Revised 12/9/2003
-Enlarged chapter 4
-Rewrote chapter 9
-Added chapter 10

Introduction to the Mathematics of Finance

From Risk Management to Options Pricing

by

Dr. Steven Roman
Professor Emeritus of Mathematics
California State University Fullerton

Contact: sroman@romanpress.com
949-854-5667

Notation Index

1 the unit vector $(1, \ldots, 1)$
1_A^S indicator function for $A \subseteq S$
$\mathcal{A} = \{\mathfrak{a}_1, \ldots, \mathfrak{a}_n\}$ assets
C (price of a call)
e_i the i-th standard unit vector
$\mathcal{E}_P(X)$: expected value of X with respect to probability P
Φ_i asset holding process
K (strike price)
μ_X: expected value of X
$\Omega = \{\omega_1, \ldots, \omega_m\}$ states of the economy
P (price of a put)
$\mathcal{P}_i = \{B_{i,1}, \ldots, B_{i,m_i}\}$ state partition
$\mathbb{P}$ probability
r risk-free interest rate
$\mathrm{RV}(\Omega)$ vector space of all random variables from Ω to $\mathbb{R}$
$\mathrm{RV}^n(\Omega)$ vector space of all random vectors from Ω to $\mathbb{R}^n$
$\rho_{X,Y}$: correlation coefficient of X and Y
S (price of stock or other asset)
$\sigma = (s_1, \ldots, s_m)$ state vector
σ_X^2: variance of X
$\sigma_{X,Y}$: covariance of X and Y
Θ_i portfolio
$\mathcal{V}_0$ initial cost function
$\mathcal{V}_T$ payoff function

Greek Alphabet

A α alpha	H η eta	N ν nu	T τ tau
B β beta	Θ θ theta	Ξ ξ xi	Υ υ upsilon
Γ γ gamma	I ι iota	O o omicron	Φ ϕ phi
Δ δ delta	K κ kappa	Π π pi	X χ chi
E ϵ epislon	Λ λ lambda	P ρ rho	Ψ ψ psi
Z ζ zeta	M μ mu	Σ σ sigma	Ω ω omega

Preface

This book covers two main areas of mathematical finance—areas that culminate in work worthy of the Noble prize in economics. One is portfolio risk management, culminating in the Capital Asset Pricing Model and the other is asset pricing theory, culminating in the Black-Scholes option pricing formula. Our discussion of portfolio risk management takes but a single chapter. The rest of the book is devoted to the study of asset pricing models.

The intended audience of the book is upper division undergraduate or beginning graduate students in mathematics, finance or economics. Accordingly, no measure theory is used in this book.

I realize that the book may be read by people with rather diverse backgrounds. On the one hand, students of mathematics may be well prepared in the ways of mathematical thinking but not so well prepared when it comes to matters related to finance (portfolios, stock options, forward contacts and so on). On the other hand, students of finance and economics may be well versed in financial topics but not as mathematically minded as students of mathematics.

Since the subject of this book is the *mathematics* of finance, I have not watered down the mathematics in any way (appropriate to the level of the book, of course). That is, I have endeavored to be mathematically rigorous *at the appropriate level*. On the other hand, the reader is not assumed to have any background in finance, so I have included the necessary background in this area (stock options and forward contracts).

I have also made an effort to make the book as mathematically self-contained as possible. Aside from a certain comfort level with mathematical thinking, a freshman/sophomore course in linear algebra is more than enough. In particular, the reader should be comfortable with matrix algebra, the notion of a vector space and the kernel and range of a linear transformation. The method of Lagrange multipliers is used in a couple of proofs related to risk management, but these proofs can be skimmed or omitted if desired.

Of course, probability theory is ever present in the area of mathematical finance. In this respect, the book is self-contained. Several chapters on probability theory are placed at appropriate places throughout the book.

The idea is to provide the necessary theory on a “need to know” basis. In this way, readers who choose not to cover the continuous pricing theory, for example, need not deal with matters related to continuous probability.

The book is organized as follows. The first chapter is devoted to the elements of discrete probability. The discussion includes such topics as random variables, independence, expectation, covariance and best linear predictors. If readers have had a course in elementary probability theory then this chapter will serve mostly as a review.

Chapter 2 is devoted to the subject of portfolio theory and risk management. The main goal is to describe the famous Capital Asset Pricing Model (CAPM). The chapter stands independent of the remainder of the book and can be omitted if desired.

The remainder of the book is devoted to asset pricing models. Chapter 3 gives the necessary background on stock options. In Chapter 4, we briefly illustrate the technique of asset pricing through the assumption of no arbitrage by pricing plain-vanilla forward contracts and discussing some simple issues related to option pricing, such as the *put-call option parity formula*, which relates the price of a put and a call on the same underlying asset with the same strike price and expiration time.

Chapter 5 continues the discussion of discrete probability, covering conditional probability along with more advanced topics such as partitions of the sample space and knowledge of random variables, conditional expectation (with respect to a partition of the sample space) stochastic processes and martingales. This material is covered at the discrete level and always with a mind to the fact that it is probably being seen by the student for the first time.

With the background on probability from Chapter 5, the reader is ready to tackle discrete-time models in Chapter 6. Chapter 7 describes the Cox-Ross-Rubinstein model. The chapter is short, but introduces the important topics of drift, volatility and random walks.

Chapter 8 introduces the very basics of continuous probability. We need the notions of convergence in distribution and the Central Limit Theorem so that we can take the limit of the Cox-Ross-Rubinstein model as the length of the time periods goes to 0. We perform this limiting process in Chapter 9 to get the famous Black-Scholes Option Pricing Formula.

Finally, in Chapter 10 we discuss optimal stopping times and American options. This chapter is perhaps a bit more mathematically challenging than the previous chapters.

A Word On Definitions

Unlike many areas of mathematics, the subject of this book, namely, the mathematics of finance, does not have an extensive literature at the undergraduate level. Put more simply, there are very few undergraduate textbooks on the mathematics of finance.

Accordingly, there has not been a lot of precedent with respect to setting down the basic theory at the undergraduate level, where pedagogy and use of intuition is (or should be) at a premium. One area in which this seems to manifest itself is the lack of terminology to cover certain situations.

Accordingly, on rare occasions I have felt it necessary to invent new terminology to cover a specific concept. Let me assure the reader that I have not done this lightly. It is not my desire to invent terminology for any other reason than as an aid to pedagogy.

In any case, the reader will encounter a few definitions that I have labeled as *nonstandard*. This label is intended to convey the fact that the definition is not likely to be found in other books nor can it be used without qualification in discussions of the subject matter outside the purview of this book.

Introduction

Portfolio Risk Management

Risk is an inevitable side effect of the effort to make more money than the next guy. To be sure, money makes money and this process can be carried out without any significant measure of risk: all an investor needs to do is buy United States Treasury bonds, generally considered to be risk-free investments. The price paid for such an investment is generally a boring rate of return.

The real problem is that if everyone makes the same rate of return, then this return serves only to maintain the status quo. Put another way, if you want to buy a Rolls Royce or a yacht or even a Rolex watch, then you need to make more money than the next guy, and this requires taking risk.

The first problem is to decide how to quantify the level of risk of an asset. This turns out to be simple. However, like the rest of life, simple answers often turn out to be incomplete. In particular, not only is it important to measure an asset's risk, but it is essential to measure the risk that results from the asset's *interaction with the other assets in a portfolio*. After all, in the end it is the performance of an investor's *entire portfolio* that separates one investor from another.

Of course, the future return of an asset is generally unknown in the present. In probabilistic terms, an asset's return is a *random variable*. So too is the return on an entire portfolio of assets. However, it is not hard to see that by combining assets in a careful manner, it is possible to engineer the overall risk of the portfolio, possibly even to a point that is below the level of risk of each individual asset. This risk-lowering process of asset selection is called *diversification*.

Speaking more mathematically, it is generally accepted that a good measure of an asset's risk is the *variance* (or standard deviation) of the return. As the reader probably knows, the variance (or standard deviation) is a measure of the spread of possible values of a random variable. The greater the variance, the greater is the probability that the risk will deviate significantly from the average. By the same token, the *covariance* of an asset's return with respect to the returns of the other assets in the portfolio provides a good measure of risk-interaction.

Thus, in portfolio management theory, one examines the expected value of an asset's return as well as its variance and covariance with other assets. Only through these statistics can one determine whether adding a particular asset to a portfolio is justified based on a risk-return analysis. As we will see, a remarkably simple procedure is provided by the Capital Asset Pricing Model (or CAPM).

The CAPM leads to the notion of a *market portfolio*, which is a portfolio of risky assets that has *perfect diversification*. In theory, such a portfolio must contain a positive amount of every available asset in the universe. This is because all investors will want to invest exclusively in this portfolio (along with the risk-free asset) and so any asset that is not in the market portfolio will wither from neglect and die.

From a practical standpoint, the market portfolio is nothing but hot air. On the other hand, studies indicate that it is possible to approximate a market portfolio by investing in a few dozen or so well-chosen assets. Fortunately, this also partially mitigates the problem of withering assets, because an asset that doesn't make it into one investor's "market portfolio" may very well make it into another's portfolio.

Once a market portfolio (or approximation thereof) has been identified, there remains only one consideration for the rational investor (at least in theory) and that is how much to invest in the risky portfolio and how much to invest instead in a risk-free asset. This is a question not for mathematics but for personal introspection and this is where our story will end.

Option Pricing Models

A *financial security* or *financial instrument* is a legal contract that conveys ownership (such as in the case of a stock), credit (such as in the case of a bond) or rights to ownership (such as in the case of a stock option).

Some financial securities have the property that their value depends upon the value of another security. In this case, the former security is called a **derivative** of the latter security, which is then called the **underlying security** for the derivative. The most well know examples of derivatives are ordinary stock options (puts and calls). In this case, the underlying security is a stock.

However, derivatives have become so popular that they now exist based on more exotic underlying financial entities, some that would not normally be considered financial securities, such as interest rates and currency exchange rates. Perhaps this is why it has become common to refer to the underlying entity simply as the **underlying**.

It is also possible to base derivatives on other derivatives. For example, one can trade options on futures contracts. Thus, a given financial entity can be a derivative under some circumstances and an underlying under other circumstances.

Indeed, the business of investors is to make money and this can only be done (arbitrage opportunities aside) by taking risk, that is, by gambling. Just as the Las Vegas casinos are always on the lookout for a new game of chance with which to increase their profits, the investment community is always on the lookout for a new financial game of chance. These games often take the form of exotic derivatives.

In this book, we shall concentrate on simple derivatives, primarily ordinary stock options. We will be interested in both the purchase and sale of such securities. When a purchase is made the buyer is said to take a **long position** in the security. When a sale is made the seller is said to take the **short position** in the security. The two positions are said to be **opposite positions** of one another.

The central theme of this portion of the book is to find ways to determine the initial value (or price) of a derivative as a function of the price of its underlying asset. This is the **derivative pricing problem**.

The only time at which the derivative pricing problem is relatively easy to solve is at the time of *expiration* of the derivative. For example, if a certain derivative gives you the right to buy a stock at \$100 per share *at this very moment* (the time of expiration) then this option is worthless if the current market price of the stock is below \$100. On the other hand, if the current market price of the stock is \$110 then the current value of the derivative is \$10. More generally, if the current stock price is S then the option is worth $\max\{S-100,0\}$ assuming, as we do, that there are no external costs or fees involved.

At any time before expiration, the connection between the current value of a derivative and the current value of its underlying asset is complex and this is why the theory of derivative pricing is also complex. At the

current state of knowledge, the only way to deal with the full complexity of this relationship is to assume it away.

Assumptions

Financial markets are complex. As with most complex systems, creating a mathematical model of the system requires making some simplifying assumptions.

In the course of our analysis, we will make several such simplifying assumptions. For example, we will assume a **perfect market**, that is, a market in which

- there are no commissions or transaction costs,
- the lending rate is equal to the borrowing rate,
- there are no restrictions on short sales.

Of course, there is no such thing as a perfect market in the real world, but this assumption will make the analysis considerably simpler and will also let us concentrate on certain key issues that appear less clearly under less restrictive conditions.

Arbitrage

As we will see, the key principle behind asset pricing is the notion that *the market tries to avoid arbitrage*. More specifically, if an arbitrage opportunity exists, then prices will be adjusted to eliminate that opportunity.

As a simple example, suppose that gold is selling for $380.10 per ounce in New York and $380.20 in London. Then investors could buy gold in New York and sell it in London, making a profit of 10 cents per ounce (assuming that transaction costs do not absorb the profit). However, the purchasing of gold in New York will drive the New York price higher and the selling of gold in London will drive the London price lower. Result: no more arbitrage.

As a consequence of this tendency to an arbitrage-free market equilibrium, it only makes sense to price securities under the assumption that there is no arbitrage.

Surprisingly, the term *arbitrage* suffers from a bit of a dichotomy. In a general, nontechnical sense, the term is often used to signify a condition

under which an investor is *guaranteed* to make a profit regardless of circumstances.

The more commonly adopted technical use of the term is a bit different. An **arbitrage opportunity** is an investment opportunity that is guaranteed not to result in a loss and *may* (with positive probability) result in a gain. Note that the gain is not guaranteed, only the lack of loss is guaranteed. Also, we must be very careful how we measure the gain. For instance, if $100 today grows to $100.01 in a year, is this true gain? Put another way, would you make this investment? Probably not, because there are undoubtedly risk-free alternatives, such as depositing the money in a Federally insured bank account that will produce a better gain.

The No-Arbitrage Principle

The no-arbitrage principle for pricing is actually quite simple. Imagine two portfolios of assets (stocks, bonds, derivatives, etc.). Let us refer to these portfolios as Portfolio A and Portfolio B. Let us also consider two time periods: the initial time $t = 0$ and a time $t = T$ in the future.

Accordingly, each portfolio has an initial value (value at time 0) and a final value (value at time T) or *payoff*. Let us denote the initial value of the two portfolios by $\mathcal{V}_{A,0}$ and $\mathcal{V}_{B,0}$ and the final values by $\mathcal{V}_{A,T}$ and $\mathcal{V}_{B,T}$. The values of Portfolio A are shown in Figure 1. A similar figure holds for Portfolio B.

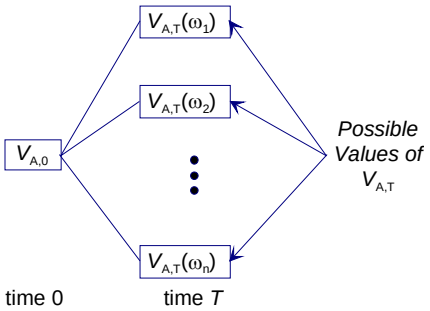


Figure 1– The values of Portfolio A

As can be seen in the figure, Portfolio A has an initial value that is either known or capable of being determined. On the other hand, the final value of portfolio A is unknown at time $t = 0$. In fact, we assume that this value depends on the state of the economy at time T, which can be one

of n possible states $\omega_1, \ldots, \omega_n$. Thus, the final value $\mathcal{V}_{A,T}$ is actually a function of these states.

Similarly, we assume that the initial value of Portfolio B is known or capable of being determined and that the final value is a function of the possible states of the economy.

Now, consider what happens if Portfolios A and B have exactly the same payoffs *regardless of the state of the economy*, that is

$$\mathcal{V}_{A,T}(\omega_i) = \mathcal{V}_{B,T}(\omega_i)$$

for $i = 1, \ldots, n$. The no-arbitrage principle then implies that the initial values must be equal, that is

$$V_{A,0} = \mathcal{V}_{B,0}$$

For suppose that $V_{A,0} > \mathcal{V}_{B,0}$. Then under the assumption of a perfect market, an investor can purchase the cheaper Portfolio B and sell the more expensive Portfolio A, pocketing the difference. At time T, *no matter what state the economy is in*, the investor receives the common final value of the portfolios and must pay out the same amount. Thus, he loses nothing at the end and can keep the initial profit. This is arbitrage.

Thus, we see that the no-arbitrage principle can be used to price a portfolio, that is, to determine an initial value of a portfolio. To price Portfolio A, for example, all we need to do is find another portfolio, say Portfolio B that has the same payoff function as Portfolio A and has a known initial value. It follows that the initial value of Portfolio A must be equal to the initial value of Portfolio B.

The no-arbitrage principle can be used in other ways to determine prices. For example, if the initial values of the two portfolios are equal, then it cannot be that one portfolio *always* yields a higher payoff than the other.

We will see many examples of the use of the no-arbitrage principle throughout the book.

Chapter 1

Probability I: An Introduction to Discrete Probability

Asset pricing involves the prediction of future events and as such relies very heavily on the mathematical theory of probability. In this chapter, we begin a discussion of basic probability. This discussion will continue in later chapters, as the need for more information arises in connection with subsequent topics to be covered in the book.

Probability seems to have had its origins in an effort to predict the outcome of games of chance and is generally considered to have begun as a formal theory in a series of letters between the two famous mathematicians Blaise Pascal and Pierre de Fermat in the summer of 1654.

Overview

In the study of probability, the typical scenario is that of an *experiment*, such as rolling a pair of dice, administering a drug to a patient or predicting the future price of a stock. The key is that the experiment must have a well-defined set of *possible outcomes*. This set is referred to as the *sample space* of the experiment.

Subsets of the sample space, that is, subsets of outcomes, are referred to as *events*. When an outcome occurs that is in a particular event, we say that the event has occurred. Thus, for example, we have the event of getting a sum of 7 on the dice, the event that a patient's temperature drops to 98.6 after receiving a drug or the event that a stock price rises by 10%.

Next, a method must be determined to measure the *probability*, or *likelihood* that various events will occur as a result of conducting the experiment. More specifically, the probability of an event is a real number between 0 and 1 that measures the likelihood that the outcome will lie in the event. A probability of 0 indicates that the event cannot occur (is impossible) and a probability of 1 indicates that the event is certain to occur.

The method that is used to determine these probabilities is not really part of the subject of probability *per se*. Two approaches are common. One is simply to assume the probabilities. For instance, consider the experiment of tossing a single coin. Assuming that the coin is fair is equivalent to assuming that the probability of heads and tails are both $1/2$. Another approach is statistical in nature, using empirical data to assign probabilities. For example, if the coin is flipped 10000 times and results in 5003 heads, we may decide to set the probability of heads equal to $5003/10000$.

The flavor of probability theory depends quite markedly on the nature of the sample space. The basic concepts of probability theory require far less mathematical machinery when dealing with *finite* sample spaces, for in this case probabilities can be assigned to *individual outcomes* in the sample space, as we did with the coin-tossing example just discussed. As we will soon see, all that is required is that the probabilities be numbers between 0 and 1 (inclusive) that add up to 1. Then the probability of an event is simply the sum of the probabilities of the outcomes that lie in that event. The term *finite probability theory* is used to refer to the theory of probability on finite sample spaces.

As an example, suppose that based on market research, we decide that a certain stock, currently selling at $\$100$ per share, will be selling at either $\$99$, $\$100$ or $\$101$ dollars by the end of the day. Thus, we have an experiment whose sample space consists of the possible stock prices

$$\Omega = \{99, 100, 101\}$$

Further, after research into the price history of the stock, we may decide to assign empirical probabilities as follows

$$\mathbb{P}(99) = 0.25, \mathbb{P}(100) = 0.5, \mathbb{P}(101) = 0.25$$

In this case, the event that the price does not fall is $\{100, 101\}$, whose probability is $\mathbb{P}(100) + \mathbb{P}(101) = 0.75$.

Probability theory for countably infinite sample spaces is also relatively approachable, at least at the beginning. Again, probabilities can be assigned to the individual outcomes in the sample space. However, the issue of convergence of an infinite sum now comes into play. The term *discrete probability* is used to refer to the probability of finite or countably infinite sample spaces. Whole books have been written on the subject of discrete probability alone.

As an example of a discrete (but nonfinite) sample space, consider the experiment of tossing a coin until the *first* heads appears. The outcome is the toss number of this first heads. At the outset, we cannot confine the set of outcomes to any finite sample space, because there is no way to tell in advance how many tosses will be necessary before a heads appears. So the sample space must be the set

$$\Omega = \{1, 2, 3, \dots\}$$

of all positive integers. Indeed, one must argue (or assume) that a heads must eventually appear, for if not then even this set does not represent all possible outcomes.

It is possible to show that if the coin is *fair*, that is, if the likelihood of heads is the same as that of tails, then the probability that the so-called *waiting time* to the first heads is k is given by

$$\mathbb{P}(\text{first heads at toss } k) = \frac{1}{2^k}$$

Since the infinite sum

$$\sum_{k \geq 1} \frac{1}{2^k}$$

converges to 1, we have a legitimate probability measure.

To get some idea of why these probabilities make sense, it should be rather obvious that the probability that the first heads occurs at toss $k = 1$ is $\frac{1}{2} = \frac{1}{2^1}$. The only way that the first heads can occur at toss $k = 2$ is if the first toss results in tails and the second in heads. But there are a total of four equally likely possibilities for the first two tosses

$$(H, H), (H, T), (T, H), (T, T)$$

so it is reasonable to set the probability of waiting till the second toss for the first heads to $\frac{1}{4} = \frac{1}{2^2}$. This reasoning can be extended to larger values of k.

We do not want to leave the reader with the impression that discrete probability is somehow “easier” than nondiscrete probability, where the sample space is uncountable. This is decidedly not the case. However, it is true that a basic understanding of discrete probability requires much

less mathematical background. For example, discrete probability does not, in general, require the notion of integrability and finite probability does not, in general, require the notion of limit.

For the nondiscrete case, things take a dramatic turn towards more sophisticated mathematics. For example, imagine the stock in a company that is headed for (or has already declared) bankruptcy. It is only a matter of time before the stock price is essentially 0 (say). Let us call it the *time to failure* of the stock. The waiting time for this event could, at least in theory, be any positive real number (assume the stock trades 24 hours per day) so the sample space is the set Ω of all positive real numbers, which is uncountable.

Unlike the case of a discrete sample space, we cannot simply assign a probability to each of the uncountably many times to failure because it is a fact of mathematics that the sum of *uncountably* many positive numbers is never finite, let alone equal to 1. So rather than attempt to determine probabilities for individual outcomes (failure times), we must limit ourselves to assigning probabilities directly to events. However, not all subsets of the sample space can qualify as events. This issue gets rather involved and we will not it here.

The most direct and elegant way to assign probabilities to events is to use a *function*. Figure 1 shows how this might be done.

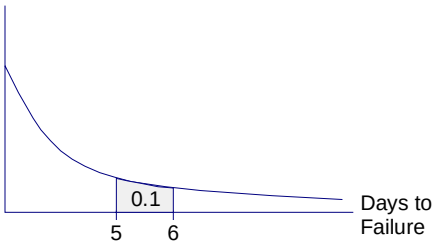


Figure 1–A probability density function

This figure shows the graph of a function that specifies the probability of failure for any *time interval*. In particular, it is the *area under the curve* that specifies the probability. For example, the probability that failure will occur sometime between the 5-th and 6-th day is the area under the curve between the vertical lines $x = 5$ and $x = 6$, which is 0.1. This function is referred to as a *probability density function*. Probability density functions, such as the well-known *bell shaped curve* that students often want professors to use in determining their grades, are often, but

not always, used to specify probabilities in the nondiscrete case. Indeed, some probability measures cannot be specified using a probability density function.

In any case, the point we wish to make at this time is that even the specification of probabilities requires much more mathematical machinery in the nondiscrete case than in the discrete case.

We will need only *finite* probability for our study of *discrete-time* pricing models. We will discuss some aspects of the general theory (including the nondiscrete case) much later in the book, as a prelude to our discussion of the Black-Scholes option pricing formula.

So let us proceed to set down the basic principles of the subject of finite probability. Since this is not, after all, a textbook on probability, we will tend to be brief, covering what we need for our immediate purposes. In a subsequent chapter, we will expand our discussion of discrete probability to cover what is necessary to make sense of the general discrete-time pricing model.

Probability Spaces

We may as well begin with the main definition.

Definition *A* ***finite probability space*** *is a pair* $(\Omega, \mathbb{P})$ *consisting of a finite nonempty set* Ω*, called the* ***sample space*** *and a real-valued function* $\mathbb{P}$ *defined on the set of all subsets of* Ω*, called a* ***probability measure*** *on* Ω*. The function* $\mathbb{P}$ *must satisfy the following properties.*

1) (Range) For all $A \subseteq \Omega$

$$0 \leq \mathbb{P}(A) \leq 1$$

2) (Probability of Ω*)*

$$\mathbb{P}(\Omega) = 1$$

3) (Additivity property) If A *and* B *are* disjoint *then*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$

In this context, subsets of Ω *are called* **events**. □

As mentioned earlier, the sample space is intended to represent the set of all possible outcomes of an experiment. The probability $\mathbb{P}(\omega)$ of a

particular outcome ω is intended to represent the likelihood that the outcome of the experiment will be ω.

On the other hand, this is all intuition, not mathematics. Formally speaking, all we care about is that Ω is a finite nonempty set and $\mathbb{P}$ is a probability measure as defined by the properties in the definition. Property 2) says that the event consisting of the entire sample space is a certain event, that is, any outcome must lie in the sample space. Property 3) says that if two events have nothing in common, then the likelihood that *either one* occurs is the sum of the likelihood of each event. Note that it is vital that the events be disjoint for this to hold.

Sometimes we will forget ourselves and engage in a common abuse of terminology by referring to the set Ω by itself as a probability space. In this case, the probability measure $\mathbb{P}$ still exists, but we just don't need to mention it explicitly at that time. The student would be well-advised to avoid this peccadillo.

Probability Mass Functions

If Ω is a finite set, then for each $\omega \in \Omega$ the event $\{\omega\}$ is called an **elementary event**. The simplest way to define a probability measure on a finite sample space Ω is just to specify the probability of all elementary events. Equivalently, we assign to each of the elements $\omega \in \Omega$ a number p_ω satisfying $0 \leq p_\omega \leq 1$ and

$$\sum_{\omega \in \Omega} p_\omega = 1$$

Then we can define a probability measure $\mathbb{P}$ by setting

$$\mathbb{P}(\{\omega\}) = p_\omega$$

and extending this to all events by finite additivity. This is a fancy way of saying that the probability of any event E is the sum of the probabilities of the elementary events contained in E.

The set $\{p_\omega \mid \omega \in \Omega\}$ is referred to as a **probability distribution** and the function $f : \Omega \to \mathbb{R}$ defined by

$$f(\omega) = p_\omega$$

is called a **probability mass function**. (Do not confuse the term *probability distribution* with the term *distribution function*, which has a different meaning that we will define in a later chapter.)

Note the subtle but important difference between the probability measure $\mathbb{P}$ and the probability mass function f, namely, $\mathbb{P}$ is defined on all *subsets* of Ω whereas f is defined on all *elements* of Ω.

When a probability distribution is given, the probability of any event $A \subseteq \Omega$ is the *sum* of the probabilities of the outcomes in the event, that is

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

Moreover, if each outcome in the sample space is equally likely, that is, if each outcome has the same probability, this probability is $\frac{1}{|\Omega|}$ and so the probability of any event E is simply the size of E divided by the size of the sample space Ω, that is

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|}$$

EXAMPLE 1 Studies of the price history of a certain stock over the last several years have shown that, for the month of January, the probability that the stock will reach a certain maximum value during the month is as follows

$$\begin{aligned}
\mathbb{P}(0{-}4.99) &= 0.65 \\
\mathbb{P}(5{-}9.99) &= 0.2 \\
\mathbb{P}(10{-}14.99) &= 0.1 \\
\mathbb{P}(15{-}19.99) &= 0.04 \\
\mathbb{P}(20{-}24.99) &= 0.01
\end{aligned}$$

What is the probability that the stock will reach 10 dollars during the month? What is the probability that the stock will either not reach 5 dollars during the month or will reach 20 dollars?

Solution The stock will reach 10 dollars during the month if and only if the maximum stock price during the month is at least 10. Hence

$$\begin{aligned}
&\mathcal{P}(\text{price reaches } 10) \\
&\qquad = \mathcal{P}(10{-}14.99) + \mathcal{P}(15{-}19.99) + \mathcal{P}(20{-}24.99) \\
&\qquad = 0.1 + 0.04 + 0.01 = 0.15
\end{aligned}$$

Similarly,

$$\begin{aligned}\mathcal{P}(\text{not reach } 5 \text{ or reach } 20) &= \mathcal{P}(0{-}4.99) + \mathcal{P}(20-24) \\ &= 0.65 + 0.01 = 0.66\end{aligned}$$ □

Probability theory tends to have its own vocabulary, even when it comes to simple concepts like the disjointness of sets.

Definition *When two events A and B are disjoint as sets, we say that they are* ***mutually exclusive****. When a collection $\{A_1, \ldots, A_n\}$ of events satisfies*

$$A_i \cap A_j = \emptyset$$

for all i,j we say that the collection is ***pairwise mutually exclusive****.* □

Some easy consequences of the definition of probability space are given below.

Theorem 1 *Let $(\Omega, \mathbb{P})$ be a finite probability space. Then*
1) (Probability of empty event)

$$\mathbb{P}(\emptyset) = 0$$

2) (Monotonicity)

$$A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$$

3) (Probability of the complement)

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$$

4) (Finite additivity property) If $\{A_1, \ldots, A_n\}$ is a finite collection of pairwise mutually exclusive *events in Ω then*

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n)$$ □

Partitions and the Theorem on Total Probabilities

The following simple concept will play a central role in our discussion of derivative pricing models.

Definition *Let Ω be a nonempty set. Then a* **partition** *of Ω is a collection $\mathcal{P} = \{B_1, \ldots, B_n\}$ of* nonempty *subsets of Ω, called the* **blocks** *of the partition, with the following properties*

1) The blocks are pairwise disjoint

$$B_i \cap B_j = \emptyset$$

for all i, j.

2) The union of the blocks is all of Ω

$$B_1 \cup \cdots \cup B_n = \Omega \qquad \square$$

The following important theorem says we can determine the probability of an event E if we can determine the probability of that portion of E that belongs to each block of a partition. We leave proof to the reader.

Theorem 2 *(***Theorem on Total Probabilities***) Let Ω be a sample space and let $E_1, \ldots, E_n$ be events that form a partition of Ω. Then for any event A in Ω,*

$$\mathbb{P}(A) = \sum_{k=1}^{n} \mathbb{P}(A \cap E_k) \qquad \square$$

Independence

A *fair* coin is one for which the probability of heads is $1/2$. Indeed, this is the definition of the term *fair coin*. Suppose we toss a fair coin 99 times and get heads each time, admittedly an unlikely event but nevertheless possible. Would you be willing to bet that the 100-th toss will result in another heads? Many people would not, reasoning (incorrectly) that since heads has occurred so many times in a row, an outcome of tails is way "overdue."

The fact is, however, that the outcome of each toss of the coin is *independent* of the outcomes of the other tosses, and so the probability of getting a heads on the 100-th toss is still $1/2$, despite the previous results.

Perhaps the reason for confusion on this point has to do with the probability of getting 99 heads in a row in the first place, which is certainly very small. But once that has happened, the extreme unlikeliness has been "factored out" so-to-speak and we are back to the likeliness of the outcome of a single toss.

Intuitively speaking, two events are independent if the knowledge that (or assumption that) one event will happen does not effect the probability of the other event happening. We will be able to make this statement

precise when we discuss conditional probability in a later chapter. In any case, we can state the formal definition of independence now.

Definition *The events* E *and* F *on the probability space* $(\Omega, \mathbb{P})$ *are* **independent** *if the probability that* both *events occur is the product of the probabilities of the events, in symbols*

$$\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F) \qquad \square$$

For example, suppose that a certain stock can move up in price or down in price over a day and a certain bond can do likewise. If we *assume* that the actions of the stock and the bond are independent then

$$\mathbb{P}(\text{stock up } \textit{and} \text{ bond down}) = \mathbb{P}(\text{stock up})\mathbb{P}(\text{bond down})$$

We can also define independence of a collection of events.

Definition *The collection of events* $\{E_1, \ldots, E_n\}$ *is* **independent** *if for any subcollection* $\{E_{i_1}, \ldots, E_{i_k}\}$ *of these events we have*

$$\mathbb{P}(E_{i_1} \cap \cdots \cap E_{i_k}) = \mathbb{P}(E_{i_1}) \cdots \mathbb{P}(E_{i_k}) \qquad \square$$

Note that to check whether or not 3 events A, B and C are independent, we must check 3 conditions

A and B are independent
A and C are independent
B and C are independent

In general, to check that a collection of k events is independent, we must check a total of $1/2^k - 1$ conditions. Thus, the number of conditions grows very rapidly with the number of events.

Binomial Probabilities

The simplest type of meaningful experiment is one that has only two outcomes. Such experiments are referred to as **Bernoulli experiments**, or **Bernoulli trials**. The two outcomes are often described by the terms **success** and **failure**, and the probability of success is usually denoted by p. Hence, the probability of failure is $1 - p$.

For example, tossing a coin is a Bernoulli experiment, where we may consider heads as success and tails as failure (or vice-versa). As a more relevant example, we will consider a derivative pricing model in which at any given time t_k the price of a certain stock may rise from its previous

value S to Su or it may fall from its previous value S to Sd where $0 < d < 1 < u$. Thus, at each time t_k we have a Bernoulli experiment.

If a Bernoulli experiment with probability of success p is repeated n times, this is called a **binomial experiment** with n **trials**. Note that since the *exact same experiment* is being repeated, the outcomes of the trials are *independent*, that is, the outcome of the k-th trial does not effect the outcome of the $(k+1)$-st trial. The **parameters** of the binomial experiment are p and n.

For example, tossing a coin n times is a binomial experiment. Drawing a card n times, with success being the drawing of an ace, is a binomial experiment *provided that we replace each card before drawing the next card*. This is necessary since we must repeat the *same* binomial experiment each time.

Because the individual Bernoulli trials in a binomial experiment are independent, it is easy to compute the probability of any particular outcome of the binomial experiment, as the following example illustrates. Indeed, we will study a generalization of the following example carefully in a later chapter.

EXAMPLE 2 Consider a stock whose price can change at any one of 6 times

$$t_0 < t_1 < \cdots < t_5$$

Suppose the stock's initial price at time t_0 is S. Moreover, during each time interval $[t_k, t_{k+1}]$ the stock price goes up by a factor of u or down by a factor of d, where $0 < d < 1 < u$, independently of the previous changes in the price. The probability that the stock price goes up is p. Thus, for each time interval we have a Bernoulli experiment with probability of success p. Moreover, the entire price history is a binomial experiment with parameters p and $n = 5$.

A typical outcome of this binomial experiment can be written as a sequence of U's and D's of length 5 and so the sample space is the set

$$\Omega = \{U, D\}^5$$

of all such sequences. For instance, the sequence $UUDUD$ says that during the intervals $[t_0, t_1]$, $[t_1, t_2]$ and $[t_3, t_4]$ the stock price went up

whereas during the intervals $[t_2, t_3]$ and $[t_4, t_5]$ the stock price when down.

To compute the probability of this outcome, we use the fact that the individual trials are independent and so the probability of their intersection is the product of their probabilities. Thus

$$\mathbb{P}(UUDUD) = pp(1-p)p(1-p) = p^3(1-p)^2$$

It is clear that the probability of an element $\omega \in \Omega$ depends only upon the *number* of U's and D's in ω and not their order. Thus, if we set

$$N_U(\omega) = \text{number of } U\text{'s in } \omega$$
$$N_D(\omega) = \text{number of } D\text{'s in } \omega$$

then

$$\mathbb{P}(\omega) = p^{N_U(\omega)}(1-p)^{N_D(\omega)}$$

Let us compute the probability of the event of having exactly 3 up-ticks in the stock price. The tedious method is to list all such price histories thusly

$UUUDD$	$UUDUD$
$UUDDU$	$UDUUD$
$UDUDU$	$UDDUU$
$DUUUD$	$DUUDU$
$DUDUU$	$DDUUU$

Since there are 10 of these histories and each one has probability $p^3(1-p)^2$ the probability is $10p^3(1-p)^2$.

The smart way to compute this probability is to observe that there are $\binom{5}{3} = 10$ such histories—one for each way to choose the 3 spots for the U's. Since each history has probability $p^3(1-p)^2$, the probability of the event is $10p^3(1-p)^2$.

It is now easy to generalize this result. The probability of having exactly k up-ticks (and thus $n-k$ down-ticks) is just

$$\binom{n}{k} p^k (1-p)^{n-k} \qquad \square$$

We have established the following useful result.

Theorem 3 *Consider a binomial experiment with parameters p and n. The sample space of this experiment is the set $\Omega = \{s, f\}^n$ of all sequences of s's and f's of length n, where s stands for success and f for failure. For any $\omega \in \Omega$ let*

$$N_s(\omega) = \textit{number of } s\textit{'s in } \omega$$

1) If $\omega \in \Omega$ then

$$\mathbb{P}(\omega) = p^{N_s(\omega)}(1-p)^{n-N_s(\omega)}$$

2) The probability of getting exactly k *successes is given by*

$$\mathbb{P}(\textit{exactly } k \textit{ successes}) = \binom{n}{k} p^k (1-p)^n \qquad \square$$

EXAMPLE 3 Four cards are drawn, with replacement, from a deck of cards. What is the probability of getting at least 3 aces?

Solution The probability of getting at least 3 aces is equal to the probability of getting exactly 3 aces plus the probability of getting exactly 4 aces. Since we are dealing with a binomial experiment, with probability of success (getting an ace) equal to $p = \frac{4}{52} = \frac{1}{13}$ we have

$$\begin{aligned}
&\mathbb{P}(\text{getting at least 3 aces}) \\
&\qquad = \mathbb{P}(\text{getting exactly 3 aces}) + \mathbb{P}(\text{getting exactly 4 aces}) \\
&\qquad = \binom{4}{3}\left(\frac{1}{13}\right)^3\left(\frac{12}{13}\right)^1 + \binom{4}{4}\left(\frac{1}{13}\right)^4\left(\frac{12}{13}\right)^0 \\
&\qquad = \frac{49}{28561} \\
&\qquad \approx 0.0017
\end{aligned}$$

which is quite small. $\square$

The probability distribution described in the previous example and theorem is extremely important.

Definition *Let $0 < p < 1$ and let n be a positive integer. Let $\Omega = \{0, \ldots, n\}$. The probability distribution on Ω with mass function*

$$b(k; n, p) = \binom{n}{k} p^k (1-p)^n$$

for $k = 0, \ldots, n$ is called the **binomial distribution**. *This distribution gives the probability of getting exactly k successes in a binomial experiment with parameters p and n.* $\square$

Figure 1 shows the graph of two binomial distributions.

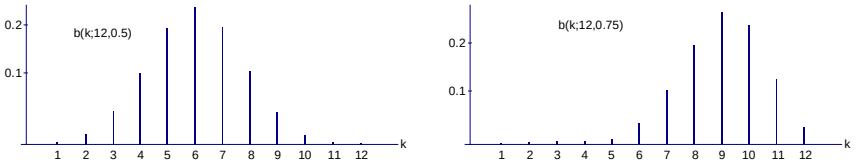


Figure 1 – Binomial distributions

Empirical Versus Theoretical Probabilities

As alluded to earlier, there are two common ways in which to assign probabilities. Consider, for example, the problem of setting the value of the probability p in Example 2. This is the probability that the stock price will rise.

One approach is to carefully examine the history of the stock's price over a substantial period of time. Then we can estimate p by taking the number of times that the stock price increased divided by the total number of times. For instance, if the stock price increased 5003 times in the last 10000 time periods, then we can set

$$p = \frac{5003}{10000}$$

Of course, it follows that the probability of a decrease is

$$1 - p = 1 - \frac{5003}{10000} = \frac{4997}{10000}$$

Because these probabilities are the result of analyzing empirical data, or at least because they are the result of some analysis of actual physical phenomena, they are referred to as **empirical probabilities**.

On the other hand, we could simply have assumed, perhaps through lack of any actual data for analysis that $p = \frac{1}{2}$. This type of probability is termed a **theoretical probability**. As we will see, both types of probabilities have their place in the mathematics of finance.

Random Variables

The following concept is key.

Definition *A real-valued function* $X: \Omega \to \mathbb{R}$ *defined on a finite sample space* Ω *is called a* **random variable** *on* Ω. *The set of all random variables on* Ω *is denoted by* $RV(\Omega)$. □

As the definition states, for finite (or discrete) probability spaces, a random variable is nothing more or less than a real-valued function. However, as we will see in a later chapter, for nondiscrete sample spaces, not all real-valued functions can qualify as random variables.

Since $\mathrm{RV}(\Omega)$ is just the set of all functions on Ω, it is a *vector space* under ordinary addition and scalar multiplication of functions. Thus, if X and Y are random variables on Ω and $a, b \in \mathbb{R}$ then

$$aX + bY$$

is a random variable on Ω. Note also that the product of two random variables on Ω is a random variable on Ω.

One of the most useful types of random variables are those that identify specific events.

Definition *Let* A *be an event in* Ω. *The function* 1_A^Ω *defined by*

$$1_A^\Omega(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

is called the **indicator function** *(or* **indicator random variable***) for* A. *When the set* Ω *is clear, we may also write* 1_A *for the indicator function for* A.□

EXAMPLE 4 Let

$$\Omega = \{0.5, 0.75, 1, 1.25, 1.5, 1.75\}$$

be a sample space of possible Federal discount rates. Consider a company whose stock price tends to fluctuate with interest rates. The stock prices can be represented by a random variable S on Ω. For example

$$\begin{aligned} S(0.5) &= 105 \\ S(0.75) &= 100 \\ S(1) &= 100 \\ S(1.25) &= 100 \\ S(1.5) &= 95 \\ S(1.75) &= 90 \end{aligned}$$

The event that $\{S = 100\}$ is the event consisting of the discount rates $\{0.75, 1, 1.25\}$, that is,

$$\{S = 100\} = \{0.75, 1, 1.25\} \qquad \square$$

EXAMPLE 5 Consider the experiment of rolling two fair dice and recording the values on each die. The sample space consists of the 36 ordered pairs

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

Since the dice are fair, each ordered pair is equally likely to occur and so the probability of each outcome is $1/36$.

However, for some games of chance, we are interested only in the *sum* of the two numbers on the dice. So let us define a random variable $S: \Omega \to \mathbb{R}$ by

$$S(a, b) = a + b$$

The event $\{S = 7\}$ of getting a sum of 7 is

$$\{S = 7\} = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$$

and

$$\mathbb{P}(\text{sum equals } 7) = \mathbb{P}(S = 7) = \frac{6}{36} = \frac{1}{6} \qquad \square$$

Perhaps the most fundamental fact about random variables is that they are used to identify events. In fact, there are times when we don't really care about the actual values of S—we only care about the events that are represented by these values. For example, in the previous example if we instead used the “doubled sum” random variable

$$D((a, b)) = 2(a + b)$$

then D serves equally well to describe the relevant events in the game of chance. For instance, $\{S = 7\} = \{D = 14\}$.

Of course, this is not always the case. The actual values of some random variables are important in their own right. Some examples are stock price random variables, the interest rate random variables, the cost random variables. All of these random variables serve to describe a particular set of events in which we are interested. Let us look at this a bit more formally.

The Probability Distribution of a Random Variable

Let X be a random variable on a probability space $(\Omega, \mathbb{P})$ where $\Omega = \{\omega_1, \ldots, \omega_n\}$. Since Ω is finite, X takes on a finite number of possible values, say $\mathcal{A} = \{x_1, \ldots, x_m\}$. For each x_i we can form the event

$$\{X = x_i\} = X^{-1}(x_i) = \{\omega_j \mid X(\omega_j) = x_i\}$$

which is simply the inverse image of x_i. The expression $\{X = x_i\}$ is the most common notation for events described by random variables. Since the range of a random variable is the set of real numbers, we can also consider events such as

$$\{X \leq x_i\} = X^{-1}((-\infty, x_i)) = \{\omega_j \mid X(\omega_j) \leq x_i\}$$

The events

$$\{X = x_1\}, \{X = x_2\}, \ldots, \{X = x_m\}$$

form a *partition* of the sample space Ω, that is, the events are pairwise disjoint and their union is all of Ω (because X must be defined on all of Ω). Thus,

$$\sum_{i=1}^{m} \mathbb{P}(X = x_i) = 1$$

Note that it is customary to replace the somewhat cumbersome notation $\mathbb{P}(\{X = x\})$ by the simpler $\mathbb{P}(X = x)$.

It follows that the numbers $\mathbb{P}(X = x_i)$ form a probability distribution on the set $\mathcal{A} = \{x_1, \ldots, x_m\}$, which is a subset of $\mathbb{R}$. Thus, the random variable X describes a probability measure $\mathbb{P}_X$ on the set $\mathcal{A}$ of values of X by

$$\mathbb{P}_X(\{x_i\}) = \mathbb{P}(X = x_i)$$

This is called the **probability measure** (or **probability distribution**)

defined by X. The corresponding probability mass function $f: \mathcal{A} \to \mathbb{R}$ defined by

$$f(x_i) = \mathbb{P}(X = x_i)$$

is called the **probability mass function** of X.

Thus, for example, to say that a random variable X has a binomial distribution with parameters p and n is to say that the values of X are $\{0, \ldots, n\}$ and that the probability mass function of X is the function

$$\mathbb{P}(X = k) = b(k; n, p) = \binom{n}{k} p^k (1-p)^n$$

It is also common to say in this case that X is *binomially distributed*.

These facts about random variables are so important that they bear repeating. Random variables are used to identify certain relevant events from the sample space. Moreover, a random variable serves to "transfer" the probability measure from the events in the sample space that it identifies to the range of the random variable in $\mathbb{R}$.

We will also have need of random vectors.

Definition *A function $X: \Omega \to \mathbb{R}^n$ from a sample space Ω to the vector space $\mathbb{R}^n$ is called a* **random vector** *on Ω.* □

The set $\mathrm{RV}^n(\Omega)$ of all random vectors on a sample space Ω is also a *vector space* under ordinary addition and scalar multiplication of functions.

EXAMPLE 6 Let

$$\Omega = \{0.5, 0.75, 1, 1.25, 1.5, 1.75\}$$

be a sample space of possible Federal discount rates. Consider a company whose stock price tends to fluctuate with interest rates. Of course, bond prices also fluctuate with respect to interest rates. We might define the price random vector $S: \Omega \to \mathbb{R}^2$ by $S(\omega) = (s, b)$ where s is the price of the stock and b is the price of the bond when the discount rate is ω. For example,

$$S(0.5) = (105, 112)$$

means that if the discount rate is 0.5% then the stock price is 105 and the bond price is 112.□

Independence of Random Variables

Two random variables X and Y on the sample space Ω are independent if the events $\{X = x\}$ and $\{Y = y\}$ are independent for all choices of x and y. Intuitively, this says that knowing the value of one of the random variables provides no knowledge of the value of the other random variable. Here is a more formal definition, where the notation $\mathbb{P}(X = x, Y = y)$ is shorthand for $\mathbb{P}(\{X = x\} \cap \{Y = y\})$.

Definition *The random variables X and Y on Ω are* **independent** *if*

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$$

for all $x, y \in \mathbb{R}$. More generally, the random variables $X_1, \ldots, X_n$ are independent if

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i)$$

for all $x_1, \ldots, x_n \in \mathbb{R}$. □

Expectation

The notion of expected value plays a central role in the mathematics of finance.

Definition *Let X be a random variable on a finite probability space $(\Omega, \mathbb{P})$ where $\Omega = \{\omega_1, \ldots, \omega_n\}$. The* **expected value** *(also called the* **expectation** *or* **mean***) of X is given by*

$$\mathcal{E}_{\mathbb{P}}(X) = \sum_{i=1}^{n} X(\omega_i)\mathbb{P}(\omega_i)$$

This is the sum of terms of the form: value of X at ω_i times probability that ω_i occurs. If X takes on the distinct *values $\{x_1, \ldots, x_m\}$ then we also have*

$$\mathcal{E}_{\mathbb{P}}(X) = \sum_{i=1}^{m} x_i \mathbb{P}(X = x_i)$$

(Note the different upper limit of summation) This is a weighted sum of

the values of X, each value weighted by its probability of occurring. The expected value of X is also denoted by μ_X.□

The expected value function $\mathcal{E}\colon \mathrm{RV}(\Omega) \to \mathbb{R}$ maps random variables on $(\Omega, \mathbb{P})$ to real numbers. One of the most important properties of this function is that it is *a linear functional.*

Theorem 4 *The expectation function $\mathcal{E}\colon RV(\Omega) \to \mathbb{R}$ is a linear functional, that is, for any random variables X and Y and real numbers a and b*

$$\mathcal{E}(aX + bY) = a\mathcal{E}(X) + b\mathcal{E}(Y)$$

Proof. Let us suppose that X has values $\{x_1, \ldots, x_n\}$ and Y has values $\{y_1, \ldots, y_m\}$. Then $aX + bY$ has values $ax_i + by_j$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. To compute the expected value of $aX + bY$ consider the events

$$E_{i,j} = \{X = x_i, Y = y_j\}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. These events form a partition of Ω with the property that $aX + bY$ has *constant* value $ax_i + by_j$ on $E_{i,j}$ and so, using the theorem on total probabilities, we have

$$\begin{aligned}
\mathcal{E}(aX + bY) &= \sum_{i=1}^{n}\sum_{j=1}^{m} (ax_i + by_j)\mathbb{P}(X = x_i, Y = y_j) \\
&= a\sum_{i=1}^{n} x_i \left[\sum_{j=1}^{m} \mathbb{P}(X = x_i, Y = y_j)\right] \\
&\quad + b\sum_{j=1}^{m} y_j \left[\sum_{i=1}^{n} \mathbb{P}(X = x_i, Y = y_j)\right] \\
&= a\sum_{i=1}^{n} x_i\mathbb{P}(X = x_i) + b\sum_{j=1}^{m} y_j\mathbb{P}(Y = y_j) \\
&= a\mathcal{E}(X) + b\mathcal{E}(Y)
\end{aligned}$$

as desired.□

Expected Value of a Function of a Random Variable

Note that if $f\colon \mathbb{R} \to \mathbb{R}$ is a real-valued function of a real variable and X is a random variable, then the composition $f(X)\colon \Omega \to \mathbb{R}$ is also a random variable. (For finite probability spaces, this is nothing more than the fact that the composition of functions is a function.)

The expected value of the random variable $f(X)$ is equal to

$$\mathcal{E}_{\mathbb{P}}(f(X)) = \sum_{i=1}^{n} f(X(\omega_i))\mathbb{P}(\omega_i)$$

or

$$\mathcal{E}_{\mathbb{P}}(f(X)) = \sum_{i=1}^{m} f(x_i)\mathbb{P}(X = x_i)$$

When there is no need to emphasize the probability measure, we will drop the subscript and write $\mathcal{E}$ instead of $\mathcal{E}_{\mathbb{P}}$ but is is important to keep in mind that the expectation depends on the probability.

EXAMPLE 7 Consider a stock whose current price is 100 and whose price at time T depends on the state of the economy, which may be one of the following states

$$\Omega = \{\mathfrak{s}_1, \mathfrak{s}_2, \mathfrak{s}_3, \mathfrak{s}_4\}$$

The probabilities of the various states are given by

$$\begin{aligned}\mathbb{P}(\mathfrak{s}_1) &= 0.2\\ \mathbb{P}(\mathfrak{s}_2) &= 0.3\\ \mathbb{P}(\mathfrak{s}_3) &= 0.3\\ \mathbb{P}(\mathfrak{s}_4) &= 0.2\end{aligned}$$

The stock price random variable is given by

$$\begin{aligned}S(\mathfrak{s}_1) &= 99\\ S(\mathfrak{s}_2) &= 100\\ S(\mathfrak{s}_3) &= 101\\ S(\mathfrak{s}_4) &= 102\end{aligned}$$

If we purchase one share of the stock now the expected return at time T is

$$\mathcal{E}(S) = 99(0.2) + 100(0.3) + 101(0.3) + 102(0.2) = 100.5$$

and so the expected profit is $100.5 - 100 = 0.5$. Consider a derivative whose return D is a function of the stock price, say

$$\begin{aligned} D(99) &= -4 \\ D(100) &= 5 \\ D(101) &= 5 \\ D(102) &= -6 \end{aligned}$$

Thus D is a random variable on Ω. The expected return of the derivative is

$$\begin{aligned} \mathcal{E}(\text{return}) &= D(99)\mathbb{P}(99) + D(100)\mathbb{P}(100) \\ &\quad + D(101)\mathbb{P}(101) + D(102)\mathbb{P}(102) \\ &= -4(0.2) + 5(0.3) + 5(0.3) - 6(0.2) \\ &= 1 \end{aligned}$$

□

The previous example points out a key property of expected values. The expected value is seldom the value expected! In this example, we never expect to get a return of 100.5. In fact, this return is impossible. The return must be one of the numbers in the sample space. The expected value is an *average*, not the value most expected. (The value most expected is called the *mode*.)

Expectation and Independence

We have seen that the expected value operator is linear, that is,

$$\mathcal{E}(aX + bY) = a\mathcal{E}(X) + b\mathcal{E}(Y)$$

It is natural to wonder about also about $\mathcal{E}(XY)$. Let us suppose that X has values $\{x_1, \ldots, x_n\}$ and Y has values $\{y_1, \ldots, y_m\}$. Then the product XY has values $x_i y_j$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

Consider the events

$$E_{i,j} = \{X = x_i, Y = y_j\}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m$, which form a partition of Ω with the property that XY has *constant* value $x_i y_j$ on $E_{i,j}$. Hence

$$\mathcal{E}(XY) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j \mathbb{P}(X = x_i, Y = y_j)$$

In general, we can do nothing with the probabilities $\mathbb{P}(X = x_i, Y = y_j)$. However, if X and Y are *independent* then

$$\begin{aligned}\mathcal{E}(XY) &= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j \mathbb{P}(X = x_i, Y = y_j) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j) \\ &= \left[\sum_{i=1}^{n} x_i \mathbb{P}(X = x_i)\right]\left[\sum_{j=1}^{m} y_j \mathbb{P}(Y = y_j)\right] \\ &= \mathcal{E}(X)\mathcal{E}(Y)\end{aligned}$$

Thus, we have an important theorem.

Theorem 5 *If X and Y are* independent *random variables on a probability space $(\Omega, \mathbb{P})$ then*

$$\mathcal{E}(XY) = \mathcal{E}(X)\mathcal{E}(Y) \qquad \square$$

This theorem can be generalized to the product of more than two independent random variables. For example, it is not hard to see that if X, Y and Z are independent, then XY and Z are also independent and so

$$\mathcal{E}(XYZ) = \mathcal{E}(XY)\mathcal{E}(Z) = \mathcal{E}(X)\mathcal{E}(Y)\mathcal{E}(Z)$$

Variance and Standard Deviation

The expectation of a random variable X is a measure of the "center" of the distribution of X. A common measure of the "spread" of the values of a random variable is the variance and its square root, which is called the standard deviation. The advantage of the standard deviation is that it has the same units as the random variable. However, its disadvantage is the often awkward presence of the square root.

Definition *Let X be a random variable with finite expected value μ. The* **variance** *of X is*

$$\sigma_X^2 = \text{Var}(X) = \mathcal{E}((X - \mu)^2)$$

and the **standard deviation** *is the positive square root of the variance*

$$\sigma_X = SD(X) = \sqrt{\text{Var}(X)} \qquad \square$$

The following theorem gives some simple properties of the variance.

Theorem *6 Let X be a random variable with finite expected value μ. Then*
1) $\text{Var}(X) = \mathcal{E}(X^2) - \mu^2 = \mathcal{E}(X^2) - \mathcal{E}(X)^2$
2) For any real number a

$$\text{Var}(aX) = a^2\text{Var}(X)$$

3) If X and Y are independent *random variables then*

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Proof. We leave proof as an exercise.□

Note that, unlike the expectation operator, the variance is *not* linear. Thus, the quantities

$$\text{Var}(aX + bY)$$

and

$$a\text{Var}(X) + b\text{Var}(Y)$$

are not the same. We will explore this matter further a bit later in the chapter.

Expected Value of a Binomial Random Variable

We can easily compute the expected value and variance of a binomial random variable.

Theorem 7 *Let X be a binomial random variable with distribution $b(k; n, p)$. Then*

$$\mathcal{E}(X) = np$$
$$\text{Var}(X) = np(1 - p)$$

Proof. Let $q = 1 - p$. For the expected value, we have

$$\begin{aligned}
\mathcal{E}(X) &= \sum_{k=0}^{n} k\mathbb{P}(X = k) \\
&= \sum_{k=1}^{n} k\binom{n}{k}p^k q^{n-k} \\
&= np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}q^{(n-1)-(k-1)} \\
&= np\sum_{k=0}^{n-1}\binom{n-1}{k}p^k q^{(n-1)-k} \\
&= np
\end{aligned}$$

We leave derivation of the variance as an exercise.□

Covariance and Correlation; Best Linear Predictor

We now wish to explore the relationship between two random variables defined on the same sample space.

Definition *If X and Y are random variables with finite means then the* ***covariance*** *of X and Y is defined by*

$$\sigma_{X,Y} = Cov(X, Y) = \mathcal{E}[(X - \mu_X)(Y - \mu_Y)] \qquad \square$$

Some properties of the covariance are given in the next theorem.

Theorem 8 *The covariance satisfies the following properties.*

1) Covariance in terms of expected values

$$Cov(X, Y) = \mathcal{E}(XY) - \mathcal{E}(X)\mathcal{E}(Y)$$

2) (Symmetry)

$$Cov(X, Y) = Cov(Y, X)$$

3) The covariance of X with itself is just the variance of X

$$Cov(X, X) = \sigma_X^2$$

4) If X is a constant random variable (that is, if $\sigma_X = 0$) then

$$Cov(X, Y) = 0$$

5) The covariance function is linear in both coordinates (that is, it is bilinear*)*

$$Cov(aX + bY, Z) = aCov(X, Z) + bCov(Y, Z)$$

6) The covariance is bounded by the product of the standard deviations

$$|Cov(X, Y)| \leq \sigma_X \sigma_Y$$

Moreover, equality holds if and only if either one of X or Y is constant or if there are constants a and b for which

$$Y = aX + b$$

Proof. We prove only part 6). If X or Y is constant then the result follows since both sides are 0, so let us assume otherwise. Let t be a real variable. Then

$$\begin{aligned} 0 &\leq \mathcal{E}((tX + Y)^2) \\ &= \mathcal{E}(t^2X^2 + 2tXY + Y^2) \\ &= t^2\mathcal{E}(X^2) + 2t\mathcal{E}(XY) + \mathcal{E}(Y^2) \\ &= f(t) \end{aligned}$$

where $f(t)$ is a quadratic function in t. Since $f(t) \geq 0$ and since the leading coefficient of $f(t)$ is positive, we conclude that the discriminant of $f(t)$ must be nonpositive (draw the graph and look at the zeros), that is,

$$[2\mathcal{E}(XY)]^2 - 4\mathcal{E}(X^2)\mathcal{E}(Y^2) \leq 0$$

or

$$\mathcal{E}(XY)^2 \leq \mathcal{E}(X^2)\mathcal{E}(Y^2)$$

Furthermore, equality holds (the discriminant is 0) if and only if there is a value of t for which $f(t) = \mathcal{E}((tX + Y)^2) = 0$. But this is possible if and only if $Y = -tX$. (Our assumption that $Y$is not constant implies that $t \neq 0$.)

Since this applies to any random variables X and Y we can also apply it to the random variables $X - \mu_X$ and $Y - \mu_Y$ to conclude that

$$[\text{Cov}(XY)]^2 \leq \sigma_X^2 \sigma_Y^2$$

that is

$$|\text{Cov}(XY)| \leq \sigma_X \sigma_Y$$

with equality holding if and only if at least one of X or Y is constant or there is a nonzero real number a such that

$$Y - \mu_Y = a(X - \mu_X)$$

that is

$$Y = aX - a\mu_X + \mu_Y = aY + b$$

This concludes the proof of part 6).□

The following definition gives a *dimensionless* version of covariance.

Definition *If X and Y have finite means and nonzero variances then the* ***correlation coefficient*** *of X and Y is*

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$ □

It follows immediately that

$$-1 \leq \rho_{X,Y} \leq 1$$

Moreover, as we will soon see, $\rho_{X,Y}$ assumes one of the boundary values ± 1 if and only if there is a linear relationship between X and Y, that is, there exist constants $a \neq 0$ and b for which

$$Y = aX + b$$

In fact, $\rho_{X,Y} = +1$ implies that the slope $a > 0$ and $\rho_{X,Y} = -1$ implies that $a < 0$. Thus, if $\rho_{X,Y} = +1$ then Y moves in the same direction as X (both increase or both decrease) whereas if $\rho_{X,Y} = -1$ then Y decreases when X increases and vice-versa.

Also, it is easy to see that if X and Y are independent then $\rho_{X,Y} = 0$. However, the converse is not true. The condition $\rho_{X,Y} = 0$ does *not* imply that X and Y are independent.

Two random variables are said to be **uncorrelated** if $\rho_{X,Y} = 0$, **perfectly positively correlated** if $\rho_{X,Y} = 1$ and **perfectly negatively correlated** if $\rho_{X,Y} = -1$. We will have much use for these terms during our study of portfolio risk management.

Best Linear Predictor

Let us examine the meaning of the correlation coefficient more closely. It is often said that the correlation coefficient is a measure of the *linear relationship* between X and Y. Indeed, we have just said that perfect correlation is equivalent to a (perfect) linear relationship between the random variables.

To explore this further, suppose we wish to approximate Y using some linear function $\beta X + \alpha$ of X. Such an approximation is called a **best linear predictor** of Y by X. The error in this approximation

$$\epsilon = Y - \beta X - \alpha$$

is called the **residual random variable**. The best fit is generally considered to be the linear predictor that minimizes the **mean squared error**, defined by

$$\text{MSE} = \mathcal{E}(\epsilon^2) = \mathcal{E}[(Y - \beta X - \alpha)^2]$$

When $\rho_{X,Y} = \pm 1$ we have said that the approximation can be made exact and so the MSE $= 0$.

In general, the MSE can be written

$$\text{MSE} = \mathcal{E}(Y^2) - 2\beta\mathcal{E}(XY) - 2\alpha\mathcal{E}(Y) + \beta^2\mathcal{E}(X^2) + 2\alpha\beta\mathcal{E}(X) + \alpha^2$$

The minimum value of this expression (which must exist) is found by setting its partial derivatives to 0. We leave it to the reader to show that the resulting equations are

$$\begin{aligned} \beta\mathcal{E}(X^2) + \alpha\mathcal{E}(X) &= \mathcal{E}(XY) \\ \beta\mathcal{E}(X) + \alpha &= \mathcal{E}(Y) \end{aligned}$$

Solving this system gives

$$\begin{aligned} \beta &= \frac{\sigma_{X,Y}}{\sigma_X^2} \\ \alpha &= \mathcal{E}(Y) - \beta\mathcal{E}(X) \end{aligned}$$

Let us summarize, beginning with a definition.

Definition *Let X and Y be random variables. Write*

$$Y = \beta X + \alpha + \epsilon$$

where β and α are constants and ϵ is the random variable defined by

$$\epsilon = Y - \beta X - \alpha$$

Thus, Y is approximated by the linear function $\beta X + \alpha$ with error random variable ϵ. The **best linear predictor** *of Y with respect to X, denoted by BLP is the linear function $\beta X + \alpha$ that minimizes the* **mean squared error** *$\mathcal{E}(\epsilon^2)$. The coefficient β is called the* **beta** *of Y with respect to X. The line $y = \beta x + \alpha$ is called the* **regression line**.□

Theorem 9 *The best linear predictor of Y with respect to X is*

$$BLP = \frac{\sigma_{X,Y}}{\sigma_X^2} X + \mu_Y - \frac{\sigma_{X,Y}}{\sigma_X^2} \mu_X$$

Moreover, the minimum mean squared error is

$$\mathcal{E}(\epsilon^2) = \sigma_Y^2 (1 - \rho_{X,Y}^2)$$ □

Now we can state the following properties of the correlation coefficient.

- $\rho_{X,Y} = \pm 1$ if and only if there is a linear relationship between X and Y.
- The closer $\rho_{X,Y}$ is to ± 1 the smaller is the mean squared error in using the best linear predictor.
- If $\rho_{X,Y}$ is positive then the BLP has positive slope. Hence, as X increases so does the BLP of Y and as X decreases so does the BLP of Y.
- If $\rho_{X,Y} = -1$ then the slope of the BLP is negative. Hence, as X increases the BLP of Y decreases and vice-versa.

It is worth mentioning that a strong correlation does not imply a *causal relationship*. Just because a random variable Y is observed to take values that are in an approximate linear relationship with the values of another random variable X does not mean that a change in X *causes* a change in Y. It only means that the two random variables are observed to behave similarly. For example, during the early 1990's the sale of personal computers rose significantly. So did the sale of automobiles. Just because there may be a positive correlation between the two does not mean that the purchase of personal computers caused the purchase of automobiles.

The Variance of a Sum

The covariance is just what we need to obtain a formula for the variance of a linear combination of random variables. Theorem 6 implies that if

the random variables X and Y are *independent* then

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$$

However, this does not hold if the random variables fail to be independent. In this case, we do have the following formula.

Theorem 10 *If X and Y are random variables on Ω and $a, b \in \mathbb{R}$ then*

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y^2) + 2ab\mathit{Cov}(X, Y)$$

More generally, if $X_1, \ldots, X_n$ are random variables on Ω and $a_1, \ldots, a_n$ are constants then

$$\text{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mathit{Cov}(X_i, X_j) \qquad \square$$

Exercises

1. A pair of fair dice are rolled. Find the probability of getting a sum that is even.
2. Three fair dice are rolled. Find the probability of getting exactly one 6.
3. A basket contains 5 red balls, 3 black balls, and 4 white balls. A ball is chosen at random from the basket.
   a) Find the probability of choosing a red ball.
   b) Find the probability of choosing a white ball or a red ball.
   c) Find the probability of choosing a ball that is not red.
4. A certain true and false test contains 10 questions. A student guesses randomly at each question.
   a) What is the probability that he will get all 10 questions correct?
   b) What is the probability that he will get at least 9 questions correct?
   c) What is the probability that he will get at least 8 questions correct?
5. A die has six sides, but two sides have only 1 dot. The other four sides have $2, 3, 4$ and 5 dots, respectively. Assume that each side is equally likely to occur.
   a) What is the probability of getting a 1?
   b) What is the probability of getting a 2?
   c) What is the probability of getting an even number?
   d) What is the probability of getting a number less than 3?
6. Four fair coins are tossed. Find the probability of getting exactly 2 heads.

7. Four fair coins are tossed. Find the probability of getting at least 2 heads.
8. A fair die is rolled and a card is chosen at random. What is the probability that the number on the die matches the number on the card? (An ace is counted as a one.)
9. Studies of the weather in a certain city over the last several decades have shown that, for the month of March, the probability of having a certain amount of sun/smog is as follows:

   $\mathcal{P}(\text{full sun/no smog}) = 0.07$ $\mathcal{P}(\text{full sun/light smog}) = 0.09$,
   $\mathcal{P}(\text{full sun/heavy smog}) = 0.12$, $\mathcal{P}(\text{haze/no smog}) = 0.09$,
   $\mathcal{P}(\text{haze/light smog}) = 0.07$, $\mathcal{P}(\text{haze/heavy smog}) = 0.11$
   $\mathcal{P}(\text{no sun/no smog}) = 0.16$, $\mathcal{P}(\text{no sun/light smog}) = 0.12$
   $\mathcal{P}(\text{no sun/heavy smog}) = 0.17$,

   What is the probability of having at a fully sunny day? What is the probability of having at day with some sun? What is the probability of having a day with no or light smog?
10. a) Consider a stock whose current price is 50 and whose price at some fixed time T in the future may be one of the following values: $48, 49, 50, 51$. Suppose we estimate that the probabilities of these stock prices are

$$\begin{aligned}\mathbb{P}(48) &= 0.2\\ \mathbb{P}(49) &= 0.4\\ \mathbb{P}(50) &= 0.3\\ \mathbb{P}(51) &= 0.1\end{aligned}$$

    If we purchase one share of the stock now, what is the expected return at time T? What is the expected profit?

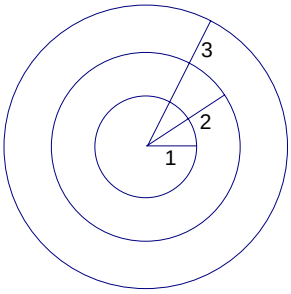
    b) Consider a derivative of the stock in part a) whose return D is a function of the stock price, say

$$\begin{aligned}D(48) &= 2\\ D(49) &= -1\\ D(50) &= 0\\ D(51) &= 3\end{aligned}$$

    Thus, the return D is a random variable on Ω. What is the expected return of the derivative?
11. Suppose that you roll a fair die once. If the number on the top face of the die is even, you win that amount, in dollars. If it is odd, you lose

that amount. What is the expected value of this game? Would you play?

12. For a cost of 1 dollar, you can roll a single fair die. If the outcome is odd, you win 2 dollars. Would you play? Why?
13. Suppose you draw a card from a deck of cards. You win the amount showing on the card if it is not a face card, and lose \$10 if it is a face card. What is your expected value? Would you play this game?
14. An American roulette wheel has 18 red numbers, 18 black numbers and two green numbers. If you bet on red, you win an amount equal to your bet (and get your original bet back) if a red number comes up, but lose your bet otherwise. What is your expected winnings in this game? Is this a fair game?
15. Consider the dart board shown below



A single dart cost \$1.50. You are paid \$3.00 for hitting the center, \$2.00 for hitting the middle ring and \$1.00 for hitting the outer ring. What is the expected value of your winnings? Would you play this game?

16. Prove that $\text{Var}(X) = \mathcal{E}(X)^2 - \mu^2$ where $\mu = \mathcal{E}(X)$.
17. Prove that for any real number a

$$\text{Var}(aX) = a^2\text{Var}(X)$$

18. Prove that if X and Y are independent random variables then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

19. Let X be a binomial random variable with distribution $b(k; n, p)$. Show that $\text{Var}(X) = np(1 - p)$. *Hint*: use the fact that $\text{Var}(X) = \mathcal{E}(X^2) - \mathcal{E}(X)^2$.
20. Prove the theorem on total probabilities.
21. Show that if X, Y and Z are independent random variables then so are XY and Z.

22. Let X and Y be independent random variables on $(\Omega, \mathbb{P})$. Let f and g be functions from $\mathbb{R}$ to $\mathbb{R}$. Then prove that $f(X)$ and $g(Y)$ are independent.
23. Show that $\rho_{X,Y} = +1$ implies that the slope $a > 0$ and $\rho_{X,Y} = -1$ implies that $a < 0$, where $Y = aX + b$.
24. Show that for any random variables $X$and Y

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y^2) + 2ab\text{Cov}(X, Y)$$

Chapter 2

Portfolio Management and the Capital Asset Pricing Model

In this chapter, we explore the issue of *risk management* in a portfolio of assets. The main issue is how to *balance* a portfolio, that is, how to choose the percentage (by value) of each asset in the portfolio so as to minimize the overall risk for a given expected return. The first lesson that we will learn is that the risks of each asset in a portfolio alone do not present enough information to understand the overall risk of the entire portfolio. It is necessary that we also consider how the assets interact, as measured by the *covariance* (or equivalently the *correlation*) of the individual risks.

Portfolios, Returns and Risk

For our model, we will assume that there are only two time periods: the initial time $t = 0$ and the final time $t = T$. Each asset $\mathfrak{a}_i$ has an **initial value** $\mathcal{V}_{i,0}$ and a **final value** $\mathcal{V}_{i,T}$.

Portfolios

A portfolio consists of a collection of assets $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ in a given proportion. Formally, we define a **portfolio** to be an ordered n-tuple of real numbers

$$\Theta = (\theta_1, \ldots, \theta_n)$$

where θ_i is the number of units of asset $\mathfrak{a}_i$. If θ_i is negative then the portfolio has a short position on that asset: a short sale of stock, a short put or call and so on. A positive value of θ_i indicates a long position: an owner of a stock, long on a put or call and so on.

Asset Weights

It is customary to measure the amount of an asset within a portfolio by its percentage by *value*. The **weight** w_i of asset $\mathfrak{a}_i$ is the percentage of the value of the asset contained in the portfolio at time $t = 0$, that is,

$$w_i = \frac{\theta_i \mathcal{V}_{i,0}}{\sum_{j=1}^{n} \theta_j \mathcal{V}_{i,0}}$$

Note that the *sum* of the weights will always be 1

$$w_1 + \cdots + w_n = 1$$

Asset Returns

The **return** R_i on asset $\mathfrak{a}_i$ is defined by the equation

$$\mathcal{V}_{i,T} = \mathcal{V}_{i,0}(1 + R_i)$$

which is equivalent to

$$R_i = \frac{\mathcal{V}_{i,T} - \mathcal{V}_{i,0}}{\mathcal{V}_{i,0}}$$

Since the value of an asset at time T in the future is a random variable, so is the return R_i. Thus, we may consider the expected value and the variance of the return. The e**xpected return** of asset $\mathfrak{a}_i$ is denoted by

$$\mu_i = \mathcal{E}(R_i)$$

The variance of the return of asset $\mathfrak{a}_i$

$$\sigma_i^2 = \text{Var}(R_i)$$

is called the **risk** of asset $\mathfrak{a}_i$. We will also consider the standard deviation as a measure of risk when appropriate.

Portfolio Return

The **return** on the portfolio itself is defined to be the weighted sum of the returns of each asset

$$R = \sum_{i=1}^{n} w_i R_i$$

For instance, suppose that a portfolio has only 2 assets, with weights 0.4 and 0.6 and returns equal to 10% and 8%, respectively. Then the return on the portfolio is

$$(0.4)(0.10) + (0.6)(0.08) = 0.088 = 8.8\%$$

Since the expected value operator is linear, the expected return of the portfolio as a whole is

$$\mu = \sum_{i=1}^{n} w_i \mu_i$$

Since the individual returns generally are *not* independent, the variance of the portfolio's return is given by the formula

$$\sigma^2 = \text{Var}\left(\sum_{i=1}^{n} w_i R_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \text{Cov}(R_i, R_j)$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \rho_{i,j} \sigma_i \sigma_j$$

where $\text{Cov}(R_i, R_j)$ is the covariance of R_i and R_j and $\rho_{i,j}$ is the correlation coefficient. Let us make some formal definitions.

Definition *The* **expected return** μ *on a portfolio is the expected value of the portfolio's return, that is*

$$\mu = \mathcal{E}\left(\sum_{i=1}^{n} w_i R_i\right) = \sum_{i=1}^{n} w_i \mu_i$$

The **risk** *of a portfolio is the variance of the portfolio's return, that is*

$$\sigma^2 = \text{Var}\left(\sum_{i=1}^{n} w_i R_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j Cov(R_i, R_j)$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \rho_{i,j} \sigma_i \sigma_j$$

An asset is **risky** if its risk σ_i^2 is positive and **risk-free** if its risk is 0.□

> Until further notice, we will assume that the all assets in a portfolio are risky, that is, $\sigma_i^2 > 0$.

More On Risk

Let us take a closer look at the notion of risk. Generally speaking, there are two forms of risk associated with an asset. The **systematic risk** of an asset is the risk that is associated with macroeconomic forces in the market as a whole and not just with any particular asset. For example, a change in interest rates affects the market as a whole. A change in the nation's money supply is another example of a contributor to systematic risk. Global acts such as those of war or terrorism would be considered part of systematic risk.

On the other hand, **unsystematic risk** or **unique risk** is the risk that is particular to an asset or group of assets. For instance, suppose that an investor decides to invest in a company that makes pogs. There are many unsystematic risks here. For example, customers may lose interest in pogs, or the pog company's factory may burn down.

A Primer on How Risks Interact

The key difference between these two types of risk is that unsystematic risk can be diversified away, whereas systematic risk cannot. For instance, an investor can reduce or eliminate the risk that the pog company's factory will burn down by investing in all pog-making companies. In this way, if one pog factory burns down, another pog company will take up the slack. More generally, an investor can reduce the risk associated with an apathy for pogs by investing in all toy and game companies. After all, when was the last time you heard a child say that he was tired of buying pogs and has decided to put his allowance in the bank instead?

To see the effect of individual assets upon risk, consider a portfolio with a single asset $\mathfrak{a}_1$, with expected return μ_1 and risk σ_1^2. The overall risk of the portfolio is also σ_1^2. Let us now add an additional asset $\mathfrak{a}_2$ to the portfolio. Assume that the asset has expected return μ_2 and risk σ_2^2.

If the weight of asset $\mathfrak{a}_1$ is t then the weight of asset $\mathfrak{a}_2$ is $1 - t$. Hence, the expected return of the portfolio is

$$\mu = t\mu_1 + (1 - t)\mu_2$$

and the risk is

$$\sigma^2 = t^2\sigma_1^2 + (1 - t)^2\sigma_2^2 + 2t(1 - t)\rho_{1,2}\sigma_1\sigma_2$$

How does this risk compare to the risks of the individual assets in the portfolio? We may assume (by reversing the numbering if necessary) that $0 < \sigma_1 \leq \sigma_2$.

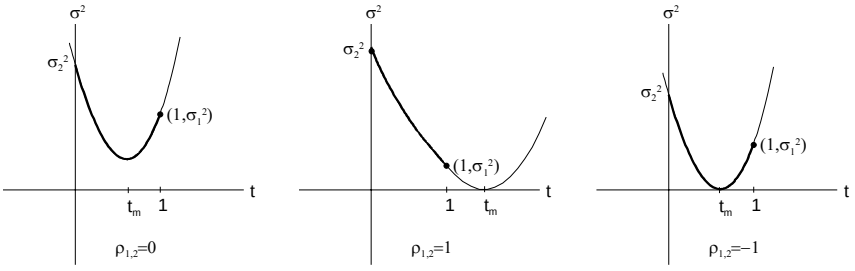


Figure 1–Some risk possibilities. Bold curves indicate no short selling.

Suppose first that the assets are uncorrelated, that is, $\rho_{1,2} = 0$. The portfolio risk is equal to

$$\sigma^2 = t^2\sigma_1^2 + (1-t)^2\sigma_2^2 = (\sigma_1^2 + \sigma_2^2)t^2 - 2\sigma_2^2 t + \sigma_2^2$$

This quadratic in t is shown on the left in Figure 1. A bit of differentiation shows that the minimum risk occurs at

$$t_m = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

and is equal to

$$\sigma_m^2 = \frac{2\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

Note that since

$$\sigma_1^2 \le \sigma_m^2 \le \sigma_2^2$$

the minimum risk lies somewhere between the risks of the individual assets.

Now suppose that the assets are perfectly positively correlated, that is, $\rho_{1,2} = 1$. Then the risk is

$$\sigma^2 = t^2\sigma_1^2 + (1-t)^2\sigma_2^2 + 2t(1-t)\sigma_1\sigma_2 = (t\sigma_1 + (1-t)\sigma_2)^2$$

This quadratic is shown in the middle of Figure 1. The minimum risk is actually 0 and occurs at

$$t_m = \frac{\sigma_2}{\sigma_2 - \sigma_1} > 0$$

Note that

$$1 - t_m = \frac{-\sigma_1}{\sigma_2 - \sigma_1} < 0$$

and so the minimum-risk portfolio must take a short position in the asset with larger risk.

Finally, suppose that the assets are perfectly negatively correlated, that is, $\rho_{1,2} = -1$. Then the risk is

$$\sigma^2 = t^2\sigma_1^2 + (1-t)^2\sigma_2^2 - 2t(1-t)\sigma_1\sigma_2 = (t\sigma_1 - (1-t)\sigma_2)^2$$

This quadratic is shown on the right side of Figure 1. The minimum risk is again 0 and occurs at

$$t_m = \frac{\sigma_1}{\sigma_1 + \sigma_2} > 0$$

In this case

$$1 - t_m = \frac{\sigma_2}{\sigma_1 + \sigma_2} > 0$$

and so the minimum-risk portfolio does not require short selling.

Thus, the case where the assets are perfectly negatively correlated seems to be the most promising, in that the risk can be reduced to 0 without short selling, which certainly has its drawbacks. Indeed, short selling may not even be possible in many cases and when it is, there can be additional costs involved. Of course, it is in general a difficult (or impossible) task to select assets that are perfectly negatively correlated with the other assets in a portfolio.

Two-Asset Portfolios

Let us now begin our portfolio analysis in earnest, starting with portfolios that contain only two assets $\mathfrak{a}_1$ and $\mathfrak{a}_2$, with weights w_1 and w_2, respectively. It is customary to draw risk-expected return curves with the risk on the horizontal axis and the expected return on the vertical axis. It is also customary to use the standard deviation as a measure of risk for graphing purposes.

The expected return of such a portfolio is given by

$$\mu = w_1\mu_1 + w_2\mu_2$$

and the risk is

$$\sigma^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{1,2}\sigma_1\sigma_2$$

As before, we assume that the assets are risky, that is, $\sigma_i > 0$.

The Case $\rho_{1,2} = \pm 1$

Let us first consider the case $\rho_{1,2} = \pm 1$. In these cases, the expression for σ^2 simplifies considerably and we have

$$\sigma = |w_1\sigma_1 \pm w_2\sigma_2|$$

(The choice of sign matches the sign of $\rho_{1,2}$.) Since $w_1 + w_2 = 1$ for convenience we can set $w_2 = s$ and $w_1 = 1 - s$ to get the *parametric equations*

$$\mu = (1-s)\mu_1 + s\mu_2$$
$$\sigma = |(1-s)\sigma_1 \pm s\sigma_2|$$

where s ranges over all real numbers. For s in the range $[0, 1]$ both weights are nonnegative and so the portfolio has no short positions. Outside this range, exactly one of the weights is negative, indicating that the corresponding asset is held short (the other asset is held long).

To help plot the points (σ, μ) in the plane, let us temporarily ignore the absolute value sign and consider the parametric equations

$$\mu = (1-s)\mu_1 + s\mu_2$$
$$\sigma' = (1-s)\sigma_1 \pm s\sigma_2$$

These are the equations of a straight line in the (σ', μ)-plane. When $\rho_{1,2} = 1$ the plus sign is taken and the line passes through the points (σ_1, μ_1) and (σ_2, μ_2). For $\rho_{1,2} = -1$ the line passes through the points (σ_1, μ_1) and $(-\sigma_2, \mu_2)$. These lines are plotted in Figure 2.

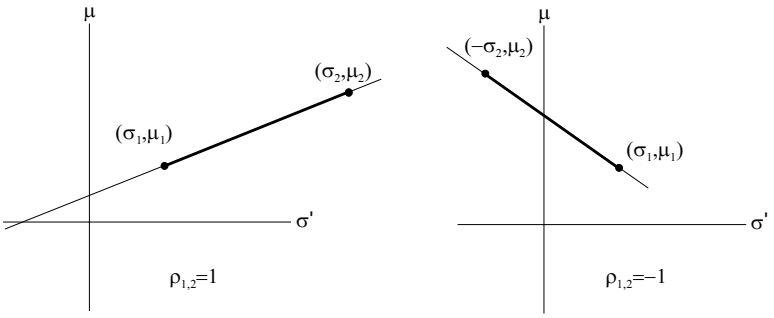


Figure 2–The graph before taking absolute values

Now, the effect of the absolute value sign is simply to flip that part of the line that lies in the left half-plane over the μ-axis (since $\sigma = |\sigma'|$). The resulting plots are shown in Figure 3. The bold portions correspond to points where both weights are nonnegative, that is, no short selling is required.

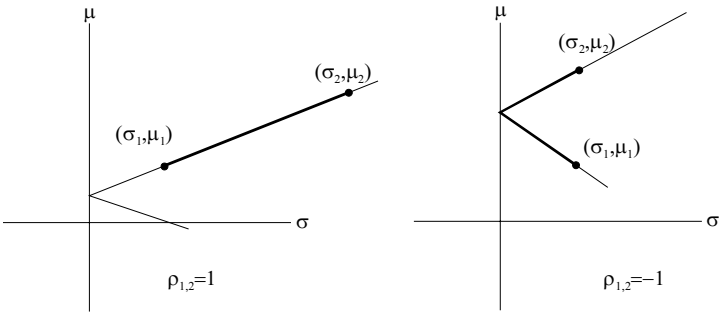


Figure 3–The risk-return lines

From the parametric equations (or from our previous discussion), we can deduce the following theorem, which shows again that there are cases where we can reduce the risk of the portfolio to 0.

Theorem 1 *For $\rho_{1,2} = \pm 1$ the risk and expected return of the portfolio are given by the parametric equations*

$$\mu = (1-s)\mu_1 + s\mu_2$$
$$\sigma = |(1-s)\sigma_1 \pm s\sigma_2|$$

where s is the weight of asset $\mathfrak{a}_2$ and ranges over all real numbers. For $s \in [0,1]$ both weights are nonnegative and the portfolio has no short positions. Outside this range, exactly one of the weights is negative, for which the corresponding asset is held short. The plots of (σ, μ) are shown in Figure 3.

Moreover, when $\rho_{1,2} = 1$ the risk σ is 0 if and only if $\sigma_1 \neq \sigma_2$ and

$$w_1 = \frac{-\sigma_2}{\sigma_1 - \sigma_2}, w_2 = \frac{\sigma_1}{\sigma_1 - \sigma_2}$$

and so short selling of asset $\mathfrak{a}_1$ is required. In this case, the expected

return is

$$\mu = \frac{\sigma_1\mu_2 - \sigma_2\mu_1}{\sigma_1 - \sigma_2}$$

When $\rho_{1,2} = -1$ *the risk* σ *is* 0 *if and only if*

$$w_1 = \frac{\sigma_2}{\sigma_1 + \sigma_2}, w_2 = \frac{\sigma_1}{\sigma_1 + \sigma_2}$$

and so no short selling is required. In this case, the expected return is

$$\mu = \frac{\sigma_1\mu_2 + \sigma_2\mu_1}{\sigma_1 + \sigma_2}$$ □

The Case $-1 < \rho_{1,2} < 1$

When $-1 < \rho_{1,2} < 1$ the parametric equations for the risk and expected return are

$$\begin{aligned} \mu &= w_1\mu_1 + w_2\mu_2 \\ \sigma^2 &= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{1,2}\sigma_1\sigma_2 \end{aligned}$$

Parametrizing as above by letting $w_1 = 1 - s, w_2 = s$ gives

$$\begin{aligned} \mu &= (\mu_2 - \mu_1)s + \mu_1 \\ \sigma^2 &= (\sigma_1^2 + \sigma_2^2 - 2\rho_{1,2}\sigma_1\sigma_2)s^2 - 2\sigma_1(\sigma_1 - \rho_{1,2}\sigma_2)s + \sigma_1^2 \end{aligned}$$

We next observe that since $\rho_{1,2} < 1$ the coefficient of s^2 in σ^2 satisfies

$$\sigma_1^2 + \sigma_2^2 - 2\rho_{1,2}\sigma_1\sigma_2 = (\sigma_1 - \sigma_2)^2 + 2\sigma_1\sigma_2(1 - \rho_{1,2}) > 0$$

and so the expression for σ^2 is truly quadratic (not linear). The graph of the points (σ^2, μ) is a parabola lying on its side, opening to the right and going through the points (σ_1^2, μ_1) and (σ_2^2, μ_2). Figure 4 shows the graph as well as two possible placements of the points (σ_1, μ_1) and (σ_2, μ_2). In the graph on the right, the minimum risk requires a short position.

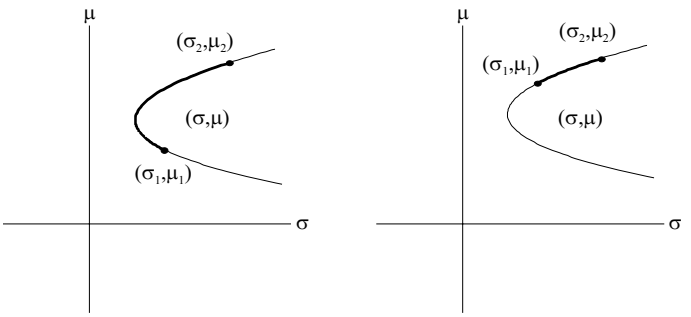


Figure 4–The risk-return graph

Let us assume again that $0 < \sigma_1 \leq \sigma_2$. Differentiating the risk σ^2 with respect to s gives

$$\frac{d}{ds}(\sigma^2) = 2(\sigma_1^2 + \sigma_2^2 - 2\rho_{1,2}\sigma_1\sigma_2)s - 2\sigma_1(\sigma_1 - \rho_{1,2}\sigma_2)$$

so the minimum-risk point occurs at

$$s_{\text{min}} = \frac{\sigma_1(\sigma_1 - \rho_{1,2}\sigma_2)}{\sigma_1^2 + \sigma_2^2 - 2\rho_{1,2}\sigma_1\sigma_2}$$

The minimum risk portfolio will have no short positions if and only if $s_{\text{min}} \in [0, 1]$.

Since $\sigma_1 \leq \sigma_2$ and $\rho_{1,2} < 1$ a little algebra shows that $s_{\text{min}} < 1$. The various possibilities are described as follows (recall that $s = w_2$ is the weight of asset $\mathfrak{a}_2$)

$s_{\text{min}} < 0$
Short selling of asset $\mathfrak{a}_2$ is required to minimize risk. This risk is less than either of σ_1^2 or σ_2^2.

$s_{\text{min}} = 0$
The minimum portfolio risk is σ_1^2, which is achieved by holding only asset $\mathfrak{a}_1$ (that is, $w_1 = 1$).

$0 < s_{\text{min}} < 1$
The minimum portfolio risk is achieved with no short selling and is less than either of σ_1^2 or σ_2^2.

$s_{\text{min}} \geq 1$
Cannot happen at minimum risk under our assumption that $0 < \sigma_1 \leq \sigma_2$.

Now, it is easy to see from the expression for $s_{\min}$ that

$$s_{\min} < 0 \Leftrightarrow \rho_{1,2} > \frac{\sigma_1}{\sigma_2}$$
$$s_{\min} = 0 \Leftrightarrow \rho_{1,2} = \frac{\sigma_1}{\sigma_2}$$
$$0 < s_{\min} < 1 \Leftrightarrow \rho_{1,2} < \frac{\sigma_1}{\sigma_2}$$

and so we have the following result, which describes when we can achieve a minimum risk with no short selling and when we can achieve a 0 overall risk.

Theorem 2 *Assume that* $0 < \sigma_1 \leq \sigma_2$.

1) *If* $-1 \leq \rho_{1,2} < \frac{\sigma_1}{\sigma_2}$ *then the minimum risk can be achieved with no short selling and is less than either of* σ_1^2 *or* σ_2^2. *For* $\rho_{1,2} = -1$ *the minimum risk is* 0.
2) *If* $\rho_{1,2} = \frac{\sigma_1}{\sigma_2}$ *then the minimum risk is* σ_1^2, *which is achieved by holding only asset* $\mathfrak{a}_1$.
3) *If* $1 \geq \rho_{1,2} > \frac{\sigma_1}{\sigma_2}$ *then short selling of asset* $\mathfrak{a}_2$ *is required to minimize risk, which is less than either of* σ_1^2 *or* σ_2^2. *For* $\rho_{1,2} = 1$ *the minimum risk is* 0.□

Multi-Asset Portfolios

Now let us turn our attention to portfolios with an arbitrary number $n \geq 2$ of assets. The weights of the portfolio can be written in matrix (or vector) form as

$$W = (\, w_1 \quad w_2 \quad \cdots \quad w_n \,)$$

It is also convenient to define the matrix (or vector) of 1's by

$$O = (\, 1 \quad 1 \quad \cdots \quad 1 \,)$$

so that the condition

$$w_1 + \cdots + w_n = 1$$

can be written as a matrix product

$$OW^t = 1$$

where W^t is the transpose of W. We will also denote the matrix of expected returns by

$$M = (\,\mu_1 \quad \mu_2 \quad \cdots \quad \mu_n\,)$$

and the *covariance matrix* by

$$C = (c_{i,j})$$

where

$$c_{i,j} = \text{Cov}(R_i, R_j)$$

Note that $c_{i,i} = \sigma_i^2$ is the variance of R_i. It can be shown, although we will not do it here, that the matrix C is **symmetric** (that is $C^t = C$) and **positive semidefinite**, which means that for any matrix $A = (a_1, \ldots, a_n)$ we have $ACA^t \geq 0$. We shall also assume that C is invertable, which in this case implies that C is **positive definite**, that is, for any matrix $A = (a_1, \ldots, a_n)$ we have $ACA^t > 0$.

The expected return can now be written as as matrix product

$$\mu = MW^t = \mu_1 w_1 + \cdots + \mu_n w_n$$

and the risk can be written as

$$\sigma^2 = \text{Var}(w_1 R_1 + \cdots + w_n R_n) = \sum_{i,j=1}^{n} c_{i,j} w_i w_j = WCW^t$$

The Markowitz Bullet

Let us examine the relationship between the weights $W = (w_1, \ldots, w_n)$ of a portfolio and the corresponding risk-expected return point (σ, μ) for that portfolio, given by the equations above. Note that we are now referring to risk in the form of the standard deviation σ.

Figure 5 describes the situation is some detail for a portfolio with three assets and this will provide some geometric intuition for the multi-asset case in general. (We will define the terms *Markowitz bullet* and *Markowitz efficient frontier* a bit later.)

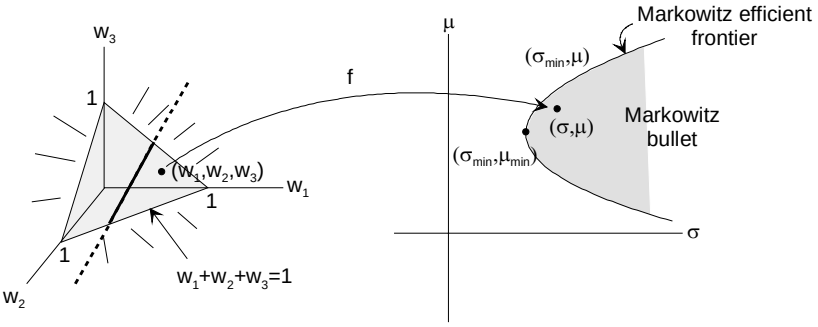


Figure 5–The Markowitz bullet

The left-hand portion of Figure 5 shows the n-dimensional space in which the weight vectors $(w_1, \ldots, w_n)$ reside. (In Figure 5 we have $n = 3$ of course.) Since the sum of the weights must equal 1, the weight vectors must lie on the *hyperplane* whose equation is

$$w_1 + \cdots + w_n = 1$$

For $n = 3$ this is an ordinary plane in 3-dimensional space, passing through the points $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$. For the sake of clarity, the figure shows only that portion of this hyperplane that lies in the positive orthant. This is the portion of the plane that corresponds to portfolios with no short selling. Let us refer to the entire hyperplane as the **weight hyperplane**.

We denote by f the function that takes each weight vector in the weight hyperplane to the risk-expected return ordered pair for the corresponding portfolio, that is,

$$f(w_1, \ldots, w_n) = (\sigma, \mu)$$

where

$$\mu = \mu_1 w_1 + \cdots + \mu_n w_n = MW^t$$
$$\sigma^2 = \sum_{i,j=1}^{n} c_{i,j} w_i w_j = WCW^t$$

The function f is also pictured in Figure 5. Our goal is to determine the image of a straight line in the weight hyperplane under the function f. This will help us get an idea of how the function f behaves in general. (It is analogous to making a contour map of a function.)

The equation of a line in n-dimensional space (whether in the weight hyperplane or not) can be written in the parametric form

$$\ell(t) = (a_1 t + b_1, \ldots, a_n t + b_n)$$

where the parameter t varies from $-\infty$ to ∞. The value $t = 0$ corresponds to the point $\ell(0) = (a_1, \ldots, a_n)$ and $t = 1$ corresponds to $\ell(1) = (b_1, \ldots, b_n)$. It is also true that any equation of this form is the equation of a line.

Now, for any point $(w_1, \ldots, w_n)$ on the line, corresponding to a particular value of t, the expected return is

$$\begin{aligned}\mu &= \mu_1 w_1 + \cdots + \mu_n w_n \\ &= \mu_1(a_1 t + b_1) + \cdots + \mu_1(a_n t + b_n) \\ &= (\mu_1 a_1 + \cdots + \mu_n a_n)t + (\mu_1 b_1 + \cdots + \mu_n b_n)\end{aligned}$$

which is a *linear function* of t. This is a critical point. Solving for t gives

$$t = \frac{\mu - (\mu_1 b_1 + \cdots + \mu_n b_n)}{\mu_1 a_1 + \cdots + \mu_n a_n} = A\mu + B$$

where we use the symbols A and B for convenience, and must assume that the denominator above is not 0.

Now let us look at the risk (in the form of the variance)

$$\begin{aligned}\sigma^2 &= \sum_{i,j=1}^{n} c_{i,j} w_i w_j \\ &= \sum_{i,j=1}^{n} c_{i,j}(a_i t + b_i)(a_j t + b_j) \\ &= \sum_{i,j=1}^{n} c_{i,j}\big(a_i a_j t^2 + (a_i b_j + a_j b_i)t + b_i b_j\big) \\ &= \sum_{i,j=1}^{n} c_{i,j} a_i a_j t^2 + \sum_{i,j=1}^{n} c_{i,j}(a_i b_j + a_j b_i)t + \sum_{i,j=1}^{n} c_{i,j} b_i b_j \\ &= \alpha t^2 + \beta t + \gamma\end{aligned}$$

where we have used the letters α, β and γ to simplify the expression, which is just a quadratic in t. Replacing t by its expression in terms of μ gives

$$\sigma^2 = \alpha(A\mu + B)^2 + \beta(A\mu + B) + \gamma$$

which is a quadratic in μ (assuming that $\alpha \neq 0$).

Thus, as t traces out a line in the weight hyperplane, the risk-expected return points (σ^2, μ) trace out a parabola (lying on its side) in the (σ, μ)-plane. Taking the square root of the first coordinate produces a curve which we will refer to as a **Markowitz curve**, although this term is not standard. Thus, straight lines in the weight hyperplane are mapped to Markowitz curves in the (σ, μ) plane under the function f. Note that Markowitz curves are *not* parabolas.

Figure 6 shows an example of a Markowitz curve generated using Microsoft Excel. For future reference, we note now that the data used to plot this curve are

$(\mu_1, \mu_2, \mu_3) = (0.1, 0.11, 0.07)$
$(\sigma_1, \sigma_2, \sigma_3) = (0.23, 0.26, 0.21)$
$\rho_{1,2} = \rho_{2,1} = -0.15$
$\rho_{1,3} = \rho_{3,1} = 0.25$
$\rho_{2,3} = \rho_{2,3} = 0.2$

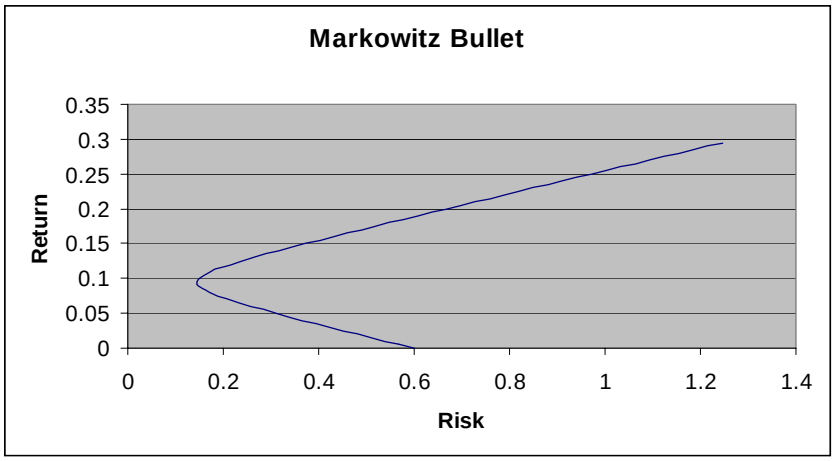


Figure 6–A Markowitz bullet

The Shape of a Markowitz Curve

It is important to make a clear distinction between the parabola traced out by (σ^2, μ) and the Markowitz curve traced out by the points (σ, μ), as

pictured in Figure 6. To get a feel for the differences in more familiar territory, consider the functions

$$y = ax^2 + bx + c$$

and

$$z = \sqrt{ax^2 + bx + c}$$

for $a > 0$. The first graph is a parabola. The slope of the tangent lines to this parabola are given by the derivative

$$y' = 2ax + b$$

and these slopes increase without bound as x tends to ∞. On the other hand, for the function z, for large values of x the first term dominates the others and so

$$z = \sqrt{ax^2 + bx + c} \approx \sqrt{ax^2} = \sqrt{a}|x|$$

The graph of the equation $z = \sqrt{a}|x|$ is a pair of *straight lines*. This shows that as x tends to ∞ the graph of z flattens, unlike the case of a parabola. In particular, the derivative is

$$z' = \frac{2ax + b}{2\sqrt{ax^2 + bx + c}}$$

Squaring this makes it easier to take the limit

$$\lim_{x\to\infty} (z')^2 = \lim_{x\to\infty} \frac{4a^2x^2 + 4abx + b^2}{4(ax^2 + bx + c)} = a$$

so we see that z' approaches $\sqrt{a}$ as x approaches ∞.

Thus, unlike parabolas *Markowitz curves flatten out* as we move to the right. One of the implications of this fact, which is important to the capital asset pricing model, is that (looking ahead to Figure 9) if μ_{rf} is too large, there is no tangent line from the point $(0, \mu_{\text{rf}})$ to the upper portion of the Markowitz curve.

The Point of Minimum Risk

Let us denote the point of minimum risk by $(\sigma_{\text{min}}, \mu_{\text{min}})$. We will be content with finding the portfolio weights (in the weight hyperplane) that correspond to this point. For any particular case, these weights can easily

be plugged into the formulas for σ and μ to get the actual point. (As the reader will see, the general formulas can get a bit messy.)

The next theorem gives the minimum risk weights. The proof uses the technique of *Lagrange multipliers*, which can be found in any standard multi-variable calculus book, so we will not go into the details here. The reader may skim over the few proofs that require this technique if desired.

Theorem 3 *A portfolio with minimum risk has weights given by*

$$W = \frac{OC^{-1}}{OC^{-1}O^t}$$

Note that the denominator is a number and is just the sum of the components in the numerator.
Proof. We seek to minimize the expression

$$\sigma^2 = \sum_{i,j=1}^{n} c_{i,j} w_i w_j = WCW^t$$

subject to the constraint

$$OW^t = w_1 + \cdots + w_n = 1$$

According to the technique of Lagrange multipliers, we must take the partial derivatives with respect to each w_i and α of the function

$$g(w_1, \ldots, w_n) = \sum_{i,j=1}^{n} c_{i,j} w_i w_j + \alpha(1 - w_1 - \cdots - w_n)$$

and set them equal to 0. We leave it as an exercise to show that this results in the equation

$$CW^t = \frac{\alpha}{2} O^t$$

and so

$$W = \frac{\alpha}{2} OC^{-1}$$

Substituting this into the constraint (and using the fact that C and C^{-1} are symmetric) gives

$$\frac{\alpha}{2} = \frac{1}{OC^{-1}O^t}$$

and so we get the desired result.□

The Markowitz Efficient Frontier

The set of points $(\sigma_{\min}, \mu)$ that gives the minimum risk *for each expected return* μ is called the **Markowitz efficient frontier** (frontier is another word for boundary). The next theorem describes this set of points. While the formula is a bit messy, there is an important lesson here. Namely, the minimum risk weights are a *linear function* of the expected return. This means that as the expected return μ takes on all values from $-\infty$ to ∞, the minimum risk weights trace out a straight line in the weight hyperplane and the corresponding points $(\sigma_{\min}, \mu)$ trace out a Markowitz curve!

In other words, the Markowitz efficient frontier is a Markowitz curve. The weight line that corresponds to the Markowitz curve is called the **minimum risk weight line**.

Theorem 4 *For a given expected return* μ*, the portfolio with minimum risk has weights given by*

$$W = \frac{\begin{vmatrix} \mu & MC^{-1}O^t \\ 1 & OC^{-1}O^t \end{vmatrix} MC^{-1} + \begin{vmatrix} MC^{-1}M^t & \mu \\ OC^{-1}M^t & 1 \end{vmatrix} OC^{-1}}{\begin{vmatrix} MC^{-1}M^t & MC^{-1}O^t \\ OC^{-1}M^t & OC^{-1}O^t \end{vmatrix}}$$

In particular, each weight w_i *is a linear function of* μ*.*
Proof. In this case, we seek to minimize the expression

$$\sigma^2 = \sum_{i,j=1}^{n} c_{i,j} w_i w_j = WCW^t$$

subject to the constraints

$$MW^t = w_1\mu_1 + \cdots + w_n\mu_n = \mu$$

and

$$OW^t = w_1 + \cdots + w_n = 1$$

This is done by setting the partial derivatives of the following function to 0

$$g = \sum_{i,j=1}^{n} c_{i,j} w_i w_j + \alpha(\mu - w_1\mu_1 - \cdots - w_n\mu_n) + \beta(1 - w_1 - \cdots - w_n)$$

This results in the matrix equation

$$2CW^t = \alpha M^t + \beta O^t$$

and so

$$W = \frac{1}{2}(\alpha M + \beta O)C^{-1}$$

Substituting the expression for W^t into the matrix form of the constraints gives the system of equations

$$(MC^{-1}M^t)\alpha + (MC^{-1}O^t)\beta = 2\mu$$
$$(OC^{-1}M^t)\alpha + (OC^{-1}O^t)\beta = 2$$

Cramer's rule can now be used to obtain a formula for α and β. Substituting this into the expression for W gives the desired result. We leave all details to the reader as an exercise.□

An ordered pair (x, y) is said to be an **attainable point**, if it is of the form (σ, μ) for some portfolio. Since the Markowitz efficient frontier contains the points of minimum risk, all attainable points must lie to the right (corresponding to greater risk) of some point on this frontier. In other words, the attainable points are contained in the shaded region on the right-hand side of Figure 5. This region (including the frontier) is known as the **Markowitz bullet**, due to its shape.

To explain the significance of the Markowitz efficient frontier, we make the following definition.

Definition *Let* $P_1 = (\sigma_1, \mu_1)$ *and* $P_2 = (\sigma_2, \mu_2)$ *be attainable points. Then* (σ_1, μ_1) ***dominates*** (σ_2, μ_2) *if*

$$\sigma_1 \leq \sigma_2 \text{ and } \mu_1 \geq \mu_2$$

in words, P_1 *has smaller or equal risk and larger or equal expected return.*□

Theorem 5 *Any attainable point is dominated by an attainable point on the Markowitz efficient frontier. Thus, investors who seek to minimize*

risk for any expected return need only look on the Markowitz efficient frontier.□

EXAMPLE 1 Let us sketch the computations needed in order to get the Markowitz bullet in Figure 6. The data are as follows

$$(\mu_1, \mu_2, \mu_3) = (0.1, 0.11, 0.07)$$
$$(\sigma_1, \sigma_2, \sigma_3) = (0.23, 0.26, 0.21)$$
$$\rho_{1,2} = \rho_{2,1} = -0.15$$
$$\rho_{1,3} = \rho_{3,1} = 0.25$$
$$\rho_{2,3} = \rho_{2,3} = 0.2$$

Since the computations are a bit tedious, they are best done with some sort of software program, such as Microsoft Excel. Figure 7 shows a portion of an Excel spreadsheet that has the required computations. The user need only fill in the gray cells and the rest will adjust automatically.

Capital Asset Pricing Model-Fill In Grey Cells

User Data		**Returns μ_i**	**Risks σ_i**	**Correlation**	
	i=1	0.1	0.23	-0.15	$=\rho_{1,2}=\rho_{2,1}$
	i=2	0.11	0.26	0.25	$=\rho_{1,3}=\rho_{3,1}$
	i=3	0.07	0.21	0.2	$=\rho_{2,3}=\rho_{3,2}$
$C=(c_{i,j})=(\rho_{i,j}\sigma_i\sigma_j)$		j=1	j=2	j=3	
	i=1	0.0529	-0.00897	0.012075	
	i=2	-0.00897	0.0676	0.01092	
	i=3	0.012075	0.01092	0.0441	
Inverse of C		21.10168374	3.888932099	-6.740815639	
		3.888932099	16.12598046	-5.05792657	
		-6.740815639	-5.05792657	25.77387544	
Min Risk Point		**OC^{-1}=**	18.2498002	14.95698599	13.97513323
		$OC^{-1}O^t$=	47.18191942		
		W=	0.386796477	0.31700673	0.296196793
		μ=	0.094284164		
		WC=	0.02119456	0.02119456	0.02119456
		WCW^t=	0.02119456		
		σ=	0.145583514		
Min Risk Line		**MC^{-1}=**	2.06609381	1.808696201	0.573717794
		$MC^{-1}O^t$=	4.448507805		
		$OC^{-1}M^t$=	4.448507805		
		$MC^{-1}M^t$=	0.445726209		
		Denom Det=	1.24099637		

Figure 7-Excel worksheet

Referring to Figure 7, the point of minimum risk is given in Theorem 3 by

$$W = \frac{OC^{-1}}{OC^{-1}O^t}$$

These weights can be used to get the expected return and risk

$$\mu = MW^t$$

and

$$\sigma^2 = WCW^t$$

Thus, the minimum risk point is

$$(\sigma_{\text{min}}, \mu_{\text{min}}) = (0.146, 0.094) = (14.6\%, 9.4\%)$$

Next, we compute the minimum risk for a given expected return μ. The formula for the minimum risk is given in Theorem 4. All matrix products are computed in Figure 7, and so is the denominator, which does not depend on μ. Figure 8 shows the computation of the minimum risk for 3 different expected returns. □

Return μ	Num Det 1	Num 1st Term		
0	-4.448507805	-9.19103444	-8.045999165	-2.55218809
0.005	-4.212598208	-8.703623082	-7.619310373	-2.41684255
0.01	-3.976688611	-8.216211723	-7.192621581	-2.28149702

Num Det 2	Num 2nd Term		
0.445726209	8.134414252	6.666720658	6.229083152
0.42348367	7.728492359	6.334039314	5.918240706
0.401241131	7.322570466	6.001357969	5.607398259

Num of W			W		
-1.056620188	-1.379278507	3.676895066	-0.851428911	-1.111428317	2.962857228
-0.975130723	-1.28527106	3.501398153	-0.785764363	-1.035676727	2.821441091
-0.893641258	-1.191263612	3.32590124	-0.720099816	-0.959925138	2.680024954

Return μ	WC			σ_2=WCW^t	Risk σ
1.11022E-16	0.000705424	-0.035140836	0.108244202	0.35916802	0.599306
0.005	0.001791987	-0.032153304	0.103627858	0.324272245	0.569449
0.01	0.00287855	-0.029165771	0.099011513	0.291277439	0.539701

Figure 8–Computing minimum risk for a given expected return

The Capital Asset Pricing Model

Now that we have discussed the so-called *Markowitz portfolio theory*, we are ready to take a look at the *Capital Asset Pricing Model*, or CAPM (pronounced “Cap M”). The major factor that turns Markowitz portfolio theory into capital market theory is the inclusion of a risk-free asset in the model. (Recall that up to now we have been assuming that all assets are risky.)

As we have said, a **risk-free asset** is one that has 0 risk, that is, variance 0. Thus, its risk-expected return point lies on the vertical axis, as shown in Figure 9.

The inclusion of a risk-free asset into the Markowitz portfolio theory is generally regarded as the contribution of William Sharpe, for which he won the Nobel Prize, but John Lintner and J. Mossin developed similar theories independently and at about the same time. For these reasons, the theory is sometimes referred to as the *Sharpe-Lintner-Mossin (SLM) capital asset pricing model*.

The basic idea behind the CAPM is that an investor can actually *improve* his or her risk/expected return balance by investing partially in a portfolio of risky assets and partially in a risk-free asset. Let us see why this is true.

Imagine a portfolio that consists of a risk-free asset $\mathfrak{a}_{\text{rf}}$ with weight w_{rf} and the risky assets $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ as before, with weights $w_1, \ldots, w_n$. Note that now the sum of the weights of the risky assets will be *at most* 1. In fact, we have

$$w_{\text{rf}} + \sum_{i=1}^{n} w_i = 1$$
$$w_{\text{risky}} = \sum_{i=1}^{n} w_i \leq 1$$

The expected return of the complete portfolio is

$$\mu = w_{\text{rf}}\mu_{\text{rf}} + \sum_{i=1}^{n} w_i\mu_i = w_{\text{rf}}\mu_{\text{rf}} + \mu_{\text{risky}}$$

and since the variance of the risk-free asset is 0, the return R_{rf} is a constant. Hence, its covariance with any other return is 0 and so

$$\begin{aligned}\sigma^2 &= \text{Var}(w_{\text{rf}}R_{\text{rf}} + \sum_{i=1}^{n} w_iR_i) \\ &= \text{Var}(\sum_{i=1}^{n} w_iR_i) \\ &= \sigma^2_{\text{risky}}\end{aligned}$$

Hence

$$\sigma = \sigma_{\text{risky}}$$

We also want to consider the portfolio formed by removing the risk-free asset and "beefing up" the weights of the risky assets *by the same factor* to make the sum of these weights equal to 1. Let us call this portfolio the **derived risky portfolio** (a nonstandard term). For example, if the original portfolio is composed of a

risk-free asset with weight $w_{\text{rf}} = 0.20$
risky asset $\mathfrak{a}_1$ with weight $w_1 = 0.30$
risky asset $\mathfrak{a}_2$ with weight $w_1 = 0.50$

then the sum of the risky weights is 0.80 so the derived risky portfolio consists of the

risky asset $\mathfrak{a}_1$ with weight $w_1 = 0.30/0.80 = 0.375$
risky asset $\mathfrak{a}_2$ with weight $w_1 = 0.50/0.80 = 0.625$

which has a total weight of 1. Let us denote the expected return of the derived risky portfolio by μ_{der} and the risk by σ^2_{der}. It follows that

$$\begin{aligned}\mu &= w_{\text{rf}}\mu_{\text{rf}} + \sum_{i=1}^{n} w_i\mu_i \\ &= w_{\text{rf}}\mu_{\text{rf}} + w_{\text{risky}}\sum_{i=1}^{n}\frac{w_i}{w_{\text{risky}}}\mu_i \\ &= w_{\text{rf}}\mu_{\text{rf}} + w_{\text{risky}}\mu_{\text{der}}\end{aligned}$$

and

$$\begin{aligned}\sigma^2 &= \text{Var}(\sum_{i=1}^{n} w_i R_i) \\ &= w^2_{\text{risky}}\text{Var}(\sum_{i=1}^{n}\frac{w_i}{w_{\text{risky}}}R_i) \\ &= w^2_{\text{risky}}\sigma^2_{\text{der}}\end{aligned}$$

Thus

$$\begin{aligned}\mu &= w_{\text{rf}}\mu_{\text{rf}} + w_{\text{risky}}\mu_{\text{der}} \\ \sigma &= w_{\text{risky}}\sigma_{\text{der}}\end{aligned}$$

or since $w_{\text{rf}} + w_{\text{risky}} = 1$

$$\begin{aligned} \mu &= \mu_{\text{rf}} + w_{\text{risky}}(\mu_{\text{der}} - \mu_{\text{rf}}) \\ \sigma &= w_{\text{risky}}\sigma_{\text{der}} \end{aligned} \tag{1}$$

As w_{risky} ranges over all real numbers, equations (1) trace out a straight line. Figure 9 shows the point $(\sigma_{\text{der}}, \mu_{\text{der}})$ corresponding to $w_{\text{risky}} = 1$. This point, being the risk-expected return point for a risky portfolio whose weights sum to 1, must lie in the Markowitz bullet. We are interested in all of the possible values of (σ, μ) given by equations (1).

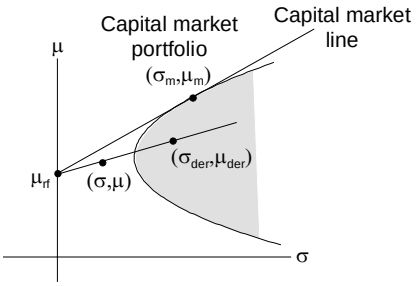


Figure 9

It is clear that if $w_{\text{risky}} = 0$ then

$$(\sigma, \mu) = (0, \mu_{\text{rf}})$$

and if $w_{\text{risky}} = 1$ then

$$(\sigma, \mu) = (\sigma_{\text{der}}, \mu_{\text{der}})$$

and so equations (1) map the line connecting $(0, \mu_{\text{rf}})$ and $(\sigma_{\text{der}}, \mu_{\text{der}})$, whose slope is

$$m = \frac{\mu_{\text{der}} - \mu_{\text{rf}}}{\sigma_{\text{der}}}$$

and whose equation (in slope-intercept form) is

$$\mu = \frac{\mu_{\text{der}} - \mu_{\text{rf}}}{\sigma_{\text{der}}}\sigma + \mu_{\text{rf}} \tag{2}$$

So where do we stand? An investor who invests in a risk-free asset along with some risky assets will have risk-expected return point lying somewhere on the line joining the points $(0, \mu_{\text{rf}})$ and $(\sigma_{\text{der}}, \mu_{\text{der}})$. But it is clear from the geometry that among all lines joining the point $(0, \mu_{\text{rf}})$ with various points $(\sigma_{\text{der}}, \mu_{\text{der}})$ in the Markowitz bullet, the line that

produces the points with the highest expected return for a given risk is the tangent line to the upper portion of the Markowitz bullet, as shown in Figure 10.

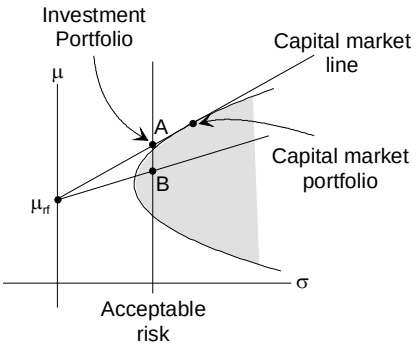


Figure 10–The investment portfolio for a given level of risk

The tangent line in Figure 10 is called the **capital market line** and the point of tangency on the Markowitz efficient frontier is called the **(capital) market portfolio**.

The reader may recall our previous discussion about the flattening out of the Markowitz curves. It follows from this discussion that if the risk-free rate is too large then there will be no capital market line and hence no market portfolio.

Assuming that a capital market line does exist, by adjusting the balance between the risk-free asset and the risky portion of the portfolio, that is, by adjusting the weights w_{rf} and w_{risky}, any point on the capital market line can be achieved. To get a point to the right of the market portfolio requires selling the risk-free asset short and using the money to buy more of the market portfolio.

We can now state the moral of this discussion:

> In order to maximize the expected return for a given level of risk the investor should invest is a portfolio consisting of the risk-free asset and the *market portfolio* (no other risky portfolio). The relative proportions of each is determined by the level of acceptable risk.

The Equation of the Capital Market Line

If the market portfolio has risk-expected return point (σ_M, μ_M) then the equation of the capital market line is

$$\mu = \frac{\mu_M - \mu_{\text{rf}}}{\sigma_M}\sigma + \mu_{\text{rf}}$$

For any point (σ, μ) on the line, the value

$$\mu - \mu_{\text{rf}} = \frac{\mu_M - \mu_{\text{rf}}}{\sigma_M}\sigma$$

which is the additional expected return above the expected return on the risk-free asset, is called the **risk premium**. It is the additional return that one may expect for assuming the risk. Of course, it is the presence of risk that implies that the investor may not actually see this additional return.

To get a better handle on this equation, we need more information about the market portfolio's risk-expected return point (σ_M, μ_M). The weights that correspond to the market portfolio's risk and expected return are given in the next theorem.

Theorem 6 *For any expected risk-free return μ_{rf}, the capital market portfolio has weights*

$$W = \frac{(M - \mu_{rf}O)C^{-1}}{(M - \mu_{rf}O)C^{-1}O^t}$$

Note that the denominator is just a number, being the sum of the coordinates of the vector in the numerator.

Proof. For any point (σ, μ) in the Markowitz bullet, the slope of the line from $(0, \mu_{\text{rf}})$ to (σ, μ) is

$$s = \frac{\mu - \mu_{\text{rf}}}{\sigma} = \frac{\Sigma\mu_i w_i - \mu_{\text{rf}}}{\Sigma c_{i,j} w_i w_j}$$

It is intuitively clear that the point of tangency is the point with the property that this slope is a maximum among all points (σ, μ) in the Markowitz bullet. So we seek to maximize s subject to the constraint that $\Sigma w_i = 1$. Using Lagrange multipliers once again, we must take the partial derivatives of the following function and set the results to 0

$$f = \frac{\sum_i \mu_i w_i - \mu_{\text{rf}}}{\sum_{i,j} c_{i,j} w_i w_j} + \lambda(1 - \sum_i w_i)$$

We leave it as an exercise to show that the resulting equations are

$$\frac{\partial f}{\partial w_k} = \frac{\mu_k \Sigma c_{i,j} w_i w_j - (\Sigma \mu_i w_i - \mu_{\text{rf}})(\Sigma c_{i,k} w_i)}{(\Sigma c_{i,j} w_i w_j)^{3/2}} - \lambda = 0$$

This can be cleaned up to get

$$(WCW^t)\mu_k - (MW^t - \mu_{\text{rf}})C_k W^t = \lambda (WCW^t)^{3/2}$$

where C_k is the k-th row of the covariance matrix C. This can be written

$$\sigma^2 \mu_k - (\mu - \mu_{\text{rf}})C_k W^t = \lambda \sigma^3$$

Since this holds for all k, we have

$$\sigma^2 M^t - (\mu - \mu_{\text{rf}})CW^t = \lambda \sigma^3 O^t$$

Taking transposes and recalling that $C^t = C$ gives

$$\sigma^2 M - (\mu - \mu_{\text{rf}})WC = \lambda \sigma^3 O$$

Multiplying on the right by W^t and recalling that $OW^t = 1$, we get

$$\sigma^2 MW^t - (\mu - \mu_{\text{rf}})WCW^t = \lambda \sigma^3$$

or

$$\sigma^2 \mu - (\mu - \mu_{\text{rf}})\sigma^2 = \lambda \sigma^3$$

and so

$$\lambda = \frac{\mu_{\text{rf}}}{\sigma}$$

We can now use this value of λ in an earlier equation to get

$$\sigma^2 M - (\mu - \mu_{\text{rf}})WC = \mu_{\text{rf}} \sigma^2 O$$

This can be rewritten as

$$\frac{\mu - \mu_{\text{rf}}}{\sigma^2} W = (M - \mu_{\text{rf}} O)C^{-1}$$

Multiplying on the right by O^t and noting that $WO^t = 1$ we get

$$\frac{\mu - \mu_{\text{rf}}}{\sigma^2} = (M - \mu_{\text{rf}} O)C^{-1}O^t$$

Using this in the previous equation gives

$$W = \frac{(M - \mu_{\text{rf}} O)C^{-1}}{(M - \mu_{\text{rf}} O)C^{-1}O^t}$$

as desired.□

To illustrate, let us continue Example 1 to derive the market portfolio.

EXAMPLE 2 Continuing Example 1, Figure 11 shows more of our Excel worksheet. This portion computes the market portfolio's risk-expected return based on various risk-free rates (in this case only three rates).

Market Portfolio

Risk-Free Rate	M-RFR*O			(M-RFR*O)C^{-1}			Den
0.01	0.09	0.1	0.06	1.883595808	1.659126341	0.433966462	3.976688611
0.02	0.08	0.09	0.05	1.701097806	1.509556481	0.29421513	3.504869417
0.03	0.07	0.08	0.04	1.518599804	1.359986621	0.154463797	3.033050222

W			Return μ	WC			σ_p=WCW^t	Risk σ
0.473659366	0.417213039	0.109127594	0.100898303	0.022631895	0.02514655	0.01508793	0.022857787	0.151187921
0.485352692	0.430702632	0.083944677	0.101788686	0.022825387	0.02567856	0.014265867	0.02333573	0.152760368
0.500684029	0.448389087	0.050926884	0.102956084	0.023079077	0.026376088	0.013188044	0.024053701	0.155092557

Figure 11–The market portfolio

For instance, a risk-free return on investment of $\mu_{\text{rf}} = 0.03$ leads to a expected return of

$$\mu_m = 1.02956084$$

and a risk of

$$\sigma_m = 0.155092557$$ □

More on the Market Portfolio

According to our theory, all rational investors will invest in the market portfolio, along with some measure of risk-free asset. This has some profound consequences for this portfolio. First, the market portfolio must contain all possible assets! For if an asset is not in the portfolio, no one will want to purchase it and so the asset will wither and die.

Since the market portfolio contains all assets, the portfolio has no unsystematic risk—this risk has been completely diversified out. Thus, all risk associated with the market portfolio is systematic risk.

In practice, the market portfolio can be approximated by a much smaller number of assets. Studies have indicated that a portfolio can achieve a degree of diversification approaching that of a true market portfolio if it contains a well-chosen set of perhaps 20-40 securities. We will use the term market portfolio to refer to an unspecified portfolio that is highly diversified and thus can be considered as essentially free of unsystematic risk.

The Risk-Return of an Asset Compared with the Market Portfolio

Let us consider any particular asset a_k in the market portfolio. We want to use the best linear predictor, discussed in Chapter 1, to approximate the return R_k of asset a_k by a linear function of the return R_M of the entire market portfolio. According to Theorem 9 of Chapter 1, we can write

$$R_k = \beta_k R_M + \alpha_k + \epsilon$$

where

$$\beta_k = \frac{\text{Cov}(R_k, R_M)}{\sigma_M^2}$$
$$\alpha_k = \mathcal{E}(R_i) - \beta_k \mathcal{E}(R_M)$$

and ϵ is the error (residual random variable). The coefficient β_k is the *beta* of the asset's return with respect to the market portfolio's return and is the slope of the linear regression line.

To get a feel for what to expect, Figure 12 shows the best linear predictor in the case of a relatively large beta and three magnitudes of error, ranging from very small to rather large.

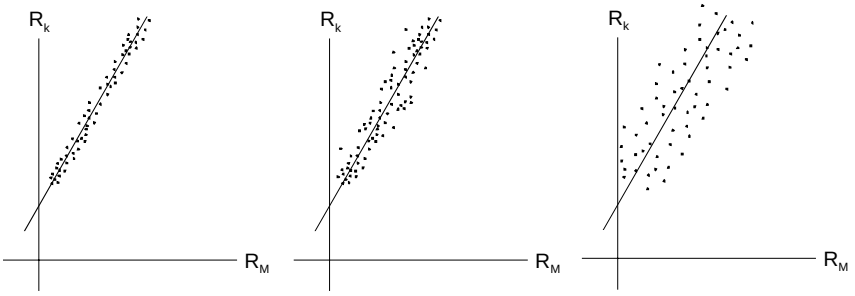


Figure 12–A large beta and different magnitudes of error

Because the beta is large, in all three cases when the market return fluctuates a certain amount, the asset's return fluctuates a relatively larger amount. Put another way, if the market returns should fluctuate over a specific range of values (as measured by the variance, for example), the asset returns will fluctuate over a larger range of values (as measured by the variance). Thus, the market risk is “magnified” in the asset risk.

Figure 13 shows the best linear predictor when the beta is small, again with three magnitues of error.

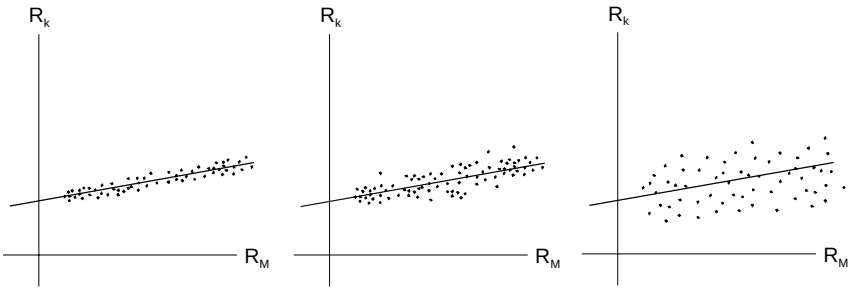


Figure 13–A small beta and different magnitudes of error

Because the beta is small, in all three cases when the market return fluctuates a certain amount, the asset's return fluctuates a relatively smaller amount. Thus, the market risk is “demagnified” in the asset risk.

It is intuitively clear then that an asset's *systematic risk*, that is, the risk that comes from the asset's relationship to the market portfolio (whose risk is purely systematic) is related in some way to the beta of the asset.

In addition, it can be seen from the graphs in Figures 12 and 13 that there is another factor that contributes to the asset's risk, a factor that has nothing whatever to do with the market risk. It is the error. The larger the error ϵ, as measured by its variance $\mathrm{Var}(\epsilon)$ for example, the larger the uncertainty in the asset's expected return.

Now let us turn to the mathematics to see if we can justify these statements. In fact, the BLP will provide formulas for the expected return and the risk of the individual asset R_k in terms of the beta.

As to the risk, we leave it as an exercise to show that

$$\text{Cov}(R_M, \epsilon) = 0$$

and so the risk associated with the asset $\mathfrak{a}_k$ is

$$\begin{aligned}\sigma_k^2 = \text{Var}(R_k) &= \text{Var}(\beta_k R_M + \alpha_k + \epsilon) \\ &= \text{Var}(\beta_k R_M + \epsilon) \\ &= \beta_k^2 \sigma_M^2 + \text{Var}(\epsilon)\end{aligned}$$

Thus, we see that our intuition is upheld. The term $\beta_k^2 \sigma_M^2$, which is referred to as the **systematic risk** of the asset $\mathfrak{a}_k$, is indeed an increasing function of the beta. Thus, it is the beta of the asset, that is, the slope of the best linear fit, that measures the proportion of the market portfolio's risk that makes up part of the asset's risk.

The remaining portion of the asset's risk is the term $\text{Var}(\epsilon)$, which is precisely the measure of the error that we discussed earlier. This is called the **unsystematic risk** or **unique risk** of the asset.

According to economic theory, when adding an asset to a *diversified* portfolio, the unique risk of that asset is canceled out by other assets in the portfolio. Accordingly, the unique risk should not be considered when evaluating the expected return of the asset. Put another way, when determining whether the expected return of an asset is sufficient to accept the risk involved, only the systematic risk of the asset should be considered. Thus, the expected return of an asset that the market will sustain under market equilibrium depends only on the asset's beta.

Let us see if we can justify this statement by turning our attention to a formula for the asset's expected return. The expected return of the market portfolio is

$$\mu_M = MW_M^t$$

and the expected return of the individual asset $\mathfrak{a}_i$ is

$$\mu_k = M{e_k}^t$$

where

$$e_k = (\,0 \quad \cdots \quad 0 \quad 1 \quad 0 \quad \cdots \quad 0\,)$$

is the matrix with a 1 in the kth position and 0s elsewhere. To relate these two quantities, we need an expression for M.

Recall that the weights of the market portfolio are given in Theorem 6 by

$$W_M = \frac{(M - \mu_{\text{rf}} O)C^{-1}}{(M - \mu_{\text{rf}} O)C^{-1}O^t}$$

Since the denominator is just a constant, let us denote its reciprocal by δ. Thus

$$W_M = \delta(M - \mu_{\text{rf}} O)C^{-1}$$

Solving for M gives

$$M = \frac{1}{\delta} W_M C + \mu_{\text{rf}} O$$

We can now write

$$\begin{aligned}
\mu_M &= M W_M^t \\
&= \left(\frac{1}{\delta} W_M C + \mu_{\text{rf}} O\right) W_M^t \\
&= \frac{1}{\delta} W_M C W_M^t + \mu_{\text{rf}} O W_M^t \\
&= \frac{1}{\delta} \sigma_M^2 + \mu_{\text{rf}}
\end{aligned}$$

Also

$$\begin{aligned}
\mu_k &= M e_k^t \\
&= \left(\frac{1}{\delta} W_M C + \mu_{\text{rf}} O\right) e_k^t \\
&= \frac{1}{\delta} W_M C e_k^t + \mu_{\text{rf}} O e_k^t \\
&= \frac{1}{\delta} \text{Cov}(R_k, R_M) + \mu_{\text{rf}}
\end{aligned}$$

Now, the reader may notice a resemblance between some of these terms and the beta

$$\beta_k = \frac{\text{Cov}(R_k, R_M)}{\sigma_M^2}$$

Solving the previous equations for the numerator and denominator of β_k gives

$$\beta_k = \frac{\text{Cov}(R_k, R_M)}{\sigma_M^2} = \frac{\delta(\mu_k - \mu_{\text{rf}})}{\delta(\mu_M - \mu_{\text{rf}})} = \frac{\mu_k - \mu_{\text{rf}}}{\mu_M - \mu_{\text{rf}}}$$

Finally, solving ths for μ_k gives

$$\mu_k = \beta_k(\mu_M - \mu_{\text{rf}}) + \mu_{\text{rf}}$$

Let us collect these important formulas in a theorem.

Theorem 7 *The expected return and risk of an asset* $\mathfrak{a}_k$ *in the market portfolio is related to the asset's beta with respect to the market portfolio as follows*

$$\mu_k = \beta_k(\mu_M - \mu_{rf}) + \mu_{rf} \qquad (3)$$

and

$$\sigma_k^2 = \beta_k^2\sigma_M^2 + \text{Var}(\epsilon)$$

where ϵ *is the error (residual random variable).*□

The expression for the expected return of the asset is very interesting, for it says that the expected return is a *linear function* of the beta. This justifies our earlier discussion to the effect that an asset's expected return under market equilibrium should depend only on the asset's systematic risk (through the asset's beta) and not on its unique risk.

Since $\mu_M - \mu_{\text{rf}}$ is positive under normal conditions, the slope of the linear relation is positive, meaning that large betas imply large expected returns and vice-versa. This makes sense—the more (systematic) risk in an asset the higher should be the expected return.

Of course, there is no *law* that says that higher risk should be rewarded by higher expected return. However, this is the condition of market equilibrium. If an asset is returning less than the market feels is reasonable with respect to the asset's perceived risk, then no one will buy that asset and its price will decline, thus increasing the asset's return. Similarly, if the asset is returning more than the market feels is required by the asset's level of risk, then more investors will buy the asset, thus raising its price and lowering its expected return.

The graph of the line in equation (3) is called the **security market line** or SML for short. The equation shows that the expected return of an asset is

equal to the return of the risk-free asset plus the so-called **risk premium** of the asset.

EXAMPLE 3 Suppose that the risk-free rate is 3% and that the market portfolio's risk is 12%. Consider the following assets and their betas

Asset	Beta
$\mathfrak{a}_1$	0.65
$\mathfrak{a}_2$	1.00
$\mathfrak{a}_3$	1.20
$\mathfrak{a}_4$	-0.20
$\mathfrak{a}_5$	-0.60

The market's risk premium is

$$\mu_M - \mu_{\text{rf}} = 0.12 - 0.03 = 0.09$$

and so the security market equation is

$$\mu_k = 0.09\beta + 0.03$$

We can now compute the expected returns under market equilibrium

Asset	Beta	Expected Return μ_k
$\mathfrak{a}_1$	0.65	8.85%
$\mathfrak{a}_2$	1.00	12%
$\mathfrak{a}_3$	1.20	13.8%
$\mathfrak{a}_4$	-0.20	1.2%
$\mathfrak{a}_5$	-0.60	-6.9%

The expected returns in the previous table are the values that the market will sustain based on the market portfolio's overall systematic risk (and the risk-free rate). For example, since asset $\mathfrak{a}_1$ has a beta less than 1, it has a smaller risk than the market portfolio. Therefore, the market will sustain a lower expected return than that of the market portfolio. Asset $\mathfrak{a}_2$ has the same systematic risk as the market portfolio so the market will sustain an expected return equal to that of the market portfolio.

Exercises

1. What is the beta of the market portfolio? Can a portfolio have any real number as its β?

2. For a risk-free rate of 4% and a market portfolio expected return of 8% calculate the equation of the security market line.
3. Show that the parametric equations

$$\mu = (1-s)\mu_1 + s\mu_2$$
$$\sigma' = (1-s)\sigma_1 \pm s\sigma_2$$

   are the equations of a straight line in the (σ', μ)-plane. (Here s is the parameter and ranges over all real numbers. Take the plus sign first and then the minus sign.)
4. For $\rho_{1,2} = 1$ we have $R_2 = aR_1 + b$ for $a > 0$. If the risk is 0 then compute the expected return.
5. Under the assumption that $\sigma_1 \leq \sigma_2$ and $\rho_{1,2} < 1$ show that

$$s_{\min} = \frac{\sigma_1(\sigma_1 - \rho_{1,2}\sigma_2)}{\sigma_1^2 + \sigma_2^2 - 2\rho_{1,2}\sigma_1\sigma_2} < 1$$

6. If ϵ is the error in the best linear predictor of an asset $\mathfrak{a}_i$ with respect to the market portfolio, show that $\mathrm{Cov}(R_M, \epsilon) = 0$.
7. Show that the regression lines for all assets in the market portfolio go through a single point. What is that point?
8. Let P_M be the market portfolio, where asset $\mathfrak{a}_i$ has weight w_i. Write the best linear predictor of R_i as

$$\mathrm{BLP}(R_i) = \beta_i R_M + \alpha_i$$

   Consider the first two assets $\mathfrak{a}_1$ and $\mathfrak{a}_2$, with their respective weights w_1 and w_2. The return from these two assets is

$$R_0 = \mathrm{w}_1 R_1 + w_2 R_2$$

   If the best linear predictor is

$$\mathrm{BLP}(R_0) = \beta_0 R_M + \alpha_M$$

   what is the relationship between β_0, β_1 and β_2 and between α_0, α_1 and α_2? Can you generalize this result to any subset of assets in the market portfolio, that is, to any subportfolio?
9. Verify the data in Figure 14 (at least to a few decimal places). For a 5% return, show that the minimum risk is 0.238952. If the risk-free rate is 3% show that the market portfolio has weights $(0.335, 0.372, 0.293)$ and risk-return $(0.194, 0.196)$.

Capital Asset Pricing Model-Fill In Grey Cells

User Data		**Returns μ_i**	**Risks σ_i**	**Correlation**	
	i=1	0.1	0.2	-0.1	$=\rho_{1,2}=\rho_{2,1}$
	i=2	0.2	0.3	0.2	$=\rho_{1,3}=\rho_{3,1}$
	i=3	0.3	0.4	0.2	$=\rho_{2,3}=\rho_{3,2}$

$C=(c_{i,j})=(\rho_{i,j}\sigma_i\sigma_j)$		j=1	j=2	j=3
	i=1	0.04	-0.006	0.016
	i=2	-0.006	0.09	0.024
	i=3	0.016	0.024	0.16

Inverse of C			
26.6075388	2.58684405	-3.048780488	
2.58684405	11.8255728	-2.032520325	
-3.048780488	-2.032520325	6.859756098	

Min Risk Point				
	$OC^{-1}=$	26.14560237	12.37989653	1.778455285
	$OC^{-1}O^t=$	40.30395418		
	$W=OC^{-1}/OC^{-1}O^t=$	0.648710602	0.307163324	0.044126074
	$\mu=MW^t=$	0.139541547		
	$WC=$	0.024811461	0.024811461	0.024811461
	$\sigma^2=WCW^t=$	0.024811461		
	$\sigma=$	0.157516543		

For Min Risk Line				
	$MC^{-1}=$	2.263488544	2.014042868	1.346544715
	$MC^{-1}O^t=$	5.624076127		
	$OC^{-1}M^t=$	5.624076127		
	$MC^{-1}M^t=$	1.033120843		
	Denom Det=	10.00862281		

Chapter 3

Background on Options

In preparation for our study of derivative pricing models, we need to discuss the basics of stock options. Readers who are familiar with these derivatives will want merely to skim through the chapters to synchronize the terminology, as it were.

Stock Options

Stock options take two forms: *put options* (*puts)* and *call options* (*calls)*. Here are the definitions.

Definition *A* **call** *is a contract between the* **writer** *(or* **seller***) of the call and the* **buyer** *of the call. The buyer has the right to buy from the writer (that is,* call for*) the stock at a fixed price called the* **exercise price** *or* **strike price**, *which we denote by* E *or* K *(both are commonly used symbols). In a* **European call**, *the right to buy can only be exercised on the expiration date of the call. In an* **American call**, *the right to buy can be exercised at any time on or before the expiration date of the call.*

A **put** *is a contract between the* **writer** *(or* **seller***) of the put and the* **buyer** *of the put. The buyer has the right to sell to (or* put to*) the writer the stock at the* **exercise price** *or* **strike price**. *In a* **European put**, *the right to sell can only be exercised on the expiration date of the call. In an* **American put**, *the right to sell can be exercised at any time on or before the expiration date of the call.*

The writer of an option has a **short position** *and the buyer of an option has a* **long position**. □

The Purpose of Options

Options are primarily used for *hedging* and for *speculation*. (Arbitrage is always good too if you can get it.) Also, options have one significant advantage over owning the underlying asset, namely, *leverage*.

A **hedge** is an investment that reduces the risk in an existing position, such as another investment. To illustrate the hedging feature of an option, suppose an investor currently (October) owns 1000 shares of a stock IBM whose current price is about $88 per share. The investor is

justifiably concerned that a significant drop is the stock's price will cause his portfolio to take a big hit.

So to hedge against this possibility the investor buy a December put with strike price $85. This gives him the right to sell the stock at $85 per share for the next 3 months. Thus, if the stock price tumbles the investor can bail out at $85 per share. The price paid for this hedge is the price of the put, which is currently selling for $1.50. So a total cost of $1500 will protect an $88,000 investment.

Leverage

At the moment of this writing, IBM is selling for about $90 per share. A small investor with $450 can purchase only 5 shares of the stock. If the investor feels that the stock price is about to rise significantly, then the use of options allows him to leverage his meager bankroll and speculate on the stock in a much more meaningful way than buying the shares.

For example, the current price of a 1 month call with strike price of $90 is $3.80. Thus, the investor is able to purchase 118 such calls (ignoring commissions). If the price of IBM is $95 at exercise time the profit on 5 shares would be only $25 whereas the profit on the calls would be $590. The return is thus over 100% on the investment in options, whereas it is less than 6% for the stock investment. This is leverage.

Of course, the downside to the call options is that if the stock does not rise, or does not rise before the expiration date, the investor will receive nothing from the options and will be out the commission on the purchase of these options, whereas the stock holder still owns the stock.

Profit and Payoff Curves

When the expiration date arrives, the owner of an option will exercise that option if and only if there is a positive return. Thus, if the strike price of the option is K and the spot price of the stock is S, the owner of a call will exercise the option (call for the stock at the price K) if and only if $K < S$. On the other hand, the owner of a put will exercise the option (put the stock at the price K) if and only if $K > S$. Some terms are used to describe the various possibilities.

Definition *A call option is*
1) **in-the-money** *if* $K < S_T$
2) **at-the-money** *if* $K = S_T$
3) **out-of-the-money** *if* $K > S_T$

A put option is
4) **in-the-money** *if* $K > S_T$
5) **at-the-money** *if* $K = S_T$
6) **out-of-the-money** *if* $K < S_T$□

It is important to note that just because an option is in the money does not mean that the owner makes a profit. The problem is that the initial cost (as well as any commissions, which we will ignore throughout this discussion) may outweigh the return gained from exercising the option. In that case, the investor will still execute because the positive return will help reduce the overall loss.

Figure 2 shows the *payoffs* (ignoring costs) for each option position. Note that the horizontal axis is the stock price at exercise time and all line segments are either horizontal or have slope ± 1.

For example, for a long call, the owner will exercise if and only if the spot price S of the stock is greater than K. In this case, the payoff to the owner is $S - K$. Otherwise, the owner will let the call expire, receiving nothing.

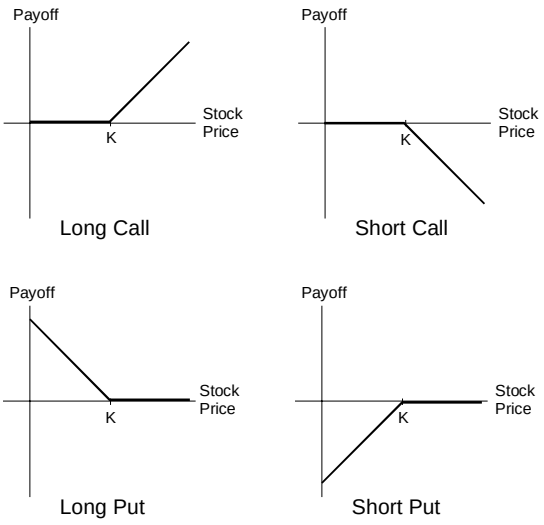


Figure 2 - Payoff Curves

Figure 3 shows the actual profit curves, which take into account the cost of the purchase or sale of the option. (As mentioned, we will ignore all commissions.)

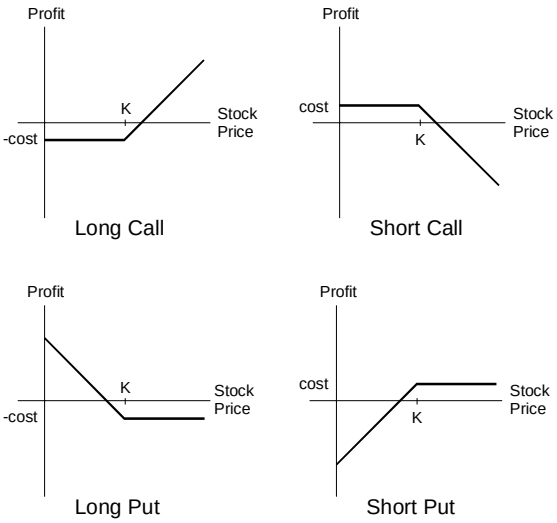


Figure 3 - Profit Curves

The payoff formulas are actually quite simple. For a long call, if the stock price S satisfies $S \geq K$ then the payoff from exercising the call is $S - K$ whereas if $S < K$ then the call will expire and so the payoff is 0. Thus, the payoff is

$$\text{Payoff(Long Call)} = \max\{S - K, 0\}$$

On the put side, we have

$$\text{Payoff(Long Put)} = \max\{K - S, 0\}$$

As mentioned, the payoff curves are very informative. Here are some of the things we can see immediately from these curves.

Long Call

- Limited Downside: The downside is limited to the cost of the call.
- Unlimited Upside: The upside is unlimited since there is no limit to the price of the stock.
- Optimistic Position: The buyer hopes the stock price will rise.
- A long call should be exercised when the stock price is above the strike price K.

Short Call

- Unlimited Downside: The downside is unlimited because there is no limit to the price of the stock.

- Limited Upside: The upside is limited to the selling price of the call.
- Pessimistic Position: The buyer hopes the stock price will fall.
- A short call should be exercised when the stock price falls below the strike price K.

Long Put

- Limited Downside: The downside is limited to the cost of the put.
- Limited Upside: The upside is also limited because the stock price can only fall to 0, in which case the profit is equal to the strike price times the number of shares minus the cost of the put.
- Pessimistic Position: The buyer hopes the stock price will fall.
- A long put should be exercised when the stock price falls below the strike price K.

Short Put

- Limited Downside: The downside is limited because the stock price can only fall to 0, in which case the loss is equal to the strike price times the number of shares $plus$ the cost of the put.
- Limited Upside: The upside is also limited to the selling price of the put.
- Optimistic Position: The buyer hopes the stock price will rise.
- A short put should be exercised when the stock price rises above the strike price K.

It is also worth noting that long calls and short puts are related in that they are both **optimistic (bullish) positions**, that is, they are profitable when the stock rises. It is the degree of risk and the degree of profit that distinguish the two in this regard, however. Similarly, long puts and short calls are both **pessimistic (bearish) positions**.

The profit curves also hold some interesting information. Setting aside for the moment the risk factor, we can say the following:

- If we believe that a stock's price will decline but only slightly, settling within the interval $(K - \text{Cost}, K]$, then a short call is the most advantageous position. In this case, a long put will still be in the red, due to the cost of the put. However, if we believe that a stock's price will decline sharply, then a long put is the most advantageous position.
- If we believe that a stock's price will rise but only slightly, then a short put is the most advantageous position. If we believe that a

stock's price will rise sharply, then a long call is the most advantageous position.

Covered Calls

We have said that a short call position has an unlimited downside because the stock price can theoretically rise indefinitely. Of course, this assumes that the writer of the call buys the shares at exercise time in order to deliver them to the owner of the call. However, if the writer of the call already owns the shares his downside is limited to the price paid for those shares, because they can be used to "cover" the call.

If the writer of a call owns the shares at the time he writes the call, then he is said to write a **covered call**. Writing covered calls is far safer than writing **uncovered** (also called **naked**) calls.

Profit Curves for Option Portfolios

An **option portfolio** consists of a collection of options. The following example shows how to obtain the profit curve for a simple portfolio.

EXAMPLE 2 Consider the purchase and sale of options, all with the same expiration date, given by the following expression

$$-P_{100} + P_{120} + 2C_{150} - C_{180}$$

This position is: short a put with strike price 100, long a put with strike price 120, long two calls with strike price 150 and short a call with strike price 180. The overall profit curve can be obtained from the individual profit curves by plotting them all on a single set of coordinates, as shown in Figure 3. Note that it is simpler to ignore all costs in drawing the curves and then simply translate the final curve up or down an amount equal to the total cost for all the options involved, which in this case is

$$-\text{Cost}(P_{100}) + \text{Cost}(P_{120}) + 2 \cdot \text{Cost}(C_{150}) - \text{Cost}(C_{180}) \qquad \square$$

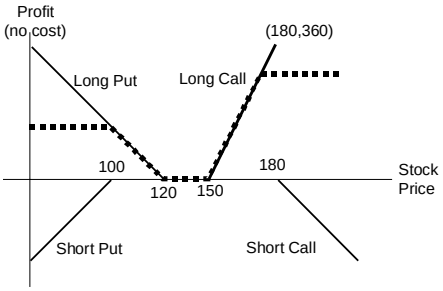


Figure 3 - Profit Curve (no costs)

Selling Short

A complete analysis of options cannot be made without also discussing the notion of selling a stock short. Simply put, to sell a stock short, the investor borrows the stock (usually from a broker) and sells it immediately (in one transaction), thus realizing an amount equal to the current price of the stock (less the ever-present commission). For this privilege, the investor must return the stock (not the money) to the lender.

As with short calls, selling a stock short incurs a potentially unlimited downside, unless the seller also owns shares of the stock with which to cover the inevitable return of the stock borrowed. Figure 3 shows the profit curve for a short sale of stock, as well as the profit curve for a long position.

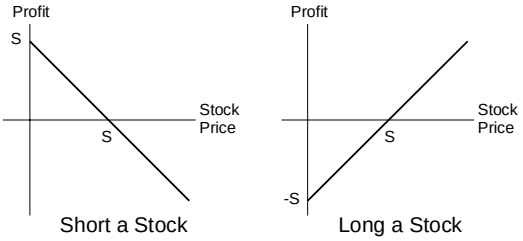


Figure 3 - Profit Curves for Short and Long Stock Positions

Exercises

1. Without looking in the book, draw the profit curves for a long put, short put, long call and short call.
2. To write a **covered put**, the investor writes a put and at the same time must be short the same quantity of the underlying stock. For example, suppose an investor writes a put for 100 shares of IBM and is also short 100 shares of IBM. This means that the investor has borrowed 100 shares of IBM and sold them. Describe the upside and downside to writing covered puts. Why would someone want to write a covered put?
3. Draw the payoff graph for the following option portfolio

$$-P_{80} + P_{100} + 2C_{130} - C_{150}$$

A **spread** is a transaction in which an investor simultaneously buys one option and sells another option, both on the same underlying asset, but

with different terms (strike price and/or expiration date). A **call spread** involves the purchase and sale of calls, and similarly for a **put spread**. The idea is that one option is used to hedge the risk of the other option.

4. In a **bull spread**, the investor buys a call at a certain strike price K_1 and sells another call at a higher strike price K_2, with the same expiration date. Draw the profit curve for a bull spread. When is a bull spread most profitable? Is this an optimistic or pessimistic investment? *Hint*: you must first decide how the costs of the two calls compare.
5. In a **bear spread**, the investor buys a call at a certain strike price K_1 and sells another call at a lower strike price K_2, with the same expiration date. Draw the profit curve for a bear spread. When is a bear spread most profitable? Is this an optimistic or pessimistic investment? *Hint*: you must first decide how the costs of the two calls compare.
6. In a **calendar spread** also called a **time spread** an investor sells a call with a certain expiration date D_1 and buys a **more distant** call, that is, a call with a longer expiration date $D_2 > D_1$. Assume that the calls have the same strike price. Consider the following calendar spread. The current (JAN) price of XYZ is \$50. Call prices as as follows:

   APR 50 call: (expiring in April at a strike price of \$50) costs \$5
   JUL 50 call: \$8
   OCT 50 call: \$10

   Suppose that in 3 months (in April) the stock price is still \$50. Then if all things else are equal the call prices should be

   APR 50 call: \$0 (expiring)
   JUL 50 call: \$5
   OCT 50 call: \$8

   Is there a profit here for the investor? Describe the reason.
7. A **butterfly spread** is a combination of a bull spread and a bear spread. A call butterfly spread consists of buying a call at strike price K_1, selling two calls at strike price $K_2 > K_1$ and selling another call at strike price $K_3 > K_2$. All calls have the same expiration date. Draw a profit curve for a butterfly spread. *Hint*: they don't call it a butterfly spread for nothing.

Chapter 4

An Aperitif on Arbitrage

As a simple introduction to the concept of arbitrage and how to use the assumption of no arbitrage to price assets, let us briefly discuss the pricing of forward contracts and some simple issues related to option pricing.

Background on Forward Contracts

We begin with the necessary background on forward contracts.

Forward Contracts

A **forward contract** is an agreement to buy a certain quantity of an asset, called the **underlying asset** at a given price K, called the **settlement price** or **delivery price** to be paid at a given time T in the future, called the **settlement date** or **delivery date**. Entering a forward contract does not require any initial purchase price—it is free.

The party that agrees to buy the asset is taking the **long position** on the contract and is said to be the **buyer** of the contract. The party that agrees to sell the asset is taking the **short position** on the contract and is said to be the **seller** of the contract.

Futures Contracts

In contrast to plain-vanilla forward contracts as described above, a **futures contract** is a forward contract with a number of constraints and a much more complicated payoff model. Indeed, futures contracts seldom come to maturity, that is, very few (perhaps on the order of 1 or 2 percent) of all futures contracts survive to the delivery date. The main properties of futures contracts are as follows.

1) Futures contracts trade on an organized exchange. For example, the Chicago Board of Trade (CBT or CBOT) is the largest futures exchange.
2) Futures contracts have standardized terms, specifying the amount and precise type of the underlying, the delivery date and delivery price. Just like you can only buy bolts of specific lengths and diameters at the hardware store, you can only buy futures contracts with specific terms.

3) Performance (delivery of losses or gains) of futures contracts is guaranteed by a *clearinghouse*.
4) The purchase of a futures contract requires that the buyer post *margin*, that is, some amount of money to cover potential day-to-day price changes.
5) Futures markets are regulated by a government agency, whereas forward contracts are largely unregulated.
6) Futures contracts can be *closed* (terminated) either by delivery, by offset (that is, by a reversing trade that cancels both contracts) or by exchange-for-physical (which is a form of "settle up early" arrangement).

We will not discuss the details of futures contracts in this book.

Forward Prices

Consider forward contracts for a given underlying (such as wheat) that have a given delivery date T (such as December 2003). At any time $t < T$, one can potentially enter into such a contract. Of course, the delivery price will depend on the time t of formation of the contract, so we will denote it by $F_{t,T}$. This would-be delivery price is called the **forward price** of the contract.

For example, on July 1 the forward price of a contract to deliver 5000 bushels of wheat in September might be 170 cents (per bushel). A week later, the forward price for such a contract might be 168 cents.

Spot Prices

In contrast to forward prices, the **spot price** S_t of an asset at a given time t is the price of the asset *at that time* for *immediate* delivery. For example, we can speak of the current spot price of a bushel of wheat. We can also speak of the spot price of wheat in one month. This is the price that investors would pay in one month for immediate delivery at that time. Of course, at the present time, this spot price is unknown.

The Pricing of Forward Contracts

To determine the forward price of a forward contract, we can use a simple no-arbitrage argument. Suppose that the forward contract is for one "share" of an asset whose initial price is S_0. (One share of a wheat contract is 5000 bushels of wheat, for example.) Consider the following two portfolios.

Portfolio A: Long the Contract

One forward contract.

Portfolio B: Cash-and-Carry the Asset
One share of the asset itself and a debt of S_0 dollars

In a perfect market it is possible to go either long or short on either portfolio. To short Portfolio A, we short (sell) the forward contract. To short Portfolio B, we sell the asset short (borrow the asset and sell it for S_0) and then lend the resulting income. This is referred to as **reverse cash-and-carry**.

The initial values for these portfolios are 0. The final payoffs are

$$\begin{aligned}\mathcal{V}(\text{long contract}) &= S_T - F_{0,T}\\ \mathcal{V}(\text{short contract}) &= F_{0,T} - S_T\\ \mathcal{V}(\text{cash-and-carry}) &= S_T - S_0e^{rT}\\ \mathcal{V}(\text{reverse cash-and-carry}) &= S_0e^{rT} - S_T\end{aligned}$$

For example, in the case of cash-and-carry, at time T the investor owns the asset, worth S_T but must repay the loan, which values S_0e^{rT}.

Now consider the following two strategies.

Strategy 1: Long the contract and reverse cash-and-carry the asset
The final payoff for this strategy is

$$\mathcal{V}(\text{long contract}) + \mathcal{V}(\text{reverse cash-and-carry}) = S_0e^{rT} - F_{0,T}$$

Strategy 2: Short the contract and cash-and-carry the asset
The final payoff for this strategy is

$$\mathcal{V}(\text{short contract}) + \mathcal{V}(\text{cash-and-carry}) = F_{0,T} - S_0e^{rT}$$

If either of these *constant* payoffs is positive, the investor has an arbitrage strategy. Hence, the lack of arbitrage implies that

$$S_0e^{rT} - F_{0,T} = 0$$

that is

$$F_{0,T} = S_0e^{rT}$$

Theorem 1 *Consider the terms of a forward contract to buy an underlying asset at time T in the future. Then under the assumption of a perfect market with no arbitrage, the forward price is*

$$F_{0,T} = S_0 e^{rT}$$

where r is the interest rate.□

The Put-Call Option Parity Formula

We can also apply the no-arbitrage principle to derive relationships between the prices of puts and calls for the same underlying under the same conditions (strike price and expiration date).

First, let us make a comment about risk-free bonds. We will assume that it is possible to buy or sell any amount of a risk-free bond (such as a U.S. Treasury bond). The bond pays a continuouly compounded interest at the rate r. We will assume that the value of 1 unit of risk-free bond at time 0 is 1 dollar.

It is important to keep separate in one's mind the notion of quantity and price of the bond. For example, if we invest in A units of risk-free bond at time 0 then at time t the quantity is still A but the value is Ae^{rt}. Also, if we invest K dollars in the risk-free bond at time t we get Ke^{-rt} units of bond each worth e^{rt}.

It will also be convenient to use the following common notation

$$X^+ = \max\{X, 0\}$$

The European Case

The *put-call option parity formula* is a formula that compares the price P of a European put to the price C of a European call on the same underlying stock and with the same expiration date and strike price K.

Assume that the underlying stock pays no dividend and is currently selling for S_0. Consider the following two portfolios.

Portfolio A
A *long* position on the put and a *short* position on the call. The initial value of this portfolio is $P - C$ and the payoff is

$$\max(K - S_T, 0) - \max(S_T - K, 0) = K - S_T$$

where S_T is the final price of the stock.

Portfolio B
A *short* position on one share of stock and Ke^{-rT} worth of risk-free bond. The initial value of this portfolio is

$$Ke^{-rT} - S_0$$

and the final payoff is

$$K - S_T$$

Since the final payoffs of the two portfolios are the same, the initial values must also be the same. Otherwise, an investor could sell the more expensive portfolio and buy the cheaper portfolio, which would produce a guaranteed profit at time T. Hence

$$P - C = Ke^{-rT} - S_0$$

This is the put-call option parity formula.

If the stock pays a dividend then the analysis is somewhat different. The reason is that the investor from whom the stock was borrowed under Portfolio B will demand not only the return of his share of stock, but also the return of the dividends that he has foregone by lending the stock.

Suppose that the time-t_0 value of the dividend is d_0. Then the final payoff of Portfolio B is

$$K - S_T - d_0 e^{rT}$$

Thus, we cannot compare the initial values of the two portfolios, since the final payoffs are not equal. This calls for an adjustment to Portfolio B so that the payoff is the same as that of Portfolio A.

Portfolio B′
A *short* position on one share of stock and $Ke^{-rT} + d_0$ worth of risk-free bond. The initial value of this portfolio is

$$Ke^{-rT} + d_0 - S_0$$

and the final payoff is

$$K + d_0 e^{rT} - S_T - d_0 e^{rT} = K - S_T$$

Now we can equate the initial values of Portfolio A and Portfolio B′ to get

$$P - C = Ke^{-rT} + d_0 - S_0$$

Theorem 2 *(European Options With Dividends) Suppose that a stock is currently selling at a price of S_0 per share, a European put on this stock sells for P dollars and a European call for C dollars, both having the same strike price K and expiration time T. Suppose that the present value of any dividends paid by the stock during the period in question is d_0. Then assuming that no arbitrage occurs, we must have*

$$C - P = S_0 - Ke^{-rT} - d_0$$

where r is the risk-free interest rate. This formula is called the **put-call option parity formula**. □

The American Case

The case of American options is more complicated. Here the price difference $C - P$ is not a constant as it was in the European case.

Consider the following strategy: We go long one put, short one call and long one share of stock (to cover the call in case it is exercised). The initial portfolio is thus

1) 1 put @ P
2) -1 call @ C
3) 1 share @ S_0

The initial value of this portfolio is thus

$$\mathcal{V}_0 = P - C + S_0$$

Now, one of two things can happen during the lifetime of the holdings: the call can be exercised against us or it can expire worthless.

If the call is exercised at time t then we give up the stock to cover the call, taking in Ke^{-rt} units of risk-free bond. Our position is thus

$$1 \text{ put}, Ke^{-rt} \text{ bonds}$$

At the final time T, our position will be

1) 1 put @ $(K - S_T)^+$
2) Ke^{-rt} bonds @ e^{rT}
3) $d_0 e^{rT}$ dollars if the call was exercised after the dividend was paid

The final value is thus

$$\mathcal{V}_{T,1} = (K - S_T)^+ + Ke^{r(T-t)} + d_0 e^{rT}\delta$$

where $\delta = 0$ or 1.

If the call is not exercised, we do nothing until the final time, when our holdings are

1) 1 put @ $(K - S_T)^+$
2) -1 call @ 0
3) 1 share @ S_T
4) $d_0 e^{rT}$ dollars

for a final value of

$$\mathcal{V}_{T,2} = (K - S_T)^+ + S_T + d_0 e^{rT} = \max\{S_T, K\} + d_0 e^{rT}$$

Now, an arbitrage situation will occur if the *discounted* final value is greater than the initial value, for this implies that a profit can be made that is guaranteed to be greater than that of the risk-free bond. (Why?)

The discounted final values under the two scenarios are

$$\overline{\mathcal{V}}_{T,1} = (K - S_T)^+ e^{-rT} + Ke^{-rt} + d_0\delta \geq Ke^{-rT}$$
$$\overline{\mathcal{V}}_{T,2} = \max\{S_T e^{-rT}, Ke^{-rT}\} + d_0 \geq Ke^{-rT}$$

Since equality may hold above, the best we can say is that arbitrage will occur if

$$Ke^{-rT} > P - C + S_0$$

Thus, to avoid arbitrage, we must have

$$P - C \geq Ke^{-rT} - S_0$$

Now let us consider the opposite strategy: We go short one put, long one call and short one share of stock. The initial portfolio is thus

1) -1 put @ P

2) 1 call @ C
3) -1 share @ S_0

The initial value of this portfolio is thus

$$\mathcal{V}_0 = -P + C - S_0$$

One of two things can happen during the lifetime of this portfolio: the put can be exercised against us or it can expire worthless.

If the put is exercised at time t then we must buy the share for K, but at least we can use that share to cover the short position on the stock. This results in the following position

$$1 \text{ call}, -Ke^{-rt} \text{ bonds}$$

At the final time T, our position will be

1) 1 call @ $(S_T - K)^+$
2) $-Ke^{-rt}$ bonds @ e^{rT}
3) $d_0 e^{rT}$ dollars if the put was exercised before the dividend was paid
4) $-d_0 e^{rt}$ dollars if the put was exercised after the dividend was paid

The final value is thus

$$\begin{aligned}\mathcal{V}_{T,1} &= (S_T - K)^+ - Ke^{r(T-t)} + d_0 e^{rT}\delta - d_0 e^{rt}(1-\delta) \\ &= (S_T - K)^+ - Ke^{r(T-t)} + d_0(e^{rT} + e^{rt})\delta - d_0 e^{rt}\end{aligned}$$

where $\delta = 0$ if the put was exercised after the dividend was paid and $\delta = 1$ if the put was exercised before the dividend was paid.

If the put is not exercised, we do nothing until the final time, when our holdings are

1) -1 put @ 0
2) 1 call @ $(S_T - K)^+$
3) -1 share @ S_T with dividends

for a final value of

$$\mathcal{V}_{T,2} = (S_T - K)^+ - S_T - d_0 e^{rT} = \max\{-S_T, -K\} - d_0 e^{rT}$$

As before, an arbitrage situation will occur if the *discounted* final value

is greater than the initial value, for this implies that a profit can be made that is guaranteed to be greater than that of the risk-free bond. (Why?)

The discounted final values under the two scenarios are

$$\begin{aligned}\overline{\mathcal{V}}_{T,1} &= (S_T - K)^+ e^{-rT} - Ke^{-rt} + d_0(1 + e^{-r(T-t)})\delta - d_0 e^{-r(T-t)} \\ &\geq -K - d_0 \\ \overline{\mathcal{V}}_{T,2} &= \max\{-S_T e^{-rT}, -Ke^{-rT}\} - d_0 \geq -K - d_0\end{aligned}$$

Hence arbitrage will occur if

$$-K - d_0 > -P + C - S_0$$

or

$$K + d_0 < P - C + S_0$$

Thus, to avoid arbitrage, we must have

$$P - C \leq K + d_0 - S_0$$

Let us summarize.

Theorem 3 *(American Options With Dividends) Suppose that a stock is currently selling at a price of S_0 per share, an American put on this stock sells for P dollars and an American call for C dollars, both having the same strike price K and expiration time T. Suppose that the present value of any dividends paid by the stock during the period in question is d_0. Then assuming that no arbitrage occurs, we must have*

$$S_0 - K - d_0 \leq C - P \leq S_0 - Ke^{-rT}$$

where r is the risk-free interest rate. Thus, in the American case, the difference $C - P$ can be no larger than in the European case, but it can be smaller.□

Option Prices

Simple arbitrage arguments, along with some common sense, can give us some information about option prices. For instance, since an American option provides all of the features of a corresponding European option and more, it seems obvious that American options should not be less expensive than their European counterparts. In symbols,

$$C^A \geq C^E,\ P^A \geq P^E$$

We leave it to the reader to produce an arbitrage argument to support these inequalitites.

It is not hard to see that the price of an American put can exceed the price of its European counterpart. The idea is that early exercise of the American put can turn a share of stock into bonds that earn the risk-free rate r. If that rate is sufficiently high, the profit can be higher than that of the European put, which is limited by the strike price, since the best case scenario for the owner of a European put is when the stock price is 0 at time T.

More specifically, suppose that the bond rate is r and a share of stock is selling for S_0. Consider an American put with strike price K. The maximum profit from a similar European put is K, which happens if the stock price drops to 0 at time T. On the other hand, suppose we exercise the American put at time 0, and invest the resulting $K - S_0$ dollars at rate r. The resulting profit is $(K - S_0)(1 + r)$. Hence, if

$$(K - S_0)(1 + r) > K$$

then the American put is more valuable than its European counterpart.

As a numerical example, there was a time when the bond rate was 12% (and even higher). Suppose a share of stock is selling for $\$5$. Consider an American put with strike price 50. The resulting profit is

$$(K - S_0)(1 + r) = 45(1.12) = 50.40 > 50$$

Thus, the American 50 put is worth more than the European 50 put.

On the other hand, it is a perhaps somewhat surprising result that an American call is worth exactly the same as a European call with the same terms. That is,

$$C^A = C^E$$

(We are assuming that the stock does not pay a dividend.) However, some reflection reveals the reason. Namely, the ownership of a European call implies that the owner can borrow a share of stock at any time and can use the call to cover the short position at time T for *at most* the strike price K. This provides protection against early exercise of the American call.

To be specific, suppose that $C^A > C^E$ and consider the following initial portfolio

1) -1 American call @ C^A
2) 1 European call @ C^E
3) $C^A - C^E > 0$ bonds @ 1

This portfolio has initial value 0. As mentioned, the ownership of the European call protects us against exercise of the American call since we can always borrow a share of stock to cover early exercise of the American call.

In particular, clearly, if the calls are never exercised then there is a guaranteed profit from the risk-free position. If the American call is exercised at time T then we can also exercise the European call in response, again resulting in a net profit from the bonds. Finally, if the American call is exercised at time $t < T$ then we borrow one share of stock and cover the call. At that time, the portfolio is

1) -1 share @ S_t
2) 1 European call @ ?
3) $C^A - C^E > 0$ bonds @ e^{rt}
4) K dollars

At time T this becomes

1) -1 share @ S_T
2) 1 European call @ $(K - S_T)^+$
3) $C^A - C^E > 0$ bonds @ e^{rT}
4) $Ke^{r(T-t)}$ dollars

Now we simply exercise the call to cover the short stock position or, if the stock price has fallen below the stock price S_T, buy the stock on the open market. In this way, we cover the short stock position at a cost of $\min\{K, S_T\}$. The final profit is thus

$$(C^A - C^E)e^{rT} + Ke^{r(T-t)} - \min\{K, S_T\} > 0$$

The essence of this inequality is that $Ke^{r(T-t)} > K$. In words, it is better to pay the strike price at the end than anywhere in the middle.

There is one more lesson to be learned here. Namely, it is never wise to exercise an American call early *if the intention is to keep the stock*. The alternative of borrowing the stock still gives the investor possession of the stock but at no immediate cost. The cost K can be deferred to the final time T by exercising the option at that time (or buying the stock on the open market if it is cheaper) to cover the short position on the stock. Of course, if the intention is to exercise the call and sell the stock, then this may be more profitable than holding the call until the end.

Theorem 4 *Assume that the underlying stock does not pay a dividend. For an American and European call under the same terms, we have*

$$C^A = C^E$$

For an American and European put under the same terms, we have

$$P^A \geq P^E$$

with strict inequality possible. Moreover, it is never wise to exercise an American call before the expiration date if the intention is to keep the stock. □

Exercises

If the underlying asset of a forward contract provides a dollar income during the life of the contract, then the long investor in the contact will lose out on this income and the cash-and-carry investor will get the income. This effects the previous no-arbitrage argument. The following exercises are *á propos* to this situation.

1. Suppose that the income from the underlying asset has present value I. What are the payoffs in this case? Assume that the annual interest rate is r compounded continuously.
2. What are the payoffs for the two strategies in this case?
3. Show that the assumption of no arbitrage implies that

$$F_{0,T} = (S_0 - I)e^{rT}$$

We have seen that in the simplest case of a forward contract that does not produce an income, the non-arbitrage forward price at time 0 is

$$F_{0,T} = S_0 e^{rT}$$

However, we derived this formula under the very idealistic assumption of a perfect market. Let us examine what happens if this restriction is lifted. In particular, suppose that the lending and borrowing rates are

different, as is almost always the case in real life. Let the lending rate for the investor be r_ℓ and the borrowing rate be r_b. Of course, life being what it is for individual investors, we have $r_\ell < r_b$.

4. Under these conditions, what are the payoffs? Assume that the annual interest rate is r compounded continuously.
5. What are the payoffs for the two strategies in this case?
6. Show that the assumption of no arbitrage implies that

$$S_0 e^{r_\ell T} \leq F_{0,T} \leq S_0 e^{r_b T}$$

Hint: To avoid arbitrage, both strategies must yield a *nonpositive* payoff.

The upper and lower bounds given in Exercise 6 are called **no-arbitrage bounds** and the range of values of the futures price that is implied by the absence of arbitrage is the **no-arbitrage spread**. Thus, in the absence of a perfect market, the lack of arbitrage implies that the futures price can lie anywhere within a *range* of values.

Upper Bounds for Option Prices

7. Prove by an arbitrage argument that the initial value of a European or American call is less than the initial price of the stock, that is

$$C^E \leq S_0$$
$$C^A \leq S_0$$

7. Solution If not then buy the share, sell the call and pocket the difference. Use the share to cover the call if and when it is exercised.
8. Prove the following by an arbitrage argument

$$P^E \leq K e^{-rT}$$
$$P^A \leq K$$

8. Solution: If not, sell the put and invest the money. The put cannot cost you more than K in either case and you will have K immediately in the case of the American put or at the end in the case of the European put.

Lower Bounds for Option Prices

9. Prove that

$$S_0 - Ke^{-rT} - d_0 \leq C^E$$
$$Ke^{-rT} + d_0 - S_0 \leq P^E$$

Solution: Use put-call option parity formula.

10. a) Prove that for a nondividend paying stock

$$S_0 - Ke^{-rT} \leq C^A$$
$$K - S_0 \leq P^A$$

b) If the stock pays a dividend whose discounted value is d_0 then

$$\max\{S_0 - Ke^{-rT} - d_0, S_0 - K\} \leq C^A$$
$$\max\{Ke^{-rT} + d_0 - S_0, K - S_0\} \leq P^A$$

Solution: a) First comes from $C^E = C^A$. Second from the fact that $K - S_0$ can be gotten by exercising the put at time t_0.

b) For the first formula, since $C^A = C^E$ we get

$$S_0 - Ke^{-rT} - d_0 \leq C^A$$

In addition, immediate exercise of the American call will return $S_0 - K$ and so the call cannot be purchased for less than this amount. For the second formula, $K - S_0 \leq P^A$ because immediate exercise is worth $K - S_0$. For the other part, consider two portfolios. Portfolio A is 1 put. Portfolio B is $(K + d_0)e^{-rT}$ in cash and a short share of stock.

Chapter 5

Probability II: More Discrete Probability

In this chapter, we cover the material on finite probability spaces that is needed for the discussion of discrete-time models in the next chapter.

Conditional Probability

When additional information is available about an experiment, the notion of conditional probability can be used to take that information into account. The idea is to "concentrate" all of the probability of Ω onto the set E, in a manner that is proportional to the original probability measure $\mathbb{P}$.

Definition *Let $(\Omega, \mathbb{P})$ be a probability space. Let E be an event with $\mathbb{P}(E) > 0$. Then for any event A, the* **conditional probability of A given** *E is*

$$\mathbb{P}(A \mid E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)} \qquad \square$$

The symbol $\mathbb{P}(A \mid E)$ is read "the probability of A given E." Note that we do not need to worry about the case $\mathbb{P}(E) = 0$, for it makes little sense to ask about a probability conditioned upon the occurrence of an impossible event.

Conditioning on an event allows us to define a new "conditional" probability measure on Ω.

Theorem 1 *Let $(\Omega, \mathbb{P})$ be a finite probability space and let E be an event for which $\mathbb{P}(E) > 0$. Then the set function $\mathbb{P}_E$ defined by*

$$\mathbb{P}_E(A) = \mathbb{P}(A \mid E)$$

is a probability measure on Ω *for which $\mathbb{P}_E(E) = 1$.*
Proof. To show that $\mathbb{P}_E$ is a probability measure on Ω we must verify a few facts. First, monotonicity of $\mathbb{P}$ implies that

$$0 \leq \mathbb{P}(A \cap E) \leq \mathbb{P}(E)$$

and so $0 \leq \mathbb{P}(A \mid E) \leq 1$, that is

$$0 \leq \mathbb{P}_E(A) \leq 1$$

Also

$$\mathbb{P}_E(\Omega) = \mathbb{P}(\Omega \mid E) = \frac{\mathbb{P}(\Omega \cap E)}{\mathbb{P}(E)} = \frac{\mathbb{P}(E)}{\mathbb{P}(E)} = 1$$

Finally, if $A \cap B = \emptyset$ then since $A \cap E$ and $B \cap E$ are also disjoint, we have

$$\begin{aligned}
\mathbb{P}_E(A \cup B) &= \mathbb{P}(A \cup B \mid E) \\
&= \frac{\mathbb{P}((A \cup B) \cap E)}{\mathbb{P}(E)} \\
&= \frac{\mathbb{P}((A \cap E) \cup (B \cap E))}{\mathbb{P}(E)} \\
&= \frac{\mathbb{P}(A \cap E) + \mathbb{P}(B \cap E)}{\mathbb{P}(E)} \\
&= \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)} + \frac{\mathbb{P}(B \cap E)}{\mathbb{P}(E)} \\
&= \mathbb{P}(A \mid E) + \mathbb{P}(B \mid E) \\
&= \mathbb{P}_E(A) + \mathbb{P}_E(B)
\end{aligned}$$

This completes the proof. □

The theorem on total probabilities takes on a nice form using conditional probabilities.

Theorem 2 *(***Theorem on Total Probabilities***) Let Ω be a sample space and let $E_1, \ldots, E_n$ form a partition of Ω. Provided that $\mathbb{P}(E_k) \neq 0$ for all k, we have for any event A in Ω,*

$$\mathbb{P}(A) = \sum_{k=l}^{n} \mathbb{P}(A \mid E_k)\mathbb{P}(E_k) \qquad \square$$

Partitions and Measurability

For convenience, let us repeat the definition of a partition.

Definition *Let Ω be a nonempty set. Then a* **partition** *of Ω is a collection $\mathcal{P} = \{B_1, \ldots, B_k\}$ of* nonempty *subsets of Ω, called the* **blocks** *of the partition, with the following properties*

1) The blocks are pairwise disjoint

$$B_i \cap B_j = \emptyset$$

2) The union of the blocks is all of Ω

$$B_1 \cup \cdots \cup B_k = \Omega \qquad \square$$

Figure 1 shows a partition of a set Ω.

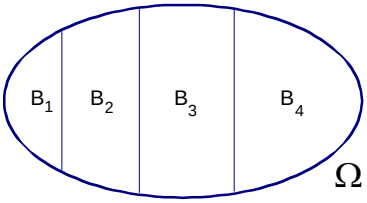


Figure 1 - A Partition of Ω

We will also have use for the notion of a refinement of a partition.

Definition *Let* $\mathcal{P} = \{B_1, \ldots, B_k\}$ *be a partition of a set* Ω*. Then a partition* $\mathcal{Q} = \{C_1, \ldots, C_n\}$ *that comes from* $\mathcal{P}$ *by breaking up some of the blocks* B_i *into smaller blocks is called a* **refinement** *of* $\mathcal{P}$*. Thus,* $\mathcal{Q}$ *is a refinement of* $\mathcal{P}$ *if each block of* $\mathcal{Q}$ *is contained in a some block of* $\mathcal{P}$ *or, equivalently, each block of* $\mathcal{P}$ *is a union of blocks of* $\mathcal{Q}$*. We denote this by* $\mathcal{P} \prec \mathcal{Q}$*.* $\square$

Note that when we say that a set A is the union of sets in the collection $\{B_1, \ldots, B_n\}$ this includes the possibility that A is the "union" of a single set B_k, that is, $A = B_k$.

Figure 2 shows a refinement of the partition in Figure 1. Note that

$$\begin{aligned} B_1 &= C_1 \cup C_2 \\ B_2 &= C_3 \cup C_4 \cup C_5 \\ B_3 &= C_6 \\ B_4 &= C_7 \end{aligned}$$

and so each block B_i is the union of blocks C_j.

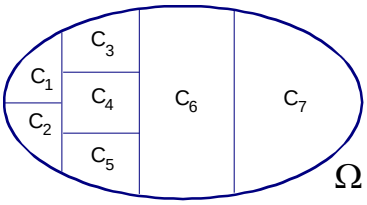


Figure 2 - A Refinement

The Partition Defined by a Random Variable

If X is a random variable it is customary to denote the inverse image of a set B under X not by $X^{-1}(B)$, as for ordinary functions, but instead by

$$\{X \in B\}$$

Also, instead of writing $X^{-1}(x)$ it is customary to write

$$\{X = x\}$$

We also remind the reader that the set of *distinct* values $\{x_1, \ldots, x_n\}$ of X is called the **image** of X and is denoted by $\text{im}(X)$.

Any random variable X on a finite sample space Ω defines a partition $\mathcal{P}_X$ of Ω, as shown in Figure 3.

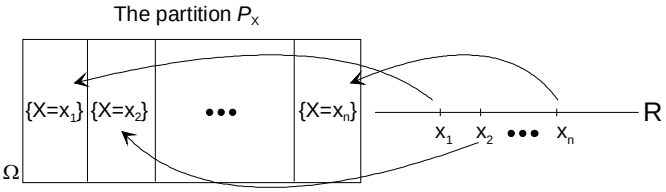


Figure 3–The partition defined by a random variable

Definition *Let X be a random variable on Ω with*

$$im(X) = \{x_1, \ldots, x_n\}$$

Then X defines a partition of Ω whose blocks are the inverse images of the elements *of $im(X)$, that is*

$$\mathcal{P}_X = \{\{X = x\} \mid x \in im(X)\} = \{\{X = x_1\}, \ldots, \{X = x_n\}\}$$

This is called the ***partition defined by*** *X.*□

Measurability Of a Random Variable with Respect to a Partition

The partition $\mathcal{P}_X$ defined by a random variable has one very important property: X is *constant* on the blocks $\{X = x\}$ of $\mathcal{P}_X$. In fact, at the risk of being redundant, X takes the constant value x on $\{X = x\}$. This property is expressed by saying that X is $\mathcal{P}_X$-**measurable**.

Definition *Let $\mathcal{P}$ be any partition of Ω. A random variable X on Ω is said to be $\mathcal{P}$-***measurable** *if X is constant on each block of $\mathcal{P}$.*□

There is another rather obvious property of X with respect to $\mathcal{P}_X$, namely, not only is X constant on the blocks of $\mathcal{P}_X$, but it is a *different* constant on each block of $\mathcal{P}_X$.

Now, given a nonconstant random variable, there are many partitions $\mathcal{Q}$ for which X is constant on each block of $\mathcal{Q}$, that is, for which X is $\mathcal{Q}$-measurable. However, $\mathcal{P}_X$ is the only partition for which X is a *different* constant on each block. We can characterize all partitions $\mathcal{Q}$ for which X is $\mathcal{Q}$-measurable quite simply. (See Figure 4.)

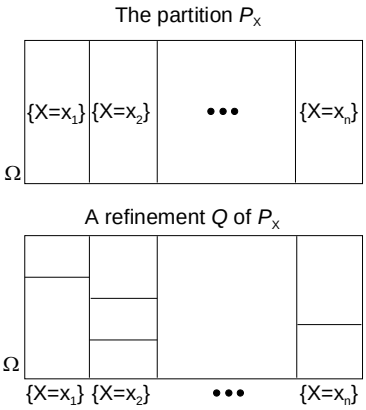


Figure 4–A refinement of $\mathcal{P}_X$

Theorem 3 *Let X be a random variable on Ω.*

1) *Then X is $\mathcal{Q}$-measurable if and only if $\mathcal{Q}$ is a refinement of $\mathcal{P}_X$.*
2) *$\mathcal{P}_X$ is the coarsest partition for which X is measurable and the only partition for which X is measurable and takes on a* different *constant value on each block.*

Proof. For 1), if X is $\mathcal{Q}$-measurable and $\mathcal{Q} = \{B_1, \ldots, B_k\}$ then for any $\omega \in B_i$ we have

$$B_i \subseteq \{X = X(\omega)\}$$

and so $\mathcal{Q}$ is a refinement of $\mathcal{P}_X$. The converse is clear. We leave proof of part 2) to the reader.□

The following very important theorem states that there is a very strong connection between a random variable X and any other random variable Y that is $\mathcal{P}_X$-measurable.

Theorem 4 *Let X and Y be random variables. Then Y is $\mathcal{P}_X$-measurable if and only if Y is a function of X, that is, if and only if there is a function $f\colon \mathbb{R} \to \mathbb{R}$ for which*

$$Y = f(X)$$

Proof. We know that Y is constant on the blocks of the partition $\mathcal{P}_X = \{B_1, \ldots, B_k\}$. Let us assume that

$$Y(B_i) = \{y_i\}$$

Of course, X is also constant on the blocks of $\mathcal{P}_X$, so let

$$X(B_i) = \{x_i\}$$

Define f by setting

$$f(x_i) = y_i$$

Then for $\omega \in B_i$

$$f(X)(\omega) = f(X(\omega)) = f(x_i) = y_i = Y(\omega)$$

and so $Y = f(X)$, as desired. The converse is much easier and we leave it as an exercise.□

Partitions and Independence

Let us take another brief look at the notion of independence. Here again is the definition.

Definition *The events E and F of $(\Omega, \mathbb{P})$ are* **independent** *if*

$$\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$$

The events $E_1, \ldots, E_k$ are independent if for any subcollection

$E_{i_1}, \ldots, E_{i_m}$ of these events

$$\mathbb{P}(E_{i_1} \cap \cdots \cap E_{i_m}) = \mathbb{P}(E_{i_1}) \cdots \mathbb{P}(E_{i_m}) \qquad \square$$

We can extend the definition of independence to families of collections of events.

Definition *The collections $\mathcal{C}_1, \ldots, \mathcal{C}_k$ of events is* **independent** *if for any choice of events $E_i \in \mathcal{C}_i$ the events $E_1, \ldots, E_k$ are independent.* $\square$

We will have reason to apply this definition when the collections are partitions of Ω. In fact, let us recall that the random variables $X_1, \ldots, X_n$ are independent if

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i)$$

for all $x_1, \ldots, x_n \in \mathbb{R}$ $\square$

This definition can be reformulated in terms of partitions as follows.

Theorem 5 *The random variables $X_1, \ldots, X_n$ are independent if and only if the partitions $\mathcal{P}(X_1), \ldots, \mathcal{P}(X_n)$ are independent collections of events.*
Proof. This follows immediately from the fact that the blocks of $\mathcal{P}(X_i)$ are precisely the sets $\{X_i = x_i\}$.$\square$

Algebras

We have seen that for a finite sample space, partitions are intimately connected with random variables. As it happens, the notion of a partition does not generalize readily for nonfinite sample spaces. For this, we need another concept called an *algebra*.

We will not use the topics discussed in this section directly, because partitions are sufficient for our analysis of discrete time pricing models, and they are a bit more intuitive than algebras. However, we do want to discuss algebras here because they provide a helpful bridge between the intuitive notion of partition and the notion of σ-algebra, which we will need for our analysis of continuous-time pricing models and the Black-Scholes option pricing formula.

Definition *Let Ω be a nonempty set. A collection $\mathcal{A}$ of subsets of Ω is called an* ***algebra of sets*** *(or just an* ***algebra****) if it satisfies the following properties*

1) (Empty set is in $\mathcal{A}$)

$$\emptyset \in \mathcal{A}$$

2) ($\mathcal{A}$ is closed under complements)

$$A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$$

3) ($\mathcal{A}$ is closed under unions)

$$A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$$ □

It is not hard to show that any algebra of sets is closed under intersections and differences as well, that is

$$A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$$
$$A, B \in \mathcal{A} \Rightarrow A \setminus B \in \mathcal{A}$$

The following concept will be very useful. It makes precise the notion of the "smallest" nonempty sets in an algebra $\mathcal{A}$.

Definition *Let $\mathcal{A}$ be an algebra of sets on Ω. An* **atom** *of $\mathcal{A}$ is a nonempty set $A \in \mathcal{A}$ with the property that no nonempty* proper *subset of A is also in $\mathcal{A}$.*□

Partitions and Algebras

Starting with a partition $\mathcal{P}$ of Ω we can generate an algebra $\mathcal{A}(\mathcal{P})$ of sets simply by taking all possible finite unions of the blocks of $\mathcal{P}$. The reverse is also possible: starting with an algebra of sets on a *finite* sample space, we can get a partition.

Theorem 6 *Let Ω be a nonempty finite set.*

1) For any partition $\mathcal{P}$ of Ω the set

$$\mathcal{A}(\mathcal{P}) = \{C \subseteq \Omega \mid C = \emptyset \text{ or } C = \text{union of blocks of } \mathcal{P}\}$$

is an algebra, called the ***algebra generated by*** *$\mathcal{P}$.*

2) If $\mathcal{A}$ is an algebra on Ω then the set of all atoms of $\mathcal{A}$

$$\mathcal{P}(\mathcal{A}) = \{\text{all atoms of } \mathcal{A}\}$$

is a partition of Ω, called the **partition defined by** *$\mathcal{A}$.*

Proof. We prove part 2), leaving part 1) as an exercise. Let

$$\mathcal{P} = \{A_1, \ldots, A_k\}$$

be a complete list of distinct atoms of $\mathcal{A}$. We must show that $\mathcal{P}$ is a partition of Ω. By definition, atoms are nonempty. If $A_i \cap A_j \neq \emptyset$ for $i \neq j$ then $A_i \cap A_j$ would be an element of $\mathcal{A}$ that was nonempty and a proper subset of A_i, which is not possible. Hence, the atoms are pairwise disjoint. Finally, suppose that $\omega \in \Omega$. The the intersection I of all elements of $\mathcal{A}$ containing ω is an element of $\mathcal{A}$ that contains ω. Moreover, I is nonempty and no proper subset of I is in $\mathcal{A}$, for any such proper subset would have been part of the intersection that defined I. Hence, I is actually an atom of $\mathcal{A}$. This shows that every element of Ω is contained in some atom of $\mathcal{A}$ and so the union of the atoms of $\mathcal{A}$ is all of Ω. Thus, $\mathcal{P}$ is a partition of $\mathcal{A}$.□

The main theme of our current discussion is that for finite sample spaces, the notions of partition and algebra are in some sense equivalent

$$\text{Partitions of } \Omega \Leftrightarrow \text{Algebras on } \Omega$$

By this we mean that, while these concepts are certainly not the same, all statements made about partitions have an analog for algebras and vice-versa. Put another way, whatever theory we can develop in the context of partitions could just as well have been developed in the context of algebras and vice-versa.

The precise connection between the two concepts is made by the correspondences

$$\mathcal{Q} \rightarrow \mathcal{A}(\mathcal{Q}) \text{ [Partition to algebra generated by partition]}$$
$$\mathcal{B} \rightarrow \mathcal{P}(\mathcal{B}) \text{ [Algebra to partition defined by algebra]}$$

described in Theorem 6. The first correspondence takes any partition of Ω and produces an algebra and the second takes any algebra and produces a partition. It is a fact that these correspondences are inverses of each other (and are therefore one-to-one).

To see this, suppose that $\mathcal{Q}$ is a partition of Ω. The algebra $\mathcal{A}(\mathcal{Q})$ is the set of all unions of blocks of $\mathcal{Q}$. Hence, the blocks of $\mathcal{Q}$ are precisely the atoms of the algebra $\mathcal{A}(\mathcal{Q})$, that is,

$$\mathcal{P}(\mathcal{A}(\mathcal{Q})) = \mathcal{Q}$$

Similarly, if we start with an algebra $\mathcal{B}$ of Ω, then the partition $\mathcal{P}(\mathcal{B})$ is

the set of atoms of $\mathcal{B}$. But all elements of $\mathcal{B}$ are unions of atoms of $\mathcal{B}$ and so $\mathcal{A}(\mathcal{P}(\mathcal{B}))$ is the same as $\mathcal{B}$, that is

$$\mathcal{A}(\mathcal{P}(\mathcal{B})) = \mathcal{B}$$

This shows that the two correspondences are one-to-one and are inverses of each other. This is a very tight connection indeed between the two concepts.

The next theorem strengthens the connection between partitions and algebras. It says that the concept of refinement of partitions corresponds to set inclusion of algebras. Notice that there are two statements in the theorem. These statements say exactly the same thing: one from the point of view of partitions and the other from the point of view of algebras. Recall that we denote the fact that $\mathcal{Q}$ is a refinement of $\mathcal{P}$ by $\mathcal{P} \prec \mathcal{Q}$.

Theorem 7

1) Let $\mathcal{P}$ and $\mathcal{Q}$ be partitions of Ω. Then

$$\mathcal{A}(\mathcal{P}) \subseteq \mathcal{A}(\mathcal{Q}) \Leftrightarrow \mathcal{P} \prec \mathcal{Q}$$

2) Let $\mathcal{A}$ and $\mathcal{B}$ be algebras on Ω. Then

$$\mathcal{A} \subseteq \mathcal{B} \Leftrightarrow \mathcal{P}(\mathcal{A}) \prec \mathcal{P}(\mathcal{B})$$

Proof. We only need to prove statement 1). (Why?) Suppose that $\mathcal{A}(\mathcal{P}) \subseteq \mathcal{A}(\mathcal{Q})$. Then the blocks of $\mathcal{P}$ are the atoms in $\mathcal{A}(\mathcal{P})$. If A is an atom of $\mathcal{P}$ then it is also in $\mathcal{A}(\mathcal{Q})$ and so it is the union of blocks of $\mathcal{Q}$. In other words, each block of $\mathcal{P}$ is the union of blocks of $\mathcal{Q}$ and so $\mathcal{Q}$ is a refinement of $\mathcal{P}$.

Conversely, suppose that $\mathcal{Q}$ is a refinement of $\mathcal{P}$. Then any block of $\mathcal{P}$ is the union of blocks of $\mathcal{Q}$. It follows that any element of $\mathcal{A}(\mathcal{P})$, being the union of blocks of $\mathcal{P}$, is also the union of blocks of $\mathcal{Q}$ and so belongs to $\mathcal{A}(\mathcal{Q})$. Thus $\mathcal{A}(\mathcal{P}) \subseteq \mathcal{A}(\mathcal{Q})$.□

The Algebra Generated by a Random Variable

We have seen the strong connection between partitions of Ω and algebras on Ω. It is now time to bring random variables into the picture.

Just as a random variable X defines a partition $\mathcal{P}_X$ of Ω

$$\mathcal{P}_X = \{\{X = x\} \mid x \in \text{im}(X)\}$$

consisting of the inverse images of the *elements* of $\text{im}(X)$, the random

variable also defines an algebra $\mathcal{A}_X$ on Ω consisting of the inverse images of the *subsets* of $\text{im}(X)$.

Definition *Let X be a random variable on Ω. Then X defines an algebra on Ω whose elements are the inverse images of the* subsets *of $im(X)$, that is*

$$\mathcal{A}_X = \{\{X \in B\} \mid B \subseteq im(X)\} \qquad \square$$

It is easy to see that $\mathcal{P}_X$ and $\mathcal{A}_X$ are connected (see Figure 5).

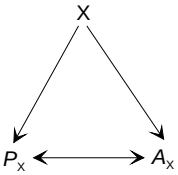


Figure 5–The partition and algebra generated by a random variable

In fact, $\mathcal{A}_X$ is nothing more than the algebra generated by $\mathcal{P}_X$, in symbols

$$\mathcal{A}_X = \mathcal{A}(\mathcal{P}_X)$$

To see this, note that if $B = \{b_1, \ldots, b_m\}$ is a subset of $\text{im}(X)$ then

$$\{X \in B\} = \bigcup_{i=1}^{m} \{X = b_i\} \in \mathcal{A}(\mathcal{P}_X)$$

and so $\mathcal{A}_X \subseteq \mathcal{A}(\mathcal{P}_X)$. But $\mathcal{A}(\mathcal{P}_X)$ is the smallest algebra that contains the blocks of $\mathcal{P}_X$, that is, the sets $\{X = b_i\}$. Hence, $\mathcal{A}_X = \mathcal{A}(\mathcal{P}_X)$.

Theorem 8 *Let X be a random variable on a finite sample space Ω. Then the algebra generated by X is the algebra generated by $\mathcal{P}_X$, in symbols*

$$\mathcal{A}_X = \mathcal{A}(\mathcal{P}_X)$$

and the partition defined by X is the partition defined by $\mathcal{A}_X$, in symbols

$$\mathcal{P}_X = \mathcal{P}(\mathcal{A}_X) \qquad \square$$

Measurability Of a Random Variable with Respect to an Algebra

We have defined measurability of a random variable X with respect to a partition $\mathcal{P}$ of Ω. This simply means that X is constant on the blocks of

$\mathcal{P}$. We have also seen that X is $\mathcal{P}$-measurable if and only if $\mathcal{P}$ is a refinement of $\mathcal{P}_X$. Now let us turn to measurability of X with respect to an algebra of sets on Ω.

It should come as no surprise that we want this concept to be defined so that a random variable X is $\mathcal{A}(\mathcal{P})$-measurable if and only if it is $\mathcal{P}$-measurable, that is, constant on the blocks of $\mathcal{P}$. In fact, since *any* algebra $\mathcal{A}$ is generated by a partition

$$\mathcal{A} = \mathcal{A}(\mathcal{P}(\mathcal{A}))$$

we can actually use this as the *defining property* of measurability with respect to an algebra. This definition reads as follows: If $\mathcal{A}$ is an algebra on Ω then X is $\mathcal{A}$-measurable if X is constant on all of the atoms of $\mathcal{A}$.

While this definition is quite intuitive, it is not standard and we would be doing the reader a disservice by adopting it. To understand the usual definition, note that the following are equivalent

1) X is $\mathcal{P}$-measurable
2) X is constant on the blocks of $\mathcal{P}$
3) X is constant on the atoms of $\mathcal{A}(\mathcal{P})$
4) Each set $\{X = x\}$ is the union of atoms of $\mathcal{A}(\mathcal{P})$
5) Each set $\{X = x\}$ is in $\mathcal{A}(\mathcal{P})$
6) For any subset $B \subseteq \text{im}(X)$, the set $\{X \in B\}$ is in $\mathcal{A}(\mathcal{P})$

Now we are ready for the standard definition, which we have just shown is equivalent to the previous intuitive definition.

Definition *Let X be a random variable on a finite sample space Ω. Let $\mathcal{A}$ be any algebra of sets on Ω. Then X is* **$\mathcal{A}$-measurable** *if*

$$\{X = B\} \in \mathcal{A}, \text{ for all } B \subseteq im(X) \qquad \square$$

We have shown that X is $\mathcal{P}$-measurable if and only if X is $\mathcal{A}(\mathcal{P})$-measurable.

Theorem 9 *Let X be a random variable on Ω.*

1) If $\mathcal{A}$ is an algebra on Ω then X is $\mathcal{A}$-measurable if and only if it is $\mathcal{P}(\mathcal{A})$-measurable.
2) If $\mathcal{P}$ is a partition of Ω then X is $\mathcal{P}$-measurable if and only if it is $\mathcal{A}(\mathcal{P})$-measurable.$\square$

Just as we have characterized measurability with respect to a partition by showing that X is $\mathcal{P}$-measurable if and only if $\mathcal{P}_X \prec \mathcal{P}$, we can characterize measurability with respect to an algebra.

Theorem 10 *A random variable X on Ω is measurable with respect to an algebra $\mathcal{A}$ if and only if $\mathcal{A}_X \subseteq \mathcal{A}$.*
Proof. The following statements are equivalent and prove the theorem
1) X is $\mathcal{A}$-measurable
2) X is $\mathcal{P}(\mathcal{A})$-measurable
3) $\mathcal{P}(\mathcal{A})$ is a refinement of $\mathcal{P}_X$
4) $\mathcal{A}(\mathcal{P}_X)$ is contained in $\mathcal{A}(\mathcal{P}(\mathcal{A}))$
5) $\mathcal{A}_X$ is contained in $\mathcal{A}$ □

Conditional Expectation

We can put together the notions of conditional probability and expectation to get *conditional expectation*, which plays a key role in derivative pricing models.

Conditional Expectation with Respect to an Event

Conditional expectation with respect to an event A with positive probability is pretty straightforward—we just take the ordinary expectation but with respect to the conditional probability measure $\mathbb{P}_A$ defined by

$$\mathbb{P}_A(B) = \mathbb{P}(B \mid A)$$

Definition Let $(\Omega, \mathbb{P})$ be a finite probability space and let A be an event for which $\mathbb{P}(A) > 0$. The **conditional expectation** of a random variable X with respect to the event A is

$$\mathcal{E}_{\mathbb{P}}(X \mid A) = \mathcal{E}_{\mathbb{P}_A}(X) \qquad \square$$

The symbol $\mathcal{E}_{\mathbb{P}}(X \mid A)$ is read "the expected value of X given A."

A little algebra gives another useful expression for the conditional expectation in terms of the nonconditional expectation.

Theorem 11 *Let $(\Omega, \mathbb{P})$ be a finite probability space and let A be an event for which $\mathbb{P}(A) > 0$. The conditional expectation of a random variable X with respect to the event A is*

$$\mathcal{E}_{\mathbb{P}}(X \mid A) = \frac{\mathcal{E}_{\mathbb{P}}(X1_A)}{\mathbb{P}(A)}$$

where 1_A *is the indicator function of* A.
Proof. We have

$$\begin{aligned}
\mathcal{E}_{\mathbb{P}_A}(X) &= \sum_{i=1}^{n} X(\omega_i)\mathbb{P}(\omega_i \mid A) \\
&= \sum_{i=1}^{n} X(\omega_i)\frac{\mathbb{P}(\omega_i \cap A)}{\mathbb{P}(A)} \\
&= \frac{1}{\mathbb{P}(A)}\sum_{i=1}^{n} X(\omega_i)\mathbb{P}(\omega_i \cap A) \\
&= \frac{1}{\mathbb{P}(A)}\sum_{i=1}^{n} X(\omega_i)1_A(\omega_i)\mathbb{P}(\omega_i) \\
&= \frac{1}{\mathbb{P}(A)}\mathcal{E}_{\mathbb{P}}(X1_A)
\end{aligned}$$

as desired. □

One simple consequence of the previous theorem is the following useful result.

Theorem 12 *If* A *and* B *are events with* $\mathbb{P}(A \cap B) > 0$ *then*

$$\mathcal{E}_{\mathbb{P}_A}(X \mid B) = \mathcal{E}(X \mid A \cap B)$$

Proof. Using the previous theorem, we have

$$\begin{aligned}
\mathcal{E}_{\mathbb{P}_A}(X \mid B) &= \frac{\mathcal{E}_{\mathbb{P}_A}(X1_B)}{\mathbb{P}_A(B)} \\
&= \frac{\mathcal{E}_{\mathbb{P}}(X1_B1_A)}{\mathbb{P}(A)\mathbb{P}_A(B)} \\
&= \frac{\mathcal{E}_{\mathbb{P}}(X1_{A\cap B})}{\mathbb{P}(A \cap B)} \\
&= \mathcal{E}_{\mathbb{P}}(X \mid A \cap B)
\end{aligned}$$

as desired.□

Next we have the expected value analog of the theorem on total probabilities.

Theorem 13 *Let $\mathcal{P} = \{B_1, \ldots, B_n\}$ be a partition of Ω. Then for any random variable X on Ω*

$$\mathcal{E}(X) = \sum_{i=1}^{n} \mathcal{E}(X \mid B_i)\mathbb{P}(B_i)$$

Moreover, if A is an event of positive probability then

$$\mathcal{E}(X \mid A) = \sum_{i=1}^{n} \mathcal{E}(X \mid B_i \cap A)\mathbb{P}(B_i \mid A)$$

These sums are valid provided that we consider each term in which the conditional probability is not defined as equal to 0. Put another way, if either of the factors in a term is 0 then the term is considered 0, that is

$$\textit{undefined} \cdot 0 = 0$$

Proof. For the first part, suppose that
a) $\mathbb{P}(B_i) > 0$ for $i = 1, \ldots, m$
b) $\mathbb{P}(B_i) = 0$ for $i = m+1, \ldots, n$
Since

$$X = \sum_{i=1}^{m} X1_{B_i} + \sum_{i=m+1}^{n} X1_{B_i}$$

applying expected values gives

$$\begin{aligned}
\mathcal{E}(X) &= \sum_{i=1}^{m} \mathcal{E}(X1_{B_i}) + \sum_{i=m+1}^{n} \mathcal{E}(X1_{B_i}) \\
&= \sum_{i=1}^{m} \frac{\mathcal{E}(X1_{B_i})}{\mathbb{P}(B_i)} \mathbb{P}(B_i) \\
&= \sum_{i=1}^{m} \mathcal{E}(X \mid B_i)\mathbb{P}(B_i)
\end{aligned}$$

which proves the first statement.

For the second statement, we apply the first statement to the conditional probability $\mathbb{P}_A$

$$\mathcal{E}_{\mathbb{P}}(X \mid A) = \mathcal{E}_{\mathbb{P}_A}(X) = \sum_{i=1}^{n} \mathcal{E}_{\mathbb{P}_A}(X \mid B_i)\mathbb{P}_A(B_i)$$

where the undefined terms are 0. But these are the terms for which $\mathcal{E}_{\mathbb{P}_A}(X \mid B_i)$ is not defined, that is, for which $\mathbb{P}_A(B_i) = 0$, or finally $\mathbb{P}(A \cap B_i) = 0$. (Note that $\mathbb{P}_A(B_i)$ is always defined because A is assumed to have positive probability.) For all other terms, we may write

$$\mathcal{E}_{\mathbb{P}_A}(X \mid B_i) = \mathcal{E}_{\mathbb{P}}(X \mid B_i \cap A)$$

to get the desired sum.□

Conditional Expectation with Respect to a Partition

Next we define conditional expectation with respect to a partition of the sample space. Unlike the conditional expectation given an event, which is a real number, the conditional expectation given a partition is a *random variable*.

By way of motivation, let us briefly revisit the ordinary expected value of a random variable. Of course, the expected value $\mathcal{E}(X)$ of a random variable X is a constant. In fact, it represents the *best possible* approximation of X by a constant. The measure that is used to judge the quality of the approximation is the **mean squared error** or **MSE**, defined by

$$\text{MSE} = \mathcal{E}[(X - c)^2]$$

where c is a constant. As it happens, for all constants c, the mean squared error is smallest if and only if c is the expected value μ_X, that is

$$\mathcal{E}[(X - \mu_X)^2] \leq \mathcal{E}[(X - c)^2]$$

with equality if and only if $c = \mu_X$. To prove this, we write

$$\begin{aligned}\mathcal{E}[(X - c)^2] &= \mathcal{E}[\{(X - \mu_X) + (\mu_X - c)\}^2] \\ &= \mathcal{E}[(X - \mu_X)^2] + \mathcal{E}[(X - \mu_X)(\mu_X - c)] + \mathcal{E}[(\mu_X - c)^2]\end{aligned}$$

But the middle term is 0 (why?) and so

$$\mathcal{E}[(X - c)^2] = \mathcal{E}[(X - \mu_X)^2] + \mathcal{E}[(\mu_X - c)^2] \geq \mathcal{E}[(X - \mu_X)^2]$$

with equality holding if and only if $c = \mu_X$.

Now, we want the expected value with respect to a partition $\mathcal{P}$ of Ω to be the best approximation to X that is constant on each block of the partition. The point is that if we are given a block B of the partition, then we can get a better constant approximation to X on B than the ordinary expected value. In fact, we get the *best* constant approximation by using

the conditional expectation $\mathcal{E}(X \mid B)$. This idea is shown in Figure 6 for a random variable X that is defined on an interval $[a, b]$ of the real line. (We chose this illustration because it is easier to picture the conditional expectation for such intervals than for random variables on a finite sample space).

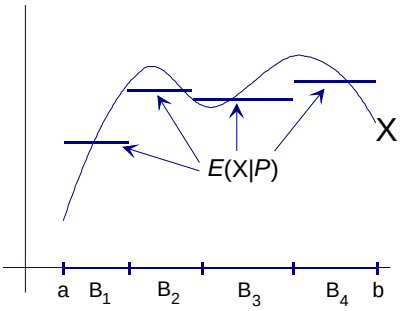


Figure 6 - Conditional Expectation

The definition of $\mathcal{E}_{\mathbb{P}}(X \mid \mathcal{P})$ should now be fairly clear: its value on each block B_i of $\mathcal{P}$ is $\mathcal{E}(X \mid B_i)$. Hence, we can define the random variable $\mathcal{E}_{\mathbb{P}}(X \mid \mathcal{P})$ as a linear combinations of the indicator functions of the blocks of $\mathcal{P}$.

Definition *Let $(\Omega, \mathbb{P})$ be a finite probability space and let $\mathcal{P} = \{B_1, \ldots, B_n\}$ be a partition of Ω for which $\mathbb{P}(B_i) > 0$ for all i. The* **conditional expectation** *of a random variable X with respect to the partition $\mathcal{P}$ is a random variable*

$$\mathcal{E}_{\mathbb{P}}(X \mid \mathcal{P}) \colon \Omega \to \mathbb{R}$$

defined by

$$\mathcal{E}_{\mathbb{P}}(X \mid \mathcal{P}) = \mathcal{E}_{\mathbb{P}}(X \mid B_1)1_{B_1} + \cdots + \mathcal{E}_{\mathbb{P}}(X \mid B_n)1_{B_n}$$

In particular, for any $\omega \in \Omega$

$$\mathcal{E}_{\mathbb{P}}(X \mid \mathcal{P})(\omega) = \mathcal{E}_{\mathbb{P}}(X \mid [\omega]_{\mathcal{P}})$$

where $[\omega]_{\mathcal{P}}$ is the block of $\mathcal{P}$ containing ω.□

Here is a formal statement of the value of the conditional expectation in approximating X.

Theorem 14 *The random variable $\mathcal{E}(X \mid \mathcal{P})$ is the best approximation to X among all functions that are constant on the blocks of $\mathcal{P}$, that is, the*

best approximation to X among all $\mathcal{P}$-measurable random variables, in the sense of mean squared error. By this, we mean that

$$\mathcal{E}[(X - \mathcal{E}(X \mid \mathcal{P}))^2] \leq \mathcal{E}[(X - Y)^2]$$

for all $\mathcal{P}$-measurable random variables Y, with equality holding if and only if $Y = \mathcal{E}(X \mid \mathcal{P})$.

Proof. As an aid to readability let us set

$$\mathcal{E}(X \mid \mathcal{P}) = \mu_{X|\mathcal{P}}$$

We begin by writing for any $\mathcal{P}$-measurable random variable Y

$$\begin{aligned}\mathcal{E}[(X - Y)^2] &= \mathcal{E}[\{(X - \mu_{X|\mathcal{P}}) + (\mu_{X|\mathcal{P}} - Y)\}^2] \\ &= \mathcal{E}[(X - \mu_{X|\mathcal{P}})^2] + \mathcal{E}[(\mu_{X|\mathcal{P}} - Y)^2] \\ &\quad + \mathcal{E}[(X - \mu_{X|\mathcal{P}})(\mu_{X|\mathcal{P}} - Y)]\end{aligned}$$

Now, we want to show that the last term is 0. This can be done by using Theorem 13. Assuming that $\mathcal{P} = \{B_1, \ldots, B_n\}$ we have

$$\mathcal{E}[(X - \mu_{X|\mathcal{P}})(\mu_{X|\mathcal{P}} - Y)] = \sum_{i=1}^{n} \mathcal{E}[(X - \mu_{X|\mathcal{P}})(\mu_{X|\mathcal{P}} - Y) \mid B_i]\mathbb{P}(B_i)$$

Let us now focus on the expressions

$$\mathcal{E}[(X - \mu_{X|\mathcal{P}})(\mu_{X|\mathcal{P}} - Y) \mid B_i] = \frac{1}{\mathbb{P}(B_i)}\mathcal{E}[(X - \mu_{X|\mathcal{P}})(\mu_{X|\mathcal{P}} - Y)1_{B_i}]$$

for the terms with $\mathbb{P}(B_i) > 0$. (The other terms are equal to 0.) Since Y is $\mathcal{P}$-measurable, the random variable $\mu_{X|\mathcal{P}} - Y$ is constant on each block B_i and can be pulled from under the expectation to get

$$\frac{1}{\mathbb{P}(B_i)}(\mu_{X|\mathcal{P}} - Y)\mathcal{E}[(X - \mu_{X|\mathcal{P}})1_{B_i}]$$

But since $\mu_{X|\mathcal{P}}1_{B_i} = \mathcal{E}(X \mid B_i)1_{B_i}$ we have

$$\begin{aligned}\frac{1}{\mathbb{P}(B_i)}\mathcal{E}[(X - \mu_{X|\mathcal{P}})1_{B_i}] &= \frac{1}{\mathbb{P}(B_i)}[\mathcal{E}(X1_{B_i}) - \mathcal{E}(\mu_{X|\mathcal{P}}1_{B_i})] \\ &= \frac{1}{\mathbb{P}(B_i)}[\mathcal{E}(X1_{B_i}) - \mathcal{E}(\mathcal{E}(X \mid B_i)1_{B_i})] \\ &= \mathcal{E}(X \mid B_i) - \mathcal{E}(X \mid B_i) \\ &= 0\end{aligned}$$

Thus, we have shown that the last term is 0 and so

$$\mathcal{E}[(X-Y)^2] = \mathcal{E}[(X-\mu_{X|\mathcal{P}})^2] + \mathcal{E}[(\mu_{X|\mathcal{P}}-Y)^2] \geq \mathcal{E}[(X-\mu_{X|\mathcal{P}})^2]$$

with equality if and only if $Y = \mu_{X|\mathcal{P}}$. This completes the proof.□

The following theorem gives some key properties of conditional expectation.

Theorem 15 *Let $(\Omega, \mathbb{P})$ be a finite probability space. Let $\mathcal{P} = \{B_1, \ldots, B_n\}$ be a partition of Ω for which $\mathbb{P}(B_i) > 0$ for all i. The conditional expectation $\mathcal{E}_{\mathbb{P}}(X \mid \mathcal{P})$ has the following properties.*

1) The function $\mathcal{E}(\cdot \mid \mathcal{P})$ is linear, that is, for random variables X and Y and real numbers a and b,

$$\mathcal{E}(aX + bY \mid \mathcal{P}) = a\mathcal{E}(X \mid \mathcal{P}) + b\mathcal{E}(Y \mid \mathcal{P})$$

2) The conditional expectation satisfies

$$\mathcal{E}(\mathcal{E}(X \mid \mathcal{P})) = \mathcal{E}(X)$$

3) The conditional expectation $\mathcal{E}(X \mid \mathcal{P})$ can be characterized as the only *random variable Y that is $\mathcal{P}$-measurable and satisfies*

$$\mathcal{E}(Y 1_{B_i}) = \mathcal{E}(X 1_{B_i})$$

for all blocks B_i of $\mathcal{P}$.

4) (Taking out what is known) If Y is a $\mathcal{P}$-measurable random variable then

$$\mathcal{E}(YX \mid \mathcal{P}) = Y\mathcal{E}(X \mid \mathcal{P})$$

5) If X is $\mathcal{P}$-measurable then

$$\mathcal{E}(X \mid \mathcal{P}) = X$$

*6) (***The Tower Properties***) If $\mathcal{Q}$ is a finer partition than $\mathcal{P}$ we have*

$$\mathcal{E}(\mathcal{E}(X \mid \mathcal{P}) \mid \mathcal{Q}) = \mathcal{E}(X \mid \mathcal{P}) = \mathcal{E}(\mathcal{E}(X \mid \mathcal{Q}) \mid \mathcal{P})$$

In words, if we take the expected values with respect to $\mathcal{P}$ and $\mathcal{Q}$ in either order *then only the expected value with respect to the* coarser *partition has any effect.*

7) (An Independent Condition Drops Out) If X and $\mathcal{P}$ are independent, that is, if $\mathcal{P}_X$ and $\mathcal{P}$ are independent then

$$\mathcal{E}(X \mid \mathcal{P}) = \mathcal{E}(X)$$

Proof. To prove 1), we have

$$\begin{aligned}
\mathcal{E}(aX+bY \mid \mathcal{P}) &= \sum_{i=1}^{k} \mathcal{E}(aX+bY \mid B_i) 1_{B_i} \\
&= \sum_{i=1}^{k} \frac{\mathcal{E}((aX+bY)1_{B_i})}{\mathbb{P}(B_i)} 1_{B_i} \\
&= \sum_{i=1}^{k} \frac{\mathcal{E}(aX1_{B_i}+bY1_{B_i})}{\mathbb{P}(B_i)} 1_{B_i} \\
&= \sum_{i=1}^{k} \frac{a\mathcal{E}(X1_{B_i})+b\mathcal{E}(Y1_{B_i})}{\mathbb{P}(B_i)} 1_{B_i} \\
&= a\sum_{i=1}^{k} \frac{\mathcal{E}(X1_{B_i})}{\mathbb{P}(B_i)} 1_{B_i} + b\sum_{i=1}^{k} \frac{\mathcal{E}(Y1_{B_i})}{\mathbb{P}(B_i)} 1_{B_i} \\
&= a\sum_{i=1}^{k} \mathcal{E}(X \mid B_i) 1_{B_i} + b\sum_{i=1}^{k} \mathcal{E}(Y \mid B_i) 1_{B_i} \\
&= a\mathcal{E}(X \mid \mathcal{P}) + b\mathcal{E}(Y \mid \mathcal{P})
\end{aligned}$$

To prove 2), we have

$$\begin{aligned}
\mathcal{E}(\mathcal{E}(X \mid \mathcal{P})) &= \mathcal{E}\left(\sum_{i=1}^{k} \mathcal{E}(X \mid B_i) 1_{B_i}\right) \\
&= \sum_{i=1}^{k} \mathcal{E}(X \mid B_i)\mathcal{E}(1_{B_i}) \\
&= \sum_{i=1}^{k} \mathcal{E}(X \mid B_i)\mathbb{P}(B_i) \\
&= \sum_{i=1}^{k} \mathcal{E}(X1_{B_i}) \\
&= \mathcal{E}\left(\sum_{i=1}^{k} X1_{B_i}\right) \\
&= \mathcal{E}(X)
\end{aligned}$$

To prove 3), let $Y = \mathcal{E}(X \mid \mathcal{P})$. Then Y is $\mathcal{P}$-measurable by definition.
Also (since $1_{B_j}1_{B_i} = 0$ unless $i = j$)

$$\begin{aligned}\mathcal{E}(Y1_{B_i}) &= \mathcal{E}(\mathcal{E}(X \mid \mathcal{P})1_{B_i}) \\ &= \mathcal{E}(\sum_{j=1}^{k}\mathcal{E}(X \mid B_j)1_{B_j}1_{B_i}) \\ &= \mathcal{E}(\mathcal{E}(X \mid B_i)1_{B_i}) \\ &= \mathcal{E}(X \mid B_i)\mathcal{E}(1_{B_i}) \\ &= \mathcal{E}(X \mid B_i)\mathbb{P}(B_i) \\ &= \mathcal{E}(X1_{B_i})\end{aligned}$$

as desired. Now we show that $Y = \mathcal{E}(X \mid \mathcal{P})$ is the only such random variable. So suppose that Z is a random variable that is $\mathcal{P}$-measurable and for which

$$\mathcal{E}(Z1_{B_i}) = \mathcal{E}(X1_{B_i})$$

for all blocks B_i of $\mathcal{P}$. Since Z is constant on B_i, suppose that $Z(\omega) = c$ for all $\omega \in B_i$. Then $Z1_{B_i} = c1_{B_i}$ and so

$$\mathcal{E}(Z1_{B_i}) = \mathcal{E}(c1_{B_i}) = c\mathcal{E}(1_{B_i})$$

It follows that

$$c\mathcal{E}(1_{B_i}) = \mathcal{E}(X1_{B_i})$$

and so

$$Z(\omega) = c = \frac{\mathcal{E}(X1_{B_i})}{\mathcal{E}(1_{B_i})} = \mathcal{E}(X \mid B_i) = \mathcal{E}(X \mid \mathcal{P})(\omega)$$

which shows that $Z(\omega) = \mathcal{E}(X \mid \mathcal{P})$, as desired.

To prove 4), suppose that Z is $\mathcal{P}$-measurable. Let $Z(\omega) = b$ for all $\omega \in B_i$. Then since $ZX1_{B_i} = bX1_{B_i}$ we have

$$\begin{aligned}\mathcal{E}(ZX \mid \mathcal{P})(\omega) &= \mathcal{E}(ZX \mid B_i) \\ &= \frac{\mathcal{E}(ZX1_{B_i})}{\mathbb{P}(B_i)} \\ &= \frac{b\mathcal{E}(X1_{B_i})}{\mathbb{P}(B_i)} \\ &= b\mathcal{E}(X \mid \mathcal{P})(\omega) \\ &= Z(\omega)\mathcal{E}(X \mid \mathcal{P})(\omega)\end{aligned}$$

and so $\mathcal{E}(ZX \mid \mathcal{P}) = Z\mathcal{E}(X \mid \mathcal{P})$, as desired.

To prove 5) take $X = 1$ in part 4), to get

$$\mathcal{E}(Z1 \mid \mathcal{P}) = Z\mathcal{E}(1 \mid \mathcal{P}) = Z$$

which is 5) with Z in place of X.

To prove 6), first we have for $\omega \in \Omega$

$$\begin{aligned}\mathcal{E}(\mathcal{E}(X \mid \mathcal{P}) \mid \mathcal{Q})(\omega) &= \mathcal{E}(\mathcal{E}(X \mid \mathcal{P}) \mid [\omega]_{\mathcal{Q}}) \\ &= \frac{\mathcal{E}(\mathcal{E}(X \mid \mathcal{P})1_{[\omega]_{\mathcal{Q}}})}{\mathbb{P}([\omega]_{\mathcal{Q}})}\end{aligned}$$

Now, since $\mathcal{E}(X \mid \mathcal{P})$ is constant on the blocks of $\mathcal{P}$ and since $\mathcal{Q}$ is finer than $\mathcal{P}$, it follows that $\mathcal{E}(X \mid \mathcal{P})$ is also constant on the blocks of $\mathcal{Q}$. Hence

$$\begin{aligned}\mathcal{E}(\mathcal{E}(X \mid \mathcal{P}) \mid \mathcal{Q})(\omega) &= \frac{\mathcal{E}(X \mid \mathcal{P})(\omega)\mathcal{E}(1_{[\omega]_{\mathcal{Q}}})}{\mathbb{P}([\omega]_{\mathcal{Q}})} \\ &= \mathcal{E}(X \mid \mathcal{P})(\omega)\end{aligned}$$

Since this holds for all ω, we have

$$\mathcal{E}(\mathcal{E}(X \mid \mathcal{P}) \mid \mathcal{Q}) = \mathcal{E}(X \mid \mathcal{P})$$

Now we must show that

$$\mathcal{E}(\mathcal{E}(X \mid \mathcal{Q}) \mid \mathcal{P}) = \mathcal{E}(X \mid \mathcal{P})$$

It is possible to do so directly, but the computation is a bit long. So instead, let us use part 3). First, set

$$Y = \mathcal{E}(X \mid \mathcal{Q})$$

Then according to 3)

$$\mathcal{E}(Y1_B) = \mathcal{E}(X1_B) \tag{1}$$

for all $B \in \mathcal{A}(\mathcal{Q})$. Since $\mathcal{Q}$ is finer than $\mathcal{P}$, it follows that $\mathcal{A}(\mathcal{P}) \subseteq \mathcal{A}(\mathcal{Q})$ and so the equation above holds *a fortiori* for all $B \in \mathcal{A}(\mathcal{P})$. Now let

$$Z = \mathcal{E}(X \mid \mathcal{P})$$

Then according to 3)

$$\mathcal{E}(Z1_B) = \mathcal{E}(X1_B) \tag{2}$$

for all $B \in \mathcal{A}(\mathcal{P})$. Putting together (1) and (2) we have

$$\mathcal{E}(Y1_B) = \mathcal{E}(Z1_B)$$

for all $B \in \mathcal{A}(\mathcal{P})$. Finally, since Z is $\mathcal{P}$-measurable, part 3) implies that

$$Z = \mathcal{E}(Y \mid \mathcal{P})$$

Substituting for Z and Y we have

$$\mathcal{E}(X \mid \mathcal{P}) = \mathcal{E}(Y = \mathcal{E}(X \mid \mathcal{Q}) \mid \mathcal{P})$$

as desired.

To prove 7) suppose that $\mathcal{P}_X$ and $\mathcal{P}$ are independent. Then for any block B of $\mathcal{P}$ we have

$$\mathbb{P}(X = r \mid B) = \mathbb{P}(X = r)$$

and so

$$\begin{aligned}\mathcal{E}(X \mid \mathcal{P})(\omega) &= \mathcal{E}(X \mid [\omega]_{\mathcal{P}}) \\ &= \sum_{r \in X(\Omega)} r\mathbb{P}(X = r \mid [\omega]_{\mathcal{P}}) \\ &= \sum_{r \in X(\Omega)} r\mathbb{P}(X = r) \\ &= \mathcal{E}(X)\end{aligned}$$

This completes the proof. □

Conditional Expectation with Respect to a Random Variable

We can use the results concerning conditional expectation with respect to a partition to define conditional expectation with respect to a random variable. Indeed, this is really nothing new at all (for *finite* sample spaces).

Definition *Let $(\Omega, \mathbb{P})$ be a finite probability space. Let Y be a random variable whose distinct values are $\{y_1, \ldots, y_k\}$. Then the* **conditional expectation** *of a random variable X with respect to Y is the conditional expectation of X with respect to the partition $\mathcal{P}_Y$ generated by Y, in symbols*

$$\mathcal{E}(X \mid Y) = \mathcal{E}(X \mid \mathcal{P}_Y) = \sum_{i=1}^{k} \mathcal{E}(X \mid \{Y = y_i\}) 1_{\{Y = y_i\}} \qquad \square$$

Stochastic Processes

We will be very interested in the price of a stock as it progresses through a sequence of times $t_0 < t_1 < t_2 < \cdots < t_N$. If the stock price at time t_i is denoted by X_i then initially these prices are unknown and so they can be thought of as random variables on some probability space. This leads to the following simple concept, which plays a very important role in many areas of applied mathematics, including the mathematics of finance.

Definition *A **(finite) stochastic process** on a sample space Ω is a sequence $X_1, \ldots, X_N$ of random variables defined on Ω.* □

Stochastic processes are used to model phenomena, like stock prices, that evolve through time. In such cases, there is a relationship between the random variables that gives substance to the stochastic process. We explore this relationship next.

Filtrations and Martingales

Let us introduce the notion of a filtration using an example.

Filtrations

Consider the following game. At each time

$$t_0 < t_1 < t_2 < \cdots < t_N$$

a coin is tossed. A record is kept of the sequence of results.

Let us denote by $\{H, T\}^k$ the set of all sequences of H's and T's of length k. These sequences are called **words** or **strings** of length k over the **alphabet** $\{H, T\}$.

Thus, at time t_i, the **current state** of the game is a string of length i over $\{H, T\}$. The **final states** of the game consist of all words of length N over $\{H, T\}$

$$\Omega = \{H, T\}^N = \{e_1 \cdots e_N \mid e_i = H \text{ or } e_i = T\}$$

The State Tree

Figure 7 gives a pictorial view of the states of the game, called the **state tree** for the game. (In this case $N = 3$.) The states are indicated on the lines of the tree. At time t_0 there is only one state, which is not shown. It is the *empty string*.

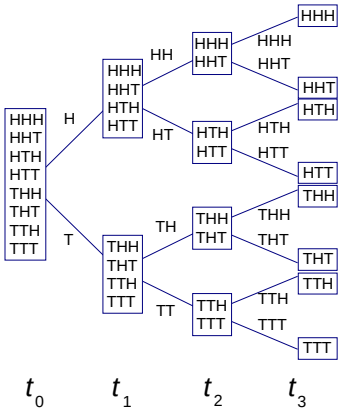


Figure 7–A state tree

The boxes (nodes of the tree) contain the set of *still possible* final states of the game given the current state. In particular, if the current state is $\delta = e_1 \cdots e_i$ then the set of still possible final states is the set of all final states with prefix δ, in symbols

$$\mathcal{F}_i(\delta) = \{\omega \in \Omega \mid [\omega]_i = \delta\}$$

where $[\omega]_i$ denotes the prefix of ω of length i. Let us make some observations about these sets.

First, at each time t_i the 2^i subsets $\mathcal{F}_i(\delta)$ form a partition $\mathcal{P}_i$ of Ω. For instance,

$$\mathcal{P}_2 = \{\mathcal{F}_2(HH), \mathcal{F}_2(HT), \mathcal{F}_2(TH), \mathcal{F}_2(TT)\}$$

and in general

$$\mathcal{P}_i = \{\mathcal{F}_i(\delta_1), \ldots, \mathcal{F}_i(\delta_{2^i})\}$$

where $\delta_1, \ldots, \delta_{2^i}$ are the 2^i elements of $\{H, T\}^i$.

Next, each block $\mathcal{F}_i(\delta)$ in $\mathcal{P}_i$ is contained in a block $\mathcal{F}_{i-1}(\epsilon)$ of the previous partition $\mathcal{P}_{i-1}$. In fact,

$$\mathcal{F}_i(\delta) \subseteq \mathcal{F}_{i-1}([\delta]_{i-1})$$

Hence $\mathcal{P}_i$ is a refinement of $\mathcal{P}_{i-1}$ and

$$\mathcal{P}_0 \prec \mathcal{P}_1 \prec \cdots \prec \mathcal{P}_N$$

is a sequence of finer and finer partitions of Ω.

Finally, note that $\mathcal{P}_0 = \{\Omega\}$ is the coarsest possible partition of Ω, with only one block consisting of Ω itself. On the other end, $\mathcal{P}_N = \Omega$ is the finest possible partition, since each block has size 1, containing just one final state. We are now ready for a definition.

Definition *A sequence* $\mathbb{F} = (\mathcal{P}_0, \ldots, \mathcal{P}_N)$ *of partitions of a set* $\Omega = \{\omega_1, \ldots, \omega_m\}$ *for which*

$$\mathcal{P}_0 \prec \mathcal{P}_1 \prec \cdots \prec \mathcal{P}_N$$

is called a ***filtration****. Moreover, if a filtration satisfies the following conditions it is called an* **information structure**

1) $\mathcal{P}_0$ *is the coarsest possible partition*

$$\mathcal{P}_0 = \{\Omega\}$$

representing no knowledge about Ω*.*

2) $\mathcal{P}_N$ *is the finest possible partition*

$$\mathcal{P}_N = \{\{\omega_1\}, \ldots, \{\omega_m\}\}$$

in which each block has size 1*, representing complete knowledge about* Ω*.* □

Thus, an information structure starts with no knowledge of the final state (other than the fact that it is in Ω), possibly gains some additional knowledge at each time instance (but never loses information) and ends with complete knowledge of the final state.

We should probably mention explicitly that the partitions in a filtration need not double in size as is the case in the example. All that is required is that $\mathcal{P}_i$ be a refinement of $\mathcal{P}_{i-1}$.

One final note. At any time t_i there is a one-to-one correspondence between the possible states δ at that time and the blocks of the partition $\mathcal{P}_i$, given by

$$\delta \leftrightarrow \{\omega \in \Omega \mid [\omega]_i = \delta\}$$

This allows us to identify the intermediate states of the game at time t_i with the blocks $\mathcal{F}_i(\delta)$ of $\mathcal{P}_i$. In fact, when we discuss discrete-time derivative pricing models, we will actually *define* the intermediate states of the model as the blocks of the partitions in a filtration.

Probabilities

Let us now assume that the probability of getting heads is p and that the coin tosses are independent. Then for any $k > 0$ we can define a probability measure on the set $\{H, T\}^k$ by setting

$$\mathbb{P}(\delta) = p^{N_H(\delta)} q^{N_T(\delta)}$$

where

$$N_H(\delta) = \text{Number of } H\text{'s in } \delta$$
$$N_T(\delta) = \text{Number of } T\text{'s in } \delta$$

It is not hard to show that $(\{H, T\}^k, \mathbb{P})$ is a finite probability space.

Theorem 16 *For any $\delta \in \{H, T\}^k$ let*

$$\mathbb{P}(\delta) = p^{N_H(\delta)} q^{N_T(\delta)}$$

1) If $\delta \in \{H, T\}^k$ and $\epsilon \in \{H, T\}^\ell$ then

$$\mathbb{P}(\delta\epsilon) = \mathbb{P}(\delta)\mathbb{P}(\epsilon)$$

2) The pair $(\{H, T\}^k, \mathbb{P})$ is a probability space.

Proof. For part 1), we have

$$\begin{aligned}
\mathbb{P}(\delta\epsilon) &= p^{N_H(\delta\epsilon)} q^{N_T(\delta\epsilon)} \\
&= p^{N_H(\delta)+N_H(\epsilon)} q^{N_T(\delta)+N_T(\epsilon)} \\
&= p^{N_H(\delta)} q^{N_H(\delta)} p^{N_T(\epsilon)} q^{N_T(\epsilon)} \\
&= \mathbb{P}(\delta)\mathbb{P}(\epsilon)
\end{aligned}$$

For part 2), it is clear that for any $\delta \in \{H, T\}^k$

$$0 \leq \mathbb{P}(\delta) \leq 1$$

so we need only show that

$$\sum_{\sigma \in \{H,T\}^k} \mathbb{P}(\sigma) = 1$$

This is clear for $k = 1$ since then we simply have

$$\mathbb{P}(H) + \mathbb{P}(T) = p + (1 - p) = 1$$

We proceed by induction on k. Assuming it is true for k then

$$\begin{aligned}\sum_{\sigma\in\{H,T\}^{k+1}}\mathbb{P}(\sigma) &= \sum_{\sigma\in\{H,T\}^{k}}\mathbb{P}(H\sigma)+\sum_{\sigma\in\{H,T\}^{k}}\mathbb{P}(T\sigma)\\ &= \sum_{\sigma\in\{H,T\}^{k}}\mathbb{P}(H)\mathbb{P}(\sigma)+\sum_{\sigma\in\{H,T\}^{k}}\mathbb{P}(T)\mathbb{P}(\sigma)\\ &= [\mathbb{P}(H)+\mathbb{P}(T)]\sum_{\sigma\in\{H,T\}^{k}}\mathbb{P}(\sigma)\\ &= \mathbb{P}(H)+\mathbb{P}(T)\\ &= 1\end{aligned}$$

and so the result is also true for $k+1$. Thus, it is true for all $k \geq 1$. □

Adapted Random Variables

Now let us suppose that for each heads, a player wins 1 dollar and for each tails the player loses 1 dollar. Let the random variable X_i denote the player's winnings at time t_i.

Thus, for a given time-t_i state $\delta \in \{H,T\}^i$, the winnings are

$$N_H(\delta) - N_T(\delta)$$

At first, it seems natural to define $X_i(\delta)$ to be $N_H(\delta) - N_T(\delta)$. The problem is that in this case each function X_i would be defined on a *different* domain $\{H,T\}^i$ and so the functions X_i would not form a stochastic process.

Instead, we define each X_i on the same set $\Omega = \{H,T\}^N$ of *final* states simply by ignoring that portion of a final state that comes after time t_i. In other words, for any $\omega \in \Omega$ we define

$$X_i(\omega) = N_H([\omega]_i) - N_T([\omega]_i)$$

where $[\omega]_i$ is the *prefix* of ω of length i. In this way, the random variables X_i have a common domain and yet no "future knowledge" of the state of the game is required in order to compute the time-t_i winnings X_i.

Moreover, under this definition the function X_i is $\mathcal{P}_i$-measurable. In fact, for any $\omega \in \mathcal{F}_i(\delta)$ we have

$$X_i(\omega) = N_H(\delta) - N_T(\delta)$$

Hence, knowledge of $\mathcal{P}_i$ implies knowledge of the value of X_i.

In summary, we have a filtration

$$\mathbb{F} = (\mathcal{P}_0, \ldots, \mathcal{P}_N)$$

on the sample space Ω and a stochastic process

$$\mathbb{X} = (X_0, X_1, \ldots, X_N)$$

on Ω (with $X_0 = 0$) for which X_i is $\mathcal{P}_i$-measurable for all i. Because X_i is $\mathcal{P}_i$-measurable for all i, we say that the stochastic process $\mathbb{X}$ is **adapted** to the filtration $\mathbb{F}$, or is $\mathbb{F}$**-adapted**.

Martingales

We would now like to compute the conditional expectation $\mathcal{E}(X_{k+1} \mid \mathcal{P}_k)$, which is the expected value of the time-t_{k+1} winnings given knowledge of the previous time-t_k partition. To this end, let us collect a few simple facts.

Theorem 17 *Let $(\{H,T\}^k, \mathbb{P}_k)$ be the probability space defined by*

$$\mathbb{P}_k(\delta) = p^{N_H(\delta)} q^{N_T(\delta)}$$

Then
1) For $\delta \in \{H,T\}^k$

$$\mathbb{P}_N(\mathcal{F}_i(\delta)) = \mathbb{P}_k(\delta)$$

2) If $\delta \in \{H,T\}^k$ and $\epsilon \in \{H,T\}^\ell$ then

$$\mathbb{P}_{k+\ell+1}(\delta H \epsilon) \pm \mathbb{P}_{k+\ell+1}(\delta T \epsilon) = \mathbb{P}_{k+\ell}(\delta\epsilon)(p \pm q)$$

Proof. For part 1), we have

$$\mathbb{P}_N(\mathcal{F}_i(\delta)) = \sum_{\sigma \in \{H,T\}^{N-k}} \mathbb{P}_N(\delta\sigma) = \mathbb{P}_k(\delta) \sum_{\sigma \in \{H,T\}^{N-k}} \mathbb{P}_{N-k}(\sigma) = \mathbb{P}_k(\delta)$$

For part 2), we have

$$\begin{aligned} \mathbb{P}_{k+\ell+1}(\delta H \epsilon) \pm \mathbb{P}_{k+\ell+1}(\delta T \epsilon) &= \mathbb{P}_{k+\ell+1}(\delta\epsilon H) \pm \mathbb{P}_{k+\ell+1}(\delta\epsilon T) \\ &= \mathbb{P}_{k+\ell}(\delta\epsilon)[\mathbb{P}_1(H) \pm \mathbb{P}_1(T)] \\ &= \mathbb{P}_{k+\ell}(\delta\epsilon)[p \pm q] \end{aligned}$$

and the proof is complete. □

Now we can proceed with our computation of $\mathcal{E}(X_{k+1} \mid \mathcal{P}_k)$. For $\delta \in \{H,T\}^k$

$$\begin{aligned}\mathcal{E}(X_{k+1} \mid \mathcal{F}_k(\delta)) &= \sum_{\omega \in \{H,T\}^N} X_{k+1}(\omega)\mathbb{P}_N(\{\omega\} \mid \mathcal{F}_k(\delta)) \\ &= \frac{1}{\mathbb{P}_N(\mathcal{F}_k(\delta))} \sum_{\omega \in \{H,T\}^N} X_{k+1}(\omega)\mathbb{P}_N(\{\omega\} \cap \mathcal{F}_k(\delta))\end{aligned}$$

But the set $\{\omega\} \cap \mathcal{F}_k(\delta)$ is empty unless $\omega \in \mathcal{F}_k(\delta)$, in which case $\{\omega\} \cap \mathcal{F}_k(\delta) = \{\omega\}$ and so

$$\mathcal{E}(X_{k+1} \mid \mathcal{F}_k(\delta)) = \frac{1}{\mathbb{P}_N(\mathcal{F}_k(\delta))} \sum_{\omega \in \mathcal{F}_k(\delta)} X_{k+1}(\omega)\mathbb{P}_N(\{\omega\})$$

As ω ranges over the set $\mathcal{F}_k(\delta)$ we can write $\omega = \delta\sigma$ where σ ranges over the set $\{H,T\}^{N-k}$ so

$$\mathcal{E}(X_{k+1} \mid \mathcal{F}_k(\delta)) = \frac{1}{\mathbb{P}_N(\mathcal{F}_k(\delta))} \sum_{\sigma \in \{H,T\}^{N-k}} X_{k+1}(\delta\sigma)\mathbb{P}_N(\delta\sigma)$$

Now, in order to evaluate $X_{k+1}(\delta\sigma)$ we need the prefix of $\delta\sigma$ of length $k+1$ so we need to know something about the first symbol in σ. This prompts us to split the sum into two parts based on the first symbol in σ to get

$$\begin{aligned}\mathcal{E}(X_{k+1} \mid \mathcal{F}_k(\delta)) = &\frac{1}{\mathbb{P}_k(\delta)} \sum_{\sigma \in \{H,T\}^{N-k-1}} X_{k+1}(\delta H\sigma)\mathbb{P}_N(\delta H\sigma) \\ &+ \frac{1}{\mathbb{P}_k(\delta)} \sum_{\sigma \in \{H,T\}^{N-k-1}} X_{k+1}(\delta T\sigma)\mathbb{P}_N(\delta T\sigma)\end{aligned}$$

We can now evaluate X_{k+1}

$$\begin{aligned}X_{k+1}(\delta H\sigma) &= N_H(\delta H) - N_T(\delta H) \\ &= 1 + N_H(\delta) - N_T(\delta) \\ &= X_k(\delta) + 1\end{aligned}$$

and

$$\begin{aligned}X_{k+1}(\delta T\sigma) &= N_H(\delta T) - N_T(\delta T) \\ &= N_H(\delta) - N_T(\delta) - 1 \\ &= X_k(\delta) - 1\end{aligned}$$

Substituting gives

$$\begin{aligned}
\mathcal{E}(X_{k+1} \mid \mathcal{F}_k(\delta)) &= \frac{1}{\mathbb{P}_k(\delta)} \sum_{\sigma \in \{H,T\}^{N-k-1}} [X_k(\delta) + 1]\mathbb{P}_N(\delta H\sigma) \\
&\quad + \frac{1}{\mathbb{P}_k(\delta)} \sum_{\sigma \in \{H,T\}^{N-k-1}} [X_k(\delta) - 1]\mathbb{P}_N(\delta T\sigma) \\
&= \frac{1}{\mathbb{P}_k(\delta)} \sum_{\sigma \in \{H,T\}^{N-k-1}} [\mathbb{P}_N(\delta H\sigma) - \mathbb{P}_N(\delta T\sigma)] \\
&\quad + \frac{X_k(\delta)}{\mathbb{P}_k(\delta)} \sum_{\sigma \in \{H,T\}^{N-k-1}} [\mathbb{P}_N(\delta H\sigma) + \mathbb{P}_N(\delta T\sigma)] \\
&= \frac{1}{\mathbb{P}_k(\delta)} \sum_{\sigma \in \{H,T\}^{N-k-1}} \mathbb{P}_{N-1}(\delta\sigma)(p - q) \\
&\quad + \frac{X_k(\delta)}{\mathbb{P}_k(\delta)} \sum_{\sigma \in \{H,T\}^{N-k-1}} \mathbb{P}_{N-1}(\delta\sigma)(p + q) \\
&= [(p - q) + X_k(\delta)] \sum_{\sigma \in \{H,T\}^{N-k-1}} \mathbb{P}(\sigma) \\
&= (p - q) + X_k(\delta)
\end{aligned}$$

and so for any $\omega \in \Omega$ we can write $\omega = \delta\sigma$ and get

$$\begin{aligned}
\mathcal{E}(X_{k+1} \mid \mathcal{P}_k)(\omega) &= \mathcal{E}(X_{k+1} \mid \mathcal{F}_k(\delta)) \\
&= (p - q) + X_k(\delta) \\
&= (p - q) + X_k(\omega)
\end{aligned}$$

and so

$$\mathcal{E}(X_{k+1} \mid \mathcal{P}_k) = X_k + (p - q)\mathbf{1}$$

where $\mathbf{1}$ is the random variable whose value is always 1.

For $p = q$ this takes on special significance, for we get

$$\mathcal{E}(X_{k+1} \mid \mathcal{P}_k) = X_k$$

which says that if we know the time-t_k partition $\mathcal{P}_k$, that is, if we know the state of the game at time t_k, then the expected value of the time-t_{k+1}winnings is the time-t_k winnings. In other words, just as we would expect from the fact that $p = q$, the game is a fair one in that the expected *gain* from one time to the next is 0.

We are ready for an important definition.

Definition *A stochastic process*

$$\mathbb{X} = (X_0, X_1, \ldots, X_N)$$

is a **martingale with respect to the filtration**

$$\mathbb{F} = (\mathcal{P}_0 \prec \mathcal{P}_1 \prec \cdots \prec \mathcal{P}_N)$$

or an $\mathbb{F}$**-martingale** *if* $\mathbb{X}$ *is adapted to* $\mathbb{F}$ *(that is,* X_i *is* $\mathcal{P}_i$*-measurable) and*

$$\mathcal{E}(X_{k+1} \mid \mathcal{P}_k) = X_k$$

that, is, given $\mathcal{P}_i$*, the expected value of* X_{k+1} *is just* X_k*.*□

Thus, martingales model fair games. The following result says that given $\mathcal{P}_i$ the expected value of *any* future random variable is just X_i.

Theorem 18 *If* $\mathbb{X} = (X_0, X_1, \ldots, X_N)$ *is an* $\mathbb{F}$*-martingale where* $\mathbb{F} = (\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_N)$ *then for any* $i > 0$

$$\mathcal{E}(X_{k+i} \mid \mathcal{P}_k) = X_k$$

Proof. We know that

$$\mathcal{E}(X_{k+2} \mid \mathcal{P}_{k+1}) = X_{k+1}$$

Taking conditional expected values gives

$$\mathcal{E}(\mathcal{E}(X_{k+2} \mid \mathcal{P}_{k+1}) \mid \mathcal{P}_k) = \mathcal{E}(X_{k+1} \mid \mathcal{P}_k) = X_k$$

But by the tower property,

$$\mathcal{E}(\mathcal{E}(X_{k+2} \mid \mathcal{P}_{k+1}) \mid \mathcal{P}_k) = \mathcal{E}(X_{k+2} \mid \mathcal{P}_k)$$

and so

$$\mathcal{E}(X_{k+2} \mid \mathcal{P}_k) = X_k$$

An induction argument can now be used to complete the proof. We leave the details to the reader. □

Exercises

1. Show that if X is a random variable on Ω then
   a) $X^{-1}(A \cup B) = X^{-1}(A) \cup X^{-1}(B)$
   b) $X^{-1}(A \cap B) = X^{-1}(A) \cap X^{-1}(B)$
   c) $X^{-1}(A \setminus B) = X^{-1}(A) \setminus X^{-1}(B)$

2. Let $\mathcal{P}$ and $\mathcal{Q}$ be partitions of a nonempty set Ω. Prove that the following are equivalent.
   a) $\mathcal{Q}$ is a refinement of $\mathcal{P}$
   b) Each block of $\mathcal{Q}$ is the union of blocks of $\mathcal{P}$
   c) Each block of $\mathcal{P}$ is contained in a block of $\mathcal{Q}$
   d) $\mathcal{A}(\mathcal{P}) \subseteq \mathcal{A}(\mathcal{Q})$.
3. Prove that if X is a *nonnegative* random variable, that is, $X(\omega) \geq 0$ for all $\omega \in \Omega$ then $\mathcal{E}(X \mid \mathcal{P}) \geq 0$.
4. Let X be a random variable on $(\Omega, \mathbb{P})$. Show that X can be written as a linear combination of indicator functions of the blocks of the partition generated by X.
5. A certain operation results in complete recovery 60% of the time, partial recovery 30% of the time and death 10% of the time. What is the probability of complete recovery, given that a patient survives the operation?
6. Imagine the following experiment. You have an unfair coin, whose probabilities are

$$\mathbb{P}(\text{heads}) = \frac{2}{3}, \mathbb{P}(\text{tails}) = \frac{1}{3}$$

   You also have two urns containing colored balls, where
   1) urn 1 has 3 blue balls and 5 red balls
   2) urn 2 has 7 blue balls and 6 red balls
   First you toss the coin. If the coin comes up heads, you draw a ball at random from urn 1. If the coin comes up tails, you draw a ball at random from urn 2. What is the probability that the ball drawn is blue? *Hint:* use the Theorem on Total Probabilities.
7. Let Ω be a sample space and let $E_1, \ldots, E_n$ form a partition of Ω with $\mathbb{P}(E_k) \neq 0$ for all k. Show that for any event A in Ω,

$$\mathbb{P}(A) = \sum_{k=l}^{n} \mathbb{P}(A \mid E_k)\mathbb{P}(\mathrm{E}_k)$$

8. Let $\mathcal{P} = \{B_1, \ldots, B_k\}$ be a partition of $(\Omega, \mathbb{P})$ with $\mathbb{P}(B_i) > 0$ for all i. Prove **Bayes' Formula**, which states that for any event A in Ω with $\mathbb{P}(A) > 0$, we have

$$\mathbb{P}(E_j \mid A) = \frac{\mathbb{P}(A \mid E_j)\mathbb{P}(E_j)}{\sum_{i=1}^{k} \mathbb{P}(A \mid E_i)\mathbb{P}(E_i)}$$

9. Show that any algebra of sets is closed under intersections and differences. What about symmetric differences? (The **symmetric difference** of two sets is the set of all elements that are in *exactly one* of the two sets.)
10. Prove that any nonempty collection of subsets of Ω is an algebra if and only if it contains Ω and is closed under differences. (The **difference** $A \setminus B$ is the set of all elements of A that are not in B.)
11. Prove that if $E_1, \ldots, E_n$ are events with $\mathbb{P}(E_1 \cap \cdots \cap E_n) > 0$ then

$$\begin{aligned}&\mathbb{P}(E_1 \cap \cdots \cap E_n)\\ &\quad = \mathbb{P}(E_1)\mathbb{P}(E_2 \mid E_1)\mathbb{P}(E_3 \mid E_1 \cap E_2)\cdots\mathbb{P}(E_n \mid E_1 \cap \cdots \cap E_{n-1})\end{aligned}$$

12. Prove in detail that for any partition $\mathcal{P}$ of Ω the set

$$\mathcal{A}(\mathcal{P}) = \{C \subseteq \Omega \mid C = \emptyset \text{ or } C = \text{union of blocks of } \mathcal{P}\}$$

    is an algebra.
13. Suppose that words of length 5 over the binary alphabet $\{0, 1\}$ are sent over a noisy communication line, such as a telephone line. Assume that, because of the noise, the probability that a bit (0 or 1) is received correctly is 0.75. Assume also that the event that one bit is received correctly is independent of the event that another bit is received correctly.
    a) What is the probability that a string will be received correctly?
    b) What is the probability that exactly 3 of the 5 bits in a string are received correctly?
14. Let X and Y be random variables on $(\Omega, \mathbb{P})$. Suppose that X and Y have the same range $\{a_1, \ldots, a_n\}$ and that

$$\mathbb{P}(X = a_i) = \mathbb{P}(Y = a_i) = p_i$$

    Compute $\mathbb{P}(X = Y)$.
15. Let $\mathcal{P}$ be a partition of a probability space $(\Omega, \mathbb{P})$. What is $\mathcal{E}(\mathbf{1} \mid \mathcal{P})$ where 1 is the constant random variable $\mathbf{1}(\omega) = 1$.
16. Let $(X_i)_{i=1,\ldots,n}$ be a martingale with respect to the filtration $(\mathcal{P}_i)_{i=1,\ldots,n}$. Prove that $\mathcal{E}(X_k) = \mathcal{E}(X_0)$ for all $k = 1, \ldots, n$. *Hint*: Use the fact that $\mathcal{E}(\mathcal{E}(X \mid \mathcal{P})) = \mathcal{E}(X)$.
17. Let $X_1, \ldots, X_n$ be random variables on $(\Omega, \mathbb{P})$ all of which have the same expected value μ and the same range $\{r_1, \ldots, r_m\}$. Let N be a random variable on $(\Omega, \mathbb{P})$ where N takes the values $1, \ldots, n$. Assume also that N is independent of the X_i's. Then we can define a random variable S by

$$S = X_1 + \cdots + X_N$$

where

$$S(\omega) = X_1(\omega) + \cdots + X_{N(\omega)}(\omega)$$

Show that

a) $\mathcal{E}(S \mid N = k) = \mu k$

b) $\mathcal{E}(S \mid N) = \mu N$

c) $\mathcal{E}(S) = \mu \mathcal{E}(N)$

Explain in words why part c) makes sense.

18. Prove if $\mathcal{Q}$ is finer than $\mathcal{P}$ then $\mathcal{E}(\mathcal{E}(X \mid \mathcal{Q}) \mid \mathcal{P}) = \mathcal{E}(X \mid \mathcal{P})$. Prove this directly.

Exercises on Submartingales and Supermartingales

Let $\mathbb{X} = (X_0, \ldots, X_N)$ be a stochastic process with respect to a filtration $\mathbb{F} = (\mathcal{P}_0, \ldots, \mathcal{P}_N)$. Then $\mathbb{X}$ is an $\mathbb{F}$**-submartingale** if $\mathbb{X}$ is adapted to $\mathbb{F}$ and

$$\mathcal{E}(X_{k+1} \mid \mathcal{P}_k) \geq X_k$$

Similarly, $\mathbb{X}$ is an $\mathbb{F}$**-supermartingale** if $\mathbb{X}$ is adapted to $\mathbb{F}$ and

$$\mathcal{E}(X_{k+1} \mid \mathcal{P}_k) \leq X_k$$

A stochastic process $\mathbb{A} = (A_0, \ldots, A_N)$ is **predictable** with respect to the filtration $\mathbb{F}$ if A_k is $\mathcal{P}_{k-1}$-measurable for all k.

19. Show that $\mathbb{X}$ is an $\mathbb{F}$-supermartingale if and only if $-\mathbb{X}$ is an $\mathbb{F}$-submartingale. Are submartingales fair games? Whom do they favor? What about supermartingales?

20. (**Doob Decomposition**) Let $(X_0, \ldots, X_N)$ be an $\mathbb{F}$-adapted stochastic process.

    a) Show that there is a unique martingale $(M_0, \ldots, M_N)$ and a unique predictable process $(A_0, \ldots, A_N)$ such that $X_k = M_k + A_k$ and $A_0 = 0$. *Hint*: Set $M_0 = X_0$ and $A_0 = 0$. Then write

    $$X_{k+1} - X_k = M_{k+1} - M_k + A_{k+1} - A_k$$

    and take the conditional expectation with respect to $\mathcal{P}_k$. Use the martingale condition to get an expression for A_{k+1} in terms of A_k.

    b) Show that if (X_k) is a supermartingale then A_k is nonincreasing (that is $A_{k+1} \leq A_k$). What if (X_k) is a submartingale?

21. Let $\mathbb{X} = (X_k \mid k = 0, \ldots, T)$ be a stochastic process.
    a) Prove that for any partition $\mathcal{P}$

$$\max_k \{\mathcal{E}(X_k \mid \mathcal{P})\} \leq \mathcal{E}(\max_k \{X_k\} \mid \mathcal{P})$$

    b) Prove that if $\mathbb{X}$ and $\mathbb{Y} = (Y_0, \ldots, Y_N)$ are submartingales then the process defined by

$$Z_k = \max\{X_k, Y_k\}$$

    is also a submartingale. What about supermartingales?
22. Define the **positive part** of a random variable X by

$$X^+ = \max\{X, 0\}$$

    If $\mathbb{X} = (X_0, \ldots, X_N)$ is a martingale show that $\mathbb{X}^+ = (X_0^+, \ldots, X_N^+)$ is a submartingale.

Chapter 6

Discrete-Time Pricing Models

We are now ready to discuss discrete-time pricing models, that is, pricing models in which all transactions take place at a series of discrete times.

The **derivative pricing problem** is to determine a fair initial value of any derivative. The difficulty is that the *final* value of the derivative is not known at time $t = 0$, since it generally depends on the final value of the underlying asset. However, we will assume that the final value of the underlying is a *known random variable* and so the set of possible final values of the asset is known. Consequently, the set of possible final values of the derivative is also known. Knowledge of this set along with the no-arbitrage principle is the key to derivative pricing.

General Assumptions

We will make the following basic assumptions for the model.

A Unit of Accounting or Numeraire

All prices are given in terms of an unspecified unit of accounting or **numeraire**. This numeraire may be dollars, Eurodollars, pounds Sterling, Yen and so on. A phrase such as "stock worth S" refers to S units of accounting. Later we shall find it useful to use one of the assets of the model as a numeraire. This will have the effect of expressing all prices in relative terms, that is, relative to the chosen asset.

Assumption of a Risk-Free Asset

We will assume that there is always available a *risk-free* asset. The *idea* of the risk-free asset is simple: for each time interval $[t_{i-1}, t_i]$, the risk-free asset is an asset that cannot decrease in value and generally increases in value. Furthermore, the amount of the increase over each interval is *known in advance*. Practical examples of securities that are generally considered risk free are U.S. Treasury bonds and Federally insured deposits.

For reasons that will become apparent as we begin to explore the discrete-time model, it is important to keep separate the notions of the *price* of an asset and the *quantity* of an asset and to assume that it is the *price* of an asset that changes with time, whereas the quantity only changes when we deliberately change it by buying or selling the asset.

Accordingly, one simple way to model the risk-free asset is to imagine a special asset with the following behavior. At time t_0 the asset's price is 1. During each interval $[t_{i-1}, t_i]$, the asset's price increases by a factor of $e^{r_i(t_i - t_{i-1})}$, where r_i is the **risk-free rate** for that interval.

It is traditional in books on the subject to model the risk-free asset as either a bank account or a risk-free bond. For a normal bank account, however, it is not the value of the units (say dollars) that change, but the quantity. If we deposit 10 dollars (10 units of dollar) in an account at time t_0 then after a period of 5% growth we have 10.5 dollars, not 10 "dollars" each worth $\$1.05$.

Whatever the nature of the risk-free asset, the important thing for our analysis is the asset's price structure, which we will define when we formally define asset prices in a moment.

Additional Assumptions

In addition to the previous assumptions, we must also make some not-so-realistic simplifying assumptions. These assumptions are very helpful to the analysis and despite their presence, we can still learn a great deal about how the market works based on these simple models.

Infinitely Divisible Market

The market is **infinitely divisible**, which means that we can speak of, for example, $\sqrt{2}$ or $-\pi$ worth of a stock or bond.

Frictionless Market

All transactions take place immediately and without any external delays.

Perfect Market

The market is **perfect**, that is

- there are no transaction fees or commissions,
- there are no restrictions on short selling,
- the borrowing rate is the same as the lending rate.

Buy-sell Parity

As an extension of the notion of a perfect market, we assume that any asset's buying price is equal to its selling price, that is, if an asset can be bought for S then it can also be sold for S. For instance, if shares of a stock can be bought for S per share then shares can also be sold for S per

share. If a bond can be purchased for S then a similar bond can be sold for S.

Prices are Determined Under the No-Arbitrage Assumption

As we have discussed, if an arbitrage opportunity exists in the market, then prices will be adjusted to eliminate that opportunity. Therefore, it makes sense to price securities under the assumption that there is no arbitrage.

Notation

We have defined the terms positive, strictly positive and strongly positive for vectors in $\mathbb{R}^n$. Let us now define these terms for random variables.

Definition *Let X be a random variable on Ω. Then*
1) X is **nonnegative**, *written $X \geq 0$ if*

$$X(\omega) \geq 0 \text{ for all } \omega \in \Omega$$

(The term **positive** *is also used in the literature for this property.)*
2) X is **strictly positive**, *written $X > 0$ if*

$$X(\omega) \geq 0 \text{ for all } \omega \in \Omega \text{ and } X(\omega) > 0 \text{ for at least one } \omega \in \Omega$$

3) X is **strongly positive**, *written $X \gg 0$ if*

$$X(\omega) > 0 \text{ for all } \omega \in \Omega$$ □

The Basic Model by Example

Before defining the discrete-time model formally, it seems like a good idea to motivate the definition with an example.

Suppose we are interested in a certain stock that is very sensitive to interest rates, in such a way that the stock price generally rises when interest rates fall and vice-versa. (A home-building company would be such a company, for example.)

Thus, we decide to track the discount rate over the next few times that the Federal Reserve Board meets to consider changes in this rate. For our purposes, a *state* of the economy will correspond to a discount rate. (The *discount rate* is the rate that the Federal government charges member banks to borrow money. This rate is often used as a starting-off point for other interest rates.)

It is important to emphasize that when setting up a model of interest rates, we can only *speculate* about future changes based on economic reports, research and other often tenuous tools. However, the model must be created at time t_0 so we have no other choice.

Referring to Figure 1, let us assume that at the current time t_0 the discount rate is 2%. At this time, the economy has only one state, denoted by w_0.

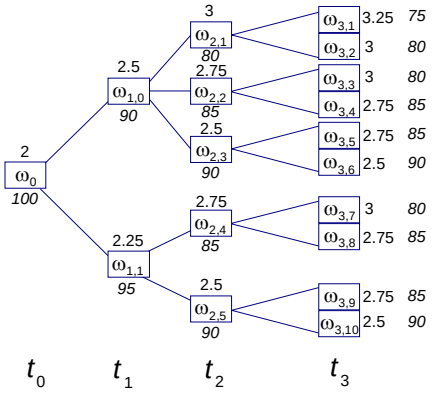


Figure 1

Now, the latest economic information leads us to believe that at time t_1 the Fed will raise interest rates either 0.25 points or 0.5 points. Thus, at time t_1 there will be two states of the economy, denoted by $w_{1,0}$ and $w_{1,1}$. The interest rates are shown next to each state in Figure 1.

We further believe upon good information that at time t_2 the Fed will be inclined to raise rates again. We speculate that if the previous rate hike was 0.5 points, there is a possibility of further hikes of 0.5 or 0.25 points and also a possibility of no change in the rate. However, if the previous hike was only 0.25 points, the strong feeling is that another rate hike of 0.5 or 0.25 points will occur.

In general, we produce a model of interest rates, or states of the economy by speculating on the path of the discount rate over a period of time. Based on predicted interest rates, we also speculate on the price of the stock. These prices are shown in Figure 1 in italics.

Note that there is a stock price for each state and each time. Thus, for example, the time-t_1 *price function* $\widehat{S}_1$ can be defined by

$$\widehat{S}_1(w_{1,0}) = 90, \widehat{S}_1(w_{1,1}) = 95$$

and the time-t_2 *price function* $\widehat{S}_2$ is

$$\begin{aligned}\widehat{S}_2(\omega_{2,1}) &= 80\\ \widehat{S}_2(\omega_{2,2}) &= 85\\ \widehat{S}_2(\omega_{2,3}) &= 90\\ \widehat{S}_2(\omega_{2,4}) &= 85\\ \widehat{S}_2(\omega_{2,5}) &= 90\end{aligned}$$

While these functions are simple to understand, they do suffer from a significant drawback when it comes to doing mathematics, namely, they are defined on different domains. In particular, $\widehat{S}_1$ is defined on $\{\omega_{1,0}, \omega_{1,1}\}$ and $\widehat{S}_2$ is defined on $\{\omega_{2,1}, \ldots, \omega_{2,5}\}$.

Accordingly, it is preferable to work with a sequence of *price random variables* defined on a single probability space. The first step in this endeavor is to take a slightly different view of the states of the economy. We begin with the set of *final states*

$$\Omega = \{w_{3,1}, \ldots, w_{3,10}\}$$

and define all intermediate states as *subsets* of the final states. This idea is pictured in Figure 2.

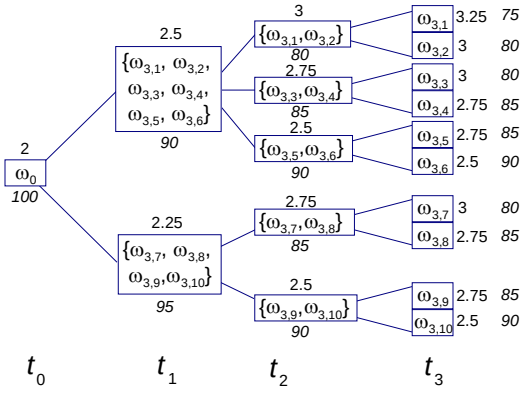


Figure 2

Thus, for example, at time t_1 there are two intermediate states

$$\mathfrak{s}_1 = \{w_{3,1}, \ldots, w_{3,6}\}$$
$$\mathfrak{s}_2 = \{w_{3,7}, \ldots, w_{3,10}\}$$

Now, for example, we can define the time-t_1 *price random variable* S_1 on the set Ω of final states by assigning the value 90 to *all* elements of $\mathfrak{s}_1$ and the value 95 to *all* elements of $\mathfrak{s}_2$. In symbols

$$S_1(w) = \begin{cases} 90 & \text{for all } w \in \mathfrak{s}_1 \\ 95 & \text{for all } w \in \mathfrak{s}_2 \end{cases}$$

It is important to emphasize that this procedure is just a mathematical expediency. It makes no economic sense to talk about the time-t_1 price of a *final* state, since the final state does not exist at time t_1. However, this expediency does no harm and is very useful.

Of course, for this to make sense, the random variable S_1 must be *constant* on each of the subsets $\mathfrak{s}_1$ and $\mathfrak{s}_2$ of Ω, as it is in our example.

Note that at each time t_i the set of intermediate states is a *partition* $\mathcal{P}_i$ of the set Ω of final states and that the time-t_i partition is a *refinement* of the previous time-t_{i-1} partition. Moreover, the price random variable S_i is $\mathcal{P}_i$-measurable.

With this example for motivation, we are ready to formally define the general discrete-time model.

The Basic Model

Here are the basic ingredients of the discrete-time model $\mathbb{M}$.

Time

The model $\mathbb{M}$ has $T + 1$ times

$$t_0 < t_1 < \cdots < t_T$$

Note that there are precisely T time *intervals* $[t_{i-1}, t_i]$ for $i = 1, \ldots, T$.

Assets

The model has a finite number of **basic assets**

$$\mathcal{A} = \{\mathfrak{a}_1, \ldots, \mathfrak{a}_n\}$$

The asset $\mathfrak{a}_1$ *is assumed to be the risk-free asset*.

States of the Economy

At the final time t_T, we assume that the economy is in one of m possible **final states**, given by the **state space**

$$\Omega = \{\omega_1, \ldots, \omega_m\}$$

Initially, that is, at time t_0 we know nothing about the final state of the economy other than the fact that lies in Ω. However, as time passes, we may gain some information (but never lose information) about the possible final state of the economy.

To model this partial knowledge, at each time t_i, we assume that there is a partition

$$\mathcal{P}_i = \{B_{i,1}, \ldots, B_{i,m_i}\}$$

of the state space Ω, called the **time-t_i state partition**. For $i < T$, the blocks of $\mathcal{P}_i$ correspond to the possible states of the economy at time t_i and are called **intermediate states**. Figure 3 shows the **state tree** or **information tree** for the model.

Thus, the term *state* can refer to either an intermediate state (which includes the initial state) or a final state. Also, we will think of both the element ω_i and the singleton *set* $\{\omega_i\}$ as a final state, whichever is more convenient at the time.

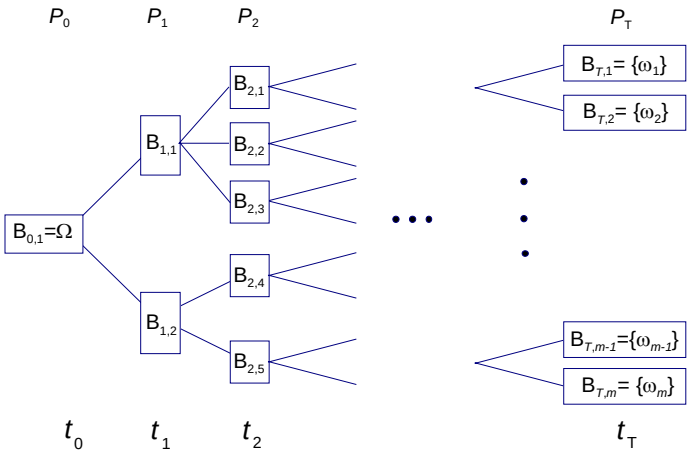


Figure 3

Since no loss of information can occur, it follows that $\mathcal{P}_i$ is a refinement of $\mathcal{P}_{i-1}$. In fact, we will assume that the **state filtration** is an information structure

$$\mathbb{F} = (\mathcal{P}_0, \ldots, \mathcal{P}_T)$$

on Ω. Thus

$$\mathcal{P}_0 = \{\Omega\} = \{\{\omega_1, \ldots, \omega_n\}\}$$

and

$$\mathcal{P}_T = \{\{\omega_1\}, \ldots, \{\omega_m\}\}$$

Natural Probabilities

It is also customary to assume the existence of a probability measure on Ω that reflects the likelihood that each final state in Ω will be the actual final state. These are called **natural probabilities**.

Asset Prices

In a discrete-time model, each asset must not only have a price at each time t_i but that price may depend on the state of the economy at that time. This calls for a price random variable for each time and for each asset. For reasons made clear in the previous example, the time-t_i price random variable should be defined on the sample space Ω and be $\mathcal{P}_i$-measurable.

Definition *For each time t_i and each asset $\mathfrak{a}_j$, the* **price random variable** $S_{i,j}\colon \Omega \to \mathbb{R}$ *is a* $\mathcal{P}_i$-measurable *random variable for which $S_{i,j}(\omega)$ is the time-t_i price of asset $\mathfrak{a}_j$ under the final state ω. The price random variables must satisfy the following properties.*

1) For the risk-free asset, the price random variables are constant, that is, they do not depend upon the state of the economy (which is precisely why they are called risk-free*). In particular,*

$$S_{0,1} = 1$$

and for all times $i > 0$

$$S_{i,1} = e^{r_i(t_i - t_{i-1})} S_{i-1,1}$$

where $r_i \geq 0$ is the risk-free rate in effect during the time interval $[t_{i-1}, t_i]$.

2) For all other assets (that is, for $j > 1$) and for all times t_i

$$S_{i,j} \geq 0$$

3) For a fixed time period t_k the **price vector** *is the random vector of time-t_k prices*

$$S_k = (S_{k,1}, \ldots, S_{k,n})$$

4) For a fixed asset $\mathfrak{a}_j$ the sequence

$$(S_{0,j}, \ldots, S_{T,j})$$

is a stochastic process, called the ***price process*** *for $\mathfrak{a}_j$. It describes the evolution of the price of $\mathfrak{a}_j$ over time.*□

Using the Risk-Free Asset as Numeraire

As we will see in some detail, rather than using dollars, yen or other constant (inflation aside) units of accounting, the use of the risk-free asset itself will provide a great simplification (although it may not seem like it now).

As an example, suppose we wish to assess the quality of various investments. Consider an investment that turns $100 into $104 in one year. Is that a good investment? It is not possible to tell because the quality of an investment must be measured relative to some guaranteed standard. For example, if the risk-free asset turns $100 into $105 in a year, then the 4% investment is not good.

Now, if we use the risk-free asset as unit of accounting instead of dollars, then it is easy to decide whether or not an investment is good (relative to the risk-free investment). For example, if an investment turns 100 risk-free asset units of value into *any number greater than* 100, then it is a good investment, at least relative to the minimal standard risk-free investment.

The **discounted asset prices** are given by

$$\overline{S}_{i,j} = \frac{S_{i,j}}{S_{i,1}}$$

Thus, the discounted price is the nondiscounted price divided by the price $S_{i,1}$ of the risk-free asset *at the same time*. Note in particular that

$$\overline{S}_{i,1} = 1$$

In words, the risk-free asset has constant unit price at all times. This is beacuse the risk-free asset is neutral (not good and not bad) compared to itself.

The **discounted price vector** is given by

$$\overline{S}_i = (\overline{S}_{i,1}, \ldots, \overline{S}_{i,n})$$

Portfolios and Trading Strategies

Portfolios are designed to model the holdings of an investor over a fixed period of time. Of course, it is reasonable to allow adjustments to the asset holdings at each intermediate time. It is also reasonable to allow these adjustments to depend on the state of the economy at that time. Here is the formal definition.

Definition *A **portfolio** for the time interval $[t_{i-1}, t_i]$ is a random vector*

$$\Theta_i = (\theta_{i,1}, \ldots, \theta_{i,n})$$

on Ω where $\theta_{i,j}(\omega_k)$ is the quantity of asset $\mathfrak{a}_j$ acquired at time t_{i-1} and held during the interval $[t_{i-1}, t_i]$ assuming state ω_k. Moreover, $\theta_{i,j}$ is required to be $\mathcal{P}_{i-1}$-measurable. This corresponds to the obvious fact that the quantities $\theta_{i,j}$ must be known at the time t_{i-1} at which the assets are acquired. □

It is worth repeating: The portfolio Θ_i is acquired at time t_{i-1} and held *up to* time t_i.

Note also that the random variables $\theta_{i,j}$ indicate the position as well as the quantity: $\theta_{i,j}$ is positive for a long position and negative for a short position.

It will be convenient to define a (nonstandard) term to denote the holdings of the *risky portion* of a portfolio, that is, all assets except the risk-free asset.

Definition *A **risky holding** for the time interval $[t_{i-1}, t_i]$ is a random vector*

$$\Theta_i^{risky} = (\theta_{i,2}, \ldots, \theta_{i,n})$$

on Ω where $\theta_{i,j}(\omega_k)$ is the quantity of the risky asset $\mathfrak{a}_j$ acquired at time t_{i-1} and held during the interval $[t_{i-1}, t_i]$ assuming state ω_k. Moreover, $\theta_{i,j}$ is required to be $\mathcal{P}_{i-1}$-measurable.□

Portfolio Rebalancing

The use of portfolios in a discrete-time model is a dynamic process that proceeds as follows. At the initial time t_0 the investor acquires the first portfolio

$$\Theta_1 = (\theta_{1,1}, \ldots, \theta_{1,n})$$

which is held through the time interval $[t_0, t_1]$. Note that the random variables $\theta_{1,j}$ are $\mathcal{P}_0$-measurable, that is, constant.

At time t_1 the investor *must* liquidate the portfolio Θ_1 and acquire a new portfolio Θ_2. Of course, there is nothing to prevent the investor from simply *rolling over* the portfolio, by which we mean that $\Theta_2 = \Theta_1$. Even in this case, however, for reasons of consistency it is simpler to think in terms of liquidation followed by acquisition. This does no harm since the model is assumed to be commission free.

In general, at time t_{i-1} the portfolio Θ_{i-1} is liquidated and a new portfolio $\Theta_i = (\theta_{i,1}, \ldots, \theta_{i,n})$ is acquired. This process is referred to as **portfolio rebalancing**.

The sequence $\Phi = (\Theta_1, \ldots, \Theta_T)$ of portfolios obtained through portfolio rebalancing has a name.

Definition *A* **trading strategy** *for a model $\mathbb{M}$ is a sequence of portfolios*

$$\Phi = (\Theta_1, \ldots, \Theta_T)$$

where Θ_i is a portfolio for the time interval $[t_{i-1}, t_i]$. □

We can isolate an individual asset from a trading strategy to obtain a stochastic process that describes the evolution of that asset's holdings. In particular, for each asset $\mathfrak{a}_j$, the **asset holding process** is the stochastic process

$$\Phi_j = (\theta_{1,j}, \ldots, \theta_{T,j})$$

A trading strategy can be represented by a matrix of random variables

$$\Phi = \begin{pmatrix} \Theta_1 \\ \vdots \\ \Theta_T \end{pmatrix} = \begin{pmatrix} \theta_{1,1} & \theta_{1,2} & \cdots & \theta_{1,n} \\ \vdots & \vdots & & \vdots \\ \theta_{T,1} & \theta_{T,2} & \cdots & \theta_{T,n} \end{pmatrix}$$

where the rows correspond to the times (the first row corresponding to time t_0) and the columns correspond to the assets. In fact, the jth column is the asset pricing process for asset $\mathfrak{a}_j$.

Let us also give a (nonstandard) name to the risky portion of a trading strategy.

Definition *A* **risky substrategy** *for a model* $\mathbb{M}$ *is a sequence of risky holdings*

$$\Phi^{risky} = (\Theta_1^{risky}, \ldots, \Theta_T^{risky})$$

where Θ_i^{risky} *is a risky holding for the time interval* $[t_{i-1}, t_i]$. □

In terms of matrices, the risky portion of a trading strategy Φ is the matrix consisting of all columns of the matrix except the first column

$$\Phi^{\text{risky}} = \begin{pmatrix} \Theta_1^{\text{risky}} \\ \vdots \\ \Theta_T^{\text{risky}} \end{pmatrix} = \begin{pmatrix} \theta_{1,2} & \cdots & \theta_{1,n} \\ \vdots & & \vdots \\ \theta_{T,2} & \cdots & \theta_{T,n} \end{pmatrix}$$

Recall that a random process $\mathbb{X} = (X_k)$ is *adapted* to a filtration $\mathbb{F} = (\mathcal{F}_k)$ if X_k is $\mathcal{F}_k$-measurable for each k. This corresponds to the idea that X_k is known at time t_k, when $\mathcal{F}_k$ is known.

On the other hand, an asset holding process $\Phi_j = (\theta_{1,j}, \ldots, \theta_{T,j})$ has the property that $\theta_{k,j}$ is $\mathcal{F}_{k-1}$-measurable. This corresponds to the fact that $\theta_{k,j}$ is known at the previous time t_{k-1}, when $\mathcal{F}_{k-1}$ is known. There are many situations in which such knowledge is common. For example, when placing bets in a game of chance, the player knows the amount X_k of the time-t_k bet before the outcome of the game at time t_k. In this often modeled by saying that X_k is known at time t_{k-1}.

Definition *A stochastic process*

$$\mathbb{X} = (X_1, \ldots, X_T)$$

is **predictable** *or* **previewable** *with respect to the filtration*

$$\mathbb{F} = (\mathcal{P}_0 \prec \mathcal{P}_1 \prec \cdots \prec \mathcal{P}_T)$$

if X_i *is* $\mathcal{P}_{i-1}$*-measurable for all* i.□

With this new language, we can say that an asset holding process is just another name for a *predictable stochastic process*. Also, a trading strategy is just a predicable stochastic process of random vectors.

The Valuation of Portfolios

If $\Phi = (\Theta_1, \ldots, \Theta_T)$ is a trading strategy then since the portfolio Θ_i exists only during the time interval $[t_{i-1}, t_i]$, it makes sense to assign a value to Θ_i only at the acquisition time t_{i-1} and the liquidation time t_i.

The **acquisition value** or **acquisition price** of the portfolio Θ_i is defined by

$$\mathcal{V}_{i-1}(\Theta_i) = \Theta_i \cdot S_{i-1} = \sum_{j=1}^{n} \theta_{i,j} S_{i-1,j}$$

and the **liquidation value** or **liquidation price** of Θ_i is defined by

$$\mathcal{V}_i(\Theta_i) = \Theta_i \cdot S_i = \sum_{j=1}^{n} \theta_{i,j} S_{i,j}$$

We can also discount the portfolio values

$$\overline{\mathcal{V}}_{i-1}(\Theta_i) = \Theta_i \cdot \overline{S}_{i-1} = \sum_{j=1}^{n} \theta_{i,j} \overline{S}_{i-1,j} = \frac{1}{S_{i-1,1}} \sum_{j=1}^{n} \theta_{i,j} S_{i-1,j}$$

and

$$\overline{\mathcal{V}}_i(\Theta_i) = \Theta_i \cdot \overline{S}_i = \sum_{j=1}^{n} \theta_{i,j} \overline{S}_{i,j} = \frac{1}{S_{i,1}} \sum_{j=1}^{n} \theta_{i,j} S_{i,j}$$

Note that the discounted value can be computed directly using the discounted price or indirectly by first computing the nondiscounted price and then discounting the result. This is because a value is computed at a single time.

Self-Financing Trading Strategies

For a trading strategy $\Phi = (\Theta_1, \ldots, \Theta_T)$ if the acquisition price of Θ_{i+1} is equal to the liquidation price of Θ_i, then no money is taken out or put into the model during the time-t_i rebalancing process.

Definition *A trading strategy* $\Phi = (\Theta_1, \ldots, \Theta_T)$ *is* **self-financing** *if for any time* t_i *(for* $i \neq 0, T$*) the acquisition price of* Θ_{i+1} *is equal to the liquidation price of* Θ_i*, that is,*

$$\mathcal{V}_i(\Theta_{i+1}) = \mathcal{V}_i(\Theta_i)$$

The set of all self-financing trading strategies is denoted by $\mathcal{T}$. □

Thus, a self-financing trading strategy is initially purchased for the acquisition value $\mathcal{V}_0(\Theta_1)$ of the first portfolio and is liquidated at time t_T, producing a payoff of $\mathcal{V}_T(\Theta_T)$. No other money is added to or removed from the model during its lifetime.

The set $\mathcal{T}$ of all *self-financing* trading strategies is a vector space under the operations of coordinate-wise addition

$$(\Theta_{1,1}, \ldots, \Theta_{1,T}) + (\Theta_{2,1}, \ldots, \Theta_{2,T}) = (\Theta_{1,1} + \Theta_{2,1}, \ldots, \Theta_{1,T} + \Theta_{2,T})$$

and scalar multiplication

$$a(\Theta_1, \ldots, \Theta_T) = (a\Theta_1, \ldots, a\Theta_T)$$

Demonstration of this is left to the reader.

We can extend the use of the symbol $\mathcal{V}_i$ to *self-financing* trading strategies by defining the time-t_i value of Φ to be the common value of the liquidation price of Θ_i and the acquisition price of Θ_{i+1}. In symbols

$$\mathcal{V}_i(\Phi) = \mathcal{V}_i(\Theta_i) = \mathcal{V}_i(\Theta_{i+1})$$

It is worth emphasizing that this extension applies only to *self-financing* trading strategies.

We will refer to $\mathcal{V}_0(\Phi)$ as the **initial cost** of the trading strategy Φ and to $\mathcal{V}_T(\Phi)$ as the **payoff** of Φ. The following theorem gives some key properties of the valuation functions $\mathcal{V}_i$.

Theorem 1

1) For the valuation function defined on portfolios

a) The i-th valuation function

$$\mathcal{V}_i : RV^n(\Omega) \to RV(\Omega)$$

is a linear transformation, that is, for portfolios $\Theta_1, \Theta_2 \in RV^n(\Omega)$ and real numbers $a, b \in \mathbb{R}$ we have

$$\mathcal{V}_i(a\Theta_1 + b\Theta_2) = a\mathcal{V}_i(\Theta_1) + b\mathcal{V}_i(\Theta_2)$$

b) The acquisition random variable $\mathcal{V}_i(\Theta_{i+1})$ is $\mathcal{P}_i$-measurable. In other words, at time t_i we know the purchase price of Θ_{i+1}.

2) The i-th valuation function defined on trading strategies

$$\mathcal{V}_i : \mathcal{T} \to RV(\Omega)$$

is a linear transformation on $\mathcal{T}$, that is, for $\Phi_1, \Phi_2 \in \mathcal{T}$ and $a, b \in \mathbb{R}$ we have

$$\mathcal{V}_i(a\Phi_1 + b\Phi_2) = a\mathcal{V}_i(\Phi_1) + b\mathcal{V}_i(\Phi_2)$$

Proof. Left to the reader. □

Discounted Gains

For self-financing trading strategies we can make the following definitions regarding the *change* in price or value.

Definition *Let Φ be a* self-financing *trading strategy. The* **discounted change in price** *from time t_{i-1} to time t_i is*

$$\Delta\overline{S}_i = \overline{S}_i - \overline{S}_{i-1} = (\Delta\overline{S}_{i,1}, \ldots, \Delta\overline{S}_{i,n})$$

The **discounted change in value** *from time t_{i-1} to time t_i is*

$$\begin{aligned}\Delta\overline{\mathcal{V}}_i(\Phi) &= \overline{\mathcal{V}}_i(\Phi) - \overline{\mathcal{V}}_{i-1}(\Phi) \\ &= \Theta_i \cdot \Delta\overline{S}_i \\ &= \Theta_i \cdot (\overline{S}_i - \overline{S}_{i-1}) \\ &= \sum_{j=1}^{n} \theta_{i,j}(\overline{S}_{i,j} - \overline{S}_{i-1,j})\end{aligned}$$

The **discounted (cumulative) gain** $\overline{G}_k$ *is*

$$\begin{aligned}\overline{G}_k(\Phi) &= \overline{\mathcal{V}}_k(\Phi) - \overline{\mathcal{V}}_0(\Phi)\\ &= \sum_{i=1}^{k} [\overline{\mathcal{V}}_i(\Phi) - \overline{\mathcal{V}}_{i-1}(\Phi)]\\ &= \sum_{i=1}^{k} \Theta_i \cdot \Delta\overline{S}_i\end{aligned}$$

For $k < \ell$ *the* **discounted gain** $\overline{G}_{k,\ell}$ *is*

$$\begin{aligned}\overline{G}_{k,\ell}(\Phi) &= \overline{\mathcal{V}}_\ell(\Phi) - \overline{\mathcal{V}}_k(\Phi)\\ &= \sum_{i=k+1}^{\ell} [\overline{\mathcal{V}}_i(\Phi) - \overline{\mathcal{V}}_{i-1}(\Phi)]\\ &= \sum_{i=k+1}^{\ell} \Theta_i \cdot \Delta\overline{S}_i\end{aligned}$$

□

A key property of the risk-free asset is the following. Suppose that we are given an initial value $\mathcal{V}_0(\Phi)$ for a self-financing trading strategy $\Phi = (\Theta_1, \ldots, \Theta_T)$ and we are also given all of the quantities $\theta_{i,j}$ of assets (for all $i = 1, \ldots, T$) in Φ *except* the quantities $\theta_{i,1}$ of the risk-free asset. Then the self-financing condition implies that there *one and only one* possibility for the quantities of the risk-free asset.

Intuitively speaking this is quite reasonable. To illustrate, suppose that the initial value of Φ is \$1000. If the risky assets of Θ_1 account for \$900 then there is one and only one choice for the quantity $\theta_{1,1}$ of risk-free asset, namely, the rest of the initial value $\theta_{1,1} = 100$. Now at time t_1 the portfolio Θ_1 is liquidated. Suppose it yields \$1100 (in time-t_1 dollars). If we are given the quantities and hence value of the risky assets in Θ_2, say \$1050 then the quantity of risk-free asset must be such that its value is \$50. Hence, the quantity is $\theta_{2,1} = 50/S_{1,1}$.

In general, if at time t_k we know the liquidation value $\mathcal{V}_k(\Theta_k)$ of Θ_k and we know the quantities and hence the value $\mathcal{V}_k^*(\Theta_{k+1})$ of the *risky* assets, then the remaining value

$$\mathcal{V}_k(\Theta_k) - \mathcal{V}_k^*(\Theta_{k+1})$$

must be spent on the risk-free asset in order to preserve the self-financing condition. Hence

$$\theta_{k+1,1} = \frac{\mathcal{V}_k(\Theta_k) - \mathcal{V}_k^*(\Theta_{k+1})}{S_{k,1}} = \overline{\mathcal{V}}_k(\Theta_k) - \overline{\mathcal{V}}_k^*(\Theta_{k+1})$$

It follows from this discussion that there is one and only one self-financing trading strategy Φ for any given

1) Initial value $\mathcal{V}_0$, which is a $\mathcal{P}_0$-measurable random variable, that is, a constant random variable
2) Risky substrategy

$$\Phi^{\text{risky}} = (\Theta_1^{\text{risky}}, \ldots, \Theta_T^{\text{risky}})$$

that is, set of asset holding processes $\Phi_2, \ldots, \Phi_m$ for the risky assets $\mathfrak{a}_2, \ldots, \mathfrak{a}_m$.

In matrix terms, we have shown that the initial value and the self-financing condition uniquely determine the missing values in the matrix

$$\Phi = \begin{pmatrix} \Theta_1 \\ \vdots \\ \Theta_T \end{pmatrix} = \begin{pmatrix} ? & \theta_{1,2} & \cdots & \theta_{1,n} \\ \vdots & \vdots & & \vdots \\ ? & \theta_{T,2} & \cdots & \theta_{T,n} \end{pmatrix}$$

Thus, all that is required to specify a self-financing trading strategy is the initial value and $n-1$ predictable processes (one for each risky asset).

Locking In a Gain

Now suppose that we are given a self-financing trading strategy Φ. Suppose further that at some intermediate time t_k we wish to "lock in" the discounted gain $\overline{G}_k(\Phi)$ at that time. This can be done simply by liquidating the portfolio Θ_k at time t_k and using all proceeds to buy only the risk-free asset. In symbols

$$\Theta_{k+1} = (\overline{\mathcal{V}}_k(\Theta_k), 0, \ldots, 0)$$

From this point forward, no changes are made to the quantities in the trading strategy. The new trading strategy $\Phi' = (\Theta_1', \ldots, \Theta_T')$ is thus defined by

$$\Theta_i' = \begin{cases} \Theta_i & \text{if } i \le k \\ (\overline{\mathcal{V}}_k(\Theta_k), 0, \ldots, 0) & \text{if } i > k \end{cases}$$

Since the discounted gain $\overline{G}_{k,T}(\Phi)$ from t_k to t_T is 0 because the portoflios contain only the risk-free asset, we have

$$\overline{G}_T(\Phi') = \overline{G}_k(\Phi) + \overline{G}_{k,T}(\Phi) = \overline{G}_k(\Phi)$$

which expresses the fact that we have locked in the discounted gain $\overline{G}_k(\Phi)$. Let us refer to the trading strategy Φ' as the trading strategy that is obtained by **locking in the discounted gain of Φ at time t_k**.

Here is a summary of what we have been discussing.

Theorem 2

1) Discounted gains are additive, that is, for $j < k < \ell$ we have

$$\overline{G}_{j,\ell}(\Phi) = \overline{G}_{j,k}(\Phi) + \overline{G}_{k,\ell}(\Phi)$$

2) The discounted gain does not depend on the quantity of risk-free asset in the trading strategy.

3) Given any constant random variable V_0 and any risky substrategy

$$\Phi^{risky} = (\Theta_1^{risky}, \ldots, \Theta_T^{risky})$$

that is, set of asset holding processes $\Phi_2, \ldots, \Phi_m$ for the risky assets $\mathfrak{a}_2, \ldots, \mathfrak{a}_m$, there is one and only one self-financing trading strategy Φ with initial value $\mathcal{V}_0(\Phi) = V_0$ that has these risky asset holdings. Thus, all that is required to specify a self-financing trading strategy is the initial value and $n-1$ predictable processes (one for each risky asset).

4) Given a self-financing trading strategy Φ and a time t_k, it is possible to find a self-financing trading strategy Φ' that locks in the discounted gain at time t_k, that is, for which

$$\overline{G}_T(\Phi') = \overline{G}_k(\Phi) + \overline{G}_{k,T}(\Phi) = \overline{G}_k(\Phi) \qquad \square$$

Value Shifting

Let

$$\Phi = (\Theta_1, \ldots, \Theta_T)$$

be a self-financing trading strategy. Let us consider what happens if we change the quantity of the risk-free asset by an amount $a \in \mathbb{R}$. In order to maintain the self-financing condition, we roll over this asset at each subsequent time. In symbols, the new portfolios are

$$\Theta_i' = (\theta_{i,1} + a1_\Omega, \theta_{i,2}, \ldots, \theta_{i,n}) = \Theta_i + a(1_\Omega, 0, \ldots, 0)$$

for $i = 1, \ldots, T$.

Let us take a moment to examine the rolling over procedure. If, for example the initial portfolio contains 100 units of the risk-free asset, then at time t_1 the portfolio is liquidated, which realizes $100e^{r_1(t_1-t_0)}$ from that asset. This money is immediately used to purchase 100 units of the risk-free asset again, so that the portfolio Θ_2 also contains exactly 100 units of the risk-free asset.

On the other hand, if the initial portfolio contains -100 units of the risk-free asset (a short position) then the investor has sold 100 units of the risk-free asset and the value of this asset is -100. (The investor is "on the hook" for 100 units.) At time t_1, the portfolio is liquidated and the risk-free asset must be *redeemed* at a *cost* of $100e^{r_1(t_1-t_0)}$ units. The risk-free asset is then immediately sold for its time-t_1 value of $100e^{r_1(t_1-t_0)}$ units. Hence, Θ_2 also has a short position of -100 units of the risk-free asset.

The self-financing condition for Φ' is

$$\mathcal{V}_i(\Theta'_i) = \mathcal{V}_i(\Theta'_{i+1})$$

which is easily verified formally and we leave the details as an exercise.

Comparing values for the trading strategies Φ' and Φ gives

$$\mathcal{V}_i(\Phi') = \mathcal{V}_i(\Phi) + aS_{i,1}$$

which shows that the shift in the initial value of a trading strategy by an amount a using the risk-free asset will ripple through the model, producing a shift in value at time t_i for all states by the amount $aS_{i,1}$. On the other hand, the *discounted* values are changed by a constant amount

$$\overline{\mathcal{V}}_i(\Phi') = \overline{\mathcal{V}}_i(\Phi) + a1_\Omega$$

and the discounted gains are not affected.

Theorem 3 *Let Φ be a trading strategy and let $a \in \mathbb{R}$. Let Φ' be the trading strategy obtained from Φ by adjusting the initial quantity of the risk-free asset by a, that is,*

$$\Theta'_i = \Theta_i + a(1_\Omega, 0, \ldots, 0)$$

for $i = 1, \ldots, T$.

1) The time-t_i value of Φ' is

$$\mathcal{V}_i(\Phi') = \mathcal{V}_i(\Phi) + aS_{i,1}$$

In particular, the initial value is changed by a constant random variable

$$\mathcal{V}_0(\Phi') = \mathcal{V}_0(\Phi) + a1_\Omega$$

and

$$\mathcal{V}_T(\Phi') = \mathcal{V}_T(\Phi) + aS_{T,1}$$

2) In discounted units

$$\overline{\mathcal{V}}_i(\Phi') = \overline{\mathcal{V}}_i(\Phi) + a1_\Omega$$

It follows that the discounted gain is not affected by value shifting

$$\overline{G}_n(\Phi') = \overline{G}_n(\Phi)$$

3) In particular, taking a to be the constant value of $-\mathcal{V}_0(\Phi)$ we get

$$\overline{\mathcal{V}}_0(\Phi') = 0$$

and

$$\overline{\mathcal{V}}_T(\Phi') = \overline{\mathcal{V}}_T(\Phi) - \overline{\mathcal{V}}_0(\Phi)$$

Thus, if Φ is a self-financing trading strategy with $\overline{\mathcal{V}}_0(\Phi) \neq 0$ then there is another self-financing trading strategy Φ' with the same discounted gain but with zero initial value.□

The Pricing Problem: Alternatives and Replication

Our goal is to price assets that are derivatives of the basic assets. By *price* we mean determine an initial price for the derivative under the assumption that the market is free of arbitrage.

To effectively price a derivative at time 0, we need to have some information about the possible payoffs of the derivative at time T. For stock options, this is not a problem, as we have seen. For example, in a two-state economy, suppose $\mathfrak{a}_2$ is a stock with initial cost 100 and final payoff vector $(120, 90)$. Then a call with strike price 95 has final payoff vector $(25, 0)$.

Consider an arbitrary derivative $\mathcal{D}$ of one (or more) of the assets in the model. This derivative is not initially part of the model, but we want to add it to the model in such a way that the no arbitrage opportunities will result. The only thing we know about the derivative is its final payoff X, which is a random variable on Ω.

Now, from the point of view of pricing the derivative, all that matters is its payoff random variable—the precise nature of the derivative (call, put, strike price, etc.) is no longer important. Thus, we are really interested in pricing *random variables* in a manner that is consistent with the absence of arbitrage. In this context, random variables have a special name.

Definition *A random variable* $X: \Omega \to \mathbb{R}$ *is called an* **alternative**, *or* **contingent claim**. □

Note that some authors require nonnegativity in this definition, the idea being that a claim based on an option will not have negative payoffs. In such cases the "claim" will simply expire. However, we do not make this additional restriction.

Thus, the pricing problem is the problem of pricing *alternatives*. Perhaps the simplest and most intuitive method for pricing an alternative X is to find a *trading strategy* Φ whose payoff vector is equal to X

$$\mathcal{V}_T(\Phi) = X$$

and set the initial price of X equal to the initial price of Φ. Indeed, any other choice will lead to arbitrage. For if the initial price P_0 of X is not equal to $\mathcal{V}_0(\Phi)$ then an investor could buy the cheaper of Φ and X and sell the more expensive one. This produces an immediate profit and at the end, the investor liquidates his long position and uses the proceeds to exactly pay off the short position.

This prompts the following definition.

Definition *Let* $X: \Omega \to \mathbb{R}$ *be an alternative. A* **replicating trading strategy** *(or* **replicating strategy** *or* **hedging strategy***) for* X *is a* self-financing *trading strategy* $\Phi = (\Theta_1, \ldots, \Theta_T)$ *whose payoff is equal to* X*, that is*

$$\mathcal{V}_T(\Phi) = \mathcal{V}_T(\Theta_T) = X$$

An alternative X that has at least one replicating strategy is said to be **attainable**. *A model is said to be* **complete** *if every alternative is attainable.* □

The set $\mathcal{M}$ of all attainable alternatives is a subspace of the vector space $\mathrm{RV}(\Omega)$ of all random variables on Ω. We leave verification of this to the reader.

The strategy of pricing an alternative X by first finding a replicating trading strategy Φ for X and then setting the initial price of X equal to the initial value of Φ is the **replicating trading strategy procedure**. We will deal with the issues involved in employing this strategy as soon as we look at an example of finding replicating trading strategies.

EXAMPLE 1 Let us consider an example of computing the replicating trading strategy for an attainable alternative. This is not hard, but does involve solving systems of linear equations, which is generally best done by computer these days.

Figure 4 shows a state tree with stock prices for a two-asset model. For convenience in doing hand computation, we assume that the risk-free rates are 0.

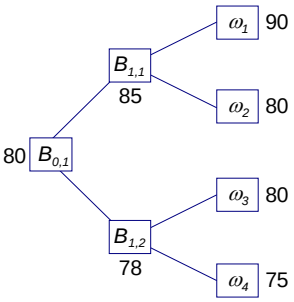


Figure 4

We will see later that this model is complete, so that all alternatives are attainable. Let us compute a self-financing trading strategy $\Phi = (\Theta_1, \Theta_2)$ that replicates the alternative

$$X(\omega_1) = 100, X(\omega_2) = 90, X(\omega_3) = 80, X(\omega_4) = 70$$

that is, for which $\mathcal{V}_2(\Theta_2) = X$, or equivalently,

$$\begin{aligned}\mathcal{V}_2(\Theta_2)(\omega_1) &= 100\\ \mathcal{V}_2(\Theta_2)(\omega_2) &= 90\\ \mathcal{V}_2(\Theta_2)(\omega_3) &= 80\\ \mathcal{V}_2(\Theta_2)(\omega_4) &= 70\end{aligned}$$

Writing out these equations gives

$$\begin{aligned}S_{2,1}(\omega_1)\theta_{2,1}(\omega_1) + S_{2,2}(\omega_1)\theta_{2,2}(\omega_1) &= 100\\ S_{2,1}(\omega_2)\theta_{2,1}(\omega_2) + S_{2,2}(\omega_2)\theta_{2,2}(\omega_2) &= 90\\ S_{2,1}(\omega_3)\theta_{2,1}(\omega_3) + S_{2,2}(\omega_3)\theta_{2,2}(\omega_3) &= 80\\ S_{2,1}(\omega_4)\theta_{2,1}(\omega_4) + S_{2,2}(\omega_4)\theta_{2,2}(\omega_4) &= 70\end{aligned}$$

Substituting the actual prices gives

$$\begin{aligned}\theta_{2,1}(\omega_1) + 90\theta_{2,2}(\omega_1) &= 100\\ \theta_{2,1}(\omega_2) + 80\theta_{2,2}(\omega_2) &= 90\\ \theta_{2,1}(\omega_3) + 80\theta_{2,2}(\omega_3) &= 80\\ \theta_{2,1}(\omega_4) + 75\theta_{2,2}(\omega_4) &= 70\end{aligned}$$

The condition that Θ_2 be $\mathcal{P}_1$-measurable is

$$\begin{aligned}\theta_{2,1}(\omega_1) &= \theta_{2,1}(\omega_2)\\ \theta_{2,1}(\omega_3) &= \theta_{2,1}(\omega_4)\\ \theta_{2,2}(\omega_1) &= \theta_{2,2}(\omega_2)\\ \theta_{2,2}(\omega_3) &= \theta_{2,2}(\omega_4)\end{aligned}$$

and so the previous system can be written using only ω_1 and ω_3 as

$$\begin{aligned}\theta_{2,1}(\omega_1) + 90\theta_{2,2}(\omega_1) &= 100\\ \theta_{2,1}(\omega_1) + 80\theta_{2,2}(\omega_1) &= 90\\ \theta_{2,1}(\omega_3) + 80\theta_{2,2}(\omega_3) &= 80\\ \theta_{2,1}(\omega_3) + 75\theta_{2,2}(\omega_3) &= 70\end{aligned}$$

The first two equations have a unique solution and so do the second two equations, giving

$$\begin{aligned}\Theta_2(\omega_1) = \Theta_2(\omega_2) &= (10, 1)\\ \Theta_2(\omega_3) = \Theta_2(\omega_4) &= (-80, 2)\end{aligned}$$

Working backwards in time, we next compute the acquisition values for Θ_2

$$\mathcal{V}_1(\Theta_2)(\omega_1) = 10 + 85 \cdot 1 = 95$$
$$\mathcal{V}_1(\Theta_2)(\omega_3) = -80 + 78 \cdot 2 = 76$$

The self-financing condition requires that these are also the liquidation values of Θ_1 and so

$$\mathcal{V}_1(\Theta_1)(\omega_1) = 95$$
$$\mathcal{V}_1(\Theta_1)(\omega_3) = 76$$

Writing these out and substituting the actual prices gives the system

$$\theta_{1,1}(\omega_1) + 85\theta_{1,2}(\omega_1) = 95$$
$$\theta_{1,1}(\omega_3) + 78\theta_{1,2}(\omega_3) = 76$$

But Θ_1 is $\mathcal{P}_0$-measurable, that is, constant on Ω, and so for any $\omega \in \Omega$

$$\theta_{1,1}(\omega) + 85\theta_{1,2}(\omega) = 95$$
$$\theta_{1,1}(\omega) + 78\theta_{1,2}(\omega) = 76$$

This system has solution

$$\Theta_1(\omega) = \left(-\frac{950}{7}, \frac{19}{7}\right)$$

which is a portfolio consisting of a short position (sale) of $\frac{950}{7} \approx 135.71$ bonds and a purchase of $\frac{19}{7} \approx 2.71$ shares of stock, for an initial cost of

$$-\frac{950}{7} + 80 \cdot \frac{19}{7} = \frac{570}{7} \approx 81.43$$

Thus, for a cost of 81.43 we can acquire a portfolio that is guaranteed to pay the following

$$X(\omega_1) = 100, X(\omega_2) = 90, X(\omega_3) = 80, X(\omega_4) = 70$$

Note that in some states we have a profit; in others a loss. This is expected in a model with no arbitrage. (We will prove that the model has no arbitrage later.) □

The Law of One Price and the Initial Pricing Functional

It is clear that the replicating strategy procedure can only be used to price *attainable* alternatives. However, there is still one potential problem, and that is the problem of multiple replicating strategies for a given alternative having different initial values. The solution is to require the Law of One Price.

Theorem 4 *The following are equivalent.*
1) ***(Law of One Price)*** *For all trading strategies* Φ_1 *and* Φ_2

$$\mathcal{V}_T(\Phi_1) = \mathcal{V}_T(\Phi_2) \Rightarrow \mathcal{V}_0(\Phi_1) = \mathcal{V}_0(\Phi_2)$$

2) For all trading strategies Φ

$$\mathcal{V}_T(\Phi) = 0 \Rightarrow \mathcal{V}_0(\Phi) = 0$$

Proof. Left to the reader. □

The Law of One Price ensures that the following initial pricing functional is well-defined.

Definition *The* **initial pricing functional** $\mathcal{I}\colon \mathcal{M} \to \mathbb{R}$ *is defined on the vector space* $\mathcal{M}$ *of all* attainable *alternatives by*

$$\mathcal{I}(X) = \mathcal{V}_0(\Phi) \text{ for any } \textit{trading strategy } \Phi \textit{ replicating } X \qquad \square$$

The existence of an initial pricing functional is the key to pricing *attainable* alternatives in a discrete-time model. For if X is an attainable alternative, that is, if there is a trading strategy Φ such that

$$\mathcal{V}_T(\Phi) = X$$

then X can be unambiguously priced at $\mathcal{I}(X) = \mathcal{V}_0(\Phi)$. In addition, we can price X at any time t_k by setting

$$\mathcal{I}_k(X) = \overline{\mathcal{V}}_k(\Phi) \text{ for any trading strategy } \Phi \text{ replicating } X$$

Note that any other pricing will lead to arbitrage. For if at time t_k we have $\mathcal{I}_k(X) \neq \overline{\mathcal{V}}_k(\Phi)$ then an investor can enter the market at this time buying the cheaper of Φ and X and selling the more expensive one. This produces a profit at time t_k and at the end, the investor can liquidate his long position and use the proceeds to exactly pay off the short position.

Arbitrage Trading Strategies

It is now time to formally consider the notion of arbitrage in a discrete-time model. The idea is simple: arbitrage is a situation in which there is no possibility of loss but there is a possibility of a gain. However, one must be careful to measure loss and gain relative to the "natural" guaranteed gain of the risk-free asset. For example, suppose that the simple annual risk-free rate is 10%. Then an investment of $100 that produces $105 in one year could hardly be considered a true gain, for the

same \$100 investment in the risk-free asset would have produced a risk-free \$110! Thus, the first investment is a *loss* relative to the guaranteed risk-free investment.

It may seem natural to define an arbitrage trading strategy Φ to be one whose *discounted* final gain is strictly positive, that is,

$$\overline{G}_T(\Phi) > 0$$

While this definition is used by some authors, the following definition seems a bit more common. It requires that the initial value be 0 as well, in which case the issue of discounting is moot. It is important to point out that while the definitions are not the same, they are equivalent in a sense we will make precise as soon as we have given the formal definition that we will adopt.

Definition *A* self-financing *trading strategy* Φ *is an* **arbitrage trading strategy** *(or* **arbitrage opportunity***) if*

$$\mathcal{V}_0(\Phi) = 0 \text{ and } \mathcal{V}_T(\Phi) > 0$$

or, equivalently in terms of gain,

$$\mathcal{V}_0(\Phi) = 0 \text{ and } \overline{G}_T(\Phi) > 0$$

This says in words that Φ *has zero initial cost, is guaranteed never to result in a loss at time* t_T *and under at least one final state, will result in a positive payoff at time* t_T.□

Let us show the equivalence of the two definitions of arbitrage mentioned earlier. We also show that a strictly positive discounted gain *at any time* will imply an arbitrage opportunity. After all, we have already seen that we can lock in any such gain until the model expires.

Theorem 5 *The following are equivalent for a model* $\mathbb{M}$.

1) $\mathbb{M}$ *has an arbitrage opportunity* Φ*, that is*

$$\mathcal{V}_0(\Phi) = 0 \text{ and } \overline{G}_T(\Phi) > 0$$

2) $\mathbb{M}$ *has a self-financing trading strategy* Φ *with strictly positive discounted final gain, that is*

$$\overline{G}_T(\Phi) > 0$$

3) $\mathbb{M}$ *has a self-financing trading strategy* Φ *with strictly positive discounted gain at some time* t_k*, that is, for some* $1 \leq k \leq T$

$$\overline{G}_k(\Phi) > 0$$

Proof. Obviously 1) implies 2) and a simple value-shift shows that 2) implies 1). (See the last statement of Theorem 3.) Clearly 2) implies 3). Finally, if 3) holds then we may lock in the discounted gain to get a trading strategy satisfying 2).□

Admissible Arbitrage Trading Strategies

Some authors require that arbitrage trading strategies never assume a negative value, as described by the following definition.

Definition *A self-financing trading strategy* Φ *is* **admissible** *if its value at all times is nonnegative, that is,*

$$\mathcal{V}_i(\Phi) \geq 0$$

for all $i = 0, \ldots, T$.□

Thus, an admissible self-financing *arbitrage* trading strategy Φ satisfies

1) $\mathcal{V}_0(\Phi) = 0$
2) $\mathcal{V}_i(\Phi) \geq 0$ all i
3) $\mathcal{V}_T(\Phi) > 0$

In terms of gain this is

1) $\mathcal{V}_0(\Phi) = 0$
2) $\overline{G}_i(\Phi) \geq 0$ all i
3) $\overline{G}_T(\Phi) > 0$

The next result shows that requiring admissibility for arbitrage strategies is not an important distinction.

Theorem 6 *A model has an arbitrage opportunity if and only if it has an* admissible *arbitrage opportunity.*
Proof. Since an admissible arbitrage strategy is an arbitrage strategy, we only need to show the converse, namely, that a model that has an arbitrage strategy

$$\Phi = (\Theta_1, \ldots, \Theta_T)$$

also has an admissible arbitrage strategy.

Of course, if Φ is admissible, then we are done, so let us assume it is not. Let t_k be the *latest* time for which the value of Φ is negative for some state $B_{k,u} \in \mathcal{P}_k$. Since $\mathcal{V}_k(\Theta_{k+1})$ is constant on $B_{k,u}$ we can write

$$a = \mathcal{V}_k(\Theta_k)(\omega) = \mathcal{V}_k(\Theta_{k+1})(\omega) < 0$$

for any $\omega \in B_{k,u}$.

Now the plan is actually quite simple: We want to isolate the holdings that produce the negative value by setting all other unrelated values to 0 and then do a value shift to bring this value to 0. The devil is in the notational details.

The first step is to do nothing before time t_k, that is

$$\Gamma_i = 0 \text{ for } i \leq k$$

From time t_k forward, we follow the strategy Φ *if and only if* the state of the economy is in $B_{k.u}$, where Φ has negative value for the last time. For other states of the economy we do nothing. Thus, Γ is defined by

$$\Gamma_i = \begin{cases} 0 & \text{for } i \leq k \\ 1_{B_{k,u}}\Theta_i & \text{for } i \geq k+1 \end{cases}$$

To examine the values, we consider two cases. For $\omega \notin B_{k,u}$ the acquisition and liquidation values are always 0, that is,

$$\mathcal{V}_i(\Gamma_i)(\omega) = \mathcal{V}_i(\Gamma_{i+1})(\omega) = 0$$

for all i. For $\omega \in B_{k,u}$ the values are 0 up to and including the *liquidation value* at time t_k. However, the *acquisition value* at time t_k is negative (equal to a). Subsequently, all values are nonnegative. Hence, for $\omega \in B_{k,u}$ we can write the sequence of values in the suggestive form

$$0, \ldots, 0, [\mathcal{V}_k(\Gamma_k)(\omega) = 0, \mathcal{V}_k(\Gamma_{k+1})(\omega) = a < 0], \geq 0, \ldots, \geq 0$$

Now we are close to our goal. It is just a matter of adjusting the trading strategy to restore the self-financing condition at time t_k (and not destroy it at subsequent times). This is done by adding the quantity $-a/S_{k,1} > 0$ of risk-free asset to the acquisition portfolio Θ_{k+1} at time t_k under the states in $B_{k,u}$ only and rolling this quantity over.

In particular, set

$$\Gamma'_i = \begin{cases} 0 & \text{for } i \le k \\ 1_{B_{k,u}}\Theta_i - (a1_{B_{k,u}}/S_{k,1}, 0, \ldots, 0) & \text{for } i \ge k+1 \end{cases}$$

For $\omega \notin B_{k,u}$ we still have

$$\mathcal{V}_i(\Gamma'_i)(\omega) = \mathcal{V}_i(\Gamma'_{i+1})(\omega) = 0$$

for all i but for $\omega \in B_{k,u}$, the values are now

$$0, \ldots, 0, [\mathcal{V}_k(\Gamma'_k)(\omega) = 0, \mathcal{V}_k(\Gamma_{k+1})(\omega) = 0], \ge -a > 0, \ldots, \ge -a > 0$$

Thus, $(\Gamma'_1, \ldots, \Gamma'_T)$ is a self-financing admissible arbitrage trading strategy, as desired.□

Characterizing Arbitrage

We now come to the issue of characterizing arbitrage in a way that can be used to price alternatives. The key concept here is the martingale measure.

Definition *Let $\mathbb{M}$ be a discrete-time model. A probability distribution Π on Ω is a* **martingale measure** *(or* **equivalent martingale measure** *or* **risk-neutral probability measure***) for $\mathbb{M}$ if*

1) The probability measure $\mathbb{P}_\Pi$ is strongly positive, that is

$$\mathbb{P}_\Pi(\omega) > 0$$

for all $\omega \in \Omega$

2) For each asset $\mathfrak{a}_j$, the discounted price process $(\overline{S}_{0,j}, \ldots, \overline{S}_{T,j})$ is an $\mathbb{F}$-martingale, that is, for all $k \ge 0$

$$\mathcal{E}_\Pi(\overline{S}_{k+1,j} \mid \mathcal{P}_k) = \overline{S}_{k,j}$$

or equivalently, for any $i, k \ge 0$

$$\mathcal{E}_\Pi(\overline{S}_{k+i,j} \mid \mathcal{P}_k) = \overline{S}_{k,j}$$ □

The next theorem characterizes martingale measures in terms of valuations and gains.

Theorem 7 *For a model $\mathbb{M}$ the following are equivalent for a strongly positive probability measure.*

1) Π *is a martingale measure, that is, the discounted price process for any asset is a martingale. In particular, for all* $k \geq 0$

$$\mathcal{E}_\Pi(\overline{S}_{k+1,j} \mid \mathcal{P}_k) = \overline{S}_{k,j}$$

or equivalently, for any $i, k \geq 0$

$$\mathcal{E}_\Pi(\overline{S}_{k+i,j} \mid \mathcal{P}_k) = \overline{S}_{k,j}$$

2) The discounted valuation process $\mathcal{V}_k(\Phi)$ *of any self-financing trading strategy* Φ *is a martingale under* Π*. In particular, for all* $k \geq 0$

$$\mathcal{E}_\Pi(\overline{\mathcal{V}}_{k+1}(\Phi) \mid \mathcal{P}_k) = \overline{\mathcal{V}}_k(\Phi)$$

or, equivalently, for all $i, k \geq 0$

$$\mathcal{E}_\Pi(\overline{\mathcal{V}}_{k+i}(\Phi) \mid \mathcal{P}_k) = \overline{\mathcal{V}}_k(\Phi)$$

3) At any time, the expected discounted value under Π *of any self-financing trading strategy* Φ *is equal to the initial value of* Φ*, that is, for any* $k \geq 0$

$$\mathcal{E}_\Pi(\overline{\mathcal{V}}_k(\Phi)) = \overline{\mathcal{V}}_0(\Phi)$$

or equivalently, the expected discounted gain under Π *is* 0*, that is*

$$\mathcal{E}_\Pi(\overline{G}_k(\Phi)) = 0$$

4) The expected discounted final payoff under Π *of any self-financing trading strategy* Φ *is equal to the initial value of* Φ*, that is,*

$$\mathcal{E}_\Pi(\overline{\mathcal{V}}_T(\Phi)) = \overline{\mathcal{V}}_0(\Phi)$$

or equivalently, the expected discounted final gain under Π *is* 0*, that is*

$$\mathcal{E}_\Pi(\overline{G}_T(\Phi)) = 0$$

Moreover, if any of these conditions holds, then for all $k \geq 0$

$$\mathcal{E}_\Pi(\overline{S}_{k,j}) = \overline{S}_{0,j}$$

that is, the initial price of asset $\mathfrak{a}_j$ *is the discounted expected price of* $\mathfrak{a}_j$*.*

Proof. Assume that 1) holds and let $\Phi = (\Theta_1, \ldots, \Theta_T)$ be a self-financing trading strategy on $\mathbb{M}$. Multiplying both sides of the martingale condition by $\theta_{k+1,j}$ gives

$$\theta_{k+1,j}\mathcal{E}_{\Pi}(\overline{S}_{k+1,j} \mid \mathcal{P}_k) = \theta_{k+1,j}\overline{S}_{k,j}$$

Since $\theta_{k+1,j}$ is $\mathcal{P}_k$-measurable, it follows from the properties of conditional expectation that we may move $\theta_{k+1,j}$ under the expectation operator to get

$$\mathcal{E}_{\Pi}(\theta_{k+1,j}\overline{S}_{k+1,j} \mid \mathcal{P}_k) = \theta_{k+1,j}\overline{S}_{k,j}$$

Summing on j and using the linearity of the conditional expectation gives

$$\mathcal{E}_{\Pi}(\sum_{j=1}^{n}\theta_{k+1,j}\overline{S}_{k+1,j} \mid \mathcal{P}_k) = \sum_{j=1}^{n}\theta_{k+1,j}\overline{S}_{k,j}$$

that is

$$\mathcal{E}_{\Pi}(\overline{\mathcal{V}}_{k+1}(\Theta_{k+1}) \mid \mathcal{P}_k) = \overline{\mathcal{V}}_k(\Theta_{k+1})$$

or, equivalently

$$\mathcal{E}_{\Pi}(\overline{\mathcal{V}}_{k+1}(\Phi) \mid \mathcal{P}_k) = \overline{\mathcal{V}}_k(\Phi)$$

which is the desired martingale condition for $\overline{\mathcal{V}}_k(\Phi)$ and so 2) holds.

If 2) holds then

$$\mathcal{E}_{\Pi}(\overline{\mathcal{V}}_{k+i}(\Phi) \mid \mathcal{P}_k) = \overline{\mathcal{V}}_k(\Phi)$$

Taking $k = 0$ gives

$$\mathcal{E}_{\Pi}(\overline{\mathcal{V}}_i(\Phi) \mid \mathcal{P}_0) = \overline{\mathcal{V}}_0(\Phi)$$

or

$$\mathcal{E}_{\Pi}(\overline{\mathcal{V}}_i(\Phi)) = \overline{\mathcal{V}}_0(\Phi)$$

or in terms of gain

$$\mathcal{E}_{\Pi}(\overline{G}_i(\Phi)) = \mathcal{E}_{\Pi}(\overline{\mathcal{V}}_i(\Phi) - \overline{\mathcal{V}}_0(\Phi)) = 0$$

which proves 3). Of course, 3) implies 4) since the latter is just a special case of the former.

Suppose now that 4) holds. Thus,

$$\mathcal{E}_{\Pi}(\overline{G}_T(\Phi)) = 0$$

for all self-financing trading strategies Φ. Consider the following trading strategy. Fix a block $B_{k,u}$ of the time-t_k partition $\mathcal{P}_k$. Also fix an asset $\mathfrak{a}_j$.

For time prior to t_k acquire nothing, that is, if $k \geq 1$ then

$$\Theta_1 = \Theta_2 = \cdots = \Theta_k = 0$$

At time t_k acquire the portfolio consisting of exactly one unit of asset $\mathfrak{a}_j$ if and only if the state is in $B_{k,u}$ and in order to maintain the self-financing condition, acquire $-\overline{S}_{k,j}1_{B_{k,u}}$ units of the risk-free asset as well. Thus,

$$\Theta_{k+1} = (-\overline{S}_{k,j}1_{B_{k,u}}, 0, \ldots, 0, 1_{B_{k,u}}, 0, \ldots, 0)$$

The acquisition value of this is

$$\mathcal{V}_k(\Theta_{k+1}) = (-\overline{S}_{k,j}1_{B_{k,u}})S_{k,1} + 1_{B_{k,u}}S_{k,j} = 0 = \mathcal{V}_k(\Theta_k)$$

and so the self-financing condition does indeed obtain at time t_k.

At time t_{k+1} liquidate Θ_{k+1} and invest only in the risk-free asset. Then roll over this asset until the end of the model. Thus, $\Theta_{k+2}, \ldots, \Theta_T$ contain only the risk-free asset and so there is no discounted gain from time t_{k+1} forward.

It follows that the only discounted gain takes place during the interval $[t_k, t_{k+1}]$. Thus, the discounted gain of the self-financing trading strategy

$$\Phi = (\Theta_1, \ldots, \Theta_T)$$

is

$$\begin{aligned}
\overline{G}_T(\Phi) &= \overline{G}_{0,k}(\Phi) + \overline{G}_{k,k+1}(\Phi) + \overline{G}_{k+1,T}(\Phi) \\
&= \overline{G}_{k,k+1}(\Phi) \\
&= \overline{\mathcal{V}}_{k+1}(\Theta_{k+1}) - \overline{\mathcal{V}}_k(\Theta_k) \\
&= \overline{\mathcal{V}}_{k+1}(\Theta_{k+1}) \\
&= (-\overline{S}_{k,j}1_{B_{k,u}})\overline{S}_{k+1,1} + 1_{B_{k,u}}\overline{S}_{k+1,j} \\
&= -\overline{S}_{k,j}1_{B_{k,u}} + 1_{B_{k,u}}\overline{S}_{k+1,j}
\end{aligned}$$

Now, by assumption, the expected value of this is 0, so we have

$$\mathcal{E}_\Pi(-\overline{S}_{k,j}1_{B_{k,u}} + \overline{S}_{k+1,j}1_{B_{k,u}}) = 0$$

or

$$\mathcal{E}_\Pi(\overline{S}_{k+1,j}1_{B_{k,u}}) = \mathcal{E}_\Pi(\overline{S}_{k,j}1_{B_{k,u}})$$

Dividing by the probability of $B_{k,u}$ gives

$$\mathcal{E}_\Pi(\overline{S}_{k+1,j} \mid B_{k,u}) = \mathcal{E}_\Pi(\overline{S}_{k,j} \mid B_{k,u})$$

and since this works equally well for any u we have

$$\mathcal{E}_\Pi(\overline{S}_{k+1,j} \mid \mathcal{P}_k) = \mathcal{E}_\Pi(\overline{S}_{k,j} \mid \mathcal{P}_k)$$

But $\overline{S}_{k,j}$ is $\mathcal{P}_k$-measurable and so this is

$$\mathcal{E}_\Pi(\overline{S}_{k+1,j} \mid \mathcal{P}_k) = \overline{S}_{k,j}$$

which is precisely the martingale condition for the discounted asset pricing process. Hence, Π is a martingale measure and 1) holds. This completes the proof.□

The First Fundamental Theorem of Asset Pricing

The preceeding theorem shows clearly that martingale measures are highly desirable. The First Fundamental Theorem of Asset Pricing tells us precisely when such probability measures exist.

Theorem 8 ***(*The First Fundamental Theorem of Asset Pricing*)*** *For a discrete-time model $\mathbb{M}$ the following are equivalent.*
1) There are no arbitrage trading strategies.
2) There is a martingale measure Π on $\mathbb{M}$.
Proof. The key to the proof of this theorem are the characterizations of arbitrage in Theorem 5 and of martingale measures in Theorem 7, along with one fact from convexity theory to connect them.

If there is a martingale measure Π for $\mathbb{M}$ then by Theorem 7

$$\mathcal{E}_\Pi(\overline{G}_T(\Phi)) = 0$$

for all self-financing trading strategies Φ. This certainly implies that there are no self-financing trading strategies Φ for which

$$\overline{G}_T(\Phi) > 0$$

because such strategies must have positive expectation under a *strongly positive* probability measure. Hence, by Theorem 5 there are no arbitrage opportunities. This proves one-half of the First Fundamental Theorem.

For the converse, the absence of arbitrage implies by Theorem 5 that all discounted gains satisfy

$$\overline{G}_T(\Phi) \not> 0$$

We must show that there is a strongly positive probability measure Π under which all *expected* discounted gains are actually 0.

In order to give the proof a more geometric flavor, we wish to view a random variable not as a function, but as a vector. This is possible because the sample space Ω is *finite*. In particular, if we fix the order of the elements of Ω, say $\Omega = (\omega_1, \ldots, \omega_m)$, then any random variable $X: \Omega \to \mathbb{R}$ can be identified with its vector of values

$$X^{\rightarrow} = (X(\omega_1), \ldots, X(\omega_m))$$

It is also clear that a random variable X is nonnegative, strictly positive or strongly positive if and only if the corresponding vector has this property.

Let us now consider the set $\mathcal{G}$ of all cumulative gain vectors

$$\mathcal{G} = \{\overline{G}_T(\Phi)^{\rightarrow} \mid \Phi \text{ is a trading strategy}\} \subseteq \mathbb{R}^m$$

Since the valuations $\overline{\mathcal{V}}_T$ and $\overline{\mathcal{V}}_0$ are linear transformations, so is $\overline{G}_T$ and so $\mathcal{G}$, being the image of $\overline{G}_T$ is a subspace of $\mathbb{R}^m$.

The absence of arbitrage condition

$$\overline{G}_T(\Phi) \not> 0$$

is equivalent to the condition that $\mathcal{G}$ is a vector space that does not intersect the nonnegative orthant in $\mathbb{R}^n$,

$$\mathbb{R}^n_+ = \{(x_1, \ldots, x_n) \mid x_i \geq 0\}$$

except at the origin, that is,

$$\mathcal{G} \cap \mathbb{R}^n_+ = \{0\}$$

It follows from Theorem 5 of the Appendix that $\mathcal{G}^{\perp}$ contains a strongly positive vector $\Pi = (\pi_1, \ldots, \pi_m)$. In other words, for any self-financing trading strategy Φ we have

$$\overline{G}_T(\Phi)^{\rightarrow} \cdot \Pi = 0$$

But the left-hand side is just the expected value we seek

$$\overline{G}_T(\Phi)^{\rightarrow} \cdot \Pi = \sum_{i=1}^{m} \overline{G}_T(\Phi)(\omega_i)\pi_i = \mathcal{E}_\Pi(\overline{G}_T(\Phi))$$

and so

$$\mathcal{E}_\Pi(\overline{G}_T(\Phi)) = 0$$

This completes the proof.□

The Second Fundamental Theorem of Asset Pricing

Let us now turn our attention to the Second Fundamental Theorem of Asset Pricing. Recall that a model $\mathbb{M}$ is *complete* if every alternative in $\mathbb{R}^m$ is attainable, that is, if for every $X \in \mathbb{R}^m$ there is a self-financing trading strategy Φ such that

$$\mathcal{V}_T(\Phi) = X$$

We will have use of the following fact from linear algebra. Any *strongly positive* probability distribution $\Gamma = (\gamma_1, \ldots, \gamma_m)$ on Ω, where $\mathbb{P}_\Gamma(\omega_k) = \gamma_k$ defines an inner product on the vector space $\mathbb{R}^m$ by

$$\langle X, Y \rangle_\Gamma = \sum_{i=1}^{m} x_i y_i \gamma_i$$

We leave it to the reader to verify that this has the properties of an inner product, which are

1) (Bilinearity)

$$\langle aX + bY, Z \rangle_\Gamma = a\langle X, Z \rangle_\Gamma + b\langle Y, Z \rangle_\Gamma$$
$$\langle X, aY + bZ \rangle_\Gamma = a\langle X, Y \rangle_\Gamma + b\langle X, Z \rangle_\Gamma$$

2) (Symmetry)

$$\langle X, Y \rangle_\Gamma = \langle Y, X \rangle_\Gamma$$

3) (Positive definiteness)

$$\langle X, X \rangle_\Gamma \geq 0$$

with equality if and only if $X = 0$.

Observe also that if $\mathbf{1} = (1, \ldots, 1)$ then for any vector (random variable) X

$$\langle X, \mathbf{1} \rangle_\Gamma = \sum_{i=1}^{m} x_i \gamma_i = \mathcal{E}_\Gamma(X)$$

Now we can turn to the theorem at hand.

Theorem 9 (**The Second Fundamental Theorem of Asset Pricing**) Let $\mathbb{M}$ be a model with no arbitrage opportunities. Then there is a unique martingale measure on $\mathbb{M}$ if and only if the model $\mathbb{M}$ is complete.
Proof. We first show that the completeness of $\mathbb{M}$ implies the uniqueness of the martingale measure on $\mathbb{M}$. Suppose that Π_1 and Π_2 are martingale measures on a complete model $\mathbb{M}$. We want to show that $\Pi_1 = \Pi_2$.

Since Π_1 is a martingale measure, Theorem 7 implies that

$$\mathcal{E}_{\Pi_1}(\overline{\mathcal{V}}_T(\Phi)) = \overline{\mathcal{V}}_0(\Phi)$$

and similarly,

$$\mathcal{E}_{\Pi_2}(\overline{\mathcal{V}}_T(\Phi)) = \overline{\mathcal{V}}_0(\Phi)$$

Hence,

$$\mathcal{E}_{\Pi_1}(\overline{\mathcal{V}}_T(\Phi)) = \mathcal{E}_{\Pi_2}(\overline{\mathcal{V}}_T(\Phi))$$

Since the discounting periods are the same, we have

$$\mathcal{E}_{\Pi_1}(\mathcal{V}_T(\Phi)) = \mathcal{E}_{\Pi_2}(\mathcal{V}_T(\Phi))$$

But since $\mathbb{M}$ is complete, all random variables on Ω have the form $\mathcal{V}_T(\Phi)$ for some self-financing trading strategy. Hence

$$\mathcal{E}_{\Pi_1}(X) = \mathcal{E}_{\Pi_2}(X)$$

for all random variables X on Ω. Taking $X = 1_{\{\omega\}}$ for $\omega \in \Omega$ gives

$$\mathbb{P}_{\Pi_1}(\omega) = \mathbb{P}_{\Pi_2}(\omega)$$

which implies that $\Pi_1 = \Pi_2$. Thus, the martingale measure on $\mathbb{M}$ is unique.

For the converse, suppose that Π is a martingale measure on $\mathbb{M}$ and that the market is not complete. We want to find a different martingale measure Π^* on $\mathbb{M}$. As with the proof of the First Fundamental Theorem,

we wish to fix the order of the elements of $\Omega = (\omega_1, \ldots, \omega_m)$ and think of a random variable on Ω as a vector in $\mathbb{R}^m$.

Since $\mathbb{M}$ is not complete, there is a vector that is not attainable. Put another way, the vector space $\mathcal{M}$ of all attainable vectors is a *proper* subspace of $\mathbb{R}^m$.

Let us consider the inner product defined on $\mathbb{R}^m$ by the martingale measure Π

$$\langle X, Y \rangle_\Pi = \sum_{i=1}^m x_i y_i \pi_i$$

It is a simple fact of linear algebra that if a subspace, such as $\mathcal{M}$, is not all of $\mathbb{R}^m$ then there is a vector $Z = (z_1, \ldots, z_m)$ that is orthogonal to every vector in the subspace. Thus, for any attainable vector $X = (x_1, \ldots, x_m)$ we have

$$\langle X, Z \rangle_\Pi = \sum_{i=1}^m x_i z_i \pi_i = 0$$

Moreover, since the vector $\mathbf{1} = (1, \ldots, 1)$ is attainable (just buy $1/S_{T,1}$ units of the risk-free asset and roll it over), we have

$$0 = \langle \mathbf{1}, Z \rangle_\Pi = \sum_{i=1}^m z_i \pi_i = \mathcal{E}_\Pi(Z)$$

Now let us attempt to define a different martingale measure $\Pi^* = (\pi_1^*, \ldots, \pi_m^*)$ on $\mathbb{M}$. This probability measure must be strongly positive, it must satisfy the martingale condition and it must be different from Π.

Of course, it must first be a probability measure. Noting that

$$\sum_{i=1}^m z_i \pi_i = 0$$

we could try something of the form

$$\pi_i^* = \pi_i + c z_i \pi_i$$

where c is a constant. At least this is a probability measure

$$\sum_{i=1}^{m} \pi_i^* = \sum_{i=1}^{m} \pi_i + c\sum_{i=1}^{m} z_i\pi_i = \sum_{i=1}^{m} \pi_i = 1$$

In addition, since $Z \perp \mathcal{M}$, for any *attainable* vector $X \in \mathcal{M}$ we have

$$\begin{aligned}
\mathcal{E}_{\Pi^*}(X) &= \sum_{i=1}^{m} x_i\pi_i^* \\
&= \sum_{i=1}^{m} x_i(\pi_i + cz_i\pi_i) \\
&= \sum_{i=1}^{m} x_i\pi_i + c\sum_{i=1}^{m} x_i z_i\pi_i \\
&= \mathcal{E}_{\Pi}(X) + \langle X, Z\rangle_{\Pi} \\
&= \mathcal{E}_{\Pi}(X)
\end{aligned}$$

Hence, for any self-financing trading strategy Φ we have

$$\mathcal{E}_{\Pi^*}(\overline{\mathcal{V}}_T(\Phi)) = \mathcal{E}_{\Pi}(\overline{\mathcal{V}}_T(\Phi))$$

and since Π is a martingale measure, Theorem 7 implies that

$$\mathcal{E}_{\Pi^*}(\overline{\mathcal{V}}_T(\Phi)) = \mathcal{E}_{\Pi}(\overline{\mathcal{V}}_T(\Phi)) = \overline{\mathcal{V}}_0(\Phi)$$

But this same theorem then tells us that Π^* is also a martingale measure, that is, provided that Π^* is strongly positive. So all we need to do to complete the proof is choose the constant c so that Π^* is strongly positive, that is

$$\pi_i^* = \pi_i + cz_i\pi_i > 0$$

or equivalently

$$1 + cz_i > 0$$

for all i. To this end, let $M = \max_i\{|z_i|\}$. Then

$$-M \le z_i \le M$$

and so

$$-1 \le \frac{z_i}{M} \le 1$$

Dividing by 2 gives and adding 1 gives

$$\frac{1}{2} \leq 1 + \frac{z_i}{2M} \leq \frac{3}{2}$$

and so we can take

$$c = \frac{1}{2M}$$

This completes the proof.□

Computing Martingale Measures

We now want to consider the issue of computing a martingale measure

$$\Pi = (\pi_1, \ldots, \pi_m)$$

for a model $\mathbb{M}$. The technique is quite simple, although writing down the details is a bit messy.

First, note that any final outcome $\omega_r \in \Omega$ lies in a sequence of blocks, one from each partition $\mathcal{P}_k$, say

$$\{\omega_r\} = B_{T,i_T} \subseteq B_{T-1,i_{T-1}} \subseteq \cdots \subseteq B_{0,i_0} = \Omega$$

Then $\pi_r = \mathbb{P}_\Pi(\omega_r)$ is just a product of conditional probabilities

$$\begin{aligned}
\pi_r &= \mathbb{P}_\Pi(\omega_r) \\
&= \mathbb{P}_\Pi(\omega_r \mid B_{T-1,i_{T-1}})\mathbb{P}_\Pi(B_{T-1,i_{T-1}}) \\
&= \mathbb{P}_\Pi(\omega_r \mid B_{T-1,i_{T-1}})\mathbb{P}_\Pi(B_{T-1,i_{T-1}} \mid B_{T-2,i_{T-2}})\mathbb{P}_\Pi(B_{T-2,i_{T-2}}) \\
&= \cdots \\
&= \mathbb{P}_\Pi(\omega_r \mid B_{T-1,i_{T-1}})\mathbb{P}_\Pi(B_{T-1,i_{T-1}} \mid B_{T-2,i_{T-2}}) \cdots \mathbb{P}_\Pi(B_{1,i_1} \mid B_{0,i_0})
\end{aligned}$$

Thus, we can compute the probabilities in Π if we can compute the conditional probabilities

$$\mathbb{P}_\Pi(B_{k+1,u} \mid B_{k,v}) \tag{1}$$

for all pairs of blocks $B_{k+1,v} \subseteq B_{k,u}$.

The state information tree gives a very intuitive picture of the conditional probabilities and how they are combined to get the martingale measure probabilities. Figure 5 shows a path from the initial block $B_{0,1}$ to a final block $\omega_r = B_{4,i_4}$. The conditional probabilities are used to label the edges of the path.

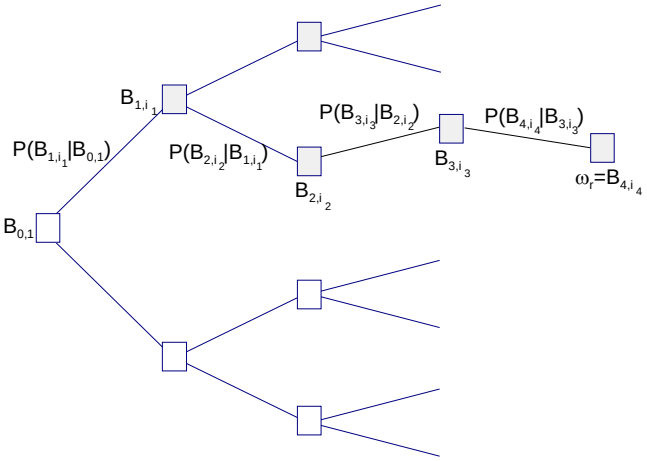


Figure 5 – A path probability

Moreover, the martingale probability $\pi_r = \mathbb{P}_\Pi(\omega_r)$ is just the product of the conditional probabilites that label the edges of the path from $B_{0,1}$ to ω_r. For this reason, we may refer to the martingale probabilities as **path probabilities**.

To actually compute the conditional probabilities in (1), we do not look at paths but rather at individual blocks and their immediate successors, as shown in Figure 6. This forms a submodel of the entire model $\mathbb{M}$.

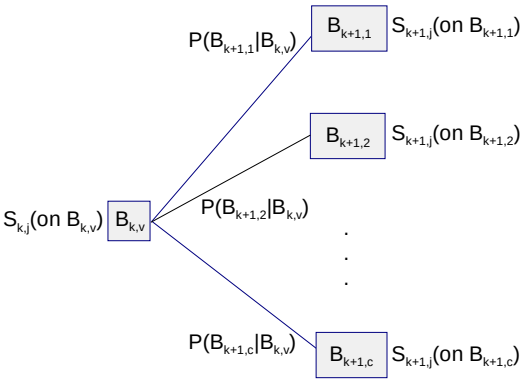


Figure 6 – The submodel starting at $B_{k,v}$

Fix a block $B_{k,v}$. Suppose that the blocks emanating from $B_{k,v}$ are

$$\mathcal{B} = \{B_{k+1,1}, \ldots, B_{k+1,c}\}$$

Then for each asset $\mathfrak{a}_j$, the martingale condition is

$$\overline{S}_{k,j} = \mathcal{E}_\Pi(\overline{S}_{k+1,j} \mid \mathcal{P}_k)$$

The random variable $\mathcal{E}_\Pi(\overline{S}_{k+1,j} \mid \mathcal{P}_k)$ is $\mathcal{P}_k$-measurable, that is, it is constant on the blocks of $\mathcal{P}_k$ so we may suggestively write

$$\overline{S}_{k,j}(\text{on } B_{k,v}) = \mathcal{E}_\Pi(\overline{S}_{k+1,j} \mid \mathcal{P}_k)(\text{on } B_{k,v})$$

or equivalently

$$\overline{S}_{k,j}(\text{on } B_{k,v}) = \mathcal{E}_\Pi(\overline{S}_{k+1,j} \mid B_{k,v})$$

Since $\overline{S}_{k+1,j}$ is constant on the blocks of $\mathcal{P}_{k+1}$ we have

$$\begin{aligned}
\overline{S}_{k,j}(\text{on } B_{k,v}) &= \mathcal{E}_\Pi(\overline{S}_{k+1,j} \mid B_{k,v}) \\
&= \frac{1}{\mathbb{P}_\Pi(B_{k,v})}\mathcal{E}_\Pi(1_{B_{k,v}}\overline{S}_{k+1,j}) \\
&= \frac{1}{\mathbb{P}_\Pi(B_{k,v})}\sum_{i=1}^{c}[1_{B_{k,v}}\overline{S}_{k+1,j}](\text{on } B_{k+1,i})\mathbb{P}_\Pi(B_{k+1,i}) \\
&= \sum_{i=1}^{c}\overline{S}_{k+1,j}(\text{on } B_{k+1,i})\mathbb{P}_\Pi(B_{k+1,i} \mid B_{k,v})
\end{aligned}$$

The equations (one for each $j = 1, \ldots, m$)

$$\overline{S}_{k,j}(\text{on } B_{k,v}) = \sum_{i=1}^{c}\overline{S}_{k+1,j}(\text{on } B_{k+1,i})\mathbb{P}_\Pi(B_{k+1,i} \mid B_{k,v})$$

along with

$$\sum_{i=1}^{c}\mathbb{P}_\Pi(B_{k+1,i} \mid B_{k,v}) = 1$$

which follow from the fact that the blocks $B_{k+1,i}$ for an partition of $B_{k,v}$, provide the means to compute the conditional probabilities.

Theorem 10 *Let $\Pi = (\pi_1, \ldots, \pi_m)$ be a martingale measure for $\mathbb{M}$. Each $\omega_r \in \Omega$ is contained in the unique chain of blocks*

$$B_{0,1} \supseteq B_{1,i_1} \supseteq \cdots \supseteq B_{T,i_T} = \{\omega_r\}$$

Then the martingale probability $\pi_r = \mathbb{P}_\Pi(\omega_r)$ is just a product of conditional probabilities

$$\pi_r = \mathbb{P}_\Pi(\omega_r) = \mathbb{P}_\Pi(B_{1,i_1} \mid B_{0,1})\mathbb{P}_\Pi(B_{2,i_2} \mid B_{1,i_1}) \cdots \mathbb{P}_\Pi(B_{T,i_T} \mid B_{T-1,i_{T-1}})$$

To compute the conditional probabilities $\mathbb{P}_\Pi(\,\cdot \mid B_{k,v})$*, suppose that the blocks emanating from* $B_{k,v}$ *are*

$$\mathcal{B} = \{B_{k+1,1}, \ldots, B_{k+1,c}\}$$

Then we have the system of equations (one for each $j = 1, \ldots, m$*)*

$$\overline{S}_{k,j}(\text{on } B_{k,v}) = \sum_{i=1}^{c} \overline{S}_{k+1,j}(\text{on } B_{k+1,i})\mathbb{P}_\Pi(B_{k+1,i} \mid B_{k,v}) \qquad (1)$$

along with

$$\sum_{i=1}^{c} \mathbb{P}_\Pi(B_{k+1,i} \mid B_{k,v}) = 1 \qquad (2)$$

Let us illustrate the computation of a martingale measure.

EXAMPLE 2 The left half of Figure 7 shows the state tree of Example 1. Recall that risk-free rates are assumed to be 0.

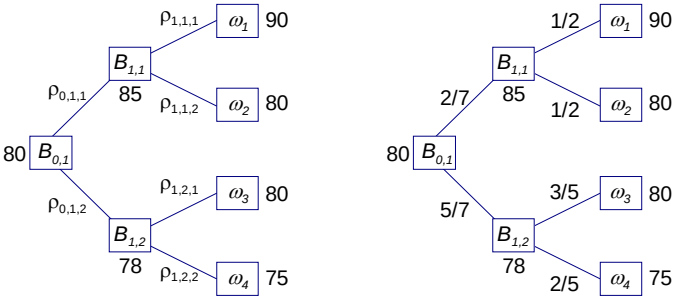


Figure 7 – Computing martingale probabilities

We can compute the conditional probabilities starting with each block of the penultimate partition $\mathcal{P}_1$. For the block $B_{1,1}$ equations (2) and (3) give

$$\begin{aligned} 90\mathbb{P}_\Pi(B_{2,1} \mid B_{1,1}) + 80\mathbb{P}_\Pi(B_{2,2} \mid B_{1,1}) &= 85 \\ \mathbb{P}_\Pi(B_{2,1} \mid B_{1,1}) + \mathbb{P}_\Pi(B_{2,2} \mid B_{1,1}) &= 1 \end{aligned}$$

Solving this system gives

$$\mathbb{P}_\Pi(B_{2,1} \mid B_{1,1}) = \mathbb{P}_\Pi(B_{2,2} \mid B_{1,1}) = \frac{1}{2}$$

as shown on the right in Figure 7. Similarly, for block $B_{1,2}$ we get

$$\begin{aligned} 80\mathbb{P}_\Pi(B_{2,3} \mid B_{1,2}) + 75\mathbb{P}_\Pi(B_{2,4} \mid B_{1,2}) &= 78 \\ \mathbb{P}_\Pi(B_{2,3} \mid B_{1,2}) + \mathbb{P}_\Pi(B_{2,4} \mid B_{1,2}) &= 1 \end{aligned}$$

with solution

$$\mathbb{P}_\Pi(B_{2,3} \mid B_{1,2}) = \frac{3}{5}, \mathbb{P}_\Pi(B_{2,4} \mid B_{1,2}) = \frac{2}{5}$$

Finally, for the block $B_{0,1}$ we have

$$\begin{aligned} 85\mathbb{P}_\Pi(B_{1,1} \mid B_{0,1}) + 78\mathbb{P}_\Pi(B_{1,2} \mid B_{0,1}) &= 80 \\ \mathbb{P}_\Pi(B_{1,1} \mid B_{0,1}) + \mathbb{P}_\Pi(B_{1,2} \mid B_{0,1}) &= 1 \end{aligned}$$

with solution

$$\mathbb{P}_\Pi(B_{1,1} \mid B_{0,1}) = \frac{2}{7}, \mathbb{P}_\Pi(B_{1,2} \mid B_{0,1}) = \frac{5}{7}$$

The right half of Figure 7 shows the conditional probabilities. We can now compute the martingale measure Π simply by taking the products along each path from the starting state to the final states

$$\begin{aligned} \mathbb{P}_\Pi(\omega_1) &= \frac{2}{7} \cdot \frac{1}{2} = \frac{2}{14} \\ \mathbb{P}_\Pi(\omega_2) &= \frac{2}{7} \cdot \frac{1}{2} = \frac{2}{14} \\ \mathbb{P}_\Pi(\omega_3) &= \frac{5}{7} \cdot \frac{3}{5} = \frac{3}{7} \\ \mathbb{P}_\Pi(\omega_4) &= \frac{5}{7} \cdot \frac{2}{5} = \frac{2}{7} \end{aligned}$$

Let us now consider once again the alternative

$$X(\omega_1) = 100, X(\omega_2) = 90, X(\omega_3) = 80, X(\omega_4) = 70$$

The payoffs for a replicating self-financing trading strategy $\Phi = (\Theta_1, \Theta_2)$ are

$$\mathcal{V}_2(\Theta_2)(\omega_1) = 100$$
$$\mathcal{V}_2(\Theta_2)(\omega_2) = 90$$
$$\mathcal{V}_2(\Theta_2)(\omega_3) = 80$$
$$\mathcal{V}_2(\Theta_2)(\omega_4) = 70$$

At this point, we can use Theorem 7, which tells us that

$$\overline{\mathcal{V}}_0(\Phi) = \mathcal{E}_\Pi(\overline{\mathcal{V}}_T(\Phi))$$

Hence,

$$\overline{\mathcal{V}}_0(\Phi) = 100 \cdot \frac{2}{14} + 90 \cdot \frac{2}{14} + 80 \cdot \frac{3}{7} + 70 \cdot \frac{2}{7} = \frac{570}{7} \approx 81.43$$

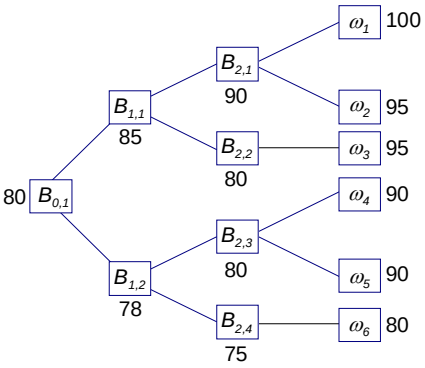
just as we found in Example 1. □

Exercises

1. For the state tree in Figure 4, compute a self-financing trading strategy $\Phi = (\Theta_1, \Theta_2)$ that replicates the alternative

   $$X(\omega_1) = 95, X(\omega_2) = 90, X(\omega_3) = 85, X(\omega_4) = 75$$

   Assume that the risk-free rates are 0.
2. For the state tree shown below



   replicate the alternative

   $$(100, 100, 95, 90, 90, 85)$$

   Assume that the risk-free rates are 0. *Hint*: there is more than one possible answer.
3. Consider the following game. Three fair coins are flipped. The player wins if three heads occur, otherwise the casino wins. For

every \$0.25 the player wagers, the casino must put up \$2.00, making the wager fair. Imagine now that the casino wants to hedge its position against a player who wishes to wager \$1 million dollars. (The casino is at risk for \$8 million.) Accordingly, the casino finds a "market maker" in coin-tossing bets and done the following: before the first toss, it bets \$1 million dollars on heads at even money; before the second toss (if there is one), it bets \$2 million dollars on heads at even money and before the third toss (if there is one), it bets \$4 million dollars on heads at even money. Track the value of the casino's and the player's portfolio during the game. Justify the statement that the casino has entered into a self-financing, replicating complete hedge.

4. Prove that the set $\mathcal{T}$ of all *self-financing* trading strategies is a vector space under the operations of coordinate-wise addition

$$(\Theta_{1,1}, \ldots, \Theta_{1,T}) + (\Theta_{2,1}, \ldots, \Theta_{2,T}) = (\Theta_{1,1} + \Theta_{2,1}, \ldots, \Theta_{1,T} + \Theta_{2,T})$$

and scalar multiplication

$$a(\Theta_1, \ldots, \Theta_T) = (a\Theta_1, \ldots, a\Theta_T)$$

5. Consider the self-financing trading strategy

$$\Phi = (\Theta_1, \ldots, \Theta_T)$$

where

$$\Theta_i = (\theta_{i,1}, \ldots, \theta_{i,n})$$

For any nonzero real number a, let

$$\Phi' = (\Theta'_1, \ldots, \Theta'_T)$$

where

$$\Theta'_i = (\theta_{i,1} + a1_\Omega, \ldots, \theta_{i,n})$$

Show that Φ' is self-financing.

6. Prove that the set $\mathcal{M}$ of all attainable alternatives is a subspace of the vector space $\mathrm{RV}(\Omega)$ of all random variables on Ω.
7. Prove Theorem 4.
8. Consider a model $\mathbb{M}$ with two assets: the risk-free asset and a stock. If the risk-free rates r_i are large enough, will there always be an arbitrage opportunity? Explain your answer. Does this apply to models with more than one risky asset?

9. Consider the following game. A set of 3 coins exists. The first coin is fair, the second coin has probability of heads equal to 0.55 and the third coin has probability of heads 0.45. Draw a state tree indicating the possible outcomes along with their probabilities. Find the path-weight probability distribution.
10. Show that the replicating relation defined by $\Phi_1 \equiv \Phi_2$ if and only if Φ_1 replicates Φ_2 is an *equivalence relation* on the set of self-financing trading strategies, that is, the relation satisfies the following conditions:
    a) (**reflexivity**) $\Phi_1 \equiv \Phi_1$
    b) (**symmetry**) $\Phi_1 \equiv \Phi_2$ implies $\Phi_2 \equiv \Phi_1$
    c) (**transitivity**) $\Phi_1 \equiv \Phi_2$ and $\Phi_2 \equiv \Phi_3$ implies $\Phi_1 \equiv \Phi_3$
11. Prove that if any *strictly positive* alternative is attainable then the market is complete.

A Single-Period Two-Asset, Two-State Model

Consider a simple single-period, two-asset, two-state model $\mathbb{M}$. The model has two assets $\mathcal{A} = (\mathfrak{a}_1, \mathfrak{a}_2)$ where $\mathfrak{a}_1$ is the risk-free bond at rate r and $\mathfrak{a}_2$ is an underlying stock with initial price S_0 and final price S_T. The model has only two states of the economy $\Omega = (\omega_1, \omega_2)$. It is customary to express the final stock price in terms of the initial price. In state ω_1 the stock price is multiplied by a factor u so that

$$S_T = S_0 u$$

and in state ω_2 the price is multiplied by a factor d so that

$$S_T = S_0 d$$

We will assume that $d \leq u$. The following exercises pertain to this model.

12. Show that $\mathbb{M}$ is complete if and only if $d < u$.
13. Consider an option with payoff X given by

$$X(\omega_1) = f_u$$
$$X(\omega_2) = f_d$$

    Find a replicating portfolio for X.
14. Find the initial price of X.
15. Set

$$\pi = \frac{e^{rT} - d}{u - d}$$

and show that the price of the derivative is

$$e^{rT}[\pi f_u + (1-\pi) f_d]$$

What does this tell you about $(\pi, 1-\pi)$?

16. Show that there is no arbitrage in this model if and only if $d < e^{rT} < u$.
17. A day trader is interested in a particular stock currently priced at $\$100$. His assessment is that by the end of the day, the stock will either be selling for $\$101$ or $\$99$. A European call is available at a strike price of $\$99.50$. How should it be priced? Assume that $r = 4\%$.
18. a) Suppose a certain security is currently selling for 160. At time T the security will sell for either 200 or 140. Price a European put on this asset with strike price 180, assuming no arbitrage and interest rate $r = 0$.
    b) Suppose you are fortunate enough to acquire the put described above for only 20. Describe the various portfolios that include the put that will guarantee a profit.

A Single-Period Two-Asset, Three-State Model

Consider now a single-period, two-asset model with three states. Assume a risk-free rate of 0. Suppose that

$$S_{0,2} = 25$$

and

$$S_{1,2}(\omega_1) = 40, S_{1,2}(\omega_2) = 30, S_{1,2}(\omega_3) = 20$$

19. Show that the model is not complete.
20. Find all martingale measures for this model.
21. Show that the following are martingale measures

$$\Pi_1 = (\frac{1}{12}, \frac{4}{12}, \frac{7}{12})$$
$$\Pi_2 = (\frac{1}{6}, \frac{1}{6}, \frac{4}{6})$$

22. Find a replicating trading strategy (portfolio) and price a call option with strike price 20 using the two martingale measures of the previous exercise.

Chapter 7

The Cox-Ross-Rubinstein Model

In this chapter, we discuss a specific discrete-time model known as the **Cox-Ross-Rubinstein** model because it was first described by these gentlemen in 1979. We will abbreviate Cox-Ross-Rubinstein by CRR. The CRR model is also referred to in the literature as the **binomial model** for reasons that will become apparent as we proceed.

In a later chapter, we will use this model to derive the famous Black-Scholes option pricing formula.

The Model

Times

The Cox-Ross-Rubinstein model is a **discrete model**, in that it has a finite number of times times

$$t_0 < t_1 < \cdots < t_T$$

Moreover, the time intervals $[t_i, t_{i-1}]$ have equal length Δt, that is

$$t_i - t_{i-1} = \Delta t$$

Thus, the entire lifetime of the model is

$$L = t_T - t_0 = T\Delta t$$

Assets

The CRR model has only two assets: the risk-free asset $\mathfrak{a}_1$ and a risky asset $\mathfrak{a}_2$.

The States of the Model

Figure 1 shows a portion of the state tree for the CRR model.

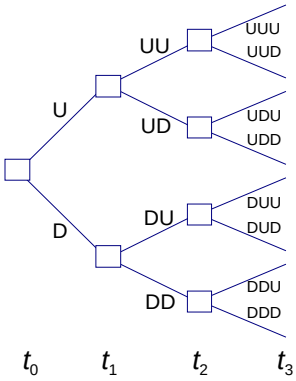


Figure 1–State tree for Cox-Ross-Rubinstein model

The CRR model assumes that during each time interval $[t_i, t_{i+1}]$ the state of the economy changes in one of two ways: it goes up or it goes down. Also, the direction of change in the economy is independent of past changes.

If we denote an up-tick in the economy by U and a down-tick by D then a final state of the economy is a *string* of U's and D's of length T. Let us denote the set of all strings of U's and D's of length k by $\{U, D\}^k$. For instance,

$$\{U, D\}^2 = \{UU, UD, DU, DD\}$$

Thus, the final state space is

$$\Omega = \{U, D\}^T$$

Note that $\{U, D\}^k$ has size 2^k, in particular, Ω has size 2^T.

Since we will be dealing regularly with strings of U's and D's, let us establish a bit of notation. For any $\omega \in \{U, D\}^T$ we denote the prefix of ω of length i by $[\omega]_i$. Thus, if $\omega = e_1 \cdots e_T$ then

$$[\omega]_i = e_1 \cdots e_i$$

for any $i \leq T$. We also set

$$N_U(\omega) = \text{number of } U\text{'s in } \omega$$
$$N_D(\omega) = \text{number of } D\text{'s in } \omega$$

The intermediate states of the model are defined as follows. There is one time-t_k intermediate state for each string in $\{U, D\}^k$. In particular, for

$\delta = e_1 \cdots e_k \in \{U, D\}^k$ the intermediate state $B_\delta \in \mathcal{P}_k$ is the set of all final states having *prefix* δ

$$B_\delta = \{\omega \in \Omega \mid [\omega]_i = \delta\}$$

Thus, $\mathcal{P}_k$ has exactly 2^k blocks (intermediate states).

For example, if $T = 4$ then $\mathcal{P}_1$ consists of the two intermediate states

$$\begin{aligned} B_U = \{&UUUU, UUUD, UUDU, UUDD, \\ &UDUU, UDUD, UDDU, UDDD\} \\ B_D = \{&DUUU, DUUD, DUDU, DUDD, \\ &DDUU, DDUD, DDDU, DDDD\} \end{aligned}$$

The partition $\mathcal{P}_2$ consists of the four intermediate states

$$\begin{aligned} B_{UU} &= \{UUUU, UUUD, UUDU, UUDD\} \\ B_{UD} &= \{UDUU, UDUD, UDDU, UDDD\} \\ B_{DU} &= \{DUUU, DUUD, DUDU, DUDD\} \\ B_{DD} &= \{DDUU, DDUD, DDDU, DDDD\} \end{aligned}$$

At time t_0 there is only one (initial) state $B_\epsilon = \Omega$. This corresponds to the empty string ϵ, which is a prefix of all strings.

It is clear that each block $B_{e_1 \cdots e_k}$ of $\mathcal{P}_k$ gives rise to exactly 2 blocks of the next partition $\mathcal{P}_{k+1}$, namely

$$B_{e_1 \cdots e_k U} \text{ and } B_{e_1 \cdots e_k D}$$

Put another way, each node of the state tree has exactly two edges emanating from it.

Natural Probabilities

We also need to consider the natural probability that the economy takes an upturn at any given time. Let us denote this probability by p. We should emphasize that the natural probability is estimated by economic, not mathematical means.

The Price Functions

To simplify the notation a bit, let us denote the time-t_k price of the risk-free asset by B_k and the time-t_k price of the risky asset, which we may think of as a stock for concreteness by S_k.

The CRR model specifies that the stock price is determined by a pair of real numbers u and d satisfying

$$0 < d < u$$

If during the time interval $[t_k, t_{k+1}]$ the economy goes up then the stock price goes up from S_k to $S_k u$ and if the economy goes down then the stock price also goes down from S_k to $S_k d$. Note that u and d are *constants*, that is, they do not depend on time.

It follows that the time-t_k stock price function S_k is given by

$$S_k(\omega) = S_0 u^{N_U([\omega]_k)} d^{N_D([\omega]_k)}$$

for any final state $\omega \in \{U, D\}^T$. In particular, the final price is

$$S_T(\omega) = S_0 u^{N_U(\omega)} d^{N_D(\omega)}$$

The fact that S_k is $\mathcal{P}_k$-measurable is reflected in the fact that the value $S_k(\omega)$ depends only on the *prefix* $[\omega]_k$ of ω and thus only on what has happened up to time t_k. Note also that the price of the stock at time t_k depends only on the *number* of U's and D's in the state up to that time, and not on their order. This is a key feature of the CRR model that is not possessed by discrete-time models in general (and is probably not very realistic as well).

Note that the stock price functions also satisfy a recurrence relation

$$S_k(\omega) = S_{k-1} u^{E_k(\omega)} d^{1-E_k(\omega)}$$

The price of the risk-free asset is, as always, given by the risk-free rate. In the CRR model, we assume that this rate r is constant throughout the lifetime of the model. Thus, for all final states ω, the price of the risk-free asset at time t_k is

$$e^{r(t_k - t_0)}$$

(Of course, the units must match. For example, if r is an annual rate then the times t_k must be measured in year.)

Martingale Measures in the CRR model

Suppose that Π is a martingale measure for a CRR model $\mathbb{M}$. Theorem 10 of Chapter 6 tells us how to compute the conditional probabilities that are used to compute Π.

Consider a block

$$B_\delta = \{\omega \in \Omega \mid [\omega]_i = \delta\}$$

of $\mathcal{P}_k$. The blocks of $\mathcal{P}_{k+1}$ that are contained in B_δ are

$$B_{\delta U} = \{\omega \in \Omega \mid [\omega]_{i+1} = \delta U\}$$

and

$$B_{\delta D} = \{\omega \in \Omega \mid [\omega]_{i+1} = \delta D\}$$

Figure 2 shows the block B_δ and its successors.

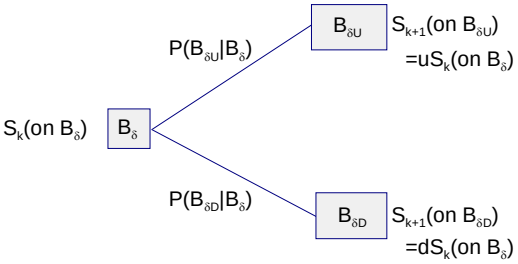


Figure 2 – The block B_δ and its successors

Let us denote the conditional probabilities by

$$\rho_{\delta,U} = \mathbb{P}_\Pi(B_{\delta U} \mid B_\delta)$$
$$\rho_{\delta,D} = \mathbb{P}_\Pi(B_{\delta D} \mid B_\delta)$$

The CRR model dictates that

$$S_{k+1}(\text{on } B_{\delta U}) = uS_k(\text{on } B_\delta)$$
$$S_{k+1}(\text{on } B_{\delta D}) = dS_k(\text{on } B_\delta)$$

or in discounted form (multiplying both sides by $e^{-(k+1)r\Delta t}$)

$$\overline{S}_{k+1}(\text{on } B_{\delta U}) = e^{-r\Delta t}u\overline{S}_k(\text{on } B_\delta)$$
$$\overline{S}_{k+1}(\text{on } B_{\delta D}) = e^{-r\Delta t}d\overline{S}_k(\text{on } B_\delta)$$

Theorem 10 of Chapter 6 then gives

$$\overline{S}_k(\text{on } B_\delta) = e^{-r\Delta t}\overline{S}_k(\text{on } B_\delta)[u\rho_{\delta,U} + d\rho_{\delta,D}]$$

or

$$e^{r\Delta t} = u\rho_{\delta,U} + d\rho_{\delta,D}$$

It follows that $\rho_{\delta,U}$ and $\rho_{\delta,D}$ are independent of δ and we may write

$$\pi_U = \rho_{\delta,U}$$

Making the substitution and solving for ρ_U gives

$$\pi_U = \frac{e^{r\Delta t} - d}{u - d}$$

and

$$1 - \pi_U = \frac{u - e^{r\Delta t}}{u - d}$$

Now, the ordered pair $(\pi_U, 1 - \pi_U)$ is a strongly positive probability distribution if and only if $0 < \pi_U < 1$. In this case, the conditional probabilities depend only on u, d and r and are unique. This implies that the martingale measure is unique and so the model is complete.

The condition $0 < \pi_U < 1$ is equivalent to

$$0 < e^{r\Delta t} - d < u - d$$

or

$$d < e^{r\Delta t} < u$$

Assuming that this is the case, the resulting unique martingale measure Π is given, for any $\omega \in \{U, D\}^T$ by

$$\mathbb{P}_\Pi(\omega) = \pi_U^{N_U(\omega)}(1 - \pi_U)^{T - N_U(\omega)}$$

We now have a very nice theorem describing martingale measures in the CRR model.

Theorem 1 *The Cox-Ross-Rubinstein model is complete and free of arbitrage if and only if*

$$d < e^{r\Delta t} < u$$

In this case, the unique martingale measure Π *on* $\mathbb{M}$ *is defined, for any* $\omega \in \{U, D\}^T$ *by*

$$\mathbb{P}_\Pi(\omega) = \pi_U^{N_U(\omega)}(1 - \pi_U)^{T - N_U(\omega)}$$

where

$$\pi_U = \frac{e^{r\Delta t} - d}{u - d}$$ □

Pricing in the Cox-Ross-Rubinstein Model

Let us assume that $\mathbb{M}$ is a complete CRR model with no arbitrage. Then the replicating strategy procedure can be used to price alternatives. In particular, if $X = (x_1, x_2)$ is an alternative then there is a replicating trading strategy Φ for X and the price of X is

$$\begin{aligned}
\mathcal{I}(X) &= \mathcal{V}_0(\Phi) \\
&= e^{-rL}\mathcal{E}_\Pi(\mathcal{V}_T(\Phi)) \\
&= e^{-rL}\mathcal{E}_\Pi(X) \\
&= e^{-rL}\sum_{\omega\in\Omega} X(\omega)\mathbb{P}_\Pi(\omega) \\
&= e^{-rL}\sum_{\omega\in\Omega} X(\omega)\pi_U^{N_U(\omega)}(1-\pi_U)^{T-N_U(\omega)}
\end{aligned}$$

EXAMPLE 1 A certain stock is currently selling for 100. The feeling is that for each month over the next 2 months, the stock's price will rise by 10% or fall by 10%. Assuming a risk-free rate of 1%, calculate the price of a European call with the various strike prices $K = 102, K = 101, K = 100, K = 99, K = 98$ and $K = 97$.
Solution The parameters of the CRR model are

$$\begin{aligned}
&t_0 = 0, t_1 = \frac{1}{12}, t_2 = \frac{1}{6} \\
&\Delta t = \frac{1}{12}, L = \frac{1}{6}, T = 2 \\
&u = 1.01, d = 0.99
\end{aligned}$$

and

$$\pi_U = \frac{e^{r\Delta t} - d}{u - d} = \frac{e^{(0.01)(1/12)} - 0.99}{0.02} \approx 0.54$$

Figure 3 shows the state tree, with stock prices and local conditional probabilities.

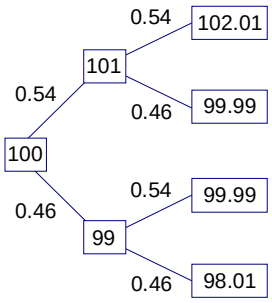


Figure 3 – State tree with conditional probabilities

We quickly check that this model is complete and arbitrage free by checking that $d \leq e^{r\Delta t} \leq u$, which is true since

$$e^{r\Delta t} = e^{(0.01)(1/12)} \approx 1.000833681$$

The payoff for the call is

$$\begin{aligned} X(\omega) &= \max\{S_2(\omega) - K, 0\} \\ &= \begin{cases} \max\{S_0u^2 - K, 0\} & \omega = UU \\ \max\{S_0ud - K, 0\} & \omega = UD \text{ or } DU \\ \max\{S_0d^2 - K, 0\} & \omega = DD \end{cases} \\ &= \begin{cases} \max\{102.01 - K, 0\} & \omega = UU \\ \max\{99.99 - K, 0\} & \omega = UD \text{ or } DU \\ \max\{98.01 - K, 0\} & \omega = DD \end{cases} \end{aligned}$$

so the discounted expected payoff is

$$\begin{aligned} S_0 = e^{-r\Delta t}\Big(& X(UU)\pi_U^2 + \\ & X(UD)\pi_U(1-\pi_U) + \\ & X(DU)\pi_U(1-\pi_U) + \\ & X(DD)(1-\pi_U)^2\Big) \end{aligned}$$

Substituting the values and noting that $X(UD) = X(DU)$ gives

$$S_0 = e^{-(0.01)(1/6)}\Big(\max\{102.01-K,0\}\pi_U^2 + 2\cdot\max\{99.99-K,0\}\pi_U(1-\pi_U) + \max\{98.01-K,0\}(1-\pi_U)^2\Big)$$

Thus, some calculation gives

K	S_0
102	0.0029
101	0.2959
100	0.5888
99	1.3725
98	2.1632
97	3.1615

Another Look at the CRR Model via Random Walks

Let us take a somewhat different perspective on the CRR model. During each time interval $[t_i, t_{i+1}]$ of length Δt the stock price takes either an up-tick or a down-tick. Thus, the individual price movements can be modeled as a sequence E_i of independent Bernoulli random variables where

$$\mathbb{P}(E_i = u) = p$$
$$\mathbb{P}(E_i = d) = 1-p$$

that is

$$E_i = \begin{cases} u & \text{with probability } p \\ d & \text{with probability } q = 1-p \end{cases}$$

where p is the natural probability of an up-tick in the economy. Hence the stock price at the final time t_T is given by

$$S_T = S_0 E_1 \cdots E_T = S_0 e^{\sum \log(E_i)} = S_0 e^{H_T}$$

where

$$H_T = \log\left(\frac{S_T}{S_0}\right) = \sum_{i=1}^{T} \log(E_i)$$

is the **logarithmic growth** of the stock price. Next, we define the constants μ and σ by

$$\mu = \frac{1}{\Delta t}\mathcal{E}(\log E_i) = \frac{1}{\Delta t}(p \log u + q \log d)$$
$$\sigma^2 = \frac{1}{\Delta t}\text{Var}(\log E_i) = \frac{1}{\Delta t} pq(\log u - \log d)^2$$

The significance of these constants will be discussed later, but in any case we can write (since $\sigma \neq 0$)

$$\log E_i = \mu\Delta t + \sigma\sqrt{\Delta t}\left[\frac{\log E_i - \mu\Delta t}{\sigma\sqrt{\Delta t}}\right] = \mu\Delta t + \sigma\sqrt{\Delta t}X_i$$

where the random variables

$$X_i = \frac{\log E_i - \mu\Delta t}{\sigma\sqrt{\Delta t}}$$

are independent Bernoulli random variables with

$$X_i = \begin{cases} \frac{q}{\sqrt{pq}} & \text{with probability } p \\ \frac{-p}{\sqrt{pq}} & \text{with probability } q \end{cases}$$

Hence

$$\mathcal{E}(X_i) = 0$$
$$\text{Var}(X_i) = 1$$

We now have

$$\begin{aligned} H_T &= \sum_{i=1}^{T} \log(E_i) \\ &= \sum_{i=1}^{T} [\mu\Delta t + \sigma\sqrt{\Delta t}X_i] \\ &= \mu L + \sigma\sqrt{\Delta t}\sum_{i=1}^{T} X_i \end{aligned}$$

that is

$$H_T = \mu L + \sigma\sqrt{\Delta t}\sum_{i=1}^{T} X_i$$

This formula expresses the logarithmic growth as a sum of a *deterministic part* μL which is a constant multiple of the lifetime L of

the model and a *random part*

$$\sigma\sqrt{\Delta t}\sum_{i=1}^{T}X_i$$

which is a constant multiple of a sum of independent Bernoulli random variables. Each term X_i describes the movement of the stock price during a subinterval of the model. Finally, the stock price itself is given by

$$S_T = S_0 e^{H_T} = S_0 e^{\mu L + \sigma\sqrt{\Delta t}\sum_{i=1}^{T}X_i}$$

The constant μ is called the **drift** and the constant σ is called the **volatility** of the stock price. These terms will be explained in a moment.

Note that the expression

$$s = \frac{1}{T}\log\left(\frac{S_T}{S_0}\right) = \frac{1}{T}H_T$$

is referred to as the **return** by some authors. The reason is that the equation above is equivalent to

$$S_T = S_0 e^{sT}$$

which shows that the stock price grows at a continuously compounded rate of s. Thus, s is the *rate of return*.

Random Walks

The sequence (X_i) that describes the behavior of the stock price over each subinterval is an example of a *random walk*. To understand random walks, imagine a flea who is constrained to jump along a straight line, say the x-axis. The flea starts at the point $x = 0$ at time $t = 0$ and during each interval of time (of length Δt) jumps randomly a distance a to the right or a distance b to the left. Assume that the probability of a jump to the right is p. This is shown in Figure 4.

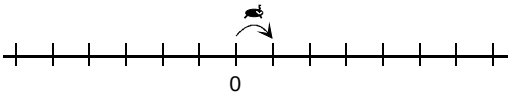


Figure 4 - The random walk of a flea

Each variable X_i in the sequence (X_i) describes a single step in the flea's perambulations and the partial sums

$$U_k = \sum_{i=1}^{k} X_i$$

represent the position of the flea at time t_k.

Figure 5 shows a couple of computer-generated random walks with $p = q = 1/2$ and $a = b$. (These are called *symmetric random walks*.) As is customary in order to see the path clearly, each position of the flea is marked by a point in the plane, where the x-axis represents time and the y-axis represents position.

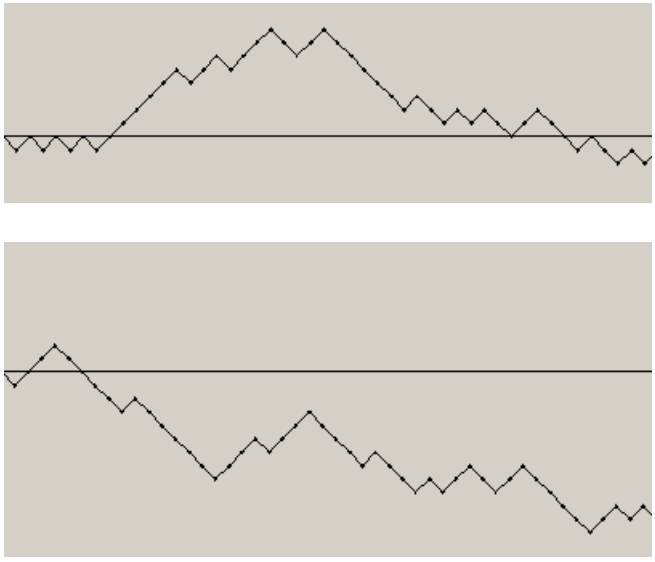


Figure 5 - Random walks

There are many formulations of the random walk scenario, involving for example, drunks who are walking randomly along a street or gambler's playing a game of chance, or the price of a stock. Indeed, entire books have been written on the subject of random walks.

There are also many questions that can be asked about the behavior of a random walk. For example, given integers a and b for which $a < 0 < b$ is it necessarily true that the flea must eventually arrive at one of these

"boundary points" or can the flea oscillate back and forth forever, never reaching either boundary?

Since the answer to the previous question is that the flea must eventually reach one of the boundary points, we can ask about the probability of reaching each of the boundary points and the expected time to reach a boundary. We might also inquire about whether the flea must return to the origin at some time in the future.

In any case, this is not a book on random walks, so let us return to the situation at hand, namely

$$H_T = \mu L + \sigma\sqrt{\Delta t}\sum_{i=1}^{T} X_i$$

The deterministic term μL is a constant multiple of the lifetime of the model and accounts for a steady change (drift) in the stock's price (if $\mu \neq 0$). It is akin to the behavior of the risk-free asset with interest rate μ. The random term is a constant multiple of the position of the random walk.

Let us summarize what we have learned about the CRR model. In a later chapter, we will use this model to derive the famous Black-Scholes option pricing formula.

Theorem 2 *For a CRR model with probability of up-tick equal to p and down-tick equal to $q = 1 - p$, lifetime L and time increments Δt the stock price is given by*

$$S_T = S_0 e^{\mu L + \sigma\sqrt{\Delta t}\sum_{i=1}^{T} X_i}$$

where the **drift** *and* **volatility** *are defined by*

$$\mu = \frac{1}{\Delta t}(p \log u + q \log d)$$
$$\sigma^2 = \frac{1}{\sqrt{\Delta t}} pq(\log u - \log d)^2$$

The random walk portion of the stock movement is given by

$$Y_T = \sum_{i=1}^{T} X_i$$

where the random variables X_i are independent and

$$X_i = \begin{cases} \frac{q}{\sqrt{pq}} & \textit{with probability } p \\ \frac{-p}{\sqrt{pq}} & \textit{with probability } q \end{cases}$$ □

Exercises

1. A certain stock is currently selling for 50. The feeling is that for each month over the next 2 months, the stock's price will rise by 10% or fall by 10%. Assuming a risk-free rate of 1%, calculate the price of a European call with strike price K given by
   a) 52 b) 51 c) 50
   d) 49 e) 48 f) 47
   What about a European put with the same strike price and expiration date?
2. A certain stock is currently selling for 10. The feeling is that for each month over the next 2 months, the stock's price will rise by 5% or fall by 10%. Assuming a risk-free rate of 1%, calculate the price of a European call with strike price K given by
   a) 11 b) 10 c) 9 d) 8
   What about a European put with the same strike price and expiration date?
3. Referring to Example 1 explain why there is a *loss* in all states except the first, that is, there is a loss with probability $3/4$.
4. Show that $\{U, D\}^k$ has size 2^k. *Hint*: use mathematical induction or the fundamental counting principle (also known as the multiplication rule).
5. Show that

$$X_i = \begin{cases} \frac{q}{\sqrt{pq}} & \text{with probability } p \\ \frac{-p}{\sqrt{pq}} & \text{with probability } q \end{cases}$$

6. Show that the two values of a Bernoulli random variable X with $p = \frac{1}{2}$ are given by $\mathcal{E}(X) \pm \sqrt{\text{Var}(X)}$.
7. An alternative X that depends on the final state only through the number of U's in the state is called a **path-independent alternative**. In particular, if $\mathcal{P}$ is the partition of Ω whose blocks are the subsets G_k of Ω that contain exactly k U's

$$G_k = \{\omega \in \Omega \mid N_U(\omega) = k\}$$

then X is path-independent if and only if there are constants X_k for which

$$X_k = X(\text{any } \omega \in G_k)$$

for $k = 0, \ldots, T$.

a) Show that

$$|G_k| = \binom{T}{k}$$

b) Show that the probability (under the martingale measure) of any $\omega \in G_k$ is

$$\pi_U^{N_U(\omega)}(1-\pi_U)^{T-N_U(\omega)} = \pi_U^k(1-\pi_U)^{T-k}$$

c) Show that the probability of G_k is

$$\mathbb{P}_\Pi(G_k) = \binom{T}{k}\pi_U^k(1-\pi_U)^{T-k}$$

d) Show that if X is a path-independent alternative then

$$\mathcal{I}(X) = e^{-rL}\sum_{k=0}^{T} X_k \binom{T}{k}\pi_U^k(1-\pi_U)^{T-k}$$

8. Write a computer program or an Excel spreadsheet to compute the price of a European call under the CRR model where $T = 2$.
9. Verify that

$$\mathcal{E}_p(\log E_i) = p\log u + q\log d$$
$$\text{Var}_p(\log E_i) = pq(\log u - \log d)^2$$

10. In a general discrete-time model, knowledge of the state of the economy at a given time implies knowledge of the asset prices at that time. Why? Is the converse necessarily true? What if at time t_k we know all previous states and asset prices? Support your answer. What happens in the case of the CRR model?

Chapter 8

Probability III: Continuous Probability

In this chapter we discuss some concepts of the general theory of probability, without restriction to finite or discrete sample spaces. This is in preparation for our discussion of the Black-Scholes derivative pricing model.

Since this is not a book in probability and since a detailed discussion of probability would take us too far from our main goals, we will need to be a bit "sketchy" in our discussion. For a more complete treatment of probability, please consult the references at the end of the book.

General Probability Spaces

Let us recall the definition of a finite probability space.

Definition *A* ***finite probability space*** *is a pair* $(\Omega, \mathbb{P})$ *consisting of a finite nonempty set* Ω*, called the* ***sample space*** *and a real-valued function* $\mathbb{P}$ *defined on the set of all subsets of* Ω*, called a* ***probability measure*** *on* Ω*. Furthermore, the function* $\mathbb{P}$ *must satisfy the following properties.*

1) (Range) For all $A \subseteq \Omega$

$$0 \leq \mathbb{P}(A) \leq 1$$

2) (Probability of Ω*)*

$$\mathbb{P}(\Omega) = 1$$

3) (Additivity property) If A *and* B *are* disjoint *then*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$

In this context, subsets of Ω *are called* **events**. □

We have also seen that the additivity property of $\mathbb{P}$ is equivalent to the finite additivity property, that is, if

$$A_1, A_2, \ldots, A_n$$

is a finite sequence of *pairwise disjoint* events then

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n)$$

Now, we would like to generalize this definition to sample spaces of arbitrary size, while preserving as much of the spirit of the current definition as possible. In particular, it is essential that $\mathbb{P}$ not only satisfy the three properties above but also that $\mathbb{P}$ be *countably additive*, that is, if

$$A_1, A_2, \ldots$$

is a *sequence* of *pairwise disjoint* events then

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

where the infinite sum on the right must converge. We must fit as much of the previous definition as possible into the context of countable additivity.

It turns out that this can be done by making only one compromise, namely, not all subsets of the sample space can be considered events. Put another way, it is not in general possible to define a *countably additive* set function on all subsets of an infinite set Ω. We would very much like to give an example to support this statement, but such examples involve more mathematical machinery than is appropriate for this book, so we must ask the reader to take this point on faith.

Given this fact, we need to consider what types of collections of subsets of the sample space can act as the collection of events of a probability measure. This leads us to the concept of a σ-algebra.

Definition *Let Ω be a nonempty set. A nonempty collection Σ of subsets of Ω is a σ-***algebra** *if*

1) $\Omega \in \Sigma$

2) Σ *is closed under* countable *unions, that is, if* $A_1, A_2, \ldots$ *is a sequence of elements of* Σ *then*

$$\bigcup_{i=1}^{\infty} A_i \in \Sigma$$

3) Σ *is closed under complements, that is, if* $A \in \Sigma$ *then* $A^c \in \Sigma$.□

Note that $\emptyset = \Omega^c \in \Sigma$. Also, DeMorgan's laws show that Σ is closed under countable intersections. (We leave details as an exercise.)

Definition *A **measurable space** is a pair* (Ω, Σ) *consisting of a nonempty set* Ω *and a* σ*-algebra* Σ *of subsets of* Ω.□

Now we can define a general probability space.

Definition *A **probability space** is a triple* $(\Omega, \Sigma, \mathbb{P})$ *consisting of a nonempty set* Ω*, called the **sample space**, a* σ*-algebra* Σ *of subsets of* Ω *whose elements are called* **events** *and a real-valued function* $\mathbb{P}$ *defined on* Σ *called a **probability measure**. The function* $\mathbb{P}$ *must satisfy the following properties.*

1) (Range) For all $A \subseteq \Omega$

$$0 \leq \mathbb{P}(A) \leq 1$$

2) (Probability of Ω*)*

$$\mathbb{P}(\Omega) = 1$$

3) (Countable additivity property)

$$A_1, A_2, \ldots$$

is a sequence of pairwise disjoint *events then*

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

A very useful property of probability measures is given in the following theorem. A **decreasing sequence** of events is a sequence of events satisfying

$$A_1 \supseteq A_2 \supseteq \cdots$$

Similarly, an **increasing sequence** of events is a sequence of events satisfying

$$A_1 \subseteq A_2 \subseteq \cdots$$

Theorem 1 *Probability measures are* **monotonically continuous** *in the following sense.*

1) If $A_1 \supseteq A_2 \supseteq \cdots$ *is a decreasing sequence of events then*

$$\lim_{i \to \infty} \mathbb{P}(A_i) = \mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right)$$

2) If $A_1 \subseteq A_2 \subseteq \cdots$ is an increasing *sequence of events then*

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

Proof. For part 1), suppose that $A_1 \supseteq A_2 \supseteq \cdots$. Then the sequence $\mathbb{P}(A_i)$ of probabilities is a nonincreasing sequence of real numbers bounded below by 0. It is a theorem of elementary real analysis that such a sequence must converge, so the limit in question does exist.

For convenience, let $A = \bigcap_{i=1}^{\infty} A_i$. We first consider the events

$$A_1 \setminus A_2, A_2 \setminus A_3, \ldots$$

These events are disjoint, since if $i < j$ then $i + 1 \leq j$ and so $A_j \subseteq A_{i+1}$ hence if

$$a \in (A_i \setminus A_{i+1}) \cap (A_j \setminus A_{j+1})$$

then a would be in A_j but not in the superset A_{i+1}. Also, each of these events is disjoint from the intersection A. Thus, A_1 is the *disjoint* union

$$A_1 = \left(\bigcup_{i=1}^{\infty} (A_i \setminus A_{i+1}) \right) \cup A$$

For if $a \in A_1$ then if $a \notin A$ then we can let $i + 1$ be the first index for which $a \notin A_{i+1}$. It follows that $a \in A_i \setminus A_{i+1}$.

Now we can apply countable additivity to get

$$\begin{aligned}
\mathbb{P}(A_1) &= \mathbb{P}\left(\bigcup_{i=1}^{\infty}(A_i \setminus A_{i-1})\right) + \mathbb{P}(A) \\
&= \sum_{i=1}^{\infty} \mathbb{P}(A_i \setminus A_{i-1}) + \mathbb{P}(A) \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(A_i \setminus A_{i-1}) + \mathbb{P}(A) \\
&= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{n}(A_i \setminus A_{i-1})\right) + \mathbb{P}(A) \\
&= \lim_{n\to\infty} \mathbb{P}(A_1 \setminus A_{n-1}) + \mathbb{P}(A) \\
&= \lim_{n\to\infty} [\mathbb{P}(A_1) - \mathbb{P}(A_{n-1})] + \mathbb{P}(A) \\
&= \mathbb{P}(A_1) - \lim_{n\to\infty} \mathbb{P}(A_{n-1}) + \mathbb{P}(A)
\end{aligned}$$

Thus

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} \mathbb{P}(A_{n-1}) = \mathbb{P}(A)$$

as desired. We leave proof of part 2) as an exercise.□

Probability Measures on $\mathbb{R}$

The most important sample space from the point of view of both theory and applications is the real line $\mathbb{R}$. In fact, the only nonfinite probability space that we will need to consider in this book is $\mathbb{R}$.

The most important σ-algebra on $\mathbb{R}$ is the so-called **Borel σ-algebra** $\mathcal{B}$. A formal definition of the Borel σ-algebra is simple to state, if not quite as simple to comprehend.

Definition The **Borel σ-algebra** $\mathcal{B}$ on $\mathbb{R}$ is the smallest σ-algebra on $\mathbb{R}$ that contains all open intervals (a, b) where $a, b \in \mathbb{R}$. □

Let us examine this definition. First, we must show that there *is* such a σ-algebra. After all, just because we use a phrase such as “the smallest set ...” doesn't mean that there is such a set.

The usual procedure for showing that there is a *smallest* set with some property is to show two things: first, that there is at least one set with the desired property and second that the intersection of any collection of sets

with the desired property also has the desired property. It follows that the intersection of all sets with the desired property exists and is the smallest set with that property.

For the case at hand, it is easy to see that there is at least one σ-algebra on $\mathbb{R}$ containing the open intervals: it is the collection of all subsets of $\mathbb{R}$. Second, it is not hard to see that the intersection of σ-algebras is also a σ-algebra. We leave the details to the reader. Hence, the Borel σ-algebra does indeed exist and is the intersection of all σ-algebras that contain the open intervals.

Note that while we have established the existence of the Borel σ-algebra, its description as the intersection of all σ-algebras that contain the open intervals is not very practical. From a practical perspective, it is more useful to consider some examples of elements of $\mathcal{B}$, that is, of **Borel sets**.

Theorem 2
1) All open, closed and half-open intervals are Borel sets.
2) All rays $(-\infty, b]$, $(-\infty, b)$, $[a, \infty)$ *and* (a, ∞) *are Borel sets.*
3) All open sets and all closed sets are Borel sets.
Proof. We sketch the proof. For 1), to see that the half-open interval $(a, b]$ is a Borel set observe that

$$(a, b] = \bigcap_{n=1}^{\infty} (a, b + \frac{1}{n})$$

and so $(a, b]$ is the countable union of open intervals and is therefore in $\mathcal{B}$.

For 3) let us briefly discuss open sets in $\mathbb{R}$. A subset A of $\mathbb{R}$ is **open** if for every $x \in A$ there is an open interval (a, b) for which

$$x \in (a, b) \subseteq A$$

A set is **closed** if its complement is open. Let A be an open set in $\mathbb{R}$. Then A is the union of all open intervals contained within A. In fact, A is the union of all maximal open intervals in A. An open interval I in A is **maximal** in A if no open interval containing I as a proper subset is also in A.

Now, we claim that any two maximal open intervals are disjoint and that there are at most a countable number of maximal open intervals. As to the former, any two distinct maximal open intervals contained in A must

be disjoint, for otherwise their union would be a strictly larger open interval contained in A. As a result, each maximal open interval in A contains a *distinct* rational number and since there are only a countable number of rational numbers, there are at most a countable number of maximal open intervals containing A.

Hence, A is the union of at most a countable number of open intervals and is therefore a Borel set.

Finally, since all open sets are Borel sets and since a closed set is the complement of an open set, all closed sets are also Borel sets.□

At first, the more one thinks about Borel sets, the more one comes to feel that all subsets of $\mathbb{R}$ are Borel sets. However, this is not the case. However, it is true that most "nonpathological" sets are Borel sets. Put another way, it is very hard (but not impossible) to describe a set that is not a Borel set. We must reluctantly ask the reader to take it on faith that there exist subsets of $\mathbb{R}$ that are not Borel sets.

From now on, the phrase "let $\mathbb{P}$ be a probability measure on $\mathbb{R}$" will carry with it the tacit understanding that the σ-algebra involved is the Borel σ-algebra.

Theorem 3 *A probability measure on $\mathbb{R}$ is uniquely determined by its values on the rays $(-\infty, t]$. That is, if $\mathbb{P}$ and $\mathbb{Q}$ are probability measures on $\mathbb{R}$ and*

$$\mathbb{P}((-\infty, t]) = \mathbb{Q}((-\infty, t])$$

for all $t \in \mathbb{R}$ then $\mathbb{P} = \mathbb{Q}$.
Proof. Since for $s \leq t$

$$(s, t] = (-\infty, t] \setminus (-\infty, s]$$

we deduce that

$$\mathbb{P}((s, t]) = \mathbb{Q}((s, t])$$

Since any open interval (s, t) is the union of an increasing sequence of half-open intervals

$$(s, t) = \bigcup_{n=1}^{\infty} (s, t - \frac{1}{n}]$$

the monotone continuity of probability measures implies that

$$\mathbb{P}((s,t)) = \lim_{n\to\infty} \mathbb{P}\left((s, t - \frac{1}{n}]\right) = \lim_{n\to\infty} \mathbb{Q}\left((s, t - \frac{1}{n}]\right) = \mathbb{Q}((s,t))$$

Thus, $\mathbb{P}$ and $\mathbb{Q}$ agree on open intervals. Now, it is possible to show that if $\mathbb{P}$ and $\mathbb{Q}$ agree on open intervals then they agree on all Borel sets. However, we will omit this part of the proof, since it requires additional concepts (such as monotone classes) that would lead us too far from our goals.□

Distribution Functions

For finite (or discrete) probability spaces, probability measures are most easily described via their mass functions

$$f(\omega) = \mathbb{P}(\{\omega\})$$

However, the concept of a mass function is not general enough to describe all possible probability measures on the real line, let alone on arbitrary sample spaces. For this, we need the concept of a *probability distribution function*.

Definition A (**probability**) **distribution function** is the function $F: \mathbb{R} \to \mathbb{R}$ with the following properties.

1) F *is nondecreasing, that is,*

$$s < t \Rightarrow F(s) \leq F(t)$$

(Note that some authors use the term increasing *for this property.)*

3) F *is right-continuous, that is, the right-hand limit exists everywhere and*

$$\lim_{t\to a} F(t) = F(a)$$

4) F *satisfies*

$$\lim_{t\to-\infty} F(t) = 0$$
$$\lim_{t\to\infty} F(t) = 1$$

□

Figure 1 shows the graph of a probability distribution function. Note that the function is nondecreasing, right continuous (but not continuous) and has the appropriate limits at $\pm\infty$.

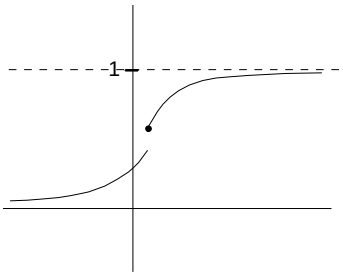


Figure 1 - A probability distribution function

The extreme importance of probability distribution functions is given in the next theorem. Basically, it implies that there is a one-to-one correspondence between probability measures on $\mathbb{R}$ and probability distribution functions. Thus, knowing one uniquely determines the other and so the two concepts are essentially equivalent. We will omit the proof of this theorem.

Theorem 4

1) Let $\mathbb{P}$ be a probability measure on $\mathbb{R}$. The function $F_{\mathbb{P}}: \mathbb{R} \to \mathbb{R}$ defined by

$$F_{\mathbb{P}}(t) = \mathbb{P}((-\infty, t])$$

is a probability distribution function, called the **distribution function** *of $\mathbb{P}$.*

2) Let $F: \mathbb{R} \to \mathbb{R}$ be a distribution function. Then there is a unique probability measure $\mathbb{P}_F$ on $\mathbb{R}$ whose distribution function is F, that is

$$\mathbb{P}_F((-\infty, t]) = F(t) \qquad \square$$

Suppose we begin with a probability measure $\mathbb{P}$, take its distribution function $F_{\mathbb{P}}$ and then form the probability measure $\mathbb{Q}$ of $F_{\mathbb{P}}$. According to the definitions,

$$\mathbb{Q}((-\infty, t]) = F_{\mathbb{P}}(t) = \mathbb{P}((-\infty, t])$$

and so $\mathbb{P}$ and $\mathbb{Q}$ agree on the rays $(-\infty, t]$. We have seen that this implies that $\mathbb{P} = \mathbb{Q}$. Consequently, the correspondence

$$\mathbb{P} \to F_{\mathbb{P}}$$

from probability measures to distribution functions and the correspondence

$$F \to \mathbb{P}_F$$

from distribution functions to probability measures are one-to-one and are inverses of one another. This establishes the fact that the notions of probability measure and distribution function are equivalent.

EXAMPLE 1 Perhaps the simplest probability measures on $\mathbb{R}$ are those that convey the notion of "equal likelihood" or "uniform probability" over an interval $[a, b]$ of $\mathbb{R}$. For example, consider the closed unit interval $[0, 1]$.

How do we convey the notion that each outcome in $[0, 1]$ is somehow equally likely? In the finite case, say of a sample space $\{1, \ldots, n\}$, we simply assign the same probability $1/n$ to each elementary event $\{k\}$. However, unlike the finite case, it is not possible to assign a positive real number p to each elementary event $\{r\}$ for all $r \in [0, 1]$ because there are an infinite number of elementary events and so the sum of their probabilities is not finite, let alone equal to 1. We must accept the fact that the probability of each elementary event is 0 and turn to more complex Borel sets.

First we observe that if B is a Borel set then that portion of B that lies outside of the interval $[0, 1]$ should not contribute anything to the probability. In other words,

$$\mathbb{P}(B \cap [0, 1]^c) = 0$$

As for the rest of B, that is, the set $B \cap [0, 1]$, the notion of uniform probability suggests that $\mathbb{P}(B \cap [0, 1])$ should be proportional to the "length" of $B \cap [0, 1]$, whatever that means.

For intervals, the concept of length is well defined

$$\text{len}([a, b]) = b - a$$

Thus, it seems reasonable to define

$$\mathbb{P}((-\infty, t]) = \mathbb{P}([0, t]) = \begin{cases} 0 & t < 0 \\ t & 0 \le t \le 1 \\ 1 & t > 1 \end{cases}$$

This is the **uniform distribution function on** $[0, 1]$. Figure 2 shows the graph of this distribution function.□

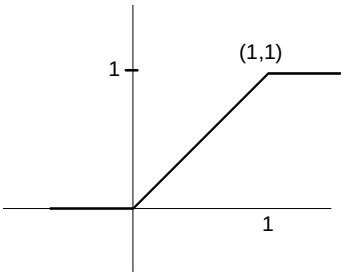


Figure 2 – The uniform distribution function for $[0, 1]$

EXAMPLE 2 The most important of all probability distributions is the **normal distribution**, whose distribution function is

$$\phi_{\mu,\sigma}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{t} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

This is quite a complicated function, but there is not much we can do about it. Nature does not always make our lives easy with simple formulas. Figure 3 shows the normal distribution function.

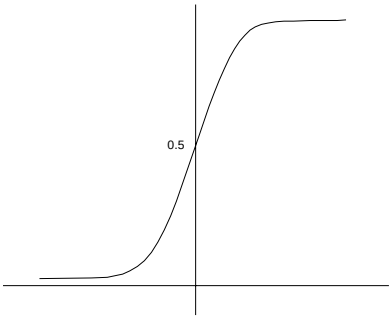


Figure 3 – The normal distribution function

The parameters μ and σ^2 are called the mean (expected value) and variance, respectively. The **standard normal distribution** is the normal distribution with mean 0 and variance 1 and thus has distribution function

$$\phi_{0,1}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx$$

The reason that the normal distribution is considered the most important goes beyond the fact that it appears often in applications. Actually, it is the *reason* why it appears so often in applications that is the key. This

reason is expressed mathematically by the most famous theorem in probability—the *central limit theorem*. We will discuss this theorem later in the chapter.□

Note that the uniform and the normal distribution functions are both continuous not just right-continuous. Put another way, their graphs have no jumps. A jump in the distribution function indicates a point at which the probability is not 0. Let us illustrate with an example.

EXAMPLE 3 A public drug manufacturing company has a new drug that is awaiting FDA approval. If the drug is approved, the company estimates that its stock will end trading that day somewhere in the range $[10, 15]$, with each price being equally likely. However, if the drug is not approved, the stock price will likely be 5. Let us assume that the probability of approval is 0.75.

We could model this situation with the sample space $\Omega = \{5\} \cup [10, 15]$ but it may be simpler to use the sample space $\mathbb{R}$ and simply assign a 0 probability outside of the set Ω. The distribution function for this probability measure is

$$F(t) = \begin{cases} 0 & t < 5 \\ 0.25 & 5 \le t < 10 \\ 0.25 + 0.75\left(\frac{t-10}{5}\right) & 10 \le t < 15 \\ 1 & t \ge 15 \end{cases}$$

The graph is shown in Figure 4. Note the jump at $t = 5$.□

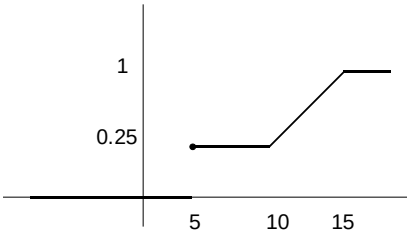


Figure 4 – A distribution function with a jump

Density Functions

The distribution function of a probability measure is extremely important, but it is not always the simplest way to describe a probability measure. Many probability measures that occur in applications have the

property that their distribution functions are differentiable and that the derivatives are very “well-behavied”.

By well-behavied, we mean that the derivative of the distribution function F can be integrated and the integral is again equal to F. (There are functions that have derivatives that are integrable, but the integral of the derivative is not the original function.) This can be expressed in symbols as follows

$$F(t) = \int_{-\infty}^{t} F'(x)\,dx$$

The function $f(x) = F'(x)$ is called a *density function* for $\mathbb{P}$. Let us have a formal definition.

Definition *A probability measure $\mathbb{P}$ or equivalently a distribution function $F_{\mathbb{P}}$ is* **absolutely continuous** *if it has a* **density function**, *which is a nonnegative function $f\colon \mathbb{R} \to \mathbb{R}$ for which*

$$F_{\mathbb{P}}(t) = \int_{-\infty}^{t} f(x)\,dx \qquad \square$$

From this definition, it follows that

$$\mathbb{P}((a,b]) = \int_{a}^{b} f(x)\,dx$$

In other words, the probability of the interval $(a,b]$ is the *area under graph of the density function* from a to b.

Note that a density function must be nonnegative and satisfy

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

that is, the area under the entire graph of f over the entire x-axis must be equal to 1. In fact, any nonnegative function f with this property is a density function for some probability measure.

Probability measures that have density functions, that is, absolutely continuous probability measures, are special. For example, their distribution functions are continuous (not just right-continuous). Thus, there are no points that have positive probability, as happened in a previous example.

EXAMPLE 4 The uniform distribution function on $[0,1]$ is absolutely continuous, with density function

$$f(x) = \begin{cases} 0 & x \notin [0,1] \\ 1 & x \in [0,1] \end{cases}$$

The graph of f is shown in Figure 5.□

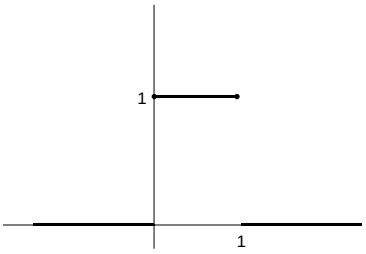


Figure 5 – The uniform density function on $[0,1]$

EXAMPLE 5 The normal distribution is absolutely continuous, with density function

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

The density of the standard normal distribution is

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

This is pictured in Figure 6. The graph of the normal density function is the oft-spoken-of *bell shaped curve*.□

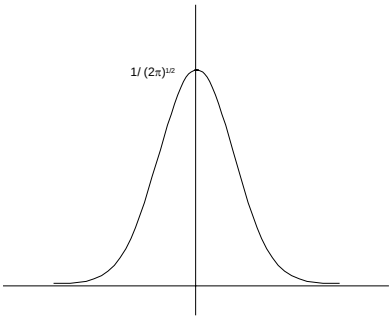


Figure 6 – The standard normal density function

Types of Probability Measures on $\mathbb{R}$

Probability measures on $\mathbb{R}$ can be classified into the following groups: finite, discrete, absolutely continuous, singular continuous and mixed. Let us take a quick look at each of these groups.

Finite Probability Measures

A probability measure $\mathbb{P}$ on $\mathbb{R}$ is **finite** if there are a finite number of real numbers $\{r_1, \ldots, r_n\}$ for which

$$\sum_{i=1}^{n} \mathbb{P}(r_i) = 1$$

Put another way, all of the probability is concentrated in a finite number of points. A finite probability measure can be described by its **probability mass function**, which is 0 everywhere except at the points of positive probability

$$f(x) = \begin{cases} \mathbb{P}(r_i) & x = r_i \\ 0 & \text{otherwise} \end{cases}$$

This is also referred to as the **density function** of $\mathbb{P}$. The distribution function of a finite probability measure has a finite number of jumps and is constant everywhere else. Figure 7 illustrates.

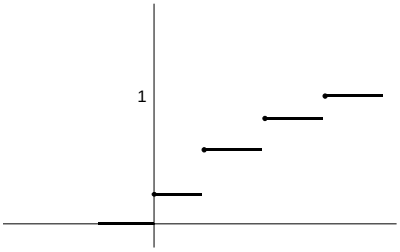


Figure 7 – The distribution function of a finite probability measure

Discrete Probability Measures

A probability measure $\mathbb{P}$ on $\mathbb{R}$ is **discrete** if there are a countable number of real numbers $\{r_1, r_2, \ldots\}$ for which

$$\sum_{i=1}^{\infty} \mathbb{P}(r_i) = 1$$

Put another way, all of the probability is concentrated in a countable number of points. (By *countable* we mean finite or countably infinite.

Hence, a finite probability measure is a discrete probability measure.) As with finite probability measures, a general discrete probability measure can be described by a probability mass function, although the term density function is more common in this case

$$f(x) = \begin{cases} \mathbb{P}(r_i) & x = r_i \\ 0 & \text{otherwise} \end{cases}$$

The distribution function of a discrete probability measure can actually be quite complex. While it is true that the function has only a countable number of jumps, these jumps can occur at sets, such as the the set of all rational numbers, that are spread "uniformly" throughout the real line.

Absolutely Continuous Probability Measures

As we have seen, a probability measure $\mathbb{P}$ on $\mathbb{R}$ is **absolutely continuous** if it has a density function, that is, a nonnegative function $f\colon \mathbb{R} \to \mathbb{R}$ for which

$$F_{\mathbb{P}}(t) = \int_{-\infty}^{t} f(x)\, dx$$

Singular Continuous Probability Measures

Singular continuous probability measures are definitely pathological in nature. A **singular continuous** (or just **singular**) probability measure is one whose distribution function is differentiable (and hence continuous) but whose derivative is 0 on "almost" the entire real line (all except a set of probability 0). Fortunately, we do not need to deal with such pathological probability measures in this book.

Mixed Probability Measures

It is a fact that any probability measure $\mathbb{P}$ on $\mathbb{R}$ can be decomposed (in a unique way) into a linear combination of a discrete (including finite), an absolutely continuous and a singular continuous probability measure, in symbols

$$\mathbb{P} = \alpha_d \mathbb{P}_d + \alpha_a \mathbb{P}_a + \alpha_s \mathbb{P}_s$$

where the coefficients α_d, α_a and α_s are nonnegative and satisfy $\alpha_d + \alpha_a + \alpha_s = 1$. Thus, all probability measures are either discrete, absolutely continuous, singular continuous or a (convex) combination of these types.

Random Variables

Just as the issue of events is more complex in the nonfinite case, so is the notion of random variable. In particular, not all functions are random variables.

Definition *Let* (Ω, Σ) *be a measurable space. A function* $X: \Omega \to \mathbb{R}$ *is* Σ**-measurable** *if the inverse image of every open interval is in* Σ*, in symbols*

$$X^{-1}((a, b)) \in \Sigma$$

A measurable function on (Ω, Σ) *is also called a* **random variable**.□

This definition says that a random variable X has the property that the set $X^{-1}((a, b))$ must be "measurable."

Here are a few facts about random variables, whose proofs we omit.

Theorem 5

1) *The sum and product of random variables are random variables, as is any constant multiple of a random variable.*
2) *The composition of random variables is a random variable.*
3) *Continuous and piecewise continuous functions are random variables.*□

The Distribution Function of a Random Variable

If $(\Omega, \Sigma, \mathbb{P})$ is an arbitrary sample space and X is a random variable on (Ω, Σ) then X defines a distribution function F_X and a corresponding probability measure $\mathbb{P}_X$ on $\mathbb{R}$ by

$$F_X(t) = \mathbb{P}_X((-\infty, t]) = \mathbb{P}(X \leq t)$$

This can be proved by showing that the function $F(t) = \mathbb{P}(X \leq t)$ is a distribution function. If $\mathbb{P}_X$ is finite, discrete or absolutely continuous then we say that the *random variable* is **finite**, **discrete** or **absolutely continuous**, respectively. Absolutely continuous random variables are often simply called **continuous random variables**.

The σ-Algebra Generated by a Random Variable

If $X: (\Omega, \Sigma) \to \mathbb{R}$ is a random variable then the inverse image of any Borel set is in Σ. However, it is not required that all elements of Σ are inverse images of X. Those elements of Σ that are inverse images form another σ-algebra that is a *sub σ-algebra of* Σ.

Definition *The* **σ-algebra generated by** *a random variable* $X\colon(\Omega,\Sigma)\to\mathbb{R}$ *is the* σ*-algebra* $\sigma(X)$ *whose elements are the inverse images of the* subsets *of the Borel sets in* $\mathbb{R}$*, that is*

$$\sigma(X)=\{\{X\in B\}\mid B\in\mathcal{B}\}$$ □

The σ-algebra $\sigma(X)$ has a unique property, namely, it is the smallest σ-algebra of Ω under which X is measurable. In loose terms, it is just what is needed and no more to make X measurable. The following theorem is little more than the definition of measurability.

Theorem 6 *Let* $X\colon\Omega\to\mathbb{R}$ *be a function and let* Σ *be a* σ*-algebra on* Ω*. Then* X *is* Ω*-measurable if and only if* Σ *contains* $\sigma(X)$.□

Independence of Random Variables

Here is the definition of independence of arbitrary random variables.

Definition *Two random variables* X *and* Y *on* $\mathbb{R}$ *are* **independent** *if*

$$\mathbb{P}(X\le t,Y\le s)=\mathbb{P}(X\le t)\mathbb{P}(Y\le s)$$

for all $s,t\in\mathbb{R}$*. More generally, a collection* $X_1,\dots,X_n$ *of random variables is* **independent** *if*

$$\mathbb{P}(X_1\le t_1,\dots,X_n\le t_n)=\prod_{i=1}^{n}\mathbb{P}(X_i\le t_i)$$ □

This definition expresses formally the feeling that if random variables are independent then the value of one random variable does not effect the value of another.

Expectation and Variance of a Random Variable

Recall that for a random variable X on a finite probability space $(\Omega,\mathbb{P})$ with $\Omega=\{\omega_1,\dots,\omega_n\}$, the expected value (or mean) is defined by

$$\mathcal{E}_{\mathbb{P}}(X)=\sum_{i=1}^{n}X(\omega_i)\mathbb{P}(\omega_i)$$

If $g\colon\mathbb{R}\to\mathbb{R}$ is a function then the expected value of the random variable $g(X)$ is

$$\mathcal{E}_{\mathbb{P}}(g(X)) = \sum_{i=1}^{n} g(X(\omega_i))\mathbb{P}(\omega_i)$$

Also, the variance is defined by

$$\text{Var}(X) = \mathcal{E}((X-\mu)^2)$$

Let us now extend these concepts to absolutely continuous random variables.

Definition *Let X be an absolutely continuous random variable, with density function f. The* **expected value** *or* **mean** *of X is the improper integral*

$$\mathcal{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

which exists provided that

$$\int_{-\infty}^{\infty} |x| f(x)\, dx < \infty$$

The **variance** *of X is*

$$\text{Var}(X) = \mathcal{E}((X-\mu)^2)$$

and the **standard deviation** *is the positive square root of the variance*

$$SD(X) = \sqrt{\text{Var}(X)}$$ □

Also, if $g: \mathbb{R} \to \mathbb{R}$ is a measurable function then the random variable $g(X)$ has expected value

$$\mathcal{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

provided that

$$\int_{-\infty}^{\infty} |g(x)| f(x)\, dx < \infty$$

Here are some basic properties of expectation and variance.

Theorem 7

1) The expected value operator is linear, that is,

$$\mathcal{E}(aX + bY) = a\mathcal{E}(X) + b\mathcal{E}(Y)$$

2) If $X_1, \ldots, X_n$ are independent *random variables on $\mathbb{R}$ then*

$$\mathcal{E}(X_1 \cdots X_n) = \prod_{i=1}^{n} \mathcal{E}(X_i)$$

3) $\text{Var}(X) = \mathcal{E}(X^2) - \mu^2 = \mathcal{E}(X^2) - \mathcal{E}(X)^2$

4) For any real number a

$$\text{Var}(aX) = a^2 \text{Var}(X)$$

and

$$\text{Var}(X - a) = \text{Var}(X)$$

5) If $X_1, \ldots, X_n$ are independent *random variables on $\mathbb{R}$ then*

$$\text{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \text{Var}(X_i)$$ □

The Normal Distribution

Let us take another look at the normal distribution, whose density function is

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

We mentioned that the parameters μ and σ^2 are the mean and variance, respectively. To calculate the mean, we need only a bit of first-year calculus. The definition is

$$\mathcal{E} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Writing $x = (x - \mu) + \mu$ and splitting the integral gives

$$\mathcal{E} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x - \mu) e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \frac{\mu}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

The second integral is just μ times the integral of $N_{\mu,\sigma}$ and since this integral is 1, we just get μ. As to the first integral, the substituting $y = x - \mu$ gives

$$\int_{-\infty}^{\infty} y e^{-\frac{y^2}{2\sigma^2}} dy$$

But the integrand $ye^{-\frac{y^2}{2\sigma^2}}$ is an odd function, from which it follows that the integral from $-\infty$ to ∞ must be 0. (We leave elaboration of this as an exercise.) Hence, $\mathcal{E} = 0 + \mu = \mu$.

Computation of the variance of the normal distribution requires the beautiful but nontrivial integral formula

$$\int_{-\infty}^{\infty} y^2 e^{-\frac{y^2}{2}}\, dy = \sqrt{2\pi}$$

From here, the rest is straightforward, especially using the formula

$$\text{Var}(X) = \mathcal{E}(X^2) - \mathcal{E}(X)^2$$

The upshot is that $\text{Var}(N_{\mu,\sigma}) = \sigma^2$. We leave the details as an exercise.

Suppose that $\mathcal{N}_{\mu,\sigma}$ is a normal random variable with mean μ and variance σ^2. Consider the random variable

$$Z = \frac{\mathcal{N}_{\mu,\sigma} - \mu}{\sigma}$$

In view of the properties of expectation and variance,

$$\mathcal{E}(Z) = \frac{1}{\sigma}\mathcal{E}(\mathcal{N}_{\mu,\sigma} - \mu) = \frac{1}{\sigma}(\mathcal{E}(\mathcal{N}_{\mu,\sigma}) - \mu) = 0$$

and

$$\text{Var}(Z) = \frac{1}{\sigma^2}\text{Var}(\mathcal{N}_{\mu,\sigma} - \mu) = \frac{1}{\sigma^2}\text{Var}(\mathcal{N}_{\mu,\sigma}) = 1$$

To compute the distribution of Z we have

$$\begin{aligned}
\mathbb{P}(Z \leq t) &= \mathbb{P}(\frac{\mathcal{N}_{\mu,\sigma} - \mu}{\sigma} \leq t) \\
&= \mathbb{P}(\mathcal{N}_{\mu,\sigma} \leq \sigma t + \mu) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\sigma t + \mu} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx
\end{aligned}$$

The substitution $y = (x - \mu)/\sigma$ gives

$$\mathbb{P}(Z \leq t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx$$

and so $Z = \mathcal{N}_{0,1}$ is a standard normal random variable. The process of going from X to Z is called **standardization**.

Theorem 8 *If $\mathcal{N}_{\mu,\sigma}$ is a normal random variable with mean μ and variance σ^2 then*

$$\mathcal{N}_{0,1} = \frac{\mathcal{N}_{\mu,\sigma} - \mu}{\sigma}$$

is a standard normal random variable. Similarly, if $\mathcal{N}_{0,1}$ is a standard normal random variable then

$$\mathcal{N}_{\tau,\sigma} = \sigma \mathcal{N}_{0,1} + \mu$$

is a normal random variable with mean μ and variance σ^2. □

A distribution related to the normal distribution that we will have use for is the *lognormal* distribution. If a random variable X has the property that its logarithm $\log X$ is normally distributed, then the random variable X is said to have a **lognormal distribution**. (Note that X is lognormal if its logarithm is normal, *not* if it is the logarithm of a normal random variable. In other words, lognormal means "log*is*normal" not "log*of*normal".)

Proof of the following is left as an exercise.

Theorem 9 *If X is lognormally distributed, say $Y = \log X$ is normal with mean a and variance b^2 then*

$$\mathcal{E}(X) = \mathcal{E}(e^Y) = e^{a+\frac{1}{2}b^2}$$
$$\text{Var}(X) = \text{Var}(e^Y) = e^{2a+b^2}(e^{b^2} - 1)$$

□

Convergence in Distribution

You may be familiar with the notion of pointwise convergence of a sequence of functions. In any case, here is the definition.

Definition *Let (f_n) be a sequence of functions from $\mathbb{R}$ to $\mathbb{R}$ and let f be another such function. Then (f_n)* **converges pointwise** *to f if for each*

real number x, the sequence of real numbers $(f_n(x))$ converges to the real number $f(x)$.□

If you are familiar with convergence of ordinary sequences of real numbers, then you are essentially familiar with pointwise convergence of functions. There is very little new here.

Now consider a sequence (X_n) of random variables. Of course, these random variables are functions, albeit special kinds of functions. Let X be another random variable. It turns out that there are several useful ways in which the notion of *convergence* of the sequence (X_n) to X can be defined (only one of which is pointwise convergence). However, we are interested in one particular form of convergence. Here is the definition.

Definition *Let (X_n) be a sequence of random variables, where we allow the possibility that each random variable may be defined on a different probability space $(\Omega_n, \mathbb{P}_n)$. Let X be a random variable on a probability space $(\Omega, \mathbb{P})$. Then (X_n)* **converges in distribution** *to X, written*

$$X_n \xrightarrow{dist} X$$

if the distribution functions (F_{X_n}) converge pointwise to the distribution function F_X at all points where F_X is continuous. Thus, if F_X is continuous at s then we must have

$$\lim_{n\to\infty} F_{X_n}(s) = F_X(s)$$

that is

$$\lim_{n\to\infty} \mathbb{P}_n(X_n \leq s) = \mathbb{P}(X \leq s)$$

Convergence in distribution is also called **weak convergence**.□

We need the following results about weak convergence.

Theorem 10 *Let (X_n) be a sequence of random variables where X_n is defined on $(\Omega_n, \mathbb{P}_n)$. Let X be a random variable defined on $(\Omega, \mathbb{P})$.*

1) We have $X_n \xrightarrow{dist} X$ if and only if

$$\mathcal{E}_{\mathbb{P}_n}(g(X_n)) \to \mathcal{E}_{\mathbb{P}}(g(X))$$

for all bounded *continuous functions* $g\colon \mathbb{R} \to \mathbb{R}$. *In particular,*

$$\mathcal{E}_{\mathbb{P}_n}(X_n) \to \mathcal{E}_{\mathbb{P}}(X)$$

2) For all continuous functions $f\colon \mathbb{R} \to \mathbb{R}$

$$f(X_n) \xrightarrow{dist} f(X)$$

Proof. We will omit the proof of part 1). As to part 2), let f be continuous. Then for any bounded continuous function g the composition $g \circ f$ is also bounded and continuous. Hence, by part 1),

$$\mathcal{E}(g(f(X_n))) = \mathcal{E}((g \circ f)(X_n)) \to \mathcal{E}((g \circ f)(X)) = \mathcal{E}(g(f(X)))$$

Part 1) then implies that

$$f(X_n) \xrightarrow{\text{dist}} f(X)$$

as desired.□

Theorem 11 *Let* (X_n) *be a sequence of random variables with*

$$X_n \xrightarrow{dist} X$$

where X *is a random variable whose distribution function is continuous. If* (a_n) *and* (b_n) *are sequences of real numbers for which*

$$a_n \to a,\ b_n \to b$$

then

$$a_n X_n + b_n \xrightarrow{dist} aX + b$$

In particular, if $a \neq 0$ *and* $X_n \xrightarrow{dist} \mathcal{N}_{0,1}$ *where* $\mathcal{N}_{0,1}$ *is a standard normal random variable then*

$$a_n X_n + b_n \xrightarrow{dist} \mathcal{N}_{a,b}$$

where $\mathcal{N}_{a,b}$ *is a normal random variable with mean* a *and variance* b^2.

Proof. The following proof requires the concepts of uniform convergence. The reader may omit this proof if these concepts are not familiar. Let F_{X_n} and F_X denote the distribution functions of X_n and X, respectively.

The first step is to show that for any $s \in \mathbb{R}$ there is an interval $(s - \lambda, s + \lambda)$ in which F_{X_n} converges to F_X *uniformly*. For this, we use

the fact that distribution functions are *nondecreasing*. So let $\epsilon > 0$ be given and write

$$\begin{aligned} F_{X_n}(t) - F_X(t) &\leq F_{X_n}(s+\alpha) - F_X(s-\alpha) \\ &= [F_{X_n}(s+\alpha) - F_X(s+\alpha)] + [F_X(s+\alpha) - F_X(s-\alpha)] \end{aligned}$$

Since F_{X_n} converges pointwise to F_X we can choose an α_1 for which

$$F_{X_n}(s+\alpha_1) - F_X(s+\alpha_1) < \epsilon/2$$

Moreover, because F_X is continuous at s we can choose α_2 such that

$$F_X(s+\alpha_2) - F_X(s-\alpha_2) < \epsilon/2$$

Hence, taking α to be the minimum of the two previous choices gives

$$F_{X_n}(t) - F_X(t) < \epsilon$$

for all $t \in (s-\alpha, s+\alpha)$. In the other direction, we also have

$$\begin{aligned} F_{X_n}(t) - F_X(t) &\geq F_{X_n}(s-\beta) - F_X(s+\beta) \\ &= [F_{X_n}(s-\beta) - F_X(s-\beta)] + [F_X(s-\beta) - F_X(s+\beta)] \end{aligned}$$

It is clear that we can choose a β_1 for which

$$F_{X_n}(s-\beta_1) - F_X(s-\beta_1) > -\epsilon/2$$

and β_2 such that

$$F_X(s-\beta_2) - F_X(s+\beta_2) > -\epsilon/2$$

Hence, taking β to be the minimum of β_1 and β_2 gives

$$-\epsilon < F_{X_n}(t) - F_X(t)$$

for all $t \in (s-\alpha, s+\alpha)$. Finally, taking λ to be the smallest of α and β we get

$$-\epsilon < F_{X_n}(t) - F_X(t) < \epsilon$$

for all $t \in (s-\lambda, s+\lambda)$. This proves the uniform convergence of F_{X_n} to F_X on $(s-\lambda, s+\lambda)$.

Now we can address the issue at hand. Let $t \in \mathbb{R}$ and choose a λ such that F_{X_n} converges uniformly to F_X in the interval

$$I = (\frac{t-b}{a} - \lambda, \frac{t-b}{a} + \lambda)$$

For any $\epsilon > 0$, there is an $N_1 > 0$ such that

$$n > N_1 \Rightarrow |F_{X_n}(s) - F_X(s)| < \frac{\epsilon}{2}$$

for all $s \in I$. Also, there is an $N_2 > 0$ such that

$$n > N_2 \Rightarrow \frac{t - b_n}{a_n} \in I$$

It follows that

$$n > \max\{N_1, N_2\} \Rightarrow \left| F_{X_n}(\frac{t - b_n}{a_n}) - F_X(\frac{t - b_n}{a_n}) \right| < \frac{\epsilon}{2}$$

Also, the continuity of F_X implies that there is an $N_3 > 0$ for which

$$n > N_3 \Rightarrow \left| F_X(\frac{t - b_n}{a_n}) - F_X(\frac{t - b}{a}) \right| < \frac{\epsilon}{2}$$

Hence,

$$n > \max\{N_1, N_2, N_3\} \Rightarrow \left| F_{X_n}(\frac{t - b_n}{a_n}) - F_X(\frac{t - b}{a}) \right| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

But

$$F_{X_n}(\frac{t - b_n}{a_n}) = \mathbb{P}(X_n \leq \frac{t - b_n}{a_n}) = \mathbb{P}(a_n X_n + b_n \leq t)$$

and

$$F_X(\frac{t - b}{a}) = \mathbb{P}(X \leq \frac{t - b}{a}) = \mathbb{P}(aX + b \leq t)$$

and so we have shown that

$$\mathbb{P}(a_n X_n + b_n \leq t) \to \mathbb{P}(aX + b \leq t)$$

that is

$$a_n X_n + b_n \xrightarrow{\text{dist}} aX + b$$

The second part follows from the fact that $a\mathcal{N}_{0,1} + b = \mathcal{N}_{a,b}$.□

The Central Limit Theorem

The Central Limit Theorem is the most famous theorem in probability and with good cause. Actually, there are several versions of the Central

Limit Theorem. We will state the most commonly seen version first and later discuss a different version that we will use in the next chapter.

Speaking intuitively, if X is a random variable then it is the distribution function of X that describes its probabilistic "behavior" or "characteristics". More precisely, if X and Y are random variables with the same distribution function then

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a \leq Y \leq b)$$

for all real numbers a and b. (Note that the *functions* X and Y need not be the same. In fact, they need not even be defined on the same sample space.) When two or more random variables have the same distribution function, they are said to be **identically distributed**.

Informally speaking, the Central Limit Theorem says that if S_n is the sum of n random variables that are

1) mutually independent
2) identically distributed

and if we standardize S_n then the resulting random variable S_n^* has very special characteristics. In particular, the distribution function of S_n^* approximates the standard normal distribution *regardless of the type of distribution of the original random variables*. Moreover, the approximation gets better and better as n gets larger and larger.

Thus, the process of summing and standardizing "washes out" the original characteristics of the individual random variables and replaces them with the characteristics of the standard normal random variable.

Here is a formal statement of the Central Limit Theorem.

Theorem 12 *(Central Limit Theorem) Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed random variables with finite mean μ and finite variance $\sigma^2 > 0$. Let*

$$S_n = \sum_{i=1}^{n} X_i$$

be the sum of the first n random variables. Thus, $\mathcal{E}(S) = n\mu$ and $\mathrm{Var}(S) = n\sigma^2$. Consider the standardized random variable

$$S_n^* = \frac{S_n - \mathcal{E}(S_n)}{\sqrt{\text{Var}(S_n)}} = \frac{S - n\mu}{\sqrt{n}\sigma}$$

The sequence of standardized random variables S_n^* converges in distribution to a standard normal random variable $\mathcal{N}_{0,1}$, that is

$$\lim_{n\to\infty} F_{S_n^*}(t) = \phi_{0,1}(t)$$

Put another way

$$P(S_n^* < t) \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx$$

where the error in the approximation tends to 0 *as* n *tends to* ∞.□

As you might expect, the proof of the Central Limit Theorem is a bit involved and we will not go into it in this book. However, the reader is advised to pause a while to consider the somewhat surprising nature of this theorem. It certainly accounts for the extreme importance of the normal distribution.

As mentioned earlier, we need a different version of the Central Limit Theorem for our work on the Black-Scholes formula. On the one hand, we need only to consider Bernoulli random variables with mean 0 and variance 1, which are among the simplest of useful random variables. On the other hand, we need to make things more complex because our Bernoulli random variables are not identically distributed!

In particular, we want to consider not just a simple sequence of random variables but a *triangular array* of random variables

$$\begin{array}{cccc} B_{1,1} & & & \\ B_{2,1} & B_{2,2} & & \\ B_{3,1} & B_{3,2} & B_{3,3} & \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

For each row, the random variables are independent, identically distributed Bernoulli random variables with mean 0 and variance 1. In

particular, $B_{n,i}$ is a Bernoulli random variable with

$$\mathbb{P}(B_{n,i} = \frac{q_n}{\sqrt{p_n q_n}}) = p_n$$
$$\mathbb{P}(B_{n,i} = \frac{-p_n}{\sqrt{p_n q_n}}) = q_n$$

where $q_n = 1 - p_n$. However, random variables from *different* rows need not be independent, nor are they necessarily identically distributed. In fact, *they need not even be defined on the same probability space*. This will turn out to be very important to us later on.

We must also assume that the probabilities p_n are “well-behavied" in the sense that they do not get close to 0 or 1. In fact, we will assume that there is a p satisfying $0 < p < 1$ for which

$$p_n \to p$$

It follows also that

$$q_n \to q = 1 - p \in (0, 1)$$

Now, there is a version of the Central Limit Theorem that addresses just this situation (even when the random variables are not Bernoulli random variables).

We begin by “standardizing” each random variable in such as way that its mean is 0 and that the *sum* of the variances in each row is 1. Since

$$\mathcal{E}(B_{n,i}) = 0$$
$$\text{Var}(B_{n,i}) = 1$$

the new array is simply

$$\begin{array}{ccccc} B_{1,1} & & & & \\ \frac{1}{\sqrt{2}}B_{2,1}(p_2) & \frac{1}{\sqrt{2}}B_{2,2}(p_2) & & & \\ \frac{1}{\sqrt{3}}B_{3,1}(p_3) & \frac{1}{\sqrt{3}}B_{3,1}(p_3) & \frac{1}{\sqrt{3}}B_{3,1}(p_3) & & \\ \vdots & \vdots & \vdots & \ddots & \end{array}$$

Now, the version of the Central Limit Theorem that covers this situation says that under a certain condition the distribution of the row sums

$$S_1 = B_{1,1}$$
$$S_2 = \frac{1}{\sqrt{2}}(B_{2,1} + B_{2,2})$$
$$S_3 = \frac{1}{\sqrt{3}}(B_{3,1} + B_{3,2} + B_{3,3})$$

converges pointwise to the distribution $\phi_{0,1}$ of a standard normal random variable.

The certain condition is a bit messy. Intuitively speaking, it says that each term in the sum S_n is "negligible" with respect to the entire sum. In the case of the Bernoulli random variables in which we are interested, the possible values of the standardized Bernoulli random variables $B^*_{n,i}(p_n)$ that appear in the sums S_n are

$$\frac{q_n}{\sqrt{np_nq_n}} \text{ and } \frac{-p_n}{\sqrt{np_nq_n}}$$

Now, as n tends to ∞, we have

$$\frac{q_n}{\sqrt{p_nq_n}} \to \frac{q}{\sqrt{pq}}$$
$$\frac{-p_n}{\sqrt{p_nq_n}} \to \frac{q}{\sqrt{pq}}$$

and since $p, q > 0$ these limits are finite. Hence, the possible values satisfy

$$\frac{q_n}{\sqrt{np_nq_n}} \to 0$$
$$\frac{-p_n}{\sqrt{np_nq_n}} \to 0$$

This turns out to be a sufficient condition for the Central Limit Theorem to apply. We have finally arrived at the theorem that we need.

Theorem 13 *Consider a triangular array of random variables*

$$\begin{array}{llll} B_{1,1} & & & \\ B_{2,1} & B_{2,2} & & \\ B_{3,1} & B_{3,2} & B_{3,3} & \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

where for each row n and $1 \leq i \leq n$, the $B_{n,i}$ are independent, identically distributed Bernoulli random variables with

$$\mathbb{P}(B_{n,i} = \frac{q_n}{\sqrt{p_n q_n}}) = p_n$$
$$\mathbb{P}(B_{n,i} = \frac{-p_n}{\sqrt{p_n q_n}}) = q_n$$

However, the random variables in different rows need not be independent or identically distributed, or even defined on the same probability space. Suppose also that $p_n \to p \in (0,1)$. Then the random variables

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} B_{n,i}$$

converge in distribution to a standard normal random variable. More specifically, if Z is a standard normal random variable on any probability space then S_n converges in distribution to Z.□

As mentioned earlier, we will use this theorem in the next chapter to help derive the Black-Scholes option pricing formula.

Exercises

1. Let $f(t)$ be a *piecewise linear* probability density function with the following properties: $f(t) = 0$ for $t \leq 0$ and $t \geq 2$, $f(1) = a$. Sketch the graph and find a. Sketch the corresponding distribution function.
2. Let X have distribution function F given by

$$F(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{2}t & 0 \leq t \leq 2 \\ 1 & t > 2 \end{cases}$$

   Let $Y = X^2$. Find
   a) $\mathbb{P}(0 \leq X \leq 1)$
   b) $\mathbb{P}(1 \leq X \leq 3)$
   c) $\mathbb{P}(Y \leq X)$
   d) $\mathbb{P}(X + Y \leq \frac{3}{4})$
   e) the distribution function of the random variable $\sqrt{X}$
3. Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and let $\mathbb{P}$ be the uniform probability measure on Ω, that is, $\mathbb{P}(\omega_i) = 1/3$ for $i = 1, 2, 3$. Consider the following random variables

$$X(\omega_1) = 1, X(\omega_2) = 2, X(\omega_3) = 3$$
$$Y(\omega_1) = 2, Y(\omega_2) = 3, Y(\omega_3) = 1$$
$$Z(\omega_1) = 3, Z(\omega_2) = 1, Z(\omega_3) = 2$$

Are these functions the same? What about their distribution functions?

4. Show that a σ-algebra is closed under countable intersections.
5. Show that all rays are Borel sets.
6. Show that all closed intervals are Borel sets.
7. Prove that

$$\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B)$$

for any events A and B. This is called the **Principle of Inclusion-Exclusion** (for two events).

8. Prove that a probability measure is **subadditive**, that is,

$$\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$$

for any events A and B.

9. Find and graph the uniform distribution function on the interval $[a, b]$.
10. Show that the most general Bernoulli random variable B with mean 0 and variance 1 is given by

$$\mathbb{P}(B = \frac{q}{\sqrt{pq}}) = p$$
$$\mathbb{P}(B = \frac{-p}{\sqrt{pq}}) = q$$

where $q = 1 - p$.

11. Fill in the details to show that the normal distribution has mean μ.
12. Compute the variance of the normal distribution using the integral formula in the text.
13. Prove that if $A_1 \subseteq A_2 \subseteq \cdots$ is an increasing sequence of events, each contained in the next event, then $\lim_{i \to \infty} \mathbb{P}(A_i)$ exists and

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

14. Let X be a random variable on $\mathbb{R}$. Prove that the function

$$f(t) = \mathbb{P}(X \leq t)$$

is a probability distribution function. *Hint*: make heavy use of monotone continuity.

15. For a probability measure $\mathbb{P}$ with distribution function F verify that
    a) $\mathbb{P}((a,b]) = F(b) - F(a)$
    b) $\mathbb{P}((a,b)) = F(b-) - F(a)$
    c) $\mathbb{P}([a,b)) = F(b-) - F(a-)$
    d) $\mathbb{P}([a,b]) = F(b) - F(a-)$
    where the negative sign means limit from below.
16. If $X_n \xrightarrow{\text{dist}} X$ show that for any real numbers $a \neq 0$ and b

$$aX_n + b \xrightarrow{\text{dist}} aX + b$$

Chapter 9

The Black-Scholes Option Pricing Formula

The models that we have been studying are *discrete-time* models, because changes take place only at discrete points in time. On the other hand, in *continuous-time* models, changes can take place (at least theoretically) at any real time during the life of the model.

The most famous continuous-time derivative pricing model culminates in the Black-Scholes option pricing formula, which gives the price of a European put or call based on five quantities

- The *initial price* of the underlying stock, which is known.
- The strike price of the option, which is known.
- The time to expiration, which is known.
- The risk-free rate during the lifetime of the option, which is assumed to be constant and can only be estimated.
- The so-called *volatility* of the stock price, a constant that provides a measure of the fluctuation in the stock's price and thus is a measure of the risk involved in the stock. This quantity can only be estimated as well.

Our goal in this chapter is to describe this continuous-time model and to derive the Black-Scholes option pricing formula. We will derive the continuous-time model as a limiting case of the Cox-Ross-Rubinstein model.

Stock Prices and Brownian Motion

In 1827, just 35 years after the New York Stock Exchange was founded, an English botanist named Robert Brown studied the motion of small pollen grains immersed in a liquid medium. Brown wrote that pollen grains exhibited a “continuous swarming motion” when viewed under the microscope.

The first scientific explanation of this phenomenon was given by Albert Einstein in 1905. He showed that this swarming motion, which is now called **Brownian motion**, could be explained as the consequence of the continual bombardment of the particle by the molecules of the liquid. A formal mathematical description of Brownian motion and its properties

was first given by the great mathematician Norbert Wiener beginning in 1918.

It is especially interesting for us to note that the phenomenon now known as Brownian motion was used in 1900 by the French mathematician Bachelier to model the movement of stock prices, for his doctoral dissertation!

Brownian Motion

Let us look a little more closely at Brownian motion. We have defined a finite stochastic process as a sequence $X_1, \ldots, X_N$ of random variables defined on a sample space Ω. A **continuous stochastic process** on an interval $I \subseteq \mathbb{R}$ of the real line is a collection $\{X_t \mid t \in I\}$ of random variables on Ω indexed by a variable t that ranges over the interval I. For us, we will generally take I to be the interval $[0, \infty)$. Often the variable t represents time and so the value of the process at time t is the value of the random variable X_t.

Definition *A continuous stochastic process* $\{W_t \mid t \geq 0\}$ *is a* **Brownian motion process** *or a* **Wiener process** *with* **volatility** σ *if*

1) $W_0 = 0$

2) W_t *is normally distributed with mean* 0 *and variance* $\sigma^2 t$

3) The process $\{W_t\}$ *has* **stationary increments**, *that is, for* $s < t$, *the* increment $W_t - W_s$ *depends only on the value* $t - s$. *Thus* $W_t - W_s$ *(which has the same distribution as* $W_{t-s} - W_0 = W_{t-s}$*) is normally distributed with mean* 0 *and variance* $\sigma^2(t - s)$.

4) The process $\{W_t\}$ *has* **independent increments**, *that is, for any times* $t_1 \leq t_2 \leq \cdots \leq t_n$, *the nonoverlapping increments*

$$W_{t_2} - W_{t_1}, W_{t_3} - W_{t_2}, \ldots, W_{t_n} - W_{t_{n-1}}$$

are independent *random variables.* □

Brownian Motion with Drift

It is also possible to define Brownian motion with drift. This is a stochastic process of the form $\{\mu t + W_t \mid t \geq 0\}$ where μ is a constant and $\{W_t\}$ is Brownian motion. Here is a formal definition.

Definition *A continuous stochastic process* $\{W_t \mid t \geq 0\}$ *is a* **Brownian motion process** *or a* **Wiener process** *with* **volatility** σ *and* **drift** μ *if*

1) $W_0 = 0$

2) W_t *is normally distributed with mean* μt *and variance* $\sigma^2 t$

3) $\{W_t\}$ *has stationary increments. Thus,* $W_t - W_s$ *is normally distributed with mean* $\mu(t-s)$ *and variance* $\sigma^2(t-s)$.
4) $\{W_t\}$ *has stationary and independent increments.*

Sample Paths

Figure 1 shows three simulated *sample paths* for Brownian motion with drift $\mu = 0.08$ and volatility $\sigma = 0.20$ on the interval $[0, 1]$. The straight line shows the drift.

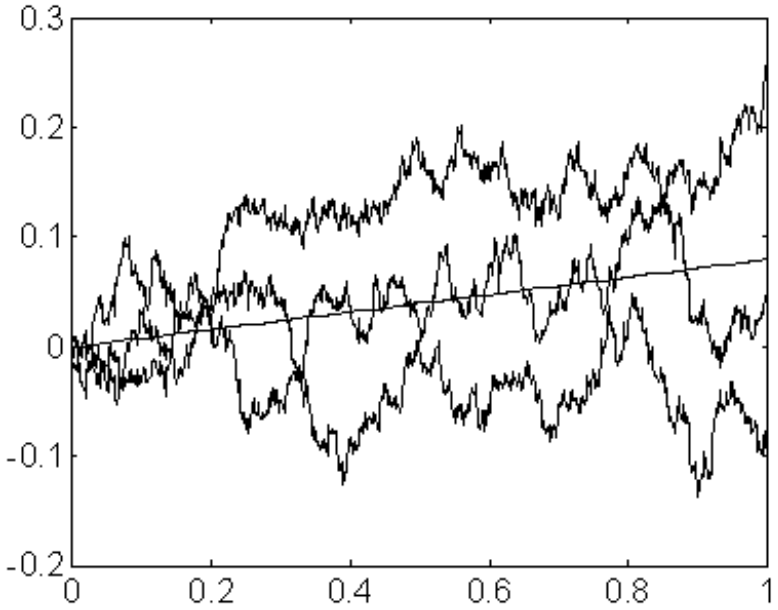


Figure 1–Brownian motion sample paths: $\mu = 0.08, \sigma = 0.20$

More specifically, if we fix an outcome $\omega \in \Omega$ then we can define a function

$$t \to W_t(\omega)$$

The graph of this function is called a **sample path**.

Figure 2 shows a discrete sample path for a Brownian motion process that is the same as the previous one except that the volatility is only $\sigma = 0.05$. As you can see, volatility is aptly named.

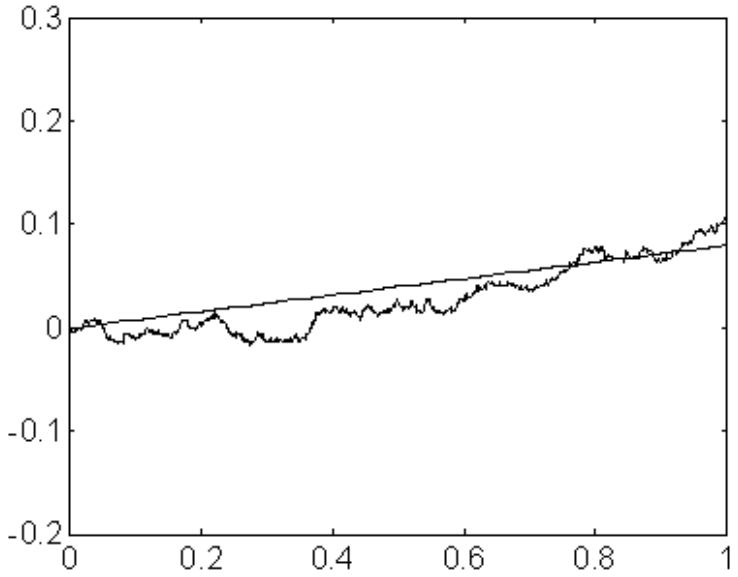


Figure 2–Brownian motion with small volatility

Brownian motion is one of the most important types of random processes and has a great many applications. This should not be a surprise in view of the Central Limit Theorem, which tells us that the normal distribution is so important. (Brownian motion is a kind of "traveling normal distribution.")

In any case, Brownian motion has some very special properties. For instance, the sample paths are always continuous functions. In other words, the sample paths do not have any jumps. On the other hand, these paths are also essentially *nowhere* differentiable, that is, it is not possible to define a tangent line at any place on the curve. Thus, a sample path has no jumps but is nonetheless very jerky, constantly changing direction abruptly. (The previous figures do not do justice to this statement because they are are not true sample paths.)

Standard Brownian Motion

A Brownian motion process $\{W_t \mid t \geq 0\}$ with $\mu = 0$ and $\sigma = 1$ is called **standard Brownian motion**. In this case W_t has mean 0 and variance t. If $\{W_t \mid t \geq 0\}$ is Brownian motion with drift μ and variance σ^2 then we can write

$$W_t = \mu t + \sigma Z_t$$

where $\{Z_t \mid t \geq 0\}$ is standard Brownian motion.

Geometric Brownian Motion and Stock Prices

How does Brownian motion with drift relate to stock prices? One *possibility* is to think of a stock's price as a "particle" that is subject to constant bombardment by "smaller particles" in the form of stock trades, or other local events. As further support of this viewpoint, we can note that the normal distribution seems like a reasonable choice to model a stock's price if we think of the vicissitudes of that price as being the result of a large number of more-or-less independent (and similarly distributed) factors.

However, there are some obvious problems with the Brownian motion viewpoint for stock prices themselves. First, a Brownian motion process can be negative whereas stock prices are never negative. Second, in a Brownian motion process, the increments

$$W_t - W_s$$

have distributions that depend only on $t - s$. Thus, if a stock's price were to behave as a Brownian motion process W_t then the expected change $\mathcal{E}(W_t - W_s)$ in the stock's price over a period of time would be $\mu(t - s)$, which does not depend on the initial price W_s. This is not very realistic. For instance, imagine a length of time $t - s$ for which the change in price is $\mu(t - s) = \$10$. A \$10 expected price change might be quite reasonable if the stock is initially priced at $W_s = \$100$ but not nearly as reasonable if the stock is initially priced at $W_s = \$1$.

To resolve these issues, it would seem to make more sense to model the *rate of return* of the stock price as a Brownian motion process, for this quantity seems more reasonably independent of the initial price. For example, to say that a stock's rate of return is 10% is to say that the price may grow from \$100 to \$110 or from \$1 to \$1.10.

This can be handled by assuming that the stock price S_t at time t is given by

$$S_t = S_0 e^{H_t}$$

where S_0 is the initial price and H_t is a Brownian motion process. The exponent H_t represents a *continuously compounded* rate of return of the

stock price over the period of time $[0, t]$. Note also that H_t, which we will refer to as the **logarithmic growth** of the stock price, satisfies

$$H_t = \log\left(\frac{S_t}{S_0}\right)$$

Definition A stochastic process of the form $\{e^{W_t} \mid t \geq 0\}$ where $\{W_t \mid t \geq 0\}$ is a Brownian motion process is called a ***geometric Brownian motion process***.□

Figure 3 shows a simulated sample path from a geometric Brownian motion process with the same drift ($\mu = 0.08$) as before but with an unnaturally large volatility in order to demonstrate the exponential nature of the growth.

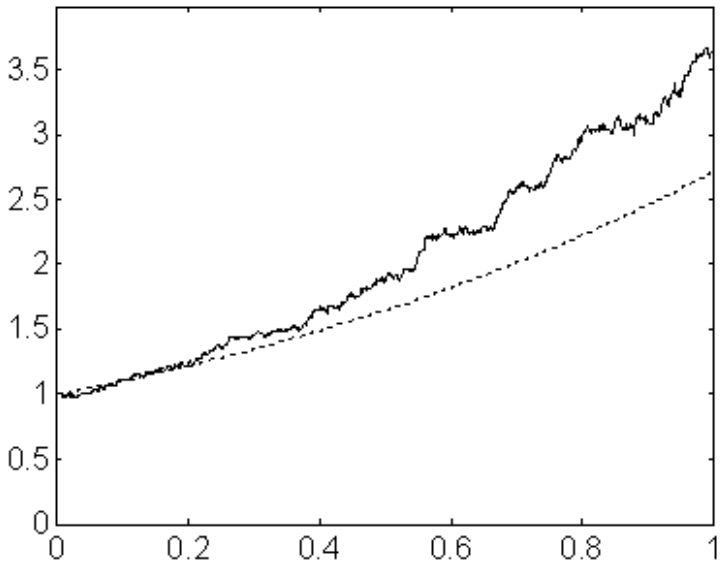


Figure 3–Geometric Brownian motion

If we assume that H_t follows a Brownian motion process with drift then we can write

$$H_t = \log\left(\frac{S_t}{S_0}\right) = \mu t + \sigma W_t$$

where $\{W_t\}$ is standard Brownian motion. Therefore, H_t has a normal distribution with

$$\mathcal{E}(H_t) = \mu t$$
$$\text{Var}(H_t) = \sigma^2 t$$

If a random variable X has the property that its logarithm $\log X$ is normally distributed, then the random variable X is said to have a **lognormal distribution**. (Note that X is lognormal if its logarithm is normal, *not* if it is the logarithm of a normal random variable. In other words, lognormal means "log*is*normal" not "log*of*normal".)

Accordingly, S_t/S_0 is lognormally distributed. Figure 4 shows a typical lognormal density function.

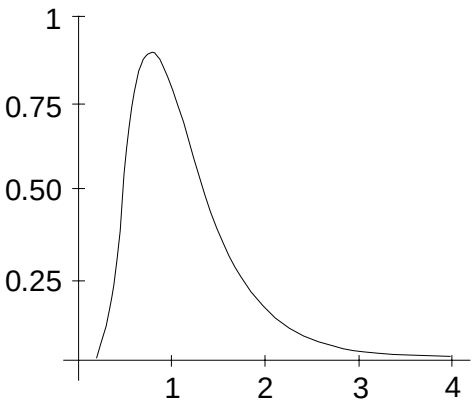


Figure 4–A lognormal density function

According to Theorem 9 of Chapter 8

$$\mathcal{E}(S_t) = S_0 e^{(\mu+\frac{1}{2}\sigma^2)t}$$
$$\text{Var}(S_t) = (S_0 e^{(\mu+\frac{1}{2}\sigma^2)t})^2(e^{\sigma^2 t} - 1)$$

This value of $\mathcal{E}(S_t)$ is quite interesting, for it tells us that the expected stock price depends not only on the drift μ of H_t but also on the volatility σ. This should not necessarily be surprising from a mathematical point of view, for there is no law that the mean of a function of a random variable should be a function *only* of the mean of the random variable.

A Different Approach to the Model

As mentioned earlier, our approach to the continuous-time model is as a limiting case of the Cox-Ross-Rubinstein model. We will endeavor to be as mathematically rigorous as possible in this approach, which is often

handled rather informally. However, we should take a few minutes to discuss what is generally considered to be a more rigorous approach to the model. Since a formal discussion would require considerably more mathematical machinery than we have at hand, we will proceed informally.

Let us begin by taking a closer look at the notion of *rate of return* on a stock price. This term has more than one meaning. We have already considered the continuously compounded rate of return H_t over the time period $[0, t]$, which satisfies the equation

$$S_t = S_0 e^{H_t}$$

The **simple rate of return** of the stock price over a short period of time $[t, t + \Delta t]$ is given by

$$\frac{\Delta S_t}{S_t} = \frac{S_{t+\Delta t} - S_t}{S_t}$$

In the limit as $\Delta t \to 0$, the simple rate of return can be thought of as a rate of return over an infinitesimal time period dt. This is more appropriately called the **instantaneous percentage return** of the stock price and is denoted by

$$\frac{dS_t}{S_t}$$

Now, the most common approach to the continuous-time model of stock prices assumes that the instantaneous percentage return is a Brownian motion process, more specifically

$$\frac{dS_t}{S_t} = \mu_0\, dt + \sigma_0\, dW_t$$

where $\{W_t \mid t \geq 0\}$ is standard Brownian motion and μ_0 and σ_0 are constants. This equation must remain somewhat vague for us, because the meaning of the differentials dS_t and dW_t are rather involved. However, we can say that the stochastic process

$$\left\{ \frac{dS_t}{S_t} \middle| t \geq 0 \right\}$$

(whatever that really is) is assumed to follow a Brownian motion process with drift μ_0 and volatility σ_0.

The previous formula can be written

$$dS_t = \mu_0 S_t dt + \sigma_0 S_t \, dW_t$$

This is an instance of what is known as a *stochastic differential equation*, whose formal solution requires some rather sophisticated mathematical machinery known as *stochastic calculus*. Fortunately, our plan of approach, through the CRR model, will lead us to an expression for the stock price itself without having to solve this differential equation.

The CRR Model in the Limit: Brownian Motion

At the end of Chapter 7 we took a look at the Cox-Ross-Rubinstein model from the point of view of the logarithmic growth in the stock price. Let us recall the pertinent results.

Recap of the CRR Model

We specify the model times

$$t_0 = 0 < t_1 < \cdots < t_{T-1} < t_T = t$$

Thus, the lifetime of the model is $[0, t]$. Note that we are using t instead of L because we want to think of t as a variable. Each of the T intervals has equal length

$$t_{k+1} - t_k = \Delta t = \frac{t}{T}$$

During each subinterval, the stock price may rise by a factor of u or fall by a factor of d. The sample space for this model is the state space

$$\Omega_T = \{U, D\}^T$$

of all sequences of U's and D's of length T.

We will be dealing with two different probabilities on Ω_T: the natural probability and the martingale measure. Let us denote the natural probability of an up-tick by ν (the Greek letter *nu*) and the martingale measure probability of an up-tick by π (dropping the subscript U used in earlier chapters). Before dealing with these probabilities and their interactions, however, it may help clarify the exposition to take another look at the logarithmic price growth using an arbitrary probability p for the up-tick in the stock price. (Then $q = 1 - p$ is the probability of a down-tick.)

Let E_i give the stock price movement over the interval $[t_i, t_{i+1}]$, that is, for any final state $\omega = e_1 \cdots e_T \in \Omega_T$ where $e_i = U$ or D we set

$$E_i(\omega) = \begin{cases} u & \text{if } e_i = U \text{ (up-tick at time } t_i) \\ d & \text{if } e_i = D \text{ (down-tick at time } t_i) \end{cases}$$

Hence the stock price at the final time t is given by

$$S_T = S_0 E_1 \cdots E_T = S_0 e^{\sum \log(E_i)} = S_0 e^{H_T}$$

where

$$H_T = \sum_{i=1}^{T} \log(E_i) = \log\left(\frac{S_T}{S_0}\right)$$

is the logarithmic growth of the stock price. If the probability of an up-tick in the stock is denoted by p then we define μ_p and σ_p by

$$\mu_p = \frac{1}{\Delta t}\mathcal{E}_p(\log E_i) = \frac{1}{\Delta t}(p \log u + q \log d)$$
$$\sigma_p^2 = \frac{1}{\Delta t}\mathrm{Var}_p(\log E_i) = \frac{1}{\Delta t} pq(\log u - \log d)^2$$

Thus, μ_p is the expected value and σ_p^2 is the variance of the logarithmic price change $\log E_i$ *per unit of time*.

Some simple algebra gives

$$\log E_i = \mu_p \Delta t + \sigma_p \sqrt{\Delta t}\left[\frac{\log E_i - \mu_p \Delta t}{\sigma_p \sqrt{\Delta t}}\right] = \mu_p \Delta t + \sigma_p \sqrt{\Delta t} X_{p,i}$$

where

$$X_{p,i} = \frac{\log E_i - \mu_p \Delta t}{\sigma_p \sqrt{\Delta t}}$$

After a bit more algebra, we see that

$$X_{p,i}(\omega) = \begin{cases} \frac{q}{\sqrt{pq}} & \text{if } e_i = U \\ \frac{-p}{\sqrt{pq}} & \text{if } e_i = D \end{cases}$$

It follows that

$$\begin{aligned} H_T &= \sum_{i=1}^{T} \log(E_i) \\ &= \sum_{i=1}^{T} [\mu_p \Delta t + \sigma_p \sqrt{\Delta t} X_{p,i}] \\ &= \mu_p t + \sigma_p \sqrt{\Delta t} \sum_{i=1}^{T} X_{p,i} \end{aligned}$$

Note that this formula

$$H_T = \mu_p t + \sigma_p \sqrt{\Delta t} \sum_{i=1}^{T} X_{p,i}$$

is valid for any $0 < p < 1$. The number μ_p is called the **drift** and the number σ_p is called the **volatility** of the stock price. These quantities are with respect to the up-tick probability p, as the subscript notation indicates. Note, however, that H_T itself does not depend on p. It is just a function on the state space Ω_T and has a different expression in terms of each probability $p \in (0, 1)$.

More on the Probabilities

As mentioned, p is the probability of an up-tick in the stock price over a subinterval. Later, we will take $p = \nu$ or $p = \pi$ but we do not want to make that restriction now.

The probability p defines a probability measure $\mathbb{P}_p$ on the state space Ω_T for which

$$\mathbb{P}_p(\omega) = p^{N_U(\omega)} q^{\text{len}(\omega) - N_U(\omega)}$$

From the definition of $X_{p,i}$ we get

$$\begin{aligned} \mathbb{P}_p(X_{p,i} = \frac{q}{\sqrt{pq}}) &= \mathbb{P}(e_i = U) = p \\ \mathbb{P}_p(X_{p,i} = \frac{-p}{\sqrt{pq}}) &= \mathbb{P}(e_i = D) = q \end{aligned}$$

and so the random variables $X_{p,i}$ have this Bernoulli distribution under $\mathbb{P}_p$.

We can also compute expected values and variances with respect to this probability measure

$$\mathcal{E}_p(X_{p,i}) = \frac{q}{\sqrt{pq}}p - \frac{p}{\sqrt{pq}}q = 0$$

and

$$\mathrm{Var}_p(X_{p,i}) = 1$$

(We leave verification of this as an exercise.)

The formula

$$H_T = \mu_p t + \sigma_p \sqrt{\Delta t} \sum_{i=1}^{T} X_{p,i}$$

can now be seen as expressing the logarithmic growth H_T as the sum of two components: a *deterministic component* $\mu_p t$ which is a constant times the time and thus grows at the fixed rate μ_p just like a risk-free asset and a *random component*

$$Q_p = \sigma_p \sqrt{\Delta t} \sum_{i=1}^{T} X_{p,i}$$

which is $\sigma_p \sqrt{\Delta t}$ times a sum of Bernoulli random variables. Since the terms in the sum are independent, we also have

$$\mathcal{E}_p(Q_p) = 0$$
$$\mathrm{Var}_p(Q_p) = \sigma_p^2 t$$

Taking the Limit as $\Delta t \to 0$

In taking the limit as $T \to \infty$, or equivalently as $\Delta t \to 0$ we want to be careful to make it clear which quantities vary with T. We also want to make it clear that some quantities depend on the lifetime t, which we now think of as a variable. So let us change the notation as follows:

- Let $S_{t,T}$ denote the final stock price and $H_{t,T}$ denote the logarithmic growth. The initial stock price S_0 does not depend on T so we do not need to change this notation.
- Let u_T and d_T denote the up-tick and down-tick factors for the stock price.
- Let p_T denote the probability of an up-tick.

- Let $\mu_{p_T,T}$ and $\sigma_{p_T,T}$ denote the drift and volatility.
- Let $X_{p_T,T,i}$ denote the random variable $X_{p_T,i}$.

With this notation at hand, the formula for H_t becomes

$$H_{t,T} = \mu_{p_T,T}t + \sigma_{p_T,T}\sqrt{\Delta t}\sum_{i=1}^{T} X_{p_T,T,i}$$

with deterministic part $\mu_{p_T,T}t$ and a random part

$$Q_{p_T,T} = \sigma_{p_T,T}\sqrt{\Delta t}\sum_{i=1}^{T} X_{p_T,T,i}$$

for which

$$\begin{aligned}\mathcal{E}_{p_T}(Q_{p_T,T}) &= 0\\ \mathrm{Var}_{p_T}(Q_{p_T,T}) &= \sigma^2_{p_T,T}t\end{aligned}$$

We now want to apply the Central Limit Theorem to the random part. To this end, consider the triangular array of Bernoulli random variables

$$\begin{array}{llll} X_{p_1,1,1} & & & \\ X_{p_2,2,1} & X_{p_2,2,2} & & \\ X_{p_3,3,1} & X_{p_3,3,2} & X_{p_3,3,3} & \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

For each fixed T, that is, for each row of the array, the random variables are independent (by assumption of the CRR model) and satisfy

$$\begin{aligned}\mathbb{P}_{p_T}(X_{p_T,T,i} = \frac{1-p_T}{\sqrt{p_T(1-p_T)}}) &= p_T\\ \mathbb{P}_{p_T}(X_{p_T,T,i} = \frac{-p_T}{\sqrt{p_T(1-p_T)}}) &= 1-p_T\end{aligned}$$

and so they are also identically distributed. Note, however, that $X_{p_T,T,i}$ is a random variable on the probability space $(\Omega_T, \mathbb{P}_{p_T})$ and so for different values of T (that is, different rows of the array) the random variables $X_{p_T,T,i}$ are defined on *different* probability spaces. This is precisely why we need the Central Limit Theorem in the form of Theorem 13 of Chapter 8, which applies provided that the probabilities p_T satisfy

Requirement 1: $p_T \to p$ for some $p \in (0,1)$

Assuming that this requirement is met, we can conclude that the random variables

$$Y_{p_T,T} = \frac{1}{\sqrt{T}}\sum_{i=1}^{T} X_{p_T,T,i} = \frac{1}{\sqrt{t}}\sqrt{\Delta t}\sum_{i=1}^{T} X_{p_T,T,i} = \frac{1}{\sigma_{p_T,T}\sqrt{t}} Q_{p_T,T}$$

converge in distribution to a standard normal random variable, that is

$$Q^*_{p_T,T} = \frac{Q_{p_T,T} - \mathcal{E}_{p_T}(Q_{p_T,T})}{\sqrt{\text{Var}_{p_T}(Q_{p_T,T})}} = \frac{Q_{p_T,T}}{\sigma_{p_T,T}\sqrt{t}} \xrightarrow{\text{dist}} Z_t$$

where Z_t is any standard normal random variable on some probability space. To be more specific, this means that for any real number s

$$\lim_{T\to\infty} \mathbb{P}_{p_T}\left(\frac{Q_{p_T,T}}{\sigma_{p_T,T}\sqrt{t}} < s\right) = \phi_{0,1}(s)$$

where $\phi_{0,1}$ is the standard normal distribution function. To emphasize the fact that the convergence involves the probability measures $\mathbb{P}_{p_T}$, we write

$$\frac{Q_{p_T,T}}{\sigma_{p_T,T}\sqrt{t}} \xrightarrow{\text{dist}(p_T)} Z_t$$

We would now like to conclude that $H_{t,T}$ itself converges in distribution to something. For this, let us recall Theorem 11 of Chapter 8, which says that if

Requirement 2: $\mu_{p_T,T} \to \mu, \sigma_{p_T,T} \to \sigma$ for real numbers μ and $\sigma \neq 0$

then we have the following limit

$$H_{t,T} = \mu_{p_T,T}t + \sigma_{p_T,T}\sqrt{t}\left(\frac{Q_{p_T,T}}{\sigma_{p_T,T}\sqrt{t}}\right) \xrightarrow{\text{dist}(p_T)} \mu t + \sigma\sqrt{t}Z_t$$

As to the stock price itself, since $S_{t,T}$ is a continuous function of $H_{t,T}$

$$S_{t,T} = S_0 e^{H_{t,T}}$$

Theorem 10 of Chapter 8 tells us that as long as requirements 1 and 2 are met we have

$$S_{t,T} \xrightarrow{\text{dist}(p_T)} S_0 e^{\mu t + \sigma\sqrt{t}Z_t}$$

where $Z_{p_T,t}$ has a standard normal distribution. Setting

$$W_t = \sqrt{t} Z_t$$

gives

$$H_{t,T} \overset{\text{dist}(p_T)}{\longrightarrow} H_t = \mu t + \sigma W_t$$

and

$$S_{t,T} \overset{\text{dist}(p_T)}{\longrightarrow} S_t = S_0 e^{\mu t + \sigma W_t}$$

where

$$\begin{aligned} \mathcal{E}(W_t) &= 0 \\ \text{Var}(W_t) &= t \end{aligned}$$

Brownian and Geometric Brownian Motion

Although we have derived these formulas for a fixed total lifetime t, as mentioned earlier, we want to think of t as a variable. Unfortunately, our derivation does not directly expose the very important relationship between the different random variables W_t as t ranges over all nonnegative real numbers.

We will not go into this issue formally. However, we can make a few informal observations. First, it should be intuitively clear that our model is "translation invariant" or "stationary" in the sense that we obtain essentially the same model over the interval $[s, t]$ as over the interval $[0, t - s]$. Second, because we assume that the changes in each subinterval are independent, we should be able to piece together models from disjoint contiguous intervals into a model for one large interval. Thus, it should not come as a surprise that as t varies, the stochastic process

$$\{W_t \mid t \geq 0\}$$

is actually a standard Brownian motion process. Hence,

$$\{H_t \mid t \geq 0\}$$

is a Brownian motion process with mean μt and variance $\sigma^2 t$, that is, with drift μ and volatility σ. Finally, the stock price process itself

$$\{S_t \mid t \geq 0\}$$

is a geometric Brownian motion process with this drift and volatility.

We can now summarize our knowledge of the behavior of the stock price in the limiting case of a CRR model.

Theorem 1 *Let $S_{t,T}$ be the final stock price for the CRR model with probability p_T of an up-tick, lifetime $[0, t]$ and T equal-sized subintervals of length Δt. Assume that*
1) $p_T \to p$ for some $p \in (0, 1)$
2) $\mu_{p_T,T} \to \mu, \sigma_{p_T,T} \to \sigma$ for real numbers μ and $\sigma \neq 0$
Let

$$H_{t,T} = \log\left(\frac{S_{t,T}}{S_0}\right)$$

be the logarithmic price growth. Then

$$H_{t,T} \overset{dist(p_T)}{\longrightarrow} H_t = \mu t + \sigma W_t$$

and

$$S_{t,T} \overset{dist(p_T)}{\longrightarrow} S_t = S_0 e^{\mu t + \sigma W_t}$$

where

$$\{W_t \mid t \geq 0\}$$

is a standard Brownian motion process. Hence, the logarithmic growth

$$\{H_t \mid t \geq 0\}$$

is a Brownian motion process with mean μt, variance $\sigma^2 t$, drift μ and volatility σ. The stock price process itself

$$\{S_t \mid t \geq 0\}$$

is a geometric Brownian motion process with drift μ and volatility σ. Note also that the stock price growth S_t/S_0 is lognormally distributed with

$$\mathcal{E}(S_t) = S_0 e^{(\mu + \frac{1}{2}\sigma^2)t}$$
$$\text{Var}(S_t) = (S_0 e^{(\mu + \frac{1}{2}\sigma^2)t})^2 (e^{\sigma^2 t} - 1)$$

□

The Natural CRR Model

We have done what we can for the general CRR model. Now it is time to consider how we should structure the model to reflect the "real" world. We will refer to the following model as the **natural probability CRR model** and denote it by **CRR$_\nu$**.

First, we will assume that there is a probability, called the **natural probability** that reflects the true probability of an up-tick in the market. Also, it is customary to make the following assumption about the natural probability.

Assumption 1
The natural probability of an up-tick is constant with respect to T throughout the lifetime of the model. We will denote this probability by ν.

It is also customary to make the following assumption.

Assumption 2
Under the natural probability, the drift and volatility

$$\mu_\nu = \frac{1}{\Delta t}(\nu \log u_T + (1-\nu)\log d_T)$$
$$\sigma_\nu^2 = \frac{1}{\Delta t}\nu(1-\nu)(\log u_T - \log d_T)^2$$

are constant with respect to Δt (and T). Thus, we can drop the subscript T and write μ_ν and σ_ν. The number μ_ν is called the **instantaneous drift** and σ_ν is called the **instantaneous volatility**.

It is important to emphasize that the second assumption has some important consequences for the up-factor, down-factor and martingale measure parameters of the model.

In the general CRR model the quantities u_T, d_T and p_T $(=\nu)$ are unrelated, whereas the drift and volatility are *defined* in terms of these quantities. However, specifying that μ_ν and σ_ν are specific *constants* amounts to specifying a relationship between u_T, d_T and ν. To draw a simple analogy, suppose that x and y are arbitrary variables and that we *define* the quantity A by

$$A = x + y$$

As soon as we postulate that $A = 5$, for instance, we have drawn a relationship between x and y.

The relationship between u_T, d_T and ν is obtained simply by solving the equations defining the drift and volatility to get

$$\log u_{\nu,T} = \mu_\nu \Delta t + \sigma_\nu \sqrt{\Delta t} \frac{1-\nu}{\sqrt{\nu(1-\nu)}}$$
$$\log d_{\nu,T} = \mu_\nu \Delta t - \sigma_\nu \sqrt{\Delta t} \frac{\nu}{\sqrt{\nu(1-\nu)}}$$

The right-hand side depends on T through $\Delta t = t/T$. Also, we write $u_{\nu,T}$ and $d_{\nu,T}$ to emphasize the dependence on ν as well. Note that this dependence flows through to the martingale measure

$$\pi_{\nu,T} = \frac{e^{rT} - d_{\nu,T}}{u_{\nu,T} - d_{\nu,T}}$$

In fact, we can express the martingale measure directly in terms of the probability ν and the drift and volatility. This also leads to an interesting limit.

Theorem 2

1) The martingale measure up-tick probability in the model CRR$_\nu$ is given by

$$\pi_{\nu,T} = \frac{e^{(r-\mu_\nu)\Delta t + \frac{\nu}{\sqrt{\nu(1-\nu)}}\sigma_\nu\sqrt{\Delta t}} - 1}{e^{\frac{1}{\sqrt{\nu(1-\nu)}}\sigma_\nu\sqrt{\Delta t}} - 1}$$

2) The martingale probability $\pi_{\nu,T}$ approaches the natural probability ν as $T \to \infty$ or equivalently as $\Delta t \to 0$, that is

$$\lim_{T\to\infty} \pi_{\nu,T} = \nu$$

Proof. For part 1), the martingale measure up-tick probability is given by

$$\pi_{\nu,T} = \frac{e^{rT} - d_{\nu,T}}{u_{\nu,T} - d_{\nu,T}} = \frac{e^{r\Delta t}(d_{\nu,T})^{-1} - 1}{u_{\nu,T}(d_{\nu,T})^{-1} - 1}$$

The previous equations for $\log u_{\nu,T}$ and $\log d_{\nu,T}$ give

$$u_{\nu,T} = e^{\mu_\nu \Delta t + \sigma_\nu \sqrt{\Delta t} \frac{1-\nu}{\sqrt{\nu(1-\nu)}}}$$
$$(d_{\nu,T})^{-1} = e^{-\mu_\nu \Delta t + \sigma_\nu \sqrt{\Delta t} \frac{\nu}{\sqrt{\nu(1-\nu)}}}$$

and so

$$e^{r\Delta t}(d_{\nu,T})^{-1} = e^{(r-\mu_\nu)\Delta t + \frac{\nu}{\sqrt{\nu(1-\nu)}}\sigma_\nu \sqrt{\Delta t}}$$

and

$$u_{\nu,T}(d_{\nu,T})^{-1} = e^{\frac{1}{\sqrt{\nu(1-\nu)}}\sigma_\nu \sqrt{\Delta t}}$$

Plugging these expressions into the right-hand side of the previous expression for $\pi_{\nu,T}$ gives the desired formula.

Part 2) is a simple application of l'Hopital's rule to evaluate the limit

$$\lim_{T\to\infty} \pi_{\nu,T} = \lim_{\Delta t \to 0} \frac{e^{(r-\mu_\nu)\Delta t + \frac{\nu}{\sqrt{\nu(1-\nu)}}\sigma_\nu \sqrt{\Delta t}} - 1}{e^{\frac{1}{\sqrt{\nu(1-\nu)}}\sigma_\nu \sqrt{\Delta t}} - 1}$$

We leave this to the reader.□

The assumption that the instantaneous drift and volatility are constant is perhaps a questionable one (as is the assumption of a constant natural probability), but is based on practical considerations (as is often the case with questionable assumptions). In fact, this assumption is usually extended into the past. In particular, it is assumed that the drift and volatility can be computed (or at least estimated) under the natural probability by looking at the *past history* of the price for the stock in question.

Specifically, for the natural probability, the instantaneous drift and volatility can be estimated empirically as follows. First we choose a small value for Δt (for example, Δt may correspond to one day). Then over a large number of these short periods of time, we compute the logarithmic growth factors

$$\log E_i = \log\left(\frac{\text{Stock price at end of period}}{\text{Stock price at start of period}}\right)$$

The average of these sample values is an estimate of $\mu_\nu \Delta t$ and the sample variance of these sample values is an estimate of $\sigma_\nu^2 \Delta t$. Of

course, the more samples we take, the better will be the estimates. Let us consider an example.

EXAMPLE 1 The following portion of an Excel spreadsheet shows closing stock prices over a 10-day period. Here we are taking Δt to be one day, that is, $1/365$ years. The initial price is \$50.

Day	Price	Growth	Log Growth	Average	Sample Var
0	50			0.000359354	0.000677264
1	50.95	1.019	0.018821754	**Per Unit**	**Per Unit**
2	49.74	0.976251	-0.024035321	0.131164046	0.247201227
3	49.46	0.994371	-0.005645176		
4	49.83	1.007481	0.00745295		
5	48.7	0.977323	-0.022938182		
6	50.2	1.030801	0.030335997		
7	49.57	0.98745	-0.012629215		
8	51.78	1.044583	0.043618162		
9	52.17	1.007532	0.007503643		
10	50.18	0.961855	-0.038891076		

The growth column contains the quotient of the stock price and the previous stock price. For example,

$$\frac{50.95}{50} = 1.019$$

The average is simply the sum of the logarithmic growths divided by the number of logarithmic growths. The sample variance is computed using the formula

$$\text{Sample var} = \frac{1}{n-1}\sum_{i=1}^{n}(i\text{th value} - \text{average})^2$$

(The reason for dividing by $n-1$ instead of n has to due with obtaining an unbiased value.) Finally, the "per unit" values are obtained by multiplying the average and variance by $1/\Delta t = 365$. It follows that $\mu_\nu \approx 0.13$ and $\sigma_\nu^2 \approx 0.25$ on an annual basis (that is, the units of time are measured in years).□

With the aforementioned assumptions concerning the natural probability in mind, the conditions of Theorem 1

1) $p_T \to p$ *for some* $p \in (0,1)$
2) $\mu_{p_T,T} \to \mu, \sigma_{p_T,T} \to \sigma$ *for real numbers* μ *and* σ

become evident. Since $p_T = \nu$ for all T, we have $p = \nu$ and since $\mu_{p_T,T} = \mu_\nu$ and $\sigma_{p_T,T} = \sigma_\nu$ for all T, we have $\mu = \mu_\nu$ and $\sigma = \sigma_\nu$. Thus, the conditions of Theorem 1 are satisfied for the natural probability and we have the following.

Theorem 3 *Let $S_{t,T}$ be the final stock price for the natural model CRR$_\nu$. Let*

$$H_{t,T} = \log\left(\frac{S_{t,T}}{S_0}\right)$$

be the logarithmic price growth. Then

$$H_{t,T} \overset{dist(\nu)}{\longrightarrow} H_t = \mu_\nu t + \sigma_\nu W_{\nu,t}$$

and

$$S_{t,T} \overset{dist(\nu)}{\longrightarrow} S_t = S_0 e^{\mu_\nu t + \sigma_\nu W_{\nu,t}}$$

where

$$\{W_{\nu,t} \mid t \geq 0\}$$

is a standard Brownian motion process. Hence, the logarithmic growth

$$\{H_t \mid t \geq 0\}$$

is a Brownian motion process with mean $\mu_\nu t$, variance $\sigma_\nu^2 t$, drift μ_ν and volatility σ_ν. The stock price process itself

$$\{S_t \mid t \geq 0\}$$

is a geometric Brownian motion process with the same drift and volatility. Note also that the stock price growth S_t/S_0 is lognormally distributed with

$$\mathcal{E}(S_t) = S_0 e^{(\mu_\nu + \frac{1}{2}\sigma_\nu^2)t}$$
$$\text{Var}(S_t) = (S_0 e^{(\mu_\nu + \frac{1}{2}\sigma_\nu^2)t})^2 (e^{\sigma_\nu^2 t} - 1)$$

□

The Martingale Measure CRR Model

Let us take a peek at our main goal to see what to do next. The payoff for a European put (for example) with strike price K is

$$\begin{aligned} X &= \max\{K - S_{t,T}, 0\} \\ &= \max\{K - S_0 e^{H_{t,T}}, 0\} \end{aligned}$$

The absence of arbitrage implies that the initial price of the put must be

$$\mathcal{I}(\text{Put}) = e^{-rt}\mathcal{E}_{\Pi_T}(\max\{K - S_0 e^{H_{t,T}}, 0\})$$

where Π_T is the martingale measure (with up-tick probability π_T). Taking limits as T tends to infinity gives

$$P_\infty = \lim_{T\to\infty} \mathcal{I}(\text{Put}) = e^{-rt} \lim_{T\to\infty} \mathcal{E}_{\Pi_T}(\max\{K - S_0 e^{H_{t,T}}, 0\})$$

where P_∞ denotes the limiting price random variable. Setting

$$g(x) = \max\{K - S_0 e^x, 0\}$$

which is bounded and continuous on $\mathbb{R}$ gives

$$P_\infty = e^{-rt} \lim_{T\to\infty} \mathcal{E}_{\Pi_T}(g(H_{t,T}))$$

Now, we would like to pass the limit inside the expectation to get

$$P_\infty = e^{-rt}\mathcal{E}(g(H_t))$$

Let us recall Theorem 10 of Chapter 8, which says that if

$$X_n \overset{\text{dist}(\pi_T)}{\longrightarrow} X$$

then

$$\mathcal{E}_{\Pi_T}(g(X_n)) \to \mathcal{E}(g(X))$$

for all bounded continuous functions $g\colon \mathbb{R} \to \mathbb{R}$. This is just what we need, but in order to apply this theorem, we need to know the weak convergence of $H_{t,T}$ under the *martingale measure* probability, not the natural probability, as given in Theorem 3.

This tells us what to do next. In particular, we need a new CRR model to do the following.

1) The probability of an up-tick should be the martingale measure up-tick probability $\pi_{\nu,T}$ so that Theorem 1 will give the weak limit of $H_{t,T}$ under the martingale measure.

2) At the same time, the model must preserve the "true" stock prices S_k from the natural model CRR_ν, which is done by using the values $u_{\nu,T}$ and $d_{\nu,T}$ from that model.

Thus, we define a new CRR model with the following parameters.

1) The up-factor $u_{\nu,T}$, down-factor $d_{\nu,T}$ and martingale measure up-tick probability

$$\pi_{\nu,T} = \frac{e^{rT} - d_{\nu,T}}{u_{\nu,T} - d_{\nu,T}}$$

are as in the natural probability model CRR_ν. It follows that the stock prices S_k are the prices are the "natural" prices, as desired.

2) The probability of an up-tick is the martingale measure up-tick probability, that is,

$$p_T = \pi_{\nu,T}$$

Hence the drift and volatility are

$$\mu_{\pi_{\nu,T},T} = \frac{1}{\Delta t}(\pi_{\nu,T} \log u_{\nu,T} + (1 - \pi_{\nu,T}) \log d_{\nu,T})$$
$$\sigma^2_{\pi_{\nu,T},T} = \frac{1}{\Delta t}\pi_{\nu,T}(1 - \pi_{\nu,T})(\log u_{\nu,T} - \log d_{\nu,T})^2$$

We will call this model the **martingale measure CRR model** and denote it by $\mathbf{CRR}_{\boldsymbol{\pi}}$.

The next theorem describes the relationship between the drift and volatility of the model CRR_π and the drift and volatility of the model CRR_ν.

Theorem 4 *The following holds*

$$\mu_{\pi_{\nu,T},T} = \mu_\nu + \sigma_\nu \frac{1}{\sqrt{\Delta t}} \frac{(\pi_{\nu,T} - \nu)}{\sqrt{\nu(1-\nu)}}$$
$$\sigma^2_{\pi_{\nu,T},T} = \sigma_v^2 \frac{\pi_{\nu,T}(1 - \pi_{\nu,T})}{\nu(1-\nu)}$$

Proof. For the sake of readability, let us write $\pi_{\nu,T} = \pi$, $u_{\nu,T} = u$ and $d_{\nu,T} = d$. Then

$$\begin{aligned}
&\frac{\mu_{\pi,T}-\mu_\nu}{\sigma_\nu}\\
&=\frac{\frac{1}{\Delta t}(\pi\log u+(1-\pi)\log d)-\frac{1}{\Delta t}(\nu\log u+(1-\nu)\log d)}{\frac{1}{\sqrt{\Delta t}}\sqrt{\nu(1-\nu)}(\log u-\log d)}\\
&=\frac{\frac{1}{\Delta t}(\pi-\nu)(\log u-\log d)}{\frac{1}{\sqrt{\Delta t}}\sqrt{\nu(1-\nu)}(\log u-\log d)}\\
&=\frac{1}{\sqrt{\Delta t}}\frac{\pi-\nu}{\sqrt{\nu(1-\nu)}}
\end{aligned}$$

From here we solve for $\mu_{\pi,T}$ to get the desired result. The computation for $\sigma_{\pi,T}$ proceeds as follows

$$\begin{aligned}
\frac{\sigma_{\pi,T}}{\sigma_v^2}&=\frac{\frac{1}{\Delta t}\pi(1-\pi)(\log u-\log d)^2}{\frac{1}{\Delta t}\nu(1-\nu)(\log u-\log d)^2}\\
&=\frac{\pi(1-\pi)}{\nu(1-\nu)}
\end{aligned}$$

as desired.□

Now we can compute the required limits in order to use Theorem 1 in the context of the model CRR$_\pi$.

Theorem 5 *The following limits hold.*

$$\begin{aligned}
\lim_{T\to\infty}\mu_{\pi_{\nu,T},T}&=r-\frac{\sigma_\nu^2}{2}\\
\lim_{T\to\infty}\sigma_{\pi_{\nu,T},T}&=\sigma_\nu
\end{aligned}$$

where r is the risk-free rate.
Proof. Theorem 4 gives

$$\mu_{\pi_{\nu,T},T}=\mu_\nu+\sigma_\nu\frac{1}{\sqrt{\Delta t}}\frac{(\pi_{\nu,T}-\nu)}{\sqrt{\nu(1-\nu)}}$$

and so

$$\lim_{T\to\infty}\mu_{\pi_{\nu,T},T}=\mu_v+\frac{\sigma_v}{\sqrt{\nu(1-\nu)}}\lim_{\Delta t\to 0}\left(\frac{\pi_{\nu,T}-\nu}{\sqrt{\Delta t}}\right)$$

Now we use the formula for $\pi_{\nu,T}$ from Theorem 2 to evaluate the limit

$$\lim_{\Delta t\to 0}\frac{1}{\sqrt{\Delta t}}\left(\frac{e^{(r-\mu_v)\Delta t+\frac{\nu}{\sqrt{\nu(1-\nu)}}\sigma_v\sqrt{\Delta t}}-1}{e^{\frac{1}{\sqrt{\nu(1-\nu)}}\sigma_v\sqrt{\Delta t}}-1}-\nu\right)$$

This is a bit messy. Either l'Hopital's rule and a strong constitution or a symbolic algebra software package gives the limit

$$\frac{2(r-\mu_v)-\sigma_v^2}{2\sigma_v}\sqrt{\nu(1-\nu)}$$

and so

$$\begin{aligned}\lim_{T\to\infty}\mu_{\pi_{\nu,T},T} &= \mu_v+\frac{\sigma_v}{\sqrt{\nu(1-\nu)}}\left[\frac{2(r-\mu_v)-\sigma_v^2}{2\sigma_v}\sqrt{\nu(1-\nu)}\right]\\ &= r-\frac{\sigma_v^2}{2}\end{aligned}$$

as desired. For the limit of the sequence $\sigma_{\pi_{\nu,T},T}$ we begin with the formula

$$\sigma^2_{\pi_{\nu,T},T}=\sigma_v^2\left(\frac{\pi_{\nu,T}(1-\pi_{\nu,T})}{\nu(1-\nu)}\right)$$

from Theorem 4. Since by Theorem 2 we have

$$\lim_{\Delta t\to 0}\pi_{\nu,T}=\nu$$

it follows that

$$\lim_{\Delta t\to 0}\frac{\pi_{\nu,T}(1-\pi_{\nu,T})}{\nu(1-\nu)}=1$$

and so

$$\lim_{\Delta t\to 0}\sigma^2_{\pi_{\nu,T},T}=\sigma_\nu^2$$

Taking square roots gives the desired result.□

Now Theorem 1 will give us the desired weak limit.

Theorem 6 *Let $S_{t,T}$ be the final stock price for the CRR$_\pi$ model with martingale measure up-tick probability $\pi_{\nu,T}$, lifetime $[0,t]$ and T equal-sized subintervals of length Δt. Let*

$$H_{t,T} = \log\left(\frac{S_{t,T}}{S_0}\right)$$

be the logarithmic price growth. Under the martingale measure we have

$$H_{t,T} \overset{dist(\pi_T)}{\longrightarrow} H_t = \left(r - \frac{\sigma_\nu^2}{2}\right)t + \sigma_\nu W_{\pi,t}$$

and

$$S_{t,T} \overset{dist(\pi_T)}{\longrightarrow} S_t = S_0 e^{\left(r - \frac{\sigma_\nu^2}{2}\right)t + \sigma_\nu W_{\pi,t}}$$

where

$$\{W_{\pi,t} \mid t \geq 0\}$$

is a standard Brownian motion process. Hence, the logarithmic growth

$$\{H_{\pi,t} \mid t \geq 0\}$$

is a Brownian motion process with

1) *mean* $\left(r - \frac{\sigma_\nu^2}{2}\right)t$
2) *variance* $\sigma_\nu^2 t$
3) *drift* $r - \frac{\sigma_\nu^2}{2}$
4) *volatility* σ_ν

The stock price process itself

$$\{S_t \mid t \geq 0\}$$

is a geometric Brownian motion process under the martingale measure with this drift and volatility. Note also that the stock price growth S_t/S_0 *is lognormally distributed with*

$$\mathcal{E}(S_t) = S_0 e^{rt}$$
$$\text{Var}(S_t) = (S_0 e^{rt})^2 (e^{\sigma_\nu^2 t} - 1)$$

□

More on the Model From a Different Perspective: Itô's Lemma

Earlier in the chapter, we spoke of the usual approach to developing a continuous-time model, namely, the assumption that stock prices behave according to the stochastic differential equation

$$\frac{dS_t}{S_t} = \mu_0\, dt + \sigma_0\, dW_t$$

where $\{W_t \mid t \geq 0\}$ is standard Brownian motion and μ_0 and σ_0 are constants (not to be confused with μ_ν, μ_{π_T}, σ_ν and σ_{π_T}). We also mentioned that the precise meaning of this equation must remain somewhat vague for us, because it requires considerably more mathematical machinery than we will develop here. While this is true, we can "wave our hands" a bit to get some further insight into how this equation is used to develop the model. This is at least worthwhile from the perspective that the reader may encounter this equation when reading the literature. However, the reader may feel free to skip this discussion without loss of continuity.

The previous equation can be written

$$dS_t = \mu_0 S_t \, dt + \sigma_0 S_t \, dW_t$$

which is a special case of the formula

$$dS_t = a(S_t, t) \, dt + b(S_t, t) \, dW_t$$

where $a(x, t)$ and $b(x, t)$ are functions of two variables. In our case

$$a(S_t, t) = \mu_0 S_t$$
$$b(S_t, t) = \sigma_0 S_t$$

A process $\{S_t\}$ that obeys the preceding equation is sometimes called an **Itô process**.

Now, if $f(x, t)$ is a function of two variables, then we may form the composition $f(S_t, t)$, which is a stochastic process since $\{S_t\}$ is a stochastic process. We are interested in finding a formula for df. This is done by applying a result from the stochastic calculus known as Itô's lemma.

Theorem 7 ***(Itô's Lemma)*** *Let $\{S_t\}$ follow an Itô process*

$$dS_t = a(S_t, t) \, dt + b(S_t, t) \, dW_t$$

where $\{W_t\}$ is standard Brownian motion and $a(\mathrm{x}, t)$ and $b(\mathrm{x}, t)$ are functions of two variables. Let $f(x, t)$ be a (sufficiently differentiable) function of two variables. Then

$$df = \left(\frac{\partial f}{\partial x} a + \frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} b^2 \right) dt + \frac{\partial f}{\partial x} b \, dW_t \qquad \square$$

In the case at hand, we have

$$\begin{aligned} a(S_t, t) &= \mu_0 S_t \\ b(S_t, t) &= \sigma_0 S_t \end{aligned}$$

and so Itô's lemma becomes

$$df = \left(\frac{\partial f}{\partial x} \mu_0 S_t + \frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} \sigma_0^2 S_t^2 \right) dt + \frac{\partial f}{\partial x} \sigma_0 S_t \, dW_t$$

Let us now apply this formula to the function

$$f(x, t) = \log x$$

In this case

$$\begin{aligned} \frac{\partial f}{\partial x}(S_t, t) &= \frac{1}{S_t} \\ \frac{\partial f}{\partial t}(S_t, t) &= 0 \\ \frac{\partial^2 f}{\partial x^2}(S_t, t) &= -\frac{1}{S_t^2} \end{aligned}$$

and we get

$$\begin{aligned} d(\log S_t) &= \left(\frac{1}{S_t} \mu_0 S_t - \frac{1}{2} \frac{1}{S_t^2} \sigma_0^2 S_t^2 \right) dt + \frac{1}{S_t} \sigma_0 S_t \, dW_t \\ &= \left(\mu_0 - \frac{\sigma_0^2}{2} \right) dt + \sigma_0 \, dW_t \end{aligned}$$

This says that $d(\log S_t)$ is normally distributed with mean $\left(\mu_0 - \frac{\sigma_0^2}{2} \right) dt$ and variance $\sigma_0^2 \, dt$. It follows that the change in $\log S_t$ over a period $[0, t]$, that is

$$\log S_t - \log S_0$$

being the sum of independent normal random variables, is also normally distributed with mean

$$\sum \left(\mu_0 - \frac{\sigma_0^2}{2} \right) dt = \left(\mu_0 - \frac{\sigma_0^2}{2} \right) t$$

and variance

$$\sum(\sigma_0^2 dt) = \sigma_0^2 t$$

Hence,

$$H_t = \log\left(\frac{S_t}{S_0}\right) = \log S_t - \log S_0$$
$$= \left(\mu_0 - \frac{\sigma_0^2}{2}\right)t + \sigma_0\sqrt{t}Z_t$$

where Z_t is standard normal. Thus we see that (in general terms) the stochastic differential equation leads to the same conclusion as that of Theorem 6. (We have not dealt specifically here with the martingale measure.)

Are the Assumptions Realistic?

We cannot continue without a short pause to comment on whether the assumption that stock prices are lognormally distributed, or equivalently that the logarithmic growth in the stock price is normally distributed, is a crealistic one. There has been much statistical work done on this question, resulting in evidence that growth rates exhibit a phenomenon known as *leptokurtosis*, which means two things:

- the probability that the logarithmic growth is near the mean is greater than it would be for a normal distribution (higher peak)
- the probability that the logarithmic growth is far away from the mean is greater than it would be for a normal distribution (fatter tails).

There is other statistical evidence that the assumption of normality is perhaps not realistic. For a further discussion, with references, we refer the reader to Chriss [1997]. Of course, this should not necessarily come as a surprise. After all, the assumptions that lead to the formula

$$H_t = \left(r - \frac{\sigma_\nu^2}{2}\right)t + \sigma_\nu W_{\pi,t}$$

are not very realistic. Nevertheless, the Black-Scholes formula, which relies on this normal model, and to which we know turn, is the most widely used formula for option pricing.

The Black-Scholes Option Pricing Formula

We now have the tools necessary to derive the Black-Scholes option pricing formula for a European option. Consider again the complete, arbitrage-free CRR_π model.

The payoff for a European put with strike price K is

$$\begin{aligned} X &= \max\{K - S_{t,T}, 0\} \\ &= \max\{K - S_0 e^{H_{t,T}}, 0\} \end{aligned}$$

The replicating pricing strategy implies that in order to avoid arbitrage, the initial price of the put must be

$$\mathcal{I}(\text{Put}) = e^{-rt}\mathcal{E}_{\Pi_T}(\max\{K - S_0 e^{H_{t,T}}, 0\})$$

where Π_T is the martingale measure, with up-tick probability π_T. Taking limits as T tends to infinity gives

$$P_\infty = \lim_{T\to\infty} \mathcal{I}(\text{Put}) = e^{-rt} \lim_{T\to\infty} \mathcal{E}_{\Pi_T}(\max\{K - S_0 e^{H_{t,T}}, 0\})$$

where P_∞ denotes the limiting price random variable. This is a case for part 2) of Theorem 10 of Chapter 8, where g is the function

$$g(x) = \max\{K - S_0 e^x, 0\}$$

which is indeed bounded and continuous on $\mathbb{R}$. In this case, we have

$$P_\infty = e^{-rt} \lim_{T\to\infty} \mathcal{E}_{\Pi_T}(g(H_{t,T}))$$

and since

$$H_{t,T} \overset{\text{dist}(\pi_T)}{\longrightarrow} H_t = \left(r - \frac{\sigma_\nu^2}{2}\right)t + \sigma_\nu \sqrt{t} Z_{\pi,t}$$

Theorem 10 of Chapter 8 implies that

$$P_\infty = e^{-rt} \lim_{T\to\infty} \mathcal{E}_{\pi_T}(g(H_{t,T})) = e^{-rt}\mathcal{E}(g(H_t))$$

Since H_t is normally distributed with mean

$$a = \left(r - \frac{\sigma_\nu^2}{2}\right)t$$

and variance

$$b^2 = \sigma_\nu^2 t$$

we have

$$P_\infty = e^{-rt}\mathcal{E}(g(H_t)) = \frac{e^{-rt}}{\sqrt{2\pi b^2}}\int_{-\infty}^{\infty} g(x)e^{-\frac{(x-a)^2}{2b^2}}dx$$

(Here the π under the square root sign is just pi, not the martingale uptick probability.)

All we need to do now is evaluate this integral. Making the substitution

$$y = \frac{x-a}{b}$$

gives

$$P_\infty = \frac{e^{-rt}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g(by+a)e^{-\frac{y^2}{2}}dy$$

Now the function

$$g(by+a) = \max\{K - S_0 e^{by+a}, 0\}$$

is nonzero if and only if

$$K - S_0 e^{by+a} > 0$$

that is, if and only if

$$by + a < \log(\frac{K}{S_0})$$

or finally

$$y < \frac{1}{b}\left[\log(\frac{K}{S_0}) - a\right]$$

Let us denote the right-hand side of this by h

$$h = \frac{1}{b}\left[\log(\frac{K}{S_0}) - a\right]$$

We can write the integral as follows

$$P_\infty = \frac{e^{-rt}}{\sqrt{2\pi}} \int_{-\infty}^{h} (K - S_0 e^{by+a}) e^{-\frac{y^2}{2}} dy$$

Splitting this into two integrals gives

$$P_\infty = \frac{Ke^{-rt}}{\sqrt{2\pi}} \int_{-\infty}^{h} e^{-\frac{y^2}{2}} dy - \frac{S_0 e^{-rt}}{\sqrt{2\pi}} \int_{-\infty}^{h} e^{by+a} e^{-\frac{y^2}{2}} dy$$

The first of these integrals is in pretty good shape because

$$\frac{Ke^{-rt}}{\sqrt{2\pi}} \int_{-\infty}^{h} e^{-\frac{y^2}{2}} dy = Ke^{-rt} \phi_{0,1}(h)$$

where $\phi_{0,1}(x)$ is the standard normal distribution function. The second integral could use some work

$$\begin{aligned}
\frac{S_0 e^{-rt}}{\sqrt{2\pi}} \int_{-\infty}^{h} e^{by+a} e^{-\frac{y^2}{2}} dy &= \frac{S_0 e^{-rt}}{\sqrt{2\pi}} \int_{-\infty}^{h} e^{-\frac{1}{2}(y^2 - 2by - 2a)} dy \\
&= \frac{S_0 e^{-rt}}{\sqrt{2\pi}} \int_{-\infty}^{h} e^{-\frac{1}{2}[(y-b)^2 - b^2 - 2a]} dy \\
&= \frac{S_0 e^{-rt}}{\sqrt{2\pi}} e^{\frac{1}{2}[b^2+2a]} \int_{-\infty}^{h} e^{-\frac{1}{2}(y-b)^2} dy \\
&= S_0 e^{-rt+\frac{1}{2}b^2+a} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{h-b} e^{-\frac{1}{2}y^2} dy \\
&= S_0 e^{-rt+\frac{1}{2}b^2+a} \phi_{0,1}(h-b)
\end{aligned}$$

Thus

$$P_\infty = Ke^{-rt} \phi_{0,1}(h) - S_0 e^{-rt+\frac{1}{2}b^2+a} \phi_{0,1}(h-b)$$

Now we have a pleasant surprise with respect to the exponent in the second term

$$\begin{aligned}
-rt + \frac{1}{2}b^2 + a &= -rt + \frac{1}{2}t\sigma_\nu^2 + \frac{t}{2}(2r - \sigma_\nu^2) \\
&= -rt + \frac{1}{2}t\sigma_\nu^2 + tr - \frac{1}{2}t\sigma_\nu^2) \\
&= 0
\end{aligned}$$

and so we arrive at our final destination

$$P_\infty = Ke^{-rt}\phi_{0,1}(h) - S_0\phi_{0,1}(h - \sqrt{t}\sigma_\nu)$$

where

$$h = \frac{1}{b}\left[\log(\frac{K}{S_0}) - a\right] = \frac{1}{\sqrt{t}\sigma_\nu}\left[\log(\frac{K}{S_0}) - tr + \frac{1}{2}t\sigma_\nu^2\right]$$

This is the famous Black-Scholes formula for the value of a European put.

We can use the put-call option parity formula to get the corresponding price of a European call. Recall that this formula says that the price of a call is given by

$$C = P + S_0 - Ke^{-rt}$$

Taking limits as T tends to ∞ gives

$$\begin{aligned} C_\infty &= P_\infty + S_0 - Ke^{-rt} \\ &= Ke^{-rt}(\phi_{0,1}(h) - 1) - S_0(\phi_{0,1}(h - \sqrt{t}\sigma_\nu) - 1) \end{aligned}$$

Since

$$\phi_{0,1}(-t) = 1 - \phi_{0,1}(t)$$

the price of the call is

$$C_\infty = -Ke^{-rt}\phi_{0,1}(-h) + S_0\phi_{0,1}(\sqrt{t}\sigma_\nu - h)$$

By setting $d_2 = -h$ and $d_1 = d_2 + \sigma\sqrt{t}$ as seems to be commonly done, we get the formulas shown in the following theorem.

Theorem 8 *(The Black-Scholes Option Pricing Formulas) For European options with strike price K and expiration time t we have*

$$\begin{aligned} C &= S_0\phi_{0,1}(d_1) - Ke^{-rt}\phi_{0,1}(d_2) \\ P &= Ke^{-rt}\phi_{0,1}(-d_2) - S_0\phi_{0,1}(-d_1) \end{aligned}$$

where S_0 is the initial price of the underlying stock, σ is the instantaneous volatility, $\phi_{0,1}$ is the standard normal distribution function and

$$d_1 = \frac{1}{\sigma\sqrt{t}}\left[\log(\frac{S_0}{K}) + t(r + \frac{1}{2}\sigma^2)\right]$$
$$d_2 = \frac{1}{\sigma\sqrt{t}}\left[\log(\frac{S_0}{K}) + t(r - \frac{1}{2}\sigma^2)\right] = d_1 - \sigma\sqrt{t}$$

where r is the risk-free rate.□

We note that these formulas do not involve the instantaneous drift. In fact, the only "unknown" quantities are the instantaneous volatility σ_ν and the risk-free rate r.

EXAMPLE 2 Consider a European call option on a stock that is currently selling for \$100. The option expires in 1 year at a strike price of \$100. Suppose that the risk-free rate is 0.05 and that the volatility is $\sigma = 0.15$ per year. Compute the value of the call.
Solution This is simply a matter of plugging into the formula. First, we have

$$d_2 = \frac{1}{0.15}\left[0 + 0.05 - \frac{1}{2}(0.15)^2\right] \approx 0.2583$$

and

$$d_1 = d_2 + \sigma\sqrt{t} \approx 0.2583 + 0.15 = 0.4083$$

Hence, with the aid of a calculator or some other means of evaluating values of $\phi_{0,1}$, we get

$$C = 100\phi_{0,1}(0.4083) - 100e^{-0.05}\phi_{0,1}(0.2583) \approx \$8.596 \qquad \square$$

How Black-Scholes is Used in Practice: Volatility Smiles and Surfaces

The assumption that the volatility σ is constant is a very unrealistic one, to say the least. This (among other things) raises questions about the quality of the Black-Scholes option pricing formula. A great deal of research has been done to determine how the Black-Scholes formula can be used in light of the questionable assumptions about the parameters. Our intention is to very briefly discuss one method that is used in practice.

The Volatility Smile

We have said that the volatility can be estimated using historical data. However, in practice, the Black-Scholes formula is not used by simply plugging in an estimate for the volatility and grinding out option prices. Instead traders usually work with a quantity known as the *implied volatility*. Loosely speaking, this is the volatility that must be used in the Black-Scholes option pricing formula in order to make the formula reflect the actual market price at a given moment in time.

Definition *Consider a European option that has a particular market price of* M. *The* **implied volatility** *of this option is the volatility that is required in the Black-Scholes option pricing formula so that the formula gives* M.□

The implied volatility is, in effect, the market's opinion about the Black-Scholes volatility of a stock. The implied volatility is a quantity that can be computed from the Black-Scholes formula by numerical methods (that is, educated guessing and reguessing).

As it happens, and as further evidence that the Black-Scholes formula is not perfect, if one computes the implied volatility of otherwise identical options at various strike prices, one gets different values. Figure 5 shows a typical graph of implied volatility versus strike price. Because of the shape of the graph, it is known as a **volatility smile**.

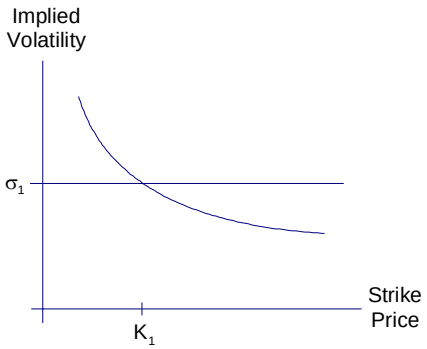


Figure 5 – A volatility smile

Now, suppose that an historical estimate of volatility for a particular stock is σ_1. Then the Black-Scholes formula gives an option price that matches the market price for only one strike price, namely K_1.

For larger strike prices, which correspond to out-of-the-money calls and in-the-money puts, the market price corresponds to a volatility that is much less than σ_1. Now, the Black-Scholes formula gives prices that vary directly with the volatility, so a smaller volatility produces a smaller price. Hence the market price is below the Black-Scholes historical price. Put another way, the Black-Scholes formula, when used with a constant volatility based on historical data, tends to *overprice* out-of-the-money calls and in-the-money puts relative to market prices.

For smaller strike prices, which correspond to in-the-money calls and out-of-the-money puts, the market price corresponds to a volatility that is much greater than σ_1. Hence the market price is above the Black-Scholes historical price. Put another way, the Black-Scholes formula, when used with a constant volatility based on historical data, tends to *underprice* in-the-money calls and out-of-the-money puts relative to market prices.

The Volatility Surface

To understand how the Black-Scholes formula may be used in practice, condier Table 1, which shows data for a *volatility surface*, that is, a set of implicit volatilities for various maturity dates as well as strike prices. The columns represent different strike prices expressed as a percent of the stock price. (The data in this table is for illustration only.)

	90%	95%	100%	105%	110%
1 month	14	13	12	113	14.4
3 months	14.2	14.1	13.6	13.8	14.1
6 months	14.5	14.3	14.2	14.5	14.6
1 year	15.1	15	14.6	14.7	14.8
2 years	16.2	16.1	16	16.1	16.2

Table 1 – Data for a volatility surface

Now, a trader who wants to price an option that has a maturity and strike price that is not in this table can interpolate from the table to get an implied volatility. For example, consider an option that matures in 9 months at a strike price of 95% of stock price. A linear interpolation between the 6 month and 1 year maturity implied volatilities gives a volatility of $(14.3 + 15)/2 = 14.65$. This volatility can be used in the Black-Scholes formula to prodce a price for the option in question.

How Dividends Effect the Use of Black-Scholes

The Black-Scholes option pricing formula assumes that the underlying stock does not pay a dividend. We should briefly discuss how dividends can be handled in this context.

First a bit of background on dividends. There are four dates that are important with respect to dividends. The *declaration date* is the date that the board of directors declares a dividend. The *record date* is the date that the registrar compiles a list of current shareholders, to whom the dividend will be paid. The key point is that an investor must be on record as owning the stock on the record date or else he will not receive a dividend. The *payment date* is the date that the dividend will be paid.

Now, normal stock purchases take 3 days to clear. This is referred to as *regular way* settlement. Thus, an owner must have purchased the stock at least 3 days prior to the record date in order to be considered an *owner of record* on the record date and hence eligible for the dividend. The first date after this date is called the *ex-dividend date*. For example, if a dividend is declared as payable to stockholders of record on a given Friday then the New York Stock Exchange (who sets the ex-dividend dates for NYSE stocks) would declare the stock “ex-dividend” as of the opening of the market on the preceding Wednesday.

We note that when the stock goes ex-dividend, typically the stock price will decline by an amount similar to the amount of the dividend. This makes sense from the perspective that the dividends are known in advance and are therefore built into the stock's price in some way.

Now, a European option on a stock that pays a dividend can be thought of as composed of two separate processes: a risky process that represents the stock price itself and a risk-free process that represents the cash dividend payments. Thus, to price an option on such a stock, we first discount all of the forthcoming dividend payments to the present. If this amount is d then we can think of the components as a risky stock that has initial value $S_0 - d$ and a risk-free asset that has initial value d. The Black-Scholes formula can then be applied to the risky stock.

Exercises

1. A stock has the initial price of \$50. Over a five-day period, the stock price at day's end is given by \$49.82, \$50.02, \$49.69, \$49.34, \$50.10. Estimate the instantaneous drift and volatility.

2. Consider a European call option on a stock that is currently selling for \$80. The option expires in 1 year at a strike price of \$80. Suppose that the risk-free rate is 0.05 and that the volatility is $\sigma = 0.1$ per year. Compute the value of the call.
3. Consider a European put option on a stock that is currently selling for \$50. The option expires in 1 year at a strike price of \$51. Suppose that the risk-free rate is 0.04 and that the volatility is $\sigma = 0.15$ per year. Compute the value of the put.
4. Prove that
$$\mathrm{Var}_p(X_{p,T,i}) = 1$$
5. Prove that
$$\mathcal{E}_p(Q_s) = \sigma_s \frac{t}{\sqrt{\Delta t}} \frac{p-s}{\sqrt{s(1-s)}}$$
$$\mathrm{Var}_p(Q_s) = \sigma_s^2 t \frac{p(1-p)}{s(1-s)}$$
6. Let X be a random variable with a lognormal distribution, that is, $Y = \log X$ is normally distributed with mean a and variance b^2. Show that
   a) $\mathcal{E}(X) = e^{a+\frac{1}{2}b^2}$
   b) $\mathrm{Var}(X) = e^{2a+b^2}(e^{b^2}-1)$
   Apply this to the random variable $X = S_t/S_0$ to deduce that
   c) $\mathcal{E}_\pi(S_t) = S_0 e^{rt}$
   d) $\mathrm{Var}_\pi(S_t) = (S_0 e^{rt})^2(e^{\sigma_\nu^2 t}-1)$
7. Show that the function $f(x) = \max\{K - S_0 e^x, 0\}$ is continuous and bounded on $\mathbb{R}$.
8. Show that $\phi_{0,1}(-t) = 1 - \phi_{0,1}(t)$ for the standard normal distribution function. *Hint*: draw a picture using the graph of the standard normal density function.
9. Show using l'Hopital's rule that
$$\lim_{T\to\infty} \pi_T = \lim_{\Delta t \to 0} \left(\frac{e^{(r-\mu)\Delta t + \frac{p}{\sqrt{pq}}\sigma\sqrt{\Delta t}} - 1}{e^{\frac{1}{\sqrt{pq}}\sigma\sqrt{\Delta t}} - 1} \right) = p$$
   Hint: write $x = (\Delta t)^2$.
10. Suppose we assume that there is a martingale probability measure in the limiting case when $T \to \infty$ and that the Fundamental Theorem of Asset pricing holds in this limiting case. If we denote this martingale measure by Π then

$$S_0 = e^{-rt}\mathcal{E}_\Pi(S_t) = e^{-rt}\mathcal{E}_\Pi(S_0 e^{\mu_\nu t + \sigma_\nu \sqrt{t} Z_t})$$

Evaluate the last expression to show that

$$\mu_\nu = r - \frac{1}{2}\sigma_\nu$$

11. Let N_U be the random variable representing the *number of* up-ticks in stock price over the lifetime of the CRR model. Show that

$$S_{t,T} = S_0 e^{N_U(\log u - \log d) + T\log d}$$

and so

$$H_{t,T} = N_U(\log u - \log d) + T\log d$$

Show that

$$\begin{aligned}\mathcal{E}(H_{t,T}) &= T\nu(\log u - \log d) + T\log d \\ \text{Var}(H_{t,T}) &= T\nu(1-\nu)\,(\log u - \log d)^2\end{aligned}$$

12. Standardize the random variable $H_{t,T}$ to show that

$$H^*_{t,T} = N^*_U$$

Hence, the random variables $H^*_{t,T}$ are standardized binomial random variables.

13. Using the Black-Scholes formulas show that the value of a put and a call increases as the volatility σ increases. Looking at the profit curves for a long put and long call, explain why this makes sense. Does the same effect obtain for the owner of a stock?
14. Show that the probability that a European call with strike price K and expiration date t will expire in or at the money is

$$\phi_{0,1}\left(\frac{\log(S_T/K) - t(r - \sigma_\nu^2/2)}{\sigma\sqrt{t}}\right)$$

where $\phi_{0,1}$ is the standard normal distribution.

Chapter 10

Optimal Stopping and American Options

The models that we have created thus far, including the Black-Scholes option pricing model, are designed to price European options, which are options that can only be exercised at the expiration time. However, in the real world, most stock options are of the American variety. In this chapter, we want to take a look at the issue of pricing American options.

American options are far more complicated than European options, because they give up nothing but add one major additional feature—they can be exercised at any time between the purchase date and the expiration date. This is clearly a significant interpolation, since there is no way to look into the future to decide when to execute. An investor cannot call his broker and say "If the stock price falls below $50 then sell the stock *before* it falls."

The mathematics used to model American options has a significantly different flavor and is a bit more sophisticated than we have thus far encountered.

An Example

To aid the discussion, let us set up a simple example to which we will refer in the sequel.

EXAMPLE 1 Figure 1 shows a CRR model state tree with stock prices (and option payoffs).

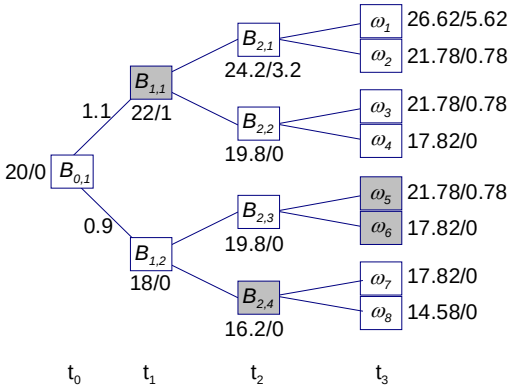


Figure 1 – A CRR model state tree

We will assume that $r = 0$. Note that for this model

$$T = 3,\ u = 1.1,\ d = 0.9$$

and the martingale measure probability is

$$\pi = \frac{1-d}{u-d} = \frac{0.1}{0.2} = \frac{1}{2}$$

Finally, let C be an American call C with strike price $K = 21$.□

Before beginning, a bit of notation. For any random variable X, it will be convenient to define the shorthand notation $[X \in A]$ as

$$[X \in A] = \{\omega \in \Omega \mid X(\omega) \in A\}$$

The Model

In general, our context will be a discrete-time model that is arbitrage-free and thus has a martingale measure Π, as is the case for Example 1. Consider an investment in an American option (also called an American *claim*). At any of the model's times

$$t_0 < t_1 < \cdots < t_T$$

the owner may exercise the option.

The Payoffs

The payoff of the option at any time t_k is a random variable Y_k. We will assume that (Y_k) is adapted to a filtration $\mathbb{F} = (\mathcal{P}_k)$. For our example, the payoffs of the American call C are

$$Y_3 = \max\{S_3 - 21, 0\} = \begin{cases} 5.62 & \text{for } \omega = \omega_1 \\ 0.78 & \text{for } \omega = \omega_2, \omega_3, \omega_5 \\ 0 & \text{otherwise} \end{cases}$$

$$Y_2 = \max\{S_2 - 21, 0\} = \begin{cases} 3.2 & \text{for } \omega \in B_{2,1} \\ 0 & \text{otherwise} \end{cases}$$

$$Y_1 = \max\{S_1 - 21, 0\} = \begin{cases} 1 & \text{for } \omega \in B_{1,1} \\ 0 & \text{otherwise} \end{cases}$$

and finally

$$Y_0 = 0$$

These payoffs are shown in Figure 1 as well.

Stopping Times

The decision about when to exercise can be modeled as a random variable with special properties, called a *stopping time*. The idea is quite simple: a stopping time is a rule, that is, a random variable that identifies, for each time t_k the outcomes in Ω for which we should exercise at time t_k. Let us refer to these outcomes as the *stopping event* for time t_k.

There is one slight flaw in this terminology, however. The stopping event for the *final* time t_T is meant to include all outcomes for which we have not yet stopped (exercised). However, in some of these cases, we may not actually exercise the option, but rather let the option expire worthless. So the notion of an "exercise event" applies literally only to the intermediate times t_k with $k < T$.

The only requirement for a stopping time is an obvious one: we must be able to tell which outcomes belong to the stopping event for time t_k *at that time*. This is an important issue. We cannot say that the exercise event for time t_2, for example, is based on what happens at time t_3. This is akin to asking our broker to sell the stock *before* its price drops below 50.

Definition *A* **(bounded) stopping time** *is a random variable*

$$\tau\colon \Omega \to \{0, \ldots, T\}$$

whose range is the set of integers from 0 *to* T*. Moreover, it is required that the* **stopping event** *"stop at time* t_k*" defined by*

$$[\tau = k] = \{\omega \in \Omega \mid \tau(\omega) = k\}$$

is in the algebra $\mathcal{A}(\mathcal{P}_k)$ *generated by* $\mathcal{P}_k$ *for all* $k = 0, \ldots, T$*. We will denote the set of all stopping times with range* $\{k, \ldots, T\}$ *by* $\mathcal{S}_{k,T}$*. These are the stopping times that cannot stop before time* t_k*.*□

Let us consider an important example of stopping times.

EXAMPLE 2 It would not be uncommon for an investor to tell his broker to "sell the stock if the price reaches $50 or more" for example. This rule is a stopping time. In fact, it is referred to mathematically as the

first entry time of the stock price process (S_k) into the set $[50, \infty)$. Formally, it is defined as follows

$$\tau(\omega) = \begin{cases} \min\{k \mid S_k \geq 50\} & \text{if } \{k \mid S_k \geq 50\} \neq \emptyset \\ T & \text{otherwise} \end{cases}$$

It is not hard to show that this is indeed a stopping time. For if $k < T$ then we have

$$\begin{aligned} [\tau = k] &= \{\omega \mid S_k \geq 50 \text{ but } S_j < 50 \text{ for } j < k\} \\ &= [S_0 < 50] \cup \cdots \cup [S_{k-1} < 50] \cup [S_k \geq 50] \end{aligned}$$

But since the price S_i is $\mathcal{P}_i$-measurable and since $(\mathcal{P}_i)$ is a filtration, we deduce that each of the events $[S_i < 50]$ and the event $[S_k \geq 50]$ are in the largest algebra $\mathcal{A}(\mathcal{P}_k)$. This is the condition required of a stopping time. Finally, for $k = T$ we have

$$[\tau = T] = [S_0 < 50] \cap \cdots \cap [S_{T-1} < 50] \in \mathcal{A}(\mathcal{P}_T)$$

Thus, τ is a stopping time. Note that the same argument will work for any value other than 50.

In fact, it is also possible to show that the first entry time inot any Borel set B is a stopping time. For example, the set

$$B = (-\infty, 17) \cup (20, \infty)$$

corresponds to the first time that the stock price drops below 17 or rises above 20. The shaded blocks in Figure 1 show the stopping events for the first entry time into B.□

Stopping the Payoff Process

Here is the scenario. Imagine that an investor has acquired an American option at time t_0. The investor sits down and looks at all possible stopping times in the set $\mathcal{S}_{0,T}$. (This is possible at least in theory since there are only a finite number of such stopping times.)

Suppose that the investor has decided upon a particular stopping time $\tau \in \mathcal{S}_{0,T}$ to use in determining when to exercise. We will discuss how this decision is made a bit later. In fact, that is the main issue of the chapter.

It may help to think of the investor as phoning his broker at time t_0 and giving him the stopping time rule τ. From this point on, the broker can

proceed without bothering the investor. In particular, at each time t_k the broker checks to see if the current state of the economy is in the stopping event $[\tau = k]$ for that time. He can do this because $[\tau = k] \in \mathcal{A}(\mathcal{P}_k)$ is just a union of the blocks of $\mathcal{P}_k$ and the broker knows which of the blocks represents the current state at that time. If the current state is in $[\tau = k]$, then the broker exercises the option; otherwise he does not.

The Stopped Value of an American Option

In order to determine how to choose a stopping time, we must first discuss the consequences of any choice of stopping time. Suppose that the investor has decided upon a particular stopping time $\tau \in \mathcal{S}_{0,T}$. Then for any $\omega \in \Omega$, the option will exercise at time $t_{\tau(\omega)}$ and give a payoff of

$$Y_{\tau(\omega)}(\omega)$$

It is customary to denote this function by Y_τ. Thus

$$[Y_\tau](\omega) = Y_{\tau(\omega)}(\omega)$$

The random variable Y_τ is refered to as the **final value** of the process (Y_k) under the stopping time τ.

EXAMPLE 3 Referring to Example 1, consider the stopping time τ shown in gray in Figure 1. This is the first entry time into

$$B = (-\infty, 17) \cup (20, \infty)$$

The (discounted) final value Y_τ is

$$Y_\tau(\omega) = \begin{cases} 1 & \text{if } \omega \in \{\omega_1, \omega_2, \omega_3, \omega_4\} \\ 0.78 & \text{if } \omega = \omega_5 \\ 0 & \text{if } \omega \in \{\omega_6, \omega_7, \omega_8\} \end{cases}$$ □

A Detail About Discounting

Now we must discuss a detail concerning discounting. If X_k is any process and $\tau \in \mathcal{S}_{0,T}$ is a stopping time then the final value is X_τ. To discount this value, we must discount each of the values $(X_\tau)(\omega) = X_{\tau(\omega)}(\omega)$ by the appropriate amount

$$\overline{X_{\tau(\omega)}(\omega)} = \frac{X_{\tau(\omega)}(\omega)}{B_{\tau(\omega)}(\omega)} = \left(\frac{X_\tau}{B_\tau}\right)(\omega)$$

Thus, we set

$$\overline{X}_\tau = \frac{X_\tau}{B_\tau}$$

Note that in Example 1 we assume that the risk-free rate is 0 and so the issue of discounting is not relevant.

The Initial Value of an American Option, or What to Do At Time t_0

At time t_0 the possible choices for stopping times are the elements of the set $\mathcal{S}_{0,T}$. If the investor stops the payoff process using a particular stopping time $\tau \in \mathcal{S}_{0,T}$ then he will realize the final value Y_τ. However, there is a subtlety here, namely, for each $\omega \in \Omega$ the payoff $Y_{\tau(\omega)}(\omega)$ comes at time $t_{\tau(\omega)}$. In keeping with the spirit of self-financing trading strategies, we will assume that the investor does not remove the funds from his brokerage account until the end of the model, at time t_T and so the payoff at time $t_{\tau(\omega)}$ is allowed to grow at the risk-free rate. This results in a **final payoff** of

$$\frac{B_T}{B_{\tau(\omega)}} Y_{\tau(\omega)}(\omega)$$

where $B_k = e^{rk}$ is the discounting factor.

Thus, the final time-t_T payoff resulting from the stopping time τ is really

$$\frac{B_T}{B_\tau} Y_\tau = B_T \overline{Y}_\tau$$

where

$$\overline{Y}_\tau = \frac{Y_\tau}{B_\tau}$$

Put another way, each stopping time turns an American option into a guaranteed sequence of payoffs, where the time-t_k payoff is

$$Y_k 1_{[\tau=k]}$$

The final value of this payoff stream is

$$\sum_{k=0}^{T} \frac{B_T}{B_k} Y_k 1_{[\tau=k]} = \frac{B_T}{B_\tau} Y_\tau = B_T \overline{Y}_\tau$$

The Initial Value of an American Option

Now let us consider what the investor should do to determine which stopping time to employ at time t_0. Keep in mind that the investor may change his mind at time t_1, but we will not worry about that problem yet.

In order to determine the best stopping time at time t_0, as mentioned earlier, the investor can look at all possible stopping times in the finite set $\mathcal{S}_{0,T}$. At first, it seems logical that the investor should choose the stopping time that maximizes the final payoff

$$\max_{\tau \in \mathcal{S}_{0,T}} \{B_T \overline{Y}_\tau\}$$

However, the payoffs $B_T \overline{Y}_\tau$ are functions (random variables) and there is no guarantee that there is a single stopping time that is best for all states $\omega \in \Omega$. Indeed, this is highly unlikely.

So an alternative is needed. This involves computing the *initial value* of the final payoff. Assuming that the final payoff $B_T \overline{Y}_\tau$ is attainable, there is a self-financing trading strategy Φ for which

$$\mathcal{V}_T(\Phi) = B_T \overline{Y}_\tau$$

and the arbitrage-free price of this final payoff is (according to the martingale measure condition)

$$\begin{aligned} V_0(\tau) &= \mathcal{I}(B_T \overline{Y}_\tau) \\ &= \mathcal{V}_0(\Phi) \\ &= \frac{B_0}{B_T} \mathcal{E}_\Pi(\mathcal{V}_T(\Phi)) \\ &= \frac{1}{B_T} \mathcal{E}_\Pi(B_T \overline{Y}_\tau) \\ &= \mathcal{E}_\Pi(\overline{Y}_\tau) \end{aligned}$$

The quantity $V_0(\tau)$ is the **initial value** of the American option under τ. We have shown that *assuming that the owner of the option follows the stopping time* τ, the non-arbitrage price of the claim must be $V_0(\tau)$.

Now, at time t_0 the investor *can* choose a stopping time that maximizes the *initial value* of the option, since these values are constants. So let us define

$$V_0 = \max_{\tau \in \mathcal{S}_{0,T}} V_0(\tau) = \max_{\tau \in \mathcal{S}_{0,T}} \mathcal{E}_\Pi(\overline{Y}_\tau)$$

Then we can assume that the investor will choose a stopping time τ^* with the property that

$$\mathcal{E}_\Pi(\overline{Y}_{\tau^*}) = V_0$$

This is the stopping time that maximizes the initial value of the option or, equivalently, the *expected* final payoff $\mathcal{E}_\Pi(\overline{Y}_\tau)$ under the martingale measure. It is called an **optimal stopping time**. With no vision into the future, this is the best that can be done.

Let us look a bit more closely at this situation as it relates to arbitrage. For a *European* option with attainable final payoff

$$\mathcal{V}_T(\Phi) = B_T \overline{Y}_{\tau^*}$$

the non-arbitrage initial value of the option must be $\mathcal{V}_0(\Phi) = V_0$ for if not, then an investor could buy the cheaper of Φ and the option and sell the more expensive of the two investments, realizing an immediate positive return. Then, at the final time t_T the *European* option must still be open and must have payoff $\mathcal{V}_T(\Phi)$. Hence, the two ending positions (one long and one short) will cancel each other out, leaving the investor with the future value of his initial profit.

However, for an American option, the situation is not as simple because the seller of the option does not really know what the final payoff will be. (The owner doesn't know either, but at least he has some control over the value.)

We can say that if the option can be *purchased* for an amount A that is less than $\mathcal{V}_0(\Phi)$ then arbitrage is available to the investor who purchases the cheaper option and shorts the more expensive trading strategy Φ. As with the European option, there is an initial profit and offsetting positions at the end, assuming that the owner does not exercise the option until expiration.

However, if the option is offered for an amount A that is greater than $\mathcal{V}_0(\Phi)$ then the immediate profit is made by the *seller* of the option (who also buys Φ). But the seller does not have control over the option and cannot be certain that the owner will not find a way to achieve a higher payoff than that given by Φ.

Thus, an arbitrage-free argument does lead to the inequality $A \geq \mathcal{V}_0(\Phi)$. On the other hand, who would be willing to pay more than $\mathcal{V}_0(\Phi)$ for the American option when there is no guarantee that a payoff greater than $\mathcal{V}_T(\Phi)$, obtained by following a time-t_0 optimal stopping time, can be arranged? Thus, we come to the conclusion that $\mathcal{V}_0(\Phi)$ is a fair price (more-or-less) for the American option.

EXAMPLE 4 Again referring to Example 1, we have seen that the final payoff for the first entry time into

$$B = (-\infty, 17) \cup (20, \infty)$$

is

$$Y_\tau(\omega) = \begin{cases} 1 & \text{if } \omega \in \{\omega_1, \omega_2, \omega_3, \omega_4\} \\ 0.78 & \text{if } \omega = \omega_5 \\ 0 & \text{if } \omega \in \{\omega_6, \omega_7, \omega_8\} \end{cases}$$

Hence,

$$\mathcal{E}_\Pi(\overline{Y}_\tau) = \frac{1}{2} \cdot 1 + \frac{1}{8} \cdot 0.78 + \frac{3}{8} \cdot 0 = 0.5975$$

Consider the stopping time σ defined by

$$\sigma(\omega) = \begin{cases} 2 & \text{if } \omega \in \{\omega_1, \omega_2, \omega_7, \omega_8\} \\ 3 & \text{otherwise} \end{cases}$$

We leave it to the reader to show that this is a stopping time. In this case, the (discounted) final value is

$$Y_\sigma(\omega) = \begin{cases} 3.2 & \text{if } \omega \in \{\omega_1, \omega_2\} \\ 0.78 & \text{if } \omega = \{\omega_3, \omega_5\} \\ 0 & \text{if } \omega \in \{\omega_4, \omega_6, \omega_7, \omega_8\} \end{cases}$$

Hence,

$$\mathcal{E}_\Pi(\overline{Y}_\sigma) = \frac{1}{4} \cdot 3.2 + \frac{1}{4} \cdot 0.78 = 0.995$$

Hence, σ is a better stopping time than τ. In fact, as we will see, σ is an optimal stopping time.□

What to Do At Time t_k

Now suppose that at time t_k the investor has not yet exercised an American option. Then the previous discussion is still valid *mutatis mutandis* (that is, with the necessary changes). In particular, the choice of stopping times must now be made from the set $\mathcal{S}_{k,T}$ since at time t_k the investor cannot exercise at any earlier time.

If the investor stops the payoff process using a particular stopping time $\tau \in \mathcal{S}_{k,T}$ then he will realize the value Y_τ, whose final time-t_T value is

$$\frac{B_T}{B_\tau} Y_\tau = B_T \overline{Y}_\tau$$

as before. Assuming that the final payoff $B_T \overline{Y}_\tau$ is attainable, there is a self-financing trading strategy Φ for which

$$\mathcal{V}_T(\Phi) = B_T \overline{Y}_\tau$$

and the arbitrage-free time-t_k price of this final payoff is

$$\begin{aligned} V_k(\tau) &= \mathcal{I}_k(B_T \overline{Y}_\tau) \\ &= \mathcal{V}_k(\Phi) \\ &= \frac{B_k}{B_T} \mathcal{E}_\Pi(\mathcal{V}_T(\Phi) \mid \mathcal{F}_k) \\ &= \mathcal{E}_\Pi(B_k \overline{Y}_\tau \mid \mathcal{F}_k) \\ &= \mathcal{E}_\Pi(\overline{Y}_\tau) \end{aligned}$$

The quantity $V_k(\tau)$ is the **time-t_k value** of the American option under τ.

Now, at time t_k we again assume that the investor makes the best possible stopping decision, which in this case amounts to choosing a stopping time τ^* for which $V_k(\tau^*)$ is maximized. Accordingly, let us define V_k by

$$V_k = \max_{\tau \in \mathcal{S}_{k,T}} V_k(\tau) = \max_{\tau \in \mathcal{S}_{k,T}} \mathcal{E}_\Pi(B_k \overline{Y}_\tau \mid \mathcal{F}_k)$$

and say that a stopping time τ^* is *optimal* if

$$\mathcal{E}_\Pi(B_k \overline{Y}_{\tau^*} \mid \mathcal{F}_k) = V_k$$

We call (V_k) the **value process** of the American option.

Let us take a moment to compare the decision process at time t_0, namely to maximize according to

$$V_0 = \max_{\tau \in \mathcal{S}_{0,T}} V_0(\tau) = \max_{\tau \in \mathcal{S}_{0,T}} \mathcal{E}_\Pi(\overline{Y}_\tau)$$

and the time t_k decision process, namely, to maximize according to

$$V_k = \max_{\tau \in \mathcal{S}_{k,T}} V_k(\tau) = \max_{\tau \in \mathcal{S}_{k,T}} \mathcal{E}_\Pi(B_k \overline{Y}_\tau \mid \mathcal{F}_k)$$

In the latter case, we are taking the maximum over a smaller set, since

$$\mathcal{S}_{k,T} \subseteq \mathcal{S}_{0,T}$$

so this would tend to make the maximum smaller. On the other hand, at time t_k we are maximizing with more information, that is, we are maximizing the *conditional* expectations $\mathcal{E}_\Pi(B_k \overline{Y}_\tau \mid \mathcal{F}_k)$, which would tend to make the maximum larger. Thus, we have two conflicting forces, the result of which is that we cannot say anything about which is larger.

Definition *For an American option with payoff process* $(Y_k \mid k = 0, \ldots, T)$ *the arbitrage-free* **value process** *is*

$$V_k = \max_{\tau \in \mathcal{S}_{k,T}} \mathcal{E}_\Pi(B_k \overline{Y}_\tau \mid \mathcal{F}_k)$$

and the **discounted value process** *is*

$$\overline{V}_k = \max_{\tau \in \mathcal{S}_{k,T}} \mathcal{E}_\Pi(\overline{Y}_\tau \mid \mathcal{F}_k)$$

The discounted value process $(\overline{V}_k)$ *is called the* **Snell envelop** *of the discounted payoff process* $(\overline{Y}_k)$.□

Definition *A stopping time* τ^* *is* **optimal** *for the interval* $[k, T]$ *if it maximizes the expected discounted payoff process* $(\overline{Y}_k)$, *that is*

$$\mathcal{E}_\Pi(\overline{Y}_{\tau^*} \mid \mathcal{F}_k) = \overline{V}_k = \max_{\tau \in \mathcal{S}_{k,T}} \mathcal{E}_\Pi(\overline{Y}_\tau \mid \mathcal{F}_k)$$

that is, if τ^* *achieves the best expected discounted payoff.*□

Optimal Stopping Times and the Snell Envelop

To simplify the notation, we will study Snell envelops in terms of arbitrary nondiscounted processes. The only difference is whether or not we need to include the overbar.

Definition *If* $\mathbb{Z} = (Z_k \mid k = 0, \ldots, T)$ *is a random process then the process* $\mathbb{U} = (U_k)$ *defined by*

$$U_k = \max_{\tau \in \mathcal{S}_{k,T}} \mathcal{E}_\Pi(Z_\tau \mid \mathcal{F}_k)$$

is called the **Snell envelop** *of* $\mathbb{Z}$*. A stopping time* τ^* *that realizes this maximum over the interval* $[k, T]$*, that is, for which*

$$\mathcal{E}_\Pi(Z_{\tau^*} \mid \mathcal{F}_k) = U_k = \max_{\tau \in \mathcal{S}_{k,T}} \mathcal{E}_\Pi(Z_\tau \mid \mathcal{F}_k)$$

is called an **optimal stopping time** *for* $\mathbb{Z}$ *over* $[k, T]$. □

Thus, if $\mathbb{Z}$ is the discounted payoff process of an American option then the Snell envelop $\mathbb{U}$ is the discounted value process of the option.

We will compute the Snell envelop of the payoff process (Y_k) from Example 1 a bit later, when we have some additional tools that will make the computation simpler.

Existence of Optimal Stopping Times

The very first question that should be addressed with respect to optimal stopping times is whether or not they exist. For $k = 0$ it is clear that optimal stopping times exist because we are simply maximizing a finite set of constants $\mathcal{E}_\Pi(Z_\tau)$. But for $k > 0$ we are maximizing nonconstant functions $\mathcal{E}_\Pi(Z_\tau \mid \mathcal{F}_k)$.

Theorem 1 *For any interval* $[k, T]$ *an optimal stopping time for* $\mathbb{Z}$ *exists.*
Proof. Recall that for

$$\mathcal{P}_k = \{B_{k,1}, \ldots, B_{k,c}\}$$

the conditional expectation is defined by

$$\mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_k) = \sum_{u=1}^{c} \mathcal{E}_\Pi(Z_\tau \mid B_{k,u}) 1_{B_{k,u}}$$

Thus, for each $B_{k,u}$ the random variable $\mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_k)$ is equal to the *constant* $\mathcal{E}_\Pi(Z_\tau \mid B_{k,u})$ on $B_{k,u}$. Hence, we can find a stopping time $\tau_{k,u}$ which maximizes these constants, that is, for which

$$\mathcal{E}_\Pi(Z_{\tau_{k,u}} \mid B_{k,u}) = \max_{\tau \in \mathcal{S}_{k,T}} \mathcal{E}_\Pi(Z_\tau \mid B_{k,u})$$

Consider the random variable

$$\tau_k^* = \sum_{u=1}^{c} \tau_{k,u} 1_{B_{k,u}}$$

which maximizes the conditional expectation on each block of $\mathcal{P}_k$. To see that τ^* is a stopping time in $\mathcal{S}_{k,T}$, note that $\tau_k^* \geq k$ and for any $h \geq k$

$$[\tau_k^* = h] = \bigcup_{v=1}^{c} ([\tau_k^* = h] \cap B_{k,v}) = \bigcup_{v=1}^{c} ([\tau_{k,v} = h] \cap B_{k,v}) \in \mathcal{A}(\mathcal{P}_h)$$

as required of a stopping time.

Now, if $\omega \in B_{k,v}$ then

$$\left[Z_{\tau_k^*}\right](\omega) = Z_{\tau_k^*(\omega)}(\omega) = Z_{\tau_{k,v}(\omega)}(\omega)$$

and so

$$Z_{\tau_k^*} = \sum_{u=1}^{c} Z_{\tau_{k,u}} 1_{B_{k,u}}$$

Hence, for any $\tau \in \mathcal{S}_{k,T}$

$$\begin{aligned}
\mathcal{E}_\Pi(Z_{\tau_k^*} \mid \mathcal{P}_k) &= \sum_{u=1}^{c} \mathcal{E}_\Pi(Z_{\tau_{k,u}} 1_{B_{k,u}} \mid \mathcal{P}_k) \\
&\geq \sum_{u=1}^{c} \mathcal{E}_\Pi(Z_\tau 1_{B_{k,u}} \mid \mathcal{P}_k) \\
&= \sum_{u=1}^{c} \mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_k) 1_{B_{k,u}} \\
&= \mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_k) \sum_{u=1}^{c} 1_{B_{k,u}} \\
&= \mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_k)
\end{aligned}$$

as required of an optimal stopping time.□

We should also prove that the Snell envelop is $\mathbb{F}$-adapted.

Theorem 2 *The Snell envelop* (U_k) *is* $\mathbb{F}$*-adapted.*

Proof. The random variable U_k is the maximum of a finite number of random variables $\mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_k)$, each of which is $\mathcal{P}_k$-measurable. Hence, so is U_k.□

Characterizing the Snell Envelop

Consider again the situation of the investor who, at time t_k needs to make a decision among the stopping times in $\mathcal{S}_{k,T}$. When casting about for an optimal stopping time, he can divide the candidates $\mathcal{S}_{k,T}$ into three subsets based on the stopping event for the current time t_k.

The investor could simply decide to stop now (at time t_k) and be done with it, he could decide not to stop at time t_k under any circumstances, or he could decide upon a stopping time that may stop now or may stop later, depending on the state of the economy. In symbols, the set $\mathcal{S}_{k,T}$ is the disjoint union

$$\mathcal{S}_{k,T} = \mathcal{S}_{k+1,T} \cup \mathcal{S}_{k,k} \cup \mathcal{S}^*_{k,T}$$

corresponding to the following:

1) Do not stop now, that is, stop at time t_{k+1} or later

$$\mathcal{S}_{k+1,T} = \{\tau \in \mathcal{S}_{k,T} \mid [\tau = k] = \emptyset\}$$

2) Stop now, at time k

$$\mathcal{S}_{k,k} = \{k1_\Omega\} = \{\tau \in \mathcal{S}_{k,T} \mid [\tau = k] = \Omega\}$$

3) May stop at time t_k or may stop later

$$\mathcal{S}^*_{k,T} = \{\tau \in \mathcal{S}_{k,T} \mid [\tau = k] \neq \emptyset, \Omega\}$$

We wish to show that the Snell envelop can be computed without the need to consider stopping times of type 3). Note that we are *not* saying that there is no optimal stopping time of type 3), but only that the values U_k can be computed without regard to stopping times of type 3).

The mathematical version of this statement is that the Snell envelop satisfies

$$U_k = \max\{Z_k, \max_{\tau \in \mathcal{S}_{k+1,T}} \mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_k)\}$$

Note that the maximum is now being taken over the set $\mathcal{S}_{k+1,T}$.

Theorem 3 The Snell envelop satisfies

$$U_k = \max\{Z_k, \max_{\tau \in \mathcal{S}_{k+1,T}} \mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_k)\}$$

for all $k = 0, \ldots, T$.

Proof. First, since $\tau = k$ is a (constant) stopping time and since $\mathcal{S}_{k+1,T} \subseteq \mathcal{S}_{k,T}$, we clearly have

$$\begin{aligned} U_k &= \max_{\tau \in \mathcal{S}_{k,T}} \mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_k) \\ &\geq \max\{\mathcal{E}_\Pi(Z_k \mid \mathcal{P}_k), \max_{\tau \in \mathcal{S}_{k+1,T}} \mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_k)\} \\ &= \max\{Z_k, \max_{\tau \in \mathcal{S}_{k+1,T}} \mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_k)\} \end{aligned}$$

We must establish the reverse inequality. Let $\tau \in \mathcal{S}_{k,T}$ and consider the stopping time τ' defined from τ by postponing any stopping from time t_k to t_{k+1}, that is

$$\tau'(\omega) = \max\{\tau, k+1\} = \begin{cases} \tau(\omega) & \text{if } \omega \in [\tau > k] \\ k+1 & \text{if } \omega \in [\tau = k] \end{cases}$$

Since the maximum of two stopping times is a stopping time, we have $\tau' \in \mathcal{S}_{k+1,T}$.

Now, since $[\tau > k] = [\tau = k]^c \in \sigma(\mathcal{P}_k)$ we have

$$\begin{aligned} \mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_k) &= \mathcal{E}_\Pi(Z_\tau 1_{[\tau=k]} \mid \mathcal{P}_k) + \mathcal{E}_\Pi(Z_\tau 1_{[\tau>k]} \mid \mathcal{P}_k) \\ &= \mathcal{E}_\Pi(Z_k 1_{[\tau=k]} \mid \mathcal{P}_k) + \mathcal{E}_\Pi(Z_{\tau'} 1_{[\tau>k]} \mid \mathcal{P}_k) \\ &\leq \max\{Z_k, \mathcal{E}_\Pi(Z_{\tau'} \mid \mathcal{P}_k)\} \\ &= \max\{Z_k, \max_{\sigma \in \mathcal{S}_{k+1,T}} \mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_k)\} \end{aligned}$$

But the left-hand side is valid for all $\tau \in \mathcal{S}_{k,T}$ and so

$$U_k = \max_{\tau \in \mathcal{S}_{k,T}} \mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_k) \leq \max\{Z_k, \max_{\sigma \in \mathcal{S}_{k+1,T}} \mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_k)\}$$

as desired.□

The most important use of the previous formula is that from it we can derive a backwards recurrence relation for U_k. Note that

$$U_T = \max_{\tau \in \mathcal{S}_{T,T}} \mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_T) = \mathcal{E}_\Pi(Z_T \mid \mathcal{P}_T) = Z_T$$

This provides the initial step in the backwards recurrence.

Let us look more closely at the random variable

$$X = \max_{\tau \in \mathcal{S}_{k+1,T}} \mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_k)$$

that appears in the Theorem 3. If the conditioning was with respect to $\mathcal{P}_{k+1}$ then X would just be U_{k+1}. This prompts us to condition further and use the tower property of conditional expectation. First, we need to mention that in general for any two random variables X and Y we have

$$\max\{\mathcal{E}(X \mid \mathcal{P}), \mathcal{E}(X \mid \mathcal{P})\} \leq \mathcal{E}(\max\{X, Y\} \mid \mathcal{P})$$

We leave proof of this as an exercise. Now we have

$$\begin{aligned} X &= \max_{\tau \in \mathcal{S}_{k+1,T}} \mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_k) \\ &= \max_{\tau \in \mathcal{S}_{k+1,T}} \mathcal{E}_\Pi(\mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_{k+1}) \mid \mathcal{P}_k) \\ &\leq \mathcal{E}_\Pi(\max_{\tau \in \mathcal{S}_{k+1,T}} \{\, \mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_{k+1})\} \mid \mathcal{P}_k)) \\ &= \mathcal{E}_\Pi(U_{k+1} \mid \mathcal{P}_k) \end{aligned}$$

and so

$$X \leq \mathcal{E}_\Pi(U_{k+1} \mid \mathcal{P}_k)$$

For the reverse inequality, let $\tau^* \in \mathcal{S}_{k+1,T}$ be an optimal stopping for the interval $[t_{k+1}, T]$, that is,

$$U_{k+1} = \mathcal{E}_\Pi(Z_{\tau^*} \mid \mathcal{P}_{k+1})$$

Then

$$\begin{aligned} \mathcal{E}_\Pi(U_{k+1} \mid \mathcal{P}_k) &= \mathcal{E}_\Pi(\mathcal{E}_\Pi(Z_{\tau^*} \mid \mathcal{P}_{k+1}) \mid \mathcal{P}_k) \\ &= \mathcal{E}_\Pi(Z_{\tau^*} \mid \mathcal{P}_k) \\ &\leq \max_{\tau \in \mathcal{S}_{k+1,T}} \mathcal{E}_\Pi(Z_\tau \mid \mathcal{P}_k) \\ &= X \end{aligned}$$

Hence,

$$X = \mathcal{E}_\Pi(U_{k+1} \mid \mathcal{P}_k)$$

and we arrive at the following recurrence relation for the Snell envelop.

Theorem 4 *The Snell envelop satisfies the backward recurrence relation*
1) $U_T = Z_T$

2)

$$U_k = \max\{Z_k, \mathcal{E}_\Pi(U_{k+1} \mid \mathcal{P}_k)\}$$

for all $k = 0, \ldots, T-1$.□

Now we can compute the Snell envelop of the payoff process from Example 1.

EXAMPLE 5 Again referring to Example 1, let us compute the Snell envelop of the payoff process (Y_k). First, we have

$$U_3 = Y_3$$

Next, we need

$$\mathcal{E}_\Pi(U_3 \mid \mathcal{P}_2) = \mathcal{E}_\Pi(Y_3 \mid \mathcal{P}_2) = \begin{cases} \frac{1}{2}(5.26 + 0.78) = 3.2 & \text{if } \omega \in B_{2,1} \\ \frac{1}{2}(0.78 + 0) = 0.39 & \text{if } \omega \in B_{2,2} \\ \frac{1}{2}(0.78 + 0) = 0.39 & \text{if } \omega \in B_{2,3} \\ 0 & \text{if } \omega \in B_{2,4} \end{cases}$$

from which we get

$$U_2 = \max\{Y_2, \mathcal{E}_\Pi(U_3 \mid \mathcal{P}_2)\} = \begin{cases} 3.2 & \text{if } \omega = \omega_1, \omega_2 \\ 0.39 & \text{if } \omega = \omega_3, \omega_4 \\ 0.39 & \text{if } \omega = \omega_5, \omega_6 \\ 0 & \text{if } \omega = \omega_7, \omega_8 \end{cases}$$

Next, we need

$$\mathcal{E}_\Pi(U_2 \mid \mathcal{P}_1) = \begin{cases} \frac{1}{2}(3.2 + 0.39) = 1.795 & \text{if } \omega = \omega_1, \omega_2, \omega_3, \omega_4 \\ \frac{1}{2}(0.39 + 0) = 0.195 & \text{if } \omega = \omega_5, \omega_6, \omega_7, \omega_8 \end{cases}$$

which gives

$$\begin{aligned} U_1 &= \max\{Y_1, \mathcal{E}_\Pi(U_2 \mid \mathcal{P}_1)\} \\ &= \mathcal{E}_\Pi(U_2 \mid \mathcal{P}_1) \\ &= \begin{cases} 1.795 & \text{if } \omega = \omega_1, \omega_2, \omega_3, \omega_4 \\ 0.195 & \text{if } \omega = \omega_5, \omega_6, \omega_7, \omega_8 \end{cases} \end{aligned}$$

Finally,

$$U_0 = \max\{Y_0, \mathcal{E}_\Pi(U_1 \mid \mathcal{P}_0)\} = \max\{0, \frac{1}{2}(1.795 + 0.195)\} = 0.995$$

Let us recall from Example 4 that the stopping time σ defined by

$$\sigma(\omega) = \begin{cases} 2 & \text{if } \omega \in \{\omega_1, \omega_2, \omega_7, \omega_8\} \\ 3 & \text{otherwise} \end{cases}$$

has expected (discounted) final value

$$\mathcal{E}_\Pi(\overline{Y}_\sigma) = \frac{1}{4} \cdot 3.2 + \frac{1}{4} \cdot 0.78 = 0.995$$

which is equal to U_0. Hence, σ is indeed an optimal stopping time. In fact, as we will see, σ is the *smallest* optimal stopping time in the sense that it stops before any other optimal stopping time. (Observe that we could have waited until time t_3 in states ω_7 and ω_8 and still achieved optimality.)□

This is a good time to emphasize a point about optimal stopping times, namely, optimal stopping times represent the best *guess* as to when to stop without being able to see into the future. Thus, an optimal stopping time is not *guaranteed* to produce the best possible payoff. Indeed, looking at Figure 1, it is clear that the best possible exercise procedure involves exercising at time t_2 if the final state is ω_2 but waiting until time t_3 if the final state is ω_1. However, at time t_2 we do not know which state will prevail: ω_1 or ω_2 so this plan is *not* a stopping time.

The Smallest Dominating Supermartingale

It is clear from condition 2) of Theorem 4 that

$$\mathcal{E}_\Pi(U_{k+1} \mid \mathcal{P}_k) \leq U_k$$

which is the condition that U_k be a supermartingale. An $\mathbb{F}$-adapted process (X_k) is an $\mathbb{F}$-**supermartingale** if

$$\mathcal{E}(X_{k+1} \mid \mathcal{P}_k) \leq X_k$$

It is also clear that

$$Z_k \leq U_k$$

that is, U_k **dominates** Z_k. It is not hard to see using the recurrence relation that U_k is the *smallest* supermartingale that dominates Z_k.

Theorem 5 *The Snell envelop U_k is the smallest $\mathbb{F}$-supermartingale that dominates Z_k.*

Proof. We have seen that U_k is a supermartingale that dominates Z_k. Suppose that V_k is a supermartingale that dominates Z_k. This is

equivalent to the single inequality

$$V_k \geq \max\{Z_k, \mathcal{E}_\Pi(V_{k+1} \mid \mathcal{P}_k)\}$$

We can now proceed by backward induction using the recurrence relation. For the basis step in the induction, we have

$$V_T \geq Z_T = U_T$$

Assuming that $V_{k+1} \geq U_{k+1}$ then

$$V_k \geq \max\{Z_k, \mathcal{E}_\Pi(V_{k+1} \mid \mathcal{P}_k)\} \geq \max\{Z_k, \mathcal{E}_\Pi(U_{k+1} \mid \mathcal{P}_k)\} = U_k$$

and we are done.□

Additional Facts About Martingales

In order to continue our discussion of optimal stopping times, we need some additional results relating to martingales and supermartingales.

Theorem 6

1) If $\mathbb{X}$ is an $\mathbb{F}$-martingale then for all $j \leq k$

$$\mathcal{E}(X_j) = \mathcal{E}(X_k)$$

2) If $\mathbb{X}$ is an $\mathbb{F}$-supermartingale then for all $j \leq k$

$$\mathcal{E}(X_j) \geq \mathcal{E}(X_k)$$

Proof. For a martingale, we have

$$\mathcal{E}(X_k \mid \mathcal{P}_j) = X_j$$

Taking expected values and using the tower property gives

$$\mathcal{E}(X_k) = \mathcal{E}(\mathcal{E}(X_k \mid \mathcal{P}_j)) = \mathcal{E}(X_j)$$

For submartingales, the proof is similar.□

Stopping a Process: Doob's Optional Stopping Theorem

We begin with a formal definition of sample path for a stochastic process.

Definition *Consider a stochastic process $\mathbb{X} = (X_0, \ldots, X_T)$. For each element $\omega \in \Omega$ the sequence*

$$(X_0(\omega), \ldots, X_T(\omega))$$

*is called a **sample path**.* □

Intuitively, to stop a stochastic process, we stop the sample path for each $\omega \in \Omega$ when τ tells us to do so, that is, at time $\tau(\omega)$. Thus, a sample path looks like this

$$X_0(\omega),\ X_1(\omega), \ldots, X_{\tau(\omega)}(\omega), X_{\tau(\omega)}(\omega), \ldots$$

Thus, the indices in this path are equal to $n \wedge \tau(\omega) = \min\{n, \tau(\omega)\}$ and we can write

$$X_{0\wedge\tau(\omega)}(\omega),\ X_{1\wedge\tau(\omega)}(\omega), \ldots, X_{n\wedge\tau(\omega)}(\omega), X_{(n+1)\wedge\tau(\omega)}(\omega), \ldots$$

Definition *Let $\mathbb{X} = (X_0, \ldots, X_T)$ be a stochastic process adapted to a filtration $\mathbb{F}$ and let τ be a stopping time τ on $\mathbb{F}$. The* **stopped process** *or* **sampled process** *is defined by*

$$\mathbb{X}^\tau = (X_k^\tau) = (X_{k\wedge\tau}) = (X_{k\wedge 0}, \ldots, X_{k\wedge T})$$

(Note that the first three expressions are just notation for the fourth.)□

Observe that for each n

$$X_{n\wedge\tau} = \sum_{i=1}^{n-1} X_i 1_{\{\tau=i\}} + X_n 1_{\{\tau\geq n\}} \tag{1}$$

The following theorem is one of the key results in martingale theory. (A stopping time is also called an *optional* random variable.)

Theorem 7 *(Doob's Optional Sampling Theorem) Let $\mathbb{X} = (X_k)$ be a martingale (or supermartingale) and τ a stopping time. Then the stopped process $\mathbb{X}^\tau$ is also a martingale (or supermartingale).*
Proof. We know that

$$\mathcal{E}(X_n \mid \mathcal{F}_{n-1}) = X_{n-1}$$

Starting from the expression (1) we have

$$\mathcal{E}(X_{n\wedge\tau} \mid \mathcal{F}_{n-1}) = \sum_{i=1}^{n-1} \mathcal{E}(X_i 1_{\{\tau=i\}} \mid \mathcal{F}_{n-1}) + \mathcal{E}(X_n 1_{\{\tau\geq n\}} \mid \mathcal{F}_{n-1})$$

Now, this simplifies quite a bit since $\{\tau = i\} \in \mathcal{A}(\mathcal{P}_{n-1})$ for $i \leq n-1$, X_i is $\mathcal{P}_{n-1}$-measurable and $\{\tau \geq n\} \in \mathcal{A}(\mathcal{P}_{n-1})$ this becomes

$$\begin{aligned}\mathcal{E}(X_{n\wedge\tau} \mid \mathcal{F}_{n-1}) &= \sum_{i=1}^{n-1} X_i 1_{\{\tau=i\}} + 1_{\{\tau\geq n\}}\mathcal{E}(X_n \mid \mathcal{F}_{n-1}) \\ &= \sum_{i=1}^{n-1} X_i 1_{\{\tau=i\}} + 1_{\{\tau\geq n\}} X_{n-1} \\ &= \sum_{i=1}^{n-2} X_i 1_{\{\tau=i\}} + 1_{\{\tau=n-1\}} X_{n-1} + 1_{\{\tau\geq n\}} X_{n-1} \\ &= \sum_{i=1}^{n-2} X_i 1_{\{\tau=i\}} + 1_{\{\tau\geq n-1\}} X_{n-1} \\ &= X_{(n-1)\wedge\tau}\end{aligned}$$

as desired. The proof for a supermartingale is almost identical.□

The Doob Decomposition

Finally, we need the following decomposition result.

Theorem 8 *(The Doob Decomposition) Let $\mathbb{X} = (X_0, \ldots, X_T)$ be an $\mathbb{F}$-adapted stochastic process.*

1) There is a unique martingale $\mathbb{M} = (M_0, \ldots, M_T)$ and a unique predictable process $\mathbb{A} = (A_1, \ldots, A_T)$ such that

$$X_k = M_k - A_k$$

with $A_0 = 0$.

2) If $\mathbb{X}$ is a supermartingale then $\mathbb{A}$ is nondecreasing, that is $A_{k+1} \geq A_k$.

Proof. For part 1), set $M_0 = X_0$, $A_0 = 0$ and for $k > 0$

$$M_k = \sum_{i=1}^{k} [X_i - \mathcal{E}_\Pi(X_i \mid \mathcal{F}_{i-1})] + X_0$$

and $A_k = M_k - X_k$. Then it is easy to check that the desired properties hold. For part 2), suppose that $\mathbb{X}$ is a supermartingale. Then

$$\begin{aligned}M_k - A_k = X_k &\geq \mathcal{E}(X_{k+1} \mid \mathcal{F}_k) \\ &= \mathcal{E}(M_{k+1} \mid \mathcal{F}_k) - \mathcal{E}(A_{k+1} \mid \mathcal{F}_k) \\ &= M_k - A_{k+1}\end{aligned}$$

and so $A_k \leq A_{k+1}$.□

Characterizing Optimal Stopping Times

Note that if two discrete random variables satisfy $X \leq Y$ and $\mathcal{E}_\Pi(X) = \mathcal{E}_\Pi(Y)$ then because Π is strongly positive, it follows that $X = Y$. This fact will prove very useful.

Armed with the previous additional facts about martingales and supermartingales, we can return to the matter at hand, namely characterizing optimal stopping times.

We begin by exploring what happens if we stop the process (U_k). Since U_k is a supermartingale, it follows by Doob's optional sampling theorem that for any stopping time $\tau \in \mathcal{S}_{0,T}$ the stopped process (U_k^τ) is also a supermartingale, that is

$$\mathcal{E}_\Pi(U_{k+1}^\tau \mid \mathcal{P}_k) \leq U_k^\tau$$

Moreover, since $Z_k \leq U_k$ it follows that the final values satisfy $Z_\tau \leq U_\tau$.

The fact that (U_k^τ) is a supermartingale and $Z_\tau \leq U_\tau$ implies the chain of inequalitites

$$\mathcal{E}_\Pi(Z_\tau) \leq \mathcal{E}_\Pi(U_\tau) = \mathcal{E}_\Pi(U_T^\tau) \leq \cdots \leq \mathcal{E}_\Pi(U_k^\tau) \leq \cdots \leq \mathcal{E}_\Pi(U_0^\tau) = U_0$$

Now, if τ^* is an *optimal* stopping time for $[0, T]$, that is,

$$U_0 = \mathcal{E}_\Pi(Z_{\tau^*})$$

then the previous sequence of inequalitites becomes a sequence of equalitites. The first of these equalities

$$\mathcal{E}_\Pi(Z_{\tau^*}) = \mathcal{E}_\Pi(U_{\tau^*})$$

implies, since $Z_{\tau^*} \leq U_{\tau^*}$ and Π is strongly positive, that $Z_{\tau^*} = U_{\tau^*}$. (See theremark at the beginning of this section.) Looking further down the chain of inequalities we also see that

$$\mathcal{E}_\Pi(U_k^{\tau^*}) = \mathcal{E}_\Pi(U_{k-1}^{\tau^*})$$

The supermartingale property now implies that $(U_k^{\tau^*})$ is in fact a martingale. To see this, we know that $(U_k^{\tau^*})$ is a supermartingale,

$$\mathcal{E}_\Pi(U_k^{\tau^*} \mid \mathcal{P}_{k-1}) \leq U_{k-1}^{\tau^*}$$

Taking the expected value of the left side gives, by the tower property

$$\mathcal{E}_\Pi(\mathcal{E}_\Pi(U_k^{\tau^*} \mid \mathcal{P}_{k-1})) = \mathcal{E}_\Pi(U_k^{\tau^*})$$

and do both sides of the inequality have the same expected value and hence are the same. In other words, $(U_k^{\tau^*})$ is a martingale

$$\mathcal{E}_\Pi(U_k^{\tau^*} \mid \mathcal{P}_{k-1}) = U_{k-1}^{\tau^*}$$

For the converse, suppose that $Z_\tau = U_\tau$ and that U_k^τ is a martingale. Then the sequence of inequalities is a sequence of equalities and in particular,

$$U_0 = \mathcal{E}_\Pi(Z_\tau)$$

which implies that τ is optimal for the interval $[0, T]$.

We now have a characterization of optimal stopping times.

Theorem 9 *A stopping time $\tau \in \mathcal{S}_{0,T}$ is optimal for the interval $[0, T]$ if and only if*
1) $Z_\tau = U_\tau$
2) U_k^τ is a martingale.□

Optimal Stopping Times and the Doob Decomposition

We have seen that a stopping time τ is optimal if and only if $Z_\tau = U_\tau$ and $\mathbb{U}^\tau = (U_k^\tau)$ is a martingale. This prompts us to take a closer look at when $\mathbb{U}^\tau$ is a martingale.

We have seen that the Snell envelop

$$\mathbb{U} = (U_0, \ldots, U_T)$$

is a supermartingale. Using Doob's decomposition, we can write

$$\mathbb{U} = \mathbb{M} - \mathbb{A}$$

where $\mathbb{M} = (M_0, \ldots, M_T)$ is a martingale and $\mathbb{A} = (A_1, \ldots, A_T)$ is predictable and nondecreasing and $A_0 = 0$.

Suppose we now stop this sequence

$$\mathbb{U}^\tau = \mathbb{M}^\tau - \mathbb{A}^\tau = (M_0^\tau - A_0^\tau, \ldots, M_T^\tau - A_T^\tau)$$

Since the difference of martingales is a martingale and since $\mathbb{M}^\tau$ is a martingale, we see that $\mathbb{U}^\tau$ is a martingale if and only if

$$\mathbb{A}^\tau = (A_0^\tau, \ldots, A_T^\tau)$$

is a (predictable) martingale. However, it is easy to see that a predictable martingale is a constant sequence, that is

$$A_0^\tau = \cdots = A_T^\tau$$

But $A_0^\tau = 0$ and so $\mathbb{U}^\tau$ is a martingale if and only if $\mathbb{A}^\tau$ is the zero process.

Now, for any $\omega \in \Omega$ the sequence $A_k^\tau(\omega)$ is

$$A_0(\omega), \ldots, A_{\tau(\omega)-1}(\omega), A_{\tau(\omega)}(\omega), \ldots, A_{\tau(\omega)}(\omega)$$

Since this sequence is nondecreasing from 0, it is the zero sequence if and only if

$$[A_\tau](\omega) = A_{\tau(\omega)}(\omega) = 0$$

Hence, $\mathbb{A}^\tau = 0$ if and only if $A_\tau = 0$, that is, U^τ is a martingale if and only if $A_\tau = 0$.

Theorem 10 *Let $\mathbb{U} = (U_k)$ be the Snell envelop of (Z_k). For a stopping time $\tau \in \mathcal{S}_{0,T}$ the process stopped process $\mathbb{U}^\tau = (U_k^\tau)$ is a martingale if and only if $A_\tau = 0$ where $\mathbb{A} = (A_k)$ is the predictable process in the Doob decomposition of $\mathbb{U}$.*□

The Smallest Optimal Stopping Time

The previous theorem makes it easy to determine the smallest optimal stopping time. First, we recall the recurrence formula

$$U_k = \max\{Z_k, \mathcal{E}_\Pi(U_{k+1} \mid \mathcal{P}_k)\}$$

Using the Doob decomposition, we notice that

$$\begin{aligned}\mathcal{E}_\Pi(U_{k+1} \mid \mathcal{P}_k) &= \mathcal{E}_\Pi(M_{k+1} \mid \mathcal{P}_k) - \mathcal{E}_\Pi(A_{k+1} \mid \mathcal{P}_k) \\ &= M_k - A_{k+1} \\ &= (M_k - A_k) - (A_{k+1} - A_k) \\ &= U_k - (A_{k+1} - A_k)\end{aligned}$$

and so

$$U_k = \max\{Z_k, U_k - (A_{k+1} - A_k)\}$$

Hence, the *strict* inequality

$$U_k > Z_k$$

implies that

$$A_{k+1} = A_k$$

It follows that *prior to* the first time t_k that $U_k = Z_k$ we do have the strict inequality $U_i > Z_i$ ($i < k$) and so $A_{i+1} = A_i$ ($i < k$). But $A_0 = 0$ and so

$$0 = A_0 = \cdots = A_k$$

This prompts us to define $\tau_{\min}$ by

$$\tau_{\min}(\omega) = \min\{k \mid Z_k(\omega) = U_k(\omega)\}$$

which exists since $Z_T(\omega) = U_T(\omega)$. In addition, $\tau_{\min}$ is the first entry time of the adaptd process $(Z_k - U_k)$ into the Borel set $\{0\}$ and so is a stopping time. By definition we have

$$Z_{\tau_{\min}} = U_{\tau_{\min}}$$

If $\tau_{\min}$ is an optimal stopping time then it must be the smallest optimal stopping time because all optimal stopping times τ satisfy $Z_\tau = U_\tau$, that is $Z_{\tau(\omega)}(\omega) = U_{\tau(\omega)}(\omega)$. Moreover, we have just seen that

$$0 = A_0(\omega) = A_1(\omega) = \cdots = A_{\tau_{\min}(\omega)}(\omega)$$

Hence $A_{\tau_{\min}} = 0$, which implies that $\mathbb{U}^{\tau_{\min}}$ is a martingale. Finally,

Theorem 11 *The smallest optimal stopping time is*

$$\tau_{min}(\omega) \in \min\{k \mid Z_k(\omega) = U_k(\omega)\} \qquad \square$$

EXAMPLE 6 In Example 4 we defined the stopping time

$$\sigma(\omega) = \begin{cases} 2 & \text{if } \omega \in \{\omega_1, \omega_2, \omega_7, \omega_8\} \\ 3 & \text{otherwise} \end{cases}$$

and showed in Example 5 that σ is optimal. In fact, it is easy to see that σ has the property

$$\sigma(\omega) \in \min\{k \mid Z_k(\omega) = U_k(\omega)\}$$

and so it is the smallest optimal stopping time.$\square$

The Largest Optimal Stopping Time

In view of Theorem 10, in casting about for the largest optimal stopping time, it is natural to consider the function

$$\tau_{\max}(\omega) = \max\{k \mid A_k(\omega) = 0\}$$

which exists since $A_0 = 0$. Since any optimal stopping time τ satisfies $A_\tau = 0$ if $\tau_{\max}$ is an optimal stopping time then it must be the largest optimal stopping time. Note that $\tau_{\max}$ can also be defined by

$$\tau_{\max}(\omega) = \begin{cases} \min\{k \mid A_{k+1}(\omega) > 0\} & \text{if } \{k \mid A_{k+1}(\omega) > 0\} \neq \emptyset \\ T & \text{otherwise} \end{cases}$$

and since this is the first entry time (into $(0, \infty)$) of the adapted process $\mathbb{A}$ we see that $\tau_{\max}$ is a stopping time. Also, since $A_{\tau_{\max}} = 0$ we know by Theorem 10 that $(U_k^{\tau_{\max}})$ is a martingale. Thus, to show that $\tau_{\max}$ is optimal, we need only show that $U_{\tau_{\max}} = Z_{\tau_{\max}}$.

Once again we look at the recurrence relation

$$U_k = \max\{Z_k, \mathcal{E}_\Pi(U_{k+1} \mid \mathcal{P}_k)\}$$

which, using Doob's decomposition can be written

$$U_k = \max\{Z_k, U_k - (A_{k+1} - A_k)\}$$

But for $k = \tau_{\max}(\omega)$ we have

$$A_{\tau_{\max}+1}(\omega) - A_{\tau_{\max}}(\omega) = A_{\tau_{\max}+1}(\omega) > 0$$

and so the maximum above is just Z_k, that is

$$U_{\tau_{\max}}(\omega) = Z_{\tau_{\max}}(\omega)$$

Thus $U_{\tau_{\max}} = Z_{\tau_{\max}}$ as desired.

Theorem 12 *The largest optimal stopping time is*

$$\begin{aligned} \tau_{max}(\omega) &= \max\{k \mid A_k(\omega) = 0\} \\ &= \begin{cases} \min\{k \mid A_{k+1}(\omega) > 0\} & \textit{if } \{k \mid A_{k+1}(\omega) > 0\} \neq \emptyset \\ T & \textit{otherwise} \end{cases} \end{aligned}$$

where

$$U_k = M_k - A_k$$

is the Doob decomposition.□

Exercises

1. Show that the first entry time into any set of the form (a, b) is a stopping time.
2. Show that the first entry time into any Borel set B is a stopping time.
3. Show that the first time that a stock's price doubles its initial price is a stopping time.
4. Show that the first time that a stock's price doubles its previous price is a stopping time, that is, the random variable

$$\tau(\omega) = \begin{cases} \min\{k \mid S_k(\omega) \geq 2S_{k-1}(\omega)\} & \text{if } \{k \mid S_k(\omega) \geq 2S_{k-1}(\omega)\} \neq \emptyset \\ T & \text{otherwise} \end{cases}$$

   is a stopping time.
5. Show that the first *exit* time from a set of the form (a, b) is a stopping time.
6. Prove that the maximum, minimum or sum of two stopping times is a stopping time. How about the difference?
7. Prove that for any random variables X and Y

$$\max\{\mathcal{E}(X \mid \mathcal{P}), \mathcal{E}(X \mid \mathcal{P})\} \leq \mathcal{E}(\max\{X, Y\} \mid \mathcal{P})$$

8. Prove that since $Z_k \leq U_k$ for all k, it follows that $Z_\tau \leq U_\tau$ for any stopping time τ.
9. Prove that if two discrete random variables satisfy $X \leq Y$ and $\mathcal{E}_\Pi(X) = \mathcal{E}_\Pi(Y)$ then because Π is strongly positive, it follows that $X = Y$.
10. Prove that if $(A_0, \ldots, A_T)$ is a martingale and $(A_1, \ldots, A_T)$ is predictable then (A_k) is a constant sequence, that is

$$A_0^\tau = \cdots = A_T^\tau$$

11. For the CRR model with

$$u = 1.2, d = 0.8, r = 0, S_0 = 20$$

    compute the price process, payoff process for an American call with $K = 21$ and the Snell envelop. Find the first optimal stopping time.
12. Write an Excel spreadsheet that given u, d, r, S_0 and K will compute the price process, payoff process for an American call/put with strike price K and the Snell envelop.

Appendix

Convexity and the Separation Theorem

In this appendix, we develop the necessary material on convexity.

Convex, Closed and Compact Sets

We shall need the following concepts.

Definition

1) Let $x_1, \ldots, x_k \in \mathbb{R}^n$. The any linear combination of the form

$$t_1 x_1 + \cdots + t_k x_k$$

where

$$t_1 + \cdots + t_k = 1$$
$$0 \leq t_i \leq 1$$

is called a **convex combination** *of the vectors $x_1, \ldots, x_k$.*

2) A subset $X \subseteq \mathbb{R}^n$ is **convex** *if whenever $x, y \in X$ then the entire line segment between x and y also lies in X, in symbols*

$$\{sx + ty \mid s + t = 1, 0 \leq s, t \leq 1\} \subseteq X$$

3) A subset $X \subseteq \mathbb{R}^n$ is a **cone** *if $x \in X$ implies that $\alpha x \in X$ for all $\alpha > 0$.*

4) A subset $X \subseteq \mathbb{R}^n$ is **closed** *if whenever $x_n \in X$ is a convergent sequence of points in X then the limit is also in X. Simply put, a subset is closed if it is closed under the taking of limits.*

5) A subset $X \subseteq \mathbb{R}^n$ is **compact** *if it is both closed and bounded.* □

We will also have need of the following facts from analysis.

1) A continuous function that is defined on a compact set X in $\mathbb{R}^n$ takes on its maximum and minimum values at some points within the set X.
2) A subset X of $\mathbb{R}^n$ is compact if and only if every sequence in X has a subsequence that converges in X.

Theorem 1 *Let X and Y be subsets of $\mathbb{R}^n$. Define*

$$X + Y = \{a + b \mid a \in X,\ b \in Y\}$$

1) If X and Y are convex then so is $X + Y$
2) If X is compact and Y is closed then $X + Y$ is closed.
Proof. For 1) let $a_0 + b_0$ and $a_1 + b_1$ be in $X + Y$. The line segment between these two points is

$$t(a_0 + b_0) + (1 - t)(a_1 + b_1) = ta_0 + (1 - t)a_1 + tb_0 + (1 - t)b_1 \in X + Y$$

and so $X + Y$ is convex.

For part 2) let $a_n + b_n$ be a convergent sequence in $X + Y$. Suppose that $a_n + b_n \to c$. We must show that $c \in X + Y$. Since a_n is a sequence in the compact set X, it has a convergent subsequence a_{n_k} whose limit α lies in X. Since $a_{n_k} + b_{n_k} \to c$ and $a_{n_k} \to \alpha$ we can can conclude that $b_{n_k} \to c - \alpha$. Since Y is closed, we must have $c - \alpha \in Y$ and so $c = \alpha + (c - \alpha) \in X + Y$, as desired. □

Convex Hulls

We will have use for the notion of convex hull.

Definition *The* **convex hull** *of a set $S = \{x_1, \ldots, x_k\}$ of vectors in $\mathbb{R}^n$ is the smallest convex set in $\mathbb{R}^n$ that contains the vectors $x_1, \ldots, x_k$. We denote the convex hull of S by $\mathcal{C}(S)$.* □

Here is a characterization of convex hulls.

Theorem 2 *Let $S = \{x_1, \ldots, x_k\}$ be a set of vectors in $\mathbb{R}^n$. Then the convex hull $\mathcal{C}(S)$ is the set Δ*

$$\Delta = \{t_1x_1 + \cdots + t_kx_k \mid 0 \leq t_i \leq 1, \Sigma t_i = 1\}$$

of all convex combinations of vectors in S.
Proof. First, we show that Δ is convex. So let

$$X = t_1x_1 + \cdots + t_kx_k$$
$$Y = s_1x_1 + \cdots + s_kx_k$$

be convex combinations of S and let $a + b = 1, 0 \leq a, b \leq 1$. Then

$$\begin{aligned} aX + bY &= a(t_1x_1 + \cdots + t_kx_k) + b(s_1x_1 + \cdots + s_kx_k) \\ &= (at_1 + bs_1)x_1 + \cdots + (at_k + bs_k)x_k \end{aligned}$$

But this is also a convex combination of S because

$$0 \leq at_i + bs_i \leq \begin{cases} (a+b)s_i = s_i \leq 1 & t_i \leq s_i \\ (a+b)t_i = t_i \leq 1 & t_i > s_i \end{cases}$$

and

$$\sum_{i=1}^{k}(at_i + bs_i) = a\sum_{i=1}^{k}t_i + b\sum_{i=1}^{k}s_i = a + b = 1$$

Thus,

$$X, Y \in \Delta \Rightarrow aX + bY \in \Delta$$

which says that Δ is convex. It is also clear that $x_i \in \Delta$ for all i and so Δ is a convex set that contains all of the vectors in S. It follows that

$$\mathcal{C}(S) \subseteq \Delta$$

To show the reverse inclusion, we must show that any convex set that contains S must also contain Δ. So suppose that D is a convex set that contains S. Then D contains all convex combinations of any two vectors in S. Consider a convex combination, say $t_1x_1 + t_2x_2 + t_3x_3$ of three such vectors. We can write

$$t_1x_1 + t_2x_2 + t_3x_3 = t_1x_1 + (t_2 + t_3)(\frac{t_2}{t_2 + t_3}x_2 + \frac{t_3}{t_2 + t_3}x_3)$$

Now, the expression in the parentheses at the far right is a convex combination of two vectors in S and so is in D. If we denote it by d then

$$t_1x_1 + t_2x_2 + t_3x_3 = t_1x_1 + (t_2 + t_3)d$$

But the expression on the right side of the equal sign is a convex combination of two elements of D and so is in D. Thus, we see that any convex combination of three vectors in S is in D. An inductive argument along these lines, which we leave as an exercise, can be used to furnish a complete proof. □

Linear and Affine Hyperplanes

We next discuss hyperplanes in $\mathbb{R}^n$. A **linear hyperplane** in $\mathbb{R}^n$ is an $(n-1)$-dimensional subspace of $\mathbb{R}^n$. As such, it is the solution set of a linear equation of the form

$$a_1x_1 + \cdots + a_nx_n = 0$$

or

$$\alpha \cdot x = 0$$

where $\alpha = (a_1, \ldots, a_n)$ and $x = (x_1, \ldots, x_n)$. Geometrically speaking, this is the set of all vectors in $\mathbb{R}^n$ that are perpendicular to the vector α.

An **(affine) hyperplane** is a linear hyperplane that has been translated by a vector $\beta = (b_1, \ldots, b_n)$. Thus, it is the solution set to an equation of the form

$$a_1(x_1 - b_1) + \cdots + a_n(x_n - b_n) = 0$$

or

$$a_1 x_1 + \cdots + a_n x_n = a_1 b_1 + \cdots a_n b_n$$

or finally

$$\alpha \cdot x = \alpha \cdot \beta$$

Let us write $\mathcal{H}(\alpha, b)$, where α is a vector in $\mathbb{R}^n$ and b is a real number to denote the hyperplane

$$\mathcal{H}(\alpha, b) = \{x \in \mathbb{R}^n \mid \alpha \cdot x = b\}$$

Note that the hyperplane

$$\mathcal{H}(\alpha, \|\alpha\|^2) = \{x \in \mathbb{R}^n \mid \alpha \cdot x = \|\alpha\|^2\}$$

contains the point α, which is the point of $\mathcal{H}(\alpha, b)$ closest to the origin, since

$$\alpha \cdot x = \|\alpha\|^2 \Rightarrow \|x\| \cos\theta = \|\alpha\| \Rightarrow \|x\| \geq \|\alpha\|$$

A hyperplane divides $\mathbb{R}^n$ into **closed halfspaces**

$$\mathcal{H}_+(\alpha, b) = \{x \in \mathbb{R}^n \mid \alpha \cdot x \geq b\}$$
$$\mathcal{H}_-(\alpha, b) = \{x \in \mathbb{R}^n \mid \alpha \cdot x \leq b\}$$

and two **open halfspaces**

$$\mathcal{H}^\circ_+(\alpha, b) = \{x \in \mathbb{R}^n \mid \alpha \cdot x > b\}$$
$$\mathcal{H}^\circ_-(\alpha, b) = \{x \in \mathbb{R}^n \mid \alpha \cdot x < b\}$$

It is not hard to show that

$$\mathcal{H}_+(\alpha, b) \cap \mathcal{H}_-(\alpha, b) = \mathcal{H}(\alpha, b)$$

and that $\mathcal{H}_+^\circ(\alpha, b)$, $\mathcal{H}_-^\circ(\alpha, b)$ and $\mathcal{H}(\alpha, b)$ are pairwise disjoint and

$$\mathcal{H}_+^\circ(\alpha, b) \cup \mathcal{H}_-^\circ(\alpha, b) \cup \mathcal{H}(\alpha, b) = \mathbb{R}^n$$

Definition *The subsets X and Y of $\mathbb{R}^n$ are* **completely separated** *by a hyperplane $\mathcal{H}(\alpha, b)$ if X lies in one open halfspace determined by $\mathcal{H}(\alpha, b)$ and Y lies in the other. Thus, one of the following holds*

1) $\alpha \cdot x < b < \alpha \cdot y$ *for all* $x \in X$, $y \in Y$
2) $\alpha \cdot y < b < \alpha \cdot x$ *for all* $x \in X$, $y \in Y$ □

Separation

Now that we have the preliminaries out of the way, we can get down to some theorems. The first is a well-known separation theorem that is the basis for many other separation theorems.

Theorem 3 *Let C be a closed convex subset of $\mathbb{R}^n$ that does* not *contain the origin, that is, $0 \notin C$. Then there is a nonzero $\alpha \in \mathbb{R}^n$ for which*

$$\alpha \cdot x \geq \|\alpha\|^2$$

for all $x \in C$. Hence, the hyperplane $\mathcal{H}(\alpha, \frac{1}{2}\|\alpha\|^2)$ completely separates 0 and C.

Proof. First we want to show that C contains a point that is closest to the origin from among all points in C. The function

$$d(x) = \|x\|$$

which measures the distance from x to the origin is a continuous function. Although C need not be compact, if we choose a real number s such that the closed ball $B_s(0) = \{z \in \mathbb{R}^n \mid \|z\| \leq s\}$ of radius s about the origin intersects C, then the intersection

$$C' = C \cap B_s(0)$$

is both closed and bounded and so is compact. The distance function therefore achieves its minimum on this set C', say at the point $\alpha \in C' \subseteq C$. We want to show that

$$\alpha \cdot x \geq \|\alpha\|^2$$

for all $x \in C$.

Suppose to the contrary that for some $x \in C$ we had

$$\alpha \cdot x < \|\alpha\|^2$$

Then since C is convex, the line segment from α to x must be contained in C

$$\{(1-t)\alpha + tx \mid 0 \leq t \leq 1\} \subseteq C$$

Let us look at the distance from a typical point on this line segment to the origin. If we discover that one point on this line is strictly closer than α we will have a contradiction because α is closest to the origin from among all points in C. This contradiction will show that $\alpha \cdot x \geq \|\alpha\|^2$ for all $x \in C$, as desired.

So we compute

$$\begin{aligned}\|(1-t)\alpha + tx\|^2 &= ((1-t)\alpha + tx) \cdot ((1-t)\alpha + tx) \\ &= (1-t)^2\|\alpha\|^2 + 2t(1-t)\alpha \cdot x + t^2\|x\|^2 \\ &= (\|\alpha\|^2 + \|x\|^2 - \alpha \cdot x)t^2 + (\alpha \cdot x - 2\|\alpha\|^2)t + \|\alpha\|^2\end{aligned}$$

Now, this is a quadratic in t that is concave up and has its minimum value at

$$t = \frac{2\|\alpha\|^2 - \alpha \cdot x}{2(\|\alpha\|^2 + \|x\|^2 - \alpha \cdot x)}$$

Since we are assuming that $\alpha \cdot x < \|\alpha\|^2$ we see that $0 < t$ and so the minimum value of $\|(1-t)\alpha + tx\|$ is strictly less than $\|\alpha\|$, which is what we wanted to show. □

The next result brings us closer to our goal.

Theorem 4 *Let C be a compact convex subset of $\mathbb{R}^n$ and let S be a subspace of $\mathbb{R}^n$ such that $C \cap S = \emptyset$. Then there exists a nonzero $\alpha \in \mathbb{R}^n$ such that*
1) $\alpha \cdot \sigma = 0$ for all $\sigma \in S$ (that is, $\alpha \in S^\perp$)
2) $\alpha \cdot \gamma \geq \|\alpha\|^2$ for all $\gamma \in C$
Hence, the hyperplane $\mathcal{H}(\alpha, \frac{1}{2}\|\alpha\|^2)$ completely separates S and C.
Proof. We consider the set

$$A = S + C$$

which is closed since S is closed and C is compact. It is also convex

since S and C are convex. Furthermore, $0 \neq A$ because if $0 = \sigma + \gamma$ then $\gamma = -\sigma$ would be in the intersection $C \cap S$, which is empty.

So we can apply Theorem 3 to deduce the existence of a nonzero $\alpha \in \mathbb{R}^n$ such that

$$\alpha \cdot x \geq \|\alpha\|^2$$

for all $x \in A = S + C$. Let $x = \sigma + \gamma$ be an arbitrary element of $S + C$. Then

$$\alpha \cdot \sigma + \alpha \cdot \gamma = \alpha \cdot (\sigma + \gamma) \geq \|\alpha\|^2$$

Now, if $\alpha \cdot \sigma$ is nonzero for any value of $\sigma \in S$, we can replace σ by a scalar multiple of σ to make the left side negative, which is impossible. Hence, we must have $\alpha \cdot \sigma = 0$ for all $\sigma \in S$, which is 1) above. Since $\alpha \cdot \sigma = 0$ we also get

$$\alpha \cdot \gamma \geq \|\alpha\|^2$$

for all $\gamma \in C$, as desired. □

Now we come to our main goal.

Theorem 5 *Let S be a subspace of $\mathbb{R}^n$ for which $S \cap \mathbb{R}^n_+ = \{0\}$, where*

$$\mathbb{R}^n_+ = \{(x_1, \ldots, x_n) \mid x_i \geq 0\}$$

is the nonnegative orthant in $\mathbb{R}^n$. Then $S^\perp$ contains a strongly positive vector $\alpha \gg 0$.

Proof. We would like to separate S from something, but we cannot separate it from $\mathbb{R}^n_+$. Consider instead the convex hull Δ of the standard basis vectors $\epsilon_1, \ldots, \epsilon_n$ in $\mathbb{R}^n_+$

$$\Delta = \{t_1\epsilon_1 + \cdots + t_n\epsilon_n \mid 0 \leq t_i \leq 1, \Sigma t_i = 1\}$$

It is clear that Δ contains only strongly positive vectors, that is, $\Delta \subseteq \mathbb{R}^n_{++}$ and is convex. It is also closed and bounded and therefore compact. Since $0 \notin \Delta$ there is a nonzero vector $\alpha = (a_1, \ldots, a_n)$ such that

1) $\alpha \in S^\perp$

2) $\alpha \cdot \delta \geq \|\alpha\|^2$ for all $\delta \in \Delta$

Taking $\delta = \epsilon_i$ to be the i-th standard basis vector, we get

$$a_i = \alpha \cdot \epsilon_i \geq \|\alpha\|^2 > 0$$

and so α is strongly positive, as desired. $\square$

Selected Solutions

Chapter 1: Probability I: Introduction to Discrete Probability

1. $1/2$
2. $25/72$
3. $5/12, 9/12, 7/12$
4. $1/1024, 11/1024, 56/1024$
5. $1/3, 1/6, 1/3, 1/2$
6. $3/8$
7. $11/16$
8. $1/13$
10. a) Consider a stock whose current price is 50 and whose price at some fixed time T in the future may be one of the following values: $48, 49, 50, 51$. Suppose we estimate that the probabilities of these stock prices are

$$\begin{aligned}\mathbb{P}(48) &= 0.2\\ \mathbb{P}(49) &= 0.4\\ \mathbb{P}(50) &= 0.3\\ \mathbb{P}(51) &= 0.1\end{aligned}$$

If we purchase one share of the stock now, what is the expected return at time T? What is the expected profit?

b) Consider a derivative of the stock in part a) whose return D is a function of the stock price, say

$$\begin{aligned}D(48) &= 2\\ D(49) &= -1\\ D(50) &= 0\\ D(51) &= 3\end{aligned}$$

Thus, the return D is a random variable on Ω. What is the expected return of the derivative?

11. $1/2$
12. The expected value is 0.
13. $25/13$ cents. Yes.
14. Not fair, $\mathcal{E} \approx -5$ cents.
21. Suppose that the values of X are $\{x_1, \ldots, x_n\}$ and the values of Y are $\{y_1, \ldots, y_m\}$. Then the values of XY are the distinct products $x_i y_j$. For a given value a of XY let

$$\{(x_{i_1}, y_{j_1}), \ldots, (x_{i_m}, y_{j_m})\}$$

be the set of all pairs whose product is equal to a. Then

$$\begin{aligned}
\mathbb{P}(XY = a, Z = z) &= \mathbb{P}\left\{\left[\bigcup_{k=1}^{m}(\{X = x_{i_k}\} \cap \{Y = y_{j_k}\})\right] \cap \{Z = z\}\right\} \\
&= \mathbb{P}\left\{\bigcup_{k=1}^{m}(\{X = x_{i_k}\} \cap \{Y = y_{j_k}\} \cap \{Z = z\})\right\} \\
&= \sum_{k=1}^{m}\mathbb{P}(X = x_{i_k})\mathbb{P}(Y = y_{j_k})\mathbb{P}(Z = z) \\
&= \left[\sum_{k=1}^{m}\mathbb{P}(X = x_{i_k})\mathbb{P}(Y = y_{j_k})\right]\mathbb{P}(Z = z) \\
&= \left[\sum_{k=1}^{m}\mathbb{P}(X = x_{i_k}, Y = y_{j_k})\right]\mathbb{P}(Z = z) \\
&= \mathbb{P}(XY = a)\mathbb{P}(Z = z)
\end{aligned}$$

22. We have

$$\begin{aligned}
\mathbb{P}(f(X) = a, g(Y) = b) &= \mathbb{P}(X \in f^{-1}\{a\}, Y \in g^{-1}\{b\}) \\
&= \mathbb{P}(X \in f^{-1}\{a\})\mathbb{P}(Y \in g^{-1}\{b\}) \\
&= \mathbb{P}(f(X) = a)\mathbb{P}(g(Y) = b)
\end{aligned}$$

Chapter 2: Portfolio Management and the Capital Asset Pricing Model

1. $\beta = 1$
2. $\mu = 0.04(\beta + 1)$
3. The equation for σ' can be solve for s to get a linear function of s. Plug this into the equation for μ to get μ as a linear function of σ'.
4. $b/(1 - a)$
6. We have

$$\begin{aligned}
\text{Cov}(R_M, \epsilon) &= \text{Cov}(R_M, R_i - \beta_k R_M) \\
&= \text{Cov}(R_M, R_i) - \beta_k \text{Cov}(R_M, R_M) \\
&= \text{Cov}(R_M, R_i) - \frac{\text{Cov}(R_i, R_M)}{\text{Cov}(R_M, R_M)}\text{Cov}(R_M, R_M) \\
&= 0
\end{aligned}$$

7. Solve for β from the equation of the capital market line and plug it into the equation $y = \beta x + \alpha$ of the regression line to get the equation

$$y = \frac{\mu_k - \mu_{\text{rf}}}{\mu_M - \mu_{\text{rf}}}(x - \mu_M) + \mu_k$$

Setting $x = \mu_{\text{rf}}$ gives $y = \mu_{\text{rf}}$.

8. $\beta_0 = w_1\beta_1 + w_1\beta_2$ and $\alpha_0 = w_1\alpha_1 + w_1\alpha_2$

Chapter 3: Background on Options

4. The cost C_1 of the call with the smaller strike price is more than the cost C_2 of the call with the higher strike price. The profit curve is shown below.

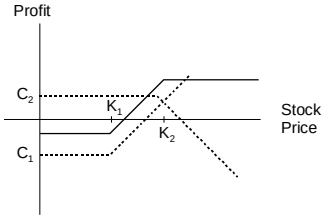


Figure 3 – A bull spread

6. The profit is made by selling the APR call and buying the JUL call. This costs $3. However, in April, the investor owns a call worth $5, thus making $2.
7. The profit curve is shown in Figure 3.

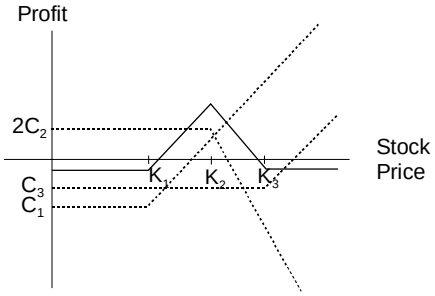


Figure 3 – The butterfly spread

Chapter 4: An Aperitif on Arbitrage

1. We still have

$$\mathcal{V}(\text{long contract}) = S_T - F_{0,T}$$
$$\mathcal{V}(\text{short contract}) = F_{0,T} - S_T$$

but the cash-and-carry investor has a final payoff of

$$\mathcal{V}(\text{cash-and-carry}) = S_T - S_0 e^{rT} + Ie^{rT}$$

and

$$\mathcal{V}(\text{reverse cash-and-carry}) = S_0 e^{rT} - S_T - Ie^{rT}$$

To explain the last term, note that the short sale of an asset requires a lender to lend that asset. This lender will demand the return of not only the asset itself, but also the income that would have come to the lender by virtue of owning the asset.

2. The two strategies now have payoffs as follows:

$$\mathcal{V}(\text{long contract}) + \mathcal{V}(\text{reverse cash-and-carry}) = S_0 e^{rT} - F_{0,T} - Ie^{rT}$$

and

$$\mathcal{V}(\text{short contract}) + \mathcal{V}(\text{cash-and-carry}) = F_{0,T} - S_0 e^{rT} + Ie^{rT}$$

3. Setting the final payoffs in exercise 2 to 0 gives

$$F_{0,T} = (S_0 - I)e^{rT}$$

4. Since no lending or borrowing is done for the buyer of the contract, we still have

$$\mathcal{V}(\text{long contract}) = S_T - F_{0,T}$$
$$\mathcal{V}(\text{short contract}) = F_{0,T} - S_T$$

The other payoffs need a slight tweaking as follows

$$\mathcal{V}(\text{cash-and-carry}) = S_T - S_0 e^{r_b T}$$
$$\mathcal{V}(\text{reverse cash-and-carry}) = S_0 e^{r_\ell T} - S_T$$

5. The final payoff for Strategy 1 is

$$\mathcal{V}(\text{long contract}) + \mathcal{V}(\text{reverse cash-and-carry}) = S_0 e^{r_\ell T} - F_{0,T}$$

For Strategy 2 we have

$$\mathcal{V}(\text{short contract}) + \mathcal{V}(\text{cash-and-carry}) = F_{0,T} - S_0 e^{r_b T}$$

6. The assumption of no arbitrage implies that both of the payoffs must be nonpositive, that is

$$S_0 e^{r_\ell T} - F_{0,T} \leq 0$$

and

$$F_{0,T} - S_0 e^{r_b T} \leq 0$$

Combining these two inequalities gives

$$S_0 e^{r_\ell T} \leq F_{0,T} \leq S_0 e^{r_b T}$$

7. If not then buy the share, sell the call and pocket the difference. Use the share to cover the call if and when it is exercised.
8. If not, sell the put and invest the money. The put cannot cost you more than K in either case and you will have K immediately in the case of the American put or at the end in the case of the European put.
9. Use put-call option parity formula.
10. a) The first inequality comes from the fact that $C^E = C^A$. The second comes from the fact that $K - S_0$ can be gotten by exercising the put at time t_0.
    b) For the first formula, since $C^A = C^E$ we get

$$S_0 - Ke^{-rT} - d_0 \leq C^A$$

In addition, immediate exercise of the American call will return $S_0 - K$ and so the call cannot be purchased for less than this amount. For the second formula, $K - S_0 \leq P^A$ because immediate exercise is worth $K - S_0$. For the other part, consider two portfolios. Portfolio A is 1 put. Portfolio B is $(K + d_0)e^{-rT}$ in cash and a short share of stock.

Chapter 5: Probability II: More Discrete Probability

3. We have

$$\begin{aligned}\mathcal{E}(X \mid \mathcal{P})(\omega) &= \mathcal{E}(X \mid [\omega]_{\mathcal{P}}) \\ &= \sum_{\sigma \in \Omega} X(\sigma)\mathbb{P}(\sigma \mid [\omega]_{\mathcal{P}}) \\ &\geq 0\end{aligned}$$

17. For $\omega \in \Omega$ let $k = N(\omega)$. Then

$$\begin{aligned}
\mathcal{E}(S \mid N)(\omega) &= \mathcal{E}(S \mid N = k) \\
&= \sum_{i=1}^{m} r_i \mathbb{P}(S = r_i \mid N = k) \\
&= \sum_{i=1}^{m} r_i \frac{\mathbb{P}((X_1 + \cdots + X_N = r_i) \cap (N = k))}{\mathbb{P}(N = k)} \\
&= \sum_{i=1}^{m} r_i \frac{\mathbb{P}((X_1 + \cdots + X_k = r_i) \cap (N = k))}{\mathbb{P}(N = k)} \\
&= \sum_{i=1}^{m} r_i \frac{\mathbb{P}(X_1 + \cdots + X_k = r_i)\mathbb{P}(N = k)}{\mathbb{P}(N = k)} \\
&= \sum_{i=1}^{m} r_i \mathbb{P}(X_1 + \cdots + X_k = r_i) \\
&= \mathcal{E}(X_1 + \cdots + X_k) \\
&= \mu k \\
&= \mu N(\omega)
\end{aligned}$$

18. The solution is

$$\begin{aligned}
\mathcal{E}(\mathcal{E}(X \mid \mathcal{Q}) \mid \mathcal{P}) &= \mathcal{E}\left(\sum_{i=1}^{m} \mathcal{E}(X \mid C_i) 1_{C_i} \Big| \mathcal{P} \right) \\
&= \sum_{i=1}^{m} \mathcal{E}(\mathcal{E}(X \mid C_i) 1_{C_i} \mid \mathcal{P}) \\
&= \sum_{i=1}^{m} \mathcal{E}(X \mid C_i) \mathcal{E}(1_{C_i} \mid \mathcal{P}) \\
&= \sum_{i=1}^{m} \mathcal{E}(X \mid C_i) \left[\sum_{j=1}^{k} \mathcal{E}(1_{C_i} \mid B_j) 1_{B_j} \right] \\
&= \sum_{i=1}^{m} \mathcal{E}(X \mid C_i) \left[\sum_{j=1}^{k} \frac{\mathcal{E}(1_{C_i} 1_{B_j})}{\mathbb{P}(B_j)} 1_{B_j} \right] \\
&= \sum_{i=1}^{m} \mathcal{E}(X \mid C_i) \left[\sum_{j=1}^{k} \frac{\mathcal{E}(1_{C_i \cup B_j})}{\mathbb{P}(B_j)} 1_{B_j} \right]
\end{aligned}$$

Now, since for each C_i there is a *unique* B_{j_i} for which $C_i \subseteq B_{j_i}$ we know that

$$1_{C_i \cup B_j} = \begin{cases} 1_{C_i} & j = j_i \\ 0 & j \neq j_i \end{cases}$$

and so we pick up with

$$\begin{aligned} \sum_{i=1}^{m} \mathcal{E}(X \mid C_i) \left[\sum_{j=1}^{k} \frac{\mathcal{E}(1_{C_i \cup B_j})}{\mathbb{P}(B_j)} 1_{B_j} \right] &= \sum_{i=1}^{m} \mathcal{E}(X \mid C_i) \frac{\mathcal{E}(1_{C_i})}{\mathbb{P}(B_{j_i})} 1_{B_{j_i}} \\ &= \sum_{i=1}^{m} \frac{\mathcal{E}(X 1_{C_i})}{\mathbb{P}(C_i)} \frac{\mathcal{E}(1_{C_i})}{\mathbb{P}(B_{j_i})} 1_{B_{j_i}} \\ &= \sum_{i=1}^{m} \frac{\mathcal{E}(X 1_{C_i})}{\mathbb{P}(B_{j_i})} 1_{B_{j_i}} \end{aligned}$$

Now we group the terms of this sum into smaller sums, each one being over just the blocks C_i that are contained in one block B_{j_i}. (In other words, group the sum by the blocks in $\mathcal{P}$.) This gives

$$\sum_{i=1}^{m} \frac{\mathcal{E}(X 1_{C_i})}{\mathbb{P}(B_{j_i})} 1_{B_{j_i}} = \sum_{j=1}^{k} \frac{\mathcal{E}(X 1_{B_j})}{\mathbb{P}(B_j)} 1_{B_j} = \mathcal{E}(X \mid \mathcal{P})$$

as desired.

Chapter 6: Discrete-Time Pricing Models

1. The system of equations is

$$\begin{aligned} \mathcal{V}_2(\Theta_2)(\omega_1) &= 95 \\ \mathcal{V}_2(\Theta_2)(\omega_2) &= 90 \\ \mathcal{V}_2(\Theta_2)(\omega_3) &= 85 \\ \mathcal{V}_2(\Theta_2)(\omega_4) &= 75 \end{aligned}$$

or

$$\begin{aligned} S_{2,1}(\omega_1)\theta_{2,1}(\omega_1) + S_{2,2}(\omega_1)\theta_{2,2}(\omega_1) &= 95 \\ S_{2,1}(\omega_2)\theta_{2,1}(\omega_2) + S_{2,2}(\omega_2)\theta_{2,2}(\omega_2) &= 90 \\ S_{2,1}(\omega_3)\theta_{2,1}(\omega_3) + S_{2,2}(\omega_3)\theta_{2,2}(\omega_3) &= 85 \\ S_{2,1}(\omega_4)\theta_{2,1}(\omega_4) + S_{2,2}(\omega_4)\theta_{2,2}(\omega_4) &= 75 \end{aligned}$$

Substituting the actual prices gives

$$\begin{aligned}
\theta_{2,1}(\omega_1) + 90\theta_{2,2}(\omega_1) &= 95 \\
\theta_{2,1}(\omega_2) + 80\theta_{2,2}(\omega_2) &= 90 \\
\theta_{2,1}(\omega_3) + 80\theta_{2,2}(\omega_3) &= 85 \\
\theta_{2,1}(\omega_4) + 75\theta_{2,2}(\omega_4) &= 75
\end{aligned}$$

The condition that Θ_2 be $\mathcal{P}_1$-measurable is

$$\begin{aligned}
\theta_{2,1}(\omega_1) &= \theta_{2,1}(\omega_2) \\
\theta_{2,1}(\omega_3) &= \theta_{2,1}(\omega_4) \\
\theta_{2,2}(\omega_1) &= \theta_{2,2}(\omega_2) \\
\theta_{2,2}(\omega_3) &= \theta_{2,2}(\omega_4)
\end{aligned}$$

and so the previous system can be written using only ω_1 and ω_3 as

$$\begin{aligned}
\theta_{2,1}(\omega_1) + 90\theta_{2,2}(\omega_1) &= 95 \\
\theta_{2,1}(\omega_1) + 80\theta_{2,2}(\omega_1) &= 90 \\
\theta_{2,1}(\omega_3) + 80\theta_{2,2}(\omega_3) &= 85 \\
\theta_{2,1}(\omega_3) + 75\theta_{2,2}(\omega_3) &= 75
\end{aligned}$$

The first two equations have a unique solution and so do the second two equations, giving

$$\begin{aligned}
\Theta_2(\omega_1) = \Theta_2(\omega_2) &= (50, \frac{1}{2}) \\
\Theta_2(\omega_3) = \Theta_2(\omega_4) &= (-75, 2)
\end{aligned}$$

Working backwards in time, we next compute the acquisition values for Θ_2

$$\begin{aligned}
\mathcal{V}_1(\Theta_2)(\omega_1) &= 50 + 85 \cdot \frac{1}{2} = \frac{185}{2} \\
\mathcal{V}_1(\Theta_2)(\omega_3) &= -75 + 78 \cdot 2 = 81
\end{aligned}$$

The self-financing condition requires that these are also the liquidation values of Θ_1 and so

$$\begin{aligned}
\mathcal{V}_1(\Theta_1)(\omega_1) &= \frac{185}{2} \\
\mathcal{V}_1(\Theta_1)(\omega_3) &= 81
\end{aligned}$$

Writing these out and substituting the actual prices gives the system

$$\theta_{1,1}(\omega_1) + 85\theta_{1,2}(\omega_1) = \frac{185}{2}$$
$$\theta_{1,1}(\omega_3) + 78\theta_{1,2}(\omega_3) = 81$$

But Θ_1 is $\mathcal{P}_0$-measurable, that is, constant on Ω, and so for any $\omega \in \Omega$

$$\theta_{1,1}(\omega) + 85\theta_{1,2}(\omega) = \frac{185}{2}$$
$$\theta_{1,1}(\omega) + 78\theta_{1,2}(\omega) = 81$$

This system has solution

$$\Theta_1(\omega) = \left(-\frac{330}{7}, \frac{23}{14}\right)$$

which is a portfolio consisting of a short position (sale) of $\frac{330}{7}$ bonds and a purchase of $\frac{23}{14}$ shares of stock, for an initial cost of

$$-\frac{330}{7} + 80 \cdot \frac{23}{14} = \frac{590}{7} \approx \$84.29$$

2. The system of equations is

$$\mathcal{V}_3(\Theta_3)(\omega_1) = 100$$
$$\mathcal{V}_3(\Theta_3)(\omega_2) = 100$$
$$\mathcal{V}_3(\Theta_3)(\omega_3) = 95$$
$$\mathcal{V}_3(\Theta_3)(\omega_4) = 90$$
$$\mathcal{V}_3(\Theta_3)(\omega_5) = 90$$
$$\mathcal{V}_3(\Theta_3)(\omega_6) = 85$$

or, since the price of the risk-free asset $\mathfrak{a}_1$ is 1

$$\theta_{3,1}(\omega_1) + S_{3,2}(\omega_1)\theta_{3,2}(\omega_1) = 100$$
$$\theta_{3,1}(\omega_2) + S_{3,2}(\omega_2)\theta_{3,2}(\omega_2) = 100$$
$$\theta_{3,1}(\omega_3) + S_{3,2}(\omega_3)\theta_{3,2}(\omega_3) = 95$$
$$\theta_{3,1}(\omega_4) + S_{3,2}(\omega_4)\theta_{3,2}(\omega_4) = 90$$
$$\theta_{3,1}(\omega_5) + S_{3,2}(\omega_5)\theta_{3,2}(\omega_5) = 90$$
$$\theta_{3,1}(\omega_6) + S_{3,2}(\omega_6)\theta_{3,2}(\omega_6) = 85$$

Substituting the actual prices gives

$$\begin{aligned}
\theta_{3,1}(\omega_1) + 100\theta_{3,2}(\omega_1) &= 100 \\
\theta_{3,1}(\omega_2) + 95\theta_{3,2}(\omega_2) &= 100 \\
\theta_{3,1}(\omega_3) + 95\theta_{3,2}(\omega_3) &= 95 \\
\theta_{3,1}(\omega_4) + 90\theta_{3,2}(\omega_4) &= 90 \\
\theta_{3,1}(\omega_5) + 90\theta_{3,2}(\omega_5) &= 90 \\
\theta_{3,1}(\omega_6) + 80\theta_{3,2}(\omega_6) &= 85
\end{aligned}$$

The condition that Θ_3 be $\mathcal{P}_2$-measurable is

$$\begin{aligned}
\theta_{3,1}(\omega_1) &= \theta_{3,1}(\omega_2) \\
\theta_{3,1}(\omega_4) &= \theta_{3,1}(\omega_5) \\
\theta_{3,2}(\omega_1) &= \theta_{3,2}(\omega_2) \\
\theta_{3,2}(\omega_4) &= \theta_{3,2}(\omega_5)
\end{aligned}$$

and so the previous system can be written using only $\omega_1, \omega_3, \omega_4$ and ω_6 as

$$\begin{aligned}
\theta_{3,1}(\omega_1) + 100\theta_{3,2}(\omega_1) &= 100 \\
\theta_{3,1}(\omega_1) + 95\theta_{3,2}(\omega_1) &= 100 \\
\theta_{3,1}(\omega_3) + 95\theta_{3,2}(\omega_3) &= 95 \\
\theta_{3,1}(\omega_4) + 90\theta_{3,2}(\omega_4) &= 90 \\
\theta_{3,1}(\omega_4) + 90\theta_{3,2}(\omega_4) &= 90 \\
\theta_{3,1}(\omega_6) + 80\theta_{3,2}(\omega_6) &= 85
\end{aligned}$$

The first two equations have a unique solution and so do the fourth and fifth equations, giving

$$\begin{aligned}
\Theta_3(\omega_1) = \Theta_3(\omega_2) &= (100, 0) \\
\Theta_3(\omega_4) = \Theta_3(\omega_5) &= (0, 1)
\end{aligned}$$

along with

$$\begin{aligned}
\Theta_3(\omega_3) &= (s, \frac{95 - s}{95}) \\
\Theta_3(\omega_6) &= (t, \frac{85 - t}{80})
\end{aligned}$$

where s and t are parameters. The acquisition values for Θ_3 are

$$\begin{aligned}\mathcal{V}_2(\Theta_3)(\omega_1) = \mathcal{V}_2(\Theta_3)(\omega_2) &= 100\\ \mathcal{V}_2(\Theta_3)(\omega_3) &= s + 80 \cdot \frac{95-s}{95} = \frac{3s}{19} + 80\\ \mathcal{V}_2(\Theta_3)(\omega_4) = \mathcal{V}_2(\Theta_3)(\omega_5) &= 80\\ \mathcal{V}_2(\Theta_3)(\omega_6) &= t + 75 \cdot \frac{95-t}{95} = \frac{4t}{19} + 75\end{aligned}$$

The self-financing condition requires that these are also the liquidation values of Θ_2 and so Θ_2 must replicate the alternative

$$\left(100, \frac{3s}{19} + 80, 80, \frac{4t}{19} + 75\right)$$

Since we are asked for only one replicating portfolio, let us choose $s = t = 0$ to get the alternative

$$(100, 80, 80, 75)$$

We have the system

$$\begin{aligned}\theta_{2,1}(\omega_1) + 90\theta_{2,2}(\omega_1) &= 100\\ \theta_{2,1}(\omega_2) + 90\theta_{2,2}(\omega_2) &= 100\\ \theta_{2,1}(\omega_3) + 80\theta_{2,2}(\omega_3) &= 80\\ \theta_{2,1}(\omega_4) + 80\theta_{2,2}(\omega_4) &= 80\\ \theta_{2,1}(\omega_5) + 80\theta_{2,2}(\omega_5) &= 80\\ \theta_{2,1}(\omega_6) + 75\theta_{2,2}(\omega_6) &= 75\end{aligned}$$

Since $\theta_{2,i}$ is constant on the blocks $\{\omega_1, \omega_2, \omega_3\}$ and $\{\omega_6, \omega_6, \omega_6\}$ this can be written

$$\begin{aligned}\theta_{2,1}(\omega_1) + 90\theta_{2,2}(\omega_1) &= 100\\ \theta_{2,1}(\omega_1) + 90\theta_{2,2}(\omega_1) &= 100\\ \theta_{2,1}(\omega_1) + 80\theta_{2,2}(\omega_1) &= 80\\ \theta_{2,1}(\omega_4) + 80\theta_{2,2}(\omega_4) &= 80\\ \theta_{2,1}(\omega_4) + 80\theta_{2,2}(\omega_4) &= 80\\ \theta_{2,1}(\omega_4) + 75\theta_{2,2}(\omega_4) &= 75\end{aligned}$$

or

$$\theta_{2,1}(\omega_1) + 90\theta_{2,2}(\omega_1) = 100$$
$$\theta_{2,1}(\omega_1) + 80\theta_{2,2}(\omega_1) = 80$$
$$\theta_{2,1}(\omega_4) + 80\theta_{2,2}(\omega_4) = 80$$
$$\theta_{2,1}(\omega_4) + 75\theta_{2,2}(\omega_4) = 75$$

giving

$$\Theta_2(\omega_1) = \Theta_2(\omega_2) = \Theta_2(\omega_3) = (-80, 2)$$
$$\Theta_2(\omega_4) = \Theta_2(\omega_5) = \Theta_2(\omega_6) = (0, 1)$$

Working backwards in time, we next compute the acquisition values for Θ_2

$$\mathcal{V}_1(\Theta_2)(\omega_1) = -80 + 85 \cdot 2 = 90$$
$$\mathcal{V}_1(\Theta_2)(\omega_4) = 0 + 78 = 78$$

The self-financing condition requires that these are also the liquidation values of Θ_1 and so

$$\mathcal{V}_1(\Theta_1)(\omega_1) = 90$$
$$\mathcal{V}_1(\Theta_1)(\omega_4) = 78$$

Writing these out and substituting the actual prices gives the system

$$\theta_{1,1}(\omega_1) + 85\theta_{1,2}(\omega_1) = 90$$
$$\theta_{1,1}(\omega_4) + 78\theta_{1,2}(\omega_4) = 78$$

But Θ_1 is $\mathcal{P}_0$-measurable, that is, constant on Ω, and so for any $\omega \in \Omega$

$$\theta_{1,1}(\omega_1) + 85\theta_{1,2}(\omega_1) = 90$$
$$\theta_{1,1}(\omega_4) + 78\theta_{1,2}(\omega_4) = 78$$

This system has solution

$$\Theta_1(\omega) = \left(-\frac{390}{7}, \frac{12}{7}\right)$$

which is a portfolio consisting of a short position (sale) of $\frac{390}{7}$ bonds and a purchase of $\frac{12}{7}$ shares of stock, for an initial cost of

$$-\frac{390}{7} + 80 \cdot \frac{12}{7} = \frac{570}{7} \approx \$81.43$$

3. Toss 1 is tails: Casino is even, player down $1 million, game over.
   Toss 1 is heads: Casino up $2 million, player down $1 million, game continues.
   Toss 2 is tails: Casino is even, player down $1 million, game over.
   Toss 2 is heads: Casino up $4 million, player down $1 million, game continues.
   Toss 3 is tails: Casino is even, player down $1 million, game over.
   Toss 3 is heads: Casino is even, player up $8 million, game continues.

   In all ending cases, the casino is even. Thus, the casino has a perfect hedge. It is self-financing because the casino never added its own money or removed money. The side bets on heads replicated the payoff to the casino, but in the opposite position, resulting in a 0 payoff to the casino.
4. $$\begin{aligned}\mathcal{V}_i(a\Theta_{1,i} + b\Theta_{2,i}) &= a\mathcal{V}_i(\Theta_{1,i}) + b\mathcal{V}_i(\Theta_{2,i}) \\ &= a\mathcal{V}_i(\Theta_{1,i+1}) + b\mathcal{V}_i(\Theta_{2,i+1}) \\ &= \mathcal{V}_i(a\Theta_{1,i+1} + b\Theta_{2,i+1})\end{aligned}$$

   and so it follows that $a\Phi_1 + b\Phi_2$ is self-financing.
5. The self-financing condition is

$$\mathcal{V}_i(\Theta'_i) = \mathcal{V}_i(\Theta'_{i+1})$$

   for all $i = 1, \ldots, T-1$. Because Φ is assumed to be self-financing, the liquidation value of Θ'_i is

$$\begin{aligned}\mathcal{V}_i(\Theta'_i) &= (\theta_{i,1} + a)S_{i,1} + \sum_{j=2}^{n}\theta_{i,j}S_{i,j} \\ &= \mathcal{V}_i(\Theta_i) + aS_{i,1}1_\Omega \\ &= \mathcal{V}_i(\Theta_{i+1}) + aS_{i,1}1_\Omega\end{aligned}$$

   and the acquisition value is

$$\begin{aligned}\mathcal{V}_i(\Theta'_{i+1}) &= (\theta_{i+1,1} + a)S_{i,1} + \sum_{j=2}^{n}\theta_{i+1,j}S_{i,j} \\ &= \mathcal{V}_i(\Theta_{i+1}) + aS_{i,1}1_\Omega\end{aligned}$$

   Thus Φ' is self-financing.
7. Let $\Phi_0 = 0$ be the zero trading strategy (where all portfolios are the zero portfolio). If the Law of One Price holds then for any trading

strategy Φ that has 0 final value we have

$$\begin{aligned}\mathcal{V}_T(\Phi) = 0 &\Rightarrow \mathcal{V}_T(\Phi) = \mathcal{V}_T(\Phi_0)\\ &\Rightarrow \mathcal{V}_0(\Phi) = \mathcal{V}_0(\Phi_0)\\ &\Rightarrow \mathcal{V}_0(\Phi) = 0\end{aligned}$$

and so 2) holds. Conversely, suppose that any trading strategy with payoff 0 has initial value 0. Then

$$\begin{aligned}\mathcal{V}_T(\Phi_1) = \mathcal{V}_T(\Phi_2) &\Rightarrow \mathcal{V}_T(\Phi_1 - \Phi_2) = 0\\ &\Rightarrow \mathcal{V}_0(\Phi_1 - \Phi_2) = 0\\ &\Rightarrow \mathcal{V}_0(\Phi_1) = \mathcal{V}_0(\Phi_2)\end{aligned}$$

and so the Law of One Price holds.

8. Yes. We simply sell one share of stock short and invest the money in the risk-free asset. If the risk-free rates are high enough, they we provide us with *more* money than necessary to buy the stock and return it at the end. More formally, let M be the maximum value of $S_{T,2}$ on Ω. Let $\Theta_1 = (S_{0,2}, -1)$. Roll this over at each time. The final portfolio is thus also $\Theta_T = (S_{0,2}, -1)$, with liquidation value $\mathcal{V}_T(\Theta_T) = S_{0,2}r - S_{T,2}$ where $r = e^{r_1(t_1-t_0)}\cdots e^{r_T(t_T-t_{T-1})}$. Thus,

$$\mathcal{V}_T(\Theta_T) = S_{0,2}r - S_{T,2} \geq S_{0,2}r - M$$

So if $r > M/S_{0,2}$ we are guaranteed a profit.

11. In this case the random variables $X_k = 1_{\{\omega_k\}}$ are attainable, say by Φ_k. Hence, any random variable $X = \Sigma X(\omega_k)X_k$ is attainable via the replicating strategy $X = \Sigma X(\omega_k)\Phi_k$.
12. Let $X\colon \Omega \to \mathbb{R}^2$. Let $\Theta_1 = (\theta_{1,1}, \theta_{1,2})$. Then Θ_1 replicates X if

$$\mathcal{V}_1(\Theta_1) = \theta_{1,1}e^{rT} + \theta_{1,2}S_T = X$$

that is,

$$\begin{aligned}\theta_{1,1}(\omega_1)e^{rT} + \theta_{1,2}(\omega_1)S_0u &= X(\omega_1)\\ \theta_{1,1}(\omega_2)e^{rT} + \theta_{1,2}(\omega_2)S_0d &= X(\omega_2)\end{aligned}$$

This system always has a solution if and only if the determinant is nonzero, that is,

$$\begin{vmatrix} e^{rT} & S_0u \\ e^{rT} & S_0d \end{vmatrix} = e^{rT}S_0(d-u) \neq 0$$

13. The solution is

$$\theta_{1,1} = e^{-rT}\frac{uf_d - df_u}{u-d}$$
$$\theta_{1,2} = \frac{f_u - f_d}{S_0(u-d)}$$

14. The price of X is

$$e^{rT}\left[\frac{e^{rT}-d}{u-d}f_u + \frac{u-e^{rT}}{u-d}f_d\right]$$

15. It is a candidate for martingale measure.
16. To avoid arbitrage, if the payoff satisfies $(f_u, f_d) > 0$ then the initial price must be positive. In other words,

$$\pi f_u + (1-\pi)f_d > 0$$

for all strictly positive vectors (f_u, f_d). This is easily seen to be the case if and only if $0 < \pi < 1$ which is equivalent to $d < e^{rT} < u$.

17. We have

$$u = \frac{101}{100} = 1.01,\ d = \frac{99}{100} = 0.99$$

and

$$f_x = \max(100x - 99.50, 0) = \begin{cases} 1.50 & x = u \\ 0 & x = d \end{cases}$$

Thus,

$$C = \frac{1-(0.99)e^{-rT}}{0.02}1.50 = 75(1-(0.99)e^{-rT}) = 0.75813654$$

18. a) If the risk-neutral probability distribution is $(\pi, 1-\pi)$ then we have

$$160 = 200\pi + 140(1-\pi)$$

and so $p = 1/3$. To price the asset, we have

$$P = (1/3)\cdot 0 + (2/3)\cdot 40 = 80/3$$

b) Suppose you invest in the portfolio $\Theta = (x, y, 1)$: x bonds, y stock and 1 put. Then the initial cost is

$$\mathcal{V}_0(\Theta) = x + 160y + 20$$

and the final payoff is

$$\mathcal{V}_T(\Theta) = \begin{pmatrix} x & y & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 200 & 140 \\ 0 & 40 \end{pmatrix} = \begin{pmatrix} x + 200y & x + 140y + 40 \end{pmatrix}$$

Now, to make a profit (since $r = 0$), we only need to make sure that

$$\begin{aligned} x + 200y &> x + 160y + 20 \\ x + 140y + 40 &> x + 160y + 20 \end{aligned}$$

This is equivalent to

$$1/2 < y < 1$$

So, a profit is guaranteed by buying any amount of stock between $1/2$ and 1.

19. A trading strategy, which amount to nothing more than a single portfolio, is only two-dimensional whereas the space containing the alternatives is three-dimensional. Hence, the valuation $\mathcal{V}_1$ cannot be surjective.
20. To obtain a risk-neutral probability distribution

$$\Pi = (p, q, 1 - p - q)$$

we must solve the equation

$$40p + 30q + 20(1 - p - q) = 25$$

or, equivalently

$$4p + 2q = 1$$

This has infinitely many solutions for which the vector

$$(p, q, 1 - p - q) = (p, \frac{1}{2} - 2p, \frac{1}{2} + p)$$

is *strongly positive.* (It is easy to forget when solving that we need strongly positive solutions!) The condition of strong positivity is

$$\begin{aligned} p &> 0 \\ \frac{1}{2} - 2p &> 0 \\ \frac{1}{2} + p &> 0 \end{aligned}$$

which is equivalent to

$$0 < p < \frac{1}{4}$$

22. The payoff vector is

$$\rho = (20, 10, 0)$$

and we must check that this is attainable, that is, that there is a solution to the system

$$(\,\theta_1 \quad \theta_2\,)\begin{pmatrix} 1 & 1 & 1 \\ 40 & 30 & 20 \end{pmatrix} = (20, 10, 0)$$

It is readily seen that the solution is

$$\theta_1 = -20,\ \theta_2 = 1$$

so the alternative is attainable and a replicating portfolio is given by *selling* 20 units of the risk-free asset and buying one share of the stock. To price the option under the two risk-neutral distributions, we have

$$\mathcal{I}_{\Pi_1}(\rho) = e^{-rT}\mathcal{E}_{\Pi_1}(\rho) = \frac{1}{12}\cdot 12 + \frac{4}{12}\cdot 10 + \frac{7}{12}\cdot 0 = \frac{13}{3}$$

and

$$\mathcal{I}_{\Pi_2}(\rho) = e^{-rT}\mathcal{E}_{\Pi_2}(\rho) = \frac{1}{6}\cdot 12 + \frac{1}{6}\cdot 10 + \frac{4}{6}\cdot 0 = \frac{11}{3}$$

Thus, the two risk-neutral probability distributions do not give the same price for the derivative. Nevertheless, we may choose either of these distributions and be assured of the absence of arbitrage!

Chapter 7: Chapter on CRR Model

1. a) 0, b) 0.0015, c) 0.2944, d) .0816, e) 2.0799, f) .0783. For the put, use the put-call option parity formula $P = Ke^{-rt} + C - S_0$. For example, when $K = 50$ we have

$$P = 50e^{0.01/6} + 0.2944 - 50 = 0.3778$$

2. a) 0.0008, b) 0.1501, c) 1.015, d) 2.0133. For the put, use the put-call option parity formula $P = Ke^{-rt} + C - S_0$.
3. A 10% gain followed by a 10% loss, or vice-versa, results is a slight loss, as shown by

$$(1+0.1)(1-0.1) = 1 - 0.01 = 0.99$$

(If the gain comes first, the loss is on a larger amount; if the loss comes first, the gain is on a smaller amount.

7. For part d), we can price a path-independent alternative X as follows

$$\begin{aligned}\mathcal{I}(X) &= \mathcal{V}_0(\Phi) \\ &= e^{-rL}\mathcal{E}_\Pi(\mathcal{V}_i(\Phi)) \\ &= e^{-rL}\mathcal{E}_\Pi(X) \\ &= e^{-rL}\sum_{k=0}^{T} X(\text{any } \omega \in G_k)\mathbb{P}_\Pi(G_k) \\ &= e^{-rL}\sum_{k=0}^{T} X_k\mathbb{P}_\Pi(G_k)\end{aligned}$$

10. This is because the time-t_k asset prices are required to be $\mathcal{P}_k$-measurable. The converse fails in general. For example, the asset prices at time t_k could be constant over the entire state space Ω. For the CRR model, assuming we know the history of stock prices and states prior to time t_k then at time t_k knowledge of the stock's price tells us the state as well. This follows from the fact that at time t_k the stock's price can only be one of two *distinct* values relative to the previous time-t_{k-1} price

$$S_k(\omega) = S_{k-1}u \text{ or } S_k(\omega) = S_{k-1}d$$

Thus, knowledge of the true price implies knowledge of the change in state from time t_{k-1} to time t_k. This, together with knowledge of the time-t_{k-1} state, tells us the time-t_k state.

Chapter 8: Continuous Probability

4. In particular, if $A_i \in \Sigma$ then

$$\left(\bigcap_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c \in \Sigma$$

and so

$$\bigcap_{i=1}^{\infty} A_i \in \Sigma$$

that is, Σ is also closed under countable intersections.

5. If $a \in \mathbb{R}$ then

$$(-\infty, a) = \bigcup_{n>a} (-n, a)$$

is a countable union of open intervals and so is a Borel set. Also,

$$(-\infty, a] = \bigcap_{n>0} (-\infty, a + \frac{1}{n}) \in \mathcal{B}$$

The right rays are complements of the left rays.

6. If $a \leq b$ then

$$[a, b] = \bigcap_{n>0} (a - \frac{1}{n}, b + \frac{1}{n}) \in \mathcal{B}$$

7. Write

$$A = (A \setminus B) \cup (A \cap B)$$
$$B = (B \setminus A) \cup (A \cap B)$$

Since these are disjoint unions, we have

$$\mathbb{P}(A) = \mathbb{P}(A \setminus B) \cup \mathbb{P}(A \cap B)$$
$$\mathbb{P}(B) = \mathbb{P}(B \setminus A) \cup \mathbb{P}(A \cap B)$$

and so

$$\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B)$$

But

$$(A \setminus B) \cup (A \cap B) \cup (B \setminus A)$$

is a disjoint union that equals $A \cup B$ and so we get the result.

8. This follows immediately from the Principle of Inclusion-Exclusion and the fact that probabilities are nonnegative.

10. If

$$\mathbb{P}(B = a) = p$$
$$\mathbb{P}(B = b) = q$$

then

$$0 = \mathcal{E}(B) = ap + bq$$

and

$$1 = \text{Var}(B) = a^2 p + b^2 q$$

Solving these equations gives

$$a = \frac{q}{\sqrt{pq}}, b = \frac{-p}{\sqrt{pq}}$$

13. Suppose that $A_1 \subseteq A_2 \subseteq \cdots$ is an increasing sequence of events and let

$$A = \bigcup_{i=1}^{\infty} A_i$$

The limit exists because it is the limit of an increasing bounded sequence of real numbers. Set $A_0 = \emptyset$ and write

$$A = \bigcup_{i=1}^{\infty} (A_i \setminus A_{i-1})$$

where the events $A_i \setminus A_{i-1}$ are disjoint. Then

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} (A_i \setminus A_{i-1})\right) \\
&= \sum_{i=1}^{\infty} \mathbb{P}(A_i \setminus A_{i-1}) \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(A_i \setminus A_{i-1}) \\
&= \lim_{n\to\infty} [\mathbb{P}(A_i) - \mathbb{P}(A_0)] \\
&= \lim_{n\to\infty} \mathbb{P}(A_i)
\end{aligned}$$

14. Let $A_t = \{X \leq t\}$. Let t_n be a sequence of real numbers for which $t_n \to -\infty$. Then the events A_{t_n} form a decreasing sequence whose intersection is $\emptyset$. The continuity of the probability implies that

$$\lim_{n\to\infty} f(t_n) = \lim_{n\to\infty} \mathbb{P}(A_{t_n}) = \mathbb{P}(\emptyset) = 0$$

Similarly, if $t_n \to \infty$ then $A_{t_n} \uparrow \Omega$ and so

$$\lim_{n\to\infty} f(t_n) = \lim_{n\to\infty} \mathbb{P}(A_{t_n}) = \mathbb{P}(\Omega) = 1$$

As to right continuity, let $t_n \downarrow t$. Then

$$(-\infty, t] = \bigcap_{n=1}^{\infty} (-\infty, t_n]$$

and so by continuity

$$f(t) = \mathbb{P}((-\infty, t]) = \lim_{n\to\infty} \mathbb{P}((-\infty, t_n]) = \lim_{n\to\infty} f(t_n)$$

which shows that f is right-continuous.

15. For part b)

$$\begin{aligned}
\mathbb{P}((a, b]) &= \mathbb{P}\left(\bigcup_{n=1}^{\infty} (a, b - \frac{1}{n}]\right) \\
&= \lim_{n\to\infty} \mathbb{P}((a, b - \frac{1}{n}]) \\
&= \lim_{n\to\infty} F(b - \frac{1}{n}) - F(a) \\
&= F(b-) - F(a)
\end{aligned}$$

16. We have $\mathbb{P}(X_n \leq t) \to \mathbb{P}(X \leq t)$ and so for $a > 0$

$$\mathbb{P}(aX_n + b \leq t) = \mathbb{P}(X_n \leq \frac{t-b}{a}) \to \mathbb{P}(X \leq \frac{t-b}{a}) = \mathbb{P}(aX + b \leq t)$$

A similar equation holds for $a < 0$.

Chapter 9: Black-Scholes

1. $\mu = 0.15, \sigma^2 = 0.03$.
2. $C = \$5.44$
3. $C = \$33.36$
4. We have

$$\begin{aligned}
\text{Var}(X_{p_T,i}) &= \mathcal{E}(X_{p_T,i}^2) - [\mathcal{E}(X_{p_T,i})]^2 \\
&= \frac{(1-p)^2}{p(1-p)} p + \frac{p^2}{p(1-p)}(1-p) + 0 \\
&= 1
\end{aligned}$$

5. For the expected value, we have

$$\begin{aligned}
\mathcal{E}_p(Q_s) &= \mathcal{E}_p(\sigma_s\sqrt{\Delta t}\sum_{i=1}^{T} X_{s,i}) \\
&= \sigma_s\sqrt{\Delta t}\sum_{i=1}^{T}\mathcal{E}_p(X_{s,i}) \\
&= \frac{\sigma_s\sqrt{\Delta t}}{\sqrt{s(1-s)}}\sum_{i=1}^{T}[(1-s)p - s(1-p)] \\
&= \frac{\sigma_s\sqrt{\Delta t}}{\sqrt{s(1-s)}}T(p-s) \\
&= \sigma_s\frac{t}{\sqrt{\Delta t}}\frac{p-s}{\sqrt{s(1-s)}}
\end{aligned}$$

10. We have

$$\begin{aligned}
S_0 &= e^{-rt}\mathcal{E}_\Pi(S_t) \\
&= e^{-rt}\mathcal{E}_\Pi(S_0 e^{\mu_\nu t + \sigma_\nu\sqrt{t}Z_t}) \\
&= S_0 e^{-rt+\mu_\nu t}\mathcal{E}_\Pi(e^{\sigma_\nu\sqrt{t}Z_t}) \\
&= S_0 e^{t(\mu_\nu - r)}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\sigma_\nu\sqrt{t}x}e^{-x^2/2}\,dx \\
&= S_0 e^{t(\mu_\nu - r)}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2 - 2\sigma_\nu\sqrt{t}x + \sigma_\nu^2 t) + \frac{1}{2}\sigma_\nu^2 t}\,dx \\
&= S_0 e^{t(\mu_\nu - r) + \frac{1}{2}\sigma_\nu^2 t}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\sigma_\nu\sqrt{t})^2}\,dx \\
&= S_0 e^{t(\mu_\nu - r) + \frac{1}{2}\sigma_\nu^2 t}
\end{aligned}$$

Thus

$$e^{t(\mu_\nu - r) + \frac{1}{2}\sigma_\nu^2 t} = 1$$

which happens if and only if

$$t(\mu_\nu - r) + \frac{1}{2}\sigma_\nu^2 t = 0$$

that is

$$\mu_\nu = r - \frac{1}{2}\sigma_\nu^2$$

11. Let N_U be the random variable representing the *number of* up-ticks in stock price over the lifetime of the model. Then, of course, the number of down-ticks is $T - N_U$. It follows that

$$\begin{aligned} S_{t,T} &= S_0 u^{N_U} d^{T-N_U} \\ &= S_0 e^{N_U \log u + (T-N_U)\log d} \\ &= S_0 e^{N_U(\log u - \log d) + T\log d} \end{aligned}$$

and so

$$H_{t,T} = N_U(\log u - \log d) + T\log d$$

Since N_U is a binomial random variable with parameters T and ν we have $\mathcal{E}(N_U) = T\nu$ and $\text{Var}(N_U) = T\nu(1-\nu)$. Thus

$$\begin{aligned} \mathcal{E}(H_{t,T}) &= T\nu(\log u - \log d) + T\log d \\ \text{Var}(H_{t,T}) &= T\nu(1-\nu)\,(\log u - \log d)^2 \end{aligned}$$

12. Standardizing the random variable H_T gives

$$\begin{aligned} H_T^* &= \frac{H_T - \mathcal{E}_\Pi(H_T)}{\sqrt{\text{Var}_\pi(H_T)}} \\ &= \frac{(2N_{T,U} - T)\log u - T(2\pi_U - 1)\log u}{\sqrt{4T\pi_U(1-\pi_U)(\log^2 u)}} \\ &= \frac{(2N_{T,U} - T) - T(2\pi_U - 1)}{\sqrt{4T\pi_U(1-\pi_U)}} \\ &= \frac{N_{T,U} - T\pi_U}{\sqrt{T\pi_U(1-\pi_U)}} \\ &= \frac{N_{T,U} - \mathcal{E}_\Pi(N_{T,U})}{\sqrt{\text{Var}_\pi(N_{T,U})}} \end{aligned}$$

Hence, since $N_{T,U}$ is a binomial random variable with parameters $n = T$ and

$$p = \pi_U = \frac{e^{r\frac{L}{T}} - d}{u - d}$$

the random variables

$$H_T^* = \frac{N_{T,U} - \mathcal{E}_\Pi(N_{T,U})}{\sqrt{\text{Var}_\pi(N_{T,U})}}$$

for $T = 1, 2, \ldots$ are standardized binomial random variables. According to Theorem 12 of Chapter 8 this sequence converges in distribution to a standard normal random variable $\mathcal{N}_{0,1}$.

13. High volatility implies that the stock price is more likely to be far from the strike price than is the case when the volatility is small. A high stock price is good for the owner of a call. On the other hand, when the stock price falls below the strike price, it doesn't really matter how far it falls—the call will expire and the owner will simply loose the purchase price. Thus, high upside volatility is good, high downside volatility is irrelevant. A similar argument obtains for a long put.

Chapter 10: Optimal Stopping and American Options

2. We have

$$\begin{aligned}[\tau = k] &= \{\omega \mid S_k \in B \text{ but } S_j \in B^c \text{ for } j < k\} \\ &= [S_0 \in B^c] \cup \cdots \cup [S_{k-1} \in B^c] \cup [S_k \in B]\end{aligned}$$

But since the price S_i is $\mathcal{P}_i$-measurable and since $(\mathcal{P}_i)$ is a filtration, we deduce that each of the events $[S_i \in B^c]$ and the event $[S_k \in B]$ are in the largest algebra $\mathcal{A}(\mathcal{P}_k)$. This is the condition required of a stopping time. Finally, for $k = T$ we have

$$[\tau = T] = [S_0 \in B^c] \cap \cdots \cap [S_{T-1} \in B^c] \in \mathcal{A}(\mathcal{P}_T)$$

3. This is the first time that $S_k \geq 2S_0$.
4. Write

$$\{k \mid S_k(\omega) \geq 2S_{k-1}(\omega)\} = \{k \mid \frac{S_k(\omega)}{S_{k-1}(\omega)} \geq 2\}$$

and consider first entry times for the adapted process $\left(\frac{S_k(\omega)}{S_{k-1}(\omega)}\right)$.

5. Exiting a set B is the same as entering the complement B^c.
6. For the maximum, we have

$$[\max\{\tau, \sigma\} = k] = \bigcup_{i=0}^{k} ([\tau = i] \cap [\sigma = k - i]) \in \mathcal{A}(\mathcal{P}_k)$$

The difference is not a stopping time, since it requirs knowledge of the future.

8. For any $\omega \in \Omega$ we have for $k = \tau(\omega)$

$$Z_{\tau(\omega)}(\omega) \leq U_{\tau(\omega)}(\omega)$$

that is

$$[Z_\tau](\omega) \leq [U_\tau](\omega)$$

or finally, $Z_\tau \leq U_\tau$.

10. The martingale condition is

$$\mathcal{E}(A_{k+1} \mid \mathcal{P}_k) = A_k$$

for $k \geq 0$. But the predictablility implies that $\mathcal{E}(A_{k+1} \mid \mathcal{P}_k) = A_{k+1}$ and so $A_{k+1} = A_k$.

11. Here is an Excel worksheet with the solution.

u=	1.2	**r=**	0	**K=**	21	**pi=**	0.5
d=	0.8	**S0=**	20	**T=**	3	**Call=1/Put=-1**	1

S0	**S1**	**S2**	**S3**	**Y0Bar**	**Y1Bar**	**Y2Bar**	**Y3Bar**
20	24	28.8	34.56	0	3	7.8	13.56
	16	19.2	23.04		0	0	2.04
		19.2	23.04			0	2.04
		12.8	15.36			0	0
			23.04				2.04
			15.36				0
			15.36				0
			10.24				0

V3Bar	**E(V3Bar|P2)**	**V2Bar**	**E(V2Bar|P1)**	**V1Bar**	**E(V1Bar|P0)**	**V0Bar**
13.56	7.8	7.8	4.41	4.41	2.46	2.46
2.04	7.8	7.8	4.41	4.41	2.46	2.46
2.04	1.02	1.02	4.41	4.41	2.46	2.46
0	1.02	1.02	4.41	4.41	2.46	2.46
2.04	1.02	1.02	0.51	0.51	2.46	2.46
0	1.02	1.02	0.51	0.51	2.46	2.46
0	0	0	0.51	0.51	2.46	2.46
0	0	0	0.51	0.51	2.46	2.46

References

Baxter, M. and Rennie, A., *Financial Calculus: An Introduction to Derivative Pricing*, Cambridge University Press, 1996, 0-521-55289-3.

Etheridge, A., *A Course in Financial Calculus*, Cambridge University Press, 2002, 0-521-89077-2.

Föllmer, H., *Stochastic Finance: An Introduction in Discrete Time*, Walter de Gruyter, 2002, 3-11-017119-8.

Kallianpur, G. and Karandikar, R., *Introduction to Option Pricing Theory*, Birkhauser, 1999, 0-8176-4108-4.

Lo, A. and MacKinlay, C., *A Non-Random Walk Down Wall Street*, Princeton University Press, 2001, 0-691-09256-7.

Medina, P. and Merino, S., *Mathematical Finance and Probability*, Birkhauser, 2003, 3-7643-6921-3.

Ross, S., *An Introduction to Mathematical Finance: Options and Other Topics*, Cambridge University Press, 1999, 0-521-77043-2.

Ross, S., *An Elementary Introduction to Mathematical Finance*, Cambridge University Press, 2002, 0-521-81429-4.

Shafer, G, and Vovk, V., *Probability and Finance: It's Only a Game*, John Wiley, 0-471-40226-5.

Wilmott, P., Howison, S. and Dewynne, J., *The Mathematics of Financial Derivatives: A Student Introduction*, Cambridge University Press, 1995, 0-521-49789-2.