.

The value of the portfolio at time $t = 2$ along all possible paths is

1. $V(2) = V(2, (u, u)) = S_0 e^{2u} - B_0 e^{2r}$, for $x = (u, u)$;
2. $V(2) = V(2, (u, d)) = V(2, (d, u)) = S_0 e^{u+d} - B_0 e^{2r}$, for $x = (u, d)$ or $x = (d, u)$;
3. $V(2) = V(2, (d, d)) = S_0 e^{2d} - B_0 e^{2r}$, for $x = (d, d)$.

Hence (2.7) gives, for $N = 2$, $t = 1$, and $x_1 = u$,

$$V(1, u) = e^{-r}(q_u V(2, (u, u)) + q_d V(2, (u, d)) = e^{-r}(q_u(S_0 e^{2u} - B_0 e^{2r}) + q_d(S_0 e^{u+d} - B_0 e^{2r})).$$

Using the identities (2.13), we obtain

$$V(1, u) = S_0 e^u - B_0 e^r, \tag{2.14}$$

which is of course correct because the right hand side is $S(1, (u, u)) - B(1)$, i.e., the value of the portfolio at time $t = 1$, assuming that the price of the stock goes up at this time. Similarly, applying (2.7) for $x_1 = d$ we obtain the (correct) formula

$$V(1, d) = S_0 e^d - B_0 e^r. \tag{2.15}$$

Finally we compute $V(0)$ with (2.7):

$$V(0) = e^{-2r}[(q_u)^2V(2,(u,u)) + q_uq_dV(2,(u,d)) + q_dq_uV(2,(d,u)) + (q_d)^2V(2,(d,d))].$$

After the proper simplifications we obtain

$$V(0) = S_0 - B_0,$$

which is of course correct by definition of portfolio value at time $t = 0$. Note that we obtain the same result by applying the recurrence formula (2.9). In fact, as $V^u(1) = V(1,u)$ and $V^d(1) = V(1,d)$, and using (2.14), (2.15), we obtain

$$V(0) = e^{-r}(q_uV^u(1) + q_dV^d(1)) = e^{-r}((q_ue^u + q_de^d)S_0 - (q_u + q_d)e^r) = S_0 - B_0.$$

2.3 Arbitrage portfolio

Our next purpose is to prove that the binomial market introduced in the previous section is arbitrage free, provided its parameters satisfy $d < r < u$ (no condition is required on the probability p). To achieve this we first need to introduce the precise definition of arbitrage portfolio.

Definition 2.3. *A portfolio process $\{(h_S(t), h_B(t)\}_{t\in\mathcal{I}}$ invested in a binomial market is called an* **arbitrage portfolio** *if its value $V(t)$ satisfies*

1) $V(0) = 0$;

2) $V(N,x) \geq 0$, for all $x \in \{u,d\}^N$;

3) There exists $y \in \{u,d\}^N$ such that $V(N,y) > 0$.

We say that the market is **arbitrage-free** *if there exist no self-financing arbitrage portfolio process invested in the stock and the risk-free asset.*

Let us comment on the previous definition. Condition 1) means that no initial wealth is required to set up the portfolio, i.e., the long and short positions on the two assets are perfectly balanced. In particular, anyone can (in principle) open such portfolio. Condition 2) means that the investor is sure not to loose money with this investment: regardless of the path followed by the stock price, the return of the portfolio is always non-negative. Condition 3) means that there is a strictly positive probability to make a profit, since along at least one path of the stock price the return of the portfolio is strictly positive.

Theorem 2.2. *The binomial market is arbitrage free if and only if $d < r < u$.*

Proof. We divide the proof in two steps. In step 1 we prove the claim for the **1-period model**, i.e., $N = 1$. The (easy) generalization to the **multiperiod model** ($N > 1$) is carried out in step 2. Note also the claim of the theorem is logically equivalent to the following: *There exists a self-financing arbitrage portfolio in the binomial market if and only if* $r \notin (d, u)$. It is the latter claim which is actually proved below.

Step 1: the 1-period model. Because of our convention (2.3), we can set

$$h_S(0) = h_S(1) = h_S, \quad h_B(0) = h_B(1) = h_B,$$

i.e., the portfolio position in the 1-period model is constant (and thus self-financing) over the interval $[0, 1]$. The value of the portfolio at time $t = 0$ is

$$V(0) = h_S S_0 + h_B B_0,$$

while at time $t = 1$ it is given by one of the following:

$$V(1) = V(1, u) = h_S S_0 e^u + h_B B_0 e^r,$$

if the stock price goes up at time $t = 1$, or

$$V(1) = V(1, d) = h_S S_0 e^d + h_B B_0 e^r,$$

if the stock price goes down at time $t = 1$. Thus the portfolio is an arbitrage if $V(0) = 0$, i.e.,

$$h_S S_0 + h_B B_0 = 0, \tag{2.16}$$

if $V(1) \geq 0$, i.e.,

$$h_S S_0 e^u + h_B B_0 e^r \geq 0, \tag{2.17}$$
$$h_S S_0 e^d + h_B B_0 e^r \geq 0, \tag{2.18}$$

and if at least one of the inequalities in (2.17)-(2.18) is strict. Now assume that (h_S, h_B) is an arbitrage portfolio. From (2.16) we have $h_B B_0 = -h_S S_0$ and therefore (2.17)-(2.18) become

$$h_S S_0(e^u - e^r) \geq 0, \tag{2.19}$$
$$h_S S_0(e^d - e^r) \geq 0. \tag{2.20}$$

Since at least one of the inequalities must be strict, then $h_S \neq 0$. If $h_S > 0$, then (2.19) gives $u \geq r$, while (2.20) gives $d \geq r$. As $u > d$, the last two inequalities are equivalent to $r \leq d$. Similarly, for $h_S < 0$ we obtain $u \leq r$ and $d \leq r$ which, again due to $u > d$, are equivalent to $r \geq u$. We conclude that the existence of an arbitrage portfolio implies $r \leq d$ or $r \geq u$, that is $r \notin (d, u)$. This proves that for $r \in (d, u)$ there is no arbitrage portfolio in the one period model, and thus the "if" part of the theorem is proved for $N = 1$. To prove the "only if" part, i.e., the fact that $r \in (d, u)$ is necessary for the absence of arbitrages, we construct an arbitrage portfolio when $r \notin (d,u)$. Assume $r \leq d$. Let us pick $h_S = 1$ and $h_B = -S_0/B_0$. Then $V(0) = 0$. Moreover (2.18) is trivially satisfied and, since $u > d$,

$$h_S S_0 e^u + h_B B_0 e^r = S_0(e^u - e^r) > S_0(e^d - e^r) \geq 0,$$

hence the inequality (2.17) is strict. This shows that one can construct an arbitrage portfolio if $r \leq d$ and a similar argument can be used to find an arbitrage portfolio when $r \geq u$. The proof of the theorem for the 1-period model is complete.

Step 2: the multiperiod model. Let $r \notin (d,u)$. As shown in the previous step, there exists an arbitrage portfolio (h_S, h_B) in the single period model. We can now build a self-financing arbitrage portfolio process $\{(h_S(t), h_B(t))\}_{t\in\mathcal{I}}$ for the multiperiod model by investing at time $t = 1$ the whole value of the portfolio (h_S, h_B) in the risk-free asset. The value of this portfolio process satisfies $V(0) = 0$ and $V(N,x) = V(1,x)e^{r(N-1)} \geq 0$ along every path $x \in \{u,d\}^N$. Moreover, since (h_S, h_B) is an arbitrage, then $V(1,y) > 0$ along some path $y \in \{u,d\}^N$ and thus $V(N,y) > 0$. Hence the constructed self-financing portfolio process $\{(h_S(t), h_B(t))\}_{t\in\mathcal{I}}$ is an arbitrage and the "if" part of the theorem is proved. To prove the "only if" part for the multiperiod model, we use that, by Theorem 2.1,

$$V(0) = e^{-rN} \sum_{x\in\{u,d\}^N} (q_u)^{N_u(x)}(q_d)^{N_d(x)} V(N,x),$$

where q_u, q_d are given by (2.6), $N_u(x)$ is the number of "ups" in x and $N_d(x) = N - N_u(x)$ the number of "downs". Now, assume that the portfolio is an arbitrage. Then $V(0) = 0$ and $V(N,x) \geq 0$. Of course, the above sum can be restricted to the paths along which $V(N,x) > 0$, which exist since the portfolio is an arbitrage. But then the sum can be zero only if either one of the factors q_u, q_d is zero, or if they have opposite sign. Since $u > d$, the denominator in the expressions (2.6) is positive, hence

$$q_u = 0, \quad \text{resp. } q_d = 0 \Rightarrow r = d, \quad \text{resp. } u = r,$$

$$(q_u > 0, q_d < 0), \quad \text{resp.} \quad (q_u < 0, q_d > 0) \Rightarrow u < r, \quad \text{resp. } r < d.$$

We conclude that the existence of a self-financing arbitrage portfolio entails $r \notin (d,u)$, which completes the proof of the theorem. □

Remark 2.8. The condition $d < r < u$ is equivalent to require that the parameters q_u, q_d given by (2.6) satisfy $q_u \in (0,1)$ and $q_d = 1 - q_u \in (0,1)$, i.e., that the pair (q_u, q_d) defines a probability, which we call **risk-neutral probability** or **martingale probability**; the reason for this terminology will become clear in Chapter 5.

2.4 Computation of the binomial stock price with Matlab

In this section we explain how to compute the binomial stock price with Matlab. More precisely, our goal is to construct a binomial tree for the stock price in some interval $[0,T]$, with $T > 0$ measured in fraction of years. Let us start by dividing the interval $[0, T]$ into N subintervals of length $h = T/N$, i.e.,

$$0 = t_1 < t_2 < \cdots < t_{N+1} = T, \quad t_{i+1} = t_i + h, \quad i = 1, \ldots, N.$$

Let $S(i) = S(t_i)$. We define the binomial stock price on the given partition of the interval $[0, T]$ as

$$S(i+1) = \begin{cases} S(i)e^u, & \text{with probability } p \\ S(i)e^d, & \text{with probability } 1-p \end{cases}, \quad i \in \mathcal{I}.$$

Now, it is clear that the smaller is h, the lower will be the change of the stock price on each subinterval, i.e., the smaller will be the parameters u, d. To make this precise, we define the **expected log-return** and the **volatility** of the stock in the interval $[0, T]$ as

$$\alpha = \frac{1}{h}[pu + (1-p)d], \quad \sigma = \frac{u-d}{\sqrt{h}}\sqrt{p(1-p)}. \tag{2.21}$$

The reason to define α and σ in this form will be explained at the end of Section 5.2.1. Note that α, σ are expressed in yearly percentage. Inverting (2.21) we obtain the relations

$$u = \alpha h + \sigma\sqrt{\frac{1-p}{p}}\sqrt{h}, \quad d = \alpha h - \sigma\sqrt{\frac{p}{1-p}}\sqrt{h}, \tag{2.22}$$

so that $u, d \to 0$ as $h \to 0$.

The following code defines the Matlab function *BinomialStock* which generates the binomial tree for the stock price on the partition $Q = \{t_1, \ldots t_{N+1}\}$ of the interval $[0, T]$:

```
function [Q,S]=BinomialStock(p,alpha,sigma,T,s,N)
h=T/N;
u=alpha*h+sigma*sqrt(h)*sqrt((1-p)/p);
d=alpha*h-sigma*sqrt(h)*sqrt(p/(1-p));
Q=zeros(N+1,1);
S=zeros(N+1);
Q(1)=0;
S(1,1)=s;
for j=1:N
Q(j+1)=j*h;
S(1,j+1)=S(1,j)*exp(u);
for i=1:j
S(i+1,j+1)=S(i,j)*exp(d);
end
end
```

The arguments of the function are the parameters p, α, σ, the time $T > 0$ (expressed in fraction of years), the initial price of the stock s and the number of steps N in the binomial model. The function returns a column vector Q containing the times $t_1, \dots t_{N+1}$ of the partition, and an upper-triangular $(N+1) \times (N+1)$ matrix S. The column j of S contains, in decreasing order along rows, the possible prices of the stock at time $t_j = (j-1)h$. A path of the stock price is obtained by moving from each column to the next one by either staying in the same row (which means that the price went up at this step) or going down one row (which means that the price went down at this step). For example, by running the command

```
[Q,S]=BinomialStock(0.5,0.01,0.02,1/12,10,5);
```

we get the output

$$Q = \begin{matrix} 0 \\ 0.0167 \\ 0.0333 \\ 0.0500 \\ 0.0667 \\ 0.0833 \end{matrix}$$

$$S = \begin{matrix} 10.0000 & 10.0275 & 10.0551 & 10.0828 & 10.1106 & 10.1384 \\ 0 & 9.9759 & 10.0033 & 10.0309 & 10.0585 & 10.0862 \\ 0 & 0 & 9.9518 & 9.9792 & 10.0067 & 10.0342 \\ 0 & 0 & 0 & 9.9278 & 9.9551 & 9.9825 \\ 0 & 0 & 0 & 0 & 9.9039 & 9.9311 \\ 0 & 0 & 0 & 0 & 0 & 9.8800 \end{matrix} \tag{2.23}$$

As $T = 1/12$ years and $N = 5$, the matrix S shows the stock prices in a period of 1 month with intervals of $1/(12 * 5)$ years, i.e., about 6 days[§].

Exercise 2.2. *Define a function RandomPath(S) that generates a random path from the binomial tree S created with the function BinomialStock and compute its probability. Plot the price of the stock as a function of time.*

We remark that it is impossible to generate all possible 2^N paths of the stock price. Even for a relatively small number of steps, like $N \approx 50$, the number of admissible paths is too big. Note also that the probability of each single path is practically zero for $N \gtrsim 20$. For instance, if we assume $p = 1/2$, then each single path has equal probability $\mathbb{P}(S^x) = (1/2)^N$, which is approximately $10^{-4}\%$ for $N = 20$. In Chapter 5 we shall learn how to compute the (much more interesting) probability that the price $S(t)$ lies within a given range $[a, b]$, i.e., $\mathbb{P}(a < S(t) < b)$.

[§] We do not adjust our calculations to take into account that markets are closed in the week-ends.

Chapter 3

European derivatives

This chapter deals with the pricing theory of European derivatives on a single stock under the assumption that the price of the underlying is given by $S(0) = S_0$ at time $t = 0$ and

$$S(t) = \begin{cases} S(t-1)e^u & \text{with probability } p, \\ S(t-1)e^d & \text{with probability } 1-p. \end{cases}, \quad t \in \mathcal{I} = \{1, \dots, N\}.$$

Here $0 < p < 1$ and $u > d$. $S(t)$ is the binomial stock price at time $t \in \mathcal{I}$. It is assumed that the initial price S_0 is known. We say that the price goes up at time t if $S(t) = S(t-1)e^u$ and that it goes down at time t if $S(t) = S(t-1)e^d$ (although this terminology is strictly correct only if $d < 0$ and $u > 0$, which is often the case in the applications). Moreover we assume the existence of a risk-free asset with value

$$B(t) = B_0 e^{rt}, \quad t \in \mathcal{I},$$

where $B_0 = B_0$ is the initial value and r is the constant interest rate. We impose $d < r < u$. In particular, the market is arbitrage-free and the risk-neutral probability is well defined:

$$q_u = \frac{e^r - e^d}{e^u - e^d}, \quad q_d = \frac{e^u - e^r}{e^u - e^d}, \qquad q_u, q_d \in (0,1), \; q_u + q_d = 1.$$

We denote by $\{(h_S(t), h_B(t)\}_{t \in \mathcal{I}}$ a portfolio process invested in the stock and the risk-free asset, where $(h_S(t), h_B(t))$ is the portfolio position in the interval $(t-1, t]$ and $h_S(0) = h_S(1)$, $h_B(0) = h_B(1)$. The value at time t of the portfolio is $V(t) = h_S(t)S(t) + h_B(t)B(t)$. Recall that a portfolio is said to be self-financing if

$$V(t-1) = h_S(t)S(t-1) + h_B(t)B(t-1),$$

which means that no cash is ever withdrawn or added to the portfolio. The value of a self-financing portfolio process at time t is uniquely determined by its value at time N. In fact we have the formula

$$V(t) = e^{-r(N-t)} \sum_{(x_{t+1}, \dots x_N) \in \{u,d\}^{N-t}} q_{x_{t+1}} \cdots q_{x_N} V(N, x). \tag{3.1}$$

which we proved in Theorem 2.1.

Definition 3.1. *A portfolio process* $\{(h_S(t), h_B(t)\}_{t\in\mathcal{I}}$ *is called* **predictable** *if there exist* N *functions* $H_1, \dots H_N$ *such that* $H_t : (0,\infty)^t \to \mathbb{R}^2$ *and*

$$(h_S(t), h_B(t)) = H_t(S_0, \dots, S(t-1)), \quad t \in \mathcal{I}.$$

Hence the portfolio position at time t of a predictable portfolio process is a **deterministic function*** of the stock price up to the time $t-1$. In particular the portfolio position in the interval $(t-1, t]$ is determined by the information available at time $t-1$.

3.1 The binomial price of European derivatives

Consider a European derivative on the stock with pay-off Y at the time of maturity $T = N$. The derivative is called **standard** if its pay-off depends only on the price of the stock at maturity, i.e., $Y = g(S(N))$, for some function $g : (0,\infty) \to \mathbb{R}$, which is called the **pay-off function** of the derivative. A European call, for instance, is a standard European derivative with pay-off function $g(z) = (z-K)_+$, where $K > 0$ is the strike price of the call. In the case of a **non-standard** European derivative, the pay-off is a function of the stock price at all times earlier and including N, i.e., $Y = g(S(1), \dots, S(N))$. Of course, in both cases the pay-off depends on the path followed by the stock price, in particular,

$$Y(x) = g(S(N,x)), \quad \text{for standard European derivatives}$$

and

$$Y(x) = g(S(1,x), S(2,x), \dots, S(N,x)), \quad \text{for non-standard European derivatives.}$$

Now, assume that a European derivative is sold at time $t < T$ for the price $\Pi_Y(t)$. The first concern of the seller is to **hedge** the derivative, i.e., to invest the premium $\Pi_Y(t)$ in such a way that the seller portfolio value at the expiration date is enough to pay-off the buyer of the derivative. We assume that the seller invests the premium in the 1+1 dimensional binomial market consisting of the underlying stock and the risk-free asset (**delta-hedging**).

Definition 3.2. *An* **hedging** *portfolio for a European derivative with pay-off* Y *at the expiration date* $T = N$ *is a portfolio process* $\{(h_S(t), h_B(t))\}_{t\in\mathcal{I}}$ *invested in the underlying stock and the risk-free asset such that its value* $V(t)$ *satisfies* $V(N) = Y$*; the latter equality must be satisfied for all possible paths of the price of the underlying stock, i.e.,* $V(N,x) = Y(x)$*, for all* $x \in \{u,d\}^N$*.*

It follows by (3.1) that the value $V(t)$ of (any) self-financing hedging portfolio at time t is given by

$$V(t) = e^{-r(N-t)} \sum_{(x_{t+1},\dots x_N)\in\{u,d\}^{N-t}} q_{x_{t+1}} \cdots q_{x_N} Y(x_1, \dots x_N). \tag{3.2}$$

*i.e., the function H_t does not depend on the path x of the stock.

Definition 3.3. *The* **binomial (fair) price** *of a European derivative with pay-off Y and maturity N is given by*

$$\Pi_Y(t) := e^{-r(N-t)} \sum_{(x_{t+1},\dots x_N)\in\{u,d\}^{N-t}} q_{x_{t+1}} \cdots q_{x_N} Y(x_1, \dots x_N), \tag{3.3}$$

for $t = 0, \dots N-1$, while $\Pi_Y(N) := Y$. In other words, the binomial price of the European derivative equals the value of self-financing portfolios hedging the derivative.

There are two reasons why it is reasonable to identify the fair price of the derivative with the value of self-financing hedging portfolios. From one hand the only purpose of the hedging portfolio is to pay-off the buyer of the derivative; the seller does not try to ensure a profit from the derivative (as it would be if $V(N) > Y$). The other reason is the absence of cash flow. In fact, if the writer needs to add cash to the portfolio in order to hedge the derivative, this would be unfair for the writer. On the other hand, if the writer could withdraw cash from the portfolio and still be able to hedge the derivative, this would be unfair for the buyer. Note carefully that we have *not* proved yet that an hedging self-financing portfolio process exists, but we know that, if it exists, its value is given by (3.2).

By Remark 2.6 in Chapter 2, the value $\Pi_Y(t)$ depends on the path of the stock price up to time t, i.e.,

$$\Pi_Y(t) = \Pi_Y(t, x_1, \dots, x_t).$$

Hence the binomial price of the European derivative at time t can be computed using the information available up to time t. In particular, for a standard European derivative with pay-off $Y = g(S(N))$ we can write, by Definition 2.1 in Chapter 2,

$$\begin{array}{l} S(N,x) = S_0 \exp(x_1 + \dots, x_N) \\ S(t,x) = S_0 \exp(x_1 + \dots x_t) \end{array} \Rightarrow S(N,x) = S(t) \exp(x_{t+1} + \dots x_N),$$

hence (3.3) becomes

$$\Pi_Y(t) := e^{-r(N-t)} \sum_{(x_{t+1},\dots x_N)\in\{u,d\}^{N-t}} q_{x_{t+1}} \cdots q_{x_N} g(S(t) \exp(x_{t+1} + \dots x_N)). \tag{3.4}$$

This shows that the fair price of a standard European derivative at time t is a deterministic function of $S(t)$, i.e., $\Pi_Y(t) = f_g(t, S(t))$, where

$$f_g(t,z) = e^{-r(N-t)} \sum_{(x_{t+1},\dots x_N)\in\{u,d\}^{N-t}} q_{x_{t+1}} \cdots q_{x_N} g(z \exp(x_{t+1} + \dots x_N)).$$

In particular, for a European call, resp. put, option with strike K and maturity $T = N$, the binomial fair price at time $t = 0, \dots, N-1$ can be written in the form $C(t, S(t), K, N)$, resp.

$P(t, S(t), K, N)$, where

$$C(t, S(t), K, N) = e^{-r(N-t)} \sum_{(x_{t+1},\dots x_N)\in\{u,d\}^{N-t}} q_{x_{t+1}} \cdots q_{x_N} (S(t)\exp(x_{t+1} + \dots x_N) - K)_+, \tag{3.5}$$

$$P(t, S(t), K, N) = e^{-r(N-t)} \sum_{(x_{t+1},\dots x_N)\in\{u,d\}^{N-t}} q_{x_{t+1}} \cdots q_{x_N} (K - S(t)\exp(x_{t+1} + \dots x_N))_+. \tag{3.6}$$

Exercise 3.1. *Use the epressions* (3.5)-(3.6) *to show that the binomial price of European calls and puts satisfy the properties in Theorem 1.1.*

Now let $\Pi^u_Y(t)$ denote the binomial price of the European derivative at time t assuming that the stock price goes up at time t (i.e., $S(t) = S(t-1)e^u$, or equivalently, $x_t = u$), and similarly for $\Pi^d_Y(t)$, with "up" replaced by "down". By the proven formula (2.9) in Theorem 2.1 we have

Theorem 3.1. *The binomial price of European derivatives satisfies the recurrence formula*

$$\Pi_Y(N) = Y, \quad \textit{and} \quad \Pi_Y(t) = e^{-r}[q_u\Pi^u_Y(t+1)+q_d\Pi^d_Y(t+1)], \quad \textit{for } t \in \{0, \dots, N-1\}. \tag{3.7}$$

The recurrence formula (3.7) is very useful to compute the binomial price of European derivatives.

Example. Assume $N = 2$ and $Y = S(2)^2 = g(S(2))$. Then, since $\Pi_Y(2) = g(S(2))$,

$$\Pi^u_Y(2) = g(S(1)e^u) = S(1)^2e^{2u}, \quad \Pi^d_Y(2) = g(S(1)e^d) = S(1)^2e^{2d}.$$

Hence (3.7) gives

$$\Pi_Y(1) = e^{-r}[q_u\Pi^u_Y(2) + q_d\Pi^d_Y(2)] = e^{-r}[q_uS(1)^2e^{2u} + q_dS(1)^2e^{2d}].$$

By the latter it follows that

$$\Pi^u_Y(1) = e^{-r}[q_u(S_0e^u)^2e^{2u} + (S_0e^u)^2e^{2d}] = e^{-r}S_0^2e^{2u}(q_ue^{2u} + q_de^{2d}),$$
$$\Pi^d_Y(1) = e^{-r}[q_u(S_0e^d)^2e^{2u} + (S_0e^d)^2e^{2d}] = e^{-r}S_0^2e^{2d}(q_ue^{2u} + q_de^{2d})$$

and therefore, again by (3.7),

$$\Pi_Y(0) = e^{-r}[q_u\Pi^u_Y(1) + q_d\Pi^d_Y(1)] = e^{-2r}S_0^2(q_ue^{2u} + q_de^{2d})^2.$$

Now, for $N = 2$ there exist 4 possible paths for the binomial stock price:

(p1) $S_0 \longrightarrow S_0e^u \longrightarrow S_0e^{2u}$, when $x = (u, u)$,

(p2) $S_0 \longrightarrow S_0e^u \longrightarrow S_0e^{u+d,}$ when $x = (u, d)$, (p3) $S_0 \longrightarrow S_0 e^d \longrightarrow S_0 e^{d+u}, \qquad \text{when } x = (d, u),$

(p4) $S_0 \longrightarrow S_0 e^d \longrightarrow S_0 e^{2d}, \qquad \text{when } x = (d, d),$

and to each of them there corresponds a path for the derivative price:

(p1) $\Pi_Y(0) = e^{-2r} S_0^2 (q_u e^{2u} + q_d e^{2d})^2 \to \Pi_Y^u(1) = e^{-r} S_0^2 e^{2u} (q_u e^{2u} + q_d e^{2d}) \to S_0^2 e^{4u}$

(p2) $\Pi_Y(0) = e^{-2r} S_0^2 (q_u e^{2u} + q_d e^{2d})^2 \to \Pi_Y^u(1) = e^{-r} S_0^2 e^{2u} (q_u e^{2u} + q_d e^{2d}) \to S_0^2 e^{2u+2d}$

(p3) $\Pi_Y(0) = e^{-2r} S_0^2 (q_u e^{2u} + q_d e^{2d})^2 \to \Pi_Y^d(1) = e^{-r} S_0^2 e^{2d} (q_u e^{2d} + q_d e^{2d}) \to S_0^2 e^{2u+2d}$

(p4) $\Pi_Y(0) = e^{-2r} S_0^2 (q_u e^{2u} + q_d e^{2d})^2 \to \Pi_Y^d(1) = e^{-r} S_0^2 e^{2d} (q_u e^{2d} + q_d e^{2d}) \to S_0^2 e^{4d}$

Note that the possible prices of the derivative in the future are known at time $t = 0$, although of course at time $t = 0$ we do not know which path will be followed. In particular, for $t = 1, 2$ the price $\Pi_Y(t)$ is *not* a deterministic function of S_0. However $\Pi_Y(t)$ is a deterministic function of $S(t)$, for $t = 0, 1, 2$.

We look at two more examples of computing the binomial price of European derivatives.

3.1.1 Example: A standard European derivative

Consider a standard European derivative with pay-off $Y = (\sqrt{S(2)} - 1)_+$ at maturity time $T = 2$. The binomial price of the stock may follow one of the following paths:

(p1) $S_0 = 1 \longrightarrow S(1) = 2 \longrightarrow S(2) = 4$

(p2) $S_0 = 1 \longrightarrow S(1) = 2 \longrightarrow S(2) = 2$

(p3) $S_0 = 1 \longrightarrow S(1) = 1 \longrightarrow S(2) = 2$

(p4) $S_0 = 1 \longrightarrow S(1) = 1 \longrightarrow S(2) = 1$

Assuming an interest rate $r > 0$ such that $e^r = 4/3$, and a probability $p = 1/4$ that the stock price goes up at any given time, compute the possible paths of the derivative price $\Pi_Y(t)$. Compute also the probability that the derivative expire in the money.

Solution: By the given paths of the stock price we derive that the binomial model

$$S(t) = \begin{cases} S(t-1)e^u \\ S(t-1)e^d \end{cases} \qquad t \in \{1, 2\}$$

has the following parameters:

$$e^u = 2, \quad e^d = 1.$$

Hence, since $e^r = 4/3$,

$$q_u = \frac{e^r - e^d}{e^u - e^d} = \frac{1}{3}, \quad q_d = 1 - q_u = \frac{2}{3}.$$

Now, let us write the binomial tree of the stock price, including the possible values of the derivative at the expiration time $T = 2$ (where we use that $\Pi_Y(2) = Y$):

$$\begin{tikzcd}
 & & S(2) = 4 \Rightarrow \Pi_Y(2) = (\sqrt{4} - 1)_+ = 1 \\
 & S(1) = 2 \arrow[ur, "u"] \arrow[dr, "d"] & \\
S_0 = 1 \arrow[ur, "u"] \arrow[dr, "d"] & & S(2) = 2 \Rightarrow \Pi_Y(2) = (\sqrt{2} - 1)_+ = \sqrt{2} - 1 \\
 & S(1) = 1 \arrow[ur, "u"] \arrow[dr, "d"] & \\
 & & S(2) = 1 \Rightarrow \Pi_Y(2) = (\sqrt{1} - 1)_+ = 0
\end{tikzcd}$$

Using the recurrence formula

$$\Pi_Y(t) = e^{-r}(q_u \Pi_Y^u(t+1) + q_d \Pi_Y^d(t+1))$$

we have, at time $t = 1$,

$$\begin{aligned}
S(1) = 2 \Rightarrow \Pi_Y(1) &= \frac{3}{4}(\frac{1}{3} \cdot 1 + \frac{2}{3}(\sqrt{2} - 1)) = \frac{1}{4}(2\sqrt{2} - 1) \\
S(1) = 1 \Rightarrow \Pi_Y(1) &= \frac{3}{4}(\frac{1}{3}(\sqrt{2} - 1) + \frac{2}{3} \cdot 0) = \frac{1}{4}(\sqrt{2} - 1)
\end{aligned}$$

while at time $t = 0$ we have

$$\Pi_Y(0) = \frac{3}{4}(\frac{1}{3} \cdot \frac{1}{4}(2\sqrt{2} - 1) + \frac{2}{3} \cdot \frac{1}{4}(\sqrt{2} - 1)) = \frac{1}{4}(\sqrt{2} - \frac{3}{4}).$$

Hence we have found the following diagram for the binomial price of the derivative

$$\begin{tikzcd}
 & & \Pi_Y(2) = 1 \\
 & \Pi_Y(1) = \frac{1}{4}(2\sqrt{2} - 1) \arrow[ur, "u"] \arrow[dr, "d"] & \\
\Pi_Y(0) = \frac{1}{4}(\sqrt{2} - \frac{3}{4}) \arrow[ur, "u"] \arrow[dr, "d"] & & \Pi_Y(2) = \sqrt{2} - 1 \\
 & \Pi_Y(1) = \frac{1}{4}(\sqrt{2} - 1) \arrow[ur, "u"] \arrow[dr, "d"] & \\
 & & \Pi_Y(2) = 0
\end{tikzcd}$$

As to the probability that the derivative expires in the money, i.e., $\mathbb{P}(\Pi_Y(2) > 0)$, we see from the above diagram that this happens along the paths $(u,u), (u,d), (d,u)$, hence

$$\mathbb{P}(\Pi(2) > 0) = \mathbb{P}(S^{(u,u)}) + \mathbb{P}(S^{(u,d)}) + \mathbb{P}(S^{(d,u)}) = \left(\frac{1}{4}\right)^2 + \frac{1}{4}\cdot\frac{3}{4} + \frac{3}{4}\cdot\frac{1}{4} = \frac{7}{16},$$

which corresponds to $43,75\%$.

3.1.2 Example: A non-standard European derivative

Consider a 3-period binomial market with the following parameters:

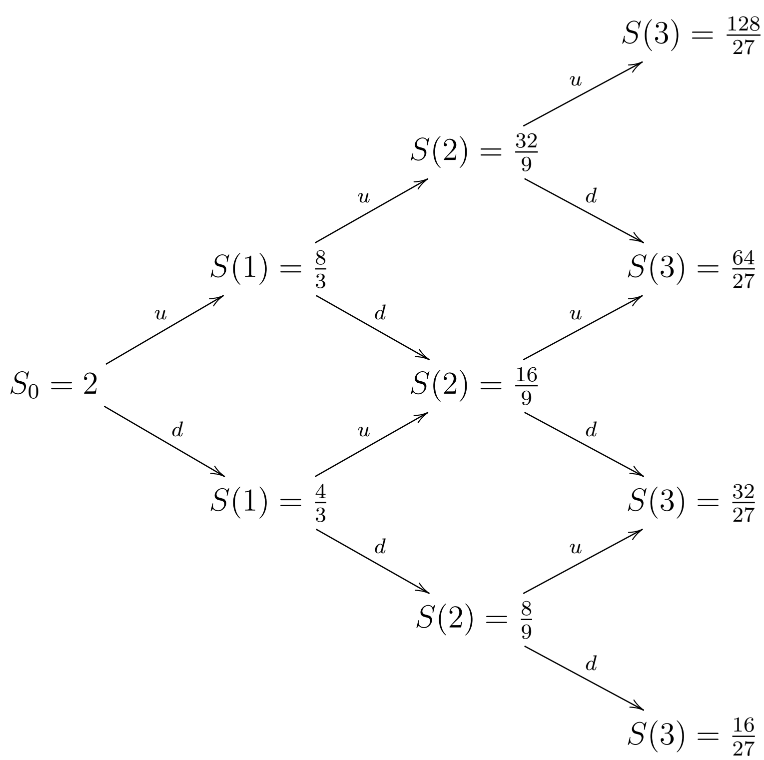
$$e^u = \frac{4}{3}, \quad e^d = \frac{2}{3}, \quad p = \frac{3}{4}.$$

Assuming $S_0 = 2$ and $r = 0$, compute the initial binomial price of the European derivative with pay-off

$$Y = \left(\frac{11}{9} - \min(S_0, S(1), S(2), S(3))\right)_+, \quad (z)_+ = \max(0, z),$$

and time of maturity $T = 3$. This is an example of **lookback option**. Compute the probability that the derivative expires in the money and the probability that the return of a constant portfolio with a long position on this derivative be positive.

Solution: We start by writing down the binomial tree of the stock price



To compute the initial binomial price of a non-standard European derivative it is convenient to use the formula

$$\Pi_Y(0) = e^{-rN} \sum_{x \in \{u,d\}^N} q_{x_1} \cdot \dots \cdot q_{x_N} Y(x), \tag{3.8}$$

where $Y(x)$ denotes the pay-off as a function of the path of the stock price. In this example we have $N = 3$, $r = 0$ and

$$q_u = q_d = \frac{1}{2}.$$

So, it remains to compute the pay-off for all possible paths of the binomial stock price, where

$$Y = \left(\frac{11}{9} - \min(S_0, S(1), S(2), S(3)) \right)_+, \quad (z)_+ = \max(0, z).$$

For instance

$$Y(u,u,u) = \left(\frac{11}{9} - \min(2, 8/3, 32/9, 128/27) \right)_+ = \left(\frac{11}{9} - 2 \right)_+ = \max(0, -\frac{7}{9}) = 0.$$

Similarly we find

$$\begin{aligned}
Y(u,u,d) &= \left(\frac{11}{9} - \min(2, 8/3, 32/9, 64/27) \right)_+ = 0 \\
Y(u,d,u) &= \left(\frac{11}{9} - \min(2, 8/3, 16/9, 64/27) \right)_+ = 0 \\
Y(u,d,d) &= \left(\frac{11}{9} - \min(2, 8/3, 16/9, 32/27) \right)_+ = 1/27 \\
Y(d,u,u) &= \left(\frac{11}{9} - \min(2, 4/3, 16/9, 64/27) \right)_+ = 0 \\
Y(d,u,d) &= \left(\frac{11}{9} - \min(2, 4/3, 16/9, 32/27) \right)_+ = 1/27 \\
Y(d,d,u) &= \left(\frac{11}{9} - \min(2, 4/3, 8/9, 32/27) \right)_+ = 1/3 \\
Y(d,d,d) &= \left(\frac{11}{9} - \min(2, 4/3, 8/9, 16/27) \right)_+ = 17/27
\end{aligned}$$

Replacing in (3.8) we obtain

$$\Pi_Y(0) = q_u (q_d)^2 Y(u,d,d) + (q_d)^2 q_u Y(d,u,d) + (q_d)^2 q_u Y(d,d,u) + (q_d)^3 Y(d,d,d),$$

the other terms being zero. Hence

$$\Pi_Y(0) = \frac{1}{8} \Big(\frac{1}{27} + \frac{1}{27} + \frac{1}{3} + \frac{17}{27} \Big) = \frac{7}{54}.$$

The probability that the derivative expires in the money is the probability that $Y > 0$. Hence we just sum the probabilities of the paths which lead to a positive pay-off:

$$\begin{aligned}
\mathbb{P}(Y > 0) &= \mathbb{P}(S^{(u,d,d)}) + \mathbb{P}(S^{(d,u,d)}) + \mathbb{P}(S^{(d,d,u)}) + \mathbb{P}(S^{(d,d,d)}) \\
&= p(1-p)^2 + (1-p)^2 p + (1-p)^2 p + (1-p)^3 \\
&= 3(1-p)^2 p + (1-p)^3 = 3\left(\frac{1}{4}\right)^2 \frac{3}{4} + \left(\frac{1}{4}\right)^3 = \frac{5}{32} \approx 15,6\%
\end{aligned}$$

Next consider a constant portfolio with a long position on the derivative. This means that we buy the derivative at time $t = 0$ and we wait (without changing the portfolio) until the expiration time $t = 3$. The return will be positive if $\Pi_Y(3) > \Pi_Y(0)$. But $\Pi_Y(3) = Y$, which, according to the computations above, is greater that $\Pi_Y(0) = 7/54$ only when the binomial stock price follows one of the paths (d, d, u) or (d, d, u). Hence

$$\mathbb{P}[\Pi_Y(3) > \Pi_Y(0)] = \mathbb{P}(S^{(d,d,u)}) + \mathbb{P}(S^{(d,d,d)}) = (1-p)^2 p + (1-p)^3 = (1-p)^2 = \frac{1}{16} \approx 6,2\%$$

3.2 Hedging portfolio

Next we treat the important problem of building a self-financing hedging portfolio for European derivatives. We only discuss standard derivatives.

Theorem 3.2. *Consider a standard European derivative with pay-off $Y = g(S(N))$ at the time of maturity N. Then the portfolio given by*

$$h_S(0) = h_S(1), \quad h_B(0) = h_B(1) \tag{3.9a}$$

and, for $t \in \mathcal{I}$,

$$h_S(t) = \frac{1}{S(t-1)} \frac{\Pi_Y^u(t) - \Pi_Y^d(t)}{e^u - e^d}, \tag{3.9b}$$

$$h_B(t) = \frac{e^{-r}}{B(t-1)} \frac{e^u \Pi_Y^d(t) - e^d \Pi_Y^u(t)}{e^u - e^d} \tag{3.9c}$$

is a self-financing, predictable, hedging portfolio.

Proof. We first show that the given portfolio hedges the derivative. We have

$$V(t) = h_S(t)S(t) + h_B(t)B(t) = \frac{S(t)}{S(t-1)} \frac{\Pi_Y^u(t) - \Pi_Y^d(t)}{e^u - e^d} + \frac{e^{-r}B(t)}{B(t-1)} \frac{e^u \Pi_Y^d(t) - e^d \Pi_Y^u(t)}{e^u - e^d}.$$

Note that $e^{-r}B(t)/B(t-1) = 1$, while $S(t)/S(t-1)$ is either e^u or e^d. By straightforward calculations we obtain $V^u(t) = \Pi_Y^u(t)$ and $V^d(t) = \Pi_Y^d(t)$, that it $V(t) = \Pi_Y(t)$, i.e., the portfolio replicates the derivative. In particular for $t = N$ we obtain $V(N) = \Pi_Y(N) = Y$, hence the portfolio is hedging the derivative. As to the self-financing property, we have

$$\begin{aligned} h_S(t)S(t-1) + h_B(t)B(t-1) &= \frac{\Pi_Y^u(t)(1-e^{d-r}) + \Pi_Y^d(t)(e^{u-r}-1)}{e^u - e^d} \\ &= e^{-r}(q_u\Pi_Y^u(t) + q_d\Pi_Y^d(t)) = \Pi_Y(t-1), \end{aligned}$$

where we used the definition of q_u, q_d, as well as the recurrence formula (3.7). By the already proven fact that the portfolio is replicating the derivative, we have $\Pi_Y(t-1) = V(t-1)$, thus

$$h_S(t)S(t-1) + h_B(t)B(t-1) = V(t-1),$$

which proves the self-financing property. Finally the portfolio is predictable because, by (3.4),

$$\Pi_Y^u(t) := e^{-r(N-t)} \sum_{(x_{t+1},\dots x_N)\in\{u,d\}^{N-t}} q_{x_{t+1}}\cdots q_{x_N} g(S(t-1)e^u \exp(x_{t+1}+\dots x_N)),$$

hence $\Pi_Y^u(t)$ is a deterministic function of $S(t-1)$ and the same property holds for $\Pi_Y^d(t)$. Therefore $h_S(t), h_B(t)$ are deterministic functions of $S(t-1)$, which proves in particular that the portfolio is predictable. □

Note that the portfolio (3.9) is more than just predictable: it is completely determined by the price of the stock at time $t-1$. It will now be shown that (3.9) is the only self-financing hedging portfolio satisfying this property (however it is in general not unique in the largest class of predictable portfolios, which are allowed to depend also on the stock price prior to the time $t-1$; see Exercise 3.3 below). We prove the claim for $N = 2$. Assume that the stock price goes up at time 2, i.e., $S(2) = S(1)e^u$. Then the value at time 2 of the portfolio is

$$V^u(2) := h_S(2)S(1)e^u + h_B(2)B(1)e^r.$$

If the price of the stock goes down at time 2 the value of the portfolio at time 2 is

$$V^d(2) := h_S(2)S(1)e^d + h_B(2)B(1)e^r.$$

Therefore, since $Y = g(S(2))$, the hedging condition is

$$\begin{aligned} h_S(2)S(1)e^u + h_B(2)B(1)e^r &= g(S(1)e^u), \\ h_S(2)S(1)e^d + h_B(2)B(1)e^r &= g(S(1)e^d). \end{aligned}$$

Since $u > d$, the previous system is solvable on $h_S(2), h_B(2)$ and the *unique* solution is given by

$$h_S(2) = \frac{1}{S(1)}\frac{g(S(1)e^u) - g(S(1)e^d)}{e^u - e^d}, \tag{3.10a}$$

$$h_B(2) = \frac{e^{-r}}{B(1)}\frac{e^u g(S(1)e^d) - e^d g(S(1)e^u)}{e^u - e^d}. \tag{3.10b}$$

In particular $h_S(2)$ and $h_B(2)$ depend only on $S(1)$ (and the given parameters r, u, d). Now, the self-financing property at time $t = 1$ is

$$V(1) = h_S(2)S(1) + h_B(2)B(1),$$

i.e., using $V(1) = h_S(1)S(1) + h_B(1)B(1)$ and (3.10),

$$h_S(1)S(1) + h_B(1)B(1) = e^{-r}[q_u g(S(1)e^u) + q_d g(S(1)e^d)].$$

Again distinguishing the cases $S(1) = S_0e^u$ and $S(1) = S_0e^d$, we derive the system

$$\begin{aligned} h_S(1)S_0e^u + h_B(1)B_0e^r &= e^{-r}[q_u g(S_0e^{2u}) + q_d g(S_0e^{u+d})] := f_u(S_0), \\ h_S(1)S_0e^d + h_B(1)B_0e^r &= e^{-r}[q_u g(S_0e^{u+d}) + q_d g(S_0e^{2d})] := f_d(S_0). \end{aligned}$$

As before, since $u > d$, the previous system is solvable in $h_S(1)$, $h_B(1)$ and the unique solution is given by

$$h_S(1) = \frac{1}{S_0}\frac{f_u(S_0) - f_d(S_0)}{e^u - e^d}, \tag{3.11a}$$

$$h_B(1) = \frac{e^{-r}}{B_0}\frac{e^u f_d(S_0) - e^d f_u(S_0)}{e^u - e^d}. \tag{3.11b}$$

Note that $h_S(1)$ and $h_B(1)$ are determined by S_0. Finally the self-financing property at time 0 is

$$V(0) = h_S(1)S_0 + h_B(1)B_0. \tag{3.12}$$

Since $V(0)$ may also be written as $V(0) = h_S(0)S_0 + h_B(0)B_0$, and we defined $h_S(0) = h_S(1)$, $h_B(0) = h_B(1)$, then (3.12) holds. Hence we have found that there exists a unique self-financing, hedging portfolio process which, at time t, depends only on $S(t-1)$. Moreover we claim that the found portfolio process is exactly (3.9). In fact, $\Pi_Y^u(2) = g(S(1)e^u)$ and $\Pi_Y^d(2) = g(S(1)e^d)$, hence we can rewrite (3.10) as

$$\begin{aligned} h_S(2) &= \frac{1}{S(1)}\frac{\Pi_Y^u(2) - \Pi_Y^d(2)}{e^u - e^d}, \\ h_B(2) &= \frac{e^{-r}}{B(1)}\frac{e^u\Pi_Y^d(2) - e^d\Pi_Y^u(2)}{e^u - e^d}, \end{aligned}$$

which is exactly (3.9b)-(3.9c) for $t = 2$. Similarly, one can show that (3.11) coincide with (3.9b)-(3.9c) for $t = 1$.

Example. Let us derive the hedging portfolio for $N = 2$ and $Y = S(2)^2$. We showed earlier that

$$\begin{aligned} &\Pi_Y^u(2) = S(1)^2e^{2u}, \quad \Pi_Y^d(2) = S(1)^2e^{2d}, \\ &\Pi_Y^u(1) = e^{-r}S_0^2e^{2u}(q_ue^{2u} + q_de^{2d}), \quad \Pi_Y^d(1) = e^{-r}S_0^2e^{2d}(q_ue^{2u} + q_de^{2d}). \end{aligned}$$

Thus

$$h_S(2) = S(1)\frac{e^{2u} - e^{2d}}{e^u + e^d} = S(1)(e^u + e^d), \quad h_S(1) = h_S(0) = e^{-r} S_0 \frac{q_u e^{4u} - q_d e^{4d}}{e^u - e^d}$$

and similarly one can compute $h_B(2)$ and $h_B(1) = h_B(0)$. Note that, corresponding to each path (p1)–(p4) of the binomial stock price, the number of shares of the stock in the writer portfolio is given by

(p1) $h_S(0) \to h_S(1) = h_S(0) \to h_S(2) = S_0 e^u(e^u + e^d)$ (replacing $S(1) = S_0 e^u$ in $h_S(2)$)

(p2) $h_S(0) \to h_S(1) = h_S(0) \to h_S(2) = S_0 e^u(e^u + e^d)$ (replacing $S(1) = S_0 e^u$ in $h_S(2)$)

(p3) $h_S(0) \to h_S(1) = h_S(0) \to h_S(2) = S_0 e^d(e^u + e^d)$ (replacing $S(1) = S_0 e^d$ in $h_S(2)$)

(p4) $h_S(0) \to h_S(1) = h_S(0) \to h_S(2) = S_0 e^d(e^u + e^d)$ (replacing $S(1) = S_0 e^d$ in $h_S(2)$)

Exercise 3.2. *Consider a 3-period binomial market with the following parameters:*

$$e^u = \frac{5}{4}, \quad e^d = \frac{1}{2}, \quad e^r = 1 \quad p = \frac{1}{2}.$$

Assume $S_0 = \frac{64}{25}$. Consider a European derivative expiring at time $T = 3$ and with pay-off

$$Y = S(3)H(S(3) - 1),$$

where H is the Heaviside function: $H(x) = 0$, if $x < 0$, $H(x) = 1$ if $x \geq 0$ (this is an example of a so called **digital option***). Compute the possible paths of the derivative price and for each of them give the number of shares of the underlying asset in the hedging portfolio process. Compute the probability that the return of a constant portfolio with a short position in the derivative be positive.*

Exercise 3.3. *Consider a standard European derivative with pay-off $Y = g(S(2))$ at the time of maturity 2. Assume that the price of the underlying stock follows the 2-period arbitrage-free binomial model*

$$S(t) = \begin{cases} S(t-1)e^u & \text{with probability } p, \\ S(t-1)e^d & \text{with probability } 1-p. \end{cases} \qquad t = 1, 2$$

and that the interest rate is a constant $r > 0$. Let

$$\Delta = g(S_0 e^{2d}) - e^{d-u} g(S_0 e^{u+d}) - g(S_0 e^{u+d}) + g(S_0 e^{2u}) e^{d-u}.$$

Show that a constant hedging portfolio (h_S, h_B) exists if and only if $\Delta = 0$ and find such portfolio

3.3 Computation of the binomial price of European derivatives with Matlab

As in Section 2.4, we consider a partition $0 = t_1 < t_2 < \cdots < t_{N+1} = T$ of the interval $[0, T]$ and the binomial stock price

$$S(i+1) = \begin{cases} S(i)e^u, & \text{with probability } p \\ S(i)e^d, & \text{with probability } 1-p \end{cases}, \quad i \in \mathcal{I} = \{1, \dots, N\},$$

where $S(i) = S(t_i)$ and

$$u = \alpha h + \sigma\sqrt{\frac{1-p}{p}}\sqrt{h}, \quad d = \alpha h - \sigma\sqrt{\frac{p}{1-p}}\sqrt{h}, \tag{3.13}$$

The value of the risk-free asset at time t_i is given by $B(t_i) = B_0 e^{rt_i} = B_0 e^{(rh)i}$. Hence the pair $(S(i), B(i))$ defines a 1+1 dimensional binomial market, in the sense of Section 2.2, with parameters u, d given by (3.13) and interest rate rh. The recurrence formula (3.7) for the price of a European option with pay-off Y at maturity T becomes

$$\Pi_Y(N+1) = Y, \quad \text{and} \quad \Pi_Y(i) = e^{-rh}[q_u \Pi_Y^u(i+1) + q_d \Pi_Y^d(i+1)], \quad \text{for } t \in \mathcal{I},$$

where $\Pi_Y(i) = \Pi_Y(t_i)$ and

$$q_u = \frac{e^{rh} - e^{\alpha h - \sigma\sqrt{\frac{p}{1-p}}\sqrt{h}}}{e^{\alpha h + \sigma\sqrt{\frac{1-p}{p}}\sqrt{h}} - e^{\alpha h - \sigma\sqrt{\frac{p}{1-p}}\sqrt{h}}}, \quad q_d = 1 - q_u.$$

The recurrence formula for standard European options is implemented by the Matlab function *BinomialEuropean* defined by the following code:

```
function P=BinomialEuropean(Q,S,r,g)
h=Q(2)-Q(1);
syms x;
f = sym(g);
N=length(Q)-1;
expu=S(1,2)/S(1,1);
expd=S(2,2)/S(1,1);
qu=(exp(r*h)-expd)/(expu-expd);
qd=(expu-exp(r*h))/(expu-expd);
if (qu<0 || qd<0)
display('Error:  the market is not arbitrage free.');
P=0;
return
end
P=zeros(N+1);
```

```
P(:,N+1)=subs(f,x,S(:,N+1));
for j=N:-1:1
for i=1:j
P(i,j)=exp(-r*h)*(qu*P(i,j+1)+qd*P(i+1,j+1));
end
end
```

The arguments are the partition $Q = \{t_1, \dots, t_{N+1}\}$ and the binomial tree S for the price of the underlying stock, computed with the function *BinomialStock* defined in Section 2.4, the interest rate r of the risk-free asset and the pay-off function g (e.g., `'max(x-10,0)'` for a call with strike $K = 10$). The function returns an upper-triangular $(N+1) \times (N+1)$ matrix which contains the binomial tree for the fair price of the derivative. The column j contains the possible prices of the derivative at time t_j. A path of the derivative price is obtained by moving from each column to next one by either stays in the same row (which means that the price of the underlying stock went up at this step) or going down one row (which means that the price of the underlying went down at this step). Note that the Matlab function also checks that (q_u, q_d) defines a probability, i.e., that the market is arbitrage free. If not the function stops.

For example, let S be the binomial tree (2.23) and run the command

```
P=European(Q,S,0.01,'max(x-10,0)')
```

which computes the binomial price of a European call with strike $K = 10$. The result is

$$P = \begin{array}{cccccc} 0.0284 & 0.0429 & 0.0623 & 0.0861 & 0.1122 & 0.1384 \\ 0 & 0.0139 & 0.0235 & 0.0386 & 0.0601 & 0.0862 \\ 0 & 0 & 0.0043 & 0.0085 & 0.0171 & 0.0342 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \tag{3.14}$$

Observe that if the price of the stock goes down in the first three steps, then the price of the call becomes zero and remains zero for all subsequent times, regardless of the future path of the stock price. This happens because the binomial model predicts that the call has no chance to expire in the money when the stock price goes down in the first three steps and hence the call becomes worthless. This of course is in contradiction with reality, as the market price of calls is never zero prior to expire[†]. This paradox of the binomial model becomes less and less important the higher is the number of steps used for the computation. On one hand if N is too low the contribution of the paths where the call price is zero is significant for the computation of the fair price of the call and will determine an underestimation of this price. On the other hand as $N \to +\infty$ the binomial price of the call converges to the Black-Scholes price (see Chapter 6), and within the Black-Scholes theory the fair price of a call is always

[†]In fact in real life we can never exclude with certainty that the call will expire in the money, irrespective of how deeply out of the money is the call today.

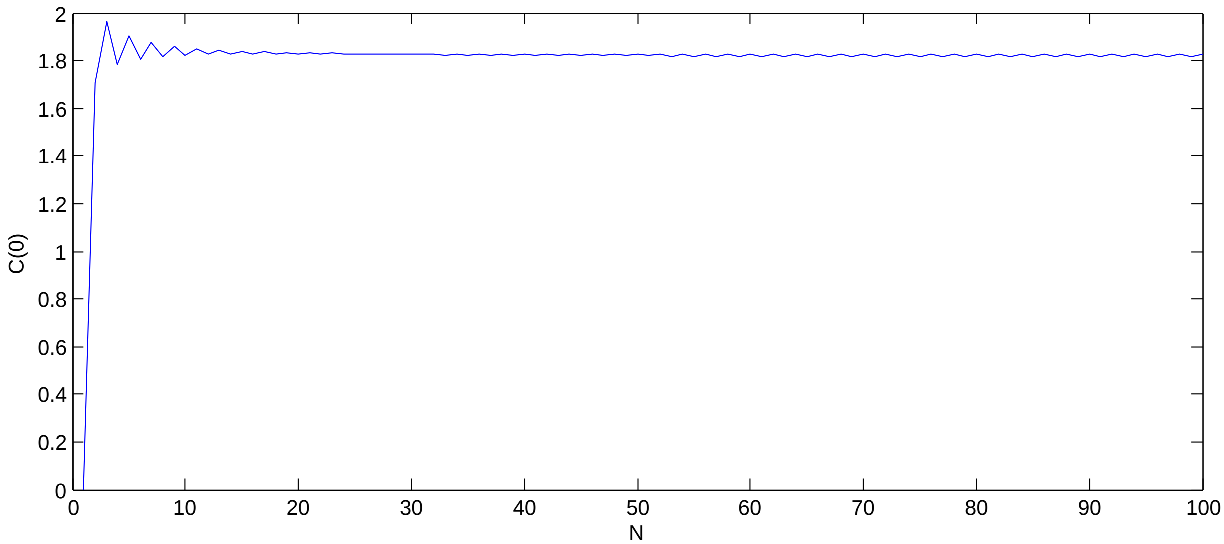


Figure 3.1: Initial price $C(0) = P(1,1)$ of the call computed for increasing values of N ($S_0 = 10, K = 10.5, T = 1, \alpha = 0.1, \sigma = 0.5, p = 1/2$). The price stabilizes around 1.8 for $N \gtrsim 20$.

positive. Moreover it can be shown that the binomial algorithm to compute the fair price of European derivatives is stable in the following sense: the larger is the number of steps, the smaller is the numerical error due, for instance, to truncations. In conclusion, *one can trust the results given by the binomial model only if N is sufficiently large.* As a way of example, Figure 3.1 shows the initial binomial price of a call as a function of $N = 2, \dots, 100$. It is clear that for a small number of steps the result is unreliable, while for sufficiently many steps, say $N \gtrsim 20$, the result becomes stable.

Figures 3.2 and 3.3 show the initial price of the call as a function of the parameters α and σ. We see that (i) the price of the call is very weakly dependent on the parameter α and (ii) the price of the call increases with the volatility of the stock. Hence the price of the call is not sensitive to the average movement of the stock price, but rather only on how uncertain is the stock price, measured by its volatility. The analytical proof of these results is more easily carried out in the Black-Scholes model, which is the time-continuum limit of the binomial model (i.e., $N \to \infty, h \to 0$ such that $Nh = T$); see Chapter 6.

Exercise 3.4 (?)**.** *Can you give an intuitive explanation for the found dependence on the parameter α, σ of the call option binomial price?*

Exercise 3.5. *Let $N = 100$ and consider a European call.*

1. *Plot the initial binomial price of the call as a function of S_0*

2. *Plot the initial binomial price of the call as a function of the strike price*

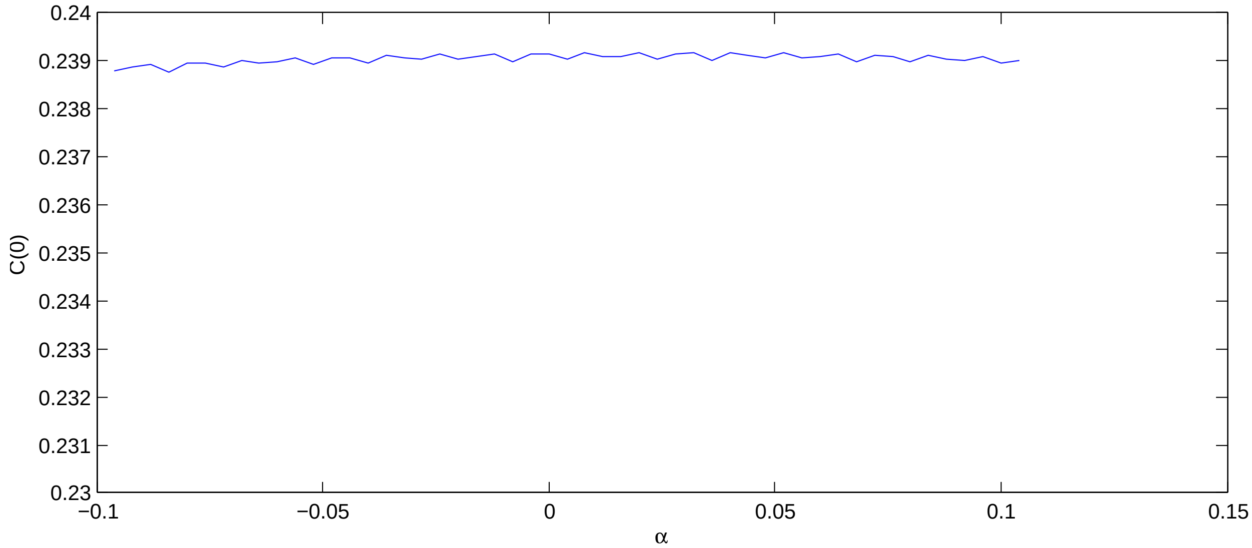


Figure 3.2: Initial price $C(0)$ of the call computed for increasing values of $\alpha \in [-0.1, 0.1]$ ($S_0 = 10, K = 10.5, T = 1, \sigma = 0.1, N = 1000, p = 1/2$). Clearly, the dependence on α of the call price is extremely weak.

3. *Verify the validity of the put-call parity*

Exercise 3.6. *Write a Matlab function which computes the hedging portfolio of a standard European derivative.*

Exercise 3.7. *Look for the market price of call options on the Apple stock which expire in the third Friday from now (the expiration date is always a Friday). Select 12 prices, 6 for options in the money and 6 for options out of the money. Let S_0 be the current value of Apple and σ_{20} the current 20-days volatility of Apple. Compile a table with the following information: the first column contains the strike price K, the second column the market price, the third column the binomial price. Use $\sigma = \sigma_{20}$ and $r = 0$ to compute the binomial price, but check that the result does not change significantly for, say, $0 \leq r \leq 0.05$ (which is a quite large value for the interest rate). Explain why. Plot the difference between the market price and the theoretical price as a function of K.*

The difference between the theoretical fair price and the market price of a call option is commonly expressed in terms of the **implied volatility** σ_{imp} of the call, which is defined as the value of the parameter σ to be used in the calculation of the binomial price in order for the latter to be equal to the market price‡. To this regard we remark that, as shown in Fig. 3.3, the binomial price is a monotone function of the volatility; moreover the market

‡Actually the computation of the implied volatility is typically done using the Black-Scholes model. However, as already mentioned, the binomial price and the Black-Scholes price are practically the same for N sufficiently large.

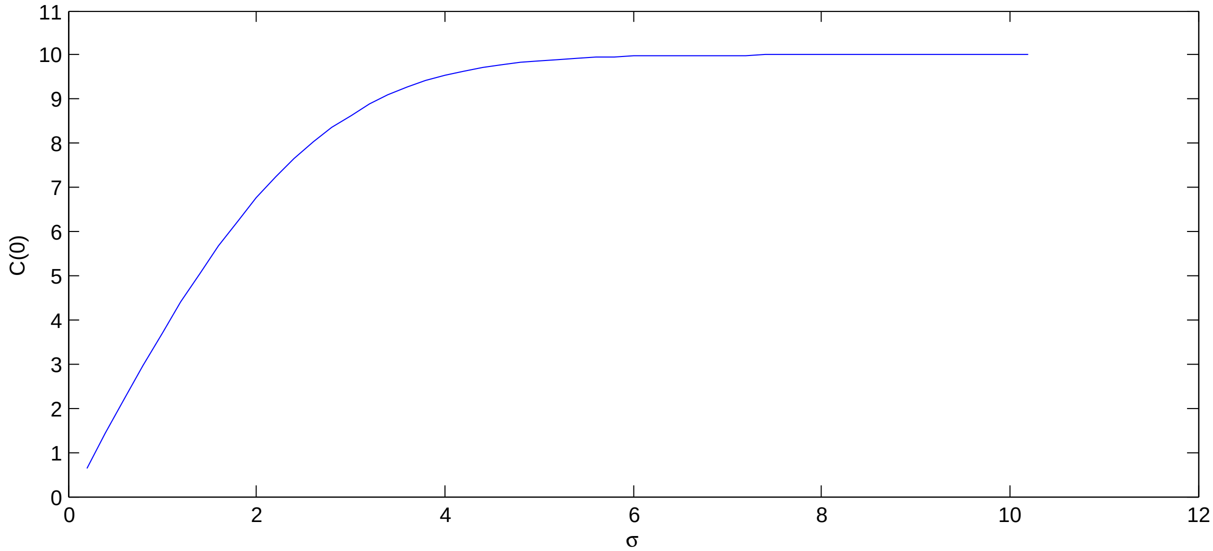


Figure 3.3: Initial price $C(0)$ of the call computed for increasing values of $\sigma > 0$ ($S_0 = 10, K = 10.5, T = 1, \alpha = 0, N = 1000, p = 1/2$). The call price increases with the volatility and approaches the stock price $S_0 = 10$ for σ large (note however that only values $0 < \sigma \lesssim 2$ are realistic).

price of an option never exceeds the price of the underlying stock, hence the option market price is always within the co-domain of the function in Fig. 3.3. It follows that σ_{imp} is a well-defined unique quantity.

Exercise 3.8. *Compute numerically the implied volatility of the options analysed in Exercise 3.7. Compare your value of σ_{imp} with the one quoted in the market. Plot the implied volatility as a function of the strike price and discuss your findings.*

We shall discuss again the important concept of implied volatility in Chapter 6.

Chapter 4

American derivatives

This chapter is concerned with American derivatives on a stock. We assume that the stock price follows the binomial model

$$S(t) = \begin{cases} S(t-1)e^u & \text{with probability } p, \\ S(t-1)e^d & \text{with probability } 1-p. \end{cases}, \quad t \in \mathcal{I} = \{1, \dots, N\},$$

where $0 < p < 1$, $u > d$. It is assumed that $S(0) = S_0$ is known. Moreover we assume that the interest rate r of the risk-free asset, so that its value at time t is

$$B(t) = B_0 e^{rt}.$$

We impose $d < r < u$. In particular, the binomial market is arbitrage-free and the risk-neutral probability is well defined:

$$q_u = \frac{e^r - e^d}{e^u - e^d}, \quad q_d = \frac{e^u - e^r}{e^u - e^d}, \qquad q_u, q_d \in (0,1), \; q_u + q_d = 1.$$

We denote by $\{(h_S(t), h_B(t)\}_{t \in \mathcal{I}}$ a portfolio process invested in the stock and the risk-free asset, where $(h_S(t), h_B(t))$ is the portfolio position in the interval $(t-1, t]$ and $h_S(0) = h_S(1)$, $h_B(0) = h_B(1)$. The value at time t of the portfolio is $V(t) = h_S(t)S(t) + h_B(t)B(t)$. Recall that a portfolio is said to be self-financing if

$$V(t-1) = h_S(t)S(t-1) + h_B(t)B(t-1), \quad t \in \mathcal{I},$$

which means that no cash is ever withdrawn or added to the portfolio. Recall also that the portfolio is called predictable if there exist N functions $H_1, \dots H_N$ such that $H_t : (0, \infty)^t \to \mathbb{R}^2$ and

$$(h_S(t), h_B(t)) = H_t(S(0), \dots, S(t-1)).$$

Another way to say this is that the portfolio position at time t is a deterministic function of the price up to time $t-1$. In particular the portfolio position in the interval $(t-1, t]$ is determined by the information available up to and included the time $t-1$.

We shall need the proven recurrence formula for the fair price $\Pi_Y(t)$ of European derivatives. Namely, denoting $\Pi_Y^u(t)$ the value of the European derivative at time t assuming that the stock price goes up at time t (i.e., $S(t) = S(t-1)e^u$, or equivalently, $x_t = u$), and similarly for $\Pi_Y^d(t)$, with "up" replaced by "down", we have seen in Chapter 3, Theorem 3.1, that $\Pi_Y(t)$ satisfies

$$\Pi_Y(N) = Y, \quad \text{and} \quad \Pi_Y(t) = e^{-r}[q_u \Pi_Y^u(t+1) + q_d \Pi_Y^d(t+1)], \quad \text{for } t \in \{0, N-1\}. \quad (4.1)$$

4.1 The binomial price of American derivatives

In contrast to European derivatives, American derivatives can be exercised at any time prior or including the expiration date. Let $Y(t)$ be the pay-off of an American derivative exercised at time t. We assume that $t \in \mathcal{I} = \{1, \ldots, N\}$, while $t = 0$ is the present time. Let $S(t)$ be the binomial stock price of the underlying stock at time t. We restrict ourselves to **standard American derivatives**, which means that

$$Y(t) = g(S(t)), \quad t \in \mathcal{I},$$

for some function $g : (0, \infty) \to [0, \infty)$, which is called pay-off function of the derivative*. For example, $g(z) = (z - K)_+$ for American call options and $g(z) = (K - z)_+$ for American put options, where $(z)_+ = \max(0, z)$ and K is the strike price of the option. $Y(t)$ is also called the **intrinsic value** of the derivative. If $Y(t)$ is zero the derivative is out of the money at time t.

Note that $Y(t)$ has been defined only for $t = 1, \ldots, N$ and not for $t = 0$. This of course makes sense, as the American derivative can be exercised only from time $t = 1$ onwards, hence it is meaningless to talk about "pay-off at time t=0". However, it is convenient to define

$$Y_0 = Y(0) := g(S(0)),$$

so that $Y(t)$ is defined for all $t \in \{0, \ldots, N\}$. Moreover $Y(t)$ is clearly path-dependent, i.e., $Y(t) = Y(t, x)$. As usual, $Y^u(t)$ denotes the intrinsic value at time t assuming that the stock price goes up at time t and similarly for $Y^d(t)$, with "up" replaced by "down".

For a European derivative with pay-off Y at the expiration date $T = N$, hedging portfolios processes have been defined by imposing that their value $V(t)$ satisfies $V(N) = Y$, see Definition 3.2 in Chapter 3. For American derivatives we define hedging portfolios as follow:

Definition 4.1. *A portfolio process $\{h_S(t), h_B(t)\}_{t \in \mathcal{I}}$ is said to be hedging an American derivative with intrinsic value $Y(t)$ if*

$$V(N) = Y(N), \quad V(t) \geq Y(t), \ t = 0, \ldots, N-1,$$

where $V(t) = h_S(t)S(t) + h_B(t)B(t)$ is the value of the portfolio process at time t.

*For **non-standard American derivatives**, the pay-off has the form $Y(t) = g_t(S(1), \ldots, S(t))$, where $g_t : (0, \infty)^t \to [0, \infty)$.

To justify this definition, recall that the goal of an hedging portfolio is to secure the writer position (i.e., the short position). Since the buyer of the American derivative has the right to exercise the derivative at any time $t \in \mathcal{I}$, then we need to ensure that the value of the writer portfolio is, at any time, sufficient to pay-off the buyer, i.e., $V(t) \geq Y(t)$.

In the American case we do *not* expect that an hedging portfolio could be self-financing. In fact, if $\{h_S(t), h_B(t)\}_{t\in\mathcal{I}}$ were self-financing, the condition $V(N) = Y(N)$ in Definition 4.1 would force $V(t)$ to be given by (2.7). However, this portfolio value need not be larger or equal to $Y(t)$, as required in Definition 4.1, which means that self-financing portfolios are in general not enough valued to pay-off an American derivative.

As in the case of European derivatives, we want to use hedging portfolio processes to justify a definition of binomial fair price for American derivatives. Consider an American derivative with intrinsic value $Y(t)$ and maturity time N, and a European derivative with pay-off $Y(N)$ at the same expiration date N. By definition of hedging portfolios in the two cases, we are led to impose $\widehat{\Pi}_Y(N) = \Pi_Y(N) = Y(N)$, where $\widehat{\Pi}_Y(t)$ denotes the binomial price of the American derivative and $\Pi_Y(t)$ the binomial price of the European counterpart. Consider now the time $t = N-1$. Since American derivatives are in general more valuable than Europeans prior to expiration, then a reasonable fair price for the American derivative at time $t = N-1$ is

$$\widehat{\Pi}_Y(N-1) = \max(Y(N-1), \Pi_Y(N-1)). \tag{4.2}$$

Using the recurrence formula (3.7), we have

$$\Pi_Y(N-1) = e^{-r}(q_u\Pi_Y^u(N) + q_d\Pi_Y^d(N)) = e^{-r}(q_u\widehat{\Pi}_Y^u(N) + q_d\widehat{\Pi}_Y^d(N)),$$

where for the second equality we use that $\widehat{\Pi}_Y(N) = \Pi_Y(N)$. Hence (4.2) becomes

$$\widehat{\Pi}_Y(N-1) = \max[Y(N-1), e^{-r}(q_u\widehat{\Pi}_Y^u(N) + q_d\widehat{\Pi}_Y^d(N))].$$

This suggests to introduce the following definition:

Definition 4.2. *The binomial (fair) price $\widehat{\Pi}_Y(t)$ of a standard American derivative with pay-off $Y(t) = g(S(t))$ at time $t \in \{0, 1, \ldots, N\}$ is defined by the recurrence formula*

$$\widehat{\Pi}_Y(N) = Y(N) \tag{4.3}$$

$$\widehat{\Pi}_Y(t) = \max(Y(t), e^{-r}(q_u\widehat{\Pi}_Y^u(t+1)) + q_d\widehat{\Pi}_Y^d(t+1)), \quad t \in \{0, \ldots, N-1\}, \tag{4.4}$$

Exercise 4.1 (?)**.** *Suppose an investor buys an American derivative at time $N-1$. Since the derivative can only be exercised at time $t = N$, why is the price of the American derivative not the same as the corresponding European derivative? Namely, why $\widehat{\Pi}_Y(N-1) = \max(Y(N-1), \Pi_Y(N-1))$ and not $\widehat{\Pi}_Y(N-1) = \Pi_Y(N-1)$?*

Exercise 4.2 (?)**.** *Why did we not define the binomial fair price of the American derivative as $\widehat{\Pi}_Y(t) = \max(Y(t), \Pi_Y(t))$?*

Exercise 4.3. *Show that the binomial price of a standard American derivative at time t is a determininistic function of $S(t)$. Moreover show that the binomial price of a standard American derivative is always greater than or equal to the binomial price of the corresponding European derivative. Finally show that an American call and a European call with the same strike and maturity have the same binomial price.*

The last claim in the previous exercise is consistent with what we have seen in Chapter 1, namely that it is never optimal to exercise an American call option prior to expiration when the underlying asset does not pay dividends. The example treated in Section 4.2 below shows that this is not true for put options. Some properties of American options when the underlying pays dividends are given in [3, Ch. 7].

4.2 Example: American put options

In this section we consider an example of American put option. We let the strike price $K = 3/4$, and so

$$Y(t) = (\frac{3}{4} - S(t))_+.$$

We present an example in $N = 2$ periods. Consider a market with the following parameters:

$$e^u = \frac{7}{4}, \quad e^d = \frac{1}{2}, \quad e^r = \frac{9}{8},$$

so that $q_u = q_d = 1/2$. Assuming $S_0 = 1$, the binomial tree for the stock price is

$$\begin{tikzcd}
 & & S(2) = \frac{49}{16} \\
 & S(1) = \frac{7}{4} \arrow[ur, "u"] \arrow[dr, "d"] & \\
S_0 = 1 \arrow[ur, "u"] \arrow[dr, "d"] & & S(2) = \frac{7}{8} \\
 & \boxed{S(1) = \frac{1}{2}} \arrow[ur, "u"] \arrow[dr, "d"] & \\
 & & \boxed{S(2) = \frac{1}{4}}
\end{tikzcd}$$

When the price of the stock in the paths above is within a box, the put option is in the money. In fact, the binomial tree for the intrinsic value $Y(t)$ of the American put is

$$\begin{tikzcd}
 & & Y(2)=0 \\
 & Y(1)=0 \arrow[ur, "u"] \arrow[dr, "d"] & \\
Y_0=0 \arrow[ur, "u"] \arrow[dr, "d"] & & Y(2)=0 \\
 & Y(1)=\frac{1}{4} \arrow[ur, "u"] \arrow[dr, "d"] & \\
 & & Y(2)=\frac{1}{2}
\end{tikzcd}$$

Let us first compute the price of the corresponding European put with pay off

$$\left(\frac{3}{4}-S(2)\right)_+ = Y(2) = \Pi_Y(2)$$

at time of maturity 2. Using the recurrence formula (4.1) we have

$$\Pi_Y(1) = \frac{4}{9}[\Pi_Y^u(2)+\Pi_Y^d(2)] = \frac{4}{9}\left[\left(\frac{3}{4}-\frac{7}{4}S(1)\right)_+ + \left(\frac{3}{4}-\frac{1}{2}S(1)\right)_+\right].$$

Hence

$$\Pi_Y^u(1) = \frac{4}{9}\left[\left(\frac{3}{4}-\frac{49}{16}\right)_+ + \left(\frac{3}{4}-\frac{7}{8}\right)_+\right] = 0,$$

$$\Pi_Y^d(1) = \frac{4}{9}\left[\left(\frac{3}{4}-\frac{7}{8}\right)_+ + \left(\frac{3}{4}-\frac{1}{4}\right)_+\right] = \frac{2}{9}.$$

Therefore, again by (4.1), we have

$$\Pi_Y(0) = \frac{4}{9}[\Pi_Y^u(1)+\Pi_Y^d(1)] = \frac{8}{81}.$$

In conclusion, we have obtained the following paths for the binomial price of the European derivative

$$
\begin{tikzcd}
 & & \Pi_Y(2)=0 \\
 & \Pi_Y(1)=0 \arrow[ur, "u"] \arrow[dr, "d"] & \\
\Pi_Y(0)=\frac{8}{81} \arrow[ur, "u"] \arrow[dr, "d"] & & \Pi_Y(2)=0 \\
 & \Pi_Y(1)=\frac{2}{9} \arrow[ur, "u"] \arrow[dr, "d"] & \\
 & & \Pi_Y(2)=\frac{1}{2}
\end{tikzcd}
$$

Exercise 4.4. *Compute the self-financing hedging portfolio process for this European put. Can you guess whether $h_S(0)$ will be positive or negative before computing the portfolio?*

Now we compute the prices of the American put option. Of course, at time of maturity it is the same as its European counterpart (by definition of hedging portfolio). At time $t = 1$ we have, by (4.4),

$$
\begin{aligned}
\widehat{\Pi}_Y(1) &= \max\left[Y(1), \frac{4}{9}(\widehat{\Pi}_Y^u(2) + \widehat{\Pi}_Y^d(2))\right] \\
&= \max\left[Y(1), \frac{4}{9}\left(\left(\frac{3}{4} - \frac{7}{4}S(1)\right)_+ + \left(\frac{3}{4} - \frac{1}{2}S(1)\right)_+\right)\right].
\end{aligned}
$$

Since

$$
Y^u(1) = \left(\frac{3}{4} - \frac{7}{4}\right)_+ = 0, \quad Y^d(1) = \left(\frac{3}{4} - \frac{1}{2}\right)_+ = \frac{1}{4},
$$

we find

$$
\widehat{\Pi}_Y^u(1) = \max[0, 0] = 0, \quad \widehat{\Pi}_Y^d(1) = \max\left[\frac{1}{4}, \frac{2}{9}\right] = \frac{1}{4}
$$

and so

$$
\widehat{\Pi}_Y(0) = \max\left[Y(0), \frac{4}{9}(\widehat{\Pi}_Y^u(1) + \widehat{\Pi}_Y^d(1))\right] = \frac{1}{9}.
$$

Hence the binomial price of the American put corresponding to the different paths of the stock price is as follows:

$$
\begin{tikzcd}
 & & \widehat{\Pi}_Y(2) = 0 \\
 & \widehat{\Pi}_Y(1) = 0 \arrow[ur, "u"] \arrow[dr, "d"] & \\
\boxed{\widehat{\Pi}_Y(0) = \frac{1}{9}} \arrow[ur, "u"] \arrow[dr, "d"] & & \widehat{\Pi}_Y(2) = 0 \\
 & \boxed{\widehat{\Pi}_Y(1) = \frac{1}{4}} \arrow[ur, "u"] \arrow[dr, "d"] & \\
 & & \widehat{\Pi}_Y(2) = \frac{1}{2}
\end{tikzcd}
$$

Note that the binomial price of the American put and of the European put are different in two instances, which are indicated in the paths above by putting the price of the American put within a box. In particular, their initial price is different. When the prices are different, the American put is more expensive than the European put.

Exercise 4.5 (?). *Would you exercise the option at time $t = 1$?*

4.3 Replicating portfolio processes of American derivatives

In this section we describe how to obtain replicating portfolio processes for American derivatives.

Definition 4.3. *A* **replicating portfolio process** *for an American derivative with intrinsic value $Y(t)$ is a portfolio process that satisfies $V(t) = \widehat{\Pi}_Y(t)$, for all $t \in \{0, \dots, N\}$ (and for all possible paths of the stock price).*

Note that replicating portfolio processes are hedging portfolios, because $\widehat{\Pi}_Y(t) \geq Y(t)$. Moreover, in the European case any self-financing hedging portfolio is (trivially) replicating, because $\Pi_Y(t)$ has been defined as the common value of any such portfolio. However, in the American case, the value at time t of an hedging portfolio process could be strictly greater than the binomial price $\widehat{\Pi}_Y(t)$ of the American derivative defined in Definition 4.2. If this happens, then the writer should withdraw cash from the portfolio in order to replicate the value of the American derivative. This leads us to introduce the important concept of portfolio processes generating a cash flow. We argue as we did for self-financing portfolios. Recall that $(h_S(t), h_B(t))$ is the investor position on the stock and the risk-free asset during the time interval $(t-1, t]$. Let $V(t) = h_S(t)S(t) + h_B(t)B(t)$ be the value of this portfolio. At the time t, the investor sells/buys shares of the two assets. Let $(h_S(t+1), h_B(t+1))$ be the new position on the stock and the risk-free asset in the interval $(t, t+1]$. Then the value of the new portfolio at time t is given by $V'(t) = h_S(t+1)S(t) + h_B(t+1)B(t)$. The cash flow $C(t)$ is defined as $V'(t) - V(t) = -C(t)$ and corresponds to cash withdrawn (if $C(t) > 0$) or added (if $C(t) < 0$) to the portfolio as a result of the change in the position on the assets (for a self-financing portfolio we have of course $C(t) = 0$, for all $t \in \{0, \dots, N\}$). This leads to the following definition.

Definition 4.4. *A portfolio process* $\{(h_S(t), h_B(t))\}_{t\in\mathcal{I}}$ *is said to generate the* **cash flow** $C(t-1)$*,* $t \in \mathcal{I}$*, if*

$$h_S(t)S(t-1) + h_B(t)B(t-1) = h_S(t-1)S(t-1) + h_B(t-1)B(t-1) - C(t-1), \quad t \in \mathcal{I},$$

or, equivalently,

$$V(t) - V(t-1) = h_S(t)(S(t) - S(t-1)) + h_B(t)(B(t) - B(t-1)) - C(t-1).$$

In particular, if $C(t-1) > 0$*, then the cash is withdrawn from the portfolio, causing a decrease of its value (and vice versa).*

Remark 4.1. As we assume $h_S(0) = h_S(1)$ and $h_B(0) = h_B(1)$, then $C(0) = 0$. Therefore the first time at which the investor can add/remove cash from the portfolio is after changing the position (instantaneously) at time $t = 1$, i.e., when passing from $(h_S(1), h_B(1))$ to $(h_S(2), h_B(2))$, generating the cash flow $C(1)$.

Example: Consider a constant portfolio process that consists of only one share of the stock, that is $h_S(t) \equiv 1$ and $h_B(t) \equiv 0$. The value at time t of this portfolio is $V(t) = h_S(t)S(t) = S(t)$. Suppose that in the interval of time $(t-1, t)$ the stock pays a dividend of 1%. This means that the fraction $S(t-1)/100$ is deposited into the account of the investor, while the price of the stock decreases of the same amount. Hence the value of the portfolio at time t is

$$V(t) = S(t) - S(t-1)/100.$$

Therefore

$$V(t) - V(t-1) = S(t) - S(t-1)/100 - S(t-1) = h_S(t)(S(t) - S(t-1)) - C(t-1),$$

where $h_S(t) = 1$ and $C(t) = S(t-1)/100$.

We now show how to build a predictable replicating portfolio for American derivatives.

Theorem 4.1. *Consider a standard American derivative with intrinsic value* $Y(t)$ *and let* $\widehat{\Pi}_Y(t)$ *be its binomial fair price. Define the portfolio process* $\{(\widehat{h}_S(t), \widehat{h}_B(t))\}_{t\in\mathcal{I}}$ *and the cash flow process* $C(t)$ *recursively as follows:*

$$C(0) = 0, \quad C(t-1) = \widehat{\Pi}_Y(t-1) - e^{-r}[q_u\widehat{\Pi}_Y^u(t) + q_d\widehat{\Pi}_Y^d(t)], \quad t \in \{2, \dots N\}, \qquad (4.5)$$

$$\widehat{h}_S(1) = \widehat{h}_S(0), \quad \widehat{h}_B(0) = \widehat{h}_B(1) \tag{4.6}$$

and, for $t = 1, \dots, N$,

$$\widehat{h}_S(t) = \frac{1}{S(t-1)} \frac{\widehat{\Pi}_Y^u(t) - \widehat{\Pi}_Y^d(t)}{e^u - e^d}, \tag{4.7a}$$

$$\widehat{h}_B(t) = \frac{e^{-r}}{B(t-1)} \frac{e^u \widehat{\Pi}_Y^d(t) - e^d \widehat{\Pi}_Y^u(t)}{e^u - e^d}. \tag{4.7b}$$

Then the value of this portfolio process satisfies

$$V(t) = \widehat{\Pi}_Y(t), \quad \textit{for all } t \in \{0, \dots N\} \tag{4.8}$$

and

$$V(t-1) = \widehat{h}_S(t)S(t-1) + \widehat{h}_B(t)B(t-1) + C(t-1), \quad \textit{for all } t \in \mathcal{I}. \tag{4.9}$$

Proof. The proof is straightforward: just replace (4.5), (4.6) and (4.7) into (4.8)-(4.9). For instance, assuming that the price goes up at time t, we compute

$$\begin{aligned}
V^u(t) &= \widehat{h}_S(t)S(t-1)e^u + \widehat{h}_B(t)B(t-1)e^r \\
&= \left(\frac{\widehat{\Pi}_Y^u(t) - \widehat{\Pi}_Y^d(t)}{e^u - e^d} \right) e^u + \left(\frac{e^u \widehat{\Pi}_Y^d(t) - e^d \widehat{\Pi}_Y^u(t)}{e^u - e^d} \right) \\
&= \Pi_Y^u(t),
\end{aligned}$$

and at the same fashion one proves that $V^d(t) = \Pi_Y^d(t)$. Hence (4.8) holds. In a similar fashion, replacing (4.5), (4.6) and (4.7) into the right hand side of (4.9) we find that the latter is equal to $\widehat{\Pi}_Y(t-1)$, which we already proved to be equal to $V(t-1)$. Hence (4.9) holds as well. □

The previous theorem is telling us that the writer can hedge the derivative and still be able to withdraw cash from the portfolio, which the writer can use to invest in other assets. Hence American derivatives could be very useful for active investors. Note that whether the writer is allowed or not to withdraw a positive amount of cash from the portfolio ($C(t) > 0$) depends on the "smartness" of the buyer. In fact, using (4.4) in (4.5), we have, for $t \in \{2, \dots, N\}$,

$$C(t-1) = \max(Y(t-1), e^{-r}(q_u \widehat{\Pi}_Y^u(t)) + q_d \widehat{\Pi}_Y^d(t)) - e^{-r}(q_u \widehat{\Pi}_Y^u(t)) + q_d \widehat{\Pi}_Y^d(t)).$$

This quantity is positive, i.e., the investor can withdraw cash from the portfolio, if the buyer does not exercise the derivative at the optimal time, which is when the value of the derivative is equal to its intrinsic value. For instance, computing the cash flow at $t = 2$ for the American put considered in the previous section, we find

$$C(1) = \Pi_Y(1) - \frac{4}{9}\left(\left(\frac{3}{4} - \frac{7}{4}S(1) \right)_+ + \left(\frac{3}{4} - \frac{1}{2}S(1) \right)_+ \right).$$

If the price goes up at time 1 this gives $C^u(1) = 0$; if the price goes down at time 1 we obtain

$$C^d(1) = \frac{1}{4} - \frac{4}{9}\left(\left(\frac{3}{4} - \frac{7}{8}\right)_+ + \left(\frac{3}{4} - \frac{1}{4}\right)_+\right) = \frac{1}{36}.$$

Hence if at time 1 the price of the stock goes down and the buyer does not exercise the American put at time 1, then the writer can withdraw the amount $\frac{1}{36}$ from the portfolio.

We can look at this also from the buyer perspective. We shall say that it is **optimal** for the buyer to exercise an American derivative at some given time $t \in 1, \dots N-1$ if $\widehat{\Pi}_Y(t) = Y(t)$. In particular, if the buyer exercises the derivative optimally, then the seller is not allowed to withdraw cash from the portfolio at that time. The reason why it is optimal for the buyer to exercise the derivative when $\widehat{\Pi}_Y(t) = Y(t)$ is that in this way the buyer is taking full advantage of the derivative: the pay-off is exactly equal to its value. If, on the other hand, $\widehat{\Pi}_Y(t) > Y(t)$, then the optimal strategy is not to exercise the derivative but, rather, to sell a derivative with the same parameters. In fact, by writing this derivative, the investor income will be higher than the profit from exercising the derivative ($\widehat{\Pi}_Y(t) > Y(t)$), and moreover the investment is risk-less, because the investor has a short and a long position on the same asset.

Exercise 4.6. *Compute the replicating portfolio for the American put given in Section 4.2.*

Exercise 4.7. *Consider an American derivative with intrinsic value*

$$Y(t) = \min(S(t), (24 - S(t))_+)$$

and expiring at time $T = 3$. The initial price of the underlying stock is $S(0) = 27$, while at future times it follows the binomial model

$$S(t+1) = \begin{cases} 4S(t)/3 & \textit{with probability } 1/2 \\ 2S(t)/3 & \textit{with probability } 1/2 \end{cases}$$

for $t = 0, 1, 2$. Assume also that the interest rate of the money market is zero. Compute the possible paths of the fair value of the derivative. In which case it is optimal for the buyer to exercise the derivative prior to expiration? What is the amount of cash that the seller can withdraw from the portfolio if the buyer does not exercise the derivative optimally?

4.4 Computation of the fair price of American derivatives with Matlab

We work under the same set-up as in Section 3.3. Namely we consider a partition $0 = t_1 < t_2 < \cdots < t_{N+1} = T$ of the interval $[0, T]$ and the binomial stock price

$$S(i+1) = \begin{cases} S(i)e^u, & \text{with probability } p \\ S(i)e^d, & \text{with probability } 1-p \end{cases}, \quad i \in \mathcal{I} = \{1, \dots, N\},$$

where $S(i) = S(t_i)$ and

$$u = \alpha h + \sigma\sqrt{\frac{1-p}{p}}\sqrt{h}, \quad d = \alpha h - \sigma\sqrt{\frac{p}{1-p}}\sqrt{h}, \tag{4.10}$$

The value of the risk-free asset at time t_i is given by $B(t_i) = B_0 e^{rt_i} = B_0 e^{(rh)i}$. Hence the pair $(S(i), B(i))$ defines a 1+1 dimensional binomial market with parameters u, d given by (4.10) and interest rate rh. The definition (4.2) of binomial price of American derivative becomes

$$\widehat{\Pi}_Y(N+1) = Y(N+1) \quad \widehat{\Pi}_Y(i) = \max(Y(i), e^{-rh}(q_u\widehat{\Pi}_Y^u(i+1)) + q_d\widehat{\Pi}_Y^d(i+1)), \quad t \in \mathcal{I},$$

where $\Pi_Y(i) = \Pi_Y(t_i)$ and

$$q_u = \frac{e^{rh} - e^{\alpha h - \sigma\sqrt{\frac{p}{1-p}}\sqrt{h}}}{e^{\alpha h + \sigma\sqrt{\frac{1-p}{p}}\sqrt{h}} - e^{\alpha h - \sigma\sqrt{\frac{p}{1-p}}\sqrt{h}}}, \quad q_d = 1 - q_u.$$

Moreover $Y(i) = g(S(i))$, $i = 1, \dots N+1$, is the intrinsic value of the American derivative. The following code defines a function *BinomialAmerican* which computes the binomial price and the cash flow of a standard American derivative with pay-off function g.

```
function [P,C]=BinomialAmerican(Q,S,r,g)
h=Q(2)-Q(1);
syms x;
f = sym(g);
N=length(Q)-1;
expu=S(1,2)/S(1,1);
expd=S(2,2)/S(1,1);
qu=(exp(r)-expd)/(expu-expd);
qd=(expu-exp(r))/(expu-expd);
if (qu<0 || qd<0)
display('Error:  the market is not arbitrage free.');
P=0;
return
end
P=zeros(N+1);
P(:,N+1)=double(subs(f,x,S(:,N+1)));
C(:,N+1)=0;
Y=double(subs(f,x,S));
for j=N:-1:1
for i=1:j
P(i,j)=max(Y(i,j),exp(-r*h)*(qu*P(i,j+1)+qd*P(i+1,j+1)));
C(i,j)=P(i,j)-exp(-r*h)*(qu*P(i,j+1)+qd*P(i+1,j+1));
end
```

```
end
```

The binomial price of the derivative is stored in the upper-triangular $(N+1)\times(N+1)$ matrix P, while the cash flow is stored in the upper-triangular $(N+1)\times(N+1)$ matrix C. We use the convention that at time of maturity the cash flow is equal to zero.

For example, using as inputs the binomial stock price (2.23), the interest rate $r = 0.01$ and the pay-off function of a put with strike 10, i.e., $g(x) = (10-x)+$, we obtain the output

$$P = \begin{matrix} 0.0208 & 0.0089 & 0.0022 & 0 & 0 & 0 \\ 0 & 0.0326 & 0.0156 & 0.0044 & 0 & 0 \\ 0 & 0 & 0.0495 & 0.0268 & 0.0087 & 0 \\ 0 & 0 & 0 & 0.0722 & 0.0449 & 0.0175 \\ 0 & 0 & 0 & 0 & 0.0961 & 0.0689 \\ 0 & 0 & 0 & 0 & 0 & 0.1200 \end{matrix}$$

$$C = \begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0017 & 0.0017 & 0 \\ 0 & 0 & 0 & 0 & 0.0017 & 0 \end{matrix}$$

Remark 4.2. Using the pay-off of the call, $g(x) = (x-10)_+$, we obtain that the price of the American call is exactly the same as the corresponding European call, see (3.14), while $C \equiv 0$. This of course is consistent with the proven fact that, in the absence of dividends, it is never optimal to exercise an American call prior to expire.

Exercise 4.8. *Write a function that sets to* 0 *each element in the matrix P for which it is not optimal to exercise the derivative. Experiment with American call options to verify that is never optimal to exercise the derivative prior to expiration.*

Remark 4.3. Letting O be the matrix generated by the function constructed in the previous exercise, the positive elements in the column j of O are the prices of the derivative for which it is optimal to exercise at time t_j.

Exercise 4.9. *Plot the pairs $(t_i, S(t_i))$ in the $t - S(t)$ plane, for each optimal exercise time t_i of an American put. Experiment for different values of K, T. Use a large number of steps ($N \approx 100$). What happens when $T \to +\infty$?*

In the time-continuum limit, $N \to +\infty$, $h \to 0$ such that $Nh = T$, the points $(t_i, S(t_i))$ for each optimal exercise time lie on a curve in the $t - S(t)$ plane which is called the **optimal exercise curve** (or **optimal exercise boundary**).

Exercise 4.10. *Show numerically that the initial binomial price of an American put with strike K and maturity T converges, as $T \to +\infty$, to the value $v(S_0)$, where v is the function*

$$v(x) = \begin{cases} K - x & 0 \le x \le L \\ (K-L)\left(\frac{x}{L}\right)^{-\frac{2r}{\sigma^2}} & x > L \end{cases}$$

and

$$L = \frac{2r}{2r + \sigma^2} K.$$

How is this result related to your findings in Exercise 4.9?

Chapter 5

Introduction to Probability Theory

The purpose of this chapter is threefold: (1) Introducing some basic concept in probability theory, (2) re-formulate the binomial options pricing model in the language of probability theory and (3) derive the geometric Brownian motion as the time-continuum limit of the binomial stock price.

5.1 Finite Probability Spaces

Let Ω be a set containing a finite number of elements $\omega_1, \omega_2, \ldots, \omega_M$. We denote Ω as

$$\Omega = \{\omega_1, \ldots, \omega_M\}, \quad \text{or} \quad \Omega = \{\omega_i\}_{i=1,\ldots,M} \tag{5.1}$$

and call it a **sample space**. The elements $\omega_i \in \Omega$, $i = 1, \ldots, M$, are called **sample points**. The sample points identify the possible outcomes of an experiment. For instance, if the experiment consists in "throwing a die", then

$$\Omega = \{1, 2, 3, 4, 5, 6\} \quad (M = 6),$$

while for the experiment "tossing a coin once", we have

$$\Omega = \Omega_1 := \{H, T\} \quad (M = 2),$$

where H stands for "Head" and T for "Tail"; the subscript 1 in Ω_1 indicates that the coin is tossed only once. In the experiment "tossing a coin twice" we have

$$\Omega = \Omega_2 := \{(H, H), (H, T), (T, H), (T, T)\} = \Omega_1 \times \Omega_1 \quad (M = 2^2 = 4)$$

and in the experiment "tossing a coin N times" we have

$$\begin{aligned} \Omega = \Omega_N &:= \{\omega = (\gamma_1, \gamma_2, \ldots, \gamma_N);\ \gamma_j = H \text{ or } T,\ j = 1, \ldots, N\} \\ &= \underbrace{\Omega_1 \times \Omega_1 \times \cdots \times \Omega_1}_{N \text{ times}} = \{H, T\}^N \quad (M = 2^N). \end{aligned}$$

We denote by 2^Ω the **power set** of Ω, i.e., the set of all subsets of Ω. It consists of the empty set $\emptyset$, the subsets containing one element, i.e., $\{\omega_1\}, \{\omega_2\}, \dots, \{\omega_M\}$, which are called **atomic sets**, the subsets containing two elements, i.e.,

$$\{\omega_1, \omega_2\}, \dots, \{\omega_1, \omega_M\}, \{\omega_2, \omega_3\}, \dots, \{\omega_2, \omega_M\}, \dots, \{\omega_{M-1}, \omega_M\},$$

the subsets containing 3 elements and so on, and the set $\Omega = \{\omega_1, \dots, \omega_M\}$ itself. Thus 2^Ω contains 2^M elements. For instance

$$2^{\Omega_1} = \{\emptyset, \{H\}, \{T\}, \{H, T\} = \Omega_1\}.$$

Exercise 5.1. *Write down* 2^{Ω_2}.

The elements of 2^Ω (i.e., the subsets of Ω) are called **events**. They identify possible events that occur in the experiment. For example

$$\{2, 4, 6\} \equiv [\text{ the result of throwing a die is an even number}],$$

$$\{(H, H), (T, T)\} \equiv [\text{tossing a coin twice gives the same outcome in both tosses}].$$

Exercise 5.2. *Write down the following events* $A, B, C \in 2^{\Omega_4}$*:*

$$A = [\textit{number heads=number tails}], \quad B = [\textit{successive tosses are different}],$$

$$C = [\textit{there exist at least three identical tosses}].$$

If $A, B \in 2^\Omega$ are events, then $A \cup B$ is the event that A *or* B happens, while $A \cap B$ is the event that both A *and* B happen. If the sets $A, B \subset \Omega$ are **disjoint**, i.e., $A \cap B = \emptyset$, the events A and B cannot occur simultaneously. For instance, $\{1, 3, 5\} \cap \{2, 4, 6\} = \emptyset$ means that the outcome of a die roll cannot be both even and odd.

The atomic set $\{\omega_i\}$ identifies the event that the outcome of the experiment is exactly ω_i. We want to assign a probability $\mathbb{P}$ to such special events. To this purpose we introduce M real numbers $p_1, p_2, \dots p_M$ such that

$$0 < p_i < 1, \text{ for all } i = 1, \dots, M, \quad \text{and} \quad \sum_{i=1}^{M} p_i = 1. \tag{5.2}$$

We define p_i to be the probability of the event $\{\omega_i\}$, that is

$$\mathbb{P}(\{\omega_i\}) = p_i, \quad i = 1, \dots, M.$$

Any generic event $A \in 2^\Omega$ can be written as the disjoint union of atomic events, e.g.,

$$\{\omega_1, \omega_3, \omega_6\} = \{\omega_1\} \cup \{\omega_3\} \cup \{\omega_6\}.$$

This leads to define the probability of a generic event $A \in 2^\Omega$ as

$$\mathbb{P}(A) = \sum_{i:\omega_i \in A} \mathbb{P}(\{\omega_i\}) = \sum_{i:\omega_i \in A} p_i. \tag{5.3}$$

For instance $\mathbb{P}(\{\omega_1, \omega_3, \omega_6\}) = p_1 + p_3 + p_6$. We shall also write the definition of $\mathbb{P}(A)$ as

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}).$$

In particular

$$\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = \sum_{i=1}^{M} p_i = 1.$$

We also set

$$\mathbb{P}(\emptyset) = 0, \tag{5.4}$$

which means that it is impossible that the experiment gives no outcome. Clearly $\emptyset$ is the only event with zero probability: any other such event is excluded *a priori* by the sample space. At this point every event has been assigned a probability.

Definition 5.1. *Given* $(p_1, \dots, p_M)$ *satisfying* (5.2) *and a set* $\Omega = \{\omega_1, \dots, \omega_M\}$*, the function* $\mathbb{P} : 2^\Omega \to [0, 1]$ *defined by* (5.3)-(5.4) *is called a* **probability measure***. The pair* $(\Omega, \mathbb{P})$*, is called a* **finite probability space***.*

Examples

- In the experiment "throwing a die", let $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6$. We say that the die is **fair**, as any number between 1 and 6 has the same probability to appear as a result of the roll. Then $\mathbb{P}(\{2, 4, 6\}) = 1/6 + 1/6 + 1/6 = 1/2$, i.e., we have 50% chances to get an even number.

- In Ω_1 we let $\mathbb{P}(\{H\}) = p_H$, $\mathbb{P}(\{T\}) = p_T$, where $p_T = 1 - p_H$, $p_H \in (0, 1)$. The coin is said to be **fair** if $p_H = p_T = 1/2$.

- In Ω_2 we assign a probability to the atomic events by setting

$$\mathbb{P}(\{(H, H)\}) = p_H \cdot p_H = p_H^2,$$

$$\mathbb{P}(\{(T, H)\}) = \mathbb{P}(\{(H, T)\}) = p_T p_H, \quad \mathbb{P}(\{(T, T)\}) = p_T^2.$$

  Notice that $\mathbb{P}(\{(H, H)\}) + \mathbb{P}(\{(T, H)\}) + \mathbb{P}(\{(H, T)\}) + \mathbb{P}(\{(T, T)\}) = (p_H + p_T)^2 = 1$, as it should be. This way of assigning probabilities is equivalent to the assumption that the two tosses are **independent**, i.e., the result of the first toss does not influence the result of the second toss (see Definition 5.4 below). As an example of computing the probability of non-atomic events, consider the event

$$A = [\textit{the outcome of the two tosses is the same}].$$

  Then

$$\mathbb{P}(A) = \mathbb{P}(\{(H, H), (T, T)\}) = \mathbb{P}(\{(H, H)\}) + \mathbb{P}(\{(T, T)\}) = p_H^2 + p_T^2.$$

- In $\Omega_N = \{\omega = (\gamma_1, \gamma_2, \dots, \gamma_N);\ \gamma_j = H \text{ or } T,\ j = 1, \dots, N\}$, we assign a probability to each of the 2^N atomic events by letting

$$\mathbb{P}(\{\omega\}) = (p_H)^{N_H(\omega)}(p_T)^{N_T(\omega)}, \quad \text{for all } \omega \in \Omega_N,$$

  where $N_H(\omega)$ and $N_T(\omega) = N - N_H(\omega)$ are respectively the number of heads and tails in the N-toss corresponding to ω. Again, this definition of probability makes the N-tosses independent. Since for all $k = 0, \dots, N$ the number of N-tosses $\omega \in \Omega_N$ having $N_H(\omega) = k$ is given by the binomial coefficient

$$\binom{N}{k} = \frac{N!}{k!(N-k)!},$$

  then

$$\begin{aligned}\sum_{\omega \in \Omega_N} \mathbb{P}(\{\omega\}) &= \sum_{\omega \in \Omega_N} (p_H)^{N_H(\omega)}(p_T)^{N_T(\omega)} = (p_T)^N \sum_{\omega \in \Omega_N} \left(\frac{p_H}{p_T}\right)^{N_H(\omega)} \\ &= (p_T)^N \sum_{k=0}^{N} \binom{N}{k} \left(\frac{p_H}{p_T}\right)^{k}.\end{aligned}$$

  By the binomial theorem, $(1+a)^N = \sum_{k=0}^{N} \binom{N}{k} a^k$, hence

$$\mathbb{P}(\Omega) = \sum_{\omega \in \Omega_N} \mathbb{P}(\{\omega\}) = (p_T)^N \left(1 + \frac{p_H}{p_T}\right)^N = (p_T + p_H)^N = 1.$$

  The probability of any other event $A \in 2^{\Omega_N}$ is the sum of the probabilities of the atomic events whose (disjoint) union forms the set A.

The last example deserves to be given a separate definition.

Definition 5.2. *Given $0 < p < 1$, the pair $(\Omega_N, \mathbb{P}_p)$ given by $\Omega_N = \{H, T\}^N$ and*

$$\mathbb{P}_p(A) = \sum_{\omega \in A} p^{N_H(\omega)}(1-p)^{N_T(\omega)}, \quad \textit{for all } A \in 2^{\Omega_N}, \tag{5.5}$$

*is called a N-**coin toss probability space**. Here $N_H(\omega)$ is the number of H in the sample ω and $N_T(\omega) = N - N_H(\omega)$ is the number of T.*

It is possible that the occurrence of an event A affects the probability that a second event B occurred. For instance, for a fair coin we have $\mathbb{P}(\{H, H\}) = 1/4$, but if we know that the first toss is a head, then $\mathbb{P}(\{H, H\}) = 0$. This simple remark leads to the definition of **conditional probability**.

Definition 5.3. *Given two events A, B such that $\mathbb{P}(B) > 0$, the conditional probability of A given B is defined as*

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

If the occurrence of B does not affect the occurrence of A, i.e., if $\mathbb{P}(A|B) = \mathbb{P}(A)$, we say that the two events are **independent**. By the previous definition, the independence property is equivalent to the following.

Definition 5.4. *Two events A, B are said to be independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.*

Exercise 5.3. *Compute the probability of the events defined in Exercise 2 and verify that there is no pair of independent events among them. Compute*

$$\mathbb{P}(A|B), \quad \mathbb{P}(B|A), \quad \mathbb{P}(A|C), \quad \mathbb{P}(B|C).$$

Give an example of two independent events defined on Ω_4.

5.2 Random Variables

In general the purpose of an experiment is to determine the value of quantities which depend on the outcome of the experiment (e.g., the velocity of a particle, which is determined by successive measurements of its position). We call such quantities random variables.

Definition 5.5. *Let $(\Omega, \mathbb{P})$ be a finite probability space. A* **random variable** *is a function $X : \Omega \to \mathbb{R}$. If $g : \mathbb{R} \to \mathbb{R}$, then the random variable $Y = g(X)$ is said to be X-measurable.*

Note that the property of Y being X-measurable means that the value attained by Y can be inferred by the value attained by X, i.e., $Y(\omega) = g(X(\omega))$.

Since $\Omega = \{\omega_1, \dots, \omega_M\}$, then X can attain only a finite number of values, which we denote $x_1, \dots, x_M$, namely

$$X(\omega_i) = x_i, \quad i = 1, \dots, M.$$

The values $x_1, \dots, x_M$ need *not* be distinct. If $X(\omega_i) = c$, for all $i = 1, \dots, M$, we say that X is a non-random, or **deterministic**, constant (the value of X is independent of the outcome of the experiment).

The **image** of X is the finite set defined as

$$\mathrm{Im}(X) = \{x \in \mathbb{R} \text{ such that } X(\omega) = x, \text{ for some } \omega \in \Omega\}.$$

Given $a \in \mathbb{R}$, we denote

$$\{X = a\} = \{\omega \in \Omega : X(\omega) = a\},$$

which is the event that X attains the value a. Of course, $\{X = a\} = \emptyset$ if $a \notin \mathrm{Im}(X)$. In general, given $I \subseteq \mathbb{R}$, we denote

$$\{X \in I\} = \{\omega \in \Omega : X(\omega) \in I\},$$

which is the event that the value attained by X lies in the set I. Moreover we denote

$$\{X = a, Y = b\} = \{X = a\} \cap \{Y = b\}, \quad \{X \in I_1, Y \in I_2\} = \{X \in I_1\} \cap \{Y \in I_2\}.$$

The probability that X takes value a is given by

$$\mathbb{P}(X=a)=\mathbb{P}(\{X=a\})=\sum_{i:X(\omega_i)=a} p_i.$$

If $a \notin \mathrm{Im}(X)$, then $\mathbb{P}(X=a)=\mathbb{P}(\emptyset)=0$. More generally, given any open subset I of $\mathbb{R}$, we write

$$\mathbb{P}(X \in I)=\mathbb{P}(\{X \in I\})=\sum_{i:X(\omega_i)\in I} p_i,$$

which is the probability that the value of X belongs to I. For example, in the probability space of a fair die consider the random variable

$$X(\omega)=(-1)^{\omega}, \quad \omega \in \{1,2,3,4,5,6\}. \tag{5.6}$$

Then $X(\omega)=1$ if ω is even and $X(\omega)=-1$ is ω is odd. Moreover

$$\mathbb{P}(X=1)=\mathbb{P}(\{2,4,6\})=1/2, \quad \mathbb{P}(X=-1)=\mathbb{P}(\{1,3,5\})=1/2,$$

whereas

$$\mathbb{P}(X \neq \pm 1)=\mathbb{P}(\emptyset)=0.$$

The set $A=\{2,4,6\}$ is said to be **resolved** by X, because the occurrence of the event A (i.e., the fact that the outcome of the throw is an even number) is equivalent to X taking value 1. In general, given a random variable $X:\Omega \to \mathbb{R}$, the events resolved by X are the sets of the form $\{X \in I\}$, for some $I \subseteq \mathbb{R}$. These events comprise the so called **information carried by** X. The idea is that even if the outcome of an experiment is unknown, measuring the value attained by a random variable gives some information on the result of the experiment.

Definition 5.6. *Given a random variable $X:\Omega\to\mathbb{R}$, the function $f_X:\mathbb{R}\to[0,1]$ defined by*

$$f_X(x)=\mathbb{P}(X=x)$$

is called the **distribution** *of X, while $F_X:\mathbb{R}\to[0,1]$ given by*

$$F_X(x)=\mathbb{P}(X\leq x), \quad x \in \mathbb{R}$$

is called the **cumulative distribution** *of X.*

Note that $f_X(x)$ is non-zero if only if $x \in \mathrm{Im}(X)$, and that F_X is a non-decreasing function. For example, for the random variable (5.6) defined on the probability space of a fair die we have

$$F_X(x)=\begin{cases} 0, & x<-1, \\ 1/2, & x \in [-1,1), \\ 1, & x \geq 1. \end{cases}$$

The probability that a random variable X takes value in the interval $[a,b]$ can be written in terms of the distribution of X as

$$\mathbb{P}(a \leq X \leq b)=\sum_{i:X(\omega_i)=x_i\in[a,b]} \mathbb{P}(X=x_i)=\sum_{i:a\leq x_i\leq b} f_X(x_i). \tag{5.7}$$

In a similar fashion, if $g : \mathbb{R} \to \mathbb{R}$ then

$$\mathbb{P}(a \leq g(X) \leq b) = \sum_{i:g(X(\omega_i))=g(x_i)\in[a,b]} \mathbb{P}(X = x_i) = \sum_{i:a\leq g(x_i)\leq b} f_X(x_i). \tag{5.8}$$

We shall use these formulas later on.

5.2.1 Expectation and Variance

Next we define the expectation and variance of random variables. We may think of the expectation of X as an estimate (or "guess") on the value of X and the variance of X as a measure of how good/bad is this estimate compared to the real value of X.

Definition 5.7. *Given a finite probability space* $(\Omega, \mathbb{P})$*, the* **expectation** *(or* **expected value***) of* $X : \Omega \to \mathbb{R}$ *is defined by*

$$\mathbb{E}[X] = \sum_{i=1}^{M} X(\omega_i)\mathbb{P}(\omega_i).$$

We shall also write the definition of $\mathbb{E}[X]$ as

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\}). \tag{5.9}$$

For instance, in the N-coin toss probability space $(\Omega_N, \mathbb{P}_p)$ we have

$$\mathbb{E}[X] = \sum_{\omega\in\Omega_N} X(\omega)p^{N_H(\omega)}(1-p)^{N_T(\omega)}, \tag{5.10}$$

where $N_H(\omega)$ is the number of heads and $N_T(\omega) = N - N_H(\omega)$ is the number of tails in the N-toss $\omega \in \Omega_N$, see Definition 5.2.

Exercise 5.4. *Let* $X : \Omega_N \to \mathbb{R}$*,* $X(\omega) = N_H(\omega) - N_T(\omega)$*. Compute* $\mathbb{E}[X]$*.*

We can rewrite the definition of expectation as

$$\mathbb{E}[X] = \sum_{x\in\mathrm{Im}(X)} x\,\mathbb{P}(X = x),$$

or equivalently,

$$\mathbb{E}[X] = \sum_{x\in\mathrm{Im}(X)} x f_X(x). \tag{5.11}$$

The importance of (5.11) is that it allows to compute the expectation of X from its distribution, without any reference to the original probability space. For instance, if we are told that a random variable X takes the following values:

$$X = \begin{cases} 1 & \text{with probability } 1/4 \\ 2 & \text{with probability } 1/4 \\ -1 & \text{with probability } 1/2 \end{cases}, \tag{5.12}$$

then we can compute $\mathbb{E}[X]$ using (5.11) as

$$\mathbb{E}[X] = 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} - 1 \cdot \frac{1}{2} = \frac{1}{4}.$$

Exercise 5.5. *Let $Y = g(X)$ be a X-measurable random variable. Show that*

$$\mathbb{E}[g(X)] = \sum_{x \in \mathrm{Im}(X)} g(x) f_X(x). \tag{5.13}$$

Definition 5.8. *The* **variance** *of a random variable $X : \Omega \to \mathbb{R}$ is defined by*

$$\mathrm{Var}[X] = \mathbb{E}[(\mathbb{E}[X] - X)^2].$$

Using the **linearity** of the expectation, i.e., the fact that $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$, for all constants $a, b \in \mathbb{R}$ and random variables X, Y, one can easily prove the formula

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2. \tag{5.14}$$

Moreover $\mathrm{Var}[aX] = a^2\mathrm{Var}[X]$ holds for all constants $a \in \mathbb{R}$. The variance of the sum of two random variables is in general different from the sum of their variances, unless the random variables are independent (see Theorem 5.1 below). Furthermore the variance of a random variable is always non-negative and it is zero if and only if the random variable is a deterministic constant. Hence we may interpret the variance as a measure of the "randomness" of a random variable.

Using (5.13) with $g(x) = x^2$, we can rewrite the definition of variance in terms of the distribution function of X as

$$\mathrm{Var}[X] = \sum_{x \in \mathrm{Im}(X)} x^2 f_X(x) - \left(\sum_{x \in \mathrm{Im}(X)} x f_X(x) \right)^2, \tag{5.15}$$

which allows to compute $\mathrm{Var}[X]$ without any reference to the original probability space. For instance for the random variable (5.12) we find

$$\mathrm{Var}[X] = 1 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} - \left(\frac{1}{4}\right)^2 = \frac{27}{16}.$$

Example: expected log return and volatility of the binomial stock price

Let $0 = t_0 < t_1 < \cdots < t_N = T$ be a partition of the interval $[0, T]$ with $t_i - t_{i-1} = h$, for all $i = 1, \dots N$. Given $u > d$, consider a random variable X such that $X = u$ with probability p and $X = d$ with probability $1 - p$. We may think of X as being defined on $\Omega_1 = \{H, T\}$, with $X(H) = u$ and $X(T) = d$. Now, the binomial stock price at time t_i can be written as $S(t_i) = S(t_{i-1}) \exp(X)$. Hence $S(T)$ is a random variable such that

$$S(T) = S(0)e^{NX} = S(0)e^{\frac{T}{h}X}.$$

It follows that the expectation and the variance of the log-return of the stock in the interval $[0, T]$ are

$$\mathbb{E}[\log S(T) - \log S(0)] = \frac{T}{h}(pu + (1-p)d),$$

$$\mathrm{Var}[\log S(T) - \log S(0)] = \frac{T}{h}[pu^2 + (1-p)d^2 - (pu + (1-p)d)^2)] = \frac{T}{h}p(1-p)(u-d)^2$$

Therefore the parameters α, σ^2 defined in Section 2.4 are given by

$$\alpha = \frac{1}{T}\mathbb{E}[\log S(T) - \log S(0)], \quad \sigma^2 = \frac{1}{T}\mathrm{Var}[\log S(T) - \log S(0)].$$

Hence, assuming that T is expressed in fraction of years, α is the expected annualized log-return of the stock in the interval $[0, T]$, while σ^2 is the annualized variance. The parameter σ itself is called volatility of the stock.

5.2.2 Independence and Correlation

We have seen before that a random variable X carries information. Now, if $Y = g(X)$ for some (non-constant) function $g : \mathbb{R} \to \mathbb{R}$, then Y carries no more information than X: any event resolved by knowing the value of Y is also resolved by knowing the value of X. The other extreme case is when two random variables carry independent information.

Definition 5.9. *Two random variables $X_1, X_2 : \Omega \to \mathbb{R}$ are said to be* **independent** *if the events $\{X_1 \in I_1\}, \{X_2 \in I_2\}$ are independent events, for all regular* sets $I_1, I_2 \subset \mathbb{R}$. This means that*

$$\mathbb{P}(X \in I_1, X_2 \in I_2) = \mathbb{P}(X_1 \in I_1)\mathbb{P}(X_2 \in I_2).$$

More generally, n random variables $X_1, \dots, X_n : \Omega \to \mathbb{R}$ are called independent if

$$\mathbb{P}(X_1 \in I_1, X_2 \in I_2, \dots, X_n \in I_n) = \mathbb{P}(X_1 \in I_1)\mathbb{P}(X_2 \in I_2) \dots \mathbb{P}(X_n \in I_n),$$

for all regular sets $I_1, I_2, \dots I_n \subset \mathbb{R}$.

Exercise 5.6. *Show that when X, Y are independent random variables, then the only events which are resolved by both variables are $\emptyset$ and Ω. Show that two deterministic constants are always independent. Finally assume $Y = g(X)$ and show that in this case the two random variables are independent if and only if Y is a deterministic constant.*

Note that the independence property is connected with the probability defined on the sample space. Thus two random variables may be independent with respect to some probability and not-independent with respect to another. We shall use later the following important result:

*By regular sets we mean intervals or sets which can be written as the union of countably many intervals.

Theorem 5.1. *Let $X_1, X_2, \dots X_n$ be independent random variables, $k \in \{1, \dots, n-1\}$ and $g : \mathbb{R}^k \to \mathbb{R}$, $f : \mathbb{R}^{n-k} \to \mathbb{R}$. Then the random variables*

$$Y = g(X_1, X_2, \dots, X_k), \quad Z = f(X_{k+1}, \cdots, X_n)$$

are independent. Moreover

$$\mathrm{Var}[X_1 + X_2 + \cdots + X_n] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2] + \cdots + \mathrm{Var}[X_n].$$

Exercise 5.7. *Prove the theorem for $n = 2$.*

When two random variables are not independent, we measure their degree of correlation using the concept of covariance.

Definition 5.10. *The* **covariance** *of two random variables $X, Y : \Omega \to \mathbb{R}$ is defined by*

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

If $\mathrm{Cov}(X, Y) = 0$, the two random variables are said to be **uncorrelated**.

Using the linearity of the expectation we can rewrite the definition of covariance as

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \tag{5.16}$$

The interpretation is the following: If $\mathrm{Cov}(X, Y) > 0$, then Y tends to increase (resp. decrease) when X increases (resp. decrease), while if $\mathrm{Cov}(X, Y) < 0$, the two variables have tendency to move in the opposition direction. For instance, assuming $\mathrm{Var}[X] > 0$,

$$\mathrm{Cov}(X, 2X) = 2\mathrm{Var}[X] > 0, \quad \mathrm{Cov}(X, -2X) = -2\mathrm{Var}[X] < 0.$$

Exercise 5.8. *Let $(\Omega, \mathbb{P})$ be a finite probability space and $X, Y : \Omega \to \mathbb{R}$ be two random variables. Prove that X, Y independent $\Rightarrow$ X, Y uncorrelated. Show with a counterexample that the opposite implication is not true. Prove the inequality*

$$-\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]} \leq \mathrm{Cov}(X, Y) \leq \sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}. \tag{5.17}$$

Show that the left (resp. right) inequality becomes an equality if and only if there exists a negative (resp. positive) constant a_0 and a real constant b_0 such that $Y = a_0 X + b_0$.

By inequality (5.17) it is convenient to introduce the following definition.

Definition 5.11. *Let $X, Y : \Omega \to \mathbb{R}$ be two random variables such that $\mathrm{Var}[X]$ and $\mathrm{Var}[Y]$ are both positive (i.e., X, Y are not deterministic constants). Then*

$$\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}} \in [-1, 1]$$

is called the **correlation** *of X, Y.*

Hence, the closer is $\mathrm{Cor}(X,Y)$ to 1 (resp. -1) the more Y has the tendency to move in the same (resp. opposite) direction of X. For instance $\mathrm{Cor}(X,2X)=1$, and $\mathrm{Cor}(X,-2X)=-1$.

Exercise 5.9. *Compute the correlation of the random variables $X,Y:\Omega_3\to\mathbb{R}$ given by $X(\omega)=N_T(\omega)-N_H(\omega)$, $Y(\omega)=N_H(\omega)$.*

Definition 5.12. *If X,Y are two random variables on a finite probability space, then the function $f_{X,Y}:\mathbb{R}^2\to[0,1]$ given by*

$$f_{X,Y}(x,y)=\mathbb{P}(X=x,Y=y)$$

is called the **joint distribution** *of X and Y.*

Note that $f_{X,Y}$ is non-zero if and only if $(x,y)\in\mathrm{Im}(X)\times\mathrm{Im}(Y)$.

Exercise 5.10. *Let X,Y have the joint distribution $f_{X,Y}$. Show that the distributions of X and Y are given by*

$$f_X(x)=\sum_{y\in\mathrm{Im}(X)}f_{X,Y}(x,y),\quad f_Y(y)=\sum_{x\in\mathrm{Im}(X)}f_{X,Y}(x,y)$$

Show that X,Y are independent if and only if $f_{X,Y}(x,y)=f_X(x)f_Y(y)$.

If the joint distribution is given, then the covariance of two random variables can easily be computed without any reference to the original probability space. In fact, since

$$\mathbb{E}[XY]=\sum_{i=1}^{M}X(\omega_i)Y(\omega_i)\mathbb{P}(\omega_i)=\sum_{x\in\mathrm{Im}(X)}\sum_{y\in\mathrm{Im}(Y)}x\,y\,f_{X,Y}(x,y),\tag{5.18}$$

then

$$\mathrm{Cov}(X,Y)=\sum_{x\in\mathrm{Im}(X)}\sum_{y\in\mathrm{Im}(Y)}x\,y\,f_{X,Y}(x,y)-\sum_{x\in\mathrm{Im}(X)}xf_X(x)\sum_{y\in\mathrm{Im}(Y)}y\,p_Y(y),\tag{5.19}$$

where the distributions f_X,f_Y are computed from $f_{X,Y}$ as shown in Exercise 5.10. In conclusion, if the joint distribution of two random variables X,Y is given, then all relevant information on X,Y (independence, expectation, variance, correlation, etc.) can be inferred without any reference to the original probability space.

Example. Consider two random variables X,Y such that $\mathrm{Im}(X)=\{-1,1,3,4\}$, $\mathrm{Im}(Y)=\{-1,0,1,2\}$ and let their joint probability distribution be defined as in the following table

$Y\downarrow$, $X\rightarrow$	-1	1	3	4
-1	1/64	2/64	1/64	4/64
0	5/64	1/64	9/64	6/64
1	6/64	10/64	1/64	12/64
2	2/64	1/64	1/64	2/64

For instance, $f_{X,Y}(-1,1) = 1/64$, $f_{X,Y}(-1,0) = 5/64$, and so on. Let us compute $\mathbb{E}[X]$, $\mathbb{E}[Y]$ and $\mathrm{Cov}(X,Y)$.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x\in\mathrm{Im}(X)} x f_X(x) = \sum_{x\in\mathrm{Im}(X)} x \sum_{y\in\mathrm{Im}(Y)} f_{X,Y}(x,y) \\
&= -\sum_{y\in\mathrm{Im}(Y)} f_{X,Y}(-1,y) + \sum_{y\in\mathrm{Im}(Y)} f_{X,Y}(1,y) + 3\sum_{y\in\mathrm{Im}(Y)} f_{X,Y}(3,y) + 4\sum_{y\in\mathrm{Im}(Y)} f_{X,Y}(4,y) \\
&= -(\frac{1}{64}+\frac{5}{64}+\frac{6}{64}+\frac{2}{64}) + (\frac{2}{64}+\frac{1}{64}+\frac{10}{64}+\frac{1}{64}) \\
&\quad + 3(\frac{1}{64}+\frac{9}{64}+\frac{1}{64}+\frac{1}{64}) + 4(\frac{4}{64}+\frac{6}{64}+\frac{12}{64}+\frac{2}{64}) = \frac{33}{16}.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_{y\in\mathrm{Im}(Y)} y f_Y(y) = \sum_{y\in\mathrm{Im}(Y)} y \sum_{x\in\mathrm{Im}(X)} f_{X,Y}(x,y) \\
&= -\sum_{x\in\mathrm{Im}(X)} f_{X,Y}(x,-1) + 0\cdot\sum_{x\in\mathrm{Im}(X)} f_{X,Y}(x,0) + \sum_{x\in\mathrm{Im}(X)} f_{X,Y}(x,1) + 2\sum_{x\in\mathrm{Im}(X)} f_{X,Y}(x,2) \\
&= -(\frac{1}{64}+\frac{2}{64}+\frac{1}{64}+\frac{4}{64}) + 0\cdot(\frac{5}{64}+\frac{1}{64}+\frac{9}{64}+\frac{6}{64}) \\
&\quad + (\frac{6}{64}+\frac{10}{64}+\frac{1}{64}+\frac{12}{64}) + 2(\frac{2}{64}+\frac{1}{64}+\frac{1}{64}+\frac{2}{64}) = \frac{33}{64}.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_{x\in\mathrm{Im}(X)}\sum_{y\in\mathrm{Im}(Y)} x\,y\,f_{X,Y}(x,y) \\
&= (-1)(-1)\frac{1}{64} + 1(-1)\frac{2}{64} + 3(-1)\frac{1}{64} \\
&\quad + \cdots\cdots\cdots\cdots = \frac{55}{64}.
\end{aligned}$$

Hence $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \approx -0.204$.

Exercise 5.11. *Compute* $\mathrm{Cor}(X,Y)$.

5.2.3 Conditional expectation

If X, Y are independent random variables, knowing the value of Y does not help to estimate the random variable X. However if X, Y are not independent, then we may use the information on the value attained by Y to find an estimate of X which is better than $\mathbb{E}[X]$. This leads to the important concept of **conditional expectation**.

Definition 5.13. *Let* $(\Omega, \mathbb{P})$ *be a finite probability space,* $X, Y : \Omega \to \mathbb{R}$ *random variables and* $y \in \mathbb{R}$. *The expectation of* X *conditional to* $Y = y$ *(or given the event* $\{Y = y\}$*) is defined as*

$$\mathbb{E}[X|Y=y] = \sum_{x\in\mathrm{Im(X)}} \mathbb{P}(X=x|Y=y)\,x$$

where $\mathbb{P}(X = x|Y = y)$ is the conditional probability of the event $\{X = x\}$, given the event $\{Y = y\}$ (see Def. 5.3), The random variable

$$\mathbb{E}[X|Y] : \Omega \to \mathbb{R}, \quad \mathbb{E}[X|Y](\omega) = \mathbb{E}[X|Y = Y(\omega)]$$

is called the expectation of X conditional to Y.

In a similar fashion one defines $\mathbb{E}[X|Y_1 = y_1, Y_2 = y_2, \dots Y_N = y_N]$ and $\mathbb{E}[X|Y_1, \dots, Y_N]$, where $X, Y_1, \dots Y_N$ are random variables and $y_1, \dots, y_N \in \mathbb{R}$. For example, in the probability space of a fair die, consider

$$X(\omega) = (-1)^\omega, \quad Y(\omega) = (\omega - 1)(\omega - 2)(\omega - 3), \quad \omega \in \{1, 2, 3, 4, 5, 6\}. \tag{5.20}$$

Then we compute

$$\begin{aligned}\mathbb{E}[X|Y = 0] &= \mathbb{P}(X = 1|Y = 0) - \mathbb{P}(X = -1|Y = 0)\\ &= \frac{\mathbb{P}(X = 1, Y = 0)}{\mathbb{P}(Y = 0)} - \frac{\mathbb{P}(X = -1, Y = 0)}{\mathbb{P}(Y = 0)}\\ &= \frac{\mathbb{P}(\{2\})}{\mathbb{P}(\{1, 2, 3\})} - \frac{\mathbb{P}(\{1, 3\})}{\mathbb{P}(\{1, 2, 3\})} = -1/3.\end{aligned}$$

Exercise 5.12. *Derive the random variable $\mathbb{E}[X|Y]$, where X, Y are given by* (5.20).

The following theorem collects a few simple properties of the conditional expectation that will be used later on.

Theorem 5.2. *Let $X, Y, Y_1, Y_2 \dots, Y_N : \Omega \to \mathbb{R}$ be random variables on the finite probability space $(\Omega, \mathbb{P})$. Then*

1. *The conditional expectation is a linear operator, i.e.,*

$$\mathbb{E}[\alpha X + \beta Y|Y_1, \dots, Y_N] = \alpha\mathbb{E}[X|Y_1, \dots, Y_N] + \beta\mathbb{E}[Y|Y_1, \dots, Y_N],$$

   for all $\alpha, \beta \in \mathbb{R}$.

2. *If X is independent of $Y_1, \dots, Y_N$, then $\mathbb{E}[X|Y_1, \dots, Y_N] = \mathbb{E}[X]$;*

3. *If X is measurable with respect to $Y_1, .., Y_N$, i.e., $X = g(Y_1, Y_2, \dots, Y_N)$ for some function g, then $\mathbb{E}[X|Y_1, \dots, Y_N] = X$;*

4. $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$.

Proof. 1. Obvious.

2. By the definitions of conditional expectation and independent random variables we have, for any fixed $\omega \in \Omega$,
$$\begin{aligned}\mathbb{E}[X|Y_1,\dots,Y_N](\omega) &= \mathbb{E}[X|Y_1 = Y_1(\omega),\cdots,Y_N = Y_N(\omega)]\\ &= \sum_{x\in \mathrm{Im}(X)} \mathbb{P}(X = x|Y_1 = Y_1(\omega),\dots,Y_N = Y_N(\omega))x\\ &= \sum_{x\in \mathrm{Im}(X)} \frac{\mathbb{P}(X = x, Y_1 = Y_1(\omega),\dots,Y_N = Y_N(\omega))}{\mathbb{P}(Y_1 = Y_1(\omega),\dots,Y_N = Y_N(\omega))}x\\ &= \sum_{x\in \mathrm{Im}(X)} \frac{\mathbb{P}(X = x)\mathbb{P}(Y_1 = Y_1(\omega))\dots\mathbb{P}(Y_N = Y_N(\omega))}{\mathbb{P}(Y_1 = Y_1(\omega))\dots\mathbb{P}(Y_N = Y_N(\omega))}x\\ &= \sum_{x\in \mathrm{Im}(X)} \mathbb{P}(X = x)x = \mathbb{E}[X].\end{aligned}$$

3. Since $X = g(Y_1,\dots,Y_N)$, then $X = x$ if and only if there exist $(y_1^{(x)},\dots y_N^{(x)})$ in the image of $(Y_1,\dots,Y_N)$ such that $x = g(y_1^{(x)},\dots,y_N^{(x)})$. Let $A_x = \{\omega \in \Omega : Y_i(\omega) = y_i^{(x)},\ \text{for all}\, i = 1,\dots,N\}$. Thus
$$X(\omega) = g(Y_1(\omega),\dots,Y_N(\omega)) = x, \quad \text{if and only if } \omega \in A_x.$$
Hence $\mathbb{P}(X = x|Y_1 = Y_1(\omega),\dots,Y_N = Y_N(\omega))$ is 1 if $\omega \in A_x$ and 0 otherwise. Letting $\mathbb{I}_{A_x}$ the characteristic function of the set A_x, the definition of conditional expectation gives
$$\mathbb{E}[X|Y_1,\dots,Y_N](\omega) = \sum_{x\in \mathrm{Im}(X)} x\mathbb{I}_{A_x}(\omega) = X(\omega).$$

4. By definition of expectation and conditional expectation we have
$$\begin{aligned}\mathbb{E}[\mathbb{E}[X|Y]] &= \sum_{\omega\in\Omega} \mathbb{E}[X|Y = Y(\omega)]\mathbb{P}(\omega) = \sum_{y\in \mathrm{Im}(Y)} \mathbb{E}[X|Y = y]\mathbb{P}(Y = y)\\ &= \sum_{(x,y)\in \mathrm{Im}(X)\times \mathrm{Im}(Y)} \mathbb{P}(X = x|Y = y)\mathbb{P}(Y = y)x\\ &= \sum_{(x,y)\in \mathrm{Im}(X)\times \mathrm{Im}(Y)} \mathbb{P}(X = x, Y = y)x = \sum_{(x,y)\in \mathrm{Im(X)}\times \mathrm{Im}(Y)} xf_{X,Y}(x,y)\\ &= \sum_{x\in \mathrm{Im}(X)} xf_X(x) = \mathbb{E}[X],\end{aligned}$$
where we used that $\sum_{y\in \mathrm{Im}(Y)} f_{X,Y}(x,y)$ equals the distribution $f_X(x)$, see Exercise 5.10. $\square$

The interpretation of 2-3 in the previous theorem is the following: If Y is independent of $X_1,\dots,X_N$, then the information carried by $X_1,\dots,X_N$ does not help to improve our estimate on Y (this estimate remaining $\mathbb{E}[Y]$). On the other hand, if Y depends on $X_1,\dots,X_N$, then by knowing $X_1,\dots,X_N$ we also know Y and thus our best estimate on Y is Y itself.

Exercise 5.13 (?). *What is the interpretation of 4 in Theorem 5.2?*

Exercise 5.14. *Let X, Y, Z be random variables on the finite probability space $(\Omega, \mathbb{P})$ and assume X is measurable with respect to Z. Show that $\mathbb{E}[XY|Z] = X\mathbb{E}[Y|Z]$ and interpret the result.*

5.3 Stochastic processes. Martingales

Definition 5.14. *Let $T > 0$. A one parameter family of random variables, $X(t) : \Omega \to \mathbb{R}$, $t \in [0, T]$, is called a* **stochastic process**. *We denote a stochastic process by $\{X(t)\}_{t\in[0,T]}$ and by $X(t, \omega)$ the value of the random variable $X(t)$ on the sample $\omega \in \Omega$. For each fixed $\omega \in \Omega$, the curve $t \to X(t, \omega)$, is called a* **path** *of the stochastic process.*

We refer to the parameter t as the time variable. If $X(t, \omega) = C(t)$, for all $\omega \in \Omega$, i.e., if the paths are the same for all sample points, we say that the stochastic process is a non-random (or **deterministic**) function of time. If t runs over a (possibly infinite) discrete set $\{t_1, t_2, \dots\} \subset [0, T]$, then we say that the stochastic process is **discrete**. Note that a discrete stochastic process is equivalent to a sequence of random variables:

$$X = \{X_1, X_2, \dots\} \text{ where } X_i = X(t_i),\ i = 1, 2, \dots.$$

As a way of example, consider the following (discrete) stochastic process defined on the N-coin toss probability space $(\Omega_N, \mathbb{P}_p)$:

$$\omega = (\gamma_1, \dots, \gamma_N) \in \Omega_N, \quad X_i(\omega) = \begin{cases} 1 & \text{if } \gamma_i = H \\ -1 & \text{if } \gamma_i = T \end{cases}$$

Clearly, the random variables $X_1, \dots, X_N$ are independent and have all the same distribution. In particular,

$$\mathbb{P}_p(X_i = 1) = p, \quad \mathbb{P}_p(X_i = -1) = 1 - p, \quad \text{for all } i = 1, \dots, N.$$

From now on we assume that the coin is fair ($p = 1/2$); we then have

$$\mathbb{E}[X_i] = 0, \quad \mathrm{Var}[X_i] = 1, \quad \text{for all } i = 1, \dots, N.$$

Moreover, for $n = 0, \dots, N$ we set

$$M_0 = 0, \quad M_n = \sum_{i=1}^{n} X_i, \quad \text{for } n \geq 1.$$

The stochastic process $(M_0, \dots, M_N)$ is called **symmetric random walk**. It is centered in zero, which means that

$$\mathbb{E}[M_n] = 0, \quad \text{for all } n = 0, \dots, N.$$

Moreover, since it is the sum of independent random variables, the symmetric random walk has variance given by

$$\mathrm{Var}(M_n) = \mathrm{Var}(X_1 + X_2 + \cdots + X_n) = n,$$

see Theorem 5.1.

To understand the meaning of the term "random walk", consider a particle moving on the real line in the following way: if $X_i = 1$ (i.e., if the i^{th} toss is a head), at time $t = i$ the particle moves one unit of length to the right, while if $X_i = -1$ (i.e., if the i^{th} toss is a tail) it moves one unit of length to the left. Hence M_n gives the total amount of units of length that the particle has traveled to the right or to the left up to the time nt. The **increments** of the random walk are defined as follows. Given $0 = k_0 < k_1 < \cdots < k_m = N$, we let

$$\Delta_i = M_{k_{i+1}} - M_{k_i} = \sum_{j=k_i+1}^{k_{i+1}} X_j, \quad i = 0, \dots, m-1,$$

i.e., $\Delta_0 = M_{k_1}$, $\Delta_1 = M_{k_2} - M_{k_1}, \dots, \Delta_{m-1} = M_{k_m} - M_{k_{m-1}}$. Hence Δ_j is the total displacement of the particle from time k_{j-1} to time k_j. It follows by Theorem 5.1 that the increments of the random walk are independent random variables, that is to say, the distance traveled by the particle in the time interval $[k_{j-1}, k_j]$ is independent of the movements during any earlier or later time interval. Moreover

$$\mathbb{E}[\Delta_i] = 0, \quad \mathrm{Var}[\Delta_i] = k_{i+1} - k_i. \tag{5.21}$$

Exercise 5.15. *Let $T > 0$ and $n \in \mathbb{N}$ be given. Define the stochastic process*

$$\{W_n(t)\}_{t\in[0,T]}, \quad W_n(t) = \frac{1}{\sqrt{n}} M_{[nt]}, \tag{5.22}$$

where $[z]$ denotes the greatest integer smaller than or equal to z and $M_k = X_1 + X_2 + \cdots + X_k$, $k = 1, \dots, N$, is a symmetric random walk. It is assumed that the stochastic process $(X_1, \dots, X_N)$ is defined for $N > [nT]$, so that $W_n(t)$ is defined for all $t \in [0, T]$. Compute $\mathbb{E}[W_n(t)]$, $\mathrm{Var}[W_n(t)]$, $\mathrm{Cov}[W_n(t), W_n(s)]$. *Show that* $\mathrm{Var}(W_n(t)) \to t$ *and* $\mathrm{Cov}(W_n(t), W_n(s)) \to \min(s, t)$ *as $n \to +\infty$.*

Remark 5.1. For large $n \in \mathbb{N}$, the process $\{W_n(t)\}_{t\in[0,T]}$ can be used an approximation for the Brownian motion, see Definition 5.18 below. An example of path of the stochastic process $\{W_n(t)\}_{t\in[0,T]}$ for $n = 1000$ is shown in Figure 5.1.

Exercise 5.16 (Matlab). *Write a Matlab function that generates a random path of the stochastic process $\{W_n(t)\}_{t\in[0,T]}$.*

To conclude this section we introduce the fundamental concept of martingale. Roughly speaking, a martingale is a stochastic process which has no tendency to rise or fall. The precise definition is the following.

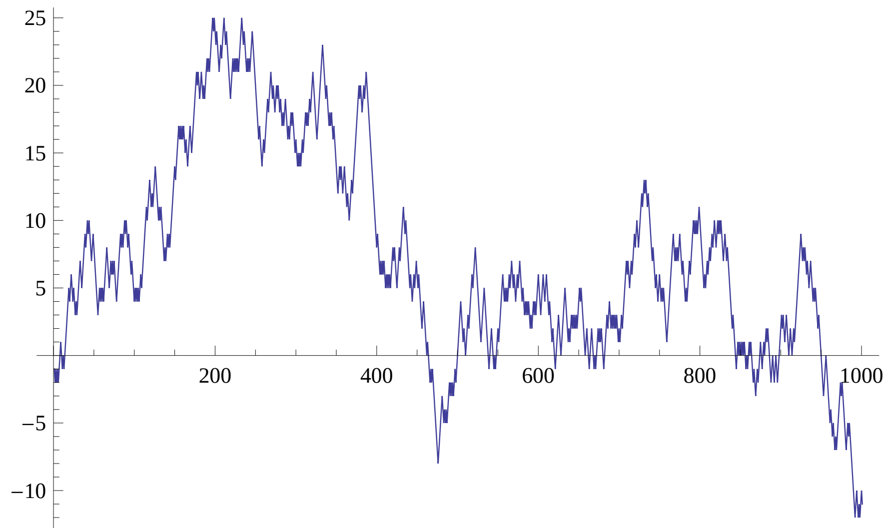


Figure 5.1: A path of the stochastic process (5.22) for $n = 1000$.

Definition 5.15. *A discrete stochastic process* $\{X_1, X_2, \dots\}$ *on the finite probability space* $(\Omega, \mathbb{P})$ *is called a* **martingale** *if*

$$\mathbb{E}[X_{i+1}|X_1, X_2, \dots X_i] = X_i, \quad \textit{for all } i \geq 1. \tag{5.23}$$

The interpretation is the following: The variables $X_1, X_2, \dots X_i$ contains the information obtained by "looking" at the stochastic process up to the step i. For a martingale process, this information is not enough to infer whether, in the next step, the process will raise or fall.

Exercise 5.17. *Let* $\{X_1, X_2, \dots\}$ *be a discrete martingale on the finite probability space* $(\Omega, \mathbb{P})$*. Show that* $\mathbb{E}[X_i] = \mathbb{E}[X_1]$ *for all* $i \geq 1$*, i.e., martingales have constant expectation.*

An example of martingale is the random walk introduced above. In fact, using the linearity of the conditional expectation we have

$$\begin{aligned}\mathbb{E}[M_n|M_1, \dots, M_{n-1}] &= \mathbb{E}[M_{n-1} + X_n|M_1, \dots, M_{n-1}] \\ &= \mathbb{E}[M_{n-1}|M_1, \dots, M_{n-1}] + \mathbb{E}[X_n|M_1, \dots, M_{n-1}].\end{aligned}$$

As M_{n-1} is measurable with respect to $M_1, \dots, M_{n-1}$, then $\mathbb{E}[M_{n-1}|M_1, \dots, M_{n-1}] = M_{n-1}$, see Theorem 5.2 (3). Moreover, as X_n is independent of $M_1, \dots, M_{n-1}$, Theorem 5.2 (2) gives $\mathbb{E}[X_n|M_1, \dots, M_{n-1}] = \mathbb{E}[X_n] = 0$. It follows that $\mathbb{E}[M_n|M_1, \dots, M_{n-1}] = M_{n-1}$, i.e., the random walk is a martingale.

5.4 Applications to the binomial model

In this section we reconsider the binomial options pricing model using the language of probability theory. We first show that the binomial stock price can be interpreted as a stochastic process defined on the N-coin toss probability space $(\Omega_N, \mathbb{P}_p)$, see Definition 5.2. Recall that, for a given $0 < p < 1$, the binomial stock price at time t is given by

$$S(t) = \begin{cases} S(t-1)e^u & \text{with probability } p \\ S(t-1)e^d & \text{with probability } 1-p, \end{cases} \tag{5.24}$$

where $u > d$. Now, for $t \in \mathcal{I} = \{1, \dots, N\}$, consider the following random variable

$$X_t : \Omega_N \to \mathbb{R}, \quad X_t(\omega) = \begin{cases} u, & \text{if the } t^{th} \text{ toss in } \omega \text{ is } H \\ d, & \text{if the } t^{th} \text{ toss in } \omega \text{ is } T \end{cases}.$$

Hence we can write

$$S(t) = S(t-1)e^{X_t} = S(t-2)e^{Z_{X-1}+X_t} = \dots = S(0)\exp(X_1 + X_2 + \dots + X_t) : \Omega_N \to \mathbb{R},$$

where $S(0)$ is the initial value of the stock price, which is a deterministic constant. Thus $S(t)$ is a random variable and therefore $\{S(t)\}_{t\in\mathcal{I}}$ is a (discrete) stochastic process. For each $\omega \in \Omega_N$, the vector $(S(1,\omega), \dots, S(N,\omega))$ is a path for the stock price.

The value at time t of the risk-free asset is the deterministic function of time $B(t) = B_0\exp(rt)$, where $r > 0$ is the (constant) interest rate and B_0 is the initial value of the risk-free asset. Recall that $\widehat{S}(t) = e^{-rt}S(t)$ is called the discounted price of the stock, see Remark 2.3.

Theorem 5.3. *If $r \notin (d,u)$, there is no probability measure $\mathbb{P}_p$ on the sample space Ω_N such that the discounted stock price process $\{\widehat{S}(t)\}_{t\in\mathcal{I}}$ is a martingale. For $r \in (d,u)$, $\{\widehat{S}(t)\}_{t\in\mathcal{I}}$ is a martingale with respect to the probability measure $\mathbb{P}_q$, where*

$$q = \frac{e^r - e^d}{e^u - e^d}.$$

Moreover $\mathbb{P}_q$ is the only probability measure on Ω_N for which $\{\widehat{S}(t)\}_{t\in\mathcal{I}}$ is a martingale.

Proof. By definition, $\{\widehat{S}(t)\}_{t\in\mathcal{I}}$ is a martingale if and only if

$$\mathbb{E}[e^{-rt}S(t)|\widehat{S}(1), \dots, \widehat{S}(t-1)] = e^{-r(t-1)}S(t-1), \quad \text{for all } t \in \mathcal{I}.$$

Clearly, taking the expectation conditional to $\widehat{S}(1), \dots, \widehat{S}(t-1)$ is the same as taking the expectation conditional to $S(1), \dots, S(t-1)$, hence the above equation is equivalent to

$$\mathbb{E}[S(t)|S(1), \dots, S(t-1)] = e^r S(t-1), \quad \text{for all } t \in \mathcal{I}, \tag{5.25}$$

where we canceled out a factor e^{-rt} in both sides of the equation. Moreover

$$\begin{aligned}\mathbb{E}[S(t)|S(1), \dots, S(t-1)] &= \mathbb{E}[\frac{S(t)}{S(t-1)}S(t-1)|S(1), \dots, S(t-1)] \\ &= S(t-1)\mathbb{E}[\frac{S(t)}{S(t-1)}|S(1), \dots, S(t-1)],\end{aligned}$$

where we used the fact that $S(t-1)$ is known and thus can be taken out from the conditional expectation (see Exercise 5.14). As

$$S(t)/S(t-1) = \begin{cases} e^u & \text{with prob. } p \\ e^d & \text{with prob. } 1-p \end{cases}$$

is independent of $S(1), \dots, S(t-1)$, then by Theorem 5.2(2) we have

$$\mathbb{E}[\frac{S(t)}{S(t-1)}|S(1), \dots, S(t-1)] = \mathbb{E}[\frac{S(t)}{S(t-1)}] = e^u p + e^d(1-p).$$

Hence (5.25) holds if and only if $e^u p + e^d(1-p) = e^r$. The latter has a solution $p \in (0,1)$ if and only if $r \in (d,u)$ and the solution, when it exists, is given by $p = q$. $\square$

Remark 5.2. Due to Theorem 5.3, $\mathbb{P}_q$ is called **martingale probability measure**. Moreover we can reformulate Theorem 2.2 as follows: *a 1+1 dimensional binomial stock market is arbitrage free if and only if there exists a martingale probability measure.*

Theorem 5.4. *Let $\mathbb{E}_p[\cdot]$ denote the expectation in the probability measure $\mathbb{P}_p$. We have*

$$\mathbb{E}_p[S(N)] = S(0)(e^u p + e^d(1-p))^N, \quad \mathbb{E}_q[S(N)] = S(0)e^{rN}.$$

Proof. The second formula follows by the first one using that $e^u q + e^d(1-q) = e^r$. To prove the first formula we use

$$\mathbb{E}_p[S(N)] = \mathbb{E}_p[S(0)\exp(X_1 + \cdots + X_N)] = S(0)\mathbb{E}_p[Y],$$

where Y is the random variable

$$Y(\omega) = \exp(X_1(\omega) + \cdots + X_N(\omega)) = \exp(uN_H(\omega) + dN_T(\omega)), \quad \omega \in \Omega_N.$$

Hence, using $N_T(\omega) = N - N_H(\omega)$ and (5.10) we obtain

$$\begin{aligned}\mathbb{E}_p[S(N)] &= S(0)\sum_{\omega\in\Omega_N} e^{(uN_H(\omega)+dN_T(\omega))}p^{N_H(\omega)}(1-p)^{N_T(\omega)} \\ &= S(0)e^{Nd}(1-p)^N \sum_{\omega\in\Omega_N} e^{(u-d)N_H(\omega)}\left(\frac{p}{(1-p)}\right)^{N_H(\omega)}.\end{aligned}$$

Now we use that for $k = 0, \dots, N$ there exist exactly $\binom{N}{k}$ sample points $\omega \in \Omega_N$ such that $N_H(\omega) = k$. Hence we can write

$$\mathbb{E}_p[S(N)] = S(0)e^{Nd}(1-p)^N \sum_{k=0}^{N} \binom{N}{k}\left(\frac{e^u p}{e^d(1-p)}\right)^k.$$

By the binomial theorem, $(1+a)^N = \sum_{k=0}^N \binom{N}{k} a^k$, hence

$$\mathbb{E}_p[S(N)] = S(0)e^{Nd}(1-p)^N\left(1 + \frac{e^u p}{e^d(1-p)}\right)^N = S(0)(e^d(1-p) + e^u p)^N.$$

$\square$

Exercise 5.18. *Compute* $\mathrm{Var}_p[S(N)]$, $\mathrm{Var}_q[S(N)]$.

Remark 5.3. The expectation value of the stock price at time $t = N$ in the probability space $(\Omega_N, \mathbb{P}_q)$ depends only on the risk-free asset and not on the stock price dynamics. For this reason, $\mathbb{P}_q$ is also called **risk neutral probability**

Let us now derive the distribution of $S(t)$, $t \in \mathcal{I}$. First we notice that the image of $S(t)$ is $\mathrm{Im}\, S(t) = \{s_0, \dots, s_t\}$, where

$$s_k = S(0)\exp(ku + (t-k)d), \quad k = 0, \dots, t.$$

Hence $f_{S(t)}(x) = 0$, if $x \neq s_k$, for all $k = 0, \dots t$ and $t \in \mathcal{I}$. The value of $S(t)$ is s_k if there are exactly k heads in the first t tosses of the N-toss. Hence we obtain that $f_{S(t)}$ is given by the so-called **binomial distribution**:

$$f_{S(t)}(s_k) = \mathbb{P}_p[S(t) = s_k] = \binom{t}{k} p^k (1-p)^{t-k}, \quad k = 0, \dots, t, \quad t \in \mathcal{I}.$$

In particular, in the case of a fair coin $p = 1/2$, we have

$$f_{S(t)}(x) = \binom{t}{k} \frac{1}{2^t} \delta(x - S(0)\exp(ku + (t-k)d)) \quad \text{(fair coin)}.$$

where $\delta(z) = 1$ if $z = 0$ and $\delta(z) = 0$ for $z \neq 0$. It is customary to express this result in terms of the log-price of the stock, i.e., $\log S(t)$, as the latter can take both positive and negative values. For a fair coin we get

$$f_{\log(S(t)/S(0))}(x) = \binom{t}{k} \frac{1}{2^t} \delta(x - (ku + (t-k)d), \tag{5.26}$$

An example of this distribution is depicted in Figure 5.2. As it is clear from the picture, for large values of $N \in \mathbb{N}$, the distribution of $\log(S(N)/S(0)) = \log S(N) - \log S(0)$ can be approximated by a **normal** distribution, i.e., a distribution of the form (5.31) below. One of the main critics to the binomial model is that it assigns very law probability to large variations of the (log-)stock price.

The binomial distribution can be used to compute the probability that the price of the stock lies in an interval $[a, b]$ at time t. By (5.7) we have

$$\mathbb{P}(a \leq S(t) \leq b) = \sum_{k: a \leq s_k \leq b} f_{S(T)}(s_k) = \sum_{k: a \leq s_k \leq b} \binom{t}{k} p^k (1-p)^{t-k} \quad k = 0, \dots, t, \quad t \in \mathcal{I}. \tag{5.27}$$

Moreover if $g : \mathbb{R} \to \mathbb{R}$ then, by (5.8),

$$\mathbb{P}(a \leq g(S(t)) \leq b) = \sum_{k: a \leq g(s_k) \leq b} f_{S(T)}(s_k) = \sum_{k: a \leq g(s_k) \leq b} \binom{t}{k} p^k (1-p)^{t-k} \quad k = 0, \dots, t, \quad t \in \mathcal{I}. \tag{5.28}$$

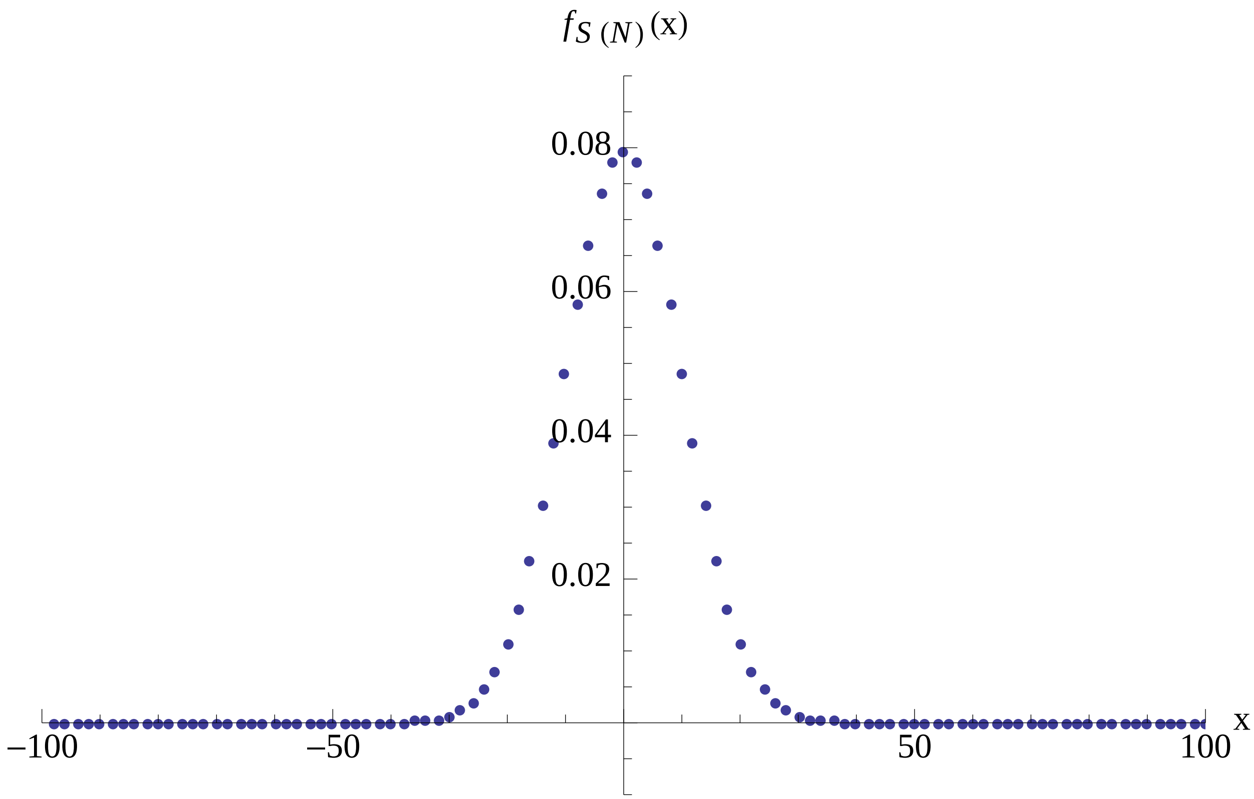


Figure 5.2: The distribution of $\log(S(N))$. It is assumed $S(0) = 1$, $u = -d = 1$, $p = 1/2$ and $N = 100$.

The formula (5.28) can be used to compute the probability that a derivative on the stock is in the money at time t. Consider a standard European or American derivative with pay-off $Y = g(S(N))$ at the expiration date $T = N$. Then the probability that the derivative is in the money at time t is given by

$$\mathbb{P}(Y(t) > 0) = \mathbb{P}(g(S(t)) > 0) = \sum_{k:g(s_k)>0} \binom{t}{k} p^k (1-p)^{t-k}.$$

We shall implement these formulas with Matlab in Section 5.4.3

The value of a portfolio position (h_S, h_B) invested on h_S shares of the stock and h_B of the risk-free asset is the stochastic process $(V(1), V(2), \dots, V(N))$, where

$$V(t) = h_B B_0 e^{rt} + h_S S(t) \, : \, \Omega_N \to \mathbb{R}, \quad t \in \mathcal{I},$$

while $V(0) = h_S S(0) + h_B B_0$ is a non-random constant. If we change the portfolio position depending on the price of the stock (i.e., depending on $\omega \in \Omega_N$), then we get a portfolio (stochastic) process $\{(h_S(t), h_B(t))\}_{t\in\mathcal{I}}$. We recall that $(h_S(t), h_B(t))$ corresponds to the portfolio position held in the interval $(t-1, t]$. The value $(V(1), V(2), \dots, V(N))$ of the portfolio process is the stochastic process given by

$$V(t) = h_B(t) B_0 e^{rt} + h_S(t) S(t) \, : \, \Omega_N \to \mathbb{R}, \quad t \in \mathcal{I}.$$

A portfolio process $\{(h_S(t), h_B(t))\}_{t\in\mathcal{I}}$ is said to be self-financing if

$$V(t-1) = h_B(t)B(t-1) + h_S(t)S(t-1).$$

We have seen in Theorem 2.1 that the value at time $t = N$ of a self-financing portfolio completely determines the value at any earlier time and in particular at time $t = 0$ it is given by (2.8). The latter result can be reformulated using the language of probability theory as follows:

Theorem 5.5. *Let $\{(h_S(t), h_B(t))\}_{t\in\mathcal{I}}$ be a self-financing portfolio with value $V(N)$ at time $t = N$. Then*

$$V(0) = e^{-rN}\mathbb{E}_q[V(N)],$$

i.e., $V(0)$ is the expectation, in the risk-neutral measure, of the discounted portfolio value at time N.

Exercise 5.19. *Prove the formula*

$$V(t) = e^{-r(N-t)}\mathbb{E}_q[V(N)|S(0),\dots,S(t)].$$

Now let $Y : \Omega_N \to \mathbb{R}$ be a random variable and consider a European-style derivative with pay-off $Y : \Omega_N \to \mathbb{R}$ at maturity time N. This means that the derivative can only be exercised at time $t = N$ (for standard European derivatives Y is a deterministic function of $S(N)$). Let $\Pi_Y(t)$ be the binomial fair price of the derivative a time t. By definition, $\Pi_Y(t)$ is the value $V(t)$ of a self-financing, hedging portfolio. In particular, $\Pi_Y(t)$ is a random variable and so $(\Pi_Y(1),\dots,\Pi_Y(N))$ is a stochastic process. Using the hedging condition $V(N) = Y$ (which means $V(N,\omega) = Y(\omega)$, for all $\omega \in \Omega_N$)) and Theorem 5.5 we have the following formula for the fair price at time $t = 0$ of the financial derivative:

$$\Pi_Y(0) = e^{-rN}\mathbb{E}_q[Y], \tag{5.29}$$

i.e., $\Pi_Y(0)$ is the expectation in the risk-neutral measure of the discounted value of the pay-off. Moreover by Exercise 5.19 we have

$$\Pi_Y(t) = e^{-r(N-t)}\mathbb{E}_q[Y|S(0),\dots,S(t)], \tag{5.30}$$

which is called the **risk-neutral pricing formula**.

Finally we remark that in the applications one is interested in hedging portfolios whose position in the interval $(t-1, t]$ can be set up *at* time $t-1$ using the information available *up to* time $t-1$. Obviously, a portfolio process whose position $(h_S(t), h_B(t))$ depends on information not yet available is useless. We say that a portfolio process $\{(h_S(t), h_B(t)\}_{t\in\mathcal{I}}$ is *predictable* if there exist functions $H_1,\dots,H_N$, where $H_t : (0,\infty)^t \to \mathbb{R}^2$, such that

$$(h_S(t), h_B(t)) = H_t(S(0),\dots,S(t-1)).$$

This means that the random variable $(h_S(t), h_B(t))$ is measurable with respect to the random variables $S(1),\dots S(t-1)$. We have seen that European derivatives admit predictable hedging portfolios. The hedging portfolio is in general *not* uniquely defined. However, by Theorem 5.5 (and Exercise 5.19), all hedging portfolios have the same value. In particular, the fair price of a derivative is unambiguously defined.

5.4.1 General discrete options pricing models

Equation (5.30) is the most important result of this chapter. In fact it leads to the following general strategy to introduce discrete options pricing models for European derivatives:

1. Fix a finite probability space $(\Omega, \mathbb{P})$
2. Define the stock price as a positive stochastic process on the probability space $(\Omega, \mathbb{P})$
3. Prove the existence of a probability measure $\mathbb{Q}$ which makes the stock price process a martingale. The existence of a martingale measure ensures that the market is arbitrage-free
4. Provided the martingale measure is unique, define the fair price of European derivatives by (5.30). If the martingale measure is not unique, then the price of the derivative is not uniquely defined and the model is called **incomplete**.

In fact, the above strategy is applied for *all* models in options pricing theory, even for the time-continuum models, which are defined on uncountable probability spaces. An example of time-continuum model is the Black-Scholes model considered in the next chapter (see [6] for an introduction to general time-continuum models). The important lesson to be learned here is the following: *probability theory is the right framework where to formulate the mathematical models in options pricing theory.*

Exercise 5.20. *Formulate an options pricing model in which, at any time step, the stock price can go up, down or stay the same (***trinomial model***). Use the language of probability theory. Show that this model is incomplete (i.e., the martingale measure is not unique).*

5.4.2 Quantitative vs fundamental analysis of a stock

To continue this section we discuss briefly the difference between the **fundamental** analysis and the **quantitative** (or statistical) analysis of a stock. Performing a quantitative analysis means that the investor tries to estimate the price of the stock using a mathematical probabilistic model, such as the binomial model. On the contrary, a fundamental analysis of the stock consists in a careful evaluation of the overall performance of the company and of other economical factors which may affect the value of the stock company. As the quantitative analysis, also the fundamental analysis can be interpreted in the language of probability theory (although in practice nobody does that!). Precisely, let $S(T)$ be the price of the stock at some future time $T > 0$ (e.g., $T = 1$ year from now) and let us start by making a list of all events which we think may affect the value of $S(T)$. Let us call $\{\omega_1\}, \{\omega_2\}, \dots, \{\omega_M\}$ such events; for instance

$$
\begin{aligned}
&\{\omega_1\} = [\text{the company will make a profit this year}]\\
&\{\omega_2\} = [\text{there will be an oil crisis this year}]\\
&\{\omega_3\} = [\text{the company will open a new branch}]\\
&\text{etc.}
\end{aligned}
$$

In this way we have constructed a sample space $\Omega = \{\omega_1, \dots, \omega_M\}$. Next we have to assign a probability to each of these atomic events, say $\mathbb{P}(\{\omega_i\}) = p_i$. Finally we have to specify how these events affect the price of the stock, that is to say, we have to give $S(T, \omega)$, for $\omega \in \Omega$. After all of this we can compute for instance

$$\mathbb{E}[S(T)] = \sum_{\omega \in \Omega} S(T, \omega)\mathbb{P}(\{\omega\}), \quad \mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2,$$

which are our "expected value" for the stock price and our "expected error" of this value. Clearly all the steps that led us to these estimates are quite subjective.

In the quantitative analysis we only try to "guess" what the distribution of the stock price will be at time T, i.e., we assign $f_{S(T)} : (0, \infty) \to [0, 1]$ and then we derive our estimates using the formulas

$$\mathbb{E}[S(T)] = \sum_{s \in I} s\, f_{S(T)}(s), \quad \mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2,$$

where I is a (finite) set of possible values that we admit for the stock price. Note that in the quantitative approach the only arbitrariness is in the choice of the distribution function of the price. In the binomial model we assume that $\log S(T)$ follows a binomial distribution, but other choices are possible and can be justified by looking at the historical data for the stock price. In fact, the advantage of the quantitative method versus the fundamental approach, is that, by employing the stock price as the only relevant information, it provides us with an objective (quantitative) way to justify, monitor and adjust our analysis.

5.4.3 Computing probabilities with Matlab

The following code defines a Matlab function *ProbStock(S,t,a,b,p)* which computes the probability that the stock price stored in the binomial tree matrix S lies in the interval $[a, b]$ at time t, where p is the probability that the stock price goes up at each time step. The matrix S is created with the function *BinomialStock* introduced in Section 2.4.

```
function Prob=ProbStock(S,t,a,b,p)
ind=find(S(:,t+1)>a & S(:,t+1)<b);
numups=t+1-ind;
k=length(numups);
probpath=0;
for j= 1:k
probpath=probpath+nchoosek(t,numups(j))*p^numups(j)*(1-p)^(t-numups(j));
end
Prob=probpath;
```

The function computes the probability $\mathbb{P}(a < S(t) < b)$ using Equation (5.27). Recall that the index k in (5.27) is the number of "ups" up to and including time t; since S_{ij} in the binomial tree S is the price at time j assuming that the price goes down i times, then the number of ups at time $t = j$ is given by $j + 1 - i$. The variable *numups* in the code above contains the number of "ups" for each price at time t (i.e., in the column $t+1$ of S) which lies in the interval (a, b).

Exercise 5.21 (Matlab). *Write a Matlab function ProbDerivative(S,t,g,p) which computes the probability that a standard derivative with pay-off function g is in the money at time t.*

5.5 Infinite Probability Spaces

In this last section we consider briefly probability spaces consisting of an infinite number of sample points. The discussion can be complemented with Chapter 4 in Ref. [1].

Assume first that Ω is a countable set. This means that

$$\Omega = \{\omega_n\}_{n\in\mathbb{N}}.$$

For countable sample spaces the definitions given in the previous sections for finite sets extend straightforwardly. Precisely, given a sequence

$$p = (p_n)_{n\in\mathbb{N}} \quad \text{such that} \quad 0 < p_n < 1, \quad \sum_{n\in\mathbb{N}} p_n = 1,$$

we define the probability of the atomic events as

$$\mathbb{P}(\{\omega_n\}) = p_n.$$

If $A \in 2^\Omega$, then we define

$$\mathbb{P}(A) = \sum_{i:\omega_i\in A} p_i = \sum_{\omega\in A} \mathbb{P}(\{\omega\}).$$

If $X : \Omega \to \mathbb{R}$ is a random variable, then

$$\mathbb{E}[X] = \sum_{n\in\mathbb{N}} X(\omega_n)p_n = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\}).$$

The remaining definitions introduced in the finite case (variance, covariance, independent random variables, etc.) continue to be valid for countable probability spaces.

When Ω is uncountable, there is no general procedure to construct a probability space, but only an abstract definition (which we shall not give). The problem is that in the uncountable case Ω admits very irregular ("wild") sets and thus defining a probability over the whole 2^Ω becomes complicated. Moreover the occurrence of non-trivial events with zero probability poses some technical problem. We restrict ourselves to present some examples.

- Let $\Omega = (0, 1)$ and let the admissible events be given by the sets which can be written as the union of countably many open subintervals of $(0, 1)$. For any admissible event $A \subseteq (0, 1)$ we define

  $$\mathbb{P}(A) = \int_0^1 \mathbb{I}_A(x)\,dx,$$

  where $\mathbb{I}_A(x) = 1$ if $x \in A$ and zero otherwise.

- Let $\Omega = \mathbb{R}$ and $p : \mathbb{R} \to [0, \infty)$ be a continuous function such that
$$\int_{\mathbb{R}} p(x)\, dx = 1.$$
As admissible events we consider all real sets which can be written as the union of countably many open intervals. For any admissible event $A \in \mathbb{R}$ we define
$$\mathbb{P}(A) = \int_A p(x)\, dx.$$

- In this example we construct a probability space that extends $(\Omega_N, \mathbb{P}_p)$ (the N-coin toss probability space) to $N = \infty$. Consider $\Omega_\infty = \{(\gamma_n)_{n\in\mathbb{N}};\, \gamma_i = H \text{ or } T\}$. Thus an element of Ω_∞ is the outcome of the experiment "tossing a coin infinitely many times". It can be shown (using a standard Cantor diagonal argument), that the set Ω_∞ is uncountable. Now, given $\bar{\omega} = (\bar{\gamma}_1, \dots, \bar{\gamma}_N) \in \Omega_N$, consider the events
$$A_N(\bar{\omega}) \subset \Omega_\infty, \quad A = \{(\gamma_n)_{n\in\mathbb{N}} \in \Omega_\infty : \gamma_i = \bar{\gamma}_i,\, i = 1, \dots, N\},$$
that is to say, $A_N(\bar{\omega})$ contains all infinity-tosses whose first N-tosses coincide with $(\bar{\gamma}_1, \dots, \bar{\gamma}_N)$. Of course, $A_N(\bar{\omega})$ is also uncountable. We define the probability of $A_N(\bar{\omega})$ as the probability of $(\bar{\gamma}_1, \dots, \bar{\gamma}_N)$ in the probability space $(\Omega_N, \mathbb{P}_p)$. For instance, assuming that the coin is fair, the event
$$\text{[the first 5 tosses in the infinite sequence are heads]}$$
has probability $(1/2)^5$. By letting $\bar{\omega}$ varies in Ω_N, we get a collection $\mathcal{A}_N$ of 2^N events in Ω_∞,
$$\mathcal{A}_N \subset 2^{\Omega_\infty}, \quad \mathcal{A}_N = \{A_N(\bar{\omega});\, \bar{\omega} \in \Omega_N\},$$
whose probability has been defined. This definition can be extended to all events that can be written as the disjoint union of events in the set $\mathcal{A}_N$, using the rule $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ if A, B are disjoint. In this way we have assigned a probability to all events in the space Ω_∞ that depends only on the first N tosses, and letting N arbitrarily large, we are able to define a probability to any event which depends on a (arbitrary) finite number of tosses (e.g., the event "there is a tail in the first 10^{100} tosses"). This probability space might seem quite large, but actually is not! For instance, the event "there exists at least one head in the infinite-toss" does not belong to this space, hence we cannot assign a probability to it. The inclusion of events which are resolved by tossing the coin infinitely many times requires advanced tools in probability theory (in particular, Carathéodory's theorem), which will not be discussed here.

Fortunately for the applications in mathematical finance the knowledge of the full probability space is usually not necessary, since in finance we are only concerned with random variables and their distributions, rather than with generic events. More precisely, we are only interested to assign a probability to events of the form $\{X \in I\}$, where X is a random variable on the (abstract) probability space and $I \subset \mathbb{R}$, that is to say, events which can be resolved by one (or more) random variables, cf. the discussion in Section 5.4.2.

Definition 5.16. *Let $f_X : \mathbb{R} \to [0, \infty)$ be a continuous function, except possibly on finitely many points. A random variable $X : \Omega \to \mathbb{R}$ is said to have* **probability density** *f_X if*

$$\mathbb{P}(X \in A) = \int_A f_X(x)\, dx,$$

for all regular sets $A \subseteq \mathbb{R}$.

Note that if f_X is the probability density of a random variable X, then necessarily

$$\int_{\mathbb{R}} f_X(x)\, dx = 1.$$

Moreover the cumulative distribution $F_X(x) = \mathbb{P}(X \leq x)$ satisfies

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\, dy, \quad \text{for all } x \in \mathbb{R}, \text{ hence } f_X = \frac{dF_X}{dx}.$$

Examples

- A random variable $X : \Omega \to \mathbb{R}$ is said to be **normal**, or to have a normal distribution, with **mean** $m \in \mathbb{R}$ and **variance** $\sigma^2 > 0$ if

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{|x - m|^2}{2\sigma^2}\right). \tag{5.31}$$

  We denote $\mathcal{N}(m, \sigma^2)$ the set of all such random variables. A variable $X \in \mathcal{N}(0, 1)$ is called a **standard** normal random variable. The cumulative distribution of standard normal random variables is denoted by $\Phi(x)$ and is called the **standard normal distribution**, i.e.,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}y^2}\, dy.$$

- A random variable $X : \Omega \to \mathbb{R}$ is said to be **exponential**, or to have an exponential distribution, with **intensity** $\lambda > 0$ if

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

  We denote $\mathcal{E}(\lambda)$ the set of all exponential random variables with intensity λ.

Definition 5.17. *Two random variables $X, Y : \Omega \to \mathbb{R}$ are said to have the* **joint probability density** $f_{X,Y} : \mathbb{R}^2 \to [0, \infty)$*, if*

$$\mathbb{P}(X \in A, Y \in B) = \int_A \int_B f_{X,Y}(x, y)\, dx\, dy,$$

for all regular sets $A, B \subseteq \mathbb{R}$.

The generalization of the previous definition to n variables is straightforward. Note that if $f_{X,Y}$ is a joint probability density, then

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f_{X,Y}(x, y)\, dx\, dy = 1.$$

Example: Jointly normally distributed random variables. Let $m \in \mathbb{R}^2$ and $C = (C_{ij})_{i,j=1,2}$ be a symmetric, positive definite $2{\times}2$ matrix. Two random variables $X_1, X_2 : \Omega \to \mathbb{R}$ are said to be jointly normally distributed with mean m and **covariance matrix** C if they admit the joint density

$$f_{X_1,X_2}(x) = \frac{1}{\sqrt{(2\pi)^2 \det C}} \exp\left(-\frac{1}{2}(x - m) \cdot C^{-1} \cdot (x - m)^T \right), \quad \text{for all } x = (x_1, x_2) \in \mathbb{R}^2, \tag{5.32}$$

where " $\cdot$ " denotes the row-by-column product of matrices. This definition extends straightforwardly to n variables.

The following theorem, which we give without proof, shows that the probability density, when it exists, provides all the relevant statistical information on a random variable.

Theorem 5.6. *The following holds for all sufficiently regular*[†] *functions $g : \mathbb{R} \to \mathbb{R}$:*

(i) Let $X : \Omega \to \mathbb{R}$ be a random variable with density f_X. Then for all regular sets $A \subseteq \mathbb{R}$,

$$\mathbb{P}(g(X) \in A) = \int_{x : g(x) \in A} f_X(x)\, dx,$$

which extends (5.8) *to general probability spaces.*

(ii) Let $X : \Omega \to \mathbb{R}$ be a random variable with density f_X. Then

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(y) f_X(y)\, dy,$$

which extends (5.13) *to general probability spaces.*

[†]In particular, for all functions g such that the integrals in the theorem are well-defined.

(iii) Let $X, Y : \Omega \to \mathbb{R}$ be random variables with joint density $f_{X,Y}$. Then

$$\mathbb{E}[XY] = \int_{\mathbb{R}^2} xy f_{X,Y}(x,y)\,dx\,dy,$$

which extends (5.18) *to general probability spaces.*

As an example of application of (i), suppose $X \in \mathcal{N}(0,1)$. Then

$$\mathbb{P}(X^2 \le 1) = \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} e^{-\frac{x^2}{2}}\,dx \approx 0.683,$$

that is to say, a standard normal random variable has about 68,3% probability to take value on the interval $(-1,1)$. Let us see some further applications of Theorem 5.6. By (ii), the expectation and the variance of a random variable X with density f_X are given by

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x)\,dx, \quad \mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \int_{\mathbb{R}} x^2 f_X(x)\,dx - \left(\int_{\mathbb{R}} x f_X(x)\,dx\right)^2.$$

These formulas generalize (5.11) and (5.15) to random variables on general sample spaces which admit a density distribution. In particular, for normal variables we obtain

$$X \in \mathcal{N}(m,\sigma^2) \implies \mathbb{E}[X] = m, \quad \mathrm{Var}[X] = \sigma^2. \tag{5.33}$$

Exercise 5.22. *Prove* (5.33). *Compute the expectation and the variance of exponential random variables.*

By (iii) of Theorem 5.6, if X_1, X_2 have the joint density f_{X_1,X_2}, then

$$\mathrm{Cov}(X_1,X_2) = \int_{\mathbb{R}^2} x_1 x_2 f_{X_1,X_2}(x_1,x_2)\,dx_1\,dx_2 - \mathbb{E}[X_1]\mathbb{E}[X_2],$$

which generalizes (5.19). In particular, if X_1, X_2 are jointly normal distributed with mean $m \in \mathbb{R}^2$ and covariance matrix $C = (C_{ij})_{i,j=1,2}$, we find

$$m = (m_1, m_2), \qquad C_{ij} = \mathrm{Cov}(X_i, X_j). \tag{5.34}$$

Exercise 5.23. *Prove* (5.34).

The next and last thing we need to know about general probability spaces is how to determine when two random variables are independent. To this regard we have the following theorem.

Theorem 5.7. *The following holds.*

(i) If two random variables X, Y admit densities f_X, f_Y and are independent, then they admit the joint density

$$f_{X,Y}(x,y) = f_X(x) f_Y(y).$$

(ii) If two random variables X, Y admit a joint density $f_{X,Y}$ of the form

$$f_X(x,y) = u(x)v(y),$$

for some functions $u, v : \mathbb{R} \to [0,\infty)$, then X, Y are independent and admit densities f_X, f_Y given by

$$f_X(x) = cu(x), \quad f_Y(y) = \frac{1}{c}v(y),$$

where

$$c = \int_{\mathbb{R}} v(x)\,dx = \Big(\int_{\mathbb{R}} u(y)\,dy\Big)^{-1}.$$

Proof. As to (i) we have

$$\begin{aligned}\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) &= \int_A f_X(x)\,dx \int_B f_Y(y)\,dy \\ &= \int_{A\times B} f_X(x) f_Y(y)\,dx\,dy.\end{aligned}$$

To prove (ii), we first write

$$\{X \in A\} = \{X \in A\} \cap \Omega = \{X \in A\} \cap \{Y \in \mathbb{R}\} = \{X \in A, Y \in \mathbb{R}\}.$$

Hence, by definition of joint density,

$$\mathbb{P}(X \in A) = \int_A \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = \int_A u(x)\,dx \int_{\mathbb{R}} v(y)\,dy = \int_A cu(x)\,dx$$

where $c = \int_{\mathbb{R}} v(x)\,dx$. Thus X admits the density $f_X(x) = c\,u(x)$. At the same fashion one proves that Y admits the density $f_Y(y) = d\,v(y)$, where $d = \int_{\mathbb{R}} u(y)dy$. Since

$$1 = \int_{\mathbb{R}} \int_{\mathbb{R}} f_{X,Y}(x,y)\,dx\,dy = \int_{\mathbb{R}} u(x)\,dx \int_{\mathbb{R}} v(y)\,dy = d\,c,$$

then $d = 1/c$. It remains to prove that X, Y are independent. This follows by

$$\begin{aligned}\mathbb{P}(X \in A, Y \in B) &= \int_A \int_B f_{X,Y}(x,y)\,dx\,dy = \int_A u(x)\,dx \int_B v(y)\,dy \\ &= \int_A cu(x)\,dx \int_B \frac{1}{c}v(y)\,dy = \int_A f_X(x)\,dx \int_B f_Y(y)\,dy \\ &= \mathbb{P}(X \in A)\mathbb{P}(Y \in B).\end{aligned}$$

□

Exercise 5.24. *Let $X \in \mathcal{N}(0,1)$ and $Y \in \mathcal{E}(1)$ be independent. Compute $\mathbb{P}(X \le Y)$.*

As an application of Theorem 5.7 we now prove the following result.

Theorem 5.8. *Let $X_1, X_2 : \Omega \to \mathbb{R}$ be normal random variables which are jointly normally distributed. Then the following properties are equivalent:*

(a) X_1, X_2 *are independent;*

(b) X_1, X_2 *are uncorrelated.*

Proof. (a)$\Longrightarrow$(b) is always true, for all random variables. As to the implication (b)$\Longrightarrow$(a), by (5.34) we have $C_{12} = C_{21} = 0$. Substituting in (5.32) we obtain the f_{X_1,X_2} has the form $f_{X1,X_2}(x_1, x_2) = u(x_1)v(x_2)$, and so the claim follows by (ii) of Theorem 5.7. □

5.6 The geometric Brownian motion

Let Ω be a sample space and $T > 0$. Recall that a stochastic process is a one parameter family $\{X(t)\}_{t\in[0,T]}$ of random variables $X(t) : \Omega \to \mathbb{R}$. We denote by $X(t, \omega)$ the value attained by the random variable $X(t)$ on the sample point $\omega \in \Omega$ (i.e., $X(t, \omega) = X(t)(\omega)$). Moreover, for each $\omega \in \Omega$ *fixed*, the function $t \to X(t, \omega)$ defined on $[0, T]$ is called a *path* of the stochastic process. In the following we refer to the parameter t as the time variable, since this is what it represents in the applications that we have in mind.

Definition 5.18. *A Brownian motion, or Wiener process, is a stochastic process* $\{W(t)\}_{t\in[0,T]}$ *with the following properties:*

1. *The paths are continuous*[‡]*, i.e.,* $t \to W(t, \omega)$ *is a continuous function on* $[0, T]$*, and* $W(0, \omega) = 0$*, for all* $\omega \in \Omega$*;*
2. *For all* $0 = t_0 < t_1 < \cdots < t_m = T$*, the increments*

$$W(t_1) = W(t_1) - W(t_0),\ W(t_2) - W(t_1), \ldots, W(T) - W(t_{m-1})$$

are independent random variables and

$$\mathbb{E}[W(t_{i+1}) - W(t_i)] = 0, \quad \mathrm{Var}[W(t_{i+1}) - W(t_i)] = t_{i+1} - t_i, \quad \textit{for all } i = 0, \ldots M - 1;$$

3. *The increments are normally distributed, that is to say, for all* $0 \le s < t \le T$ *and* $x \in \mathbb{R}$*:*

$$\mathbb{P}(W(t) - W(s) \in A) = \frac{1}{\sqrt{2\pi(t-s)}} \int_A e^{-\frac{y^2}{2(t-s)}}\, dy,$$

for all regular sets $A \in \mathbb{R}$*.*

It can be shown that Brownian motions, i.e., stochastic processes with the properties listed above, exist. One way to construct a Brownian motion is suggested in Exercise 5.15, namely by running a properly rescaled symmetric random walk for infinitely many steps. A

[‡]More precisely, the paths are continuous with probability 1.

formal construction of Brownian motions is technically quite difficult and beyond the purpose of this note. It is useful to think of a Brownian motion as time-continuum generalization of the symmetric random walk. Note to this regard that the increments of a symmetric random walk also satisfy the independence property 2 in Definition 5.18.

Exercise 5.25. *Let* $\{W(t)\}_{t\in[0,T]}$ *be a Brownian motion. Show that* $\mathrm{Cov}[W(s), W(t)] = \min(s,t)$*, for all* $s, t \in [0,T]$*. (Compare this with Exercise 5.15.)*

Can we use a Brownian motion to model the evolution in time of a stock price, i.e., can we set $S(t) = W(t)$? The answer is of course *no*, because $W(t)$ can also take negative values. In fact, since $W(t)$ is normally distributed, then it has a non-zero probability to take value in any interval within the real line. However this deficiency is easily corrected.

Definition 5.19. *Let* $\{W(t)\}_{t\in[0,T]}$ *be a Brownian motion,* $\alpha \in \mathbb{R}$ *and* $\sigma > 0$*. The positive stochastic process* $\{S(t)\}_{t\in[0,T]}$*,*

$$S(t) = S(0)e^{\alpha t+\sigma W(t)} \tag{5.35}$$

is called a **geometric Brownian motion**.

As the notation used in (5.35) suggests, we shall use geometric Brownian motions to model the dynamics of stock prices in the time-continuum case. Within this application, the parameter α is called the **expected log-return** of the stock, while σ is called the **volatility** of the stock. To justify the terminology for α, notice that, by (ii) of Theorem 5.6, and since $W(t) \in \mathcal{N}(0,t)$,

$$\mathbb{E}[\log S(t) - \log S(0)] = \mathbb{E}[\alpha t + \sigma W(t)] = \alpha t + \sigma\mathbb{E}[W(t)] = \alpha t.$$

Hence

$$\alpha = \frac{1}{t}\mathbb{E}[\log S(t) - \log S(0)],$$

i.e., α is the annualized value of the expected log-return of the stock (assuming of course that t is measured in fraction of years). The reason for calling σ the volatility of the stock is given in the following theorem.

Theorem 5.9. *Suppose that at time* $t = 0$ *it is assumed that the stock price is described by a geometric Brownian motion in the interval* $[0,T]$*:*

$$S(t) = S(0)e^{\alpha t+\sigma W(t)}, \quad t \in [0,T].$$

Given any arbitrary subinterval $[t_0, t] \subset [0,T]$ *with length* $\tau = t - t_0$*, define the random variable*

$$\sigma_\tau^2(t) = \frac{1}{h(n-1)}\sum_{i=1}^{n}(R_i - \bar{R})^2, \tag{5.36}$$

where $t_0 < t_1 < t_2 < \cdots < t_n = t$ *is a partition of* $[t_0, t]$ *with* $h = (t_i - t_{i-1})$ *and* $R_i, \bar{R}$ *are given by* (1.2), (1.5)*. Then*

$$\mathbb{E}[\sigma_\tau(t)^2] = \sigma^2.$$

In other words, σ^2 *is the expected value of the* τ*-historical variance at any time* $t \in [0,T]$*.*

Proof. In (5.36) we replace

$$\bar{R} = \frac{1}{n}(\alpha(t - t_0) + \sigma(W(t) - W(t_0)) = \alpha h + \frac{\sigma}{n}(W(t) - W(t_0)), \text{ and}$$

$$R_i = \alpha(t_i - t_{i-1}) + \sigma(W(t_i) - W(t_{i-1}) = \alpha h + \sigma(W(t_i) - W(t_{i-1}).$$

So doing we obtain

$$\begin{aligned}
\sigma_\tau^2(t) &= \frac{\sigma^2}{h(n-1)} \sum_{i=1}^{n} \left[(W(t_i) - W(t_{i-1}) - \frac{1}{n}(W(t) - W(t_0)) \right]^2 \\
&= \frac{\sigma^2}{h(n-1)} \left[\sum_{i=1}^{n} (W(t_i) - W(t_{i-1}))^2 - \frac{1}{n}(W(t) - W(t_0))^2 \right].
\end{aligned}$$

Taking the expectation and using that $\mathbb{E}[(W(t) - W(s))^2] = \mathrm{Var}[W(t) - W(s)] = t - s$ we obtain

$$\mathbb{E}[\sigma_\tau^2] = \frac{\sigma^2}{h(n-1)} \left[\sum_{i=1}^{n} (t_i - t_{i-1}) - \frac{1}{n}(t - t_0) \right] = \frac{\sigma^2}{h(n-1)}(t - t_0)\left(1 - \frac{1}{n}\right) = \sigma^2.$$

□

Theorem 5.10. *The density of the random variable $S(t)$ is given by*

$$f_{S(t)}(x) = \frac{\mathbb{I}_{x>0}}{\sqrt{2\pi\sigma^2 t}} \frac{1}{x} \exp\left(- \frac{(\log x - \log S(0) - \alpha t)^2}{2\sigma^2 t} \right), \tag{5.37}$$

where $\mathbb{I}_{x>0}$ is the indicator function of the set $\{x > 0\}$.

Proof. The density of $S(t)$ is given by

$$f_{S(t)}(x) = \frac{d}{dx} F_{S(t)}(x),$$

where $F_{S(t)}$ is the distribution of $S(t)$, i.e.,

$$F_{S(t)}(x) = \mathbb{P}(S(t) \le x).$$

Clearly, $f_{S(t)}(x) = F_{S(t)}(x) = 0$, for $x \le 0$. For $x > 0$ we use that

$$S(t) \le x \quad \text{if and only if} \quad W(t) \le \frac{1}{\sigma}\left(\log \frac{x}{S(0)} - \alpha t\right) := A(x).$$

Thus

$$\mathbb{P}(S(t) \le x) = \mathbb{P}(-\infty < W(t) \le A(x)) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{A(x)} e^{-\frac{y^2}{2t}}\, dy,$$

where for the second equality we used that $W(t) \in \mathcal{N}(0,t)$. Hence

$$f_{S(t)}(x) = \frac{d}{dx}\left(\frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{A(x)} e^{-\frac{y^2}{2t}}\,dy\right) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{A(x)^2}{2t}}\frac{dA(x)}{dx},$$

for $x > 0$, that is

$$f_{S(t)}(x) = \frac{1}{\sqrt{2\pi\sigma^2 t}}\frac{1}{x}\exp\left\{-\frac{(\log x - \log S(0) - \alpha t)^2}{2\sigma^2 t}\right\}, \quad x > 0.$$

The proof is complete. $\square$

Exercise 5.26. *Express $\mathbb{P}(a < S(t) < b)$ in terms of the standard normal distribution $\Phi(x)$.*

Next we study the relation between the discrete binomial asset pricing model and the time-continuous geometric Brownian motion model. Consider an interval of time $[0,t]$ and a partition

$$0 = t_0 < t_1 < \cdots < t_N = t, \quad t_{i+1} - t_i = h > 0, \quad \text{for all } i = 0, \dots, N-1.$$

We may define a binomial model on this partition by setting, for $u > d$,

$$S(t_i) = \begin{cases} S(t_{i-1})e^u & \text{with probability } 1/2 \\ S(t_{i-1})e^d & \text{with probability } 1/2 \end{cases}.$$

Now, consider the N independent and identically distributed random variables $X_1, \dots X_N$ defined as

$$X_i = \begin{cases} 1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2 \end{cases}.$$

We can rewrite the binomial model above as

$$S(t_i) = S(t_{i-1})\exp\left[\left(\frac{u+d}{2}\right) + \left(\frac{u-d}{2}\right)X_i\right].$$

Iterating the previous identity we obtain

$$S(t) = S(0)\exp\left[N\left(\frac{u+d}{2}\right) + \left(\frac{u-d}{2}\right)M_N\right],$$

where $M_N = X_1 + \dots X_N$ is the symmetric random walk stochastic process. Substituting $N = \frac{t}{h}$, the previous expression becomes

$$S(t) = S(0)\exp\left[\left(\frac{u+d}{2h}\right)t + \left(\frac{u-d}{2\sqrt{h}}\right)\sqrt{h}M_{t/h}\right]. \tag{5.38}$$

As indicated in Exercise 5.15, we have

$$\sqrt{h}M_{t/h} \approx W(t), \quad \text{for } h \text{ small}.$$

Therefore, with the identifications,

$$\left(\frac{u-d}{2\sqrt{h}}\right) \sim \sigma, \quad \frac{u+d}{2h} \sim \alpha, \tag{5.39}$$

we can approximate (5.38) by the geometric Brownian motion

$$S(t) \approx S(0) \exp\left[\alpha t + \sigma W(t)\right].$$

This argument can be made rigorous. Namely, using more advanced tools in probability theory (such as the Central Limit Theorem) it can be shown that the the binomial asset pricing model converges (in probability) to the geometric Brownian motion as the time step $h \to 0$ and $u, d \to 0$ such that

$$\left(\frac{u-d}{2\sqrt{h}}\right) \to \sigma, \quad \frac{u+d}{2h} \to \alpha.$$

Chapter 6

Black-Scholes options pricing Theory

This final chapter is concerned with the most famous of all models in options pricing theory, namely the Black-Scholes model, which first appeared in the seminal paper [1]. The way we introduce it in this chapter is very different from the original argument in [1]. Our strategy is to derive the Black-Scholes model from the binomial options pricing model in the time-continuum limit. More precisely we let the number of steps N in the binomial model goes to infinity and, at the same time, the time interval h between each step tends to zero, keeping constant the time of maturity $T = Nh$.

6.1 The Black-Scholes formula

In Section 5.4 we have seen that when the price of the underlying stock is given by the binomial model $S(0) = S_0 > 0$,

$$S(j) = \begin{cases} S(j-1)e^u, & \text{with probability } p \\ S(j-1)e^d, & \text{with probability } p = 1-p \end{cases}, \quad j \in \mathcal{I} = \{1, \dots, N\},$$

and the value of the risk-free asset is

$$B(j) = B_0 e^{rj}, \quad j \in \mathcal{I},\ r \in (d, u),$$

then the binomial fair price at time $t = 0$ of the standard European derivative with pay-off $Y = g(S(N))$ can be written as $\Pi_Y(0) = e^{-rN}\mathbb{E}_q[g(S(N))]$, where $\mathbb{E}_q[\cdot]$ denotes the expectation in the risk-neutral probability measure; see (5.29). As a preparation to derive the Black-Scholes model in the time-continuum limit, we first rewrite $\Pi_Y(0)$ in terms of the distribution of $S(N)$. Using the definition of risk-neutral probabiliy measure in the N-coin toss space we have

$$\begin{aligned}\Pi_Y(0) = e^{-rN}\mathbb{E}_q[g(S(N))] &= e^{-rN}\sum_{\omega\in\Omega_N} q^{N_H(\omega)}(1-q)^{N_T(\omega)} g(S(N)) \\ &= e^{-rN}\sum_{k=0}^{N}\binom{N}{k} q^k (1-q)^{N-k} g(s_k),\end{aligned}$$

where $s_0, \dots, s_N$ are the possible values of $S(N)$, i.e.,

$$s_k = S(0)\exp(ku + (N-u)d), \quad k = 0 \dots N,$$

and

$$q = \frac{e^r - e^d}{e^u - e^d}. \tag{6.1}$$

Note the important formula

$$k = \frac{1}{u-d}\left(\log\left(\frac{s_k}{S(0)}\right) - Nd\right). \tag{6.2}$$

We can rewrite $\Pi_Y(0)$ as

$$\Pi_Y(0) = e^{-rN}\sum_{k=0}^{N}\left(\frac{q}{p}\right)^k\left(\frac{1-q}{1-p}\right)^{N-k}\binom{N}{k}p^k(1-p)^{N-k}g(s_k).$$

Recalling that the distribution of $S(N)$ is given by $\binom{N}{k}p^k(1-p)^{N-k}$ (see Section 5.4), we obtain

$$\Pi_Y(0) = e^{-rN}\sum_{k=0}^{N} Z_N(s_k)f_{S(N)}(s_k)g(s_k), \tag{6.3}$$

where

$$Z_N(s_k) = \left(\frac{q}{p}\right)^k\left(\frac{1-q}{1-p}\right)^{N-k} = \left(\frac{q}{p}\right)^{\frac{1}{u-d}\left[\log\left(\frac{s_k}{S(0)}\right)-Nd\right]}\left(\frac{1-q}{1-p}\right)^{N-\frac{1}{u-d}\left[\log\left(\frac{s_k}{S(0)}\right)-Nd\right]}, \tag{6.4}$$

the second equality in (6.4) following by (6.2). From now on we assume $p = 1/2$ for simplicity. We rewrite (6.3) as

$$\Pi_Y(0) = e^{-rN}\mathbb{E}_q[g(S(N))] = e^{-rN}\sum_{x\in \mathrm{Im}S(N)}^{N} Z_N(x)f_{S(N)}(x)g(x), \tag{6.5}$$

where $\mathrm{Im}S(N)$ is the image of $S(N)$, $f_{S(N)}$ is the probability distribution of $S(N)$, and

$$Z_N(x) = (2q)^{\frac{1}{u-d}\left[\log\left(\frac{x}{S(0)}\right)-Nd\right]}(2(1-q))^{N-\frac{1}{u-d}\left[\log\left(\frac{x}{S(0)}\right)-Nd\right]}. \tag{6.6}$$

We have also seen in Section 5.6 that, in a suitable limit, the binomial stock price converges to the geometric Brownian motion. We shall obtain the Black-Scholes formula in the same limit applied to (6.3). To this purpose, let $T > 0$ and consider the partition of the interval $[0, T]$ given by

$$0 = t_0 < t_1 < \cdots < t_N = T, \quad t_j - t_{j-1} = h > 0, \quad j \in \mathcal{I}.$$

Let us define the binomial model $S(0) = S_0 > 0$,

$$S(t_j) = \begin{cases} S(t_{j-1})e^u, & \text{with probability } p = 1/2 \\ S(t_{j-1})e^d, & \text{with probability } p = 1/2 \end{cases}, \quad j \in \mathcal{I}, \tag{6.7}$$

where

$$u = \alpha h + \sigma\sqrt{h}, \quad d = \alpha h - \sigma\sqrt{h}.$$

Note that

$$\sigma = \frac{u-d}{2\sqrt{h}}, \quad \alpha = \frac{u+d}{2h},$$

hence, as pointed out in Section 5.6, in the limit $h \to 0$ the price of the stock follows the geometric Brownian motion $S(t) = S(0)\exp(\alpha t + \sigma W(t))$. Moreover $B(t_j) = B_0 e^{rt_j} = B_0 e^{rhj}$. Hence the pair $(\tilde{S}(j) = S(t_j), \tilde{B}(j) = B(t_j))$ is equivalent to a binomial market with parameters (u, d, rh) and $p = 1/2$. In particular the parameter q defined by (6.1) becomes

$$q = \frac{e^{rh} - e^{\alpha h - \sigma\sqrt{h}}}{e^{\alpha h + \sigma\sqrt{h}} - e^{\alpha h - \sigma\sqrt{h}}} \tag{6.8}$$

and since

$$\alpha h - \sigma\sqrt{h} < rh < \alpha h + \sigma\sqrt{h}$$

holds for h small, then we can assume that $q \in (0,1)$ and that the market is arbitrage-free. Therefore the initial price of the European derivative with pay-off $Y = g(\tilde{S}(N)) = g(S(T))$ is given by (6.5). Using $Nh = T$, we rewrite (6.5) as

$$\Pi_Y(0) = e^{-rhN} \sum_{x \in \mathrm{Im}\tilde{S}(N)}^{N} Z_N(x) f_{\tilde{S}(N)}(x) g(x) = e^{-rT} \sum_{x \in \mathrm{Im}S(T)}^{N} Q_h(x) f_{S(T)}(x) g(x), \tag{6.9}$$

where $Q_h(x) = Z_{T/N}(x)$; by (6.6), (6.8) and the definitions of u, d we have $Q_h(x) = \eta_h(x)\xi_h(h)$, where

$$\eta_h(x) = \left(2\frac{e^{rh} - e^{\alpha h - \sigma\sqrt{h}}}{e^{\alpha h + \sigma\sqrt{h}} - e^{\alpha h - \sigma\sqrt{h}}}\right)^{\frac{1}{2\sigma\sqrt{h}}\left(\log\frac{x}{S(0)} - \alpha T\right) + \frac{T}{2h}},$$

$$\xi_h(x) = \left(2\frac{e^{\alpha h + \sigma\sqrt{h}} - e^{rh}}{e^{\alpha h + \sigma\sqrt{h}} - e^{\alpha h - \sigma\sqrt{h}}}\right)^{-\frac{1}{2\sigma\sqrt{h}}\left(\log\frac{x}{S(0)} - \alpha T\right) + \frac{T}{2h}}.$$

Theorem 6.1. *The following holds:*

$$\lim_{h \to 0} Q_h(x) = e^{-\frac{(\alpha + \frac{\sigma^2}{2} - r)^2 T}{2\sigma^2}} e^{-\frac{1}{\sigma^2}\left(\log\frac{x}{S(0)} - \alpha T\right)\left(\alpha + \frac{\sigma^2}{2} - r\right)} := Q(x).$$

Exercise 6.1. *Prove the theorem.*

We are now in the position to justify the definition of the Black-Scholes price of the derivative. Recalling the analogies between finite and uncountable probability spaces pointed out in Theorem 5.6, the natural generalization of (6.9) in the time-continuum case is

$$\Pi_Y(0) = e^{-rT}\int_{\mathbb{R}} Q(x) f_{S(T)}(x) g(x)\, dx,$$

where $f_{S(T)}$ is now the probability density of the random variable $S(T) = S(0)e^{\alpha T + \sigma W(T)}$, i.e.,

$$f_{S(T)}(x) = \frac{\mathbb{I}_{x>0}}{\sqrt{2\pi\sigma^2 T}}\frac{1}{x}\exp\left(-\frac{(\log\frac{x}{S(0)} - \alpha T)^2}{2\sigma^2 T}\right), \tag{6.10}$$

see (5.37). Explicitly, $\Pi_Y(0)$ is given by

$$\Pi_Y(0) = \frac{e^{-rT}}{\sqrt{2\pi\sigma^2 T}}\int_0^\infty e^{-\frac{(\alpha+\frac{\sigma^2}{2}-r)^2 T}{2\sigma^2}} e^{-\frac{1}{\sigma^2}(\log\frac{x}{S(0)}-\alpha T)(\alpha+\frac{\sigma^2}{2}-r)} e^{-\frac{\left(\log\frac{x}{S(0)}-\alpha T\right)^2}{2\sigma^2 T}}\frac{g(x)}{x}\, dx.$$

Now, the exponents in the exponential functions inside the integral form a perfect square:

$$\begin{aligned}\Pi_Y(0) &= \frac{e^{-rT}}{\sqrt{2\pi\sigma^2 T}}\int_0^\infty e^{-\frac{1}{2\sigma^2 T}\left[\left(\log\frac{x}{S(0)}-\alpha T\right)+\left(\alpha+\frac{\sigma^2}{2}-r\right)T\right]^2}\frac{g(x)}{x}\, dx\\ &= \frac{e^{-rT}}{\sqrt{2\pi\sigma^2 T}}\int_0^\infty e^{-\frac{1}{2\sigma^2 T}\left[\log\frac{x}{S(0)}+\left(\frac{\sigma^2}{2}-r\right)T\right]^2}\frac{g(x)}{x}\, dx.\end{aligned}$$

Note that the dependence on α is gone! Finally, the change of variable

$$y = \frac{1}{\sigma\sqrt{T}}\left(\log\frac{x}{S(0)} + \left(\frac{\sigma^2}{2} - r\right)T\right)$$

gives

$$\Pi_Y(0) = \frac{e^{-rT}}{\sqrt{2\pi}}\int_{\mathbb{R}} g\left(S(0)e^{(r-\frac{\sigma^2}{2})T}e^{\sigma\sqrt{T}y}\right)e^{-\frac{y^2}{2}}\, dy.$$

The Black-Scholes price $\Pi_Y(t_0)$ at a generic time $t_0 \in [0,T]$ is obtained by assuming that the present time is $t = t_0$ (instead of $t = 0$) and replacing the expiration date T with the time left to maturity, i.e., $\tau = T - t_0$. The latter is of course quite reasonable and can be justified by the same argument applied to derive the formula for $\Pi_Y(0)$. We are led to introduce the following definition.

Definition 6.1. *Consider a European derivative with pay-off $Y = g(S(T))$ at the maturity time $T > 0$. Assume that the price of the underlying stock is given by the geometric Brownian motion $\{S(t)\}_{t\in[0,T]}$, $S(t) = S(0)e^{\alpha t + \sigma W(t)}$. The Black-Scholes price $\Pi_Y(t)$ of the derivative at time $t \in [0,T]$ is*

$$\Pi_Y(t) = v(t, S(t)),$$

where

$$v(t,x) = \frac{e^{-r\tau}}{\sqrt{2\pi}}\int_{\mathbb{R}} g\left(xe^{(r-\frac{\sigma^2}{2})\tau}e^{\sigma\sqrt{\tau}\,y}\right)e^{-\frac{y^2}{2}}\, dy, \quad \tau = T - t. \tag{6.11}$$

Remark 6.1. Of course we are tacitly assuming that the pay-off function g is such that the integral in the right hand side of (6.11) is well-defined. Note also that the Black-Scholes price at time t is a deterministic function of $S(t)$ and thus can be computed with the information available at time t.

Remark 6.2. The fact that the Black-Scholes price is independent of the expected log-return α of the stock is consistent with the numerical observation made in Section 3.3 that the binomial price of options is weakly dependent on the parameter α. In the time-continuum limit, this dependence disappears completely.

Note that the definition of the Black-Scholes price can be written also in the probabilistic form

$$\Pi_Y(t) = e^{-r\tau}\mathbb{E}[g(S(t)e^{(r-\frac{\sigma^2}{2})\tau+\sigma(W(T)-W(t))}]. \tag{6.12}$$

In fact, since $G = (W(T) - W(t))/\sqrt{T-t} \in \mathcal{N}(0,1)$, then

$$\begin{aligned}\mathbb{E}[g(S(t)e^{(r-\frac{\sigma^2}{2})\tau+\sigma(W(t)-W(t))}] &= \mathbb{E}[g(S(t)e^{(r-\frac{\sigma^2}{2})\tau+\sigma\sqrt{\tau}G}] \\ &= \frac{e^{-r\tau}}{\sqrt{2\pi}}\int_{\mathbb{R}} g\left(S(t)e^{(r-\frac{\sigma^2}{2})\tau}e^{\sigma\sqrt{\tau}\,y}\right)e^{-\frac{y^2}{2}}\,dy,\end{aligned}$$

where we used Theorem 5.6.

It can be shown that, even in the time-continuum case, the Black-Scholes price at time t equals the value of self-financing portfolios hedging the derivative and thus Definition 6.1 can be justified as in the time-discrete case. However this approach requires the use of stochastic calculus and is therefore beyond the purpose of the present note. Moreover it can be shown that an hedging portfolio process $\{(h_S(t), h_B(t))\}_{t\in[0,T]}$ for the derivative priced according to the Black-Scholes formula is given by

$$h_S(t) = \Delta(t, S(t)), \quad \Delta(t,x) = \partial_x v(t,x), \quad h_B(t) = \frac{1}{B(t)}(\Pi_Y(t) - h_S(t)S(t)). \tag{6.13}$$

Note that this portfolio process is predictable. For a heuristic derivation of (6.13), and for a different justification of the Black-Scholes price, see [3]. In the next two sections we compute the Black-Scholes price of some simple derivatives.

6.2 Black-Scholes price of European call and put options

In this section we focus the discussion on call/put options. We thereby assume that the pay-off of the derivative is given by

$$Y = (S(T) - K)_+,\ \text{i.e., } Y = g(S(T)),\ g(z) = (z-K)_+, \quad \text{for a call option,}$$

$$Y = (K - S(T))_+,\ \text{i.e., } Y = g(S(T)),\ g(z) = (K-z)_+, \quad \text{for a put option.}$$

The function v given by (6.11) will be denoted by C, for a call option, and by P, for a put option.

Theorem 6.2. *The Black-Scholes price at time t of a European call option with strike price $K > 0$ and maturity $T > 0$ is given by $C(t, S(t))$, where*

$$C(t,x) = x\Phi(d_1) - Ke^{-r\tau}\Phi(d_2), \tag{6.14a}$$

$$d_2 = \frac{\log\left(\frac{x}{K}\right) + \left(r - \frac{1}{2}\sigma^2\right)\tau}{\sigma\sqrt{\tau}}, \quad d_1 = d_2 + \sigma\sqrt{\tau}, \tag{6.14b}$$

and where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}y^2}\,dy$ is the standard normal distribution. The Black-Scholes price of the corresponding put option is given by $P(t, S(t))$, where

$$P(t,x) = \Phi(-d_2)Ke^{-r\tau} - \Phi(-d_1)x. \tag{6.15}$$

Moreover the **put-call parity identity** *holds:*

$$C(t,S(t)) - P(t,S(t)) = S(t) - Ke^{-r\tau}. \tag{6.16}$$

Proof. We derive the Black-Scholes price of call options only, the argument for put options being similar (see Exercise 6.2 below). We substitute $g(z) = (z-K)_+$ into the right hand side of (6.11) and obtain

$$C(t,x) = \frac{e^{-r\tau}}{\sqrt{2\pi}}\int_{\mathbb{R}} \left(xe^{(r-\frac{1}{2}\sigma^2)\tau}e^{\sigma\sqrt{\tau}y} - K\right)_+ e^{-\frac{y^2}{2}}\,dy.$$

Now we use that $xe^{(r-\frac{1}{2}\sigma^2)\tau}e^{\sigma\sqrt{\tau}y} > K$ if and only if $y > -d_2$. Hence

$$C(t,x) = \frac{e^{-r\tau}}{\sqrt{2\pi}}\left[\int_{-d_2}^{\infty} xe^{(r-\frac{1}{2}\sigma^2)\tau}e^{\sigma\sqrt{\tau}y}e^{-\frac{y^2}{2}} - K\int_{-d_2}^{\infty} e^{-\frac{y^2}{2}}\,dy\right].$$

Using $-\frac{1}{2}y^2 + \sigma\sqrt{\tau}y = -\frac{1}{2}(y-\sigma\sqrt{\tau})^2 + \frac{\sigma^2}{2}\tau$ and changing variable in the integrals we obtain

$$\begin{aligned}
C(t,x) &= \frac{e^{-r\tau}}{\sqrt{2\pi}}\left[xe^{r\tau}\int_{-d_2}^{\infty} e^{-\frac{1}{2}(y-\sigma\sqrt{\tau})^2}\,dy - K\int_{-d_2}^{\infty} e^{-\frac{y^2}{2}}\,dy\right] \\
&= \frac{e^{-r\tau}}{\sqrt{2\pi}}\left[xe^{r\tau}\int_{-\infty}^{d_2+\sigma\sqrt{\tau}} e^{-\frac{1}{2}y^2}\,dy - K\int_{-\infty}^{d_2} e^{-\frac{y^2}{2}}\,dy\right] \\
&= s\Phi(d_1) - Ke^{-r\tau}\Phi(d_2).
\end{aligned}$$

As to the put-call parity, we have

$$\begin{aligned}
C(t,x) - P(t,x) &= s\Phi(d_1) - Ke^{-r\tau}\Phi(d_2) - \Phi(-d_2)Ke^{-r\tau} + s\Phi(-d_1) \\
&= x(\Phi(d_1) + \Phi(-d_1)) - Ke^{-r\tau}(\Phi(d_2) + \Phi(-d_2)).
\end{aligned}$$

As $\Phi(z) + \Phi(-z) = 1$, the claim follows. ☐

Exercise 6.2. *Derive the Black-Scholes price $P(t, S(t))$ of European put options claimed in Theorem 6.2.*

Remark 6.3. The formulas (6.14)-(6.15) appeared for the first time in the seminal paper [1], where they were derived by a completely different argument than the one presented here.

Concerning the self-financing hedging portfolio for the call/put option, we have $h_S(t) = \partial_x C(t, S(t))$ for call options and $h_S(t) = \partial_x P(t, S(t))$ for put options, see (6.13), while the number of shares of the risk-free asset in the hedging portfolio is given by

$$h_B(t) = (C(t, S(t)) - S(t)\partial_x C(t, S(t)))/B(t), \quad \text{for call options,}$$

and

$$h_B(t) = (C(t, S(t)) - S(t)\partial_x P(t, S(t)))/B(t), \quad \text{for put options.}$$

Let us compute $\partial_x C$:

$$\partial_x C = \Phi(d_1) + x\Phi'(d_1)\partial_x d_1 - Ke^{-r\tau}\Phi'(d_2)\partial_x d_2.$$

As $\partial_x d_1 = \partial_x d_2 = \frac{1}{\sigma\sqrt{\tau}x}$, and $\Phi'(x) = e^{-\frac{1}{2}x^2}/\sqrt{2\pi}$, we obtain

$$\partial_x C = \Phi(d_1) + \frac{1}{\sigma\sqrt{2\pi\tau}}\left(e^{-\frac{1}{2}d_1^2} - \frac{K}{x}e^{-r\tau}e^{-\frac{1}{2}d_2^2}\right).$$

Replacing $d_1 = d_2 + \sigma\sqrt{\tau}$ we obtain

$$\partial_x C = \Phi(d_1) + \frac{e^{-\frac{1}{2}d_1^2}}{\sigma\sqrt{2\pi\tau}}\left(e^{-\frac{1}{2}\sigma^2\tau - d_2\sigma\sqrt{\tau}} - \frac{K}{x}e^{-r\tau}\right).$$

Using the definition of d_1, the term within round brackets in the previous expression is easily found to be zero, hence

$$\partial_x C = \Phi(d_1).$$

By the put-call parity we find also

$$\partial_x P = \Phi(d_1) - 1 = -\Phi(-d_1).$$

Note that $\partial_x C > 0$, while $\partial_x P < 0$. This agrees with the fact that call options are bought to protect a short position on the underlying stock, while put options are bought to protect a long position on the underlying stock.

The greeks

The Black-Scholes price of a call (or put) option derived in Theorem 6.2 depends on the price of the underlying stock, the time to maturity, the strike price, as well as on the (constant) market parameters r, σ (it does not depend on α). The partial derivatives of the price function c with respect to these variables are called **greeks**. We collect the most important ones (for call options) in the following theorem.

Theorem 6.3. *The price function C of call options satisfies the following:*

$$\Delta := \partial_x C = \Phi(d_1), \tag{6.17}$$

$$\Gamma := \partial_x^2 C = \frac{\phi(d_1)}{x\sigma\sqrt{\tau}}, \tag{6.18}$$

$$\rho := \partial_r C = K\tau e^{-r\tau}\Phi(d_2), \tag{6.19}$$

$$\Theta := \partial_t C = -\frac{x\phi(d_1)\sigma}{2\sqrt{\tau}} - rKe^{-r\tau}\Phi(d_2), \tag{6.20}$$

$$\nu := \partial_\sigma C = x\phi(d_1)\sqrt{\tau} \quad \textit{(called “vega”)}, \tag{6.21}$$

where $\phi(z) = \Phi'(z) = (\sqrt{2\pi})^{-1}e^{-\frac{z^2}{2}}$. In particular:

- $\Delta > 0$*, i.e., the price of a call is increasing on the price of the underlying stock;*
- $\Gamma > 0$*, i.e., the price of a call is convex on the price of the underlying stock;*
- $\rho > 0$*, i.e., the price of the call is increasing on the interest rate of the risk-free asset;*
- $\Theta < 0$*, i.e., the price of the call is decreasing in time;*
- $\nu > 0$*, i.e., the price of the call is increasing on the volatility of the stock.*

Exercise 6.3. *Use the put-call parity to derive the greeks of put options.*

Let us comment on the fact that vega is positive. It implies that the wish of an investor with a long position on a call option is that the volatility of the underlying stock increased. As usual, since this might not happen, the buyer of the call may incur in a loss if the stock volatility decreases (since the call option will loose value). This exposure to volatility can be secured by adding volatility swaps into the portfolio.

Exercise 6.4. *Prove that*

$$\lim_{\sigma\to 0^+} C(t,x) = (x - Ke^{-r\tau})_+, \quad \lim_{\sigma\to\infty} C(t,x) = x.$$

Compute also the following limits:

$$\lim_{K\to 0^+} C(t,x), \qquad \lim_{K\to+\infty} C(t,x), \qquad \lim_{T\to+\infty} C(t,x), \qquad \lim_{x\to 0^+} C(t,x).$$

Repeat all the above for put options.

6.3 The Black-Scholes price of other standard European derivatives

In this section we present a few more examples of standard European derivatives and compute their Black-Scholes price.

6.3.1 Binary call option

A **binary** (or **digital**) call option with strike K and maturity T pays-off the buyer if and only if $S(T) > K$. If the pay-off is a fixed amount of cash L, then the binary call option is said to be "cash-settled", while if the pay-off is the stock itself then the option is said to be "physically settled". Let us compute the Black-Scholes price of a cash-settled binary call option. The pay-off is $Y = LH(S(T)-K)$, where $H(z)$ is the Heaviside function. Replacing $g(z) = LH(z-K)$ into (6.11) we obtain

$$\begin{aligned} v(t,x) &= \frac{e^{-r\tau}}{\sqrt{2\pi}} L \int_{\mathbb{R}} H(xe^{(r-\frac{\sigma^2}{2})\tau} e^{\sigma\sqrt{\tau}\,y} - K) e^{-\frac{y^2}{2}}\, dy \\ &= \frac{e^{-r\tau}}{\sqrt{2\pi}} L \int_{-d_2}^{\infty} e^{-\frac{y^2}{2}}\, dy, \end{aligned}$$

where we recall that $d_2 = \frac{1}{\sigma\sqrt{\tau}}[\log(x/K) + (r - \frac{\sigma^2}{2})]$. With the change of variable $y \to -y$ we obtain

$$\Pi_Y(t) = e^{-r\tau} L\Phi(d_2).$$

Exercise 6.5. *Compute the Black-Scholes price of a physically-settled binary call.*

6.3.2 Butterfly options strategy

A **butterfly option** is a European style option with pay-off

$$Y = (S(T) - K + \Delta K)_+ - 2(S(T) - K)_+ + (S(T) - K - \Delta K)_+,$$

where $\Delta K, \Delta K > 0$. Note that the pay-off is positive, i.e., the option expires in the money, if and only if $S(T) \in (K - \Delta K, K + \Delta K)$. Hence the purpose of a butterfly option is to fix the price of the stock at time T within an interval. Moreover the pay-off Y is the value at time T of a portfolio containing the following assets:

- Long 1 call with a strike price $K - \Delta K$ and maturity T;
- Short 2 calls with a strike price K and maturity T;
- Long 1 call with a strike price $K + \Delta K$ and maturity T.

Therefore the Black-Scholes price of the butterfly option is given by

$$\Pi_Y(t) = C(t, S(t), K - \Delta K) - 2C(t, S(t), K, T) + C(t, S(t), K + \Delta K),$$

where we denoted by $C(t, S(t), K, T)$ the Black-Scholes price of a call with strike K and maturity T. This notation, and the analogous one $P(t, S(t), K, T)$ for put options, will be used in the rest of this section.

6.3.3 Chooser option

This is an example of "second derivative", i.e., of a financial derivative whose underlying is another derivative. More precisely, given $T_2 > T_1$, a **chooser option** with maturity T_1 is a contract which gives to the buyer the right to choose at time T_1 whether the derivative becomes a call or a put option with strike K and maturity T_2. Hence the pay-off at time T_1 for this derivative is

$$Y = \max(C(t_1, S(T_1), K, T_2), P(t_1, S(T_1), K, T_2)).$$

Using the identity $\max(a, b) = a + \max(0, b - a)$ we obtain

$$Y = C(t_1, S(T_1), K, T_2) + \max(0, P(t_1, S(T_1), K, T_2) - C(t_1, S(T_1), K, T_2)).$$

By the put-call parity,

$$Y = C(t_1, S(T_1), K, T_2) + \max(0, Ke^{-r(T_2-T_1)} - S(T_1)) = Z + U. \tag{6.22}$$

Hence $\Pi_Y(t) = \Pi_Z(t) + \Pi_U(t)$. Since U is the pay-off of a put option with strike $Ke^{-r(T_2-T_1)}$ expiring at time T_1 then

$$\Pi_U(t) = P(t, S(t), Ke^{-r(T_2-T_1)}, T_1). \tag{6.23}$$

To find $\Pi_Z(t)$ we need the following theorem:

Theorem 6.4. *Consider a standard European derivative with pay-off $Y = g(S(T))$ at maturity T and another derivative with pay-off $Z = \Pi_Y(t_*)$ at maturity $t_* < T$. Then $\Pi_Z(t) = \Pi_Y(t)$, $t \in [0, t_*]$.*

Proof. See [3, Th.5.1.1]. □

Applying this theorem with $Z = C(t_1, S(T_1), K, T_2) = \Pi_{(S(T_2)-K)_+}(T_1)$ we obtain

$$\Pi_Z(t) = C(t, S(t), K, T_2). \tag{6.24}$$

Replacing (6.23) and (6.24) into (6.22) we finally obtain, for the Black-Scholes price of the chooser option,

$$\Pi_Y(t) = C(t, S(t), K, T_2) + P(t, S(t), Ke^{-r(T_2-T_1)}, T_1).$$

Exercise 6.6. *Consider a European derivative with maturity T and pay-off Y given by*

$$Y = k + S(T) \log S(T),$$

where $k > 0$ is a constant. Find the Black-Scholes price of the derivative at time $t < T$ and the hedging self-financing portfolio. Find the probability that the derivative expires in the money.

6.4 Implied volatility

Let $C(t, S(t), K, T)$ denote the Black-Scholes price of a European call with strike price K, maturity time T on a stock with price $S(t)$ at time t. In fact, assuming that the underlying pays no dividend, this is also the Black-Scholes price of the corresponding American call, which is the one actually traded in the option markets. We recall that in the derivation of the Black-Scholes price it is assumed that the price of the stock follows the geometric Brownian motion

$$S(t) = S(0)e^{\alpha t+\sigma W(t)},$$

where $\{W(t)\}_{t\in[0,T]}$ is a Brownian motion stochastic process, $\sigma > 0$ is the volatility of the stock, α is the expected log-return of the stock. The function $C(t, s, K, T)$ is given by

$$C(t, s, K, T) = s\Phi(d_1) - Ke^{-r\tau}\Phi(d_2),$$

where $r > 0$ is the (constant) interest rate of the risk-free asset, $\tau = T - t$ is the time left to the expiration of the call,

$$d_2 = \frac{\log\frac{s}{K} + (r - \frac{\sigma^2}{2})\tau}{\sigma\sqrt{\tau}}, \quad d_1 = d_2 + \sigma\sqrt{\tau}$$

and Φ denotes the standard normal distribution,

$$\Phi(u) = \int_{-\infty}^{u} e^{-\frac{x^2}{2}}\frac{dx}{\sqrt{2\pi}}.$$

Remarkably, $C(t, S(t), K, T)$ does not depend on the expected log-return α of the stock. However it depends on the parameters (σ, r) and since here we are particularly interested in the dependence on the volatility, we re-denote the Black-Scholes price of the call as

$$C(t, S(t), K, T, \sigma).$$

Moreover, as shown in Theorem 6.3,

$$\frac{\partial C}{\partial \sigma} = \text{vega} = \frac{s}{\sqrt{2\pi}}e^{-\frac{d_1^2}{2}}\sqrt{\tau} > 0.$$

Hence the Black-Scholes price of the option is an increasing function of the volatility. Furthermore,

$$\lim_{\sigma\to 0^+} C(t, S(t), K, T) = (S(t) - Ke^{-r\tau})_+, \quad \lim_{\sigma\to+\infty} C(t, S(t), K, T) = S(t),$$

see Exercise 6.4. Therefore the function $C(t, S(t), K, T, \cdot)$ is a one-to-one map from $(0, \infty)$ into the interval $I = ((S(t) - Ke^{-r\tau})_+, S(t))$, see Figure 6.1. Now suppose that at some given *fixed* time t the real market price of the call is $\tilde{C}(t)$. Clearly, the option is always cheaper than the stock (otherwise we would buy directly the stock, and not the option) and

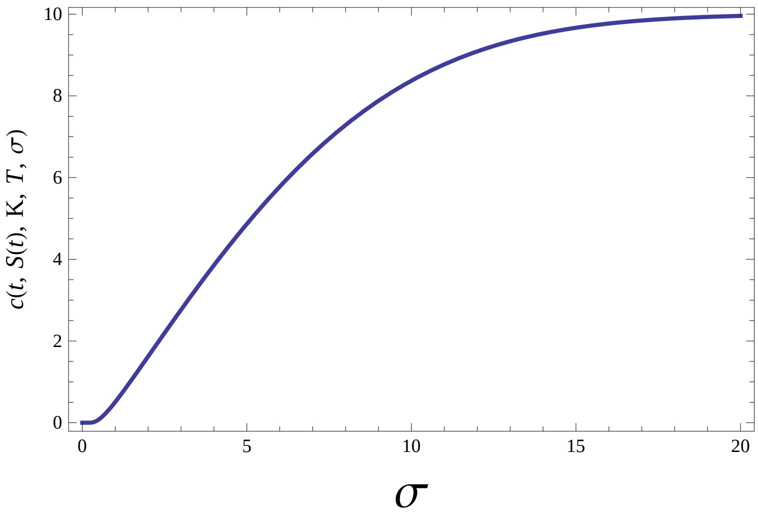


Figure 6.1: We fix $S(t) = 10$, $K = 12$, $r = 0.01$, $\tau = 1/12$ and depict the Black-Scholes price of the call as a function of the volatility. Note that in practice only the very left part of this picture is of interest, because typically $0 < \sigma < 1$.

typically we also have $\tilde{C}(t) > \max(0, S(t) - Ke^{-r\tau})$. The latter is always true if $S(t) < Ke^{-r\tau}$ (the market price of options is positive), while if $S(t) > Ke^{-r\tau}$ this follows by the fact that $S(t) - Ke^{-r\tau} \approx S(t) - K$ and real calls are always more expansive than their intrinsic value. This being said, we can safely assume that $\tilde{C}(t) \in I$.

Thus given the value of $\tilde{C}(t)$ there exists a unique value of σ, which depends on the fixed parameters T, K, r and which we denote by σ_{imp}, such that

$$C(t, S(t), K, T, \sigma_{\mathrm{imp}}) = \tilde{C}(t).$$

σ_{imp} is called the **implied volatility** of the option. The implied volatility must be computed numerically (for instance using Newton's method), since there is no close formula for it. Moreover it is usually computed using "nearly in the money" calls.

The implied volatility of an option (in this example of a call option) is a very important parameter and it is often quoted together with the price of the option. If the market followed exactly the assumptions in the Black-Scholes theory, then the implied volatility would be a constant, equal to the volatility of the underlying asset. In this respect, σ_{imp} may be viewed as a quantitative measure of how real markets deviate from ideal Black-Scholes markets. The implied volatility may also be viewed as the market consensus on the future value of the volatility of the underlying stock. Recall in fact that in order for the Black-Scholes price of the option to be $C(t, S(t), K, T, \sigma_{\mathrm{imp}})$, the volatility of the stock should be equal to σ_{imp} in the time interval $[t, T]$ in the future. Hence by pricing the option at the price $\tilde{C}(t) = C(t, S(t), K, T, \sigma_{\mathrm{imp}})$, the market is telling us that the buyers and sellers of the option believe that the volatility of the stock in the future will be σ_{imp}.

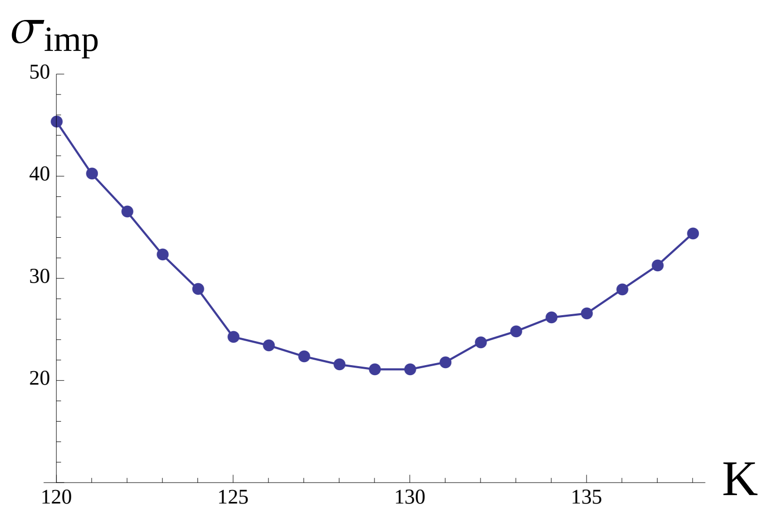


Figure 6.2: Volatility Smile of a call option on Apple expiring May 15th, 2015. The data were taken on May 12th, when the Apple stock quoted 126.34 dollars.

Volatility smile

As mentioned before, the implied volatility depends on the parameters T, K, r. Here we are particularly interested in the dependence on the strike price, hence we re-denote the implied volatility as $\sigma_{\text{imp}}(K)$. If the market behaved exactly as in the Black-Scholes theory, then $\sigma_{\text{imp}}(K) = \sigma$ for all values of K, hence the graph of $K \to \sigma_{\text{imp}}(K)$ would be just a straight horizontal line. Given that real markets do not satisfy exactly the assumptions in the Black-Scholes theory, what can we say about the graph of the function $K \to \sigma_{\text{imp}}(K)$? Remarkably, it has been found that there exists a recurrent convex shape for the graph of this function, which is known as **volatility smile**. The minimum of the graph is reached at the strike price $K \approx S(t)$, i.e., when the call is at the money. This behavior indicates that the more the call is far from being at the money, the more it will be overpriced. Volatility smiles have been found in the market only after the crash in 1987 (Black Monday), indicating that this event led investors to be more cautious when trading on options that are in or out of the money. Devise mathematical models of stochastic volatility and asset prices able to reproduce volatility smiles is an active research topic in mathematical finance.

6.5 Standard European derivatives on a dividend-paying stock

Consider a standard European derivative with pay-off $Y = g(S(T))$ and maturity T. In this section we assume that the underlying stock pays a dividend at the time $t_0 \in (0, T)$. This means that at the time t_0, the price per share of the stock decreases of a fraction $a \in (0, 1)$ of its price immediately before t_0, the difference being deposited in the account of the share holder*. Letting $S(t_0^-) = \lim_{t \to t_0^-} S(t)$, we then have

$$S(t_0) = S(t_0^-) - aS(t_0^-) = (1-a)S(t_0^-). \tag{6.25}$$

We assume that on each of the intervals $[0, t_0)$, $[t_0, T]$, the stock price follows a geometric Brownian motion, namely,

$$S(s) = S(t)e^{\alpha(s-t)+\sigma(W(s)-W(t))}, \quad t \in [0, t_0),\ s \in [t, t_0) \tag{6.26}$$
$$S(s) = S(u)e^{\alpha(s-u)+\sigma(W(s)-W(u))}, \quad u \in [t_0, T],\ s \in [u, T]. \tag{6.27}$$

Theorem 6.5. *Let $\Pi_Y^{(a,t_0)}(t)$ denote the Black-Scholes price at time $t \in [0, T]$ of a European derivative with pay-off $Y = g(S(T))$ assuming that the underlying stock pays the dividend $aS(t_0^-)$ at time $t_0 \in [0, T]$. Then*

$$\Pi_Y^{(a,t_0)}(t) = \begin{cases} v(t, (1-a)S(t)), & \text{for } t < t_0, \\ v(t, S(t)), & \text{for } t \geq t_0, \end{cases}$$

where $v(t, x)$ is the pricing function in the absence of dividends, which is given by (6.11).

Proof. We first observe that, using $\frac{S(T)}{S(t)} = e^{\alpha\tau+\sigma(W(T)-W(t))}$, we can rewrite (6.12) in the form

$$\Pi_Y(t) = e^{-r\tau}\mathbb{E}[g(S(T)e^{(r-\frac{\sigma^2}{2}-\alpha)\tau}]. \tag{6.28}$$

Taking the limit $s \to t_0^-$ in (6.26) and using the continuity of the paths of the Brownian motion, we find

$$S(t_0^-) = S(t)e^{\alpha(t_0-t)+\sigma(W(t_0)-W(t))}, \quad t \in [0, t_0).$$

Replacing in (6.25) we obtain

$$S(t_0) = (1-a)S(t)e^{\alpha(t_0-t)+\sigma(W(t_0)-W(t))}, \quad t \in [0, t_0).$$

Hence, letting $(s, u) = (T, t_0)$ and $(s, u) = (T, t)$ into (6.27), we find

$$S(T) = \begin{cases} (1-a)S(t)e^{\alpha\tau+\sigma(W(T)-W(t))} & \text{for } t \in [0, t_0), \\ S(t)e^{\alpha\tau+\sigma(W(T)-W(t))} & \text{for } t \in [t_0, T]. \end{cases} \tag{6.29}$$

Using the definition of Black-Scholes price in the form (6.28) and denoting $G = (W(T) - W(t))/\sqrt{\tau}$, we obtain

$$\Pi_Y^{(a,t_0)}(t) = e^{-r\tau}\mathbb{E}[g((1-a)S(t)e^{(r-\frac{\sigma^2}{2})\tau+\sigma\sqrt{\tau}G}], \quad \text{for } t \in [0, t_0),$$

$$\Pi_Y^{(a,t_0)}(t) = e^{-r\tau}\mathbb{E}[g(S(t)e^{(r-\frac{\sigma^2}{2})\tau+\sigma\sqrt{\tau}G}], \quad \text{for } t \in [t_0, T].$$

As $G \in \mathcal{N}(0, 1)$, the result follows. □

*The dividend is expressed as $100\,a$ percentage of the price of the stock. For instance, $a = 0.03$ means that the dividend paid is 3%.

We conclude that for $t \geq t_0$, i.e., after the dividend has been paid, the Black-Scholes price function of the derivative is again given by (6.11), while for $t < t_0$ it is obtained by replacing x with $(1-a)x$ in (6.11). To see the effect of this change, suppose that the derivative is a call option; let $C(t,x)$ be the Black-Scholes price function in the absence of dividends and $C_a(t,x)$ be the price function in the case that a dividend is paid at time t_0. Then, according to Theorem 6.5,

$$C_a(t,x) = \begin{cases} C(t,(1-a)x), & \text{for } t < t_0, \\ C(t,x), & \text{for } t \geq t_0. \end{cases}$$

Since $\partial_x C > 0$ (see Theorem 6.3), it follows that $C_a(t,x) < C(t,x)$, for $t < t_0$, that is to say, the payment of a dividend makes the call option on the stock less valuable (i.e., cheaper) than in the absence of dividends until the dividend is paid.

Exercise 6.7 (?). *Explain why the property just proved makes sense.*

Exercise 6.8. *Derive the Black-Scholes price of a derivative with pay-off $Y = g(S(T))$, assuming that the underlying pays a dividend at each time $t_1 < t_2 < \cdots < t_M \in [0,T]$. Denote by a_i the dividend paid at time t_i, $i = 1, \dots, M$.*

Appendix A

The Markowitz portfolio theory

Consider a constant portfolio position in some time interval $[0, T]$. We assume that the portfolio is invested in n risky assets $\mathcal{U}_1, \dots \mathcal{U}_n$ with prices $\Pi^{\mathcal{U}_i}(t)$, $i = 1, \dots, N$ and a risk-free asset $\mathcal{U}_{n+1}$ with value $B(t) = B_0 e^{rt}$, where $r > 0$ is the instantaneous interest rate. For notational convenience we let $\Pi^{\mathcal{U}_{n+1}}(t) = B(t)$, so that $\Pi^{\mathcal{U}_i}(t)$, $i = 1, \dots, n+1$ now denotes the price of every asset in the portfolio. The (relative) return of the asset i is defined by

$$R_i = \frac{\Pi^{\mathcal{U}_i}(T) - \Pi^{\mathcal{U}_i}(0)}{\Pi^{\mathcal{U}_i}(0)}, \quad i = 1, \dots, n+1.$$

For $i = 1, \dots, n$, R_i is a random variable, while for $i = n+1$ it is the deterministic constant

$$R_{n+1} = \frac{B_0 e^{rT} - B_0}{B_0} = e^{rT-1} := \rho > 0.$$

As usual, the price of the assets at time $t = 0$ are supposed to be known.

The problem under study in this chapter is the following: Given an investor with initial capital $K > 0$, what is the best way to distribute this wealth among the $n+1$ assets in order to maximize the expected return and, at the same time, minimize the expected risk? To solve this problem we need first to introduce some notation. Let $a_i \in \mathbb{R}$ denote the number of shares of the asset $\mathcal{U}_i$ in the portfolio. The initial value of the portfolio is

$$V(0) = \sum_{i=1}^{n+1} a_i \Pi^{\mathcal{U}_i}(0) = K.$$

Letting

$$\pi_i = \frac{a_i \Pi^{\mathcal{U}_i}(0)}{K}, \quad i = 1, \dots n+1, \tag{A.1}$$

we get

$$\sum_{i=1}^{n+1} \pi_i = 1. \tag{A.2}$$

The value of the portfolio at time $t = T$ is

$$V(T) = \sum_{i=1}^{n+1} a_i \Pi^{\mathcal{U}_i}(T) = \sum_{i=1}^{n+1} a_i \Pi^{\mathcal{U}_i}(0)(1 + R_i).$$

The (relative) return of the portfolio in the interval $[0, T]$ is given by

$$R = \frac{V(T) - V(0)}{V(0)} = \frac{V(T)}{K} - 1.$$

After simple calculations we obtain

$$R = \sum_{i=1}^{n+1} \pi_i R_i, \tag{A.3}$$

where we used (A.1)-(A.2). The return of the portfolio is a random variable and taking its expectation we obtain

$$\mathbb{E}[R] = \sum_{i=1}^{n+1} \pi_i \mathbb{E}[R_i] = \sum_{i=1}^{n} \pi_i \mathbb{E}[R_i] + \pi_{n+1} \mathbb{E}[R_{n+1}].$$

Using (A.2) and that $\mathbb{E}[R_{n+1}] = R_{n+1} = \rho$ we obtain

$$\mathbb{E}[R] = \sum_{i=1}^{n} \pi_i \mu_i + \left(1 - \sum_{i=1}^{n} \pi_i\right) \rho, \tag{A.4}$$

where we set

$$\mu_i = \mathbb{E}[R_i], \quad \text{for } i = 1, \ldots, n.$$

Hence the portfolio should be chosen to maximize (A.4) and at the same time minimize the portfolio risk. We measure the latter with the variance $\mathrm{Var}[R]$ of the return. To compute $\mathrm{Var}[R]$ we first observe that

$$R - \mathbb{E}[R] = \sum_{i=1}^{n+1} \pi_i R_i - \sum_{i=1}^{n+1} \pi_i \mathbb{E}[R_i] = \sum_{i=1}^{n} \pi_i (R_i - \mathbb{E}[R_i]),$$

hence

$$\begin{aligned}(R - \mathbb{E}[R])^2 &= \left(\sum_{i=1}^{n} \pi_i (R_i - \mathbb{E}[R_i])\right)^2 = \left(\sum_{i=1}^{n} \pi_i (R_i - \mathbb{E}[R_i])\right)\left(\sum_{j=1}^{n} \pi_j (R_j - \mathbb{E}[R_j])\right) \\ &= \sum_{i,j=1}^{n} \pi_i \pi_j (R_i - \mu_i)(R_j - \mu_j).\end{aligned}$$

Letting $C = (c_{ij})_{i,j=1\dots n}$, $c_{ij} = \mathrm{Cov}(R_i, R_j)$, denote the covariance matrix of the assets returns and $\pi = (\pi_1\ \pi_2 \dots \pi_n)^T$, we obtain

$$\mathrm{Var}[R] = \mathbb{E}[(R - \mathbb{E}[R])^2] = \sum_{i,j=1}^{n} \pi_i \pi_j c_{ij} = \pi^T C \pi. \tag{A.5}$$

We take $\mathrm{Var}[R]$, i.e., the "randomness" of the portfolio return, as a measure of the portfolio risk. Our purpose is then to find a portfolio which maximize $\mathbb{E}[R]$ and minimize $\mathrm{Var}[R]$. To this end we shall need the following simple result.

Theorem A.1. *Let $X_1, \dots X_n : \Omega \to \mathbb{R}$ be random variables and $c_{ij} = \mathrm{Cov}(X_i, X_j)$. The covariance matrix $C = (c_{ij})_{i,j=1\dots,n}$ is symmetric and positive-semidefinite, i.e., $C = C^T$ and $\xi^T C \xi \geq 0$, for all $\xi \in \mathbb{R}^n$.*

Exercise A.1. *Prove the theorem.*

Definition A.1. *Given $\theta > 0$, let*

$$g_\theta(\pi_1, \dots, \pi_n) = \mathbb{E}[R] - \theta \mathrm{Var}[R]. \tag{A.6}$$

A portfolio $(a_1, \dots a_n, a_{n+1})$ is called a **Markowitz portfolio** *if the function g_θ has a maximum on $\widehat{\pi} = (\widehat{\pi}_1, \dots, \widehat{\pi}_n)$, where*

$$\widehat{\pi}_1 = \frac{a_1 S_1(0)}{K}, \dots, \widehat{\pi}_n = \frac{a_n S_n(0)}{K}.$$

The parameter $\theta > 0$ is called the **risk aversion** of the investor. We shall see that the higher is θ the more portfolio value is invested in the risk-free asset. Hence a large θ corresponds to a very cautious investor, who favors "small risks" to "higher returns".

Theorem A.2. *Assume that the covariance matrix of the assets returns is positive-definite, i.e., $\xi^T C \xi > 0$, for all $\xi \neq 0$. Let Ω be the vector*

$$\Omega = \frac{1}{2\theta}(\mu_1 - \rho \quad \mu_2 - \rho \ \dots \ \mu_n - \rho)^T.$$

Then the function (A.6) *has a unique maximum, which is attained at $\widehat{\pi} = C^{-1}\Omega$.*

Proof. Using (A.4) and (A.5) the function g_θ is

$$g_\theta(\pi) = \sum_{i=1}^{n} \mu_i \pi_i + \left(1 - \sum_{i=1}^{n} \pi_i\right) \rho - \theta \sum_{i,j=1}^{n} \pi_i \pi_j c_{ij}.$$

Hence

$$\frac{\partial g_\theta}{\partial \pi_k} = \mu_k - \rho - 2\theta \sum_{j=1}^{n} c_{kj} \pi_j.$$

Thus the stationary points, i.e., the solutions $\widehat{\pi}$ of $\nabla_\pi g_\theta(\widehat{\pi}) = 0$, satisfy

$$\sum_{i=1}^{n} c_{kj}\widehat{\pi}_j = \frac{1}{2\theta}(\mu_k - \rho), \quad \text{i.e.,} \quad C\widehat{\pi} = \Omega.$$

As C is assumed to positive definite, then it is invertible and thus $C\widehat{\pi} = \Omega$ has the unique solution $\widehat{\pi} = C^{-1}\Omega$. To show that this stationary point is a maximum of g_θ we compute

$$\frac{\partial^2 g}{\partial \pi_k \partial \pi_l} = -2\theta c_{kl}.$$

Since C is positive definite, then the Hessian $\nabla^2_\pi g_\theta$ is negative definite and therefore the point $\widehat{\pi}$ is a maximum. This concludes the proof of the theorem. □

Remark A.1. Note that $\widehat{\pi}_i \to 0$, and so also $a_i \to 0$, as $\theta \to +\infty$, for all $i = 1, \dots, n$. Hence in this limit all initial capital K is invested in the risk-free asset: $\pi_{n+1} \to 1$, as $\theta \to +\infty$, which is equivalent to $a_{n+1} \to K/\Pi^{\mathcal{U}_{n+1}}(0) = KB_0$, as $\theta \to +\infty$. Conversely, the smaller is θ, the larger is the exposure of the investor on the risky assets.

As a special case, assume that $\Pi^{\mathcal{U}_1(T)}, \dots \Pi^{\mathcal{U}_n(T)}$ are independent. Then $R_1(T), \dots, R_n(T)$ are also independent and therefore $\mathrm{Cov}(R_i, R_j) = 0$, for $i \neq j$. Hence

$$C = \mathrm{diag}(\mathrm{Var}[R_1], \mathrm{Var}[R_2], \dots, \mathrm{Var}[R_n]),$$

where $\mathrm{diag}(z_1, \dots, z_n)$ is the diagonal $n \times n$ matrix A with diagonal elements $a_{ii} = z_i$, $i = 1, \dots n$. Therefore in this case the maximum of g_θ is attained at

$$\widehat{\pi}_i = \frac{\mathbb{E}[R_i] - \rho}{2\theta \mathrm{Var}[R_i]}. \tag{A.7}$$

Example

Suppose that the risky assets are stocks and that the prices $\Pi^{\mathcal{U}_i}(t) = S_i(t)$ are independent and that each of them follows a binomial model:

$$S_i(j) = \begin{cases} S_i(j-1)e^{u_i}, & \text{with probability } p_i \\ S_i(j-1)e^{d_i}, & \text{with probability } = 1 - p_i \end{cases}, \quad j \in \mathcal{I} = \{1, \dots, N\}, \quad i = 1, \dots n.$$

Let us compute the Markowitz portfolio in the N days period. We have

$$\mathbb{E}[R_i] = \mathbb{E}[S_i(0)^{-1}(S_i(N) - S_i(0))] = \frac{1}{S_i(0)}\mathbb{E}[S_i(N)] - 1.$$

By Theorem 4.1 in Chapter 4 we have $\mathbb{E}[S_i(N)] = S_i(0)(e^{d_i}p_i + e^{u_i}(1-p_i))^N$. Hence

$$\mathbb{E}[R_i] = (e^{d_i}p_i + e^{u_i}(1-p_i))^N - 1. \tag{A.8}$$

Moreover
$$\mathbb{E}[S_i(N)^2] = S_i(0)^2(e^{2d_i}p_i + e^{2u_i}(1-p_i))^N,$$
since $S_i(t)$ follows a binomial model with parameters $2u, 2d$. Hence

$$\mathrm{Var}[S_i(N)] = \mathbb{E}[S_i(N)^2] - \mathbb{E}[S_i(N)]^2 = S_i(0)^2[(e^{2d_i}p_i + e^{2u_i}(1-p_i))^N - (e^{d_i}p_i + e^{u_i}(1-p_i))^{2N}]$$

and therefore

$$\mathrm{Var}[R_i] = \frac{1}{S_i(0)^2}\mathrm{Var}[S_i(N)] = (e^{2d_i}p_i + e^{2u_i}(1-p_i))^N - (e^{d_i}p_i + e^{u_i}(1-p_i))^{2N}. \qquad \text{(A.9)}$$

Replacing (A.8) and (A.9) into (A.7), and then inverting (A.1), we obtain the desired Markowitz portfolio $(a_1, \dots a_{n+1})$.

Exercise A.2. *Consider a 2-period binomial model with the following parameters*

$$e^u = \frac{4}{3}, \quad e^d = \frac{2}{3}, \quad p_u = \frac{1}{2}.$$

Assume further that $S(0) = 36$, and that the interest rate is zero. Consider also an option with pay-off
$$Y = (S(2) - 28)_+ - 2(S(2) - 32)_+ + (S(2) - 36)_+$$
and time of maturity $T = 2$. Compute the fair value of the option for $t = 0$. Assume now that an investor with risk aversion $\theta = \frac{1}{36}$ wants to distribute the initial wealth $K = 1000$ in the following assets: the stock, the option, and a risk-free asset with interest r such that $e^{2r} = 10/9$. Derive the corresponding Markowitz portfolio.

Exercise A.3. *Consider two stocks with prices*

$$S_1(t) = S_1(0)e^{\alpha_1 t + \sigma_1 W_1(t)}, \quad S_2(t) = S_2(0)e^{\alpha_2 t + \sigma_2 W_2(t)}$$

where $\sigma_1, \sigma_2 > 0$, $\alpha_1, \alpha_2 \in \mathbb{R}$ and $W_1(t), W_2(t)$ are two Brownian motions. Let $T > 0$ and assume that $W_1(T)$ and $W_2(T)$ are independent random variables. Compute the Markowitz portfolio of an investor with initial capital $K > 0$ and risk aversion θ who wants to invest in the stocks and in a money market with interest $r > 0$ during the interval of time $[0, T]$.

Bibliography

[1] F. Black, M. Scholes: The Pricing of Options and Corporate Liabilities. The Journal of Political Economy **81**, 637–654 (1973)

[2] Z. Bodie, A. Kane, A. J. Marcus: Investments, 5^{th} ed. The McGraw-Hill Primis (2003)

[3] C. Borell: Introduction to the Black-Scholes theory.

[4] J. C. Cox, S. A. Ross, M. Rubinstein: Option pricing: A simplified approach. Journal of Financial Economics 7, 229 (1979)

[5] Shreve, E. S.: Stochastic calculus for finance I. The binomial asset pricing model. Springer Finance, New York (2004)

[6] Shreve, E. S.: Stochastic calculus for finance II. Continuous-time models Springer Finance, New York (2004)