Lectures on
Stochastic Calculus with Applications to Finance

Prepared for use in Statistics 441 at the University of Regina

Michael J. Kozdron
kozdron@stat.math.uregina.ca
http://stat.math.uregina.ca/∼kozdron

Contents

Preface

This set of lecture notes was used for *Statistics 441: Stochastic Calculus with Applications to Finance* at the University of Regina in the winter semester of 2009. It was the first time that the course was ever offered, and so part of the challenge was deciding what exactly needed to be covered. The end result: the first three-quarters of the course focused on the theory of option pricing while the final quarter focused on the mathematical analysis of risk. The official course textbook by Higham [12] was used for the first half of the course (Lecture #1 through Lecture #19). The advantage of that book is the inclusion of several MATLAB programs which illustrate many of the ideas in the development of the option pricing solution; unfortunately, Higham does not cover any stochastic calculus. The book by Čížek et al. [5] was used as the basis for a number of lectures on more advanced topics in option pricing including how to use the Feynman-Kac representation theorem to derive a characteristic function for a diffusion without actually solving a stochastic differential equation (Lecture #20 through Lecture #24). The final quarter of the course discussed risk measures and was mostly based on the book by Föllmer and Schied [8] (Lecture #25 through Lecture #31).

I would like to thank the students of STAT 441 from Winter 2009, namely Bridget Fortowsky, Boxiao Huang, Adam Kehler, LaToya Khan, Dustin Kidby, Bo Li, Fan Nie, Penny Roy, Jacob Schwartz, Kali Spencer, Jinghua Tang, Linh Van, Nathan Wollbaum, and Carter Woolhouse, for their patience and careful reading of earlier drafts of the lectures which resulted in many errors being eliminated.

I would also like to thank those students who worked with me as research associates in previous summers, often through the NSERC Undergraduate Student Research Award program, including Marina Christie (2005), Linshu Wang (2005), Sarah Wist (2006), and Bridget Fortowsky (2008). Many of the ideas incorporated into these notes were first worked out during those summers.

Regina, Saskatchewan — *Michael Kozdron*
April 2009

1
Introduction to Financial Derivatives

The primary goal of this course is to develop the *Black-Scholes option pricing formula* with a certain amount of mathematical rigour. This will require learning some *stochastic calculus* which is fundamental to the solution of the option pricing problem. The tools of stochastic calculus can then be applied to solve more sophisticated problems in finance and economics. As we will learn, the general Black-Scholes formula for pricing options has had a profound impact on the world of finance. In fact, trillions of dollars worth of options trades are executed each year using this model and its variants. In 1997, Myron S. Scholes (originally from Timmins, ON) and Robert C. Merton were awarded the Nobel Prize in Economics† for this work. (Fischer S. Black had died in 1995.)

Exercise 1.1. Read about these Nobel laureates at

`http://nobelprize.org/nobel_prizes/economics/laureates/1997/index.html`

and read the prize lectures *Derivatives in a Dynamic Environment* by Scholes and *Applications of Option-Pricing Theory: Twenty-Five Years Later* by Merton also available from this website.

As noted by McDonald in the Preface of his book *Derivative Markets* [18]:

"Thirty years ago the Black-Scholes formula was new, and derivatives was an esoteric and specialized subject. Today, a basic knowledge of derivatives is necessary to understand modern finance."

Before we proceed any further, we should be clear about what exactly a *derivative* is.

Definition 1.2. A *derivative* is a financial instrument whose value is determined by the value of something else.

That is, a *derivative* is a financial object *derived* from other, usually more basic, financial objects. The basic objects are known as *assets*. According to Higham [12], the term asset is used to describe any financial object whose value is known at present but is liable to change over time. A *stock* is an example of an asset.

† Technically, Scholes and Merton won *The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel.*

A *bond* is used to indicate cash invested in a risk-free savings account earning continuously compounded interest at a known rate.

Note. The term asset does not seem to be used consistently in the literature. There are some sources that consider a derivative to be an asset, while others consider a bond to be an asset. We will follow Higham [12] and use it primarily to refer to stocks (and not to derivatives or bonds).

Example 1.3. A *mutual fund* can be considered as a derivative since the mutual fund is composed of a range of investments in various stocks and bonds. Mutual funds are often seen as a good investment for people who want to hedge their risk (i.e., diversify their portfolio) and/or do not have the capital or desire to invest heavily in a single stock. Chartered banks, such as TD Canada Trust, sell mutual funds as well as other investments; see

`http://www.tdcanadatrust.com/mutualfunds/mffh.jsp`

for further information.

Other examples of derivatives include *options*, *futures*, and *swaps*. As you probably guessed, our goal is to develop a theory for pricing options.

Example 1.4. An example that is particularly relevant to residents of Saskatchewan is the *Guaranteed Delivery Contract* of the Canadian Wheat Board (CWB). See

`http://www.cwb.ca/public/en/farmers/contracts/guaranteed/`

for more information. The basic idea is that a farmer selling, say, barley can enter into a contract in August with the CWB whereby the CWB agrees to pay the farmer a fixed price per tonne of barley in December. The farmer is, in essence, betting that the price of barley in December will be lower that the contract price, in which case the farmer earns more for his barley than the market value. On the other hand, the CWB is betting that the market price per tonne of barley will be higher than the contract price, in which case they can immediately sell the barely that they receive from the farmer for the current market price and hence make a profit. This is an example of an option, and it is a fundamental problem to determine how much this option should be worth. That is, how much should the CWB charge the farmer for the opportunity to enter into an option contract. The Black-Scholes formula will tell us how to price such an option.

Thus, an *option* is a contract entered at time 0 whereby the buyer has the right, but not the obligation, to purchase, at time T, shares of a stock for the fixed value $\$E$. If, at time T, the actual price of the stock is greater than $\$E$, then the buyer exercises the option, buys the stocks for $\$E$ each, and immediately sells them to make a profit. If, at time T, the actual price of the stock is less than $\$E$, then the buyer does not exercise the option and the option becomes worthless. The question, therefore, is "How much should the buyer pay at time 0 for this contract?" Put another way, "What is the *fair price* of this contract?"

Technically, there are *call options* and *put options* depending on one's perspective.

Definition 1.5. A *European call option* gives its *holder* the right (but not the obligation) to purchase from the *writer* a prescribed asset for a prescribed price at a prescribed time in the future.

Definition 1.6. A *European put option* gives its *holder* the right (but not the obligation) to sell to the *writer* a prescribed asset for a prescribed price at a prescribed time in the future.

The prescribed price is known as the *exercise price* or the *strike price*. The prescribed time in the future is known as the *expiry date*.

The adjective *European* is to be contrasted with *American*. While a European option can be exercised only on the expiry date, an *American option* can be exercised at *any* time between the start date and the expiry date. In Chapter 18 of Higham [12], it is shown that American call options have the same value as European call options. American put options, however, are more complicated.

Hence, our primary goal will be to systematically develop a fair value of a European call option at time $t = 0$. (The so-called *put-call parity* for European options means that our solution will also apply to European put options.)

Finally, we will use the term *portfolio* to describe a combination of

(i) assets (i.e., stocks),
(ii) options, and
(iii) cash invested in a bank, i.e., bonds.

We assume that it is possible to hold negative amounts of each at no penalty. In other words, we will be allowed to *short sell* stocks and bonds freely and for no cost.

To conclude these introductory remarks, I would like to draw your attention to the recent book *Quant Job Interview Questions and Answers* by M. Joshi, A. Downes, and N. Denson [15]. To quote from the book description,

> "Designed to get you a job in quantitative finance, this book contains over 225 interview questions taken from actual interviews in the City and Wall Street. Each question comes with a full detailed solution, discussion of what the interviewer is seeking and possible follow-up questions. Topics covered include option pricing, probability, mathematics, numerical algorithms and C++, as well as a discussion of the interview process and the non-technical interview."

The "City" refers to "New York City" which is, arguably, the financial capital of the world. (And yes, at least one University of Regina actuarial science graduate has worked in New York City.) You can see a preview of this book at

`http://www.lulu.com/content/2436045`

and read questions (such as this one on page 17).

> "In the Black-Scholes world, price a European option with a payoff of $\max\{S_T^2 - K, 0\}$ at time T."

We will answer this question in Example 17.2.

2

Financial Option Valuation Preliminaries

Recall that a *portfolio* describes a combination of

(i) assets (i.e., stocks),
(ii) options, and
(iii) cash invested in a bank, i.e., bonds.

We will write S_t to denote the value of an asset at time $t \geq 0$. Since an asset is defined as a financial object whose value is known at present but is liable to change over time, we see that it is reasonable to *model* the asset price (i.e., stock price) by a stochastic process $\{S_t, t \geq 0\}$. There will be much to say about this later.

Suppose that $D(t)$ denotes the value at time t of an investment which grows according to a continuously compounded interest rate r. That is, suppose that an amount D_0 is invested at time 0. Its value at time $t \geq 0$ is given by

$$D(t) = e^{rt} D_0. \tag{2.1}$$

There are a couple of different ways to derive this formula for compound interest. One way familiar to actuarial science students is as the solution of a *constant force of interest equation.* That is, $D(t)$ is the solution of the equation

$$\delta_t = r \text{ with } r > 0$$

where

$$\delta_t = \frac{\mathrm{d}}{\mathrm{d}t} \log D(t)$$

and initial condition $D(0) = D_0$. In other words,

$$\frac{\mathrm{d}}{\mathrm{d}t} \log D(t) = r \quad \text{implies} \quad \frac{D'(t)}{D(t)} = r$$

so that $D'(t) = rD(t)$. This differential equation can then be solved by separation-of-variables giving (2.1).

Remark. We will use $D(t)$ as our model of the risk-free savings account, or bond. Assuming that such a bond exists means that having \$1 at time 0 or \$$e^{rt}$ at time t are both of equal value.

Equivalently, having \$1 at time t or $\$e^{-rt}$ at time 0 are both of equal value. This is sometimes known as the *time value of money.* Transferring money in this way is known as *discounting for interest* or *discounting for inflation.*

The word *arbitrage* is a fancy way of saying "money for nothing." One of the fundamental assumptions that we will make is that of *no arbitrage* (informally, we might call this the *no free lunch* assumption).

The form of the *no arbitrage* assumption given in Higham [12] is as follows.

"There is never an opportunity to make a risk-free profit that gives a greater return than that provided by interest from a bank deposit."

Note that this only applies to *risk-free* profit.

Example 2.1. Suppose that a company has offices in Toronto and London. The exchange rate between the dollar and the pound must be the same in both cities. If the exchange rate were \$1.60 = £1 in Toronto but only \$1.58 = £1 in London, then the company could instantly sell pounds in Toronto for \$1.60 each and buy them back in London for only \$1.58 making a risk-free profit of \$0.02 per pound. This would lead to unlimited profit for the company. Others would then execute the same trades leading to more unlimited profit and a total collapse of the market! Of course, the market would never allow such an *obvious discrepancy* to exist for any period of time.

The scenario described in the previous example is an illustration of an economic law known as *the law of one price* which states that "in an *efficient* market all identical goods must have only one price." An obvious violation of the efficient market assumption is found in the pricing of gasoline. Even in Regina, one can often find two gas stations on opposite sides of the street selling gas at different prices! (Figuring out how to legally take advantage of such a discrepancy is another matter altogether!)

The job of *arbitrageurs* is to scour the markets looking for arbitrage opportunities in order to make risk-free profit. The website

`http://www.arbitrageview.com/riskarb.htm`

lists some arbitrage opportunities in pending merger deals in the U.S. market. The following quote from this website is also worth including.

"It is important to note that merger arbitrage is not a complete risk free strategy. Profiting on the discount spread may look like the closest thing to a free lunch on Wall Street, however there are number of risks such as the probability of a deal failing, shareholders voting down a deal, revising the terms of the merger, potential lawsuits, etc. In addition the trading discount captures the time value of money for the period between the announcement and the closing of the deal. Again the arbitrageurs face the risk of a deal being prolonged and achieving smaller rate of return on an annualized basis."

Nonetheless, in order to derive a reasonable mathematical model of a financial market we must not allow for arbitrage opportunities.

A neat little argument gives the relationship between the value (at time 0) of a European call option C and the value (at time 0) of a European put option P (with both options being on the same asset S at the same expiry date T and same strike price E). This is known as the so-called *put-call parity for European options.*

Consider two portfolios Π_1 and Π_2 where (at time 0)

(i) Π_1 consists of one call option plus Ee^{-rT} invested in a risk-free bond, and
(ii) Π_2 consists of one put option plus one unit of the asset S_0.

At the expiry date T, the portfolio Π_1 is worth $\max\{S_T - E, 0\} + E = \max\{S_T, E\}$, and the portfolio Π_2 is worth $\max\{E - S_T, 0\} + S_T = \max\{S_T, E\}$. Hence, since both portfolios always give the same payoff, the no arbitrage assumption (or simply common sense) dictates that they have the same value at time 0. Thus,

$$C + Ee^{-rT} = P + S_0. \tag{2.2}$$

It is important to note that we have not figured out a fair value at time 0 for a European call option (or a European put option). We have only concluded that it is sufficient to price the European call option, because the value of the European put option follows immediately from (2.2). We will return to this result in Lecture #17.

Summary. We assume that it is possible to hold a portfolio of stocks and bonds. Both can be freely traded, and we can hold negative amounts of each without penalty. (That is, we can short-sell either instrument at no cost.) The stock is a risky asset which can be bought or sold (or even short-sold) in arbitrary units. Furthermore, it does not pay dividends. The bond, on the other hand, is a risk-free investment. The money invested in a bond is secure and grows according to a continuously compounded interest rate r. Trading takes place in continuous time, there are no transaction costs, and we will not be concerned with the bid-ask spread when pricing options. We trade in an efficient market in which arbitrage opportunities do not exist.

Example 2.2 (Pricing a forward contract). As already noted, our primary goal is to determine the fair price to pay (at time 0) for a European call option. The call option is only one example of a financial derivative. The oldest derivative, and arguably the most natural claim on a stock, is the *forward.*

If two parties enter into a *forward contract* (at time 0), then one party (the seller) agrees to give the other party (the holder) the specified stock at some prescribed time in the future for some prescribed price.

Suppose that T denotes the expiry date, F denotes the strike price, and the value of the stock at time $t > 0$ is S_t.

Note that a forward is not the same as a European call option. The stock *must* change hands at time T for $\$F$. The contract dictates that the seller is obliged to produce the stock at time T and that the holder is obliged to pay $\$F$ for the stock. Thus, the time T value of the forward contract for the holder is $S_T - F$, and the time T value for the seller is $F - S_T$.

Since money will change hands at time T, to determine the *fair value* of this contract means to determine the value of F.

Suppose that the distribution of the stock at time T is known. That is, suppose that S_T is a random variable having a known continuous distribution with density function f. The expected value of S_T is therefore

$$\mathbb{E}[S_T] = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x.$$

Thus, the expected value at time T of the forward contract is

$$\mathbb{E}[S_T - F]$$

(which is *calculable exactly* since the distribution of S_T is known). This suggests that the fair value of the strike price should satisfy

$$0 = \mathbb{E}[S_T - F] \text{ so that } F = \mathbb{E}[S_T].$$

In fact, the strong law of large numbers *justifies* this calculation—in the long run, the average of outcomes tends towards the expected value of a single outcome. In other words, the law of large numbers *suggests* that the fair strike price is $F = \mathbb{E}[S_T]$.

The problem is that this price is not enforceable. That is, although our calculation is not incorrect, it does lead to an arbitrage opportunity. Thus, in order to show that expectation pricing is not enforceable, we need to construct a portfolio which allows for an arbitrage opportunity.

Consider the seller of the contract obliged to deliver the stock at time T in exchange for \$$F$. The seller borrows S_0 now, buys the stock, puts it in a drawer, and waits. At time T, the seller then repays the loan for $S_0 e^{rT}$ but has the stock ready to deliver. Thus, if the strike price is less that $S_0 e^{rT}$, the seller will lose money with certainty. If the strike price is more than $S_0 e^{rT}$, the seller will make money with certainty.

Of course, the holder of the contract can run this scheme in reverse. Thus, writing more than $S_0 e^{rT}$ will mean that the holder will lose money with certainty.

Hence, the only *fair value* for the strike price is $F = S_0 e^{rT}$.

Remark. To put it quite simply, if there is an arbitrage price, then any other price is too dangerous to quote. Notice that the no arbitrage price for the forward contract completely ignores the randomness in the stock. If $\mathbb{E}(S_T) > F$, then the holder of a forward contract *expects* to make money. However, so do holders of the stock itself!

Remark. Both a forward contract and a *futures contract* are contracts whereby the seller is obliged to deliver the prescribed asset to the holder at the prescribed time for the prescribed price. There are, however, two main differences. The first is that futures are traded on an exchange, while forwards are traded over-the-counter. The second is that futures are margined, while forwards are not. These matters will not concern us in this course.

3

Normal and Lognormal Random Variables

The purpose of this lecture is to remind you of some of the key properties of normal and lognormal random variables which are basic objects in the mathematical theory of finance. (Of course, you already know of the ubiquity of the normal distribution from your elementary probability classes since it arises in the central limit theorem, and if you have studied any actuarial science you already realize how important lognormal random variables are.)

Recall that a continuous random variable Z is said to have a *normal distribution with mean* 0 *and variance* 1 if the density function of Z is

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad -\infty < z < \infty.$$

If Z has such a distribution, we write $Z \sim \mathcal{N}(0,1)$.

Exercise 3.1. Show directly that if $Z \sim \mathcal{N}(0,1)$, then $\mathbb{E}(Z) = 0$ and $\text{Var}(Z) = 1$. That is, calculate

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-\frac{z^2}{2}} \, \mathrm{d}z \quad \text{and} \quad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{z^2}{2}} \, \mathrm{d}z$$

using only results from elementary calculus. This calculation justifies the use of the "mean 0 and variance 1" phrase in the definition above.

Let $\mu \in \mathbb{R}$ and let $\sigma > 0$. We say that a continuous random variable X has a *normal distribution with mean* μ *and variance* σ^2 if the density function of X is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

If X has such a distribution, we write $X \sim \mathcal{N}(\mu, \sigma^2)$.

Shortly, you will be asked to prove the following result which establishes the relationship between the random variables $Z \sim \mathcal{N}(0,1)$ and $X \sim \mathcal{N}(\mu, \sigma^2)$.

Theorem 3.2. *Suppose that $Z \sim \mathcal{N}(0,1)$, and let $\mu \in \mathbb{R}$, $\sigma > 0$ be constants. If the random variable X is defined by $X = \sigma Z + \mu$, then $X \sim \mathcal{N}(\mu, \sigma^2)$. Conversely, if $X \sim \mathcal{N}(\mu, \sigma^2)$, and* *the random variable Z is defined by*

$$Z = \frac{X - \mu}{\sigma},$$

then $Z \sim \mathcal{N}(0, 1)$.

Let

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, \mathrm{d}x$$

denote the standard normal cumulative distribution function. That is, $\Phi(z) = \mathbf{P}\{Z \le z\} = F_Z(z)$ is the distribution function of a random variable $Z \sim \mathcal{N}(0, 1)$.

Remark. Higham [12] writes N instead of Φ for the standard normal cumulative distribution function. The notation Φ is far more common in the literature, and so we prefer to use it instead of N.

Exercise 3.3. Show that $1 - \Phi(z) = \Phi(-z)$.

Exercise 3.4. Show that if $X \sim \mathcal{N}(\mu, \sigma^2)$, then the distribution function of X is given by

$$F_X(x) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Exercise 3.5. Use the result of Exercise 3.4 to complete the proof of Theorem 3.2.

The next two exercises are extremely important for us. In fact, these exercises ask you to prove special cases of the Black-Scholes formula.

Notation. We write $x^+ = \max\{0, x\}$ to denote the *positive part* of x.

Exercise 3.6. Suppose that $Z \sim \mathcal{N}(0, 1)$, and let $c > 0$ be a constant. Compute

$$\mathbb{E}[\,(e^Z - c)^+\,].$$

You will need to express your answer in terms of Φ.

Answer. $e^{1/2}\,\Phi(1 - \log c) - c\,\Phi(-\log c)$

Exercise 3.7. Suppose that $Z \sim \mathcal{N}(0, 1)$, and let $a > 0$, $b > 0$, and $c > 0$ be constants. Compute

$$\mathbb{E}[\,(ae^{bZ} - c)^+\,].$$

You will need to express your answer in terms of Φ.

Answer. $ae^{b^2/2}\,\Phi\left(b + \frac{1}{b}\log\frac{a}{c}\right) - c\,\Phi\left(\frac{1}{b}\log\frac{a}{c}\right)$

Recall that the characteristic function of a random variable X is the function $\varphi_X : \mathbb{R} \to \mathbb{C}$ given by $\varphi_X(t) = \mathbb{E}(e^{itX})$.

Exercise 3.8. Show that if $Z \sim \mathcal{N}(0, 1)$, then the characteristic function of Z is

$$\varphi_Z(t) = \exp\left\{-\frac{t^2}{2}\right\}.$$

Exercise 3.9. Show that if $X \sim \mathcal{N}(\mu, \sigma^2)$, then the characteristic function of X is

$$\varphi_X(t) = \exp\left\{ i\mu t - \frac{\sigma^2 t^2}{2} \right\}.$$

The importance of characteristic functions is that they completely characterize the distribution of a random variable since the characteristic function always exists (unlike moment generating functions which do not always exist).

Theorem 3.10. *Suppose that X and Y are random variables. The characteristic functions φ_X and φ_Y are equal if and only if X and Y are equal in distribution (that is, $F_X = F_Y$).*

Proof. For a proof, see Theorem 4.1.2 on page 160 of [10]. □

Exercise 3.11. One consequence of this theorem is that it allows for an alternative solution to Exercise 3.5. That is, use characteristic functions to complete the proof of Theorem 3.2.

We will have occasion to analyze sums of normal random variables. The purpose of the next several exercises and results is to collect all of the facts that we will need. The first exercise shows that a linear combination of independent normals is again normal.

Exercise 3.12. Suppose that $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are independent. Show that for any a, $b \in \mathbb{R}$,

$$aX_1 + bX_2 \sim \mathcal{N}\left(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2\right).$$

Of course, whenever two random variables are independent, they are necessarily uncorrelated. However, the converse is not true in general, even in the case of normal random variables. As the following example shows, uncorrelated normal random variables need not be independent.

Example 3.13. Suppose that $X_1 \sim \mathcal{N}(0, 1)$ and suppose further that Y is independent of X_1 with $\mathbf{P}\{Y = 1\} = \mathbf{P}\{Y = -1\} = 1/2$. If we set $X_2 = YX_1$, then it follows that $X_2 \sim \mathcal{N}(0, 1)$. (Verify this fact.) Furthermore, X_1 and X_2 are uncorrelated since

$$\mathrm{Cov}(X_1, X_2) = \mathbb{E}(X_1X_2) = \mathbb{E}(X_1^2 Y) = \mathbb{E}(X_1^2)\mathbb{E}(Y) = 1 \cdot 0 = 0$$

using the fact that X_1 and Y are independent. However, X_1 and X_2 are not independent since

$$\mathbf{P}\{X_1 \geq 1, X_2 \geq 1\} = \mathbf{P}\{X_1 \geq 1, Y = 1\} = \mathbf{P}\{X_1 \geq 1\}\mathbf{P}\{Y = 1\} = \frac{1}{2}\mathbf{P}\{X_1 \geq 1\}$$

whereas

$$\mathbf{P}\{X_1 \geq 1\}\mathbf{P}\{X_2 \geq 1\} = [\mathbf{P}\{X_1 \geq 1\}]^2.$$

Since $\mathbf{P}\{X_1 \geq 1\}$ does not equal either 0 or 1/2 (it actually equals $\doteq 0.1587$) we see that

$$\frac{1}{2}\mathbf{P}\{X_1 \geq 1\} \neq [\mathbf{P}\{X_1 \geq 1\}]^2.$$

An extension of this same example also shows that the sum of uncorrelated normal random variables need not be normal.

Example 3.13 (continued). We will now show that $X_1 + X_2$ is not normally distributed. If $X_1 + X_2$ were normally distributed, then it would necessarily be the case that for any $x \in \mathbb{R}$, we would have $\mathbf{P}\{X_1 + X_2 = x\} = 0$. Indeed, this is true for any continuous random variable. But we see that $\mathbf{P}\{X_1 + X_2 = 0\} = \mathbf{P}\{Y = -1\} = 1/2$ which shows that $X_1 + X_2$ cannot be a normal random variable (let alone a continuous random variable).

However, if we have a bivariate normal random vector $\mathbf{X} = (X_1, X_2)'$, then independence of the components and no correlation between them are equivalent.

Theorem 3.14. *Suppose that $\mathbf{X} = (X_1, X_2)'$ has a bivariate normal distribution so that the components of $\mathbf{X}$, namely X_1 and X_2, are each normally distributed. Furthermore, X_1 and X_2 are uncorrelated if and only if they are independent.*

Proof. For a proof, see Theorem V.7.1 on page 133 of Gut [9]. □

Two important variations on the previous results are worth mentioning.

Theorem 3.15 (Cramér). *If X and Y are independent random variables such that $X + Y$ is normally distributed, then X and Y themselves are each normally distributed.*

Proof. For a proof of this result, see Theorem 19 on page 53 of [6]. □

In the special case when X and Y are also identically distributed, Cramér's theorem is easy to prove.

Exercise 3.16. Suppose that X and Y are independent and identically distributed random variables such that $X + Y \sim \mathcal{N}(2\mu, 2\sigma^2)$. Prove that $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y \sim \mathcal{N}(\mu, \sigma^2)$.

Example 3.13 showed that uncorrelated normal random variables need not be independent and need not have a normal sum. However, if uncorrelated normal random variables are known to have a normal sum, then it must be the case that they are independent.

Theorem 3.17. *If $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are normally distributed random variables with $\operatorname{Cov}(X_1, X_2) = 0$, and if $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, then X_1 and X_2 are independent.*

Proof. In order to prove that X_1 and X_2 are independent, it is sufficient to prove that the characteristic function of $X_1 + X_2$ equals the product of the characteristic functions of X_1 and X_2. Since $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ we see using Exercise 3.9 that

$$\varphi_{X_1+X_2}(t) = \exp\left\{ i(\mu_1 + \mu_2)t - \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2} \right\}.$$

Furthermore, since $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ we see that

$$\varphi_{X_1}(t)\varphi_{X_2}(t) = \exp\left\{ i\mu_1 t - \frac{\sigma_1^2 t^2}{2} \right\} \cdot \exp\left\{ i\mu_2 t - \frac{\sigma_2^2 t^2}{2} \right\} = \exp\left\{ i(\mu_1 + \mu_2)t - \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2} \right\}.$$

In other words, $\varphi_{X_1}(t)\varphi_{X_2}(t) = \varphi_{X_1+X_2}(t)$ which establishes the result. □

Remark. Actually, the assumption that $\operatorname{Cov}(X_1, X_2) = 0$ is unnecessary in the previous theorem. The same proof shows that if $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are normally distributed random variables, and if $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, then X_1 and X_2 are independent. It is now a consequence that $\operatorname{Cov}(X_1, X_2) = 0$.

A variation of the previous result can be proved simply by equating variances.

Exercise 3.18. If $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are normally distributed random variables, and if $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)$, then $\operatorname{Cov}(X_1, X_2) = \rho\sigma_1\sigma_2$ and $\operatorname{Corr}(X_1, X_2) = \rho$.

Our final result gives conditions under which normality is preserved for limits in distribution. Before stating this theorem, we need to recall the definition of *convergence in distribution.*

Definition 3.19. Suppose that $X_1, X_2, \ldots$ and X are random variables with distribution functions F_n, $n = 1, 2, \ldots$, and F, respectively. We say that X_n *converges in distribution* to X as $n \to \infty$ if

$$\lim_{n \to \infty} F_n(x) = F(x)$$

for all $x \in \mathbb{R}$ at which F is continuous.

The relationship between convergence in distribution and characteristic functions is extremely important for us.

Theorem 3.20. *Suppose that $X_1, X_2, \ldots$ are random variables with characteristic functions φ_{X_n}, $n = 1, 2, \ldots$. It then follows that $\varphi_{X_n}(t) \to \varphi_X(t)$ as $n \to \infty$ for all $t \in \mathbb{R}$ if and only if X_n converges in distribution to X.*

Proof. For a proof of this result, see Theorem 5.9.1 on page 238 of [10]. □

It is worth noting that in order to apply the result of the previous theorem we must know *a priori* what the limiting random variable X is. In the case when we only know that the characteristic functions converge to something, we must be a bit more careful.

Theorem 3.21. *Suppose that $X_1, X_2, \ldots$ are random variables with characteristic functions φ_{X_n}, $n = 1, 2, \ldots$. If $\varphi_{X_n}(t)$ converges to some function $\varphi(t)$ as $n \to \infty$ for all $t \in \mathbb{R}$ and $\varphi(t)$ is continuous at 0, then there exists a random variable X with characteristic function φ such that X_n converges in distribution to X.*

Proof. For a proof of this result, see Theorem 5.9.2 on page 238 of [10]. □

Remark. The statement of the central limit theorem is really a statement about convergence in distribution, and its proof follows after a careful analysis of characteristic functions from Theorems 3.10 and 3.21.

We are now ready to prove that normality is preserved under convergence in distribution. The proof uses a result known as Slutsky's theorem, and so we will state and prove this first.

Theorem 3.22 (Slutsky). *Suppose that the random variables X_n, $n = 1, 2, \ldots$, converge in distribution to X and that the sequence of real numbers a_n, $n = 1, 2, \ldots$, converges to the finite real number a. It then follows that $X_n + a_n$ converges in distribution to $X + a$ and that $a_n X_n$ converges in distribution to aX.*

Proof. We begin by observing that for $\varepsilon > 0$ fixed, we have

$$\begin{aligned}\mathbf{P}\{X_n + a_n \le x\} &= \mathbf{P}\{X_n + a_n \le x, |a_n - a| < \varepsilon\} + \mathbf{P}\{X_n + a_n \le x, |a_n - a| > \varepsilon\} \\ &\le \mathbf{P}\{X_n + a_n \le x, |a_n - a| < \varepsilon\} + \mathbf{P}\{|a_n - a| > \varepsilon\} \\ &\le \mathbf{P}\{X_n \le x - a + \varepsilon\} + \mathbf{P}\{|a_n - a| > \varepsilon\}\end{aligned}$$

That is,

$$F_{X_n + a_n}(x) \le F_{X_n}(x - a + \varepsilon) + \mathbf{P}\{|a_n - a| > \varepsilon\}.$$

Since $a_n \to a$ as $n \to \infty$ we see that $\mathbf{P}\{|a_n - a| > \varepsilon\} \to 0$ as $n \to \infty$ and so

$$\limsup_{n\to\infty} F_{X_n + a_n}(x) \le F_X(x - a + \varepsilon)$$

for all points $x - a + \varepsilon$ at which F is continuous. Similarly,

$$\liminf_{n\to\infty} F_{X_n + a_n}(x) \ge F_X(x - a - \varepsilon)$$

for all points $x - a - \varepsilon$ at which F is continuous. Since $\varepsilon > 0$ can be made arbitrarily small and since F_X has at most countably many points of discontinuity, we conclude that

$$\lim_{n\to\infty} F_{X_n + a_n}(x) = F_X(x - a) = F_{X+a}(x)$$

for all $x \in \mathbb{R}$ at which F_{X+a} is continuous. The proof that $a_n X_n$ converges in distribution to aX is similar. □

Exercise 3.23. Complete the details to show that $a_n X_n$ converges in distribution to aX.

Theorem 3.24. *Suppose that $X_1, X_2, \ldots$ is a sequence of random variables with $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots$. If the limits*

$$\lim_{n\to\infty} \mu_n \quad \textit{and} \quad \lim_{n\to\infty} \sigma_n^2$$

each exist and are finite, then the sequence $\{X_n, n = 0, 1, 2, \ldots\}$ converges in distribution to a random variable X. Furthermore, $X \sim \mathcal{N}(\mu, \sigma^2)$ where

$$\mu = \lim_{n\to\infty} \mu_n \quad \textit{and} \quad \sigma^2 = \lim_{n\to\infty} \sigma_n^2.$$

Proof. For each n, let

$$Z_n = \frac{X_n - \mu_n}{\sigma_n}$$

so that $Z_n \sim \mathcal{N}(0, 1)$ by Theorem 3.2. Clearly, Z_n converges in distribution to some random variable Z with $Z \sim \mathcal{N}(0, 1)$. By Slutsky's theorem, since Z_n converges in distribution to Z, it follows that $X_n = \sigma_n Z_n + \mu_n$ converges in distribution to $\sigma Z + \mu$. If we now define $X = \sigma Z + \mu$, then X_n converges in distribution to X and it follows from Theorem 3.2 that $X \sim \mathcal{N}(\mu, \sigma^2)$. □

We end this lecture with a brief discussion of lognormal random variables. Recall that if $X \sim \mathcal{N}(\mu, \sigma^2)$, then the *moment generating function* of X is

$$m_X(t) = \mathbb{E}(e^{tX}) = \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}.$$

Exercise 3.25. Suppose that $X \sim \mathcal{N}(\mu, \sigma^2)$ and let $Y = e^X$.

(a) Determine the density function for Y

(b) Determine the distribution function for Y. *Hint: You will need to express your answer in terms of* Φ.

(c) Compute $\mathbb{E}(Y)$ and $\mathrm{Var}(Y)$. *Hint: Use the moment generating function of* X.

Answer. (c) $\mathbb{E}(Y) = \exp\{\mu + \frac{\sigma^2}{2}\}$ and $\mathrm{Var}(Y) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$.

Definition 3.26. We say that a random variable Y has a *lognormal distribution with parameters* μ *and* σ^2, written

$$Y \sim \mathcal{LN}(\mu, \sigma^2),$$

if $\log(Y)$ is normally distributed with mean μ and variance σ^2. That is, $Y \sim \mathcal{LN}(\mu, \sigma^2)$ iff $\log(Y) \sim \mathcal{N}(\mu, \sigma^2)$. Equivalently, $Y \sim \mathcal{LN}(\mu, \sigma^2)$ iff $Y = e^X$ with $X \sim \mathcal{N}(\mu, \sigma^2)$.

Exercise 3.27. Suppose that $Y_1 \sim \mathcal{LN}(\mu_1, \sigma_1^2)$ and $Y_2 \sim \mathcal{LN}(\mu_2, \sigma_2^2)$ are independent lognormal random variables. Prove that $Z = Y_1 \cdot Y_2$ is lognormally distributed and determine the parameters of Z.

Remark. As shown in STAT 351, if a random variable Y has a lognormal distribution, then the moment generating function of Y does not exist.

4

Discrete-Time Martingales

The concept of a martingale is fundamental to modern probability and is one of the key tools needed to study mathematical finance. Although we saw the definition in STAT 351, we are now going to need to be a little more careful than we were in that class. This will be especially true when we study continuous-time martingales.

Definition 4.1. A sequence $X_0, X_1, X_2, \ldots$ of random variables is said to be a *martingale* if $\mathbb{E}(X_{n+1}|X_0, X_1, \ldots, X_n) = X_n$ for every $n = 0, 1, 2, \ldots$.

Technically, we need all of the random variables to have finite expectation in order that conditional expectations be defined. Furthermore, we will find it useful to introduce the following notation. Let $\mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n)$ denote the *information* contained in the sequence $\{X_0, X_1, \ldots, X_n\}$ up to (and including) time n. We then call the sequence $\{\mathcal{F}_n, n = 0, 1, 2, \ldots\} = \{\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \ldots\}$ a *filtration*.

Definition 4.2. A sequence $\{X_n, n = 0, 1, 2 \ldots\}$ of random variables is said to be a *martingale* with respect to the filtration $\{\mathcal{F}_n, n = 0, 1, 2, \ldots\}$ if

(i) $X_n \in \mathcal{F}_n$ for every $n = 0, 1, 2, \ldots$,
(ii) $\mathbb{E}|X_n| < \infty$ for every $n = 0, 1, 2, \ldots$, and
(iii) $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$ for every $n = 0, 1, 2, \ldots$.

If $X_n \in \mathcal{F}_n$, then we often say that X_n is *adapted*. The intuitive idea is that if X_n is adapted, then X_n is "known" at time n. In fact, you are already familiar with this notion from STAT 351.

Remark. Suppose that n is fixed, and let $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$. Clearly $\mathcal{F}_{n-1} \subseteq \mathcal{F}_n$ and so $X_1 \in \mathcal{F}_n$, $X_2, \in \mathcal{F}_n, \ldots, X_n \in \mathcal{F}_n$.

Theorem 4.3. *Let $X_1, X_2, \ldots, X_n$, Y be random variables, let $g : \mathbb{R}^n \to \mathbb{R}$ be a function, and let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. It then follows that*

(a) $\mathbb{E}(g(X_1, X_2, \ldots, X_n)\, Y|\mathcal{F}_n) = g(X_1, X_2, \ldots, X_n)\mathbb{E}(Y|\mathcal{F}_n)$ *(taking out what is known),*
(b) $\mathbb{E}(Y|\mathcal{F}_n) = \mathbb{E}(Y)$ *if Y is independent of $\mathcal{F}_n$, and*
(c) $\mathbb{E}(\mathbb{E}(Y|\mathcal{F}_n)) = \mathbb{E}(Y)$.

One useful fact about martingales is that they have stable expectation.

Theorem 4.4. *If* $\{X_n,\, n = 0, 1, 2, \ldots\}$ *is a martingale, then* $\mathbb{E}(X_n) = \mathbb{E}(X_0)$ *for every* $n = 0, 1, 2, \ldots$.

Proof. Since

$$\mathbb{E}(X_{n+1}) = \mathbb{E}(\mathbb{E}(X_{n+1}|\mathcal{F}_n)) = \mathbb{E}(X_n),$$

we can iterate to conclude that

$$\mathbb{E}(X_{n+1}) = \mathbb{E}(X_n) = \mathbb{E}(X_{n-1}) = \cdots = \mathbb{E}(X_0)$$

as required. □

Exercise 4.5. Suppose that $\{X_n,\, n = 1, 2, \ldots\}$ is a discrete-time stochastic process. Show that $\{X_n,\, n = 1, 2, \ldots\}$ is a martingale with respect to the filtration $\{\mathcal{F}_n, n = 0, 1, 2, \ldots\}$ if and only if

(i) $X_n \in \mathcal{F}_n$ for every $n = 0, 1, 2, \ldots$,
(ii) $\mathbb{E}|X_n| < \infty$ for every $n = 0, 1, 2, \ldots$, and
(iii) $\mathbb{E}(X_n|\mathcal{F}_m) = X_m$ for every integer m with $0 \le m < n$.

We are now going to study several examples of martingales. Most of them are variants of simple random walk which we define in the next example.

Example 4.6. Suppose that $Y_1, Y_2, \ldots$ are independent, identically distributed random variables with $\mathbf{P}\{Y_1 = 1\} = \mathbf{P}\{Y = -1\} = 1/2$. Let $S_0 = 0$, and for $n = 1, 2, \ldots$, define $S_n = Y_1 + Y_2 + \cdots + Y_n$. The sequence $\{S_n, n = 0, 1, 2, \ldots\}$ is called a *simple random walk (starting at* 0*)*. Before we show that $\{S_n, n = 0, 1, 2, \ldots\}$ is a martingale, it will be useful to calculate $\mathbb{E}(S_n)$, $\mathrm{Var}(S_n)$, and $\mathrm{Cov}(S_n, S_{n+1})$. Observe that

$$(Y_1 + Y_2 + \cdots + Y_n)^2 = Y_1^2 + Y_2^2 + \cdots + Y_n^2 + \sum_{i \neq j} Y_i Y_j.$$

Since $\mathbb{E}(Y_1) = 0$ and $\mathrm{Var}(Y_1) = \mathbb{E}(Y_1^2) = 1$, we find

$$\mathbb{E}(S_n) = \mathbb{E}(Y_1 + Y_2 + \cdots + Y_n) = \mathbb{E}(Y_1) + \mathbb{E}(Y_2) + \cdots + \mathbb{E}(Y_n) = 0$$

and

$$\begin{aligned}\mathrm{Var}(S_n) = \mathbb{E}(S_n^2) = \mathbb{E}(Y_1 + Y_2 + \cdots + Y_n)^2 &= \mathbb{E}(Y_1^2) + \mathbb{E}(Y_2^2) + \cdots + \mathbb{E}(Y_n^2) + \sum_{i \neq j} \mathbb{E}(Y_i Y_j) \\ &= 1 + 1 + \cdots + 1 + 0 \\ &= n\end{aligned}$$

since $\mathbb{E}(Y_i Y_j) = \mathbb{E}(Y_i)\mathbb{E}(Y_j)$ when $i \neq j$ because of the assumed independence of $Y_1, Y_2, \ldots$. Since $S_{n+1} = S_n + Y_{n+1}$ we see that

$$\mathrm{Cov}(S_n, S_{n+1}) = \mathrm{Cov}(S_n, S_n + Y_{n+1}) = \mathrm{Cov}(S_n, S_n) + \mathrm{Cov}(S_n, Y_{n+1}) = \mathrm{Var}(S_n) + 0$$

using the fact that Y_{n+1} is independent of S_n. Furthermore, since $\mathrm{Var}(S_n) = n$, we conclude $\mathrm{Cov}(S_n, S_{n+1}) = n$.

Exercise 4.7. As a generalization of this covariance calculation, show that $\operatorname{Cov}(S_n, S_m) = \min\{n, m\}$.

Example 4.6 (continued). We now show that the simple random walk $\{S_n, n = 0, 1, 2, \ldots\}$ is a martingale. This also illustrates the usefulness of the $\mathcal{F}_n$ notation since

$$\mathcal{F}_n = \sigma(S_0, S_1, \ldots, S_n) = \sigma(Y_1, \ldots, Y_n).$$

Notice that

$$\mathbb{E}(S_{n+1}|\mathcal{F}_n) = \mathbb{E}(Y_{n+1} + S_n|\mathcal{F}_n) = \mathbb{E}(Y_{n+1}|\mathcal{F}_n) + \mathbb{E}(S_n|\mathcal{F}_n).$$

Since Y_{n+1} is independent of $\mathcal{F}_n$ we conclude that

$$\mathbb{E}(Y_{n+1}|\mathcal{F}_n) = \mathbb{E}(Y_{n+1}) = 0.$$

If we condition on $\mathcal{F}_n$, then S_n is *known*, and so

$$\mathbb{E}(S_n|\mathcal{F}_n) = S_n.$$

Combined we conclude

$$\mathbb{E}(S_{n+1}|\mathcal{F}_n) = \mathbb{E}(Y_{n+1}|\mathcal{F}_n) + \mathbb{E}(S_n|\mathcal{F}_n) = 0 + S_n = S_n$$

which proves that $\{S_n, n = 0, 1, 2, \ldots\}$ is a martingale.

Example 4.6 (continued). Next we show that $\{S_n^2 - n, n = 0, 1, 2, \ldots\}$ is also a martingale. Let $M_n = S_n^2 - n$. We must show that $\mathbb{E}(M_{n+1}|\mathcal{F}_n) = M_n$ since

$$\mathcal{F}_n = \sigma(M_0, M_1, \ldots, M_n) = \sigma(S_0, S_1, \ldots, S_n).$$

Notice that

$$\mathbb{E}(S_{n+1}^2|\mathcal{F}_n) = \mathbb{E}((Y_{n+1} + S_n)^2|\mathcal{F}_n) = \mathbb{E}(Y_{n+1}^2|\mathcal{F}_n) + 2\mathbb{E}(Y_{n+1}S_n|\mathcal{F}_n) + \mathbb{E}(S_n^2|\mathcal{F}_n).$$

However, $\mathbb{E}(Y_{n+1}^2|\mathcal{F}_n) = \mathbb{E}(Y_{n+1}^2) = 1$,

$$\mathbb{E}(Y_{n+1}S_n|\mathcal{F}_n) = S_n\mathbb{E}(Y_{n+1}|\mathcal{F}_n) = S_n\mathbb{E}(Y_{n+1}) = 0,$$

and $\mathbb{E}(S_n^2|\mathcal{F}_n) = S_n^2$ from which we conclude that $\mathbb{E}(S_{n+1}^2|\mathcal{F}_n) = S_n^2 + 1$. Therefore,

$$\begin{aligned}\mathbb{E}(M_{n+1}|\mathcal{F}_n) = \mathbb{E}(S_{n+1}^2 - (n+1)|\mathcal{F}_n) = \mathbb{E}(S_{n+1}^2|\mathcal{F}_n) - (n+1) &= S_n^2 + 1 - (n+1) \\ &= S_n^2 - n \\ &= M_n\end{aligned}$$

and so we conclude that $\{M_n, n = 0, 1, 2, \ldots\} = \{S_n^2 - n, n = 0, 1, 2, \ldots\}$ is a martingale.

Example 4.6 (continued). We are now going to construct one more martingale related to simple random walk. Suppose that $\theta \in \mathbb{R}$ and let

$$Z_n = (\operatorname{sech}\theta)^n e^{\theta S_n}, \quad n = 0, 1, 2, \ldots,$$

where the hyperbolic secant is defined as

$$\operatorname{sech}\theta = \frac{2}{e^\theta + e^{-\theta}}.$$

We will show that $\{Z_n, n = 0, 1, 2, \ldots\}$ is a martingale. Thus, we must verify that

$$\mathbb{E}(Z_{n+1}|\mathcal{F}_n) = Z_n$$

since

$$\mathcal{F}_n = \sigma(Z_0, Z_1, \ldots, Z_n) = \sigma(S_0, S_1, \ldots, S_n).$$

Notice that $S_{n+1} = S_n + Y_{n+1}$ which implies

$$\begin{aligned} Z_{n+1} = (\operatorname{sech}\theta)^{n+1} e^{\theta S_{n+1}} = (\operatorname{sech}\theta)^{n+1} e^{\theta(S_n + Y_{n+1})} &= (\operatorname{sech}\theta)^n e^{\theta S_n} \cdot (\operatorname{sech}\theta) e^{\theta Y_{n+1}} \\ &= Z_n \cdot (\operatorname{sech}\theta) e^{\theta Y_{n+1}}. \end{aligned}$$

Therefore,

$$\mathbb{E}(Z_{n+1}|\mathcal{F}_n) = \mathbb{E}(Z_n \cdot (\operatorname{sech}\theta) e^{\theta Y_{n+1}}|\mathcal{F}_n) = Z_n \mathbb{E}((\operatorname{sech}\theta) e^{\theta Y_{n+1}}|\mathcal{F}_n) = Z_n \mathbb{E}((\operatorname{sech}\theta) e^{\theta Y_{n+1}})$$

where the second equality follows by "taking out what is known" and the third equality follows by independence. The final step is to compute $\mathbb{E}((\operatorname{sech}\theta) e^{\theta Y_{n+1}})$. Note that

$$\mathbb{E}(e^{\theta Y_{n+1}}) = e^{\theta \cdot 1} \cdot \frac{1}{2} + e^{\theta \cdot -1} \cdot \frac{1}{2} = \frac{e^{\theta} + e^{-\theta}}{2} = \frac{1}{\operatorname{sech}\theta}$$

and so

$$\mathbb{E}((\operatorname{sech}\theta) e^{\theta Y_{n+1}}) = (\operatorname{sech}\theta)\mathbb{E}(e^{\theta Y_{n+1}}) = (\operatorname{sech}\theta) \cdot \frac{1}{\operatorname{sech}\theta} = 1.$$

In other words, we have shown that

$$\mathbb{E}(Z_{n+1}|\mathcal{F}_n) = Z_n$$

which implies that $\{Z_n, n = 0, 1, 2 \ldots\}$ is a martingale.

The following two examples give more martingales derived from simple random walk.

Example 4.8. As in the previous example, let $Y_1, Y_2, \ldots$ be independent and identically distributed random variables with $\mathbf{P}\{Y_1 = 1\} = \mathbf{P}\{Y_1 = -1\} = \frac{1}{2}$, set $S_0 = 0$, and for $n = 1, 2, 3, \ldots$, define the random variable S_n by $S_n = Y_1 + \cdots + Y_n$ so that $\{S_n, n = 0, 1, 2, \ldots\}$ is a simple random walk starting at 0. Define the process $\{M_n, n = 0, 1, 2, \ldots\}$ by setting

$$M_n = S_n^3 - 3nS_n.$$

Show that $\{M_n, n = 0, 1, 2, \ldots\}$ is a martingale.

Solution. If $M_n = S_n^3 - 3nS_n$, then

$$\begin{aligned} M_{n+1} = S_{n+1}^3 - 3(n+1)S_{n+1} &= (S_n + Y_{n+1})^3 - 3(n+1)(S_n + Y_{n+1}) \\ &= S_n^3 + 3S_n^2 Y_{n+1} + 3S_n Y_{n+1}^2 + Y_{n+1}^3 - 3(n+1)S_n - 3(n+1)Y_{n+1} \\ &= M_n + 3S_n(Y_{n+1}^2 - 1) + 3S_n^2 Y_{n+1} - 3(n+1)Y_{n+1} + Y_{n+1}^3. \end{aligned}$$

Thus, we see that we will be able to conclude that $\{M_n, n = 0, 1, \ldots\}$ is a martingale if we can show that

$$\mathbb{E}\left(3S_n(Y_{n+1}^2 - 1) + 3S_n^2 Y_{n+1} - 3(n+1)Y_{n+1} + Y_{n+1}^3 | \mathcal{F}_n\right) = 0.$$

Now

$$3\mathbb{E}(S_n(Y_{n+1}^2 - 1)|\mathcal{F}_n) = 3S_n\mathbb{E}(Y_{n+1}^2 - 1) \quad \text{and} \quad 3\mathbb{E}(S_n^2 Y_{n+1}|\mathcal{F}_n) = 3S_n^2\mathbb{E}(Y_{n+1})$$

by "taking out what is known," and using the fact that Y_{n+1} and $\mathcal{F}_n$ are independent. Furthermore,

$$3(n+1)\mathbb{E}(Y_{n+1}|\mathcal{F}_n) = 3(n+1)\mathbb{E}(Y_{n+1}) \quad \text{and} \quad \mathbb{E}(Y_{n+1}^3|\mathcal{F}_n) = \mathbb{E}(Y_{n+1}^3)$$

using the fact that Y_{n+1} and $\mathcal{F}_n$ are independent. Since $\mathbb{E}(Y_{n+1}) = 0$, $\mathbb{E}(Y_{n+1}^2) = 1$, and $\mathbb{E}(Y_{n+1}^3) = 0$, we see that

$$\begin{aligned}\mathbb{E}(M_{n+1}|\mathcal{F}_n) &= M_n + 3S_n\mathbb{E}(Y_{n+1}^2 - 1) + 3S_n^2\mathbb{E}(Y_{n+1}) - 3(n+1)\mathbb{E}(Y_{n+1}) + \mathbb{E}(Y_{n+1}^3)\\ &= M_n + 3S_n \cdot (1-1) + 3S_n^2 \cdot 0 - 3(n+1)\cdot 0 + 0\\ &= M_n\end{aligned}$$

which proves that $\{M_n, n = 0, 1, 2, \ldots\}$ is, in fact, a martingale.

The following example is the most important discrete-time martingale calculation that you will do. The process $\{I_j, j = 0, 1, 2, \ldots\}$ defined below is an example of a discrete stochastic integral. In fact, stochastic integration is one of the greatest achievements of 20th century probability and, as we will see, is fundamental to the mathematical theory of finance and option pricing.

Example 4.9. As in the previous example, let $Y_1, Y_2, \ldots$ be independent and identically distributed random variables with $\mathbf{P}\{Y_1 = 1\} = \mathbf{P}\{Y_1 = -1\} = \frac{1}{2}$, set $S_0 = 0$, and for $n = 1, 2, 3, \ldots$, define the random variable S_n by $S_n = Y_1 + \cdots + Y_n$ so that $\{S_n, n = 0, 1, 2, \ldots\}$ is a simple random walk starting at 0. Now suppose that $I_0 = 0$ and for $j = 1, 2, \ldots$ define I_j to be

$$I_j = \sum_{n=1}^{j} S_{n-1}(S_n - S_{n-1}).$$

Prove that $\{I_j, j = 0, 1, 2, \ldots\}$ is a martingale.

Solution. If

$$I_j = \sum_{n=1}^{j} S_{n-1}(S_n - S_{n-1}).$$

then

$$I_{j+1} = I_j + S_j(S_{j+1} - S_j).$$

Therefore,

$$\begin{aligned}\mathbb{E}(I_{j+1}|\mathcal{F}_j) = \mathbb{E}(I_j + S_j(S_{j+1} - S_j)|\mathcal{F}_j) &= \mathbb{E}(I_j|\mathcal{F}_j) + \mathbb{E}(S_j(S_{j+1} - S_j)|\mathcal{F}_j)\\ &= I_j + S_j\mathbb{E}(S_{j+1}|\mathcal{F}_j) - S_j^2\end{aligned}$$

where we have "taken out what is known" three times. Furthermore, since $\{S_j, j = 0, 1, \ldots\}$ is a martingale,

$$\mathbb{E}(S_{j+1}|\mathcal{F}_j) = S_j.$$

Combining everything gives

$$\mathbb{E}(I_{j+1}|\mathcal{F}_j) = I_j + S_j\mathbb{E}(S_{j+1}|\mathcal{F}_j) - S_j^2 = I_j + S_j^2 - S_j^2 = I_j$$

which proves that $\{I_j, j = 0, 1, 2, \ldots\}$ is, in fact, a martingale.

Exercise 4.10. Suppose that $\{I_j, j = 0, 1, 2, \ldots\}$ is defined as in the previous example. Show that

$$\text{Var}(I_j) = \frac{j(j-1)}{2}$$

for all $j = 0, 1, 2, \ldots$.

This next example gives several martingales derived from *biased random walk*.

Example 4.11. Suppose that $Y_1, Y_2, \ldots$ are independent and identically distributed random variables with $\mathbf{P}\{Y_1 = 1\} = p$, $\mathbf{P}\{Y_1 = -1\} = 1-p$ for some $0 < p < 1/2$. Let $S_n = Y_1 + \cdots + Y_n$ denote their partial sums so that $\{S_n, n = 0, 1, 2, \ldots\}$ is a *biased random walk*. (Note that $\{S_n, n = 0, 1, 2, \ldots\}$ is no longer a *simple* random walk.)

(a) Show that $X_n = S_n - n(2p-1)$ is a martingale.
(b) Show that $M_n = X_n^2 - 4np(1-p) = [S_n - n(2p-1)]^2 - 4np(1-p)$ is a martingale.
(c) Show that $Z_n = \left(\frac{1-p}{p}\right)^{S_n}$ is a martingale.

Solution. We begin by noting that

$$\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n) = \sigma(S_0, \ldots, S_n) = \sigma(X_0, \ldots, X_n) = \sigma(M_0, \ldots, M_n) = \sigma(Z_0, \ldots, Z_n).$$

(a) The first step is to calculate $\mathbb{E}(Y_1)$. That is,

$$\mathbb{E}(Y_1) = 1 \cdot \mathbf{P}\{Y = 1\} + (-1) \cdot \mathbf{P}\{Y = -1\} = p - (1-p) = 2p - 1.$$

Since $S_{n+1} = S_n + Y_{n+1}$, we see that

$$\begin{aligned}\mathbb{E}(S_{n+1}|\mathcal{F}_n) = \mathbb{E}(S_n + Y_{n+1}|\mathcal{F}_n) &= \mathbb{E}(S_n|\mathcal{F}_n) + \mathbb{E}(Y_{n+1}|\mathcal{F}_n)\\ &= S_n + \mathbb{E}(Y_{n+1})\\ &= S_n + 2p - 1\end{aligned}$$

by "taking out what is known" and using the fact that Y_{n+1} and $\mathcal{F}_n$ are independent. This implies that

$$\begin{aligned}\mathbb{E}(X_{n+1}|\mathcal{F}_n) = \mathbb{E}(S_{n+1} - (n+1)(2p-1)|\mathcal{F}_n) &= \mathbb{E}(S_{n+1}|\mathcal{F}_n) - (n+1)(2p-1)\\ &= S_n + 2p - 1 - (n+1)(2p-1)\\ &= S_n - n(2p-1)\\ &= X_n,\end{aligned}$$

and so we conclude that $\{X_n, n = 1, 2, \ldots\}$ is, in fact, a martingale.

(b) Notice that we can write X_{n+1} as

$$\begin{aligned} X_{n+1} = S_{n+1} - (n+1)(2p-1) &= S_n + Y_{n+1} - n(2p-1) - (2p-1) \\ &= X_n + Y_{n+1} - (2p-1) \end{aligned}$$

and so

$$\begin{aligned} X_{n+1}^2 &= (X_n + Y_{n+1})^2 + (2p-1)^2 - 2(2p-1)(X_n + Y_{n+1}) \\ &= X_n^2 + Y_{n+1}^2 + 2X_nY_{n+1} + (2p-1)^2 - 2(2p-1)(X_n + Y_{n+1}). \end{aligned}$$

Thus,

$$\begin{aligned} &\mathbb{E}(X_{n+1}^2|\mathcal{F}_n) \\ &\quad = \mathbb{E}(X_n^2|\mathcal{F}_n) + \mathbb{E}(Y_{n+1}^2|\mathcal{F}_n) + 2\mathbb{E}(X_nY_{n+1}|\mathcal{F}_n) + (2p-1)^2 - 2(2p-1)\mathbb{E}(X_n + Y_{n+1}|\mathcal{F}_n) \\ &\quad = X_n^2 + \mathbb{E}(Y_{n+1})^2 + 2X_n\mathbb{E}(Y_{n+1}) + (2p-1)^2 - 2(2p-1)(X_n + \mathbb{E}(Y_{n+1})) \\ &\quad = X_n^2 + 1 + 2(2p-1)X_n + (2p-1)^2 - 2(2p-1)(X_n + (2p-1)) \\ &\quad = X_n^2 + 1 + 2(2p-1)X_n + (2p-1)^2 - 2(2p-1)X_n - 2(2p-1)^2 \\ &\quad = X_n^2 + 1 - (2p-1)^2, \end{aligned}$$

by again "taking out what is known" and using the fact that Y_{n+1} and $\mathcal{F}_n$ are independent. Hence, we find

$$\begin{aligned} \mathbb{E}(M_{n+1}|\mathcal{F}_n) &= \mathbb{E}(X_{n+1}^2|\mathcal{F}_n) - 4(n+1)p(1-p) \\ &= X_n^2 + 1 - (2p-1)^2 - 4(n+1)p(1-p) \\ &= X_n^2 + 1 - (4p^2 - 4p + 1) - 4np(1-p) - 4p(1-p) \\ &= X_n^2 + 1 - 4p^2 + 4p - 1 - 4np(1-p) - 4p + 4p^2 \\ &= X_n^2 - 4np(1-p) \\ &= M_n \end{aligned}$$

so that $\{M_n, n = 1, 2, \ldots\}$ is, in fact, a martingale.

(c) Notice that

$$Z_{n+1} = \left(\frac{1-p}{p}\right)^{S_{n+1}} = \left(\frac{1-p}{p}\right)^{S_n + Y_{n+1}} = \left(\frac{1-p}{p}\right)^{S_n} \left(\frac{1-p}{p}\right)^{Y_{n+1}} = Z_n \left(\frac{1-p}{p}\right)^{Y_{n+1}}.$$

Therefore,

$$\begin{aligned} \mathbb{E}(Z_{n+1}|\mathcal{F}_n) = \mathbb{E}\left(Z_n \left(\frac{1-p}{p}\right)^{Y_{n+1}} \middle| \mathcal{F}_n\right) &= Z_n\mathbb{E}\left(\left(\frac{1-p}{p}\right)^{Y_{n+1}} \middle| \mathcal{F}_n\right) \\ &= Z_n\mathbb{E}\left(\left(\frac{1-p}{p}\right)^{Y_{n+1}}\right) \end{aligned}$$

where the second equality follows from "taking out what is known" and the third equality follows from the fact that Y_{n+1} and $\mathcal{F}_n$ are independent. We now compute

$$\mathbb{E}\left(\left(\frac{1-p}{p}\right)^{Y_{n+1}}\right) = p\left(\frac{1-p}{p}\right)^1 + (1-p)\left(\frac{1-p}{p}\right)^{-1} = (1-p) + p = 1$$

and so we conclude

$$\mathbb{E}(Z_{n+1}|\mathcal{F}_n) = Z_n.$$

Hence, $\{Z_n, n = 0, 1, 2, \ldots\}$ is, in fact, a martingale.

We now conclude this section with one final example. Although it is unrelated to simple random walk, it is an easy martingale calculation and is therefore worth including. In fact, it could be considered as a generalization of (c) of the previous example.

Example 4.12. Suppose that $Y_1, Y_2, \ldots$ are independent and identically distributed random variables with $\mathbb{E}(Y_1) = 1$. Suppose further that $X_0 = Y_0 = 1$ and for $n = 1, 2, \ldots$, let

$$X_n = Y_1 \cdot Y_2 \cdots Y_n = \prod_{j=1}^{n} Y_j.$$

Verify that $\{X_n, n = 0, 1, 2, \ldots\}$ is a martingale with respect to $\{\mathcal{F}_n = \sigma(Y_0, \ldots, Y_n), n = 0, 1, 2, \ldots\}$.

Solution. We find

$$\begin{aligned}
\mathbb{E}(X_{n+1}|\mathcal{F}_n) &= \mathbb{E}(X_n \cdot Y_{n+1}|\mathcal{F}_n) \\
&= X_n \mathbb{E}(Y_{n+1}|\mathcal{F}_n) \quad \text{(by taking out what is known)} \\
&= X_n \mathbb{E}(Y_{n+1}) \quad \text{(since } Y_{n+1} \text{ is independent of } \mathcal{F}_n) \\
&= X_n \cdot 1 \\
&= X_n
\end{aligned}$$

and so $\{X_n, n = 0, 1, 2, \ldots\}$ is, in fact, a martingale.

5

Continuous-Time Martingales

Let $\{X_t, t \geq 0\}$ be a continuous-time stochastic process. Recall that this implies that there are uncountably many random variables, one for each value of the time index t.

For $t \geq 0$, let $\mathcal{F}_t$ denote the information contained in the process up to (and including) time t. Formally, let

$$\mathcal{F}_t = \sigma(X_s, 0 \leq s \leq t).$$

We call $\{\mathcal{F}_t, t \geq 0\}$ a *filtration*, and we say that X_t is *adapted* if $X_t \in \mathcal{F}_t$. Notice that if $s \leq t$, then $\mathcal{F}_s \subseteq \mathcal{F}_t$ so that $X_s \in \mathcal{F}_t$ as well.

The definition of a continuous-time martingale is analogous to the definition in discrete time.

Definition 5.1. A collection $\{X_t, t \geq 0\}$ of random variables is said to be a *martingale* with respect to the filtration $\{\mathcal{F}_t, t \geq 0\}$ if

(i) $X_t \in \mathcal{F}_t$ for every $t \geq 0$,
(ii) $\mathbb{E}|X_t| < \infty$ for every $t \geq 0$, and
(iii) $\mathbb{E}(X_t|\mathcal{F}_s) = X_s$ for every $0 \leq s < t$.

Note that in the third part of the definition, the present time t must be strictly larger than the past time s. (This is clearer in discrete time since the present time $n+1$ is always strictly larger than the past time n.)

The theorem from discrete time about independence and "taking out what is known" is also true in continuous time.

Theorem 5.2. *Let $\{X_t, t \geq 0\}$ be a stochastic process and consider the filtration $\{\mathcal{F}_t, t \geq 0\}$ where $\mathcal{F}_t = \sigma(X_s, 0 \leq s \leq t)$. Let Y be a random variable, and let $g : \mathbb{R}^n \to \mathbb{R}$ be a function. Suppose that $0 \leq t_1 < t_2 < \cdots < t_n$ are n times, and let s be such that $0 \leq s < t_1$. (Note that if $t_1 = 0$, then $s = 0$.) It then follows that*

(a) $\mathbb{E}(g(X_{t_1}, \ldots, X_{t_n})\, Y|\mathcal{F}_s) = g(X_{t_1}, \ldots, X_{t_n})\mathbb{E}(Y|\mathcal{F}_s)$ *(taking out what is known),*
(b) $\mathbb{E}(Y|\mathcal{F}_s) = \mathbb{E}(Y)$ *if Y is independent of $\mathcal{F}_s$, and*
(c) $\mathbb{E}(\mathbb{E}(Y|\mathcal{F}_s)) = \mathbb{E}(Y)$.

As in the discrete case, continuous-time martingales have stable expectation.

Theorem 5.3. *If $\{X_t, t \geq 0\}$ is a martingale, then $\mathbb{E}(X_t) = \mathbb{E}(X_0)$ for every $t \geq 0$.*

Proof. Since

$$\mathbb{E}(X_t) = \mathbb{E}(\mathbb{E}(X_t|\mathcal{F}_s)) = \mathbb{E}(X_s)$$

for any $0 \leq s < t$, we can simply choose $s = 0$ to complete the proof. □

You are already familiar with one example of a continuous-time stochastic process, namely the Poisson process. This will lead us to our first continuous-time martingale.

Example 5.4. As in STAT 351, the *Poisson process with intensity* λ is a continuous-time stochastic process $\{X_t, t \geq 0\}$ satisfying the following properties.

(i) The increments $\{X_{t_k} - X_{t_{k-1}}, k = 1, \dots, n\}$ are independent for all $0 \leq t_0 < \cdots < t_n < \infty$ and all n,
(ii) $X_0 = 0$, and
(iii) there exists a $\lambda > 0$ such that

$$X_t - X_s \in \text{Po}(\lambda(t-s))$$

for $0 \leq s < t$.

Consider the filtration $\{\mathcal{F}_t, t \geq 0\}$ where $\mathcal{F}_t = \sigma(X_s, 0 \leq s \leq t)$. In order to show that $\{X_t, t \geq 0\}$ is a martingale, we must verify that

$$\mathbb{E}(X_t|\mathcal{F}_s) = X_s$$

for every $0 \leq s < t$. The trick, much like for simple random walk in the discrete case, is to add-and-subtract the correct thing. Notice that $X_t = X_t - X_s + X_s$ so that

$$\mathbb{E}(X_t|\mathcal{F}_s) = \mathbb{E}(X_t - X_s + X_s|\mathcal{F}_s) = \mathbb{E}(X_t - X_s|\mathcal{F}_s) + \mathbb{E}(X_s|\mathcal{F}_s).$$

By assumption, $X_t - X_s$ is independent of $\mathcal{F}_s$ so that

$$\mathbb{E}(X_t - X_s|\mathcal{F}_s) = \mathbb{E}(X_t - X_s) = \lambda(t-s)$$

since $X_t - X_s \in \text{Po}(\lambda(t-s))$. Furthermore, since X_s is "known" at time s we have

$$\mathbb{E}(X_s|\mathcal{F}_s) = X_s.$$

Combined, this shows that

$$\mathbb{E}(X_t|\mathcal{F}_s) = X_s + \lambda(t-s) = \lambda t + X_s - \lambda s.$$

In other words, $\{X_t, t \geq 0\}$ is NOT a martingale. However, if we consider $\{X_t - \lambda t, t \geq 0\}$ instead, then this IS a martingale since

$$\mathbb{E}(X_t - \lambda t|\mathcal{F}_s) = X_s - \lambda s.$$

The process $\{N_t, t \geq 0\}$ given by $N_t = X_t - \lambda t$ is sometimes called the *compensated Poisson process with intensity* λ. (In other words, the compensated Poisson process is what you need to *compensate* the Poisson process by in order to have a martingale!)

Remark. In some sense, this result is like the biased random walk. If $S_0 = 0$ and $S_n = Y_1 + \cdots + Y_n$ where $\mathbf{P}\{Y_1 = 1\} = 1 - \mathbf{P}\{Y_1 = -1\} = p$, $0 < p < 1/2$, then $\mathbb{E}(S_n) = (2p-1)n$. Hence, S_n does NOT have stable expectation so that $\{S_n, n = 0, 1, 2, \ldots\}$ cannot be a martingale. However, if we consider $\{S_n - (2p-1)n, n = 0, 1, \ldots\}$ instead, then this is a martingale. Similarly, since X_t has mean $\mathbb{E}(X_t) = \lambda t$ which depends on t (and is therefore not stable), it is not possible for $\{X_t, t \geq 0\}$ to be a martingale. By subtracting this mean we get $\{X_t - \lambda t, t \geq 0\}$ which is a martingale.

Remark. Do not let this previous remark fool you into thinking you can always take a stochastic process and subtract the mean to get a martingale. This is NOT TRUE. The previous remark is meant to simply provide some intuition. There is no substitute for checking the definition of martingale as is shown in the next example.

Example 5.5. Suppose that the distribution of the random variable X_0 is

$$\mathbf{P}\{X_0 = 2\} = \mathbf{P}\{X_0 = 0\} = \frac{1}{2}$$

so that $\mathbb{E}(X_0) = 1$. For $n = 1, 2, 3, \ldots$ define the random variable X_n by setting

$$X_n = nX_{n-1}.$$

Now consider the stochastic process $\{X_n, n = 0, 1, 2, \ldots\}$. The claim is that the process $\{M_n, n = 0, 1, 2, \ldots\}$ defined by setting

$$M_n = X_n - \mathbb{E}(X_n)$$

is NOT a martingale. Notice that

$$\mathbb{E}(X_n) = n\mathbb{E}(X_{n-1})$$

which implies that (just iterate) $\mathbb{E}(X_n) = n!$. Furthermore,

$$\mathbb{E}(X_n|\mathcal{F}_{n-1}) = \mathbb{E}(nX_{n-1}|\mathcal{F}_{n-1}) = nX_{n-1}.$$

Now, if we consider $M_n = X_n - \mathbb{E}(X_n) = X_n - n!$, then

$$\mathbb{E}(M_n|\mathcal{F}_{n-1}) = \mathbb{E}(X_n|\mathcal{F}_{n-1}) - n! = nX_{n-1} - n! = n[X_{n-1} - (n-1)!] = nM_{n-1}.$$

This shows that $\{M_n, n = 0, 1, 2, \ldots\}$ is NOT a martingale.

Exercise 5.6. Suppose that $\{N_t, t \geq 0\}$ is a compensated Poisson process with intensity λ. Let $0 \leq s < t$. Show that the moment generating function of the random variable $N_t - N_s$ is

$$m_{N_t - N_s}(\theta) = \mathbb{E}[\, e^{\theta(N_t - N_s)}\,] = \exp\left\{\lambda(t-s)(e^\theta - 1 - \theta)\right\}.$$

Conclude that

$$\mathbb{E}(N_t - N_s) = 0, \quad \mathbb{E}[\,(N_t - N_s)^2\,] = \lambda(t-s), \quad \mathbb{E}[\,(N_t - N_s)^3\,] = \lambda(t-s),$$

and

$$\mathbb{E}[\,(N_t - N_s)^4\,] = \lambda(t-s) + 3\lambda^2(t-s)^2.$$

Exercise 5.7. Suppose that $\{N_t, t \geq 0\}$ is a compensated Poisson process with intensity λ. Define the process $\{M_t, t \geq 0\}$ by setting $M_t = N_t^2 - \lambda t$. Show that $\{M_t, t \geq 0\}$ is a martingale with respect to the filtration $\{\mathcal{F}_t, t \geq 0\} = \{\sigma(N_s, 0 \leq s \leq t), t \geq 0\}$.

We are shortly going to learn about Brownian motion, the most important of all stochastic processes. Brownian motion will lead us to many, many more examples of martingales. (In fact, there is a remarkable theorem which tells us that any continuous-time martingale with continuous paths must be Brownian motion in disguise!)

In particular, for a simple random walk $\{S_n, n = 0, 1, 2, \ldots\}$, we have seen that

(i) $\{S_n, n = 0, 1, 2, \ldots\}$ is a martingale,
(ii) $\{M_n, n = 0, 1, 2, \ldots\}$ where $M_n = S_n^2 - n$ is a martingale, and
(iii) $\{I_j, j = 0, 1, 2, \ldots\}$ where

$$I_j = \sum_{n=1}^{j} S_{n-1}(S_n - S_{n-1}) \tag{5.1}$$

is a martingale.

As we will soon see, there are natural Brownian motion analogues of each of these martingales, particularly the stochastic integral (5.1).

6

Brownian Motion as a Model of a Fair Game

Suppose that we are interested in setting up a model of a fair game, and that we are going to place bets on the outcomes of the individual rounds of this game. If we assume that a round takes place at discrete times, say at times $1, 2, 3, \ldots$, and that the game pays even money on unit stakes per round, then a reasonable probability model for encoding the outcome of the jth game is via a sequence $\{X_j, j = 1, 2, \ldots\}$ of independent and identically distributed random variables with

$$\mathbf{P}\{X_1 = 1\} = \mathbf{P}\{X_1 = -1\} = \frac{1}{2}.$$

That is, we can view X_j as the outcome of the jth round of this fair game. Although we will assume that there is no game played at time 0, it will be necessary for our notation to consider what "happens" at time 0; therefore, we will simply define $X_0 = 0$.

Notice that the sequence $\{X_j, j = 1, 2, \ldots\}$ tracks the outcomes of the individual games. We would also like to track our net number of "wins"; that is, we care about

$$\sum_{j=1}^{n} X_j,$$

the net number of "wins" after n rounds. (If this sum is negative, we realize that a negative number of "wins" is an interpretation of a net "loss.") Hence, we define the process $\{S_n, n = 0, 1, 2, \ldots\}$ by setting

$$S_n = \sum_{j=0}^{n} X_j.$$

Of course, we know that $\{S_n, n = 0, 1, 2, \ldots\}$ is called a *simple random walk*, and so we use a simple random walk as our model of a fair game being played in discrete time.

If we write $\mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n)$ to denote the *information* contained in the first n rounds of this game, then we showed in Lecture #4 that $\{S_n, n = 0, 1, 2, \ldots\}$ is a martingale with respect to the filtration $\{\mathcal{F}_n, n = 0, 1, 2, \ldots\}$.

Notice that $S_j - S_{j-1} = X_j$ and so the *increment* $S_j - S_{j-1}$ is exactly the outcome of the jth round of this fair game.

Suppose that we bet on the outcome of the jth round of this game and that (as assumed above) the game pays even money on unit stakes; for example, if we flip a fair coin betting \$5 on "heads" and "heads" does, in fact, appear, then we win \$5 plus our original \$5, but if "tails" appears, then we lose our original \$5.

If we denote our betting strategy by Y_{j-1}, $j = 1, 2, \ldots$, so that Y_{j-1} represents the bet we make on the jth round of the game, then I_n, our *fortune* after n rounds, is given by

$$I_n = \sum_{j=1}^{n} Y_{j-1}(S_j - S_{j-1}). \tag{6.1}$$

We also *define* $I_0 = 0$. The process $\{I_n, n = 0, 1, 2, \ldots\}$ is called a *discrete stochastic integral* (or the *martingale transform of* Y *by* S).

Remark. If we choose unit bets each round so that $Y_{j-1} = 1$, $j = 1, 2, \ldots$, then

$$I_n = \sum_{j=1}^{n} (S_j - S_{j-1}) = S_n$$

and so our "fortune" after n rounds is simply the position of the random walk S_n. We are interested in what happens when Y_{j-1} is *not* constant in time, but rather varies with j.

Note that it is reasonable to assume that the bet you make on the jth round can only depend on the outcomes of the previous $j - 1$ rounds. That is, you cannot "look into the future and make your bet on the jth round based on what the outcome of the jth round will be." In mathematical language, we say that Y_{j-1} must be *previsible* (also called *predictable*).

Remark. The concept of a *previsible stochastic process* was intensely studied in the 1950s by the French school of probability that included P. Lévy. Since the French word *prévisible* is translated into English as *foreseeable*, there is no consistent English translation. Most probabilists use previsible and predictable interchangeably. (Although, unfortunately, not all do!)

A slight modification of Example 4.9 shows that $\{I_n, n = 0, 1, 2, \ldots\}$ is a martingale with respect to the filtration $\{\mathcal{F}_n, n = 0, 1, \ldots\}$. Note that the requirement that Y_{j-1} be previsible is exactly the requirement that allows $\{I_n, n = 0, 1, 2, \ldots\}$ to be a martingale.

It now follows from Theorem 4.4 that $\mathbb{E}(I_n) = 0$ for all n since $\{I_n, n = 0, 1, 2, \ldots\}$ is a martingale with $I_0 = 0$. As we saw in Exercise 4.10, calculating the variance of the random variable I_n is more involved. The following exercise generalizes that result and shows precisely how the variance depends on the choice of the sequence Y_{j-1}, $j = 1, 2, \ldots$.

Exercise 6.1. Consider the martingale transform of Y by S given by (6.1). Show that

$$\mathrm{Var}(I_n) = \sum_{j=1}^{n} \mathbb{E}(Y_{j-1}^2).$$

Suppose that instead of playing a round of the game at times $1, 2, 3, \ldots$, we play rounds more frequently, say at times $0.5, 1, 1.5, 2, 2.5, 3, \ldots$, or even more frequently still. In fact, we can imagine playing a round of the game at *every* time $t \geq 0$.

If this is hard to visualize, imagine the round of the game as being the price of a (fair) stock at time t. The stock is assumed, equally likely, to move an infinitesmal amount up or an infinitesmal amount down in every infinitesmal period of time.

Hence, if we want to model a fair game occurring in continuous time, then we need to find a *continuous limit* of the simple random walk. This continuous limit is Brownian motion, also called the *scaling limit* of simple random walk. To explain what this means, suppose that $\{S_n, n = 0, 1, 2, \ldots\}$ is a simple random walk. For $N = 1, 2, 3, \ldots$, define the scaled random walk $B_t^{(N)}$, $0 \le t \le 1$, to be the continuous process on the time interval $[0, 1]$ whose value at the fractional times $0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}, 1$ is given by setting

$$B_{\frac{j}{N}}^{(N)} = \frac{1}{\sqrt{N}} S_j, \quad j = 0, 1, 2, \ldots, N,$$

and for other times is defined by linear interpolation. As $N \to \infty$, the distribution of the process $\{B_t^{(N)}, 0 \le t \le 1\}$ converges to the distribution of a process $\{B_t, 0 \le t \le 1\}$ satisfying the following properties:

(i) $B_0 = 0$,
(ii) for any $0 \le s \le t \le 1$, the random variable $B_t - B_s$ is normally distributed with mean 0 and variance $t - s$; that is, $B_t - B_s \sim \mathcal{N}(0, t - s)$,
(iii) for any integer k and any partition $0 \le t_1 \le t_2 \le \cdots \le t_k \le 1$, the random variables $B_{t_k} - B_{t_{k-1}}, \ldots, B_{t_2} - B_{t_1}, B_{t_1}$ are independent, and
(iv) the trajectory $t \mapsto B_t$ is continuous.

By piecing together independent copies of this process, we can construct a Brownian motion $\{B_t, t \ge 0\}$ defined for all times $t \ge 0$ satisfying the above properties (without, of course, the restriction in (b) that $t \le 1$ and the restriction in (c) that $t_k \le 1$). Thus, we now suppose that $\{B_t, t \ge 0\}$ is a Brownian motion with $B_0 = 0$.

Exercise 6.2. Deduce from the definition of Brownian motion that for each $t > 0$, the random variable B_t is normally distributed with mean 0 and variance t. Why does this imply that $\mathbb{E}(B_t^2) = t$?

Exercise 6.3. Deduce from the definition of Brownian motion that for $0 \le s < t$, the distribution of the random variable $B_t - B_s$ is the *same* as the distribution of the random variable B_{t-s}.

Exercise 6.4. Show that if $\{B_t, t \ge 0\}$ is a Brownian motion, then $\mathbb{E}(B_t) = 0$ for all t, and $\text{Cov}(B_s, B_t) = \min\{s, t\}$. *Hint: Suppose that $s < t$ and write $B_sB_t = (B_sB_t - B_s^2) + B_s^2$.* The result of this exercise actually shows that Brownian motion is not a stationary process, although it does have stationary increments.

Note. One of the problems with using either simple random walk or Brownian motion as a model of an asset price is that the value of a real stock is never allowed to be negative—it can equal 0, but can never be strictly less than 0. On the other hand, both a random walk and a Brownian motion can be negative. Hence, neither serves as an adequate model for a stock. Nonetheless, Brownian motion is the key ingredient for building a reasonable model of a stock and the stochastic integral that we are about to construct is *fundamental* to the analysis. At this point, we must be content with modelling (and betting on) fair games whose values can be either positive or negative.

If we let $\mathcal{F}_t = \sigma(B_s, 0 \le s \le t)$ denote the "information" contained in the Brownian motion up to (and including) time t, then it easily follows that $\{B_t, t \ge 0\}$ is a continuous-time martingale with respect to the Brownian filtration $\{\mathcal{F}_t, t \ge 0\}$. That is, suppose that $s < t$, and so

$$\mathbb{E}(B_t|\mathcal{F}_s) = \mathbb{E}(B_t - B_s + B_s|\mathcal{F}_s) = \mathbb{E}(B_t - B_s|\mathcal{F}_s) + \mathbb{E}(B_s|\mathcal{F}_s) = \mathbb{E}(B_t - B_s) + B_s = B_s$$

since the Brownian increment $B_t - B_s$ has mean 0 and is independent of $\mathcal{F}_s$, and B_s is "known" at time s (using the "taking out what is known" property of conditional expectation).

In analogy with simple random walk, we see that although $\{B_t^2, t \ge 0\}$ is not a martingale with respect to $\{\mathcal{F}_t, t \ge 0\}$, the process $\{B_t^2 - t, t \ge 0\}$ *is* one.

Exercise 6.5. Let the process $\{M_t, t \ge 0\}$ be defined by setting $M_t = B_t^2 - t$. Show that $\{M_t, t \ge 0\}$ is a (continuous-time) martingale with respect to the Brownian filtration $\{\mathcal{F}_t, t \ge 0\}$.

Exercise 6.6. The same "trick" used to solve the previous exercise can also be used to show that both $\{B_t^3 - 3tB_t, t \ge 0\}$ and $\{B_t^4 - 6tB_t^2 + 3t^2, t \ge 0\}$ are martingales with respect to the Brownian filtration $\{\mathcal{F}_t, t \ge 0\}$. Verify that these are both, in fact, martingales. (Once we have learned Itô's formula, we will discover a much easier way to "generate" such martingales.)

Assuming that our fair game is modelled by a Brownian motion, we need to consider appropriate betting strategies. For now, we will allow only deterministic betting strategies that do not "look into the future" and denote such a strategy by $\{g(t), t \ge 0\}$. This notation might look a little strange, but it is meant to be suggestive for when we allow certain random betting strategies. Hence, at this point, our betting strategy is simply a real-valued function $g : [0, \infty) \to \mathbb{R}$. Shortly, for technical reasons, we will see that it is necessary for g to be at least bounded, piecewise continuous, and in $L^2([0, \infty))$. Recall that $g \in L^2([0, \infty))$ means that

$$\int_0^\infty g^2(s)\,\mathrm{d}s < \infty.$$

Thus, if we fix a time $t > 0$, then, in analogy with (6.1), our "fortune process" up to time t is given by the (yet-to-be-defined) *stochastic integral*

$$I_t = \int_0^t g(s)\,\mathrm{d}B_s. \tag{6.2}$$

Our goal, now, is to try and define (6.2) in a reasonable way. A natural approach, therefore, is to try and relate the stochastic integral (6.2) with the discrete stochastic integral (6.1) constructed earlier. Since the discrete stochastic integral resembles a Riemann sum, that seems like a good place to start.

7

Riemann Integration

Suppose that $g : [a, b] \to \mathbb{R}$ is a real-valued function on $[a, b]$. Fix a positive integer n, and let

$$\pi_n = \{a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b\}$$

be a *partition* of $[a, b]$. For $i = 1, \cdots, n$, define $\Delta t_i = t_i - t_{i-1}$ and let $t_i^* \in [t_{i-1}, t_i]$ be distinguished points; write $\tau_n^* = \{t_1^*, \ldots, t_n^*\}$ for the set of distinguished points. If π_n is a partition of $[a, b]$, define the *mesh* of π_n to be the width of the largest subinterval; that is,

$$\text{mesh}(\pi_n) = \max_{1 \le i \le n} \Delta t_i = \max_{1 \le i \le n} (t_i - t_{i-1}).$$

Finally, we call

$$S(g; \pi_n; \tau_n^*) = \sum_{i=1}^{n} g(t_i^*) \Delta t_i$$

the *Riemann sum for g corresponding to the partition π_n with distinguished points τ_n^*.*

We say that $\pi = \{\pi_n, n = 1, 2, \ldots\}$ is a *refinement* of $[a, b]$ if π is a sequence of partitions of $[a, b]$ with $\pi_n \subset \pi_{n+1}$ for all n.

Definition 7.1. We say that g is *Riemann integrable over* $[a, b]$ and define the *Riemann integral* of g to be I if for every $\varepsilon > 0$ and for every refinement $\pi = \{\pi_n, n = 1, 2, \ldots\}$ with $\text{mesh}(\pi_n) \to 0$ as $n \to \infty$, there exists an N such that

$$|S(g; \pi_m; \tau_m^*) - I| < \varepsilon$$

for all choices of distinguished points τ_m^* and for all $m \ge N$. We then define

$$\int_a^b g(s)\, \mathrm{d}s$$

to be this limiting value I.

Remark. There are various equivalent definitions of the Riemann integral including Darboux's version using upper and lower sums. The variant given in Definition 7.1 above will be the most useful one for our construction of the stochastic integral.

The following theorem gives a sufficient condition for a function to be Riemann integrable.

Theorem 7.2. *If $g : [a,b] \to \mathbb{R}$ is bounded and piecewise continuous, then g is Riemann integrable on $[a,b]$.*

Proof. For a proof, see Theorem 6.10 on page 126 of Rudin [21]. □

The previous theorem is adequate for our purposes. However, it is worth noting that, in fact, this theorem follows from a more general result which completely characterizes the class of Riemann integrable functions.

Theorem 7.3. *Suppose that $g : [a,b] \to \mathbb{R}$ is bounded. The function g is Riemann integrable on $[a,b]$ if and only if the set of discontinuities of g has Lebesgue measure 0.*

Proof. For a proof, see Theorem 11.33 on page 323 of Rudin [21]. □

There are two particular Riemann sums that are studied in elementary calculus—the so-called *left-hand Riemann sum* and *right-hand Riemann sum.*

For $i = 0, 1, \ldots, n$, let $t_i = a + \frac{i(b-a)}{n}$. If $t_i^* = t_{i-1}$, then

$$\frac{b-a}{n} \sum_{i=1}^{n} g\left(a + \frac{(i-1)(b-a)}{n}\right)$$

is called the *left-hand Riemann sum.* The *right-hand Riemann sum* is obtained by choosing $t_i^* = t_i$ and is given by

$$\frac{b-a}{n} \sum_{i=1}^{n} g\left(a + \frac{i(b-a)}{n}\right).$$

Remark. It is a technical matter that if $t_i = a + \frac{i(b-a)}{n}$, then $\pi = \{\pi_n, n = 1, 2, \ldots\}$ with $\pi_n = \{t_0 = a < t_1 < \cdots < t_{n-1} < t_n = b\}$ is not a refinement. To correct this, we simply restrict to those n of the form $n = 2^k$ for some k in order to have a refinement of $[a,b]$. Hence, from now on, we will not let this concern us.

The following example shows that even though the limits of the left-hand Riemann sums and the right-hand Riemann sums might both exist and be equal for a function g, that is not enough to guarantee that g is Riemann integrable.

Example 7.4. Suppose that $g : [0,1] \to \mathbb{R}$ is defined by

$$g(x) = \begin{cases} 0, & \text{if } x \in \mathbb{Q} \cap [0,1], \\ 1, & \text{if } x \notin \mathbb{Q} \cap [0,1]. \end{cases}$$

Let $\pi_n = \{0 < \frac{1}{n} < \frac{2}{n} < \cdots < \frac{n-1}{n} < 1\}$ so that $\Delta t_i = \frac{1}{n}$ and $\mathrm{mesh}(\pi_n) = \frac{1}{n}$. The limit of the left-hand Riemann sums is therefore given by

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g\left(\frac{i-1}{n}\right) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g(0) = 0$$

since $\frac{i-1}{n}$ is necessarily rational. Similarly, the limit of the right-hand Riemann sums is given by

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g\left(\frac{i}{n}\right) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g(0) = 0.$$

However, define a sequence of partitions as follows:

$$\pi_n = \left\{0 < \frac{1}{n\sqrt{2}} < \cdots < \frac{n-1}{n\sqrt{2}} < \frac{1}{\sqrt{2}} < \frac{1}{\sqrt{2}} + \frac{\sqrt{2}-1}{n\sqrt{2}} < \cdots < \frac{1}{\sqrt{2}} + \frac{(n-1)(\sqrt{2}-1)}{n\sqrt{2}} < 1\right\}.$$

In this case, $\operatorname{mesh}(\pi_n) = \frac{\sqrt{2}-1}{n\sqrt{2}}$ so that $\operatorname{mesh}(\pi_n) \to 0$ as $n \to \infty$. If t_i^* is chosen to be the mid-point of each interval, then t_i^* is necessarily irrational so that $g(t_i^*) = 1$. Therefore,

$$\sum_{i=1}^{n} g(t_i^*)\Delta t_i = \sum_{i=1}^{n} \Delta t_i = \sum_{i=1}^{n} (t_i - t_{i-1}) = t_n - t_0 = 1 - 0 = 1$$

for each n. Hence, we conclude that g is *not* Riemann integrable on $[0, 1]$ since there is no unique limiting value.

However, we can make the following postive assertion about the limits of the left-hand Riemann sums and the right-hand Riemann sums.

Remark. Suppose that $g : [a, b] \to \mathbb{R}$ is Riemann integrable on $[a, b]$ so that

$$I = \int_a^b g(s)\,\mathrm{d}s$$

exists. Then, the limit of the left-hand Riemann sums and the limit of the right-hand Riemann sums both exist, and furthermore

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g\left(\frac{i}{n}\right) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g\left(\frac{i-1}{n}\right) = I.$$

8

The Riemann Integral of Brownian Motion

Before *integrating with respect to* Brownian motion it seems reasonable to try and *integrate* Brownian motion itself. This will help us get a feel for some of the technicalities involved when the integrand/integrator in a stochastic process.

Suppose that $\{B_t, 0 \le t \le 1\}$ is a Brownian motion. Since Brownian motion is continuous with probability one, it follows from Theorem 7.2 that Brownian motion is Riemann integrable. Thus, at least theoretically, we can integrate Brownian motion, although it is not so clear what the Riemann integral of it is. To be a bit more precise, suppose that $B_t(\omega)$, $0 \le t \le 1$, is a realization of Brownian motion (a so-called *sample path* or *trajectory*) and let

$$I = \int_0^1 B_s(\omega)\,\mathrm{d}s$$

denote the Riemann integral of the function $B(\omega)$ on $[0,1]$. (By this notation, we mean that $B(\omega)$ is the function and $B(\omega)(t) = B_t(\omega)$ is the value of this function at time t. This is analogous to our notation in calculus in which g is the function and $g(t)$ is the value of this function at time t.)

Question. What can be said about I?

On the one hand, we know from elementary calculus that the Riemann integral represents the area under the curve, and so we at least have that interpretation of I. On the other hand, since Brownian motion is nowhere differentiable with probability one, there is no hope of using the fundamental theorem of calculus to evaluate I. Furthermore, since the value of I depends on the realization $B(\omega)$ observed, we should really be viewing I as a function of ω; that is,

$$I(\omega) = \int_0^1 B_s(\omega)\,\mathrm{d}s.$$

It is now clear that I is itself a random variable, and so the best that we can hope for in terms of "calculating" the Riemann integral I is to determine its distribution.

As noted above, the Riemann integral I necessarily exists by Theorem 7.2, which means that in order to determine its distribution, it is sufficient to determine the distribution of the limit of the right-hand sums

$$I = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} B_{i/n}.$$

(See the final remark of Lecture #7.) Therefore, we begin by calculating the distribution of

$$I^{(n)} = \frac{1}{n} \sum_{i=1}^{n} B_{i/n}. \tag{8.1}$$

We know that for each $i = 1, \ldots, n$, the distribution of $B_{i/n}$ is $\mathcal{N}(0, i/n)$. The problem, however, is that the sum in (8.1) is not a sum of independent random variables—only Brownian *increments* are independent. However, we can use a little algebraic trick to express this as the sum of independent increments. Notice that

$$\sum_{i=1}^{n} Y_i = nY_1 + (n-1)(Y_2 - Y_1) + (n-2)(Y_3 - Y_2) + \cdots + 2(Y_{n-1} - Y_{n-2}) + (Y_n - Y_{n-1}).$$

We now let $Y_i = B_{i/n}$ so that $Y_i \sim \mathcal{N}(0, i/n)$. Furthermore, $Y_i - Y_{i-1} \sim \mathcal{N}(0, 1/n)$, and the sum above is the sum of independent normal random variables, so it too is normal. Let $X_i = Y_i - Y_{i-1} \sim \mathcal{N}(0, 1/n)$ so that $X_1, X_2, \ldots, X_n$ are independent and

$$\sum_{i=1}^{n} Y_i = nX_1 + (n-1)X_2 + \cdots + 2X_{n-1} + X_n = \sum_{i=1}^{n} (n-i+1)X_i \sim \mathcal{N}\left(0, \frac{1}{n}\sum_{i=1}^{n}(n-i+1)^2\right)$$

by Exercise 3.12. Since

$$\sum_{i=1}^{n} (n-i+1)^2 = n^2 + (n-1)^2 + \cdots + 2^2 + 1 = \frac{n(n+1)(2n+1)}{6},$$

we see that

$$\sum_{i=1}^{n} Y_i \sim \mathcal{N}\left(0, \frac{(n+1)(2n+1)}{6}\right),$$

and so finally piecing everything together we have

$$I^{(n)} = \frac{1}{n}\sum_{i=1}^{n} B_{i/n} \sim \mathcal{N}\left(0, \frac{(n+1)(2n+1)}{6n^2}\right) = \mathcal{N}\left(0, \frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2}\right).$$

Hence, we now conclude that as $n \to \infty$, the variance of $I^{(n)}$ approaches $1/3$ so by Theorem 3.24, the distribution of I is

$$I \sim \mathcal{N}\left(0, \tfrac{1}{3}\right).$$

In summary, this result says that if we consider the area under a Brownian path up to time 1, then that (random) area is normally distributed with mean 0 and variance 1/3.

Weird.

Remark. Theorem 7.2 tells us that for any fixed $t > 0$ we can, in theory, "compute" (i.e., determine the distribution of) any Riemann integral of the form

$$\int_0^t h(B_s)\,\mathrm{d}s$$

where $h : \mathbb{R} \to \mathbb{R}$ is a continuous function. Unless h is relatively simple, however, it is not so straightforward to determine the resulting distribution. Exercise 10.5 outlines one case in which such a calculation is possible.

9

Wiener Integration

Having successfully determined the Riemann integral of Brownian motion, we will now learn how to integrate with respect to Brownian motion; that is, we will study the (yet-to-be-defined) stochastic integral

$$I_t = \int_0^t g(s)\,\mathrm{d}B_s.$$

Our experience with integrating Brownian motion suggests that I_t is really a random variable, and so one of our goals will be to determine the distribution of I_t.

Assume that g is bounded, piecewise continuous, and in $L^2([0,\infty))$, and suppose that we partition the interval $[0,t]$ by $0 = t_0 < t_1 < \cdots < t_n = t$. Consider the left-hand Riemann sum

$$\sum_{j=1}^{n} g(t_{j-1})(B_{t_j} - B_{t_{j-1}}).$$

Notice that our experience with the discrete stochastic integral suggests that we should choose a left-hand Riemann sum; that is, our discrete-time betting strategy Y_{j-1} needed to be previsible and so our continuous-time betting strategy $g(t)$ should also be previsible. When working with the Riemann sum, the previsible condition translates into taking the left-hand Riemann sum. We do, however, remark that when following a deterministic betting strategy, this previsible condition will turn out to not matter at all. On the other hand, when we follow a random betting strategy, it will be of the utmost importance.

To begin, let

$$I_t^{(n)} = \sum_{j=1}^{n} g(t_{j-1})(B_{t_j} - B_{t_{j-1}})$$

and notice that as in the discrete case, we can easily calculate $\mathbb{E}(I_t^{(n)})$ and $\mathrm{Var}(I_t^{(n)})$. Since $B_{t_j} - B_{t_{j-1}} \sim \mathcal{N}(0, t_j - t_{j-1})$, we have

$$\mathbb{E}(I_t^{(n)}) = \sum_{j=1}^{n} g(t_{j-1})\mathbb{E}(B_{t_j} - B_{t_{j-1}}) = 0,$$

and since the increments of Brownian motion are independent, we have

$$\mathrm{Var}(I_t^{(n)}) = \sum_{j=1}^{n} g^2(t_{j-1})\mathbb{E}(B_{t_j} - B_{t_{j-1}})^2 = \sum_{j=1}^{n} g^2(t_{j-1})(t_j - t_{j-1}).$$

We now make a crucial observation. The variance of $I_t^{(n)}$, namely

$$\sum_{j=1}^{n} g^2(t_{j-1})(t_j - t_{j-1}),$$

should look familiar. Since $0 = t_0 < t_1 < \cdots < t_n = t$ is a partition of $[0, t]$ we see that this sum is the left-hand Riemann sum approximating the Riemann integral

$$\int_0^t g^2(s)\,\mathrm{d}s.$$

We also see the reason to assume that g is bounded, piecewise continuous, and in $L^2([0,\infty))$. By Theorem 7.2, this condition is sufficient to guarantee that the limit

$$\lim_{n\to\infty} \sum_{j=1}^{n} g^2(t_{j-1})(t_j - t_{j-1})$$

exists and equals

$$\int_0^t g^2(s)\,\mathrm{d}s.$$

(Although by Theorem 7.3 it is possible to weaken the conditions on g, we will not concern ourselves with such matters.)

In summary, we conclude that

$$\lim_{n\to\infty} \mathbb{E}(I_t^{(n)}) = 0$$

and

$$\lim_{n\to\infty} \mathrm{Var}(I_t^{(n)}) = \int_0^t g^2(s)\,\mathrm{d}s.$$

Therefore, if we can somehow construct I_t as an appropriate limit of $I_t^{(n)}$, then it seems reasonable that $\mathbb{E}(I_t) = 0$ and

$$\mathrm{Var}(I_t) = \int_0^t g^2(s)\,\mathrm{d}s.$$

As in the previous section, however, examining the Riemann sum

$$I_t^{(n)} = \sum_{j=1}^{n} g(t_{j-1})(B_{t_j} - B_{t_{j-1}})$$

suggests that we can determine more than just the mean and variance of $I_t^{(n)}$. Since disjoint Brownian increments are independent and normally distributed, and since $I_t^{(n)}$ is a sum of disjoint Brownian increments, we conclude that $I_t^{(n)}$ is normally distributed. In fact, combined with our earlier calculations, we see from Exercise 3.12 that

$$I_t^{(n)} \sim \mathcal{N}\left(0, \sum_{j=1}^{n} g^2(t_{j-1})(t_j - t_{j-1})\right).$$

It now follows from Theorem 3.24 that $I_t^{(n)}$ converges in distribution to the random variable I_t where

$$I_t \sim \mathcal{N}\left(0, \int_0^t g^2(s)\,\mathrm{d}s\right)$$

since the limit in distribution of normal random variables whose means and variances converge must itself be normal. Hence, we define

$$\int_0^t g(s)\,\mathrm{d}B_s$$

to be this limit I_t so that

$$I_t = \int_0^t g(s)\,\mathrm{d}B_s \sim \mathcal{N}\left(0, \int_0^t g^2(s)\,\mathrm{d}s\right).$$

Definition 9.1. Suppose that $g : [0,\infty) \to \mathbb{R}$ is a bounded, piecewise continuous function in $L^2([0,\infty))$. The *Wiener integral* of g with respect to Brownian motion $\{B_t, t \geq 0\}$, written

$$\int_0^t g(s)\,\mathrm{d}B_s,$$

is a random variable which has a

$$\mathcal{N}\left(0, \int_0^t g^2(s)\,\mathrm{d}s\right)$$

distribution.

Remark. We have taken the approach of defining the Wiener integral in a distributional sense. It is possible, with a lot more technical machinery, to define it as the L^2 limit of a sequence of random variables. In the case of a random g, however, in order to the define the Itô integral of g with respect to Brownian motion, we will need to follow the L^2 approach. Furthermore, we will see that the Wiener integral is actually a special case of the Itô integral. Thus, it seems pedagogically more appropriate to define the Wiener integral in the distributional sense since this is a much simpler construction and, arguably, more intuitive.

10

Calculating Wiener Integrals

Now that we have defined the Wiener integral of a bounded, piecewise continuous deterministic function in $L^2([0,\infty))$ with respect to Brownian motion as a normal random variable, namely

$$\int_0^t g(s)\,\mathrm{d}B_s \sim \mathcal{N}\left(0, \int_0^t g^2(s)\,\mathrm{d}s\right),$$

it might seem like we are done. However, as our ultimate goal is to be able to integrate random functions with respect to Brownian motion, it seems useful to try and develop a *calculus* for Wiener integration. The key computational tool that we will develop is an integration-by-parts formula. But first we need to complete the following exercise.

Exercise 10.1. Verify that the Wiener integral is a linear operator. That is, show that if α, $\beta \in \mathbb{R}$ are constants, and g and h are bounded, piecewise continuous functions in $L^2([0,\infty))$, then

$$\int_0^t [\alpha g(s) + \beta h(s)]\,\mathrm{d}B_s = \alpha \int_0^t g(s)\,\mathrm{d}B_s + \beta \int_0^t h(s)\,\mathrm{d}B_s.$$

Theorem 10.2. *Let $g : [0,\infty) \to \mathbb{R}$ be a bounded, continuous function in $L^2([0,\infty))$. If g is differentiable with g' also bounded and continuous, then the integration-by-parts formula*

$$\int_0^t g(s)\,\mathrm{d}B_s = g(t)B_t - \int_0^t g'(s)B_s\,\mathrm{d}s$$

holds.

Remark. Since all three objects in the above expression are random variables, the equality is interpreted to mean that the distribution of the random variable on the left side and the distribution of the random variable on the right side are the same, namely

$$\mathcal{N}\left(0, \int_0^t g^2(s)\,\mathrm{d}s\right).$$

Also note that the second integral on the right side, namely

$$\int_0^t g'(s)B_s\,\mathrm{d}s, \tag{10.1}$$

is the Riemann integral of a function of Brownian motion. Using the notation of the final remark of Lecture #8, we have $h(B_s) = g'(s)B_s$. In Exercise 10.5 you will determine the distribution of (10.1).

Proof. We begin by writing

$$\sum_{j=1}^{n} g(t_{j-1})(B_{t_j} - B_{t_{j-1}}) = \sum_{j=1}^{n} g(t_{j-1})B_{t_j} - \sum_{j=1}^{n} g(t_{j-1})B_{t_{j-1}}. \tag{10.2}$$

Since g is differentiable, the mean value theorem implies that there exists some value $t_j^* \in [t_{j-1}, t_j]$ such that

$$g'(t_j^*)(t_j - t_{j-1}) = g(t_j) - g(t_{j-1}).$$

Substituting this for $g(t_{j-1})$ in the previous expression (10.2) gives

$$\begin{aligned}
\sum_{j=1}^{n} g(t_{j-1})B_{t_j} - \sum_{j=1}^{n} g(t_{j-1})B_{t_{j-1}} &= \sum_{j=1}^{n} g(t_j)B_{t_j} - \sum_{j=1}^{n} g'(t_j^*)(t_j - t_{j-1})B_{t_j} - \sum_{j=1}^{n} g(t_{j-1})B_{t_{j-1}} \\
&= \sum_{j=1}^{n} [g(t_j)B_{t_j} - g(t_{j-1})B_{t_{j-1}}] - \sum_{j=1}^{n} g'(t_j^*)(t_j - t_{j-1})B_{t_j} \\
&= g(t_n)B_{t_n} - g(t_0)B_{t_0} - \sum_{j=1}^{n} g'(t_j^*)B_{t_j}(t_j - t_{j-1}) \\
&= g(t)B_t - \sum_{j=1}^{n} g'(t_j^*)B_{t_j}(t_j - t_{j-1})
\end{aligned}$$

since $t_n = t$ and $t_0 = 0$. Notice that we have established an equality between random variables, namely that

$$\sum_{j=1}^{n} g(t_{j-1})(B_{t_j} - B_{t_{j-1}}) = g(t)B_t - \sum_{j=1}^{n} g'(t_j^*)B_{t_j}(t_j - t_{j-1}). \tag{10.3}$$

The proof will be completed if we can show that the distribution of the limiting random variable on the left-side of (10.3) and the distribution of the limiting random variable on the right-side of (10.3) are the same. Of course, we know that

$$\sum_{j=1}^{n} g(t_{j-1})(B_{t_j} - B_{t_{j-1}}) \to I_t = \int_0^t g(s)\, \mathrm{d}B_s \sim \mathcal{N}\left(0, \int_0^t g^2(s)\, \mathrm{d}s\right)$$

from our construction of the Wiener integral in Lecture #9. Thus, we conclude that

$$g(t)B_t - \sum_{j=1}^{n} g'(t_j^*)B_{t_j}(t_j - t_{j-1}) \to I_t \sim \mathcal{N}\left(0, \int_0^t g^2(s)\, \mathrm{d}s\right)$$

in distribution as well. We now observe that since g' is bounded and piecewise continuous, and since Brownian motion is continuous, the function $g'(t)B_t$ is necessarily Riemann integrable. Thus,

$$\lim_{n\to\infty} \sum_{j=1}^{n} g'(t_j^*)B_{t_j}(t_j - t_{j-1}) = \int_0^t g'(s)B_s\, \mathrm{d}s$$

in distribution as in Lecture #8. In other words, we have shown that the distribution of

$$\int_0^t g(s)\,\mathrm{d}B_s$$

and the distribution of

$$g(t)B_t - \int_0^t g'(s)B_s\,\mathrm{d}s$$

are the same, namely

$$\mathcal{N}\left(0, \int_0^t g^2(s)\,\mathrm{d}s\right)$$

and so the proof is complete. □

Example 10.3. Suppose that $t > 0$. It might seem obvious that

$$B_t = \int_0^t \mathrm{d}B_s.$$

However, since Brownian motion is nowhere differentiable, and since we have only defined the Wiener integral as a normal random variable, this equality needs a proof. Since $B_t \sim \mathcal{N}(0, t)$ and since

$$\int_0^t \mathrm{d}B_s \sim \mathcal{N}\left(0, \int_0^t 1^2\,\mathrm{d}s\right) = \mathcal{N}(0, t),$$

we conclude that

$$B_t = \int_0^t \mathrm{d}B_s$$

in distribution. Alternatively, let $g \equiv 1$ so that the integration-by-parts formula implies

$$\int_0^t \mathrm{d}B_s = g(t)B_t - \int_0^t g'(s)B_s\,\mathrm{d}s = B_t - 0 = B_t.$$

Example 10.4. Suppose that we choose $t = 1$ and $g(s) = s$. The integration-by-parts formula implies that

$$\int_0^1 s\,\mathrm{d}B_s = B_1 - \int_0^1 B_s\,\mathrm{d}s.$$

If we now write

$$B_1 = \int_0^1 \mathrm{d}B_s$$

and use linearity of the stochastic integral, then we find

$$\int_0^1 B_s\,\mathrm{d}s = B_1 - \int_0^1 s\,\mathrm{d}B_s = \int_0^1 \mathrm{d}B_s - \int_0^1 s\,\mathrm{d}B_s = \int_0^1 (1-s)\,\mathrm{d}B_s.$$

Since

$$\int_0^1 (1-s)\,\mathrm{d}B_s$$

is normally distributed with mean 0 and variance

$$\int_0^1 (1-s)^2 \, \mathrm{d}s = \frac{1}{3},$$

we conclude that

$$\int_0^1 B_s \, \mathrm{d}s \sim \mathcal{N}(0, 1/3).$$

Thus, we have a different derivation of the fact that we proved in Lecture #8.

Exercise 10.5. Show that

$$\int_0^1 g'(s) B_s \, \mathrm{d}s = \int_0^1 [g(1) - g(s)] \, \mathrm{d}B_s$$

where g is any antiderivative of g'. Conclude that

$$\int_0^1 g'(s) B_s \, \mathrm{d}s \sim \mathcal{N}\left(0, \int_0^1 [g(1) - g(s)]^2 \, \mathrm{d}s\right).$$

In general, this exercise shows that for fixed $t > 0$, we have

$$\int_0^t g'(s) B_s \, \mathrm{d}s \sim \mathcal{N}\left(0, \int_0^t [g(t) - g(s)]^2 \, \mathrm{d}s\right).$$

Exercise 10.6. Use the result of Exercise 10.5 to establish the following generalization of Example 10.4. Show that if $n = 0, 1, 2, \ldots$ is a non-negative integer, then

$$\int_0^1 s^n B_s \, \mathrm{d}s \sim \mathcal{N}\left(0, \frac{2}{(2n+3)(n+2)}\right).$$

11

Further Properties of the Wiener Integral

Recall that we have defined the Wiener integral of a bounded, piecewise continuous deterministic function in $L^2([0,\infty))$ with respect to Brownian motion as a normal random variable, namely

$$\int_0^t g(s)\,\mathrm{d}B_s \sim \mathcal{N}\left(0, \int_0^t g^2(s)\,\mathrm{d}s\right),$$

and that we have derived the integration-by-parts formula. That is, if $g : [0,\infty) \to \mathbb{R}$ is a bounded, continuous function in $L^2([0,\infty))$ such that g is differentiable with g' also bounded and continuous, then

$$\int_0^t g(s)\,\mathrm{d}B_s = g(t)B_t - \int_0^t g'(s)B_s\,\mathrm{d}s$$

holds as an equality in distribution of random variables. The purpose of today's lecture is to give some further properties of the Wiener integral.

Example 11.1. Recall from Example 10.4 that

$$\int_0^1 B_s\,\mathrm{d}s = B_1 - \int_0^1 s\,\mathrm{d}B_s.$$

We know from that example (or from Lecture #8) that

$$\int_0^1 B_s\,\mathrm{d}s \sim \mathcal{N}(0, 1/3).$$

Furthermore, we know that $B_1 \sim \mathcal{N}(0,1)$, and we can easily calculate that

$$\int_0^1 s\,\mathrm{d}B_s \sim \mathcal{N}\left(0, \int_0^1 s^2\,\mathrm{d}s\right) = \mathcal{N}(0, 1/3).$$

If B_1 and

$$\int_0^1 s\,\mathrm{d}B_s$$

were independent random variables, then from Exercise 3.12 the distribution of

$$B_1 - \int_0^1 s\,\mathrm{d}B_s$$

would be $\mathcal{N}(0, 1 + 1/3) = \mathcal{N}(0, 4/3)$. However,

$$B_1 - \int_0^1 s \, \mathrm{d}B_s = \int_0^1 B_s \, \mathrm{d}s$$

which we know is $\mathcal{N}(0, 1/3)$. Thus, we are forced to conclude that B_1 and

$$\int_0^1 s \, \mathrm{d}B_s$$

are *not* independent.

Suppose that g and h are bounded, piecewise continuous functions in $L^2([0,\infty))$ and consider the random variables

$$I_t(g) = \int_0^t g(s) \, \mathrm{d}B_s$$

and

$$I_t(h) = \int_0^t h(s) \, \mathrm{d}B_s.$$

As the previous example suggests, these two random variables are not, in general, independent. Using linearity of the Wiener integral, we can now calculate their covariance. Since

$$I_t(g) = \int_0^t g(s) \, \mathrm{d}B_s \sim \mathcal{N}\left(0, \int_0^t g^2(s) \, \mathrm{d}s\right),$$

$$I_t(h) = \int_0^t h(s) \, \mathrm{d}B_s \sim \mathcal{N}\left(0, \int_0^t h^2(s) \, \mathrm{d}s\right),$$

and

$$I_t(g+h) = \int_0^t [g(s) + h(s)] \, \mathrm{d}B_s \sim \mathcal{N}\left(0, \int_0^t [g(s) + h(s)]^2 \, \mathrm{d}s\right),$$

and since

$$\mathrm{Var}(I_t(g+h)) = \mathrm{Var}(I_t(g) + I_t(h)) = \mathrm{Var}(I_t(g)) + \mathrm{Var}(I_t(h)) + 2\,\mathrm{Cov}(I_t(g), I_t(h)),$$

we conclude that

$$\int_0^t [g(s) + h(s)]^2 \, \mathrm{d}s = \int_0^t g^2(s) \, \mathrm{d}s + \int_0^t h^2(s) \, \mathrm{d}s + 2\,\mathrm{Cov}(I_t(g), I_t(h)).$$

Expanding the square on the left-side and simplifying implies that

$$\mathrm{Cov}(I_t(g), I_t(h)) = \int_0^t g(s)h(s) \, \mathrm{d}s.$$

Note that taking $g = h$ gives

$$\mathrm{Var}(I_t(g)) = \mathrm{Cov}(I_t(g), I_t(g)) = \int_0^t g(s)g(s) \, \mathrm{d}s = \int_0^t g^2(s) \, \mathrm{d}s$$

in agreement with our previous work. This suggests that the covariance formula should not come as a surprise to you!

Exercise 11.2. Suppose that $g(s) = \sin s$, $0 \le s \le \pi$, and $h(s) = \cos s$, $0 \le s \le \pi$.

(a) Show that $\mathrm{Cov}(I_\pi(g), I_\pi(h)) = 0$.
(b) Prove that $I_\pi(g)$ and $I_\pi(h)$ are independent. *Hint: Theorem 3.17 will be useful here.*

The same proof you used for (b) of the previous exercise holds more generally.

Theorem 11.3. *If g and h are bounded, piecewise continuous functions in $L^2([0,\infty))$ with*

$$\int_0^t g(s)h(s)\,\mathrm{d}s = 0,$$

then the random variables $I_t(g)$ and $I_t(h)$ are independent.

Exercise 11.4. Prove this theorem.

We end this lecture with two extremely important properties of the Wiener integral I_t, namely that $\{I_t, t \ge 0\}$ is a martingale and that the trajectories $t \mapsto I_t$ are continuous. The proof of the following theorem requires some facts about convergence in L^2 and is therefore beyond our present scope.

Theorem 11.5. *Suppose that $g : [0,\infty) \to \mathbb{R}$ is a bounded, piecewise continuous function in $L^2([0,\infty))$. If the process $\{I_t, t \ge 0\}$ is defined by setting $I_0 = 0$ and*

$$I_t = \int_0^t g(s)\,\mathrm{d}B_s$$

for $t > 0$, then

(a) *the process $\{I_t, t \ge 0\}$ is a continuous-time martingale with respect to the Brownian filtration $\{\mathcal{F}_t, t \ge 0\}$, and*
(b) *the trajectory $t \mapsto I_t$ is continuous.*

That is, $\{I_t, t \ge 0\}$ is a continuous-time continuous martingale.

12

Itô Integration (Part I)

Recall that for bounded, piecewise continuous deterministic $L^2([0,\infty))$ functions, we have defined the Wiener integral

$$I_t = \int_0^t g(s)\,\mathrm{d}B_s$$

which satisfied the following properties:

(i) $I_0 = 0$,
(ii) for fixed $t > 0$, the random variable I_t is normally distributed with mean 0 and variance

$$\int_0^t g^2(s)\,\mathrm{d}s,$$

(iii) the stochastic process $\{I_t,\ t \geq 0\}$ is a martingale with respect to the Brownian filtration $\{\mathcal{F}_t, t \geq 0\}$, and
(iv) the trajectory $t \mapsto I_t$ is continuous.

Our goal for the next two lectures is to define the integral

$$I_t = \int_0^t g(s)\,\mathrm{d}B_s. \tag{12.1}$$

for *random* functions g.

We understand from our work on Wiener integrals that for fixed $t > 0$ the stochastic integral I_t must be a random variable depending on the Brownian sample path. Thus, the interpretation of (12.1) is as follows. Fix a realization (or sample path) of Brownian motion $\{B_t(\omega), t \geq 0\}$ and a realization (depending on the Brownian sample path observed) of the stochastic process $\{g(t,\omega), t \geq 0\}$ so that, for fixed $t > 0$, the integral (12.1) is really a random variable, namely

$$I_t(\omega) = \int_0^t g(s,\omega)\,\mathrm{d}B_s(\omega).$$

We begin with the example where g is a Brownian motion. This seemingly simple example will serve to illustrate more of the subtleties of integration with respect to Brownian motion.

Example 12.1. Suppose that $\{B_t,\ t \ge 0\}$ is a Brownian motion with $B_0 = 0$. We would like to compute

$$I_t = \int_0^t B_s \,\mathrm{d}B_s$$

for this particular realization $\{B_t, t \ge 0\}$ of Brownian motion. If Riemann integration were valid, we would expect, using the fundamental theorem of calculus, that

$$I_t = \int_0^t B_s \,\mathrm{d}B_s = \frac{1}{2}(B_t^2 - B_0^2) = \frac{1}{2}B_t^2. \tag{12.2}$$

Motivated by our experience with Wiener integration, we expect that I_t has mean 0. However, if I_t is given by (12.2), then

$$\mathbb{E}(I_t) = \frac{1}{2}\mathbb{E}(B_t^2) = \frac{t}{2}.$$

We might also expect that the stochastic process $\{I_t,\ t \ge 0\}$ is a martingale; of course, $\{B_t^2/2,\ t \ge 0\}$ is *not* a martingale, although,

$$\left\{\frac{1}{2}B_t^2 - \frac{t}{2},\ t \ge 0\right\} \tag{12.3}$$

is a martingale. Is it possible that the value of I_t is given by (12.3) instead? We will now show that yes, in fact,

$$\int_0^t B_s \,\mathrm{d}B_s = \frac{1}{2}B_t^2 - \frac{t}{2}.$$

Suppose that $\pi_n = \{0 = t_0 < t_1 < t_2 < \cdots < t_n = t\}$ is a partition of $[0,t]$ and let

$$L_n = \sum_{i=1}^n B_{t_{i-1}}(B_{t_i} - B_{t_{i-1}}) \quad \text{and} \quad R_n = \sum_{i=1}^n B_{t_i}(B_{t_i} - B_{t_{i-1}})$$

denote the left-hand and right-hand Riemann sums, respectively. Observe that

$$R_n - L_n = \sum_{i=1}^n B_{t_i}(B_{t_i} - B_{t_{i-1}}) - \sum_{i=1}^n B_{t_{i-1}}(B_{t_i} - B_{t_{i-1}}) = \sum_{i=1}^n (B_{t_i} - B_{t_{i-1}})^2. \tag{12.4}$$

The next theorem shows that

$$(R_n - L_n) \not\to 0 \ \text{ as } \ \mathrm{mesh}(\pi_n) = \max_{i \le i \le n}(t_i - t_{i-1}) \to 0$$

which implies that the attempted Riemann integration (12.2) is not valid for Brownian motion.

Theorem 12.2. *If* $\{\pi_n,\ n = 1, 2, 3, \ldots\}$ *is a refinement of* $[0,t]$ *with* $\mathrm{mesh}(\pi_n) \to 0$*, then*

$$\sum_{i=1}^n \left(B_{t_i} - B_{t_{i-1}}\right)^2 \to t \quad \text{in } L^2$$

as $\mathrm{mesh}(\pi_n) \to 0$.

Proof. To begin, notice that

$$\sum_{i=1}^{n}(t_i - t_{i-1}) = t.$$

Let

$$Y_n = \sum_{i=1}^{n}\left(B_{t_i} - B_{t_{i-1}}\right)^2 - t = \sum_{i=1}^{n}\left[\left(B_{t_i} - B_{t_{i-1}}\right)^2 - (t_i - t_{i-1})\right] = \sum_{i=1}^{n} X_i$$

where

$$X_i = \left(B_{t_i} - B_{t_{i-1}}\right)^2 - (t_i - t_{i-1}),$$

and note that

$$Y_n^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} X_i\, X_j = \sum_{i=1}^{n} X_i^2 + 2\sum_{i<j} X_i\, X_j.$$

The independence of the Brownian increments implies that $\mathbb{E}(X_iX_j) = 0$ for $i \neq j$; hence,

$$\mathbb{E}(Y_n^2) = \sum_{i=1}^{n}\mathbb{E}(X_i^2).$$

But

$$\begin{aligned}\mathbb{E}(X_i^2) &= \mathbb{E}\left[(B_{t_i} - B_{t_{i-1}})^4\right] - 2(t_i - t_{i-1})\mathbb{E}\left[(B_{t_i} - B_{t_{i-1}})^2\right] + (t_i - t_{i-1})^2 \\ &= 3(t_i - t_{i-1})^2 - 2(t_i - t_{i-1})^2 + (t_i - t_{i-1})^2 \\ &= 2(t_i - t_{i-1})^2\end{aligned}$$

since the fourth moment of a normal random variable with mean 0 and variance $t_i - t_{i-1}$ is $3(t_i - t_{i-1})^2$. Therefore,

$$\mathbb{E}(Y_n^2) = \sum_{i=1}^{n}\mathbb{E}(X_i^2) = 2\sum_{i=1}^{n}(t_i - t_{i-1})^2 \le 2\ \mathrm{mesh}(\pi_n)\ \sum_{i=1}^{n}(t_i - t_{i-1}) = 2t\ \mathrm{mesh}(\pi_n) \to 0$$

as $\mathrm{mesh}(\pi_n) \to 0$ from which we conclude that $\mathbb{E}(Y_n^2) \to 0$ as $\mathrm{mesh}(\pi_n) \to 0$. However, this is exactly what it means for $Y_n \to 0$ in L^2 as $\mathrm{mesh}(\pi_n) \to 0$, and the proof is complete. □

As a result of this theorem, we define the *quadratic variation* of Brownian motion to be this limit in L^2.

Definition 12.3. The *quadratic variation* of a Brownian motion $\{B_t, t \geq 0\}$ on the interval $[0, t]$ is defined to be

$$Q_2(B[0,t]) = t \quad (\text{in } L^2).$$

Since

$$(R_n - L_n) \to t \text{ in } L^2 \text{ as } \ \mathrm{mesh}(\pi_n) \to 0$$

we see that L_n and R_n cannot possibly have the same limits in L^2. This is not necessarily surprising since $B_{t_{i-1}}$ is independent of $B_{t_i} - B_{t_{i-1}}$ from which it follows that $\mathbb{E}(L_n) = 0$ while $\mathbb{E}(R_n) = t$.

Exercise 12.4. Show that $\mathbb{E}(L_n) = 0$ and $\mathbb{E}(R_n) = t$.

On the other hand,

$$\begin{aligned} R_n + L_n = \sum_{i=1}^{n} B_{t_i}(B_{t_i} - B_{t_{i-1}}) + \sum_{i=1}^{n} B_{t_{i-1}}(B_{t_i} - B_{t_{i-1}}) &= \sum_{i=1}^{n}(B_{t_i} + B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}}) \\ &= \sum_{i=1}^{n}(B_{t_i}^2 - B_{t_{i-1}}^2) \\ &= B_{t_n}^2 - B_{t_0}^2 \\ &= B_t^2 - B_0^2 \\ &= B_t^2. \end{aligned} \tag{12.5}$$

Thus, from (12.4) and (12.5) we conclude that

$$L_n = \frac{1}{2}\left(B_t^2 - \sum_{i=1}^{n}(B_{t_i} - B_{t_{i-1}})^2\right) \quad \text{and} \quad R_n = \frac{1}{2}\left(B_t^2 + \sum_{i=1}^{n}(B_{t_i} - B_{t_{i-1}})^2\right)$$

and so

$$L_n \to \frac{1}{2}(B_t^2 - t) \quad \text{in } L^2 \quad \text{and} \quad R_n \to \frac{1}{2}(B_t^2 + t) \quad \text{in } L^2.$$

Unlike the usual Riemann integral, the limit of these sums *does* depend on the intermediate points used (i.e., left– or right-endpoints). However, $\{B_t^2 + t,\ t \geq 0\}$ is not a martingale, although $\{B_t^2 - t,\ t \geq 0\}$ is a martingale. Therefore, while *both* of these limits are valid ways to define the integral I_t, it is reasonable to use as the definition the limit for which a martingale is produced. And so we make the following definition:

$$\begin{aligned} \int_0^t B_s \, \mathrm{d}B_s &= \lim L_n \text{ in } L^2 \\ &= \frac{1}{2}B_t^2 - \frac{t}{2}. \end{aligned} \tag{12.6}$$

13

Itô Integration (Part II)

Recall from last lecture that we defined the Itô integral of Brownian motion as

$$\int_0^t B_s \, \mathrm{d}B_s = \lim L_n \text{ in } L^2$$
$$= \frac{1}{2}B_t^2 - \frac{t}{2}. \tag{13.1}$$

where $\{\pi_n, n = 1, 2, \ldots\}$ is a refinement of $[0, t]$ with $\text{mesh}(\pi_n) \to 0$ and

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

denotes the left-hand Riemann sum corresponding to the partition $\pi_n = \{0 = t_0 < \cdots < t_n = t\}$.

We saw that the definition of I_t depended on the intermediate point used in the Riemann sum, and that the reason for *choosing* the left-hand sum was that it produced a martingale.

We now present another example which shows some of the dangers of a naïve attempt at stochastic integration.

Example 13.1. Let $\{B_t, \ t \geq 0\}$ be a realization of Brownian motion with $B_0 = 0$, and suppose that for any fixed $0 \leq t < 1$ we define the random variable I_t by

$$I_t = \int_0^t B_1 \, \mathrm{d}B_s.$$

Since B_1 is constant (for a given realization), we might expect that

$$I_t = \int_0^t B_1 \, \mathrm{d}B_s = B_1 \int_0^t \mathrm{d}B_s = B_1(B_t - B_0) = B_1 B_t.$$

However,

$$\mathbb{E}(I_t) = \mathbb{E}(B_1 B_t) = \min\{1, t\} = t$$

which is not constant. Therefore, if we want to obtain martingales, this is not how we should define the integral I_t. The problem here is that the random variable B_1 is not adapted to $\mathcal{F}_t = \sigma(B_s, \ 0 \leq s \leq t)$ for any fixed $0 \leq t < 1$.

From the previous example, we see that in order to define

$$I_t = \int_0^t g(s)\,\mathrm{d}B_s$$

the stochastic process $\{g(s),\ 0 \le s \le t\}$ will necessarily need to be adapted to the Brownian filtration $\{\mathcal{F}_s,\ 0 \le s \le t\} = \{\sigma(B_r,\ 0 \le r \le s),\ 0 \le s \le t\}$.

Definition 13.2. Let L^2_{ad} denote the space of stochastic processes $g = \{g(t), t \ge 0\}$ such that

(i) g is adapted to the Brownian filtration $\{\mathcal{F}_t,\ t \ge 0\}$ (i.e., $g(t) \in \mathcal{F}_t$ for every $t > 0$), and

(ii) $\displaystyle\int_0^T \mathbb{E}[g^2(t)]\,\mathrm{d}t < \infty$ for every $T > 0$.

Our goal is to now define

$$I_t(g) = \int_0^t g(s)\,\mathrm{d}B_s$$

for $g \in L^2_{\text{ad}}$. This is accomplished in a more technical manner than the construction of the Wiener integral, and the precise details will therefore be omitted. Complete details may by found in [17], however.

The first step involves defining the integral for *step stochastic processes*, and the second step is to then pass to a limit.

Suppose that $g = \{g(t), t \ge 0\}$ is a stochastic process. We say that g is a *step stochastic process* if for every $t \ge 0$ we can write

$$g(s,\omega) = \sum_{i=1}^{n-1} X_{i-1}(\omega)\mathbb{1}_{[t_{i-1},t_i)}(s) + X_{n-1}(\omega)\mathbb{1}_{[t_{n-1},t_n]}(s) \tag{13.2}$$

for $0 \le s \le t$ where $\{0 = t_0 < t_1 < \cdots < t_n = t\}$ is a partition of $[0,t]$ and $\{X_j, j = 0,1,\ldots,n-1\}$ is a finite collection of random variables. Define the integral of such a g as

$$I_t(g)(\omega) = \int_0^t g(s,\omega)\,\mathrm{d}B_s(\omega) = \sum_{i=1}^{n} X_{i-1}(\omega)(B_{t_i}(\omega) - B_{t_{i-1}}(\omega)), \tag{13.3}$$

and note that (13.3) is simply a discrete stochastic integral as in (6.1), the so-called *martingale transform* of X by B.

The second, and more difficult, step is show that it is possible to approximate an arbitrary $g \in L^2_{\text{ad}}$ by a sequence of step processes $g_n \in L^2_{\text{ad}}$ such that

$$\lim_{n\to\infty} \int_0^t \mathbb{E}(|g_n(s) - g(s)|^2)\,\mathrm{d}s = 0.$$

We then define $I_t(g)$ to be the limit in L^2 of the approximating Itô integrals $I_t(g_n)$ defined by (13.3), and show that the limit does not depend of the choice of step processes $\{g_n\}$; that is,

$$I_t(g) = \lim_{n\to\infty} I_t(g_n) \quad \text{in } L^2 \tag{13.4}$$

and so we have the following definition.

Definition 13.3. If $g \in L^2_{\text{ad}}$, define the Itô integral of g to be

$$I_t(g) = \int_0^t g(s)\,\mathrm{d}B_s$$

where $I_t(g)$ is defined as the limit in (13.4).

Notice that the definition of the Itô integral did not use any approximating Riemann sums. However, in Lecture #12 we calculated $\int_0^t B_s\,\mathrm{d}B_s$ directly by taking the limit in L^2 of the approximating Riemann sums. It is important to know when both approaches give the same answer which is the content of the following theorem. For a proof, see Theorem 4.7.1 of [17].

Theorem 13.4. *If the stochastic process $g \in L^2_{ad}$ and $\mathbb{E}(g(s)g(t))$ is a continuous function of s and t, then*

$$\int_0^t g(s)\,\mathrm{d}B_s = \lim \sum_{i=1}^n g(t_{i-1})(B_{t_i} - B_{t_{i-1}}) \quad \text{in } L^2.$$

Example 13.5. For example, if the stochastic process g is a Brownian motion, then B_t is necessarily $\mathcal{F}_t$-measurable with $\mathbb{E}(B_t^2) = t < \infty$ for every $t > 0$. Since $\mathbb{E}(B_sB_t) = \min\{s,t\}$ is a continuous function of s and t, we conclude that Theorem 13.4 can be applied to calculate $\int_0^t B_s\,\mathrm{d}B_s$. This is exactly what we did in (12.6).

The following result collects together a number of properties of the Itô integral. It is relatively straightforward to prove all of these properties when g is a step stochastic process. It is rather more involved to pass to the appropriate limits to obtain these results for general $g \in L^2_{\text{ad}}$.

Theorem 13.6. *Suppose that g, $h \in L^2_{ad}$, and let*

$$I_t(g) = \int_0^t g(s)\,\mathrm{d}B_s \quad \text{and} \quad I_t(h) = \int_0^t h(s)\,\mathrm{d}B_s.$$

(a) *If α, $\beta \in \mathbb{R}$ are constants, then $I_t(\alpha g + \beta h) = \alpha I_t(g) + \beta I_t(h)$.*
(b) *$I_t(g)$ is a random variable with $I_0(g) = 0$, $\mathbb{E}(I_t(g)) = 0$ and*

$$\operatorname{Var}(I_t(g)) = \mathbb{E}[I_t^2(g)] = \int_0^t \mathbb{E}[g^2(s)]\,\mathrm{d}s. \tag{13.5}$$

(c) *The covariance of $I_t(g)$ and $I_t(h)$ is given by*

$$\mathbb{E}[I_t(g)I_t(h)] = \int_0^t \mathbb{E}[g(s)h(s)]\,\mathrm{d}s.$$

(d) *The process $\{I_t,\ t \geq 0\}$ is a martingale with respect to the Brownian filtration.*
(e) *The trajectory $t \mapsto I_t$ is a continuous function of t.*

Remark. The equality (13.5) in the second part of this theorem is sometimes known as the *Itô isometry*.

Remark. It is important to observe that the Wiener integral is a special case of the Itô integral. That is, if g is a bounded, piecewise continuous deterministic $L^2([0,\infty))$ function, then $g \in L^2_{\text{ad}}$ and so the Itô integral of g with respect to Brownian motion can be constructed. The fact that g is deterministic means that we recover the properties for the Wiener integral from the properties in Theorem 13.6 for the Itô integral. Theorem 10.2, the integration-by-parts formula for Wiener integration, will follow from Theorem 15.1, the generalized version of Itô's formula.

Remark. It is also important to observe that, unlike the Wiener integral, there is no general form of the distribution of $I_t(g)$. In general, the Riemann sum approximations to $I_t(g)$ contain terms of the form

$$g(t_{i-1})(B_{t_i} - B_{t_{i-1}}). \tag{13.6}$$

When g is deterministic, the distribution of the $I_t(g)$ is normal as a consequence of the fact that the sum of independent normals is normal. However, when g is random, the distribution of (13.6) is not necessarily normal. The following exercises illustrates this point.

Exercise 13.7. Consider

$$I = \int_0^1 B_s \, \mathrm{d}B_s = \frac{B_1^2}{2} - \frac{1}{2}.$$

Since $B_1 \sim \mathcal{N}(0,1)$, we know that $B_1^2 \sim \chi^2(1)$, and so we conclude that

$$2I + 1 \sim \chi^2(1).$$

Simulate 10000 realizations of I and plot a histogram of $2I+1$. Does your simulation match the theory?

Exercise 13.8. Suppose that $\{B_t, t \geq 0\}$ is a standard Brownian motion, and let the stochastic process $\{g(t), t \geq 0\}$ be defined as follows. At time $t = 0$, flip a fair coin and let $g(0) = 2$ if the coin shows heads, and let $g(0) = 3$ if the coin shows tails. At time $t = \sqrt{2}$, roll a fair die and let $g(\sqrt{2}\,)$ equal the number of dots showing on the die. If $0 < t < \sqrt{2}$, define $g(t) = g(0)$, and if $t > \sqrt{2}$, define $g(t) = g(\sqrt{2}\,)$. Note that $\{g(t), t \geq 0\}$ is a step stochastic process.

(a) Express g in the form (13.2).
(b) Sketch a graph of the stochastic process $\{g(t), t \geq 0\}$.
(c) Determine the mean and the variance of

$$\int_0^5 g(s) \, \mathrm{d}B_s.$$

(d) If possible, determine the distribution of

$$\int_0^5 g(s) \, \mathrm{d}B_s.$$

14

Itô's Formula (Part I)

In this section, we develop Itô's formula which may be called the "fundamental theorem of stochastic integration." It allows for the explicit calculation of certain Itô integrals in much the same way that the fundamental theorem of calculus gives one a way to compute definite integrals. In fact, recall that if $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are differentiable functions, then

$$\frac{\mathrm{d}}{\mathrm{d}t}(f \circ g)(t) = f'(g(t)) \cdot g'(t),$$

which implies that

$$\int_0^t f'(g(s)) \cdot g'(s) \, \mathrm{d}s = (f \circ g)(t) - (f \circ g)(0). \tag{14.1}$$

Our experience with the Itô integral that we computed earlier, namely

$$\int_0^t B_s \, \mathrm{d}B_s = \frac{1}{2}(B_t^2 - t),$$

tells us that we do not expect a formula quite like the fundamental theorem of calculus given by (14.1).

In order to explain Itô's formula, we begin by recalling Taylor's theorem. That is, if $f : \mathbb{R} \to \mathbb{R}$ is infinitely differentiable, then f can be expressed as an infinite polynomial expanded around $a \in \mathbb{R}$ as

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots .$$

We now let $x = t + \Delta t$ and $a = t$ so that

$$f(t + \Delta t) = f(t) + f'(t)\Delta t + \frac{f''(t)}{2!}(\Delta t)^2 + \frac{f'''(t)}{3!}(\Delta t)^3 + \cdots$$

which we can write as

$$\frac{f(t + \Delta t) - f(t)}{\Delta t} = f'(t) + \frac{f''(t)}{2!}\Delta t + \frac{f'''(t)}{3!}(\Delta t)^2 + \cdots .$$

At this point we see that if $\Delta t \to 0$, then

$$\lim_{\Delta t \to 0} \frac{f(t + \Delta t) - f(t)}{\Delta t} = \lim_{\Delta t \to 0} \left[f'(t) + \frac{f''(t)}{2!}\Delta t + \frac{f'''(t)}{3!}(\Delta t)^2 + \cdots \right] = f'(t)$$

which is exactly the definition of derivative.

The same argument can be used to prove the chain rule. That is, suppose that f and g are infinitely differentiable. Let $x = g(t) + \Delta g(t)$ and $a = g(t)$ so that Taylor's theorem takes the form

$$f(g(t) + \Delta g(t)) = f(g(t)) + f'(g(t))\Delta g(t) + \frac{f''(g(t))}{2!}(\Delta g(t))^2 + \frac{f'''(g(t))}{3!}(\Delta g(t))^3 + \cdots .$$

We now write $\Delta g(t) = g(t + \Delta t) - g(t)$ so that

$$f(g(t + \Delta t)) - f(g(t)) = f'(g(t))(g(t + \Delta t) - g(t)) + \frac{f''(g(t))}{2!}(g(t + \Delta t) - g(t))^2 + \frac{f'''(g(t))}{3!}(g(t + \Delta t) - g(t))^3 + \cdots .$$

Dividing both sides by Δt implies

$$\frac{f(g(t + \Delta t)) - f(g(t))}{\Delta t} = f'(g(t)) \cdot \frac{g(t + \Delta t) - g(t)}{\Delta t} + \frac{f''(g(t))}{2!} \cdot \frac{(g(t + \Delta t) - g(t))^2}{\Delta t} + \frac{f'''(g(t))}{3!} \cdot \frac{(g(t + \Delta t) - g(t))^3}{\Delta t} + \cdots . \tag{14.2}$$

The question now is what happens when $\Delta t \to 0$. Notice that the limit of the left-side of the previous equation (14.2) is

$$\lim_{\Delta t \to 0} \frac{f(g(t + \Delta t)) - f(g(t))}{\Delta t} = \lim_{\Delta t \to 0} \frac{(f \circ g)(t + \Delta t) - (f \circ g)(t)}{\Delta t} = \frac{\mathrm{d}}{\mathrm{d}t}(f \circ g)(t).$$

As for the right-side of (14.2), we find for the first term that

$$\lim_{\Delta t \to 0} \left[f'(g(t)) \cdot \frac{g(t + \Delta t) - g(t)}{\Delta t} \right] = f'(g(t)) \cdot \lim_{\Delta t \to 0} \frac{g(t + \Delta t) - g(t)}{\Delta t} = f'(g(t)) \cdot g'(t).$$

For the second term, however, we have

$$\begin{aligned} \lim_{\Delta t \to 0} & \left[\frac{f''(g(t))}{2!} \cdot \frac{(g(t + \Delta t) - g(t))^2}{\Delta t} \right] \\ &= \frac{f''(g(t))}{2!} \cdot \lim_{\Delta t \to 0} \frac{g(t + \Delta t) - g(t)}{\Delta t} \cdot \lim_{\Delta t \to 0} [g(t + \Delta t) - g(t)] \\ &= \frac{f''(g(t))}{2!} \cdot g'(t) \cdot 0 = 0 \end{aligned}$$

which follows since g is differentiable (and therefore continuous). Similarly, the higher order terms all approach 0 in the $\Delta t \to 0$ limit. Combining everything gives

$$\frac{\mathrm{d}}{\mathrm{d}t}(f \circ g)(t) = f'(g(t)) \cdot g'(t).$$

In fact, this proof of the chain rule illustrates precisely why the fundamental theorem of calculus fails for Itô integrals. Brownian motion is nowhere differentiable, and so the step of the proof of the chain rule where the second order term vanishes as $\Delta t \to 0$ is not valid. Indeed, if we take

$g(t) = B_t$ and divide by Δt, then we find

$$\begin{aligned}\frac{\Delta f(B_t)}{\Delta t} &= \frac{f(B_{t+\Delta t}) - f(B_t)}{\Delta t} \\ &= f'(B_t)\frac{\Delta B_t}{\Delta t} + \frac{f''(B_t)}{2!} \cdot \frac{(\Delta B_t)^2}{\Delta t} + \frac{f'''(B_t)}{3!} \cdot \frac{(\Delta B_t)^3}{\Delta t} + \cdots .\end{aligned}$$

In the limit as $\Delta t \to 0$, the left-side of the previous equation is

$$\lim_{\Delta t \to 0} \frac{\Delta f(B_t)}{\Delta t} = \frac{\mathrm{d}}{\mathrm{d}t} f(B_t).$$

As for the right-side, we are tempted to say that the first term approaches

$$\lim_{\Delta t \to 0} \left[f'(B_t) \cdot \frac{\Delta B_t}{\Delta t} \right] = f'(B_t) \cdot \frac{\mathrm{d}B_t}{\mathrm{d}t}$$

so that

$$\frac{\mathrm{d}}{\mathrm{d}t} f(B_t) = f'(B_t) \cdot \frac{\mathrm{d}B_t}{\mathrm{d}t} + \frac{f''(B_t)}{2!} \cdot \left[\lim_{\Delta t \to 0} \frac{(\Delta B_t)^2}{\Delta t} \right] + \frac{f'''(B_t)}{3!} \cdot \left[\lim_{\Delta t \to 0} \frac{(\Delta B_t)^3}{\Delta t} \right] + \cdots .$$

(Even though Brownian motion is nowhere differentiable so that $\mathrm{d}B_t / \mathrm{d}t$ does not exist, bear with us.) We know that $\Delta B_t = B_{t+\Delta t} - B_t \sim \mathcal{N}(0, \Delta t)$ so that

$$\mathrm{Var}(\Delta B_t) = \mathbb{E}\left[(\Delta B_t)^2 \right] = \Delta t$$

or, approximately,

$$(\Delta B_t)^2 \approx \Delta t.$$

This suggests that

$$\lim_{\Delta t \to 0} \frac{(\Delta B_t)^2}{\Delta t} = \lim_{\Delta t \to 0} \frac{\Delta t}{\Delta t} = 1$$

but if $k \geq 3$, then

$$\lim_{\Delta t \to 0} \frac{(\Delta B_t)^k}{\Delta t} = \lim_{\Delta t \to 0} \frac{\left(\sqrt{\Delta t}\right)^k}{\Delta t} = \lim_{\Delta t \to 0} \frac{\Delta t \left(\sqrt{\Delta t}\right)^{k-2}}{\Delta t} = 0.$$

(In fact, these approximations can be justified using our result on the quadratic variation of Brownian motion.) Hence, we conclude that

$$\frac{\mathrm{d}}{\mathrm{d}t} f(B_t) = f'(B_t) \cdot \frac{\mathrm{d}B_t}{\mathrm{d}t} + \frac{f''(B_t)}{2!}.$$

Multiplying through by $\mathrm{d}t$ gives

$$\mathrm{d}f(B_t) = f'(B_t)\,\mathrm{d}B_t + \frac{f''(B_t)}{2}\,\mathrm{d}t$$

and so if we integrate from 0 to T, then

$$\int_0^T \mathrm{d}f(B_t) = \int_0^T f'(B_t)\,\mathrm{d}B_t + \frac{1}{2}\int_0^T f''(B_t)\,\mathrm{d}t.$$

Since

$$\int_0^T \mathrm{d}f(B_t) = f(B_T) - f(B_0)$$

we have motivated the following result.

Theorem 14.1 (K. Itô, 1944). *If $f(x) \in C^2(\mathbb{R})$, then*

$$f(B_t) - f(B_0) = \int_0^t f'(B_s)\,\mathrm{d}B_s + \frac{1}{2}\int_0^t f''(B_s)\,\mathrm{d}s. \tag{14.3}$$

Notice that the first integral in (14.3) is an Itô integral, while the second integral is a Riemann integral.

Example 14.2. Let $f(x) = x^2$ so that $f'(x) = 2x$ and $f''(x) = 2$. Therefore, Itô's formula implies

$$B_t^2 - B_0^2 = \int_0^t 2B_s\,\mathrm{d}B_s + \frac{1}{2}\int_0^t 2\,\mathrm{d}s = 2\int_0^t B_s\,\mathrm{d}B_s + t.$$

Rearranging we conclude

$$\int_0^t B_s\,\mathrm{d}B_s = \frac{1}{2}(B_t^2 - t)$$

which agrees with our earlier result for this integral.

Example 14.3. Let $f(x) = x^3$ so that $f'(x) = 3x^2$ and $f''(x) = 6x$. Therefore, Itô's formula implies

$$B_t^3 - B_0^3 = \int_0^t 3B_s^2\,\mathrm{d}B_s + \frac{1}{2}\int_0^t 6B_s\,\mathrm{d}s$$

so that rearranging yields

$$\int_0^t B_s^2\,\mathrm{d}B_s = \frac{1}{3}B_t^3 - \int_0^t B_s\,\mathrm{d}s, \tag{14.4}$$

and hence we are able to evaluate another Itô integral explictly. We can determine the distribution of the Riemann integral in the above expression by recalling from the integration-by-parts formula for Wiener integrals that (for fixed $t > 0$)

$$\int_0^t B_s\,\mathrm{d}s = \int_0^t (t-s)\,\mathrm{d}B_s \sim \mathcal{N}\left(0, \int_0^t (t-s)^2\,\mathrm{d}s\right) \sim \mathcal{N}\left(0, \frac{t^3}{3}\right).$$

Example 14.4. We can take another approach to the previous example by using the integration-by-parts formula for Wiener integrals in a slightly different way, namely

$$\int_0^t s\,\mathrm{d}B_s = tB_t - \int_0^t B_s\,\mathrm{d}s.$$

If we then substitute this into (14.4) we find

$$\int_0^t B_s^2\,\mathrm{d}B_s = \frac{1}{3}B_t^3 - tB_t + \int_0^t s\,\mathrm{d}B_s$$

and so using the linearity of the Itô integral we are able to evaluate another integral explicitly, namely

$$\int_0^t (B_s^2 - s)\,\mathrm{d}B_s = \frac{1}{3}B_t^3 - tB_t.$$

Example 14.5. Let $f(x) = x^4$ so that $f'(x) = 4x^3$ and $f''(x) = 12x^2$. Therefore, Itô's formula implies

$$B_t^4 - B_0^4 = \int_0^t 4B_s^3\,\mathrm{d}B_s + \frac{1}{2}\int_0^t 12B_s^2\,\mathrm{d}s \tag{14.5}$$

and so we can rearrange (14.5) to compute yet another Itô integral:

$$\int_0^t B_s^3\,\mathrm{d}B_s = \frac{1}{4}B_t^4 - \frac{3}{2}\int_0^t B_s^2\,\mathrm{d}s.$$

15

Itô's Formula (Part II)

Recall from last lecture that we derived Itô's formula, namely if $f(x) \in C^2(\mathbb{R})$, then

$$f(B_t) - f(B_0) = \int_0^t f'(B_s)\,\mathrm{d}B_s + \frac{1}{2}\int_0^t f''(B_s)\,\mathrm{d}s. \tag{15.1}$$

The derivation of Itô's formula involved carefully manipulating Taylor's theorem for the function $f(x)$. (In fact, the actual proof of Itô's formula follows a careful analysis of Taylor's theorem for a function of one variable.) As you may know from MATH 213, there is a version of Taylor's theorem for functions of two variables. Thus, by writing down Taylor's theorem for the function $f(t, x)$ and carefully checking which higher order terms disappear, one can derive the following generalized version of Itô's formula.

Consider those functions of two variables, say $f(t, x)$, which have one continuous derivative in the "t-variable" for $t \geq 0$, and two continuous derivatives in the "x-variable." If f is such a function, we say that $f \in C^1([0, \infty)) \times C^2(\mathbb{R})$.

Theorem 15.1 (Generalized Version of Itô's Formula). *If $f \in C^1([0, \infty)) \times C^2(\mathbb{R})$, then*

$$f(t, B_t) - f(0, B_0) = \int_0^t \frac{\partial}{\partial x} f(s, B_s)\,\mathrm{d}B_s + \frac{1}{2}\int_0^t \frac{\partial^2}{\partial x^2} f(s, B_s)\,\mathrm{d}s + \int_0^t \frac{\partial}{\partial t} f(s, B_s)\,\mathrm{d}s. \tag{15.2}$$

Remark. It is traditional to use the variables t and x for the function $f(t, x)$ of two variables in the generalized version of Itô's formula. This has the unfortunate consequence that the letter t serves both as a dummy variable for the function $f(t, x)$ and as a time variable in the upper limit of integration. One way around this confusion is to use the prime ($'$) notation for derivatives in the space variable (the x-variable) and the dot ($\dot{}$) notation for derivatives in the time variable (the t-variable). That is,

$$f'(t, x) = \frac{\partial}{\partial x} f(t, x), \quad f''(t, x) = \frac{\partial^2}{\partial x^2} f(t, x), \quad \dot{f}(t, x) = \frac{\partial}{\partial t} f(t, x),$$

and so (15.2) becomes

$$f(t, B_t) - f(0, B_0) = \int_0^t f'(s, B_s)\,\mathrm{d}B_s + \frac{1}{2}\int_0^t f''(s, B_s)\,\mathrm{d}s + \int_0^t \dot{f}(s, B_s)\,\mathrm{d}s.$$

Example 15.2. Let $f(t,x) = tx^2$ so that

$$f'(t,x) = 2xt, \quad f''(t,x) = 2t, \quad \text{and} \quad \dot{f}(t,x) = x^2.$$

Therefore, the generalized version of Itô's formula implies

$$tB_t^2 = \int_0^t 2sB_s \, \mathrm{d}B_s + \frac{1}{2}\int_0^t 2s \, \mathrm{d}s + \int_0^t B_s^2 \, \mathrm{d}s.$$

Upon rearranging we conclude

$$\int_0^t sB_s \, \mathrm{d}B_s = \frac{1}{2}\left(tB_t^2 - \frac{t^2}{2} - \int_0^t B_s^2 \, \mathrm{d}s \right).$$

Example 15.3. Let $f(t,x) = \frac{1}{3}x^3 - xt$ so that

$$f'(t,x) = x^2 - t, \quad f''(t,x) = 2x, \quad \text{and} \quad \dot{f}(t,x) = -x.$$

Therefore, the generalized version of Itô's formula implies

$$\frac{1}{3}B_t^3 - tB_t = \int_0^t (B_s^2 - s) \, \mathrm{d}B_s + \frac{1}{2}\int_0^t 2B_s \, \mathrm{d}s - \int_0^t B_s \, \mathrm{d}s = \int_0^t (B_s^2 - s) \, \mathrm{d}B_s$$

which gives the same result as was obtained in Example 14.4.

Example 15.4. If we combine our result of Example 14.5, namely

$$\int_0^t B_s^3 \, \mathrm{d}B_s = \frac{1}{4}B_t^4 - \frac{3}{2}\int_0^t B_s^2 \, \mathrm{d}s,$$

with our result of Example 15.2, namely

$$\int_0^t sB_s \, \mathrm{d}B_s = \frac{1}{2}\left(tB_t^2 - \frac{t^2}{2} - \int_0^t B_s^2 \, \mathrm{d}s \right),$$

then we conclude that

$$\int_0^t (B_s^3 - 3sB_s) \, \mathrm{d}B_s = \frac{1}{4}B_t^4 - \frac{3}{2}tB_t^2 + \frac{3}{4}t^2.$$

Example 15.5. If we re-write the results of Example 14.4 and Example 15.4 slightly differently, then we see that

$$\int_0^t 3(B_s^2 - s) \, \mathrm{d}B_s = B_t^3 - 3tB_t$$

and

$$\int_0^t 4(B_s^3 - 3sB_s) \, \mathrm{d}B_s = B_t^4 - 6tB_t^2 + 3t^2.$$

The reason for doing this is that Theorem 13.6 tells us that Itô integrals are martingales. Hence, we see that $\{B_t^3 - 3tB_t, t \geq 0\}$ and $\{B_t^4 - 6tB_t^2 + 3t^2, t \geq 0\}$ must therefore be martingales with respect to the Brownian filtration $\{\mathcal{F}_t, t \geq 0\}$. Look back at Exercise 6.6; you have already verified that these are martingales. Of course, using Itô's formula makes for a much easier proof.

Exercise 15.6. Prove that the process $\{M_t, t \geq 0\}$ defined by setting

$$M_t = \exp\left\{\theta B_t - \frac{\theta^2 t}{2}\right\}$$

where $\theta \in \mathbb{R}$ is a constant is a martingale with respect to the Brownian filtration $\{\mathcal{F}_t, t \geq 0\}$.

Example 15.7. We will now show that Theorem 10.2, the integration-by-parts formula for Wiener integrals, is a special case of the generalized version of Itô's formula. Suppose that $g : [0,\infty) \to \mathbb{R}$ is a bounded, continuous function in $L^2([0,\infty))$. Suppose further that g is differentiable with g' also bounded and continuous. Let $f(t,x) = xg(t)$ so that

$$f'(t,x) = g(t), \quad f''(t,x) = 0, \quad \text{and} \quad \dot{f}(t,x) = xg'(t).$$

Therefore, the generalized version of Itô's formula implies

$$g(t)B_t = \int_0^t g(s)\,\mathrm{d}B_s + \frac{1}{2}\int_0^t 0\,\mathrm{d}s + \int_0^t g'(s)B_s\,\mathrm{d}s.$$

Rearranging gives

$$\int_0^t g(s)\,\mathrm{d}B_s = g(t)B_t - \int_0^t g'(s)B_s\,\mathrm{d}s$$

as required.

There are a number of versions of Itô's formula that we will use; the first two we have already seen. The easiest way to remember all of the different versions is as a *stochastic differential equation* (or SDE).

Theorem 15.8 (Version I). *If $f \in C^2(\mathbb{R})$, then*

$$\mathrm{d}f(B_t) = f'(B_t)\,\mathrm{d}B_t + \frac{1}{2}f''(B_t)\,\mathrm{d}t.$$

Theorem 15.9 (Version II). *If $f \in C^1([0,\infty)) \times C^2(\mathbb{R})$, then*

$$\begin{aligned}\mathrm{d}f(t,B_t) &= f'(t,B_t)\,\mathrm{d}B_t + \frac{1}{2}f''(t,B_t)\,\mathrm{d}t + \dot{f}(t,B_t)\,\mathrm{d}t \\ &= f'(t,B_t)\,\mathrm{d}B_t + \left[\dot{f}(t,B_t) + \frac{1}{2}f''(t,B_t)\right]\,\mathrm{d}t.\end{aligned}$$

Example 15.10. Suppose that $\{B_t, t \geq 0\}$ is a standard Brownian motion. Determine the SDE satisfied by

$$X_t = \exp\{\sigma B_t + \mu t\}.$$

Solution. Consider the function $f(t,x) = \exp\{\sigma x + \mu t\}$. Since

$$f'(t,x) = \sigma\exp\{\sigma x + \mu t\}, \quad f''(t,x) = \sigma^2\exp\{\sigma x + \mu t\}, \quad \dot{f}(t,x) = \mu\exp\{\sigma x + \mu t\},$$

it follows from Version II of Itô's formula that

$$\mathrm{d}f(t,B_t) = \sigma\exp\{\sigma B_t + \mu t\}\,\mathrm{d}B_t + \frac{\sigma^2}{2}\exp\{\sigma B_t + \mu t\}\,\mathrm{d}t + \mu\exp\{\sigma B_t + \mu t\}\,\mathrm{d}t.$$

In other words,

$$\mathrm{d}X_t = \sigma X_t \,\mathrm{d}B_t + \left(\frac{\sigma^2}{2} + \mu\right) X_t \,\mathrm{d}t.$$

Suppose that the stochastic process $\{X_t, t \geq 0\}$ is defined by the stochastic differential equation

$$\mathrm{d}X_t = a(t, X_t)\,\mathrm{d}B_t + b(t, X_t)\,\mathrm{d}t$$

where a and b are *suitably smooth* functions. We call such a stochastic process a *diffusion* (or an *Itô diffusion* or an *Itô process*).

Again, a careful analysis of Taylor's theorem provides a version of Itô's formula for diffusions.

Theorem 15.11 (Version III). *Let X_t be a diffusion defined by the SDE*

$$\mathrm{d}X_t = a(t, X_t)\,\mathrm{d}B_t + b(t, X_t)\,\mathrm{d}t.$$

If $f \in C^2(\mathbb{R})$, then

$$\mathrm{d}f(X_t) = f'(X_t)\,\mathrm{d}X_t + \frac{1}{2} f''(X_t)\,\mathrm{d}\langle X\rangle_t$$

where $\mathrm{d}\langle X\rangle_t$ is computed as

$$\mathrm{d}\langle X\rangle_t = (\mathrm{d}X_t)^2 = [a(t, X_t)\,\mathrm{d}B_t + b(t, X_t)\,\mathrm{d}t]^2 = a^2(t, X_t)\,\mathrm{d}t$$

using the rules $(\mathrm{d}B_t)^2 = \mathrm{d}t$, $(\mathrm{d}t)^2 = 0$, $(\mathrm{d}B_t)(\mathrm{d}t) = (\mathrm{d}t)(\mathrm{d}B_t) = 0$. That is,

$$\begin{aligned}\mathrm{d}f(X_t) &= f'(X_t)\,[a(t, X_t)\,\mathrm{d}B_t + b(t, X_t)\,\mathrm{d}t] + \frac{1}{2} f''(X_t) a^2(t, X_t)\,\mathrm{d}t \\ &= f'(X_t) a(t, X_t)\,\mathrm{d}B_t + f'(X_t) b(t, X_t)\,\mathrm{d}t + \frac{1}{2} f''(X_t) a^2(t, X_t)\,\mathrm{d}t \\ &= f'(X_t) a(t, X_t)\,\mathrm{d}B_t + \left[f'(X_t) b(t, X_t) + \frac{1}{2} f''(X_t) a^2(t, X_t)\right]\,\mathrm{d}t.\end{aligned}$$

And finally we give the version of Itô's formula for diffusions for functions $f(t, x)$ of two variables.

Theorem 15.12 (Version IV). *Let X_t be a diffusion defined by the SDE*

$$\mathrm{d}X_t = a(t, X_t)\,\mathrm{d}B_t + b(t, X_t)\,\mathrm{d}t.$$

If $f \in C^1([0, \infty)) \times C^2(\mathbb{R})$, then

$$\begin{aligned}\mathrm{d}f(t, X_t) &= f'(t, X_t)\,\mathrm{d}X_t + \frac{1}{2} f''(t, X_t)\,\mathrm{d}\langle X\rangle_t + \dot{f}(t, X_t)\,\mathrm{d}t \\ &= f'(t, X_t) a(t, X_t)\,\mathrm{d}B_t + \left[\dot{f}(t, X_t) + f'(t, X_t) b(t, X_t) + \frac{1}{2} f''(t, X_t) a^2(t, X_t)\right]\,\mathrm{d}t\end{aligned}$$

again computing $\mathrm{d}\langle X\rangle_t = (\mathrm{d}X_t)^2$ using the rules $(\mathrm{d}B_t)^2 = \mathrm{d}t$, $(\mathrm{d}t)^2 = 0$, and $(\mathrm{d}B_t)(\mathrm{d}t) = (\mathrm{d}t)(\mathrm{d}B_t) = 0$.

16

Deriving the Black–Scholes Partial Differential Equation

Our goal for today is to use Itô's formula to derive the Black-Scholes partial differential equation. We will then solve this equation next lecture.

Recall from Lecture #2 that $D(t)$ denotes the value at time t of an investment which grows according to a continuously compounded interest rate r. We know its value at time $t \geq 0$ is given by $D(t) = e^{rt} D_0$ which is the solution to the differential equation $D'(t) = rD(t)$ with initial condition $D(0) = D_0$. Written in differential form, this becomes

$$\mathrm{d}D(t) = rD(t)\,\mathrm{d}t. \tag{16.1}$$

We now assume that our stock price is modelled by geometric Brownian motion. That is, let S_t denote the price of the stock at time t, and assume that S_t satisfies the stochastic differential equation

$$\mathrm{d}S_t = \sigma S_t\,\mathrm{d}B_t + \mu S_t\,\mathrm{d}t. \tag{16.2}$$

We can check using Version II of Itô's formula (Theorem 15.9) that the solution to this SDE is geometric Brownian motion $\{S_t, t \geq 0\}$ given by

$$S_t = S_0 \exp\left\{\sigma B_t + \left(\mu - \frac{\sigma^2}{2}\right) t\right\}$$

where S_0 is the initial value.

Remark. There are two, equally common, ways to parametrize the drift of the geometric Brownian motion. The first is so that the process is simpler,

$$S_t = S_0 \exp\left\{\sigma B_t + \mu t\right\},$$

and leads to the more complicated SDE

$$\mathrm{d}S_t = \sigma S_t\,\mathrm{d}B_t + \left(\mu + \frac{\sigma^2}{2}\right) S_t\,\mathrm{d}t.$$

The second is so that the SDE is simpler,

$$\mathrm{d}S_t = \sigma S_t\,\mathrm{d}B_t + \mu S_t\,\mathrm{d}t,$$

and leads to the more complicated process

$$S_t = S_0 \exp\left\{\sigma B_t + \left(\mu - \frac{\sigma^2}{2}\right) t\right\}.$$

We choose the parametrization given by (16.2) to be consistent with Higham [12].

We also recall from Lecture #1 the definition of a European call option.

Definition 16.1. A *European call option* with strike price E at time T gives its holder an opportunity (i.e., the right, but not the obligation) to buy from the writer one share of the prescribed stock at time T for price E.

Notice that if, at time T, the value of the stock is less than E, then the option is worthless and will not be exercised, but if the value of the stock is greater than E, then the option is valuable and will therefore be exercised.

That is,

(i) if $S_T \le E$, then the option is worthless, but
(ii) if $S_T > E$, then the option has the value $S_T - E$.

Thus, the value of the option at time T is $(S_T - E)^+ = \max\{0, S_T - E\}$. Our goal, therefore, is to determine the value of this option at time 0.

We will write V to denote the value of the option. Since V depends on both time and on the underlying stock, we see that $V(t, S_t)$ denotes the value of the option at time t, $0 \le t \le T$.

Hence,

(i) $V(T, S_T) = (S_T - E)^+$ is the value of the option at the expiry time T, and
(ii) $V(0, S_0)$ denotes the value of option at time 0.

Example 16.2. Assuming that the function $V \in C^1([0,\infty)) \times C^2(\mathbb{R})$, use Itô's formula on $V(t, S_t)$ to compute $\mathrm{d}V(t, S_t)$.

Solution. By Version IV of Itô's formula (Theorem 15.12), we find

$$\mathrm{d}V(t, S_t) = \dot V(t, S_t)\,\mathrm{d}t + V'(t, S_t)\,\mathrm{d}S_t + \frac{1}{2}V''(t, S_t)\,\mathrm{d}\langle S\rangle_t.$$

From (16.2), the SDE for geometric Brownian motion is

$$\mathrm{d}S_t = \sigma S_t\,\mathrm{d}B_t + \mu S_t\,\mathrm{d}t$$

and so we find

$$\mathrm{d}\langle S\rangle_t = (\mathrm{d}S_t)^2 = \sigma^2 S_t^2\,\mathrm{d}t$$

using the rules $(\mathrm{d}B_t)^2 = \mathrm{d}t$, $(\mathrm{d}t)^2 = (\mathrm{d}B_t)(\mathrm{d}t) = (\mathrm{d}t)(\mathrm{d}B_t) = 0$. Hence, we conclude

$$\begin{aligned}\mathrm{d}V(t, S_t) &= \dot V(t, S_t)\,\mathrm{d}t + V'(t, S_t)\big[\sigma S_t\,\mathrm{d}B_t + \mu S_t\,\mathrm{d}t\big] + \frac{1}{2}V''(t, S_t)\big[\sigma^2 S_t^2\,\mathrm{d}t\big]\\ &= \sigma S_t V'(t, S_t)\,\mathrm{d}B_t + \left[\dot V(t, S_t) + \mu S_t V'(t, S_t) + \frac{\sigma^2}{2}S_t^2 V''(t, S_t)\right]\mathrm{d}t. \qquad (16.3)\end{aligned}$$

We now recall the no arbitrage assumption from Lecture #2 which states that "there is never an opportunity to make a risk-free profit that gives a greater return than that provided by interest from a bank deposit."

Thus, to find the fair value of the option $V(t, S_t)$, $0 \le t \le T$, we will set up a *replicating portfolio* of assets and bonds that has precisely the *same risk* at time t as the option does at time t. The portfolio consists of a cash deposit D and a number A of assets.

We assume that we can vary the number of assets and the size of our cash deposit at time t so that both D and A are allowed to be functions of both the time t and the asset price S_t. (Technically, our *trading strategy* needs to be previsible; we can only alter our portfolio depending on what has happened already.)

That is, if Π denotes our portfolio, then the value of our portfolio at time t is given by

$$\Pi(t, S_t) = A(t, S_t)S_t + D(t, S_t). \tag{16.4}$$

Recall that we are allowed to short-sell both the stocks and the bonds and that there are no transaction costs involved. Furthermore, it is worth noting that, although our strategy for buying bonds may depend on both the time and the behaviour of the stock, the bond is *still* a risk-free investment which evolves according to (16.1) as

$$\mathrm{d}D(t, S_t) = rD(t, S_t)\,\mathrm{d}t. \tag{16.5}$$

The assumption that the portfolio is replicating means precisely that the portfolio is *self-financing*; in other words, the value of the portfolio one time step later is financed entirely by the current wealth. In terms of stochastic differentials, the self-financing condition is

$$\mathrm{d}\Pi(t, S_t) = A(t, S_t)\,\mathrm{d}S_t + \mathrm{d}D(t, S_t),$$

which, using (16.2) and (16.5), is equivalent to

$$\begin{aligned}\mathrm{d}\Pi(t, S_t) &= A(t, S_t)\big[\sigma S_t\,\mathrm{d}B_t + \mu S_t\,\mathrm{d}t\big] + rD(t, S_t)\,\mathrm{d}t \\ &= \sigma A(t, S_t)S_t\,\mathrm{d}B_t + \big[\mu A(t, S_t)S_t + rD(t, S_t)\big]\,\mathrm{d}t. \end{aligned} \tag{16.6}$$

The final step is to consider $V(t, S_t) - \Pi(t, S_t)$. By the no arbitrage assumption, the change in $V(t, S_t) - \Pi(t, S_t)$ over any time step is non-random. Furthermore, it must equal the corresponding growth offered by the continuously compounded risk-free interest rate. In terms of differentials, if we write

$$U_t = V(t, S_t) - \Pi(t, S_t),$$

then U_t must be non-random and grow according to (16.1) so that

$$\mathrm{d}U_t = rU_t\,\mathrm{d}t.$$

That is,

$$\mathrm{d}\big[V(t, S_t) - \Pi(t, S_t)\big] = r\big[V(t, S_t) - \Pi(t, S_t)\big]\,\mathrm{d}t. \tag{16.7}$$

The logic is outlined by Higham [12] on page 79.

Using (16.3) for $\mathrm{d}V(t, S_t)$ and (16.6) for $\mathrm{d}\Pi(t, S_t)$, we find

$$\begin{aligned}
&\mathrm{d}\big[V(t,S_t) - \Pi(t,S_t)\big] \\
&\quad = \left(\sigma S_t V'(t,S_t)\,\mathrm{d}B_t + \left[\dot{V}(t,S_t) + \mu S_t V'(t,S_t) + \frac{\sigma^2}{2}S_t^2 V''(t,S_t)\right]\mathrm{d}t\right) \\
&\qquad\quad - \left(\sigma A(t,S_t)S_t\,\mathrm{d}B_t + \big[\mu A(t,S_t)S_t + rD(t,S_t)\big]\,\mathrm{d}t\right) \\
&\quad = \sigma S_t\big[V'(t,S_t) - A(t,S_t)\big]\,\mathrm{d}B_t \\
&\qquad\quad + \left[\dot{V}(t,S_t) + \mu S_t V'(t,S_t) + \frac{\sigma^2}{2}S_t^2 V''(t,S_t) - \mu A(t,S_t)S_t - rD(t,S_t)\right]\mathrm{d}t \\
&\quad = \sigma S_t\big[V'(t,S_t) - A(t,S_t)\big]\,\mathrm{d}B_t \\
&\qquad\quad + \left[\dot{V}(t,S_t) + \frac{\sigma^2}{2}S_t^2 V''(t,S_t) - rD(t,S_t) + \mu S_t\big[V'(t,S_t) - A(t,S_t)\big]\right]\mathrm{d}t. \qquad (16.8)
\end{aligned}$$

Since we assume that the change over any time step is non-random, it must be the case that the $\mathrm{d}B_t$ term is 0. In order for the $\mathrm{d}B_t$ term to be 0, we simply choose

$$A(t,S_t) = V'(t,S_t).$$

This means that that $\mathrm{d}t$ term

$$\dot{V}(t,S_t) + \frac{\sigma^2}{2}S_t^2 V''(t,S_t) - rD(t,S_t) + \mu S_t\big[V'(t,S_t) - A(t,S_t)\big]$$

reduces to

$$\dot{V}(t,S_t) + \frac{\sigma^2}{2}S_t^2 V''(t,S_t) - rD(t,S_t)$$

since we already need $A(t, S_t) = V'(t, S_t)$ for the $\mathrm{d}B_t$ piece. Looking at (16.7) therefore gives

$$\dot{V}(t,S_t) + \frac{\sigma^2}{2}S_t^2 V''(t,S_t) - rD(t,S_t) = r\big[V(t,S_t) - \Pi(t,S_t)\big]. \qquad (16.9)$$

Using the facts that

$$\Pi(t,S_t) = A(t,S_t)S_t + D(t,S_t)$$

and

$$A(t,S_t) = V'(t,S_t)$$

therefore imply that (16.9) becomes

$$\dot{V}(t,S_t) + \frac{\sigma^2}{2}S_t^2 V''(t,S_t) - rD(t,S_t) = rV(t,S_t) - rS_tV'(t,S_t) - rD(t,S_t)$$

which, upon simplification, reduces to

$$\dot{V}(t,S_t) + \frac{\sigma^2}{2}S_t^2 V''(t,S_t) + rS_tV'(t,S_t) - rV(t,S_t) = 0.$$

In other words, we must find a function $V(t,x)$ which satisfies the *Black-Scholes partial differential equation*

$$\dot{V}(t,x) + \frac{\sigma^2}{2} x^2 V''(t,x) + rxV'(t,x) - rV(t,x) = 0. \tag{16.10}$$

Remark. We have finally arrived at what Higham [12] calls "the famous *Black-Scholes* partial differential equation (PDE)" given by equation (8.15) on page 79.

We now mention two important points.

(i) The drift parameter μ in the asset model does NOT appear in the Black-Scholes PDE.
(ii) Actually, we have not yet specified what type of option is being valued. The PDE given in (16.10) must be satisfied by ANY option on the asset S whose value can be expressed as a smooth function, i.e., a function in $C^1([0,\infty)) \times C^2(\mathbb{R})$.

In view of the second item, we really want to price a European call option with strike price E. This amounts to requiring $V(T, S_T) = (S_T - E)^+$. Our goal, therefore, in the next lecture is to solve the Black-Scholes partial differential equation

$$\dot{V}(t,x) + \frac{\sigma^2}{2} x^2 V''(t,x) + rxV'(t,x) - rV(t,x) = 0$$

for $V(t,x)$, $0 \le t \le T$, $x \in \mathbb{R}$, *subject to the boundary condition*

$$V(T,x) = (x-E)^+.$$

17
Solving the Black–Scholes Partial Differential Equation

Our goal for this lecture is to solve the Black-Scholes partial differential equation

$$\dot{V}(t,x) + \frac{\sigma^2}{2}x^2V''(t,x) + rxV'(t,x) - rV(t,x) = 0 \tag{17.1}$$

for $V(t,x)$, $0 \le t \le T$, $x \in \mathbb{R}$, subject to the boundary condition

$$V(T,x) = (x-E)^+.$$

The first observation is that it suffices to solve (17.1) when $r = 0$. That is, if W satisfies

$$\dot{W}(t,x) + \frac{\sigma^2}{2}x^2W''(t,x) = 0, \tag{17.2}$$

and $V(t,x) = e^{r(t-T)}W(t, e^{r(T-t)}x)$, then $V(t,x)$ satisfies (17.1) and $V(T,x) = W(T,x)$.

This can be checked by differentiation. There is, however, an "obvious" reason why it is true, namely due to the *time value of money* mentioned in Lecture #2. If money invested in a cash deposit grows at continuously compounded interest rate r, then \$$x$ at time T is equivalent to \$$e^{r(t-T)}x$ at time t.

Exercise 17.1. Verify (using the multivariate chain rule) that if $W(t,x)$ satisfies (17.2) and $V(t,x) = e^{r(t-T)}W(t, e^{r(T-t)}x)$, then $V(t,x)$ satisfies (17.1) and $V(T,x) = W(T,x)$.

Since we have already seen that the Black-Scholes partial differential equation (17.1) does not depend on μ, we can assume that $\mu = 0$. We have also just shown that it suffices to solve (17.1) when $r = 0$. Therefore, we will use W to denote the Black-Scholes solution in the $r = 0$ case, i.e., the solution to (17.2), and we will then use V as the solution in the $r > 0$ case, i.e., the solution to (17.1), where

$$V(t,x) = e^{r(t-T)}W(t, e^{r(T-t)}x). \tag{17.3}$$

We now note from (16.3) that the SDE for $W(t,S_t)$ is

$$\mathrm{d}W(t,S_t) = \sigma S_t W'(t,S_t)\,\mathrm{d}B_t + \left[\dot{W}(t,S_t) + \mu S_t W'(t,S_t) + \frac{\sigma^2}{2}S_t^2W''(t,S_t)\right]\mathrm{d}t.$$

We are assuming that $\mu = 0$ so that

$$\mathrm{d}W(t, S_t) = \sigma S_t W'(t, S_t)\,\mathrm{d}B_t + \left[\dot{W}(t, S_t) + \frac{\sigma^2}{2} S_t^2 W''(t, S_t)\right] \mathrm{d}t.$$

We are also assuming that $W(t, x)$ satisfies the Black-Scholes PDE given by (17.2) which is exactly what is needed to make the $\mathrm{d}t$ term equal to 0. Thus, we have reduced the SDE for $W(t, S_t)$ to

$$\mathrm{d}W(t, S_t) = \sigma S_t W'(t, S_t)\,\mathrm{d}B_t.$$

We now have a stochastic differential equation with no $\mathrm{d}t$ term which means, using Theorem 13.6, that $W(t, S_t)$ is a martingale. Formally, if $M_t = W(t, S_t)$, then the stochastic process $\{M_t, t \geq 0\}$ is a martingale with respect to the Brownian filtration $\{\mathcal{F}_t, t \geq 0\}$.

Next, we use the fact that martingales have stable expectation (Theorem 5.3) to conclude that

$$\mathbb{E}(M_0) = \mathbb{E}(M_T).$$

Remark. The expiry date T is a *fixed* time (and not a random time). This allows us to use Theorem 5.3 directly.

Since we know the value of the European call option at time T is $W(T, S_T) = (S_T - E)^+$, we see that

$$M_T = W(T, S_T) = (S_T - E)^+.$$

Furthermore, $M_0 = W(0, S_0)$ is non-random (since S_0, the stock price at time 0, is known), and so we conclude that $M_0 = \mathbb{E}(M_T)$ which implies

$$W(0, S_0) = \mathbb{E}[\,(S_T - E)^+\,]. \tag{17.4}$$

The final step is to actually calculate the expected value in (17.4). Since we are assuming $\mu = 0$, the stock price follows geometric Brownian motion $\{S_t, t \geq 0\}$ where

$$S_t = S_0 \exp\left\{\sigma B_t - \frac{\sigma^2}{2} t\right\}.$$

Hence, at time T, we need to consider the random variable

$$S_T = S_0 \exp\left\{\sigma B_T - \frac{\sigma^2}{2} T\right\}.$$

We know $B_T \sim \mathcal{N}(0, T)$ so that we can write

$$S_T = S_0\, e^{-\frac{\sigma^2 T}{2}}\, e^{\sigma\sqrt{T} Z}$$

for $Z \sim \mathcal{N}(0, 1)$. Thus, we can now use the result of Exercise 3.7, namely if $a > 0$, $b > 0$, $c > 0$ are constants and $Z \sim \mathcal{N}(0, 1)$, then

$$\mathbb{E}[\,(ae^{bZ} - c)^+\,] = ae^{b^2/2}\,\Phi\left(b + \frac{1}{b}\log\frac{a}{c}\right) - c\,\Phi\left(\frac{1}{b}\log\frac{a}{c}\right), \tag{17.5}$$

with

$$a = S_0\, e^{-\frac{\sigma^2 T}{2}}, \quad b = \sigma\sqrt{T}, \quad c = E$$

to conclude

$$\begin{aligned}
&\mathbb{E}[\,(S_T - E)^+\,] \\
&\quad = S_0\, e^{-\frac{\sigma^2 T}{2}} e^{\frac{\sigma^2 T}{2}}\, \Phi\left(\sigma\sqrt{T} + \frac{1}{\sigma\sqrt{T}} \log \frac{S_0\, e^{-\frac{\sigma^2 T}{2}}}{E}\right) - E\,\Phi\left(\frac{1}{\sigma\sqrt{T}} \log \frac{S_0\, e^{-\frac{\sigma^2 T}{2}}}{E}\right) \\
&\quad = S_0\, \Phi\left(\frac{1}{\sigma\sqrt{T}} \log \frac{S_0}{E} + \frac{\sigma\sqrt{T}}{2}\right) - E\,\Phi\left(\frac{1}{\sigma\sqrt{T}} \log \frac{S_0}{E} - \frac{\sigma\sqrt{T}}{2}\right).
\end{aligned}$$

To account for the *time value of money*, we can use Exercise 17.1 to give the solution for $r > 0$. That is, if $V(0, S_0)$ denotes the fair price (at time 0) of a European call option with strike price E, then using (17.3) we conclude

$$\begin{aligned}
V(0, S_0) &= e^{-rT} W(0, e^{rT} S_0) \\
&= e^{-rT} e^{rT} S_0\, \Phi\left(\frac{1}{\sigma\sqrt{T}} \log \frac{e^{rT} S_0}{E} + \frac{\sigma\sqrt{T}}{2}\right) - E e^{-rT}\,\Phi\left(\frac{1}{\sigma\sqrt{T}} \log \frac{e^{rT} S_0}{E} - \frac{\sigma\sqrt{T}}{2}\right) \\
&= S_0\, \Phi\left(\frac{\log(S_0/E) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) - E e^{-rT}\,\Phi\left(\frac{\log(S_0/E) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) \\
&= S_0\, \Phi\,(d_1) - E e^{-rT}\, \Phi\,(d_2)
\end{aligned}$$

where

$$d_1 = \frac{\log(S_0/E) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \quad \text{and} \quad d_2 = \frac{\log(S_0/E) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}.$$

AWESOME!

Remark. We have now arrived at equation (8.19) on page 80 of Higham [12]. Note that Higham only *states* the answer; he never actually goes through the solution of the Black-Scholes PDE.

Summary. Let's summarize what we did. We assumed that the asset S followed geometric Brownian motion given by

$$\mathrm{d}S_t = \sigma S_t\, \mathrm{d}B_t + \mu S_t\, \mathrm{d}t,$$

and that the risk-free bond D grew at continuously compounded interest rate r so that

$$\mathrm{d}D(t, S_t) = rD(t, S_t)\, \mathrm{d}t.$$

Using Version IV of Itô's formula on the value of the option $V(t, S_t)$ combined with the self-financing portfolio implied by the no arbitrage assumption led to the Black-Scholes partial differential equation

$$\dot{V}(t, x) + \frac{\sigma^2}{2} x^2 V''(t, x) + rxV'(t, x) - rV(t, x) = 0.$$

We also made the important observation that this PDE does not depend on μ. We then saw that it was sufficient to consider $r = 0$ since we noted that if $W(t,x)$ solved the resulting PDE

$$\dot{W}(t,x) + \frac{\sigma^2}{2} x^2 W''(t,x) = 0,$$

then $V(t,x) = e^{r(t-T)} W(t, e^{r(T-t)}x)$ solved the Black-Scholes PDE for $r > 0$. We then assumed that $\mu = 0$ and we found the SDE for $W(t, S_t)$ which had only a $\mathrm{d}B_t$ term (and no $\mathrm{d}t$ term). Using the fact that Itô integrals are martingales implied that $\{W(t,S_t), t \geq 0\}$ was a martingale, and so the stable expectation property of martingales led to the equation

$$W(0, S_0) = \mathbb{E}(W(T, S_T)).$$

Since we knew that $V(T, S_T) = W(T, S_T) = (S_T - E)^+$ for a European call option, we could compute the resulting expectation. We then translated back to the $r > 0$ case via

$$V(0, S_0) = e^{-rT} W(0, e^{rT} S_0).$$

This previous observation is extremely important since it tells us precisely how to price European call options with different payoffs. In general, if the payoff function at time T is $\Lambda(x)$ so that

$$V(T,x) = W(T,x) = \Lambda(x),$$

then, since $\{W(t,S_t), t \geq 0\}$ is a martingale,

$$W(0,S_0) = \mathbb{E}(W(T,S_T)) = \mathbb{E}(\Lambda(S_T)).$$

By assuming that $\mu = 0$, we can write S_T as

$$S_T = S_0 \exp\left\{\sigma B_T - \frac{\sigma^2}{2} T\right\} = S_0\, e^{-\frac{\sigma^2 T}{2}}\, e^{\sigma\sqrt{T} Z}$$

with $Z \sim \mathcal{N}(0,1)$. Therefore, if Λ is sufficiently nice, then $\mathbb{E}(\Lambda(S_T))$ can be calculated explicitly, and we can use

$$V(0, S_0) = e^{-rT} W(0, e^{rT} S_0)$$

to determine the required fair price to pay at time 0.

In particular, we can follow this strategy to answer the following question posed at the end of Lecture #1.

Example 17.2. In the Black-Scholes world, price a European option with a payoff of $\max\{S_T^2 - K, 0\}$ at time T.

Solution. The required time 0 price is $V(0,S_0) = e^{-rT} W(0, e^{rT} S_0)$ where $W(0,S_0) = \mathbb{E}[\,(S_T^2 - K)^+\,]$. Since we can write

$$S_T^2 = S_0^2\, e^{-\sigma^2 T}\, e^{2\sigma\sqrt{T} Z}$$

with $Z \sim \mathcal{N}(0,1)$, we can use (17.5) with $a = S_0^2\, e^{-\sigma^2 T}$, $b = 2\sigma\sqrt{T}$, and $c = K$ to conclude

$$V(0,S_0) = S_0^2\, e^{(\sigma^2 + r)T}\, \Phi\left(\frac{\log(S_0^2/K) + (2r + 3\sigma^2)T}{2\sigma\sqrt{T}}\right) - Ke^{-rT}\, \Phi\left(\frac{\log(S_0^2/K) + (2r - \sigma^2)T}{2\sigma\sqrt{T}}\right).$$

18

The Greeks

Recall that if $V(0,S_0)$ denotes the fair price (at time 0) of a European call option with strike price E and expiry date T, then the Black-Scholes option valuation formula is

$$\begin{aligned} V(0,S_0) &= S_0\,\Phi\left(\frac{\log(S_0/E)+(r+\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) - Ee^{-rT}\,\Phi\left(\frac{\log(S_0/E)+(r-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) \\ &= S_0\,\Phi\left(d_1\right) - Ee^{-rT}\,\Phi\left(d_2\right) \end{aligned}$$

where

$$d_1 = \frac{\log(S_0/E)+(r+\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \quad \text{and} \quad d_2 = \frac{\log(S_0/E)+(r-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}.$$

We see that this formula depends on the initial price of the stock S_0, the expiry date T, the strike price E, the risk-free interest rate r, and the stock's volatility σ.

The partial derivatives of $V = V(0,S_0)$ with respect to these variables are extremely important in practice, and we will now compute them; for ease, we will write $S = S_0$. In fact, some of these partial derivatives are given special names and referred to collectively as "the Greeks":

(a) $\Delta = \dfrac{\partial V}{\partial S}$ (delta),

(b) $\Gamma = \dfrac{\partial^2 V}{\partial S^2} = \dfrac{\partial \Delta}{\partial S}$ (gamma),

(c) $\rho = \dfrac{\partial V}{\partial r}$ (rho),

(d) $\Theta = \dfrac{\partial V}{\partial T}$ (theta),

(e) vega $= \dfrac{\partial V}{\partial \sigma}$.

Note. Vega is not actually a Greek letter. Sometimes it is written as ν (which is the Greek letter nu).

Remark. On page 80 of [12], Higham changes from using $V(0,S_0)$ to denote the fair price at time 0 of a European call option with strike price E and expiry date T to using $C(0,S_0)$. Both notations seem to be widely used in the literature.

The financial use of each of "The Greeks" is as follows.

- Delta measures sensitivity to a small change in the price of the underlying asset.
- Gamma measures the rate of change of delta.
- Rho measures sensitivity to the applicable risk-free interest rate.
- Theta measures sensitivity to the passage of time. Sometimes the financial definition of Θ is
$$-\frac{\partial V}{\partial T}.$$
With this definition, if you are "long an option, then you are short theta."
- Vega measures sensitivity to volatility.

In addition to the five "Greeks" that we have just defined (which are in widespread use), there are many other partial derivatives that have been given special names.

- Lambda, the percentage change in the option value per unit change in the underlying asset price, is given by
$$\lambda = \frac{1}{V}\frac{\partial V}{\partial S} = \frac{\partial \log V}{\partial S}.$$
- Vega gamma, or volga, measures second-order sensitivity to volatility and is given by
$$\frac{\partial^2 V}{\partial \sigma^2}.$$
- Vanna measures cross-sensitivity of the option value with respect to change in the underlying asset price and the volatility and is given by
$$\frac{\partial^2 V}{\partial S \partial \sigma} = \frac{\partial \Delta}{\partial \sigma}.$$
It is also the sensitivity of delta to a unit change in volatility.
- Delta decay, or charm, given by
$$\frac{\partial^2 V}{\partial S \partial T} = \frac{\partial \Delta}{\partial T},$$
measures time decay of delta. (This can be important when hedging a position over the weekend.)
- Gamma decay, or colour, given by
$$\frac{\partial^3 V}{\partial S^2 \partial T},$$
measures the sensitivity of the charm to the underlying asset price.
- Speed, given by
$$\frac{\partial^3 V}{\partial S^3},$$
measures third-order sensitivity to the underlying asset price.

In order to actually perform all of the calculations of the Greeks, we need to recall that

$$\Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Furthermore, we observe that

$$\log\left(\frac{S\Phi'(d_1)}{Ee^{-rT}\Phi'(d_2)}\right) = 0 \tag{18.1}$$

which implies that

$$S\Phi'(d_1) - Ee^{-rT}\Phi'(d_2) = 0. \tag{18.2}$$

Exercise 18.1. Verify (18.1) and deduce (18.2).

Since

$$d_1 = \frac{\log(S/E) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}$$

we find

$$\frac{\partial d_1}{\partial S} = \frac{1}{S\sigma\sqrt{T}}, \quad \frac{\partial d_1}{\partial r} = \frac{\sqrt{T}}{\sigma}, \quad \frac{\partial d_1}{\partial \sigma} = \frac{\sigma^2 T - \left[\log(S/E) + (r + \frac{1}{2}\sigma^2)T\right]}{\sigma^2\sqrt{T}} = -\frac{d_2}{\sigma}, \quad \text{and}$$

$$\frac{\partial d_1}{\partial T} = \frac{-\log(S/E) + (r + \frac{1}{2}\sigma^2)T}{2\sigma T^{3/2}}.$$

Furthermore, since

$$d_2 = d_1 - \sigma\sqrt{T},$$

we conclude

$$\frac{\partial d_2}{\partial S} = \frac{1}{S\sigma\sqrt{T}}, \quad \frac{\partial d_2}{\partial r} = \frac{\sqrt{T}}{\sigma}, \quad \frac{\partial d_2}{\partial \sigma} = \frac{\partial d_1}{\partial \sigma} - \sqrt{T} = -\frac{d_2}{\sigma} - \sqrt{T}, \quad \text{and}$$

$$\frac{\partial d_2}{\partial T} = \frac{\partial d_1}{\partial T} - \frac{\sigma}{2\sqrt{T}} = \frac{-\log(S/E) + (r + \frac{1}{2}\sigma^2)T}{2\sigma T^{3/2}} - \frac{\sigma}{2\sqrt{T}} = \frac{-\log(S/E) + (r - \frac{1}{2}\sigma^2)T}{2\sigma T^{3/2}}.$$

(a) Delta. Since $V = S\,\Phi(d_1) - Ee^{-rT}\,\Phi(d_2)$, we find

$$\begin{aligned}
\Delta = \frac{\partial V}{\partial S} &= \Phi(d_1) + S\,\frac{\partial \Phi(d_1)}{\partial S} - Ee^{-rT}\,\frac{\partial \Phi(d_2)}{\partial S} \\
&= \Phi(d_1) + S\,\Phi'(d_1)\,\frac{\partial d_1}{\partial S} - Ee^{-rT}\,\Phi'(d_2)\,\frac{\partial d_2}{\partial S} \\
&= \Phi(d_1) + \frac{\Phi'(d_1)}{\sigma\sqrt{T}} - Ee^{-rT}\,\frac{\Phi'(d_2)}{S\sigma\sqrt{T}} \\
&= \Phi(d_1) + \frac{1}{S\sigma\sqrt{T}}\left[S\Phi'(d_1) - Ee^{-rT}\Phi'(d_2)\right] \\
&= \Phi(d_1)
\end{aligned}$$

where the last step follows from (18.2).

(b) Gamma. Since $\Delta = \Phi(d_1)$, we find

$$\Gamma = \frac{\partial^2 V}{\partial S^2} = \frac{\partial \Delta}{\partial S} = \Phi'(d_1)\,\frac{\partial d_1}{\partial S} = \frac{\Phi'(d_1)}{S\sigma\sqrt{T}}.$$

(c) Rho. Since $V = S\,\Phi\left(d_1\right) - Ee^{-rT}\,\Phi\left(d_2\right)$, we find

$$\begin{aligned}
\rho = \frac{\partial V}{\partial r} &= S\,\frac{\partial \Phi\left(d_1\right)}{\partial r} + ETe^{-rT}\,\Phi\left(d_2\right) - Ee^{-rT}\,\frac{\partial \Phi\left(d_2\right)}{\partial r} \\
&= S\,\Phi'\left(d_1\right)\,\frac{\partial d_1}{\partial r} + ETe^{-rT}\,\Phi\left(d_2\right) - Ee^{-rT}\,\Phi'\left(d_2\right)\,\frac{\partial d_2}{\partial r} \\
&= \frac{S\sqrt{T}}{\sigma}\,\Phi'\left(d_1\right) + ETe^{-rT}\,\Phi\left(d_2\right) - \frac{Ee^{-rT}\sqrt{T}}{\sigma}\,\Phi'\left(d_2\right) \\
&= \frac{\sqrt{T}}{\sigma}\left[S\Phi'(d_1) - Ee^{-rT}\Phi'(d_2)\right] + ETe^{-rT}\,\Phi\left(d_2\right) \\
&= ETe^{-rT}\,\Phi\left(d_2\right)
\end{aligned}$$

where, as before, the last step follows from (18.2).

(d) Theta. Since $V = S\,\Phi\left(d_1\right) - Ee^{-rT}\,\Phi\left(d_2\right)$, we find

$$\begin{aligned}
\Theta = \frac{\partial V}{\partial T} &= S\,\frac{\partial \Phi\left(d_1\right)}{\partial T} + Ere^{-rT}\,\Phi\left(d_2\right) - Ee^{-rT}\,\frac{\partial \Phi\left(d_2\right)}{\partial T} \\
&= S\,\Phi'\left(d_1\right)\,\frac{\partial d_1}{\partial T} + Ere^{-rT}\,\Phi\left(d_2\right) - Ee^{-rT}\,\Phi'\left(d_2\right)\,\frac{\partial d_2}{\partial T} \\
&= S\,\Phi'\left(d_1\right)\,\frac{\partial d_1}{\partial T} + Ere^{-rT}\,\Phi\left(d_2\right) - Ee^{-rT}\,\Phi'\left(d_2\right)\left[\frac{\partial d_1}{\partial T} - \frac{\sigma}{2\sqrt{T}}\right] \\
&= \left[S\,\Phi'\left(d_1\right) - Ee^{-rT}\,\Phi'\left(d_2\right)\right]\frac{\partial d_1}{\partial T} + Ere^{-rT}\,\Phi\left(d_2\right) + \frac{\sigma}{2\sqrt{T}}Ee^{-rT}\,\Phi'\left(d_2\right) \\
&= Ere^{-rT}\,\Phi\left(d_2\right) + \frac{\sigma}{2\sqrt{T}}Ee^{-rT}\,\Phi'\left(d_2\right)
\end{aligned}$$

where, as before, the last step follows from (18.2). However, (18.2) also implies that we can write Θ as

$$\Theta = Ere^{-rT}\,\Phi\left(d_2\right) + \frac{\sigma S}{2\sqrt{T}}\Phi'\left(d_1\right). \tag{18.3}$$

(e) Vega. Since $V = S\,\Phi\left(d_1\right) - Ee^{-rT}\,\Phi\left(d_2\right)$, we find

$$\begin{aligned}
\text{vega } = \frac{\partial V}{\partial \sigma} = S\,\frac{\partial \Phi\left(d_1\right)}{\partial \sigma} - Ee^{-rT}\,\frac{\partial \Phi\left(d_2\right)}{\partial \sigma} &= S\,\Phi'\left(d_1\right)\,\frac{\partial d_1}{\partial \sigma} - Ee^{-rT}\,\Phi'\left(d_2\right)\,\frac{\partial d_2}{\partial \sigma} \\
&= -\frac{d_2}{\sigma}\,S\,\Phi'\left(d_1\right) - \left(-\frac{d_2}{\sigma} - \sqrt{T}\right)E\,e^{-rT}\,\Phi'\left(d_2\right) \\
&= -\frac{d_2}{\sigma}\left[S\,\Phi'\left(d_1\right) - Ee^{-rT}\,\Phi'\left(d_2\right)\right] + \sqrt{T}\,E\,e^{-rT}\,\Phi'\left(d_2\right) \\
&= \sqrt{T}\,E\,e^{-rT}\,\Phi'\left(d_2\right)
\end{aligned}$$

where, as before, the last step follows from (18.2). However, (18.2) also implies that we can write vega as

$$\text{vega } = S\sqrt{T}\,\Phi'\left(d_1\right).$$

Remark. Our definition of Θ is slightly different than the one in Higham [12]. We are differentiating V with respect to the expiry date T as opposed to an arbitrary time t with $0 \le t \le T$. This accounts for the discrepancy in the minus signs in (10.5) of [12] and (18.3).

Exercise 18.2. Compute lambda, volga, vanna, charm, colour, and speed for the Black-Scholes option valuation formula for a European call option with strike price E.

We also recall the put-call parity formula for European call and put options from Lecture #2:

$$V(0, S_0) + Ee^{-rT} = P(0, S_0) + S_0. \tag{18.4}$$

Here $P = P(0, S_0)$ is the fair price (at time 0) of a European put option with strike price E.

Exercise 18.3. Using the formula (18.4), compute the Greeks for a European put option. That is, compute

$$\Delta = \frac{\partial P}{\partial S}, \quad \Gamma = \frac{\partial^2 P}{\partial S^2}, \quad \rho = \frac{\partial P}{\partial r}, \quad \Theta = \frac{\partial P}{\partial T}, \quad \text{and} \quad \text{vega} = \frac{\partial P}{\partial \sigma}.$$

Note that gamma and vega for a European put option with strike price E are the same as gamma and vega for a European call option with strike price E.

19

Implied Volatility

Recall that if $V(0, S_0)$ denotes the fair price (at time 0) of a European call option with strike price E and expiry date T, then the Black-Scholes option valuation formula is

$$\begin{aligned} V(0, S_0) &= S_0 \, \Phi\left(\frac{\log(S_0/E) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) - Ee^{-rT}\, \Phi\left(\frac{\log(S_0/E) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) \\ &= S_0 \, \Phi(d_1) - Ee^{-rT}\, \Phi(d_2) \end{aligned}$$

where

$$d_1 = \frac{\log(S_0/E) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \quad \text{and} \quad d_2 = \frac{\log(S_0/E) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}.$$

Suppose that at time 0 a first investor buys a European call option on the stock having initial price S_0 with strike price E and expiry date T. Of course, the fair price for the first investor to pay is $V(0, S_0)$.

Suppose that some time later, say at time t, a second investor wants to buy a European call option on the same stock with the same strike price E and the same expiry date T. What is the fair price for this second investor to pay at time t?

Since it is now time t, the value of the underlying stock, namely S_t, is known. The expiry date T is time $T - t$ away. Thus, we simply re-scale our original Black-Scholes solution so that t is the new time 0, the new initial price of the stock is S_t, and $T - t$ is the new expiry date. This implies that the fair price (at time t) of a European call option with strike price E and expiry date T is given by the Black-Scholes option valuation formula

$$\begin{aligned} &V(t, S_t) \\ &= S_t \, \Phi\left(\frac{\log(S_t/E) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}\right) - Ee^{-r(T-t)}\, \Phi\left(\frac{\log(S_t/E) + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}\right) \\ &= S_t \, \Phi(d_1) - Ee^{-r(T-t)}\, \Phi(d_2) \end{aligned}$$

where

$$d_1 = \frac{\log(S_t/E) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{(T-t)}} \text{ and } d_2 = \frac{\log(S_t/E) + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} = d_1 - \sigma\sqrt{T-t}.$$

In fact, the rigorous justification for this is exactly the same as in (17.4) except now we view $M_t = W(t, S_t)$ as non-random since S_t, the stock price at time t, is known at time t. Note that the formula for $V(t, S_t)$ holds for $0 \leq t \leq T$. In particular, for $t = 0$ we recover our original result.

Remark. Given the general Black-Scholes formula $V(t, S_t)$, $0 \leq t \leq T$, we can define Θ as the derivative of V with respect to t. (In Lecture #18, we defined Θ as the derivative of V with respect to the expiry date T.) With this revised definition, we compute

$$\Theta = \frac{\partial V}{\partial t} = -Ere^{-r(T-t)}\,\Phi\left(d_2\right) - \frac{\sigma S_t}{2\sqrt{T-t}}\Phi'\left(d_1\right)$$

as in (10.5) of [12]. Note that there is a sign difference between this result and (18.3). All of the other Greeks, namely delta, gamma, rho, and vega, are the same as in Lecture #18 except that T is replaced with $T - t$.

The practical advantage of the Black-Scholes formula $V(t, S_t)$ is that it allows for the fast and easy calculation of option prices. It is worth noting, however, that "exact" calculations are not actually possible since the formula is given in terms of Φ, the normal cumulative distribution function. In order to "evaluate" $\Phi\left(d_1\right)$ or $\Phi\left(d_2\right)$ one must resort to using a computer (or table of normal values). Computationally, it is quite easy to evaluate Φ to many decimal places accuracy; and so this is the reason that we say the Black-Scholes formulation gives an exact formula. (In fact, programs like MATLAB or MAPLE can easily give values of Φ accurate to 10 decimal places.)

However, the limitations of the Black-Scholes model are numerous. Assumptions such as the asset price process following a geometric Brownian motion (so that an asset price at a fixed time has a lognormal distribution), or that the asset's volatility is constant, are not justified by actual market data.

As such, one of the goals of modern finance is to develop more sophisticated models for the asset price process, and to then develop the necessary stochastic calculus to produce a "solution" to the pricing problem. Unfortunately, there is no other model that produces as compact a solution as Black-Scholes. This means that the "solution" to any other model involves numerical analysis—and often quite involved analysis at that.

Suppose, for the moment, that we assume that the Black-Scholes model is valid. In particular, assume that the stock price $\{S_t, t \geq 0\}$ follows geometric Brownian motion. The fair price $V(t, S_t)$ to pay at time t depends on the parameters S_t, E, $T - t$, r, and σ^2. Of these, only the asset volatility σ cannot be directly observed.

There are two distinct approaches to extracting a value of σ from market data. The first is known as *implied volatility* and is obtained by using a quoted option value to recover σ. The second is known as *historical volatility* and is essentially maximum likelihood estimation of the parameter σ.

We will discuss only implied volatility. For ease, we will focus on the time $t = 0$ case. Suppose that S_0, E, T, and r are all known, and consider $V(0, S_0)$. Since we are assuming that only σ is unknown, we will emphasis this by writing $V(\sigma)$.

Thus, if we have a quoted value of the option price, say V^*, then we want to solve the equation $V(\sigma) = V^*$ for σ.

We will now show there is a unique solution to this equation which will be denoted by σ^* so that $V(\sigma^*) = V^*$.

To begin, note that we are only interested in positive volatilities so that $\sigma \in [0, \infty)$. Furthermore, $V(\sigma)$ is continuous on $[0, \infty)$ with

$$\lim_{\sigma \to \infty} V(\sigma) = S_0 \quad \text{and} \quad \lim_{\sigma \to 0^+} V(\sigma) = \max\{S_0 - Ee^{-rT}, 0\}. \tag{19.1}$$

Recall that from Lecture #18 that

$$\text{vega } = \frac{\partial V}{\partial \sigma} = V'(\sigma) = S_0 \sqrt{T}\, \Phi'(d_1)\,.$$

Since

$$\Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

we immediately conclude that $V'(\sigma) > 0$.

The fact that $V(\sigma)$ is continuous on $[0, \infty)$ with $V'(\sigma) > 0$ implies that $V(\sigma)$ is strictly increasing on $[0, \infty)$. Thus, we see that $V(\sigma) = V^*$ has a solution if and only if

$$\max\{S_0 - Ee^{-rT}, 0\} \le V^* \le S_0 \tag{19.2}$$

and that if a solution exists, then it must be unique. The no arbitrage assumption (i.e., the put-call parity) implies that the condition (19.2) always holds.

We now calculate

$$V''(\sigma) = \frac{\partial^2 V}{\partial \sigma^2} = \frac{d_1 d_2}{\sigma} \frac{\partial V}{\partial \sigma} = \frac{d_1 d_2}{\sigma} V'(\sigma) \tag{19.3}$$

which shows that the only inflection point of $V(\sigma)$ on $[0, \infty)$ is at

$$\hat{\sigma} = \sqrt{2 \left| \frac{\log(S_0/E) + rT}{T} \right|}. \tag{19.4}$$

Notice that we can write

$$V''(\sigma) = \frac{T}{4\sigma^3} (\hat{\sigma}^4 - \sigma^4) V'(\sigma) \tag{19.5}$$

which implies that $V(\sigma)$ is convex (i.e., concave up) for $\sigma < \hat{\sigma}$ and concave (i.e., concave down) for $\sigma > \hat{\sigma}$.

Exercise 19.1. Verify (19.1), (19.2), (19.3), (19.4), and (19.5).

The consequence of all of this is that *Newton's method* will be globally convergent for a suitably chosen initial value. Recall that Newton's method tells us that in order to solve the equation $F(x) = 0$, we consider the sequence of iterates $x_0, x_1, x_2, \ldots$ where

$$x_{n+1} = x_n - \frac{F(x_n)}{F'(x_n)}.$$

If we define

$$x^* = \lim_{n\to\infty} x_n,$$

then $F(x^*) = 0$. Of course, there are assumptions needed to ensure that Newton's method converges and produces the correct solution.

If we now consider $F(\sigma) = V(\sigma) - V^*$, then we have already shown that the conditions needed to guarantee that Newton's method converges have been satisfied.

It can also be shown that

$$0 < \frac{\sigma_{n+1} - \sigma^*}{\sigma_n - \sigma^*} < 1$$

for all n which implies that the error in the approximation is strictly decreasing as n increases. Thus, if we choose $\sigma_0 = \hat{\sigma}$, then the error must always converge to 0. Moreover, it can be shown that the convergence is quadratic. Thus, choosing $\sigma_0 = \hat{\sigma}$ is a foolproof (and deterministic) way of starting Newton's method. We can then stop iterating when our error is within some pre-specified tolerance, say $< 10^{-8}$.

Remark. Computing implied volatility using Newton's method is rather easy to implement in MATLAB. See, for instance, the program `ch14.m` from Higham [12].

Consider obtaining data that reports the option price V^* for a variety of values of the strike price E while at the same time holds r, S_0, and T fixed. An example of such data is presented in Section 14.5 of [12]. If the Black-Scholes formula were valid, then the volatility would be the same for each strike price. That is, the graph of strike price vs. implied volatility would be a horizontal line passing through $\sigma = \sigma^*$.

However, in this example, and in numerous other examples, the implied volatility curve appears to bend in the shape of either a *smile* or a *frown*.

Remark. More sophisticated analyses of implied volatility involve data that reports the option price V^* for a variety of values of the strike price E and expiry dates T while at the same time holding r and S_0 fixed. This produces a graph of strike price vs. expiry date vs. implied volatility, and the result is an *implied volatility surface*. The functional data analysis needed to in this case requires a number of statistical tools including *principal components analysis*. If the Black-Scholes formula were valid, then the resulting implied volatility surface would be a plane. Market data, however, typically results in bowl-shaped or hat-shaped surfaces. For details, see [5] which is also freely available online.

This implies, of course, that the Black-Scholes formula is not a perfect description of the option values that arise in practice. Many attempts have been made to "fix" this by considering stock price models that do not have constant volatility. We will investigate some such models over the next several lectures.

20

The Ornstein-Uhlenbeck Process as a Model of Volatility

The Ornstein-Uhlenbeck process is a diffusion process that was introduced as a model of the velocity of a particle undergoing Brownian motion. We know from Newtonian physics that the velocity of a (classical) particle in motion is given by the time derivative of its position. However, if the position of a particle is described by Brownian motion, then the time derivative does not exist. The Ornstein-Uhlenbeck process is an attempt to overcome this difficulty by modelling the velocity directly. Furthermore, just as Brownian motion is the scaling limit of simple random walk, the Ornstein-Uhlenbeck process is the scaling limit of the Ehrenfest urn model which describes the diffusion of particles through a permeable membrane.

In recent years, however, the Ornstein-Uhlenbeck process has appeared in finance as a model of the volatility of the underlying asset price process.

Suppose that the price of a stock $\{S_t, t \geq 0\}$ is modelled by geometric Brownian motion with volatility σ and drift μ so that S_t satisfies the SDE

$$\mathrm{d}S_t = \sigma S_t \,\mathrm{d}B_t + \mu S_t \,\mathrm{d}t.$$

However, market data indicates that implied volatilities for different strike prices and expiry dates of options are not constant. Instead, they appear to be *smile* shaped (or *frown* shaped).

Perhaps the most natural approach is to allow for the volatility $\sigma(t)$ to be a deterministic function of time so that S_t satisfies the SDE

$$\mathrm{d}S_t = \sigma(t) S_t \,\mathrm{d}B_t + \mu S_t \,\mathrm{d}t.$$

This was already suggested by Merton in 1973. Although it does explain the different implied volatility levels for different expiry dates, it does not explain the smile shape for different strike prices.

Instead, Hull and White in 1987 proposed to use a stochastic volatility model where the underlying stock price $\{S_t, t \geq 0\}$ satisfies the SDE

$$\mathrm{d}S_t = \sqrt{v_t}\, S_t \,\mathrm{d}B_t + \mu S_t \,\mathrm{d}t$$

and the variance process $\{v_t, t \geq 0\}$ is given by geometric Brownian motion

$$\mathrm{d}v_t = c_1 v_t \,\mathrm{d}B_t + c_2 v_t \,\mathrm{d}t$$

with c_1 and c_2 known constants. The problem with this model is that geometric Brownian motion tends to increase exponentially which is an undesirable property for volatility.

Market data also indicates that volatility exhibits mean-reverting behaviour. This lead Stein and Stein in 1991 to introduce the mean-reverting Ornstein-Uhlenbeck process satisfying

$$\mathrm{d}v_t = \sigma\,\mathrm{d}B_t + a(b - v_t)\,\mathrm{d}t$$

where a, b, and σ are known constants. This process, however, allows negative values of v_t.

In 1993 Heston overcame this difficulty by considering a more complex stochastic volatility model. Before investigating the Heston model, however, we will consider the Ornstein-Uhlenbeck process separately and prove that negative volatilities are allowed thereby verifying that the Stein and Stein stock price model is flawed.

We say that the process $\{X_t, t \geq 0\}$ is an *Ornstein-Uhlenbeck process* if X_t satisfies the *Ornstein-Uhlenbeck stochastic differential equation* given by

$$\mathrm{d}X_t = \sigma\,\mathrm{d}B_t + aX_t\,\mathrm{d}t \tag{20.1}$$

where σ and a are constants and $\{B_t, t \geq 0\}$ is a standard Brownian motion.

Remark. Sometimes (20.1) is called the *Langevin equation*, especially in physics contexts.

Remark. The Ornstein-Uhlenbeck SDE is very similar to the SDE for geometric Brownian motion; the only difference is the absence of X_t in the $\mathrm{d}B_t$ term of (20.1). However, this slight change makes (20.1) more challenging to solve.

The "trick" for solving (20.1) is to multiply both sides by the *integrating factor* e^{-at} and to compare with $\mathrm{d}(e^{-at}X_t)$. The chain rule tells us that

$$\mathrm{d}(e^{-at}X_t) = e^{-at}\,\mathrm{d}X_t + X_t\,\mathrm{d}(e^{-at}) = e^{-at}\,\mathrm{d}X_t - ae^{-at}X_t\,\mathrm{d}t \tag{20.2}$$

and multiplying (20.1) by e^{-at} gives

$$e^{-at}\mathrm{d}X_t = \sigma e^{-at}\,\mathrm{d}B_t + ae^{-at}X_t\,\mathrm{d}t \tag{20.3}$$

so that substituting (20.3) into (20.2) gives

$$\mathrm{d}(e^{-at}X_t) = \sigma e^{-at}\,\mathrm{d}B_t + ae^{-at}X_t\,\mathrm{d}t - ae^{-at}X_t\,\mathrm{d}t = \sigma e^{-at}\,\mathrm{d}B_t.$$

Since $\mathrm{d}(e^{-at}X_t) = \sigma e^{-at}\mathrm{d}B_t$, we can now integrate to conclude that

$$e^{-at}X_t - X_0 = \sigma \int_0^t e^{-as}\,\mathrm{d}B_s$$

and so

$$X_t = e^{at}X_0 + \sigma \int_0^t e^{a(t-s)}\,\mathrm{d}B_s. \tag{20.4}$$

Observe that the integral in (20.4) is a Wiener integral. Definition 9.1 tells us that

$$\int_0^t e^{a(t-s)}\,\mathrm{d}B_s \sim \mathcal{N}\left(0, \int_0^t e^{2a(t-s)}\,\mathrm{d}s\right) = \mathcal{N}\left(0, \frac{e^{2at}-1}{2a}\right).$$

In particular, choosing $X_0 = x$ to be constant implies that

$$X_t = e^{at}x + \sigma \int_0^t e^{a(t-s)}\,\mathrm{d}B_s \sim \mathcal{N}\left(xe^{at}, \frac{\sigma^2(e^{2at}-1)}{2a}\right).$$

Actually, we can generalize this slightly. If we choose $X_0 \sim \mathcal{N}(x,\tau^2)$ independently of the Brownian motion $\{B_t, t \geq 0\}$, then Exercise 3.12 tells us that

$$\begin{aligned} X_t = e^{at}X_0 + \sigma \int_0^t e^{a(t-s)}\,\mathrm{d}B_s &\sim \mathcal{N}\left(xe^{at}, \tau^2 e^{2at} + \frac{\sigma^2(e^{2at}-1)}{2a}\right) \\ &= \mathcal{N}\left(xe^{at}, \left(\tau^2 + \frac{\sigma^2}{2a}\right)e^{2at} - \frac{\sigma^2}{2a}\right). \end{aligned}$$

Exercise 20.1. Suppose that $\{X_t, t \geq 0\}$ is an Ornstein-Uhlenbeck process given by (20.4) with $X_0 = 0$. If $s < t$, compute $\mathrm{Cov}(X_s, X_t)$.

We say that the process $\{X_t, t \geq 0\}$ is a *mean-reverting Ornstein-Uhlenbeck process* if X_t satisfies the SDE

$$\mathrm{d}X_t = \sigma\,\mathrm{d}B_t + (b - X_t)\,\mathrm{d}t \tag{20.5}$$

where σ and b are constants and $\{B_t, t \geq 0\}$ is a standard Brownian motion.

The trick for solving the mean-reverting Ornstein-Uhlenbeck process is similar. That is, we multiply by e^t and compare with $\mathrm{d}(e^t(b - X_t))$. The chain rule tells us that

$$\mathrm{d}(e^t(b - X_t)) = -e^t\,\mathrm{d}X_t + e^t(b - X_t)\,\mathrm{d}t \tag{20.6}$$

and multiplying (20.5) by e^t gives

$$e^t\mathrm{d}X_t = \sigma e^t\,\mathrm{d}B_t + e^t(b - X_t)\,\mathrm{d}t \tag{20.7}$$

so that substituting (20.7) into (20.6) gives

$$\mathrm{d}(e^t(b - X_t)) = -\sigma e^t\,\mathrm{d}B_t - e^t(b - X_t)\,\mathrm{d}t + e^t(b - X_t)\,\mathrm{d}t = -\sigma e^t\,\mathrm{d}B_t.$$

Since $\mathrm{d}(e^t(b - X_t)) = -\sigma e^t\,\mathrm{d}B_t$, we can now integrate to conclude that

$$e^t(b - X_t) - (b - X_0) = -\sigma \int_0^t e^s\,\mathrm{d}B_s$$

and so

$$X_t = (1 - e^{-t})b + e^{-t}X_0 + \sigma \int_0^t e^{s-t}\,\mathrm{d}B_s. \tag{20.8}$$

Exercise 20.2. Suppose that $X_0 \sim \mathcal{N}(x, \tau^2)$ is independent of $\{B_t, t \geq 0\}$. Determine the distribution of X_t given by (20.8).

Exercise 20.3. Use an appropriate integrating factor to solve the mean-reverting Ornstein-Uhlenbeck SDE considered by Stein and Stein, namely $\mathrm{d}X_t = \sigma\,\mathrm{d}B_t + a(b - X_t)\,\mathrm{d}t$. Assuming that $X_0 = x$ is constant, determine the distribution of X_t and conclude that $\mathbf{P}\{X_t < 0\} > 0$ for every $t > 0$. *Hint: X_t has a normal distribution.* This then explains our earlier claim that the Stein and Stein model is flawed.

As previous noted, Heston introduced a stochastic volatility model in 1993 that overcame this difficulty. Assume that the asset price process $\{S_t, t \ge 0\}$ satisfies the SDE

$$\mathrm{d}S_t = \sqrt{v_t}\, S_t \,\mathrm{d}B_t^{(1)} + \mu S_t \,\mathrm{d}t$$

where the variance process $\{v_t, t \ge 0\}$ satisfies

$$\mathrm{d}v_t = \sigma\sqrt{v_t}\,\mathrm{d}B_t^{(2)} + a(b - v_t)\,\mathrm{d}t \tag{20.9}$$

and the two driving Brownian motions $\{B_t^{(1)}, t \ge 0\}$ and $\{B_t^{(2)}, t \ge 0\}$ are correlated with rate ρ, i.e.,

$$\mathrm{d}\langle B^{(1)}, B^{(2)}\rangle_t = \rho\,\mathrm{d}t.$$

The $\sqrt{v_t}$ term in (20.9) is needed to guarantee positive volatility—when the process touches zero the stochastic part becomes zero and the non-stochastic part will push it up. The parameter a measures the speed of the mean-reversion, b is the average level of volatility, and σ is the volatility of volatility. Market data suggests that the correlation rate ρ is typically negative. The negative dependence between returns and volatility is sometimes called the *leverage effect.*

Heston's model involves a system of stochastic differential equations. The key tool for analyzing such a system is the multidimensional version of Itô's formula.

Theorem 20.4 (Version V). *Suppose that* $\{X_t, t \ge 0\}$ *and* $\{Y_t, t \ge 0\}$ *are diffusions defined by the stochastic differential equations*

$$\mathrm{d}X_t = a_1(t, X_t, Y_t)\,\mathrm{d}B_t^{(1)} + b_1(t, X_t, Y_t)\,\mathrm{d}t$$

and

$$\mathrm{d}Y_t = a_2(t, X_t, Y_t)\,\mathrm{d}B_t^{(2)} + b_2(t, X_t, Y_t)\,\mathrm{d}t,$$

respectively, where $\{B_t^{(1)}, t \ge 0\}$ *and* $\{B_t^{(2)}, t \ge 0\}$ *are each standard one-dimensional Brownian motions. If* $f \in C^1([0,\infty)) \times C^2(\mathbb{R}^2)$*, then*

$$\begin{aligned}\mathrm{d}f(t, X_t, Y_t) = \dot{f}(t, X_t, Y_t)\,\mathrm{d}t + f_1(t, X_t, Y_t)\,\mathrm{d}X_t + \frac{1}{2}f_{11}(t, X_t, Y_t)\,\mathrm{d}\langle X\rangle_t \\ + f_2(t, X_t, Y_t)\,\mathrm{d}Y_t + \frac{1}{2}f_{22}(t, X_t, Y_t)\,\mathrm{d}\langle Y\rangle_t + f_{12}(t, X_t, Y_t)\,\mathrm{d}\langle X, Y\rangle_t\end{aligned}$$

where the partial derivatives are defined as

$$\dot{f}(t,x,y) = \frac{\partial}{\partial t}f(t,x,y), \quad f_1(t,x,y) = \frac{\partial}{\partial x}f(t,x,y), \quad f_{11}(t,x,y) = \frac{\partial^2}{\partial x^2}f(t,x,y)$$

$$f_2(t,x,y) = \frac{\partial}{\partial y}f(t,x,y), \quad f_{22}(t,x,y) = \frac{\partial^2}{\partial y^2}f(t,x,y), \quad f_{12}(t,x,y) = \frac{\partial^2}{\partial x\partial y}f(t,x,y),$$

and $\mathrm{d}\langle X, Y\rangle_t$ *is computed according to the rule*

$$\mathrm{d}\langle X, Y\rangle_t = (\mathrm{d}X_t)(\mathrm{d}Y_t) = a_1(t, X_t, Y_t)a_2(t, X_t, Y_t)\,\mathrm{d}\langle B^{(1)}, B^{(2)}\rangle_t.$$

Remark. In a typical problem involving the multidimensional version of Itô's formula, the quadratic covariation process $\langle B^{(1)}, B^{(2)} \rangle_t$ will be specified. However, two particular examples are worth mentioning. If $B^{(1)} = B^{(2)}$, then $\mathrm{d}\langle B^{(1)}, B^{(2)} \rangle_t = \mathrm{d}t$, whereas if $B^{(1)}$ and $B^{(2)}$ are independent, then $\mathrm{d}\langle B^{(1)}, B^{(2)} \rangle_t = 0$.

Exercise 20.5. Suppose that $f(t, x, y) = xy$. Using Version V of Itô's formula (Theorem 20.4), verify that the *product rule for diffusions* is given by

$$\mathrm{d}(X_t Y_t) = X_t \,\mathrm{d}Y_t + Y_t \,\mathrm{d}X_t + \mathrm{d}\langle X, Y \rangle_t.$$

Thus, our goal in the next few lectures is to price a European call option assuming that the underlying stock price follows Heston's model of geometric Brownian motion with a stochastic volatility, namely

$$\begin{cases} \mathrm{d}S_t = \sqrt{v_t}\, S_t \,\mathrm{d}B_t^{(1)} + \mu S_t \,\mathrm{d}t, \\ \mathrm{d}v_t = \sigma \sqrt{v_t} \,\mathrm{d}B_t^{(2)} + a(b - v_t)\,\mathrm{d}t, \\ \mathrm{d}\langle B^{(1)}, B^{(2)} \rangle_t = \rho \,\mathrm{d}t. \end{cases}$$

21

The Characteristic Function for a Diffusion

Recall that the characteristic function of a random variable X is the function $\varphi_X : \mathbb{R} \to \mathbb{C}$ defined by $\varphi_X(\theta) = \mathbb{E}(e^{i\theta X})$. From Exercise 3.9, if $X \sim \mathcal{N}(\mu, \sigma^2)$, then the characteristic function of X is

$$\varphi_X(\theta) = \exp\left\{ i\mu\theta - \frac{\sigma^2\theta^2}{2} \right\}.$$

Suppose that $\{X_t, t \geq 0\}$ is a stochastic process. For each $T \geq 0$, we know that X_T is a random variable. Thus, we can consider $\varphi_{X_T}(\theta)$.

In the particular case that $\{X_t, t \geq 0\}$ is a diffusion defined by the stochastic differential equation

$$\mathrm{d}X_t = \sigma(t, X_t)\,\mathrm{d}B_t + \mu(t, X_t)\,\mathrm{d}t \tag{21.1}$$

where $\{B_t, t \geq 0\}$ is a standard Brownian motion with $B_0 = 0$, if we can solve the SDE, then we can determine $\varphi_{X_T}(\theta)$ for any $T \geq 0$.

Example 21.1. Consider the case when both coefficients in (21.1) are constant so that

$$\mathrm{d}X_t = \sigma\,\mathrm{d}B_t + \mu\,\mathrm{d}t$$

where $\{B_t, t \geq 0\}$ is a standard Brownian motion with $B_0 = 0$, In this case, the SDE is trivial to solve. If $X_0 = x$ is constant, then for any $T \geq 0$, we have

$$X_T = x + \sigma B_T + \mu T$$

which is simply arithmetic Brownian motion started at x. Therefore,

$$X_T \sim \mathcal{N}(x + \mu T, \sigma^2 T)$$

so that

$$\varphi_{X_T}(\theta) = \exp\left\{ i(x + \mu T)\theta - \frac{\sigma^2 T\theta^2}{2} \right\}.$$

Example 21.2. Consider the Ornstein-Uhlenbeck stochastic differential equation given by

$$\mathrm{d}X_t = \sigma\,\mathrm{d}B_t + aX_t\,\mathrm{d}t$$

where σ and a are constants. As we saw in Lecture #20, if $X_0 = x$ is constant, then for any $T \geq 0$, we have

$$X_T = e^{aT}x + \sigma \int_0^T e^{a(T-s)}\,\mathrm{d}B_s \sim \mathcal{N}\left(xe^{aT}, \frac{\sigma^2(e^{2aT}-1)}{2a}\right).$$

Therefore,

$$\varphi_{X_T}(\theta) = \exp\left\{ixe^{aT}\theta - \frac{\sigma^2(e^{2aT}-1)\theta^2}{4a}\right\}.$$

Now it might seem like the only way to determine the characteristic function $\varphi_{X_T}(\theta)$ if $\{X_t, t \geq 0\}$ is a diffusion defined by (21.1) is to solve this SDE. Fortunately, this is not true. In many cases, the characteristic function for a diffusion defined by a SDE can be found using the *Feynman-Kac representation theorem* without actually solving the SDE.

Consider the diffusion

$$\mathrm{d}X_t = \sigma(t, X_t)\,\mathrm{d}B_t + \mu(t, X_t)\,\mathrm{d}t. \tag{21.2}$$

We know from Version IV of Itô's formula (Theorem 15.12) that if $f \in C^1([0,\infty)) \times C^2(\mathbb{R})$, then

$$\begin{aligned}\mathrm{d}f(t, X_t) &= f'(t, X_t)\,\mathrm{d}X_t + \frac{1}{2}f''(t, X_t)\,\mathrm{d}\langle X\rangle_t + \dot{f}(t, x)\,\mathrm{d}t \\ &= \sigma(t, X_t)f'(t, X_t)\,\mathrm{d}B_t + \left[\mu(t, X_t)f'(t, X_t) + \frac{1}{2}\sigma^2(t, X_t)f''(t, X_t) + \dot{f}(t, x)\right]\mathrm{d}t.\end{aligned}$$

We also know from Theorem 13.6 that any Itô integral is a martingale. Therefore, if we can find a particular function $f(t, x)$ such that the $\mathrm{d}t$ term is zero, then $f(t, X_t)$ will be a martingale. We define the *differential operator* (sometimes called the *generator* of the diffusion) to be the operator $\mathcal{A}$ given by

$$(\mathcal{A}f)(t, x) = \mu(t, x)f'(t, x) + \frac{1}{2}\sigma^2(t, x)f''(t, x) + \dot{f}(t, x).$$

Note that Version IV of Itô's formula now takes the form

$$\mathrm{d}f(t, X_t) = \sigma(t, X_t)f'(t, X_t)\,\mathrm{d}B_t + (\mathcal{A}f)(t, X_t)\,\mathrm{d}t.$$

This shows us the first connection between stochastic calculus and differential equations, namely that if $\{X_t, t \geq 0\}$ is a diffusion defined by (21.2) and if $f \in C^1([0,\infty)) \times C^2(\mathbb{R})$, then $f(t, X_t)$ is a martingale if and only if f satisfies the partial differential equation

$$(\mathcal{A}f)(t, x) = 0.$$

The Feynman-Kac representation theorem extends this idea by providing an explicit formula for the solution of this partial differential equation subject to certain boundary conditions.

Theorem 21.3 (Feynman-Kac Representation Theorem). *Suppose that $u \in C^2(\mathbb{R})$, and let $\{X_t, t \geq 0\}$ be defined by the SDE*

$$\mathrm{d}X_t = \sigma(t, X_t)\,\mathrm{d}B_t + \mu(t, X_t)\,\mathrm{d}t.$$

The unique bounded function $f : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ satisfying the partial differential equation

$$(\mathcal{A}f)(t,x) = \mu(t,x)f'(t,x) + \frac{1}{2}\sigma^2(t,x)f''(t,x) + \dot{f}(t,x) = 0, \quad 0 \le t \le T, \;\; x \in \mathbb{R},$$

subject to the terminal condition

$$f(T,x) = u(x), \quad x \in \mathbb{R},$$

is given by

$$f(t,x) = \mathbb{E}[u(X_T)|X_t = x].$$

Example 21.4. We will now use the Feynman-Kac representation theorem to derive the characteristic function for arithmetic Brownian motion satisfying the SDE

$$\mathrm{d}X_t = \sigma\,\mathrm{d}B_t + \mu\,\mathrm{d}t$$

where σ, μ, and $X_0 = x$ are constants. Let $u(x) = e^{i\theta x}$ so that the Feynman-Kac representation theorem implies

$$f(t,x) = \mathbb{E}[u(X_T)|X_t = x] = \mathbb{E}[e^{i\theta X_T}|X_t = x]$$

is the unique bounded solution of the partial differential equation

$$\mu f'(t,x) + \frac{1}{2}\sigma^2 f''(t,x) + \dot{f}(t,x) = 0, \quad 0 \le t \le T, \;\; x \in \mathbb{R}, \tag{21.3}$$

subject to the terminal condition

$$f(T,x) = e^{i\theta x}, \quad x \in \mathbb{R}.$$

Note that $f(0,x) = \mathbb{E}[e^{i\theta X_T}|X_0 = x] = \varphi_{X_T}(\theta)$ is the characteristic function of X_T.

In order to solve (21.3) we use separation of variables. That is, we guess that $f(t,x)$ can be written as a function of x only times a function of t only so that

$$f(t,x) = \chi(x)\tau(t), \quad 0 \le t \le T, \;\; x \in \mathbb{R}. \tag{21.4}$$

Therefore, we find

$$f'(t,x) = \chi'(x)\tau(t), \quad f''(t,x) = \chi''(x)\tau(t), \quad \dot{f}(t,x) = \chi(x)\tau'(t)$$

so that (21.3) implies

$$\mu\chi'(x)\tau(t) + \frac{1}{2}\sigma^2\chi''(x)\tau(t) + \chi(x)\tau'(t) = 0, \quad 0 \le t \le T, \;\; x \in \mathbb{R},$$

or equivalently,

$$\frac{\mu\chi'(x)}{\chi(x)} + \frac{\sigma^2\chi''(x)}{2\chi(x)} = -\frac{\tau'(t)}{\tau(t)}.$$

Since the left side of this equation which is a function of x only equals the right side which is a function of t only, we conclude that both sides must be constant. For ease, we will write the constant as $-\lambda^2$. Thus, we must solve the two ordinary differential equations

$$\frac{\mu\chi'(x)}{\chi(x)} + \frac{\sigma^2\chi''(x)}{2\chi(x)} = -\lambda^2 \quad \text{and} \quad -\frac{\tau'(t)}{\tau(t)} = -\lambda^2.$$

The ODE for τ is easy to solve; clearly $\tau'(t) = \lambda^2 \tau(t)$ implies

$$\tau(t) = C \exp\{\lambda^2 t\}$$

where C is an arbitrary constant. The ODE for χ is

$$\mu\chi'(x) + \frac{\sigma^2 \chi''(x)}{2} = -\lambda^2 \chi(x),$$

or equivalently,

$$\sigma^2 \chi''(x) + 2\mu\chi'(x) + 2\lambda^2 \chi(x) = 0. \tag{21.5}$$

Although this ODE is reasonably straightforward to solve for χ, it turns out that we do not need to actually solve it. This is because of our terminal condition. We know that

$$f(T, x) = e^{i\theta x}$$

and we also have assumed that

$$f(t, x) = \chi(x)\tau(t).$$

This implies that

$$f(T, x) = \chi(x)\tau(T) = e^{i\theta x}$$

which means that

$$\tau(T) = 1 \quad \text{and} \quad \chi(x) = e^{i\theta x}.$$

We now realize that we can solve for the arbitrary constant C; that is,

$$\tau(t) = C \exp\{\lambda^2 t\} \quad \text{and} \quad \tau(T) = 1$$

implies

$$C = \exp\{-\lambda^2 T\} \quad \text{so that} \quad \tau(t) = \exp\{-\lambda^2 (T - t)\}.$$

We are also in a position to determine the value of λ^2. That is, we know that $\chi(x) = e^{i\theta x}$ must be a solution to the ODE (21.5). Thus, we simply need to choose λ^2 so that this is true. Since

$$\chi'(x) = i\theta e^{i\theta x} \quad \text{and} \quad \chi''(x) = -\theta^2 e^{i\theta x},$$

we conclude that

$$-\sigma^2\theta^2 e^{i\theta x} + 2i\mu\theta e^{i\theta x} + 2\lambda^2 e^{i\theta x} = 0,$$

and so factoring out $e^{i\theta x}$ gives

$$-\sigma^2\theta^2 + 2i\mu\theta + 2\lambda^2 = 0.$$

Thus,

$$-\lambda^2 = i\mu\theta - \frac{\sigma^2\theta^2}{2}$$

so that substituting in for $\tau(t)$ gives

$$\tau(t) = \exp\left\{ i\mu(T - t)\theta - \frac{\sigma^2 (T - t)\theta^2}{2} \right\}$$

and so from (21.4) we conclude

$$f(t,x) = \chi(x)\tau(t) = e^{i\theta x}\exp\left\{i\mu(T-t)\theta - \frac{\sigma^2(T-t)\theta^2}{2}\right\}$$
$$= \exp\left\{i(x+\mu(T-t))\theta - \frac{\sigma^2(T-t)\theta^2}{2}\right\}.$$

Taking $t=0$ gives

$$\varphi_{X_T}(\theta) = f(0,x) = \exp\left\{i(x+\mu T)\theta - \frac{\sigma^2 T\theta^2}{2}\right\}$$

in agreement with Example 21.1.

Example 21.5. We will now use the Feynman-Kac representation theorem to derive the characteristic function for a process satisfying the Ornstein-Uhlenbeck SDE

$$\mathrm{d}X_t = \sigma\,\mathrm{d}B_t + aX_t\,\mathrm{d}t$$

where σ, a, and $X_0 = x$ are constants. Let $u(x) = e^{i\theta x}$ so that the Feynman-Kac representation theorem implies

$$f(t,x) = \mathbb{E}[u(X_T)|X_t = x] = \mathbb{E}[e^{i\theta X_T}|X_t = x]$$

is the unique bounded solution of the partial differential equation

$$axf'(t,x) + \frac{1}{2}\sigma^2 f''(t,x) + \dot{f}(t,x) = 0, \quad 0 \le t \le T, \;\; x \in \mathbb{R}, \tag{21.6}$$

subject to the terminal condition

$$f(T,x) = e^{i\theta x}, \quad x \in \mathbb{R}.$$

Note that $f(0,x) = \mathbb{E}[e^{i\theta X_T}|X_0 = x] = \varphi_{X_T}(\theta)$ is the characteristic function of X_T.

If we try to use separation of variables to solve (21.6), then we soon discover that it does not produce a solution. Thus, we are forced to conclude that the solution $f(t,x)$ is not separable and is necessarily more complicated. Guided by the form of the terminal condition, we guess that $f(t,x)$ can be written as

$$f(t,x) = \exp\{i\theta\alpha(t)x + \beta(t)\}, \quad 0 \le t \le T, \;\; x \in \mathbb{R}, \tag{21.7}$$

for some functions $\alpha(t)$ and $\beta(t)$ of t only satisfying $\alpha(T) = 1$ and $\beta(T) = 0$. Differentiating we find

$$f'(t,x) = i\theta\alpha(t)\exp\{i\theta\alpha(t)x + \beta(t)\} = i\theta\alpha(t)f(t,x),$$

$$f''(t,x) = -\theta^2\alpha^2(t)\exp\{i\theta\alpha(t)x + \beta(t)\} = -\theta^2\alpha^2(t)f(t,x), \quad \text{and}$$

$$\dot{f}(t,x) = [i\theta\alpha'(t)x + \beta'(t)]\exp\{i\theta\alpha(t)x + \beta(t)\} = [i\theta\alpha'(t)x + \beta'(t)]f(t,x)$$

so that (21.6) implies

$$i\theta ax\alpha(t)f(t,x) - \frac{\sigma^2\theta^2}{2}\alpha^2(t)f(t,x) + [i\theta\alpha'(t)x + \beta'(t)]f(t,x) = 0, \quad 0 \le t \le T, \;\; x \in \mathbb{R}.$$

Factoring out the common $f(t,x)$ reduces the equation to

$$i\theta a x\alpha(t) - \frac{\sigma^2\theta^2}{2}\alpha^2(t) + i\theta\alpha'(t)x + \beta'(t) = 0,$$

or equivalently,

$$i\theta[a\alpha(t) + \alpha'(t)]x + \beta'(t) - \frac{\sigma^2\theta^2}{2}\alpha^2(t) = 0.$$

Since this equation must be true for all $0 \le t \le T$ and $x \in \mathbb{R}$, the only way that is possible is if the coefficient of x is zero *and* the constant term is 0. Thus, we must have

$$a\alpha(t) + \alpha'(t) = 0 \quad \text{and} \quad \beta'(t) - \frac{\sigma^2\theta^2}{2}\alpha^2(t) = 0. \tag{21.8}$$

This first equation in (21.8) involves only $\alpha(t)$ and is easily solved. That is, $\alpha'(t) = -a\alpha(t)$ implies $\alpha(t) = Ce^{-at}$ for some arbitrary constant C. The terminal condition $\alpha(T) = 1$ implies that $C = e^{aT}$ so that

$$\alpha(t) = e^{a(T-t)}.$$

Since we have solved for $\alpha(t)$, we can now solve the second equation in (21.8); that is,

$$\beta'(t) = \frac{\sigma^2\theta^2}{2}\alpha^2(t) = \frac{\sigma^2\theta^2}{2}e^{2a(T-t)}.$$

We simply integrate from 0 to t to find $\beta(t)$:

$$\beta(t) - \beta(0) = \frac{\sigma^2\theta^2}{2}\int_0^t e^{2a(T-s)}\,\mathrm{d}s = \frac{\sigma^2\theta^2}{4a}(e^{2aT} - e^{2a(T-t)}).$$

The terminal condition $\beta(T) = 0$ implies that

$$\beta(0) = \frac{\sigma^2\theta^2}{4a}(1 - e^{2aT})$$

and so

$$\beta(t) = \frac{\sigma^2\theta^2}{4a}(1 - e^{2aT}) + \frac{\sigma^2\theta^2}{4a}(e^{2aT} - e^{2a(T-t)}) = \frac{\sigma^2(1 - e^{2a(T-t)})\theta^2}{4a} = -\frac{\sigma^2(e^{2a(T-t)} - 1)\theta^2}{4a}.$$

Thus, from (21.7) we are now able to conclude that

$$f(t,x) = \exp\{i\theta\alpha(t)x + \beta(t)\} = \exp\left\{i\theta e^{a(T-t)}x - \frac{\sigma^2(e^{2a(T-t)} - 1)\theta^2}{4a}\right\}$$

for $0 \le t \le T$ and $x \in \mathbb{R}$. Taking $t = 0$ gives

$$\varphi_{X_T}(\theta) = f(0,x) = \exp\left\{i\theta e^{aT}x - \frac{\sigma^2(e^{2aT} - 1)\theta^2}{4a}\right\}$$

in agreement with Example 21.2.

22

The Characteristic Function for Heston's Model

As we saw last lecture, it is sometimes possible to determine the characteristic function of a random variable defined via a stochastic differential equation without actually solving the SDE. The computation involves the Feynman-Kac representation theorem, but it does require the solution of a partial differential equation. In certain cases where an explicit solution does not exist for the SDE, computing the characteristic function might still be possible as long as the resulting PDE is solvable.

Recall that the Heston model assumes that the asset price process $\{S_t, t \geq 0\}$ satisfies the SDE

$$\mathrm{d}S_t = \sqrt{v_t}\, S_t \,\mathrm{d}B_t^{(1)} + \mu S_t \,\mathrm{d}t$$

where the variance process $\{v_t, t \geq 0\}$ satisfies

$$\mathrm{d}v_t = \sigma\sqrt{v_t}\,\mathrm{d}B_t^{(2)} + a(b - v_t)\,\mathrm{d}t$$

and the two driving Brownian motions $\{B_t^{(1)}, t \geq 0\}$ and $\{B_t^{(2)}, t \geq 0\}$ are correlated with rate ρ, i.e.,

$$\mathrm{d}\langle B^{(1)}, B^{(2)}\rangle_t = \rho\,\mathrm{d}t.$$

In order to analyze the Heston model, it is easier to work with

$$X_t = \log(S_t)$$

instead. Itô's formula implies that $\{X_t, t \geq 0\}$ satisfies the SDE

$$\mathrm{d}X_t = \mathrm{d}\log S_t = \frac{\mathrm{d}S_t}{S_t} - \frac{\mathrm{d}\langle S\rangle_t}{2S_t^2} = \sqrt{v_t}\,\mathrm{d}B_t^{(1)} + \left(\mu - \frac{v_t}{2}\right)\mathrm{d}t.$$

We will now determine the characteristic function of X_T for any $T \geq 0$. The multidimensional version of Itô's formula (Theorem 20.4) implies that

$$\begin{aligned}
\mathrm{d}f(t,X_t,v_t) &= \dot f(t,X_t,v_t)\,\mathrm{d}t + f_1(t,X_t,v_t)\,\mathrm{d}X_t + \frac{1}{2}f_{11}(t,X_t,v_t)\,\mathrm{d}\langle X\rangle_t \\
&\quad + f_2(t,X_t,v_t)\,\mathrm{d}v_t + \frac{1}{2}f_{22}(t,X_t,v_t)\,\mathrm{d}\langle v\rangle_t + f_{12}(t,X_t,v_t)\,\mathrm{d}\langle X,v\rangle_t \\
&= \dot f(t,X_t,v_t)\,\mathrm{d}t + f_1(t,X_t,v_t)\left(\sqrt{v_t}\,\mathrm{d}B_t^{(1)} + \left(\mu - \frac{v_t}{2}\right)\mathrm{d}t\right) + \frac{1}{2}f_{11}(t,X_t,v_t)v_t\,\mathrm{d}t \\
&\quad + f_2(t,X_t,v_t)\big(\sigma\sqrt{v_t}\,\mathrm{d}B_t^{(2)} + a(b-v_t)\,\mathrm{d}t\big) + \frac{1}{2}f_{22}(t,X_t,v_t)\sigma^2 v_t\,\mathrm{d}t \\
&\quad + f_{12}(t,X_t,v_t)\sigma\rho v_t\,\mathrm{d}t \\
&= f_1(t,X_t,v_t)\sqrt{v_t}\,\mathrm{d}B_t^{(1)} + f_2(t,X_t,v_t)\sigma\sqrt{v_t}\,\mathrm{d}B_t^{(2)} + (\mathcal{A}f)(t,X_t,v_t)\,\mathrm{d}t
\end{aligned}$$

where the differential operator $\mathcal{A}$ is defined as

$$\begin{aligned}
(\mathcal{A}f)(t,x,y) &= \dot f(t,x,y) + \left(\mu - \frac{y}{2}\right) f_1(t,x,y) + \frac{y}{2}f_{11}(t,x,y) + a(b-y)f_2(t,x,y) \\
&\quad + \frac{\sigma^2 y}{2}f_{22}(t,x,y) + \sigma\rho y f_{12}(t,x,y).
\end{aligned}$$

If we now let $u(x) = e^{i\theta x}$, then the (multidimensional form of the) Feynman-Kac representation theorem implies

$$f(t,x,y) = \mathbb{E}[u(X_T)|X_t = x, v_t = y] = \mathbb{E}[e^{i\theta X_T}|X_t = x, v_t = y]$$

is the unique bounded solution of the partial differential equation

$$(\mathcal{A}f)(t,x,y) = 0, \quad 0 \le t \le T, \;\; x \in \mathbb{R}, \;\; y \in \mathbb{R}, \tag{22.1}$$

subject to the terminal condition

$$f(T,x,y) = e^{i\theta x}, \quad x \in \mathbb{R}, \;\; y \in \mathbb{R}.$$

Note that $f(0,x,y) = \mathbb{E}[e^{i\theta X_T}|X_0 = x, v_0 = y] = \varphi_{X_T}(\theta)$ is the characteristic function of X_T.

Guided by the form of the terminal condition and by our experience with the Ornstein-Uhlenbeck characteristic function, we guess that $f(t,x,y)$ can be written as

$$f(t,x,y) = \exp\{\alpha(t)y + \beta(t)\}\exp\{i\theta x\} \tag{22.2}$$

for some functions $\alpha(t)$ and $\beta(t)$ of t only satisfying $\alpha(T) = 0$ and $\beta(T) = 0$. Differentiating we find

$$\dot f(t,x,y) = [\alpha'(t)y + \beta'(t)]f(t,x,y), \quad f_1(t,x,y) = i\theta f(t,x,y), \quad f_{11}(t,x,y) = -\theta^2 f(t,x,y),$$

$$f_2(t,x,y) = \alpha(t)f(t,x,y), \quad f_{22}(t,x,y) = \alpha^2(t)f(t,x,y), \quad f_{12}(t,x,y) = i\theta\alpha(t)f(t,x,y),$$

so that substituting into the explicit form of $(\mathcal{A}f)(t,x,y) = 0$ and factoring out the common $f(t,x,y)$ gives

$$[\alpha'(t)y + \beta'(t)] + i\theta\left(\mu - \frac{y}{2}\right) - \frac{\theta^2}{2}y + a\alpha(t)(b-y) + \frac{\sigma^2\alpha^2(t)}{2}y + i\sigma\rho\theta\alpha(t)y = 0,$$

or equivalently,

$$\left[\alpha'(t) + (i\sigma\rho\theta - a)\alpha(t) + \frac{\sigma^2\alpha^2(t)}{2} - \frac{i\theta}{2} - \frac{\theta^2}{2}\right] y + \beta'(t) + i\theta\mu + ab\alpha(t) = 0.$$

Since this equation must be true for all $0 \le t \le T$, $x \in \mathbb{R}$, and $y \in \mathbb{R}$, the only way that is possible is if the coefficient of y is zero *and* the constant term is 0. Thus, we must have

$$\alpha'(t) + (i\sigma\rho\theta - a)\alpha(t) + \frac{\sigma^2\alpha^2(t)}{2} - \frac{i\theta}{2} - \frac{\theta^2}{2} = 0 \quad \text{and} \quad \beta'(t) + i\theta\mu + ab\alpha(t) = 0. \tag{22.3}$$

The first equation in (22.3) involves $\alpha(t)$ only and is of the form

$$\alpha'(t) = A\alpha(t) + B\alpha^2(t) + C$$

with

$$A = a - i\sigma\rho\theta, \quad B = -\frac{\sigma^2}{2}, \quad C = \frac{i\theta}{2} + \frac{\theta^2}{2}. \tag{22.4}$$

This ordinary differential equation can be solved by integration; see Exercise 22.1 below. The solution is given by

$$\alpha(t) = D + E\tan(Ft + G)$$

where

$$D = -\frac{A}{2B}, \quad E = \sqrt{\frac{C}{B} - \frac{A^2}{4B^2}}, \quad F = BE = B\sqrt{\frac{C}{B} - \frac{A^2}{4B^2}}, \tag{22.5}$$

and G is an arbitrary constant. The terminal condition $\alpha(T) = 0$ implies

$$0 = D + E\tan(FT + G) \quad \text{so that} \quad G = \arctan\left(-\frac{D}{E}\right) - FT$$

which gives

$$\alpha(t) = D + E\tan\left(\arctan\left(-\frac{D}{E}\right) - F(T-t)\right). \tag{22.6}$$

Exercise 22.1. Suppose that a, b, and c are non-zero real constants. Compute

$$\int \frac{\mathrm{d}x}{ax^2 + bx + c}.$$

Hint: Complete the square in the denominator. The resulting function is an antiderivative of an arctangent function.

In order to simplify the expression for $\alpha(t)$ given by (22.6) above, we begin by noting that

$$\cos\left(\arctan\left(-\frac{D}{E}\right)\right) = \frac{E}{\sqrt{D^2 + E^2}} \quad \text{and} \quad \sin\left(\arctan\left(-\frac{D}{E}\right)\right) = -\frac{D}{\sqrt{D^2 + E^2}}. \tag{22.7}$$

Using the sum of angles identity for cosine therefore gives

$$\begin{aligned}
&\cos\left(\arctan\left(-\frac{D}{E}\right) - F(T-t)\right) \\
&= \cos\left(\arctan\left(-\frac{D}{E}\right)\right)\cos\left(F(T-t)\right) + \sin\left(\arctan\left(-\frac{D}{E}\right)\right)\sin\left(F(T-t)\right) \\
&= \frac{E}{\sqrt{D^2+E^2}}\cos\left(F(T-t)\right) - \frac{D}{\sqrt{D^2+E^2}}\sin\left(F(T-t)\right) \\
&= \frac{E\cos\left(F(T-t)\right) - D\sin\left(F(T-t)\right)}{\sqrt{D^2+E^2}}. \qquad (22.8)
\end{aligned}$$

Similarly, the sum of angles identity for sine yields

$$\sin\left(\arctan\left(-\frac{D}{E}\right) - F(T-t)\right) = \frac{-D\cos\left(F(T-t)\right) - E\sin\left(F(T-t)\right)}{\sqrt{D^2+E^2}}. \qquad (22.9)$$

Writing $\tan(z) = \frac{\sin(z)}{\cos(z)}$ and using (22.8) and (22.9) implies

$$\begin{aligned}
\tan\left(\arctan\left(-\frac{D}{E}\right) - F(T-t)\right) &= \frac{-D\cos\left(F(T-t)\right) - E\sin\left(F(T-t)\right)}{E\cos\left(F(T-t)\right) - D\sin\left(F(T-t)\right)} \\
&= \frac{-D\cot\left(F(T-t)\right) - E}{E\cot\left(F(T-t)\right) - D}
\end{aligned}$$

so that substituting the above expression into (22.6) for $\alpha(t)$ gives

$$\alpha(t) = D + E\left[\frac{-D\cot\left(F(T-t)\right) - E}{E\cot\left(F(T-t)\right) - D}\right] = \frac{-(D^2+E^2)}{E\cot\left(F(T-t)\right) - D}.$$

The next step is to substitute back for D, E, and F in terms of the original parameters. It turns out, however, that it is useful to write them in terms of

$$\gamma = \sqrt{\sigma^2(\theta^2 + i\theta) + (a - i\sigma\rho\theta)^2}. \qquad (22.10)$$

Thus, substituting (22.4) into (22.5) gives

$$D = \frac{a - i\sigma\rho\theta}{\sigma^2}, \quad E = \frac{i\gamma}{\sigma^2}, \quad \text{and} \quad F = -\frac{i\gamma}{2}. \qquad (22.11)$$

Since

$$D^2 + E^2 = -\frac{i\theta + \theta^2}{\sigma^2}$$

we conclude that

$$\alpha(t) = \frac{i\theta + \theta^2}{i\gamma\cot\left(-\frac{i\gamma(T-t)}{2}\right) - (a - i\sigma\rho\theta)}.$$

The final simplification is to note that

$$\cos(-iz) = \cosh(z) \quad \text{and} \quad \sin(-iz) = -i\sinh(z)$$

so that

$$\cot(iz) = \frac{\cos(iz)}{\sin(iz)} = \frac{\cosh(z)}{-i\sinh(z)} = i\coth(z)$$

which gives

$$\alpha(t) = \frac{i\theta + \theta^2}{i^2\gamma\coth\left(\frac{\gamma(T-t)}{2}\right) - (a - i\sigma\rho\theta)} = -\frac{i\theta + \theta^2}{\gamma\coth\left(\frac{\gamma(T-t)}{2}\right) + (a - i\sigma\rho\theta)}.$$

Finally, we find

$$\exp\{\alpha(t)y\} = \exp\left\{-\frac{(i\theta + \theta^2)y}{\gamma\coth\left(\frac{\gamma(T-t)}{2}\right) + (a - i\sigma\rho\theta)}\right\}. \tag{22.12}$$

Having determined $\alpha(t)$, we can now consider the second equation in (22.3) involving $\beta'(t)$. It is easier, however, to manipulate this expression using $\alpha(t)$ in the form (22.6). Thus, the expression for $\beta'(t)$ now becomes

$$\beta'(t) = -abD - i\theta\mu - abE\tan\left(\arctan\left(-\frac{D}{E}\right) - F(T-t)\right)$$

which can be solved by integrating from 0 to t. Recall that

$$\int \tan(z)\,\mathrm{d}z = \log(\sec(z)) = -\log(\cos(z))$$

and so

$$\begin{aligned}\beta(t) &= \beta(0) - abDt - i\theta\mu t - abE\int_0^t \tan\left(\arctan\left(-\frac{D}{E}\right) - F(T-s)\right)\mathrm{d}s \\ &= \beta(0) - abDt - i\theta\mu t - \frac{abE}{F}\log\left(\frac{\cos(\arctan\left(-\frac{D}{E}\right) - FT)}{\cos(\arctan\left(-\frac{D}{E}\right) - F(T-t))}\right).\end{aligned}$$

The terminal condition $\beta(T) = 0$ implies that

$$\beta(0) = abDT + i\theta\mu T + \frac{abE}{F}\log\left(\frac{\sqrt{E^2 + D^2}\cos(\arctan\left(-\frac{D}{E}\right) - FT)}{E}\right)$$

using (22.7), and so we now have

$$\beta(t) = abD(T-t) + i\theta\mu(T-t) + \frac{abE}{F}\log\left(\frac{\sqrt{E^2 + D^2}\cos(\arctan\left(-\frac{D}{E}\right) - F(T-t))}{E}\right).$$

As in the calculation of $\alpha(t)$, we can simplify this further using (22.8) so that

$$\beta(t) = abD(T-t) + i\theta\mu(T-t) + \frac{abE}{F}\log\left(\cos\left(F(T-t)\right) - \frac{D}{E}\sin\left(F(T-t)\right)\right)$$

which implies

$$\exp\{\beta(t)\} = \exp\{abD(T-t) + i\theta\mu(T-t)\}\left(\cos\left(F(T-t)\right) - \frac{D}{E}\sin\left(F(T-t)\right)\right)^{\frac{abE}{F}}. \tag{22.13}$$

Substituting the expressions given by (22.11) for D, E, and F in terms of the original parameters into (22.13) gives

$$\exp\{\beta(t)\} = \frac{\exp\left\{\frac{ab(a-i\sigma\rho\theta)(T-t)}{\sigma^2} + i\theta\mu(T-t)\right\}}{\left(\cos\left(-\frac{i\gamma}{2}(T-t)\right) - \frac{a-i\sigma\rho\theta}{i\gamma}\sin\left(-\frac{i\gamma}{2}(T-t)\right)\right)^{\frac{2ab}{\sigma^2}}}.$$

As in the calculation of $\alpha(t)$, the final simplification is to note that $\cos(-iz) = \cosh(z)$ and $\sin(-iz) = -i\sinh(z)$ so that

$$\exp\{\beta(t)\} = \frac{\exp\left\{\frac{ab(a-i\sigma\rho\theta)(T-t)}{\sigma^2} + i\theta\mu(T-t)\right\}}{\left(\cosh\left(\frac{\gamma(T-t)}{2}\right) + \frac{a-i\sigma\rho\theta}{\gamma}\sinh\left(\frac{\gamma(T-t)}{2}\right)\right)^{\frac{2ab}{\sigma^2}}}. \tag{22.14}$$

We can now substitute our expression for $\exp\{\alpha(t)y\}$ given by (22.12) and our expression for $\exp\{\beta(t)\}$ given by (22.14) into our guess for $f(t,x,y)$ given by (22.2) to conclude

$$\begin{aligned} f(t,x,y) &= \exp\{\alpha(t)y + \beta(t)\}\exp\{i\theta x\} \\ &= \frac{\exp\left\{i\theta x - \frac{(i\theta+\theta^2)y}{\gamma\coth\left(\frac{\gamma(T-t)}{2}\right)+(a-i\sigma\rho\theta)} + \frac{ab(a-i\sigma\rho\theta)(T-t)}{\sigma^2} + i\theta\mu(T-t)\right\}}{\left(\cosh\left(\frac{\gamma(T-t)}{2}\right) + \frac{a-i\sigma\rho\theta}{\gamma}\sinh\left(\frac{\gamma(T-t)}{2}\right)\right)^{\frac{2ab}{\sigma^2}}}. \end{aligned}$$

Taking $t = 0$ gives

$$\varphi_{X_T}(\theta) = f(0,x,y) = \frac{\exp\left\{i\theta x - \frac{(i\theta+\theta^2)y}{\gamma\coth\frac{\gamma T}{2}+(a-i\sigma\rho\theta)} + \frac{abT(a-i\sigma\rho\theta)}{\sigma^2} + i\theta\mu T\right\}}{\left(\cosh\frac{\gamma T}{2} + \frac{a-i\sigma\rho\theta}{\gamma}\sinh\frac{\gamma T}{2}\right)^{\frac{2ab}{\sigma^2}}}.$$

and we are done!

23
Risk Neutrality

We will now use the Feynman-Kac representation theorem to derive a general solution to the Black-Scholes option pricing problem for European call options. This representation of the solution will be needed next lecture when we explain how to use the characteristic function of a diffusion to price an option.

Suppose that the asset price process $\{S_t, t \geq 0\}$ satisfies the stochastic differential equation

$$\mathrm{d}S_t = \sigma(t, S_t)S_t \, \mathrm{d}\tilde{B}_t + \mu(t, S_t)S_t \, \mathrm{d}t, \tag{23.1}$$

or equivalently,

$$\frac{\mathrm{d}S_t}{S_t} = \sigma(t, S_t) \, \mathrm{d}\tilde{B}_t + \mu(t, S_t) \, \mathrm{d}t,$$

where $\{\tilde{B}_t, t \geq 0\}$ is a standard Brownian motion with $\tilde{B}_0 = 0$, and that the risk-free investment $D(t, S_t)$ evolves according to

$$\mathrm{d}D(t, S_t) = rD(t, S_t) \, \mathrm{d}t$$

where $r > 0$ is the risk-free interest rate.

Remark. We are writing $\{\tilde{B}_t, t \geq 0\}$ for the Brownian motion that drives the asset price process since the formula that we are going to derive for the fair price at time $t = 0$ of a European call option on this asset involves a one-dimensional Brownian motion distinct from this one.

Remark. The asset price process given by (23.1) is similar to geometric Brownian motion, except that the volatility and drift do not necessarily need to be constant. Instead, they can be stochastic, but the randomness is assumed to come from the asset price itself. In this general form, there is no explicit form for $\{S_t, t \geq 0\}$ as the solution of the SDE (23.1).

Suppose further that we write $V(t, S_t)$, $0 \leq t \leq T$, to denote the price at time t of a European call option with expiry date T on the asset $\{S_t, t \geq 0\}$. If the payoff function is given by $\Lambda(x)$, $x \in \mathbb{R}$, then

$$V(T, S_T) = \Lambda(S_T).$$

(Recall that a European call option can be exercised only on the expiry date T and not earlier.) Our goal is to determine $V(0, S_0)$, the fair price to pay at time $t = 0$.

Furthermore, assume that there are no arbitrage opportunities so that there exists a replicating portfolio

$$\Pi(t, S_t) = A(t, S_t)S_t + D(t, S_t)$$

consisting of a cash deposit D and a number A of assets which is self-financing:

$$\mathrm{d}\Pi(t, S_t) = A(t, S_t)\,\mathrm{d}S_t + rD(t, S_t)\,\mathrm{d}t.$$

As in Lecture #16, this implies that the change in $V(t, S_t) - \Pi(t, S_t)$ over any time step is non-random and must equal the corresponding growth offered by the continuously compounded risk-free interest rate. That is,

$$\mathrm{d}\big[V(t, S_t) - \Pi(t, S_t)\big] = r\big[V(t, S_t) - \Pi(t, S_t)\big]\,\mathrm{d}t.$$

By repeating the calculations in Lecture #16 assuming that the asset price movement satisfies (23.1) leads to the following conclusion. The function $V(t, x)$, $0 \le t \le T$, $x \in \mathbb{R}$, satisfies the *Black-Scholes partial differential equation*

$$\dot{V}(t, x) + \frac{\sigma^2(t, x)}{2} x^2 V''(t, x) + rxV'(t, x) - rV(t, x) = 0, \quad 0 \le t \le T, \;\; x \in \mathbb{R}, \tag{23.2}$$

subject to the terminal condition

$$V(T, x) = \Lambda(x).$$

Note. The only difference between (23.2) and the Black-Scholes PDE that we derived in Lecture #16, namely (16.10), is the appearance of the function $\sigma(t, x)$ instead of the constant σ. Thus, (23.2) reduces to (16.10) when $\sigma(t, x) = \sigma$ is constant.

At this point, we observe that the formulation of the option pricing problem sounds rather similar to the formulation of the Feynman-Kac representation theorem which we now recall.

Theorem 23.1 (Feynman-Kac Representation Theorem). *Suppose that $u \in C^2(\mathbb{R})$, and let $\{X_t, t \ge 0\}$ be defined by the SDE*

$$\mathrm{d}X_t = a(t, X_t)\,\mathrm{d}B_t + b(t, X_t)\,\mathrm{d}t.$$

The unique bounded function $f : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ satisfying the partial differential equation

$$(\mathcal{A}f)(t, x) = b(t, x)f'(t, x) + \frac{1}{2}a^2(t, x)f''(t, x) + \dot{f}(t, x) = 0, \quad 0 \le t \le T, \;\; x \in \mathbb{R}, \tag{23.3}$$

subject to the terminal condition

$$f(T, x) = u(x), \quad x \in \mathbb{R},$$

is given by

$$f(t, x) = \mathbb{E}[u(X_T)|X_t = x].$$

However, the differential equation (23.2) that we need to solve is of the form

$$b(t, x)g'(t, x) + \frac{1}{2}a^2(t, x)g''(t, x) + \dot{g}(t, x) = rg(t, x), \quad 0 \le t \le T, \;\; x \in \mathbb{R}, \tag{23.4}$$

subject to the terminal condition

$$g(T, x) = u(x), \quad x \in \mathbb{R}.$$

In other words, we need to solve a non-homogeneous partial differential equation. Although the theory for non-homogeneous PDEs is reasonably well-established, it is not too difficult to guess what the solution to our particular equation (23.4) must be. If we let

$$g(t, x) = e^{-r(T-t)} f(t, x), \quad 0 \le t \le T, \ \ x \in \mathbb{R},$$

where $f(t, x)$ is the solution to the homogeneous partial differential equation $(\mathcal{A}f)(t, x) = 0$ given in (23.3) subject to the terminal condition $f(T, x) = u(x)$, then $g(T, x) = f(T, x) = u(x)$, and

$$g'(t, x) = e^{-r(T-t)} f'(t, x), \quad g''(t, x) = e^{-r(T-t)} f''(t, x), \quad \text{and}$$

$$\dot{g}(t, x) = e^{-r(T-t)} \dot{f}(t, x) + re^{-r(T-t)} f(t, x)$$

so that

$$\begin{aligned}
& b(t, x)g'(t, x) + \frac{1}{2}a^2(t, x)g''(t, x) + \dot{g}(t, x) \\
&= b(t, x)e^{-r(T-t)} f'(t, x) + \frac{1}{2}a^2(t, x)e^{-r(T-t)} f''(t, x) + e^{-r(T-t)} \dot{f}(t, x) + re^{-r(T-t)} f(t, x) \\
&= e^{-r(T-t)} \left[b(t, x)f'(t, x) + \frac{1}{2}a^2(t, x)f''(t, x) + \dot{f}(t, x) \right] + re^{-r(T-t)} f(t, x) \\
&= e^{-r(T-t)} (\mathcal{A}f)(t, x) + rg(t, x) \\
&= rg(t, x)
\end{aligned}$$

using the assumption that $(\mathcal{A}f)(t, x) = 0$. In other words, we have established the following extension of the Feynman-Kac representation theorem.

Theorem 23.2 (Feynman-Kac Representation Theorem). *Suppose that $u \in C^2(\mathbb{R})$, and let $\{X_t, t \ge 0\}$ be defined by the SDE*

$$\mathrm{d}X_t = a(t, X_t)\,\mathrm{d}B_t + b(t, X_t)\,\mathrm{d}t.$$

The unique bounded function $g : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ satisfying the partial differential equation

$$b(t, x)g'(t, x) + \frac{1}{2}a^2(t, x)g''(t, x) + \dot{g}(t, x) - rg(t, x) = 0, \quad 0 \le t \le T, \ \ x \in \mathbb{R},$$

subject to the terminal condition

$$g(T, x) = u(x), \quad x \in \mathbb{R},$$

is given by

$$g(t, x) = e^{-r(T-t)} \mathbb{E}[u(X_T) | X_t = x].$$

At this point, let's recall where we are. We are assuming that the asset price $\{S_t, t \ge 0\}$ evolves according to

$$\mathrm{d}S_t = \sigma(t, S_t)S_t\,\mathrm{d}\tilde{B}_t + \mu(t, S_t)S_t\,\mathrm{d}t$$

and we want to determine $V(0, S_0)$, the fair price at time $t = 0$ of a European call option with expiry date T and payoff $V(T, S_T) = \Lambda(S_T)$ where $\Lambda(x)$, $x \in \mathbb{R}$, is given. We have also shown that $V(t, x)$ satisfies the PDE

$$\dot{V}(t,x) + \frac{\sigma^2(t,x)}{2} x^2 V''(t,x) + rxV'(t,x) - rV(t,x) = 0, \quad 0 \le t \le T, \;\; x \in \mathbb{R}, \tag{23.5}$$

subject to the terminal condition

$$V(T,x) = \Lambda(x).$$

The generalized Feynman-Kac representation theorem tells us that the solution to

$$b(t,x)g'(t,x) + \frac{1}{2}a^2(t,x)g''(t,x) + \dot{g}(t,x) - rg(t,x) = 0, \quad 0 \le t \le T, \;\; x \in \mathbb{R}, \tag{23.6}$$

subject to the terminal condition $g(T, x) = \Lambda(x)$ is

$$g(t,x) = e^{-r(T-t)}\mathbb{E}[\Lambda(X_T)|X_t = x], \tag{23.7}$$

where X_t satisfies the SDE

$$\mathrm{d}X_t = a(t, X_t)\,\mathrm{d}B_t + b(t, X_t)\,\mathrm{d}t$$

and $\{B_t, t \ge 0\}$ is a standard one-dimensional Brownian motion with $B_0 = 0$. Note that the expectation in (23.7) is with respect to the process $\{X_t, t \ge 0\}$ driven by the Brownian motion $\{B_t, t \ge 0\}$.

Comparing (23.5) and (23.6) suggests that

$$b(t,x) = rx \quad \text{and} \quad a^2(t,x) = \sigma^2(t,x)x^2$$

so that

$$\mathrm{d}X_t = \sigma(t, X_t)X_t\,\mathrm{d}B_t + rX_t\,\mathrm{d}t, \tag{23.8}$$

or equivalently,

$$\frac{\mathrm{d}X_t}{X_t} = \sigma(t, X_t)\,\mathrm{d}B_t + r\,\mathrm{d}t,$$

Hence, we have developed a complete solution to the European call option pricing problem which we summarize in Theorem 23.3 below.

Remark. The process $\{X_t, t \ge 0\}$ defined by the SDE (23.8) is sometimes called the *risk-neutral process* associated with the asset price process $\{S_t, t \ge 0\}$ defined by (23.1). As with the asset price process, the associated risk-neutral process is similar to geometric Brownian motion. Note that the function $\sigma(t, x)$ is the same in both equations. However, $\{\tilde{B}_t, t \ge 0\}$, the Brownian motion that drives $\{S_t, t \ge 0\}$ is NOT the same as $\{B_t, t \ge 0\}$, the Brownian motion that drives $\{X_t, t \ge 0\}$.

Remark. The approach we have used to develop the associated risk-neutral process was via the Feynman-Kac representation theorem. An alternative approach which we do not discuss is to use the Girsanov-Cameron-Martin theorem to construct an *equivalent martingale measure.*

Theorem 23.3. *Let $V(t, S_t)$, $0 \le t \le T$, denote the fair price to pay at time t of a European call option having payoff $V(T, S_T) = \Lambda(S_T)$ on the asset price $\{S_t, t \ge 0\}$ which satisfies the stochastic differential equation*

$$\mathrm{d}S_t = \sigma(t, S_t)S_t \,\mathrm{d}\tilde{B}_t + \mu(t, S_t)S_t \,\mathrm{d}t$$

where $\{\tilde{B}_t, t \ge 0\}$ is a standard Brownian motion with $B_0 = 0$. The fair price to pay at time $t = 0$ is given by

$$V(0, S_0) = e^{-rT}\mathbb{E}[\Lambda(X_T)|X_0 = S_0]$$

where the associated risk-neutral process $\{X_t, t \ge 0\}$ satisfies the stochastic differential equation

$$\mathrm{d}X_t = \sigma(t, X_t)X_t \,\mathrm{d}B_t + rX_t \,\mathrm{d}t$$

and $\{B_t, t \ge 0\}$ is a standard one-dimensional Brownian motion distinct from $\{\tilde{B}_t, t \ge 0\}$.

Remark. Since S_0, the value of the underlying asset at $t = 0$, is known, in order to calculate the expectation $\mathbb{E}[\Lambda(X_T)|X_0 = S_0]$, you need to know something about the distribution of X_T. There is no general formula for determining the distribution of X_T in terms of the distribution of S_T unless some additional structure is known about $\sigma(t, x)$. For instance, assuming that $\sigma(t, x) = \sigma$ is constant leads to the Black-Scholes formula from Lecture #17, while assuming $\sigma(t, x) = \sigma(t)$ is a deterministic function of time leads to an explicit formula which, though similar, is more complicated to write down.

Example 23.4. We now explain how to recover the Black-Scholes formula in the case that $\{S_t, t \ge 0\}$ is geometric Brownian motion and $\Lambda(x) = (x - E)^+$. Since the asset price process SDE is

$$\mathrm{d}S_t = \sigma S_t \,\mathrm{d}\tilde{B}_t + \mu S_t \,\mathrm{d}t$$

we conclude that the risk-neutral process is

$$\mathrm{d}X_t = \sigma X_t \,\mathrm{d}B_t + rX_t \,\mathrm{d}t.$$

The risk-neutral process is also geometric Brownian motion (but with drift r) so that

$$X_T = X_0 \exp\left\{\sigma B_T + \left(r - \frac{\sigma^2}{2}\right) T\right\} = S_0 \exp\left\{\sigma B_T + \left(r - \frac{\sigma^2}{2}\right) T\right\}$$

since we are assuming that $X_0 = S_0$. Since $B_T \sim \mathcal{N}(0, T)$, we can write

$$X_T = S_0 \exp\left\{\left(r - \frac{\sigma^2}{2}\right) T\right\} \exp\left\{\sigma\sqrt{T}Z\right\}$$

for $Z \sim \mathcal{N}(0, 1)$. Therefore,

$$V(0, S_0) = e^{-rT}\mathbb{E}[\Lambda(X_T)|X_0 = S_0] = e^{-rT}\mathbb{E}[(X_T - E)^+|X_0 = S_0]$$

can be evaluated using Exercise 3.7, namely if $a > 0$, $b > 0$, $c > 0$ are constants and $Z \sim \mathcal{N}(0, 1)$, then

$$\mathbb{E}[\,(ae^{bZ} - c)^+\,] = ae^{b^2/2}\,\Phi\left(b + \frac{1}{b}\log\frac{a}{c}\right) - c\,\Phi\left(\frac{1}{b}\log\frac{a}{c}\right),$$

with

$$a = S_0\, e^{\left(r - \frac{\sigma^2}{2}\right)T}, \quad b = \sigma\sqrt{T}, \quad c = E.$$

Doing this, and noting that $ae^{b^2/2} = S_0 e^{rT}$, gives

$$\begin{aligned}
&V(0, S_0) \\
&\quad = e^{-rT}\mathbb{E}[(X_T - E)^+ | X_0 = S_0] \\
&\quad = e^{-rT}\left[S_0\, e^{rT}\, \Phi\left(\sigma\sqrt{T} + \frac{1}{\sigma\sqrt{T}} \log \frac{S_0\, e^{\left(r-\frac{\sigma^2}{2}\right)T}}{E}\right) - E\, \Phi\left(\frac{1}{\sigma\sqrt{T}} \log \frac{S_0\, e^{\left(r-\frac{\sigma^2}{2}\right)T}}{E}\right)\right] \\
&\quad = S_0\, \Phi\left(\frac{\log(S_0/E) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) - Ee^{-rT}\, \Phi\left(\frac{\log(S_0/E) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right)
\end{aligned}$$

in agreement with Lecture #17.

24

A Numerical Approach to Option Pricing Using Characteristic Functions

As we discussed in Lectures #21 and #22, it is sometimes possible to determine the characteristic function $\varphi_{X_T}(\theta)$ for the random variable X_T, which is defined via a diffusion. We then discussed risk-neutrality in Lecture #23, and derived a complete solution to the problem of pricing European call options.

At this point, it is time to address the following question. How does knowing the characteristic function help us determine the value of an option?

In order to keep our notation straight, we will write $\{S_t, t \geq 0\}$ for the underlying asset price process driven by the Brownian motion $\{\tilde{B}_t, t \geq 0\}$, and we will assume that

$$\frac{\mathrm{d}S_t}{S_t} = \sigma(t, S_t)\,\mathrm{d}\tilde{B}_t + \mu(t, S_t)\,\mathrm{d}t.$$

We will then write $\{X_t, t \geq 0\}$ for the associated risk-neutral process driven by the Brownian motion $\{B_t, t \geq 0\}$. Guided by Lecture #23, we will phrase all of our results in terms of the risk-neutral process $\{X_t, t \geq 0\}$.

Note. The purpose of Example 21.4 with arithmetic Brownian motion and Example 21.5 with the Ornstein-Uhlenbeck process was to illustrate how the characteristic function could be found without actually solving the SDE. Of course, neither of these is an adequate model of the asset price movement. Heston's model, however, is an adequate model for the underlying asset price, and in Lecture 22 we found the characteristic function without solving the defining SDE.

Suppose that we are interested in determining the fair price at time $t = 0$ of a European call option on the asset price $\{S_t, t \geq 0\}$ with strike price E and expiry date T assuming a risk-free interest rate r. The payoff function is therefore $\Lambda(x) = (x - E)^+$. If $V(0, S_0)$ denotes the fair price at time $t = 0$, then from Theorem 23.3, we can express the solution as

$$V(0, S_0) = e^{-rT}\mathbb{E}[(X_T - E)^+ | X_0 = S_0]$$

where the expectation is with respect to the associated risk-neutral process $\{X_t, t \geq 0\}$ driven by the Brownian motion $\{B_t, t \geq 0\}$. As we saw in Lecture #23, the associated risk-neutral process is a geometric-type Brownian motion given by

$$\frac{\mathrm{d}X_t}{X_t} = \sigma(t, X_t)\,\mathrm{d}B_t + r\,\mathrm{d}t.$$

It turns out that for the following calculations it is more convenient to consider the process $\{Z_t, t \geq 0\}$ where

$$Z_t = \log X_t.$$

Exercise 24.1. Suppose that $\{X_t, t \geq 0\}$ satisfies the associated risk-neutral SDE

$$\mathrm{d}X_t = \sigma(t, X_t)X_t\,\mathrm{d}B_t + rX_t\,\mathrm{d}t.$$

Use Itô's formula to determine the SDE satisfied by $\{Z_t, t \geq 0\}$ where $Z_t = \log X_t$.

We will now write the strike price E as e^k (so that $k = \log E$) and $X_T = e^{Z_T}$. Therefore,

$$V(0, S_0) = e^{-rT}\mathbb{E}[(X_T - E)^+ | X_0 = S_0] = e^{-rT}\mathbb{E}[(e^{Z_T} - e^k)^+]$$

where $Z_0 = \log X_0 = \log S_0$ is known.

Suppose further that we are able to determine the density function of the random variable Z_T which we write as $f_{Z_T}(z)$ so that

$$V(0, S_0) = e^{-rT}\mathbb{E}[(e^{Z_T} - e^k)^+] = e^{-rT}\int_{-\infty}^{\infty}(e^z - e^k)^+ f_{Z_T}(z)\,\mathrm{d}z = e^{-rT}\int_{k}^{\infty}(e^z - e^k) f_{Z_T}(z)\,\mathrm{d}z.$$

We will now view the fair price at time 0 as a function of the logarithm of the strike price k so that

$$V(k) = e^{-rT}\int_{k}^{\infty}(e^z - e^k) f_{Z_T}(z)\,\mathrm{d}z.$$

As $k \to -\infty$ (so that $E \to 0$) we see that $V(k) \to S_0$ which implies that $V(k)$ is *not* integrable:

$$\int_{-\infty}^{\infty} V(k)\,\mathrm{d}k \quad \text{does not exist.}$$

It necessarily follows that $V(k)$ is *not* square-integrable:

$$\int_{-\infty}^{\infty} V^2(k)\,\mathrm{d}k \quad \text{does not exist.}$$

However, if we consider

$$W(k) = e^{ck}V(k), \tag{24.1}$$

then W *is* square-integrable for a suitable $c > 0$ which may depend on the model for $\{S_t, t \geq 0\}$. The *Fourier transform* of W is the function $\hat{W}$ defined by

$$\hat{W}(\xi) = \int_{-\infty}^{\infty} e^{i\xi k}W(k)\,\mathrm{d}k. \tag{24.2}$$

Remark. The existence of the Fourier transform requires that the function $W(k)$ be in L^2.

Therefore, substituting in for $W(k)$ gives

$$\begin{aligned}\hat{W}(\xi) &= e^{-rT}\int_{-\infty}^{\infty}\int_{k}^{\infty} e^{i\xi k}e^{ck}(e^z - e^k) f_{Z_T}(z)\,\mathrm{d}z\,\mathrm{d}k \\ &= e^{-rT}\int_{-\infty}^{\infty}\int_{k}^{\infty}\left(e^z e^{(i\xi + c)k} - e^{(i\xi + c + 1)k}\right) f_{Z_T}(z)\,\mathrm{d}z\,\mathrm{d}k.\end{aligned}$$

Switching the order of integration, we find

$$\hat{W}(\xi) = e^{-rT} \int_{-\infty}^{\infty} f_{Z_T}(z) \int_{-\infty}^{z} \left(e^z e^{(i\xi+c)k} - e^{(i\xi+c+1)k} \right) \mathrm{d}k \, \mathrm{d}z.$$

Since

$$\int_{-\infty}^{z} \left(e^z e^{(i\xi+c)k} - e^{(i\xi+c+1)k} \right) \mathrm{d}k = \frac{e^z e^{(i\xi+c)z}}{i\xi + c} - \frac{e^{(i\xi+c+1)z}}{i\xi + c + 1} = \frac{e^{(i\xi+c+1)z}}{(i\xi + c)(i\xi + c + 1)},$$

we conclude that

$$\begin{aligned} \hat{W}(\xi) &= e^{-rT} \int_{-\infty}^{\infty} f_{Z_T}(z) \frac{e^{(i\xi+c+1)z}}{(i\xi + c)(i\xi + c + 1)} \, \mathrm{d}z \\ &= \frac{e^{-rT}}{(i\xi + c)(i\xi + c + 1)} \int_{-\infty}^{\infty} e^{(i\xi+c+1)z} f_{Z_T}(z) \, \mathrm{d}z \\ &= \frac{e^{-rT}}{(i\xi + c)(i\xi + c + 1)} \int_{-\infty}^{\infty} e^{i(\xi-i(c+1))z} f_{Z_T}(z) \, \mathrm{d}z \\ &= \frac{e^{-rT}}{(i\xi + c)(i\xi + c + 1)} \mathbb{E}[e^{i(\xi-i(c+1))Z_T}] \\ &= \frac{e^{-rT}}{(i\xi + c)(i\xi + c + 1)} \varphi_{Z_T}(\xi - i(c+1)). \end{aligned} \tag{24.3}$$

Remark. It can be shown that a sufficient condition for $W(k)$ to be square-integrable is for $\hat{W}(0)$ to be finite. This is equivalent to

$$\mathbb{E}(S_T^{c+1}) < \infty.$$

The choice $c = 0.75$ can be shown to work for the Heston model.

Given the Fourier transform $\hat{W}(\xi)$, one recovers the original function $W(k)$ via the *inverse Fourier transform* defined by

$$W(k) = \frac{1}{\pi} \int_0^{\infty} e^{-i\xi k} \hat{W}(\xi) \, \mathrm{d}\xi. \tag{24.4}$$

Substituting (24.1) and (24.3) into (24.4) implies

$$e^{ck} V(k) = \frac{1}{\pi} \int_0^{\infty} e^{-i\xi k} \frac{e^{-rT}}{(i\xi + c)(i\xi + c + 1)} \varphi_{Z_T}(\xi - i(c+1)) \, \mathrm{d}\xi$$

so that

$$V(0, S_0) = V(k) = \frac{e^{-ck} e^{-rT}}{\pi} \int_0^{\infty} \frac{e^{-i\xi k}}{(i\xi + c)(i\xi + c + 1)} \varphi_{Z_T}(\xi - i(c+1)) \, \mathrm{d}\xi \tag{24.5}$$

is the fair price at time $t = 0$ of a European call option with strike price $E = e^k$ and expiry date T assuming the risk-free interest rate is $r > 0$.

Remark. Notice that (24.5) expresses the required price of a European call option in terms of $\varphi_{Z_T}(\theta)$, the characteristic function of $Z_T = \log X_T$, the logarithm of X_T defined via the risk-neutral SDE. The usefulness of this formula is that it can be approximated numerically in an extremely efficient manner using the *fast Fourier transform* (FFT). In fact, it is shown in [5] that the FFT approach to option pricing for Heston's model is over 300 times faster than by pricing options using Monte Carlo simulations. There are, however, a number of other practical issues to implementation that the FFT approach to option pricing raises; for further details, see [5].

Example 24.2. The Heston model assumes that the asset price process $\{S_t, t \geq 0\}$ satisfies the SDE

$$\mathrm{d}S_t = \sqrt{v_t}\, S_t\, \mathrm{d}\tilde{B}_t^{(1)} + \mu S_t\, \mathrm{d}t$$

where the variance process $\{v_t, t \geq 0\}$ satisfies

$$\mathrm{d}v_t = \sigma\sqrt{v_t}\, \mathrm{d}B_t^{(2)} + a(b - v_t)\, \mathrm{d}t$$

and the two driving Brownian motions $\{\tilde{B}_t^{(1)}, t \geq 0\}$ and $\{B_t^{(2)}, t \geq 0\}$ are correlated with rate ρ, i.e.,

$$\mathrm{d}\langle \tilde{B}^{(1)}, B^{(2)} \rangle_t = \rho\, \mathrm{d}t.$$

Although this is a two-dimensional example, the risk-neutral process can be worked out in a similar manner to the one-dimensional case. The result is that

$$\mathrm{d}X_t = \sqrt{v_t}\, X_t\, \mathrm{d}B_t^{(1)} + rX_t\, \mathrm{d}t.$$

If we now consider

$$Z_t = \log(X_t),$$

then

$$\mathrm{d}Z_t = \sqrt{v_t}\, \mathrm{d}B_t^{(1)} + \left(r - \frac{v_t}{2}\right) \mathrm{d}t,$$

and so

$$\varphi_{Z_T}(\theta) \frac{\exp\left\{ i\theta x - \frac{(i\theta+\theta^2)y}{\gamma \coth \frac{\gamma T}{2} + (a - i\sigma\rho\theta)} + \frac{abT(a-i\sigma\rho\theta)}{\sigma^2} + i\theta r T \right\}}{\left(\cosh \frac{\gamma T}{2} + \frac{a - i\sigma\rho\theta}{\gamma} \sinh \frac{\gamma T}{2}\right)^{\frac{2ab}{\sigma^2}}}$$

where $x = Z_0 = \log(X_0)$ and $y = v_0$.

25

An Introduction to Functional Analysis for Financial Applications

For the remainder of the course, we are going to discuss some approaches to risk analysis. We will do this, however, with some formality. As such, we need to learn a little bit of functional analysis.

In calculus, we analyze individual functions and study particular properties of these individual functions.

Example 25.1. Consider the function $f(x) = x^2$. We see that the domain of f is all real numbers, and the range of f is all non-negative real numbers. The graph of f is a parabola with its vertex at $(0, 0)$ and opening up. We can also compute

$$f'(x) = \frac{\mathrm{d}}{\mathrm{d}x}x^2 = 2x \quad \text{and} \quad \int f(x)\,\mathrm{d}x = \int x^2\,\mathrm{d}x = \frac{x^3}{3} + C.$$

In functional analysis we study sets of functions with a view to properties possessed by every function in the set. Actually, you would have seen a glimpse of this in calculus.

Example 25.2. Let $\mathcal{X}$ be the set of all differentiable functions with domain $\mathbb{R}$. If $f \in \mathcal{X}$, then f is necessarily **(i)** continuous, and **(ii)** Riemann integrable on every finite interval $[a, b]$.

In order to describe a function acting on a set of functions such as $\mathcal{X}$ in the previous example, we use the word *functional* (or *operator*).

Example 25.3. As in the previous example, let $\mathcal{X}$ denote the set of differentiable functions on $\mathbb{R}$. Define the functional D by setting $D(f) = f'$ for $f \in \mathcal{X}$. That is,

$$f \in \mathcal{X} \mapsto f'.$$

Formally, we define D by

$$(Df)(x) = f'(x)$$

for every $x \in \mathbb{R}$, $f \in \mathcal{X}$.

Example 25.4. We have already seen a number of differential operators in the context of the Feynman-Kac representation theorem. If we let $\mathcal{X}$ denote the space of all functions f of two variables, say $f(t, x)$, such that $f \in C^1([0, \infty)) \times C^2(\mathbb{R})$, and $a(t, x)$ and $b(t, x)$ are given functions, then we can define the functional $\mathcal{A}$ by setting

$$(\mathcal{A}f)(t, x) = b(t, x)f'(t, x) + \frac{1}{2}a^2(t, x)f''(t, x) + \dot{f}(t, x).$$

The next definition is of fundamental importance to functional analysis.

Definition 25.5. Let $\mathcal{X}$ be a space. A *norm* on $\mathcal{X}$ is a function $\|\cdot\| : \mathcal{X} \to \mathbb{R}$ satisfying the following properties:

(i) $\|x\| \geq 0$ for every $x \in \mathcal{X}$,
(ii) $\|x\| = 0$ if and only if $x = 0$,
(iii) $\|\alpha x\| = |\alpha|\|x\|$ for every $\alpha \in \mathbb{R}$ and $x \in \mathcal{X}$, and
(iv) $\|x + y\| \leq \|x\| + \|y\|$ for every $x, y \in \mathcal{X}$.

Remark. We often call (iv) the *triangle inequality.*

Example 25.6. The idea of a norm is that it generalizes the usual absolute value on $\mathbb{R}$. Indeed, let $\mathcal{X} = \mathbb{R}$ and for $x \in \mathcal{X}$ define $\|x\| = |x|$. Properties of absolute value immediately imply that

(i) $\|x\| = |x| \geq 0$ for every $x \in \mathcal{X}$,
(ii) $\|x\| = |x| = 0$ if and only if $x = 0$, and
(iii) $\|\alpha x\| = |\alpha x| = |\alpha||x| = |\alpha|\|x\|$ for every $\alpha \in \mathbb{R}$ and $x \in \mathcal{X}$.

The triangle inequality $|x + y| \leq |x| + |y|$ is essentially a fact about right-angle triangles. The proof is straightforward. Observe that $xy \leq |x||y|$. Therefore,

$$x^2 + 2xy + y^2 \leq x^2 + 2|x||y| + y^2 \quad \text{or, equivalently,} \quad (x + y)^2 \leq x^2 + 2|x||y| + y^2.$$

Using the fact that $x^2 = |x|^2$ implies

$$|x + y|^2 \leq |x|^2 + 2|x||y| + |y|^2 = (|x| + |y|)^2.$$

Taking square roots of both sides yields the result.

Exercise 25.7. Assume that $x, y \in \mathbb{R}$. Show that the triangle inequality $|x + y| \leq |x| + |y|$ is equivalent to the following statements:

(i) $|x - y| \leq |x| + |y|$,
(ii) $|x + y| \geq |x| - |y|$,
(iii) $|x - y| \geq |x| - |y|$, and
(iv) $|x - y| \geq |y| - |x|$.

Example 25.8. Let $\mathcal{X} = \mathbb{R}^2$. If $x \in \mathbb{R}^2$, we can write $x = (x_1, x_2)$. If we define

$$\|x\| = \sqrt{x_1^2 + x_2^2},$$

then $\|\cdot\|$ is a norm on $\mathcal{X}$.

Exercise 25.9. Verify that $\|x\| = \sqrt{x_1^2 + x_2^2}$ is, in fact, a norm on $\mathbb{R}^2$.

Example 25.10. More generally, let $\mathcal{X} = \mathbb{R}^n$. If $x \in \mathbb{R}^n$, we can write $x = (x_1, \dots, x_n)$. If we define

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2},$$

then $\|\cdot\|$ is a norm on $\mathcal{X}$.

Example 25.11. Let $\mathcal{X}$ denote the space of continuous functions on $[0, 1]$. In calculus, we would write such a function as $f(x)$, $0 \le x \le 1$. In functional analysis, we prefer to write such a function as $x(t)$, $0 \le t \le 1$. That is, it is traditional to use a lower case x to denote an arbitrary point in a space. It so happens that our space $\mathcal{X}$ consists of individual points x which happen themselves to be functions. If we define

$$\|x\| = \max_{0 \le t \le 1} |x(t)|,$$

then $\|\cdot\|$ is a norm on $\mathcal{X}$. Indeed,

(i) $|x(t)| \ge 0$ for every $0 \le t \le 1$ and $x \in \mathcal{X}$ so that $\|x\| \ge 0$,
(ii) $\|x\| = 0$ if and only if $x(t) = 0$ for every $0 \le t \le 1$ (i.e., $x = 0$), and
(iii) $\|\alpha x\| = \max_{0 \le t \le 1} |\alpha x(t)| = |\alpha| \max_{0 \le t \le 1} |x(t)| = |\alpha| \|x\|$ for every $\alpha \in \mathbb{R}$ and $x \in \mathcal{X}$.

As for **(iv)**, notice that

$$\|x + y\| = \max_{0 \le t \le 1} |x(t) + y(t)| \le \max_{0 \le t \le 1} (|x(t)| + |y(t)|)$$

by the usual triangle inequality. Since

$$\max_{0 \le t \le 1} (|x(t)| + |y(t)|) \le \max_{0 \le t \le 1} |x(t)| + \max_{0 \le t \le 1} |y(t)| = \|x\| + \|y\|,$$

we conclude $\|x + y\| \le \|x\| + \|y\|$ as required. Note that we sometimes write $C[0, 1]$ for the space of continuous functions on $[0, 1]$.

Example 25.12. Let $\mathcal{X}$ denote the space of all random variables with finite variance. In keeping with the traditional notation for random variables, we prefer to write $X \in \mathcal{X}$ instead of the traditional functional analysis notation $x \in \mathcal{X}$. As we will see next lecture, if we define

$$\|X\| = \sqrt{\mathbb{E}(X^2)},$$

then this defines a norm on $\mathcal{X}$. (Actually, this is not quite precise. We will be more careful next lecture.) A *risk measure* will be a functional $\rho : \mathcal{X} \to \mathbb{R}$ satisfying certain natural properties.

26

A Linear Space of Random Variables

Let $\mathcal{X}$ denote the space of all random variables with finite variance. If $X \in \mathcal{X}$, define

$$\|X\| = \sqrt{\mathbb{E}(X^2)}.$$

Question. Is $\|\cdot\|$ a norm on $\mathcal{X}$?

In order to answer this question, we need to verify four properties, namely

(i) $\|X\| \geq 0$ for every $X \in \mathcal{X}$,
(ii) $\|X\| = 0$ if and only if $X = 0$,
(iii) $\|\alpha X\| = |\alpha|\|X\|$ for every $\alpha \in \mathbb{R}$ and $X \in \mathcal{X}$, and
(iv) $\|X + Y\| \leq \|X\| + \|Y\|$ for every $X, Y \in \mathcal{X}$.

We see that **(i)** is obviously true since $X^2 \geq 0$ for any $X \in \mathcal{X}$. (Indeed the square of any real number is non-negative.) As for **(iii)**, we see that if $\alpha \in \mathbb{R}$, then

$$\|\alpha X\| = \sqrt{\mathbb{E}[(\alpha X)^2]} = \sqrt{\alpha^2 \mathbb{E}(X^2)} = |\alpha| \sqrt{\mathbb{E}(X^2)} = |\alpha|\|X\|.$$

The trouble comes when we try to verify **(ii)**. One direction is true, namely that if $X = 0$, then $\mathbb{E}(X^2) = 0$ so that $\|X\| = 0$. However, if $\|X\| = 0$ so that $\mathbb{E}(X^2) = 0$, then it need not be the case that $X = 0$.

Here is one such counterexample. Suppose that we define the random variable X to be 0 if a head appears on a toss of a fair coin and to be 0 if a tail appears. If the coin lands on its side, define X to be 1. It then follows that

$$\mathbb{E}(X^2) = 0^2 \cdot \mathbf{P}\{\text{head}\} + 0^2 \cdot \mathbf{P}\{\text{tail}\} + 1^2 \cdot \mathbf{P}\{\text{side}\} = 0 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} + 1 \cdot 0 = 0.$$

This shows that it is theoretically possible to define a random variable $X \neq 0$ such that $\|X\| = 0$. In other words, $X = 0$ with probability 1, but X is not identically 0. Since **(ii)** fails, we see that $\|\cdot\|$ is not a norm.

However, it turns out that **(iv)** actually holds. In order to verify that this is so, we need the following lemma.

Lemma. *If a, $b \in \mathbb{R}$, then*

$$ab \leq \frac{a^2}{2} + \frac{b^2}{2}.$$

Proof. Clearly $(a-b)^2 \geq 0$. Expanding gives $a^2 + b^2 - 2ab \geq 0$ so that $a^2 + b^2 \geq 2ab$ as required. □

Let X, $Y \in \mathcal{X}$ and consider

$$\tilde{X} = \frac{X}{\|X\|} \quad \text{and} \quad \tilde{Y} = \frac{Y}{\|Y\|}$$

so that

$$\|\tilde{X}\| = \sqrt{\mathbb{E}(\tilde{X}^2)} = \sqrt{\mathbb{E}\left(\frac{X^2}{\|X\|^2}\right)} = \sqrt{\frac{\mathbb{E}(X^2)}{\|X\|^2}} = \frac{\sqrt{\mathbb{E}(X^2)}}{\|X\|} = \frac{\|X\|}{\|X\|} = 1$$

and, similarly, $\|\tilde{Y}\| = 1$. By the lemma,

$$\tilde{X}\tilde{Y} \leq \frac{\tilde{X}^2}{2} + \frac{\tilde{Y}^2}{2}$$

so that

$$\mathbb{E}(\tilde{X}\tilde{Y}) \leq \frac{1}{2}\left[\mathbb{E}(\tilde{X}^2) + \mathbb{E}(\tilde{Y}^2)\right] = \frac{1}{2}(1+1) = 1$$

since $\|\tilde{X}\|^2 = \mathbb{E}(\tilde{X}^2) = 1$ and $\|\tilde{Y}\|^2 = \mathbb{E}(\tilde{Y}^2) = 1$. In other words,

$$\mathbb{E}(\tilde{X}\tilde{Y}) = \mathbb{E}\left[\frac{X}{\|X\|}\frac{Y}{\|Y\|}\right] \leq 1$$

implies

$$\mathbb{E}(XY) \leq \|X\|\|Y\|. \tag{26.1}$$

We now use the fact that $(X+Y)^2 = X^2 + Y^2 + 2XY$ so that

$$\mathbb{E}[(X+Y)^2] = \mathbb{E}(X^2) + \mathbb{E}(Y^2) + 2\mathbb{E}(XY),$$

or equivalently, $\|X+Y\|^2 = \|X\|^2 + \|Y\|^2 + 2\mathbb{E}(XY)$. Using (26.1) we find

$$\|X+Y\|^2 \leq \|X\|^2 + \|Y\|^2 + 2\|X\|\|Y\| = (\|X\| + \|Y\|)^2.$$

Taking square roots of both sides gives

$$\|X+Y\| \leq \|X\| + \|Y\|$$

which establishes the triangle inequality **(iv)**.

Remark. We have shown that $\|X\| = \sqrt{\mathbb{E}(X^2)}$ satisfies properties **(i)**, **(iii)**, and **(iv)** only. As a result, we call $\|\cdot\|$ a *seminorm*. If we identify random variables X_1 and X_2 whenever $\mathbf{P}\{X_1 = X_2\} = 1$, then $\|\cdot\|$ also satisfies **(ii)** and is truly a norm.

It is common to write L^2 to denote the space of random variables of finite variance. In fact, if L^2 is equipped with the norm $\|X\| = \sqrt{\mathbb{E}(X^2)}$, then it can be shown that L^2 is a both a *Banach space* and a *Hilbert space*.

27

Value at Risk

Suppose that Ω denotes the set of all possible financial scenarios up to a given expiry date T. In other words, if we write $\{S_t, t \geq 0\}$ to denote the underlying asset price process that we care about, then Ω consists of all possible trajectories on $[0, T]$. We will abbreviate such a trajectory simply by ω.

Therefore, we will let the random variable X denote our financial position at time T. In other words, $X : \Omega \to \mathbb{R}$ is given by $\omega \mapsto X(\omega)$ where $X(\omega)$ describes our financial position at time T (or our resulting net worth already discounted) assuming the trajectory ω was realized.

Our goal is to quantify the *risk* associated with the financial position X. Arbitrarily, we could use $\mathrm{Var}(X)$ to measure risk. Although this is easy to work with, it is symmetric. This is not desirable in a financial context since upside risk is fine; it is perfectly acceptable to make more money than expected!

As a first example, we will consider the so-called *value at risk at level* α. Recall that $\mathcal{X}$ denotes the space of all random variables of finite variance. As we saw in Lecture #26, if we define $\|X\| = \sqrt{\mathbb{E}(X^2)}$ for $X \in \mathcal{X}$, then $\|\cdot\|$ defines a norm on $\mathcal{X}$ as long as we identify random variables which are equal with probability one.

Example 27.1. Let X be a given financial position and suppose that $\alpha \in (0, 1)$. We will say that X is *acceptable* if and only if

$$\mathbf{P}\{X < 0\} \leq \alpha.$$

We then define $\mathrm{VaR}_\alpha(X)$, the *value at risk of the position X at level $\alpha \in (0, 1)$*, to be

$$\mathrm{VaR}_\alpha(X) = \inf\{m : \mathbf{P}\{X + m < 0\} \leq \alpha\}.$$

In other words, if X is not acceptable, then the value at risk is the minimal amount m of capital that is required to be added to X in order to make it acceptable. For instance, suppose that we declare X to be acceptable if $\mathbf{P}\{X < 0\} \leq 0.1$. If X is known to have a $\mathcal{N}(1, 1)$ distribution, then X is not acceptable since $\mathbf{P}\{X < 0\} = 0.1587$. However, we find (accurate to 4 decimal places) that $\mathbf{P}\{X < -0.2816\} = 0.1$. Therefore, if $X \sim \mathcal{N}(1, 1)$, then

$$\mathrm{VaR}_{0.1}(X) = \inf\{m : \mathbf{P}\{X + m < 0\} \leq 0.1\} = 0.2816.$$

Since X was not acceptable, we see that the minimal capital we must add to make it acceptable is 0.2816. On the other hand, if $X \sim \mathcal{N}(3,1)$, then X is already acceptable and so

$$\mathrm{VaR}_{0.1}(X) = \inf\{m : \mathbf{P}\{X + m \ < 0\} \le 0.1\} = -1.7184$$

since $\mathbf{P}\{X < 1.7184\} = 0.1$. Since X was already acceptable, our value at risk is negative. This indicates that we could afford to lower our capital by 1.7184 and our position would still be acceptable. We can write $\mathrm{VaR}_\alpha(X)$ in terms of the distribution function F_X of X as follows. Let $c = -m$ so that $\mathbf{P}\{X + m < 0\} = \mathbf{P}\{X - c < 0\}$, and so

$$\mathbf{P}\{X - c < 0\} = \mathbf{P}\{X < c\} = \mathbf{P}\{X \le c\} - \mathbf{P}\{X = c\} = F_X(c-)$$

where $c-$ denotes the limit from the left. Therefore,

$$\mathrm{VaR}_\alpha(X) = \inf\{-c : F_X(c-) \le \alpha\} = -\sup\{c : F_X(c-) \le \alpha\}.$$

Although value at risk is widely used, it has a number of drawbacks. For instance, it pays attention only to shortfalls ($X < 0$), but never to how bad they are. It may also penalize diversification. Mathematically, value at risk requires a probability measure $\mathbf{P}$ to be known in advance, and it does not behave in a convex manner.

Exercise 27.2. Show that if $X \le Y$, then $\mathrm{VaR}_\alpha(X) \ge \mathrm{VaR}_\alpha(Y)$.

Exercise 27.3. Show that if $r \in \mathbb{R}$, then $\mathrm{VaR}_\alpha(X + r) = \mathrm{VaR}_\alpha(X) - r$.

As we will learn next lecture, any functional satisfying the properties given in the previous two exercises will be called a *monetary risk measure*. That is, value at risk is an example of a monetary risk measure. As we will see in Lecture #31, however, it is not a *coherent risk measure*.

28
Monetary Risk Measures

As we saw last lecture, value at risk has a number of drawbacks, including the fact that it requires a probability measure $\mathbf{P}$ to be known in advance.

Instead of using value at risk, we would like to have a measure of risk that does not require an *a priori* probability measure. Motivated by the point-of-view of functional analysis, we will consider a risk measure to be a functional on a space of random variables.

Unfortunately, we cannot work with the space of all random variables of finite variance. This is because the calculation of the variance of a random variable X requires one to compute $\mathbb{E}(X^2)$. However, expectation is computed with respect to a given probability measure, and so to compute $\mathbb{E}(X^2)$, one is required to know $\mathbf{P}$ in advance.

Thus, we need a more general setup. Suppose that Ω is the set of all possible financial scenarios, and let $X : \Omega \to \mathbb{R}$ be a function. Denote by $\mathcal{X}$ the space of all real-valued bounded functions on Ω. That is, if we define

$$\|X\|_\infty = \sup_{\omega \in \Omega} |X(\omega)|,$$

then

$$\mathcal{X} = \{X : \|X\|_\infty < \infty\}.$$

It follows from Example 25.11 that $\|\cdot\|_\infty$ defines a norm on the space of bounded functions which is sometimes called the *sup norm.*

Since $\mathcal{X}$ with the sup norm does not require a probability to be known, this is the space that we will work with from now on.

Definition 28.1. We will call a functional $\rho : \mathcal{X} \to \mathbb{R}$ a *monetary risk measure* if it satisfies

(i) monotonicity, namely $X \leq Y$ implies $\rho(X) \geq \rho(Y)$, and
(ii) translation invariance, namely $\rho(X + r) = \rho(X) - r$ for every $r \in \mathbb{R}$.

Remark. Notice that $\rho(X) \in \mathbb{R}$ so that translation invariance implies (with $r = \rho(X)$)

$$\rho(X + \rho(X)) = \rho(X) - \rho(X) = 0.$$

Furthermore, translation invariance implies (with $X = 0$)

$$\rho(r) = \rho(0) - r.$$

In many situations, there is no loss of generality in assuming that $\rho(0) = 0$. In fact, we say that a monetary risk measure is *normalized* if $\rho(0) = 0$. Notice that if ρ is normalized, then

$$\rho(r) = -r$$

for any $r \in \mathbb{R}$.

Example 28.2. Define the *worst-case risk measure* $\rho_{\max}$ by

$$\rho_{\max}(X) = -\inf_{\omega\in\Omega} X(\omega)$$

for all $X \in \mathcal{X}$. The value $\rho_{\max}(X)$ is the least upper bound for the potential loss that can occur in any scenario. Clearly

$$\inf_{\omega\in\Omega} X(\omega) \le X.$$

If ρ is *any* monetary risk measure, then monotonicity implies

$$\rho\left(\inf_{\omega\in\Omega} X(\omega)\right) \ge \rho(X).$$

As in the previous remark, translation invariance invariance implies

$$\rho\left(\inf_{\omega\in\Omega} X(\omega)\right) = \rho(0) - \inf_{\omega\in\Omega} X(\omega) = \rho(0) + \rho_{\max}(X).$$

Combined, we see

$$\rho(X) \le \rho(0) + \rho_{\max}(X).$$

Thus, if ρ is a normalized monetary risk measure, then

$$\rho(X) \le \rho_{\max}(X).$$

In this sense, $\rho_{\max}$ is the most conservative measure of risk.

Exercise 28.3. Verify that $\rho_{\max}$ is a monetary risk measure.

Theorem 28.4. *Suppose that $\rho : \mathcal{X} \to \mathbb{R}$ is a monetary risk measure. If $X, Y \in \mathcal{X}$, then*

$$|\rho(X) - \rho(Y)| \le \|X - Y\|_\infty.$$

Proof. Clearly $X \le Y + |X - Y|$ so that

$$X \le Y + \sup_{\omega\in\Omega} |X(\omega) - Y(\omega)| = Y + \|X - Y\|_\infty.$$

Since $\|X - Y\|_\infty \in \mathbb{R}$, we can use monotonicity and translation invariance to conclude

$$\rho(X) \ge \rho(Y + \|X - Y\|_\infty) = \rho(Y) - \|X - Y\|_\infty.$$

In other words,

$$\rho(Y) - \rho(X) \le \|X - Y\|_\infty.$$

Switching X and Y implies $\rho(X) - \rho(Y) \leq \|X - Y\|_\infty$. from which we conclude

$$|\rho(X) - \rho(Y)| \leq \|X - Y\|_\infty.$$

as required. □

One of the basic tenets of measuring risk is that diversification should not increase risk. This is expressed through the idea of convexity.

Definition 28.5. We say that the monetary risk measure ρ is *convex* if

$$\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y) \tag{28.1}$$

for any $0 < \lambda \leq 1$.

Remark. If ρ is a normalized, convex risk measure, then choosing $Y = 0$ in (28.1) implies

$$\rho(\lambda X) \leq \lambda\rho(X) \tag{28.2}$$

for any $0 < \lambda \leq 1$.

Remark. If $\lambda > 1$, then $\lambda^{-1} \in (0,1)$. This means that (28.2) can be written as

$$\rho(\lambda^{-1}X) \leq \lambda^{-1}\rho(X) \quad \text{or, equivalently,} \quad \lambda\rho(\lambda^{-1}X) \leq \rho(X) \tag{28.3}$$

for any $\lambda > 1$. If we now replace X in (28.3) by λX, then we obtain

$$\rho(\lambda X) \geq \lambda\rho(X) \tag{28.4}$$

for any $\lambda > 1$.

If we want to replace the inequalities in (28.2) and (28.4) with equalities, then we need something more than just convexity.

Definition 28.6. A monetary risk measure $\rho : \mathcal{X} \to \mathbb{R}$ is called *positively homogeneous* if

$$\rho(\lambda X) = \lambda\rho(X)$$

for any $\lambda > 0$.

Remark. Suppose that ρ is positively homogeneous. It then follows that $\rho(0) = 0$; in other words, ρ is normalized. Indeed, let $\lambda > 0$ be arbitrary and take $X = 0$ so that $\rho(0) = \lambda\rho(0)$. The only way that this equality can be true is if either $\lambda = 1$ or $\rho(0) = 0$. Since $\lambda > 0$ is arbitrary, we must have $\rho(0) = 0$.

We now say that a convex, positively homogeneous monetary risk measure is a *coherent risk measure.*

Definition 28.7. We will call a functional $\rho : \mathcal{X} \to \mathbb{R}$ a *coherent risk measure* if it satisfies

(i) monotonicity, namely $X \leq Y$ implies $\rho(X) \geq \rho(Y)$,
(ii) translation invariance, namely $\rho(X + r) = \rho(X) - r$ for every $r \in \mathbb{R}$,
(iii) convexity, namely $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y)$ for any $0 < \lambda \leq 1$, and
(iv) positive homogeneity, namely $\rho(\lambda X) = \lambda\rho(X)$ for any $\lambda > 0$.

Remark. It is possible to replace **(iii)** in the definition of coherent risk measure with the following:

(iii)$'$ subadditivity, namely $\rho(X + Y) \leq \rho(X) + \rho(Y)$.

Exercise 28.8. Show that a convex, positively homogeneous monetary risk measure is subadditive.

Exercise 28.9. Show that a subadditive, positively homogeneous monetary risk measure is convex.

As a result we have the following equivalent definition of coherent risk measure.

Definition 28.10. We will call a functional $\rho : \mathcal{X} \to \mathbb{R}$ a *coherent risk measure* if it satisfies

(i) monotonicity, namely $X \leq Y$ implies $\rho(X) \geq \rho(Y)$,
(ii) translation invariance, namely $\rho(X + r) = \rho(X) - r$ for every $r \in \mathbb{R}$,
(iii)$'$ subadditivity, namely $\rho(X + Y) \leq \rho(X) + \rho(Y)$, and
(iv) positive homogeneity, namely $\rho(\lambda X) = \lambda\rho(X)$ for any $\lambda > 0$.

29

Risk Measures and Their Acceptance Sets

Recall. We write $\mathcal{X}$ to denote the space of all bounded random variables $X : \Omega \to \mathbb{R}$ with norm

$$\|X\| = \sup_{\omega \in \Omega} |X(\omega)|.$$

A monetary risk measure $\rho : \mathcal{X} \to \mathbb{R}$ is a functional which satisfies

(i) monotonicity, namely $X \leq Y$ implies $\rho(X) \geq \rho(Y)$, and
(ii) translation invariance, namely $\rho(X + r) = \rho(X) - r$ for every $r \in \mathbb{R}$.

If ρ also satisfies

(iii) convexity, namely $\rho(\lambda X + (1 - \lambda)Y) \leq \lambda\rho(X) + (1 - \lambda)\rho(Y)$ for any $0 < \lambda \leq 1$, and
(iv) positive homogeneity, namely $\rho(\lambda X) = \lambda\rho(X)$ for any $\lambda > 0$,

then we say that ρ is a coherent risk measure.

Given a monetary risk measure ρ, we can define its associated *acceptance set* $\mathcal{A}_\rho$ to be

$$\mathcal{A}_\rho = \{X \in \mathcal{X} : \rho(X) \leq 0\}.$$

In other words, $\mathcal{A}_\rho \subseteq \mathcal{X}$ consists of those financial positions X for which no extra capital is needed to make them acceptable when ρ is used to measure risk.

Theorem 29.1. *If ρ is a monetary risk measure with associated acceptance set $\mathcal{A}_\rho$, then the following properties hold:*

(a) *if $X \in \mathcal{A}_\rho$ and $Y \in \mathcal{X}$ with $Y \geq X$, then $Y \in \mathcal{A}_\rho$,*
(b) $\inf\{m \in \mathbb{R} : m \in \mathcal{A}_\rho\} > -\infty$, *and*
(c) *if $X \in \mathcal{A}_\rho$ and $Y \in \mathcal{X}$, then*

$$\{\lambda \in [0, 1] : \lambda X + (1 - \lambda)Y \in \mathcal{A}_\rho\}$$

is a closed subset of $[0, 1]$.

Proof. The verification of both **(a)** and **(b)** is straightforward. As for **(c)**, consider the function

$$\lambda \mapsto \rho(\lambda X + (1-\lambda)Y).$$

It follows from Theorem 28.4 that this function is continuous. That is, for $\lambda \in [0,1]$, let

$$f(\lambda) = \rho(\lambda X + (1-\lambda)Y),$$

and note that if λ_1, $\lambda_2 \in [0,1]$, then

$$\begin{aligned} |f(\lambda_1) - f(\lambda_2)| &= |\rho(\lambda_1 X + (1-\lambda_1)Y) - \rho(\lambda_2 X + (1-\lambda_2)Y)| \\ &\le \|(\lambda_1 - \lambda_2)X + (\lambda_2 - \lambda_1)Y\|_\infty \\ &\le \|(\lambda_1 - \lambda_2)X\|_\infty + \|(\lambda_2 - \lambda_1)Y\|_\infty \\ &= |\lambda_1 - \lambda_2|\|X\|_\infty + |\lambda_2 - \lambda_1|\|Y\|_\infty \\ &= |\lambda_1 - \lambda_2|(\|X\|_\infty + \|Y\|_\infty) \end{aligned}$$

where the first inequality follows from Theorem 28.4 and the second inequality follows from the triangle inequality. Since X, $Y \in \mathcal{X}$, we have $\|X\|_\infty + \|Y\|_\infty < \infty$. Therefore, if λ_2 is fixed and $\lambda_1 \to \lambda_2$, then $f(\lambda_1) \to f(\lambda_2)$ so that f is indeed continuous. We now note that the inverse image of a closed set under a continuous function is closed. Hence, the set of $\lambda \in [0,1]$ such that $\rho(\lambda X + (1-\lambda)Y) \le 0$ is closed. □

Remark. The intuition for **(b)** is that some negative constants might be acceptable, but we cannot go arbitrarily far to $-\infty$.

Alternatively, suppose that we are given a set $\mathcal{A} \subseteq \mathcal{X}$ with the following two properties:

(a) if $X \in \mathcal{A}$ and $Y \in \mathcal{X}$ with $Y \ge X$, then $Y \in \mathcal{A}$, and
(b) $\inf\{m \in \mathbb{R} : m \in \mathcal{A}\} > -\infty$.

If we then *define* $\rho_\mathcal{A} : \mathcal{X} \to \mathbb{R}$ by setting

$$\rho_\mathcal{A}(X) = \inf\{m \in \mathbb{R} : X + m \in \mathcal{A}\},$$

then $\rho_\mathcal{A}$ is a monetary risk measure.

Exercise 29.2. Verify that $\rho_\mathcal{A}(X) = \inf\{m \in \mathbb{R} : X + m \in \mathcal{A}\}$ is, in fact, a monetary risk measure. *Both monotonicity and translation invariance are relatively straightforward to verify. The only tricky part is showing that $\rho_\mathcal{A}(X)$ is finite.*

Theorem 29.3. *If $\rho : \mathcal{X} \to \mathbb{R}$ is a monetary risk measure, then*

$$\rho_{\mathcal{A}_\rho} = \rho.$$

Proof. Suppose that $\rho : \mathcal{X} \to \mathbb{R}$ is given and let $\mathcal{A}_\rho = \{X \in \mathcal{X} : \rho(X) \le 0\}$ be its acceptance set. By definition, if $X \in \mathcal{X}$, then

$$\rho_{\mathcal{A}_\rho}(X) = \inf\{m \in \mathbb{R} : X + m \in \mathcal{A}_\rho\}.$$

However, by definition again

$$\inf\{m \in \mathbb{R} : X + m \in \mathcal{A}_\rho\} = \inf\{m \in \mathbb{R} : \rho(X + m) \le 0\}.$$

Translation invariance implies $\rho(X+m)=\rho(X)-m$ so that

$$\inf\{m\in\mathbb{R}:\rho(X+m)\le 0\}=\inf\{m\in\mathbb{R}:\rho(X)\le m\}.$$

However, $\inf\{m\in\mathbb{R}:\rho(X)\le m\}$ is precisely equal to $\rho(X)$. In other words,

$$\rho_{\mathcal{A}_\rho}(X)=\rho(X)$$

for every $X\in\mathcal{X}$ and the proof is complete. □

Theorem 29.4. *If $\mathcal{A}\subseteq\mathcal{X}$ is given, then $\mathcal{A}\subseteq\mathcal{A}_{\rho_\mathcal{A}}$.*

Proof. In order to prove $\mathcal{A}\subseteq\mathcal{A}_{\rho_\mathcal{A}}$ we must show that if $X\in\mathcal{A}$, then $X\in\mathcal{A}_{\rho_\mathcal{A}}$. Therefore, suppose that $X\in\mathcal{A}$ so that $\rho_\mathcal{A}(X)=\inf\{m:X+m\in\mathcal{A}\}\le 0$. By definition

$$\mathcal{A}_\rho=\{X\in\mathcal{X}:\rho(X)\le 0\}$$

for *any* monetary risk measure ρ. Thus, we must have $X\in\mathcal{A}_{\rho_\mathcal{A}}$ since $\rho_\mathcal{A}(X)\le 0$. □

Remark. It turns out that the converse, however, is not necessarily true. In order to conclude that $\mathcal{A}=\mathcal{A}_{\rho_\mathcal{A}}$ there must be more structure on $\mathcal{A}$. It turns out that the closure property **(c)** is sufficient.

Theorem 29.5. *Suppose $\mathcal{A}\subseteq\mathcal{X}$ satisfies the following properties:*

(a) *if $X\in\mathcal{A}_\rho$ and $Y\in\mathcal{X}$ with $Y\ge X$, then $Y\in\mathcal{A}_\rho$,*
(b) $\inf\{m\in\mathbb{R}:m\in\mathcal{A}_\rho\}>-\infty$*, and*
(c) *if $X\in\mathcal{A}_\rho$ and $Y\in\mathcal{X}$, then*

$$\{\lambda\in[0,1]:\lambda X+(1-\lambda)Y\in\mathcal{A}_\rho\}$$

is a closed subset of $[0,1]$.

It then follows that $\mathcal{A}=\mathcal{A}_{\rho_\mathcal{A}}$.

Remark. As a consequence of Therorems 29.3 and 29.5, we have a dual view of risk measures and their acceptance sets. Instead of proving a result directly for a monetary risk measure, it might be easier to work with the corresponding acceptance set. For instance, it is proved in [8] that $\rho_\mathcal{A}$ is a coherent risk measure if and only if $\mathcal{A}$ is a convex cone. Thus, if one can find an acceptance set $\mathcal{A}$ which is a convex cone, then the resulting risk measure $\rho_\mathcal{A}$ is coherent.

30

A Representation of Coherent Risk Measures

Recall from Example 28.2 that the worst-case risk measure $\rho_{\max}$ was defined by

$$\rho_{\max}(X) = -\inf_{\omega\in\Omega} X(\omega)$$

for $X \in \mathcal{X}$ where

$$\mathcal{X} = \left\{ X : \Omega \to \mathbb{R} \text{ such that } \|X\|_\infty = \sup_{\omega\in\Omega} |X(\omega)| < \infty \right\}.$$

We then showed that if ρ is any normalized monetary risk measure, then

$$\rho(X) \le \rho_{\max}(X).$$

Fact. It turns out the corresponding acceptance set for $\rho_{\max}$ is a convex cone so that $\rho_{\max}$ is actually a coherent risk measure. Since a coherent risk measure is necessarily normalized, we conclude that if ρ is a coherent risk measure, then

$$\rho(X) \le \rho_{\max}(X)$$

for any $X \in \mathcal{X}$.

We also recall that it was necessary to introduce the space $\mathcal{X}$ of bounded financial positions X since we wanted to analyze risk without regard to any underlying distribution for X.

It turns out, however, that we can introduce distributions back into our analysis of risk! Using two of the most important theorems in functional analysis, namely the Hahn-Banach theorem and the Riesz-Markov theorem, it can be shown that $\rho_{\max}(X)$ can be represented as

$$\rho_{\max}(X) = \sup_{\mathbf{P}\in\mathcal{P}} \mathbb{E}_{\mathbf{P}}(-X)$$

where $\mathcal{P}$ denotes the class of *all* probability measures on Ω and $\mathbb{E}_{\mathbf{P}}$ denotes expectation assuming that the distribution of X is induced by $\mathbf{P}$.

In other words, the representation of $\rho_{\max}(X)$ is

$$-\inf_{\omega\in\Omega} X(\omega) = \sup_{\mathbf{P}\in\mathcal{P}} \mathbb{E}_{\mathbf{P}}(-X). \tag{30.1}$$

In order to motivate this representation, we will assume that $X \leq 0$ and show that both sides of (30.1) actually equal $\|X\|_\infty$. For the left side, notice that

$$-\inf_{\omega\in\Omega} X(\omega) = \sup_{\omega\in\Omega}(-X(\omega)) = \sup_{\omega\in\Omega} |X(\omega)| = \|X\|_\infty.$$

As for the right side, it is here that we need to use the Hahn-Banach and Riesz-Markov theorems. The basic idea is the following. Suppose that the distribution $\mathbf{P}$ is given and consider $\mathbb{E}_\mathbf{P}(-X)$. Since $X \leq 0$, we see that

$$-X \leq \sup_{\omega\in\Omega}(-X(\omega)) = \|X\|_\infty$$

as above, and so

$$\mathbb{E}_\mathbf{P}(-X) \leq \mathbb{E}_\mathbf{P}(\|X\|_\infty) = \|X\|_\infty \mathbb{E}_\mathbf{P}(1) = \|X\|_\infty.$$

If we now take the supremum over all $\mathbf{P} \in \mathcal{P}$, then

$$\sup_{\mathbf{P}\in\mathcal{P}} \mathbb{E}_\mathbf{P}(-X) \leq \|X\|_\infty.$$

The Hahn-Banach and Riesz-Markov theorems say that the supremum is actually achieved. That is, there exists *some* $\mathbf{P} \in \mathcal{P}$ for which $\mathbb{E}_\mathbf{P}(-X) = \|X\|_\infty$; in other words,

$$\sup_{\mathbf{P}\in\mathcal{P}} \mathbb{E}_\mathbf{P}(-X) = \|X\|_\infty.$$

The extension to a general bounded function X (as opposed to just $X \leq 0$) is similar, but more technical.

This also motivates the representation theorem that we are about to state. Since *any* coherent risk measure ρ is bounded above by the coherent risk measure $\rho_{\max}$, and since $\rho_{\max}(X)$ can be represented as the supremum of $\mathbb{E}_\mathbf{P}(-X)$ over *all* $\mathbf{P} \in \mathcal{P}$, it seems reasonable that ρ can be represented as the supremum of $\mathbb{E}_\mathbf{P}(-X)$ over *some suitable set of* $\mathbf{P} \in \mathcal{P}$.

Theorem 30.1. *A functional $\rho : \mathcal{X} \to \mathbb{R}$ is a coherent risk measure if and only if there exists a subset $\mathcal{Q} \subseteq \mathcal{P}$ such that*

$$\rho(X) = \sup_{\mathbf{P}\in\mathcal{Q}} \mathbb{E}_\mathbf{P}(-X)$$

for $X \in \mathcal{X}$. Moreover, $\mathcal{Q}$ can be chosen as a convex set so that the supremum is attained.

This theorem says that if ρ is a given coherent risk measure, then there exists some subset $\mathcal{Q}$ of $\mathcal{P}$ such that

$$\rho(X) = \sup_{\mathbf{P}\in\mathcal{Q}} \mathbb{E}_\mathbf{P}(-X).$$

Of course, if $\mathcal{Q} \subseteq \mathcal{P}$, then

$$\sup_{\mathbf{P}\in\mathcal{Q}} \mathbb{E}_\mathbf{P}(-X) \leq \sup_{\mathbf{P}\in\mathcal{P}} \mathbb{E}_\mathbf{P}(-X)$$

which just says that $\rho(X) \leq \rho_{\max}(X)$.

31

Further Remarks on Value at Risk

We introduced the concept of value at risk in Lecture #27. One of the problems with value at risk is that it requires a probability measure to be known in advance. This is the reason that we studied monetary risk measures in general culminating with Theorem 30.1, a representation theorem for coherent risk measures.

Assume for the rest of this lecture that a probability measure is known so that we can compute the probabilities and expectations required for value at risk. Let $\mathcal{X}$ denote the space of random variables of finite variance; that is,

$$\mathcal{X} = \{X : \Omega \to \mathbb{R} \text{ such that } \|X\| = \sqrt{\mathbb{E}(X^2)} < \infty\}.$$

Recall that if $X \in \mathcal{X}$, then

$$\text{VaR}_\alpha(X) = \inf\{-c : F_X(c-) \leq \alpha\} = -\sup\{c : F_X(c-) \leq \alpha\}.$$

If X is a continuous random variable, then $F_X(c-) = F_X(c)$ and F_X is strictly increasing so that there exists a unique c such that $F_X(c) = \alpha$ or, equivalently, $c = F_X^{-1}(\alpha)$. Thus,

$$\text{VaR}_\alpha(X) = -F_X^{-1}(\alpha).$$

Example 31.1. If $X \sim \mathcal{N}(\mu, \sigma)$, then it follows from Exercise 3.4 that

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Therefore, to determine $\text{VaR}_\alpha(X)$ we begin by solving

$$\Phi\left(\frac{c-\mu}{\sigma}\right) = \alpha$$

for c. Doing so gives $c = \mu + \Phi^{-1}(\alpha)\sigma$ and so

$$\text{VaR}_\alpha(X) = -\mu - \Phi^{-1}(\alpha)\sigma.$$

We can write this in a slightly different way by noting that

$$-\Phi^{-1}(\alpha) = \Phi^{-1}(1-\alpha).$$

Indeed Exercise 3.3 implies that

$$\Phi(-\Phi^{-1}(1-\alpha)) = 1 - \Phi(\Phi^{-1}(1-\alpha)) = 1 - (1-\alpha) = \alpha$$

and so $-\Phi^{-1}(1-\alpha)) = \Phi^{-1}(\alpha)$ as required. That is,

$$\mathrm{VaR}_\alpha(X) = -\mu + \Phi^{-1}(1-\alpha)\sigma.$$

Finally, since $\mathbb{E}(X) = \mu$ and $\mathrm{SD}(X) = \sigma$ we have

$$\mathrm{VaR}_\alpha(X) = \mathbb{E}(-X) + \Phi^{-1}(1-\alpha)\,\mathrm{SD}(X).$$

Example 31.2. Suppose that X has a Pareto distribution with scale parameter $\theta > 0$ and shape parameter $p > 1$ so that

$$F_X(x) = 1 - \left(\frac{\theta}{x+\theta}\right)^p$$

for $x > 0$. Solving

$$1 - \left(\frac{\theta}{c+\theta}\right)^p = \alpha$$

for c implies

$$c = \theta(1-\alpha)^{-1/p} - \theta$$

so that

$$\mathrm{VaR}_\alpha(X) = \theta - \theta(1-\alpha)^{-1/p}.$$

Exercise 31.3. If X has a Raleigh distribution with parameter $\theta > 0$ so that

$$f_X(x) = \frac{2}{\theta} x e^{-x^2/\theta}, \quad x > 0,$$

determine $\mathrm{VaR}_\alpha(X)$.

Exercise 31.4. If X has a binomial distribution with parameters $n = 3$ and $p = 1/2$, determine $\mathrm{VaR}_\alpha(X)$ for $\alpha = 0.1$ and $\alpha = 0.3$.

If X is a random variable, we define the *Sharpe ratio* to be

$$\frac{\mathbb{E}(X)}{\mathrm{SD}(X)}.$$

We will say that X is *acceptable at level* ℓ if

$$\frac{\mathbb{E}(X)}{\mathrm{SD}(X)} \geq \ell$$

so that the corresponding acceptance set is

$$\mathcal{A}_\ell = \left\{X \in L^2 : \frac{\mathbb{E}(X)}{\mathrm{SD}(X)} \geq \ell\right\} = \{X \in L^2 : \mathbb{E}(X) \geq \ell\,\mathrm{SD}(X)\}.$$

As in Lecture #29, the associated risk measure is

$$\begin{aligned}\rho_\ell(X) = \inf\{m \in \mathbb{R} : X + m \in \mathcal{A}_\ell\} &= \inf\{m \in \mathbb{R} : \mathbb{E}(X+m) \geq \ell\,\mathrm{SD}(X+m)\}\\ &= \inf\{m \in \mathbb{R} : m \geq \mathbb{E}(-X) + \ell\,\mathrm{SD}(X)\}\\ &= \mathbb{E}(-X) + \ell\,\mathrm{SD}(X).\end{aligned}$$

Remark. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\mathrm{VaR}_\alpha(X)$ is of the form specified by the Sharpe ratio, namely

$$\rho_\ell(X) = \mathbb{E}(-X) + \ell\,\mathrm{SD}(X)$$

with $\ell = \Phi^{-1}(1-\alpha)$.

We will now show that $\rho_\ell(X)$ is *not*, in general, a monetary risk measure.

Let $X = e^Z$ with $Z \sim \mathcal{N}(0, \sigma^2)$ so that X has a lognormal distribution. Using Exercise 3.25, we find

$$\rho_\ell(X) = \mathbb{E}(-X) + \ell\,\mathrm{SD}(X) = -e^{\sigma^2/2} + \ell e^{\sigma^2/2}\sqrt{e^{\sigma^2}-1} = -e^{\sigma^2/2}\left[1 - \ell\sqrt{e^{\sigma^2}-1}\right].$$

Monetary risk measures must satisfy monotonicity; in particular, if $X \geq 0$, then $\rho(X) \leq 0$. However, with $X = e^Z$ we see that $X \geq 0$, but we can choose σ sufficiently large to guarantee $\rho_\ell(X) \geq 0$. That is,

$$1 - \ell\sqrt{e^{\sigma^2}-1} \leq 0$$

if and only if

$$\sigma \geq \sqrt{\log(\ell^{-2}+1)}.$$

Thus, $\rho_\ell(X)$ is not a monetary risk measure.

Remark. This does not show that value at risk is not a monetary risk measure. As we saw in the exercises at the end of Lecture #27, value at risk *is* a monetary risk measure. However, it can be shown that value at risk is *not* a coherent risk measure.

Finally, value at risk is the basis for the following coherent risk measure which has been called *tail value at risk*, *conditional tail expectation*, *tail conditional expectation*, *expected shortfall*, *conditional value at risk*, and *average value at risk*.

Let $X \in \mathcal{X}$ be given. For $0 < \alpha \leq 1$, the *average value at risk at level* α is given by

$$\mathrm{AVaR}_\alpha(X) = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_x(X)\,\mathrm{d}x.$$

It is shown in [8] that average value at risk is a coherent risk measure.

Exercise 31.5. If $X \sim \mathcal{N}(\mu, \sigma^2)$, determine $\mathrm{AVaR}_\alpha(X)$.

Bibliography

[1] Martin Baxter and Andrew Rennie. *Financial Calculus.* Cambridge University Press, Cambridge, United Kingdom, 1996.

[2] Tomas Björk. *Arbitrage Theory in Continuous Time.* Oxford University Press, Oxford, Great Britain, 1998.

[3] Nicolas Bouleau. *Financial Markets and Martingales.* Springer-Verlag, New York, NY, 2004.

[4] Marek Capiński and Tomasz Zastawniak. *Mathematics for Finance.* Springer, London, UK, 2003.

[5] Pavel Čížek, Wolfgang Härdle, and Rafał Weron. *Statistical Tools for Finance and Insurance.* Springer, Berlin, Germany, 2005.

[6] Harold Cramér. *Random Variables and Probability Distributions.* Cambridge University Press, London, England, 1963.

[7] Alison Etheridge. *A Course in Financial Calculus.* Cambridge University Press, Cambridge, United Kingdom, 2002.

[8] Hans Föllmer and Alexander Schied. *Stochastic Finance*, volume 27 of *de Gruyter Studies in Mathematics.* Second Edition. de Gruyter, Berlin, Germany, 2004.

[9] Allan Gut. *An Intermediate Course in Probability.* Springer, New York, NY, 1995.

[10] Allan Gut. *Probability: A Graduate Course.* Springer, New York, NY, 2005.

[11] David G. Heath and Glen Swindle, editors. *Introduction to Mathematical Finance (San Diego, California, January 1997)*, volume 57 of *Proceedings of Symposia in Applied Mathematics.* American Mathematical Society, Providence, RI, 1999.

[12] Desmond J. Higham. *An Introduction to Financial Option Valuation.* Cambridge University Press, Cambridge, United Kingdom, 2004.

[13] John C. Hull. *Options, Futures, and Other Derivatives.* Seventh Edition. Pearson Prentice Hall, Upper Saddle River, NJ, 2009.

[14] Robert Jarrow and Philip Protter. A short history of stochastic integration and mathematical finance: The early years, 1880–1970. In Anirban DasGupta, editor, *A Festschrift for Herman Rubin*, volume 45 of *Institute of Mathematical Statistics Lecture Notes–Monograph Series*, pages 75–91. Institute of Mathematical Statistics, Beachwood, OH, 2004.

[15] Mark Joshi, Andrew Downes, and Nick Denson. *Quant Job Interview Questions and Answers.* CreateSpace, Seattle, WA, 2008.

[16] Ralf Korn and Elke Korn. *Option Pricing and Portfolio Optimization: Modern Methods of Financial Mathematics*, volume 31 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2001.

[17] Hui-Hsing Kuo. *Introduction to Stochastic Integration.* Springer, New York, NY, 2006.

[18] Robert McDonald. *Derivative Markets.* Second Edition. Addison Wesley, Reading, MA, 2006.

[19] A. V. Mel'nikov. *Financial Markets*, volume 184 of *Translations of Mathematical Monographs.* American Mathematical Society, Providence, RI, 1999.

[20] Steven Roman. *Introduction to the Mathematics of Finance.* Springer-Verlag, New York, NY, 2004.

[21] Walter Rudin. *Principles of Mathematical Analysis.* Third Edition. McGraw-Hill, New York, NY, 1976.