Mathematical Modeling
in Economics and Finance:
Probability, Stochastic Processes
and Differential Equations

Steven R. Dunbar

Department of Mathematics, University of Nebraska-Lincoln, Lincoln, Nebraska 68588
E-mail address: sdunbar1@unl.edu
2010 Mathematics Subject Classification. Primary 91Bxx, 91Gxx, 97Mxx;
Secondary 60-01, 65Cxx, 35Q91

Key words and phrases. mathematical finance, economics, mathematical
modeling, probability, stochastic processes, differential equations
To my wife Charlene, who manages finances so well.
Contents

Preface ix

Chapter 1. Background 1
  1.1. Brief History of Mathematical Finance 1
  1.2. Options and Derivatives 5
  1.3. Speculation and Hedging 10
  1.4. Arbitrage 14
  1.5. Mathematical Modeling 17
  1.6. Randomness 27
  1.7. Stochastic Processes 31
  1.8. A Model of Collateralized Debt Obligations 36

Chapter 2. Binomial Models 45
  2.1. Single Period Binomial Models 45
  2.2. Multiperiod Binomial Tree Models 50

Chapter 3. First Step Analysis 57
  3.1. A Coin Tossing Experiment 57
  3.2. Ruin Probabilities 61
  3.3. Duration of the Gambler’s Ruin 69
  3.4. A Stochastic Process Model of Cash Management 75

Chapter 4. Limit Theorems for Coin Tossing 87
  4.1. Laws of Large Numbers 87
  4.2. Moment Generating Functions 91
  4.3. The Central Limit Theorem 95

Chapter 5. Brownian Motion 103
  5.1. Intuitive Introduction to Diffusions 103
  5.2. The Definition of Brownian Motion and the Wiener Process 107
  5.3. Approximation of Brownian Motion by Coin-Flipping Sums 114
  5.4. Transformations of the Wiener Process 118
  5.5. Hitting Times and Ruin Probabilities 122
  5.6. Path Properties of Brownian Motion 126
  5.7. Quadratic Variation of the Wiener Process 130

Chapter 6. Stochastic Calculus 137
  6.1. Stochastic Differential Equations 137
  6.2. Itô’s Formula 144
  6.3. Properties of Geometric Brownian Motion 148
  6.4. Models of Stock Market Prices 155
| Chapter 7. The Black-Scholes Equation | 165 |
| 7.1. Derivation of the Black-Scholes Equation | 165 |
| 7.2. Solution of the Black-Scholes Equation | 169 |
| 7.3. Put-Call Parity | 178 |
| 7.4. Implied Volatility | 185 |
| 7.5. Sensitivity, Hedging and the Greeks | 189 |
| 7.6. Limitations of the Black-Scholes Model | 196 |

| Chapter 8. Notes | 203 |
| Bibliography | 207 |
| Index | 211 |
Preface

**History of the Book.** This book started with one purpose and ended with a different purpose. In 2002, a former student, one of the best I had taught, approached me with a book about mathematical finance in his hand. He wanted a reading course about the subject because he was thinking about a career in the area. I flipped through the book briefly, and saw that it was too advanced for a reading course even with a very good student. I was aware that the topic combined probability theory and partial differential equations, both interests of mine. Instead of a reading course, I agreed to conduct a seminar on mathematical finance, if enough others had an interest. There were others, including undergraduates, graduate students in finance and economics and even some faculty.

I soon found that there were no books or introductions to the subject suitable for mathematics students at the upper undergraduate level. I began to gather my seminar notes and organize them.

After three years of the seminar, it grew into a popular course for senior-level students from mathematics, finance, actuarial science, computer science and engineering. The variety of students and their backgrounds refined the content of the course. The course focus was on combining finance concepts, especially derivative securities, with probability theory, difference equations and differential equations to derive consequences, primarily about option prices.

In late 2008, security markets convulsed and the U. S. economy went into a deep recession. The causes were many, and are still debated, but one underlying cause was because mathematical methods had been applied in financial situations where they did not apply [66]. At the same time for different reasons, mathematical professional organizations urged a new emphasis on mathematical modeling. The course and the associated notes evolved in response, with an emphasis on uses and abuses of modeling.

Additionally, a new paradigm in mathematical sciences combining modeling, statistics, visualization, and computing with large data sets, sometimes called “big data” or more formally data analytics, was maturing and becoming common. Data analytics is now a source of employment for many mathematics majors. The topic of finance and economics is a leader in data analytics because of the existing large data sets and the measurable value in exploiting the data.

The result is the current book combining modeling, probability theory, difference and differential equations focused on quantitative reasoning, data analysis, probability, and statistics for economics and finance. The book uses all of these topics to investigate modern financial instruments that have enormous economic influence, but are hidden from popular view because many people wrongly believe these topics are esoteric and difficult.


**Purpose of the Book.** The purpose is to provide a textbook for a capstone course focusing on mathematical modeling in economics and finance. There are already many fine books about mathematical modeling in physical and biological sciences. This text is for an alternative course for students interested in the economic sciences instead of the classical sciences. This book combines mathematical modeling, probability theory, difference and differential equations, numerical solution and simulation and mathematical analysis in a single course for undergraduates in mathematical sciences. I hope the style is engaging enough that it can also be enjoyably read as an introduction by any individual interested in these topics.

I understand that this introductory modeling approach makes serious concessions to completeness and depth, financial accuracy and mathematical rigor. Phillip Protter is an expert on mathematical finance and in a review of an elementary text on mathematical finance [51] he makes the following remarks:

Mathematical finance . . . is a difficult subject, requiring a broad array of knowledge of subjects that are traditionally considered hard to learn.

The mathematics involved in the Black-Scholes paradigm is measure-theoretic probability theory, Brownian motion, stochastic processes including Markov processes and martingale theory, Ito’s stochastic calculus, stochastic differential equations, and partial differential equations. Those prerequisites give one entry to the subject, which is why it is best taught to advanced Ph.D. students. One might expect an American undergraduate to know calculus-based probability theory and to have had some exposure to PDEs and perhaps, if one is lucky, an economics course or two, but not much more. Therefore, any attempt to teach such a subject to undergraduates is fraught with compromise . . .

Perhaps it is the same with mathematical finance: it simply is not (yet?) meant to be an undergraduate subject. In a way that is too bad, because the subject is beautiful and powerful, and expertise in it is much needed in industry.

Combining economic and financial modeling with probability, stochastic processes, and differential equations along with quantitative reasoning, and data analysis with some simulation and computing provides an inviting entry into deeper aspects of this “beautiful and powerful” subject.

The goals of the book are:

1. Understand the properties of stochastic processes such as sequences of random variables, coin-flipping games, Brownian motion and the solutions of stochastic differential equations as a means for modeling financial instruments for the management of risk.

2. Use financial instruments for the management of risk as motivations for the detailed study of stochastic processes and solutions of stochastic differential equations.

3. Introduce standard stochastic processes at the level of the classic references by Karlin and Taylor, and Feller. The book proves some mathematical statements at the level of elementary analysis, some more advanced statements have heuristic motivation without proof, and some advanced results are stated without proof.
(4) Emphasize the mathematical modeling process applied to a modern area that is not based on physical science yet still leads to classical partial differential equations and numerical methods. The field of mathematical finance is only 50 years old, uses leading-edge mathematical and economic ideas, and has some controversial foundational hypotheses. Mathematical finance is also data-rich and even advanced results are testable in the market. Using ideas illustrated daily in financial news, the book applies the full cycle of mathematical modeling and analysis in a non-trivial, but still accessible, way that has economic applications.

(5) The goal of the book is to reach a point where the students thoroughly understand the derivation and modeling of financial instruments, advanced financial models, advanced stochastic processes, partial differential equations, and numerical methods at a level sufficient for beginning graduate study in mathematics, finance, economics, actuarial science, and for entry-level positions in the sophisticated financial services industry.

The general area of stochastic processes and mathematical finance has many textbooks and monographs already. This book differs from them in the following ways:

(1) Most books on stochastic processes have a variety of applications, while this book concentrates on financial instruments for the management of risk as motivations for the detailed study of mathematical modeling with stochastic processes. The emphasis is on the modeling process, not the financial instruments.

(2) Most books on mathematical finance assume either prerequisite knowledge about financial instruments or sophisticated mathematical methods, especially measure-based probability theory and martingale theory. This book serves as a introductory preparation for those texts.

(3) This book emphasizes the practice of mathematical modeling, including post-modeling analysis and criticism, making it suitable for a wider audience.

Overall, this book is an extended essay in mathematical modeling applied to financial instruments from the simplest binomial option models to the Black-Scholes-Merton model, with some excursions along the way.

**Intended Audience and Background.** This book is primarily for undergraduate students in mathematics, economics, finance, and actuarial science. Students in physical sciences, computer science and engineering will also benefit from the book with its emphasis on modeling and the uses and limits of modeling. Graduate students in economics, finance and business benefit from the non-measure theoretic based introduction to mathematical finance and mathematical modeling.

This book is for students after a course on calculus-based probability theory. To understand the explanations and complete the exercises:

(1) The reader should be able to calculate joint probabilities of independent events.

(2) The reader should be able to calculate binomial probabilities and normal probabilities using direct calculation, tables and computer or calculator applications.
(3) The reader should be able to recognize common probability distributions such as binomial probabilities and calculate probabilities from them.

(4) The reader should be able to calculate means and variances for common probability distributions.

(5) The reader should be familiar with common statistical concepts of parameter point evaluations and confidence intervals and hypothesis testing.

(6) The reader should have a familiarity with compound interest calculations, both continuous compounding and periodic compounding.

(7) The reader should be able to perform interest calculations to find present values, future values, and simple annuities.

The text also assumes general knowledge of linear algebra, especially about solutions of linear non-homogeneous equations in linear spaces. A familiarity with solving difference equations, also called recurrence equations and recursions, is helpful, but not essential. Where needed, the solution of the specific difference equations uses elementary methods without reference to the general theory. Likewise, a familiarity with differential equations is helpful but not essential since the text derives specific solutions when necessary, again without reference to the general theory. Naturally, a course in differential equations will deepen understanding and provide another means for discussing mathematical modeling, since that is often the course where many students first encounter significant mathematical modeling of physical and biological phenomena. Concepts from linear algebra also enter into the discussions about Markov processes, but this text does not make the deeper connections. Ideas from linear algebra pervade the operators and data structures in the program scripts.

Program Scripts. An important feature of this book is the simulation scripts in the R language that accompany most sections. Scripts are in the R language because it is a popular open source language widely used for data analysis. The simulation scripts illustrate the concepts and theorems of the section with numerical and graphical evidence. The scripts are part of the “Rule of 3” teaching philosophy of presenting mathematical ideas symbolically, numerically and graphically.

The programs are springboards for further experimentation, not finished apps for casual everyday use. The scripts are minimal in size, in scope of implementation and with minimal output. The scripts are not complete, stand-alone, polished applications, rather they are proof-of-concept starting points. The reader should run the scripts to illustrate the ideas and provide numerical examples of the results in the section. The scripts provide a seed for new scripts to increase the size and scope of the simulations. Increasing the size can often demonstrate convergence or increase confidence in the results of the section. Increasing the size can also demonstrate that although convergence is mathematically guaranteed, sometimes the rate of convergence is slow. The reader is also encouraged to change the output, to provide more information, to add different graphical representations, and to investigate rates of convergence.

The scripts are not specifically designed to be efficient, either in program language implementation or in mathematical algorithm. Efficiency is not ignored, but it is not the primary consideration in the construction of the scripts. Similarity of the program algorithm to the mathematical idea takes precedence over efficiency. One noteworthy aspect of both the similarity and efficiency is the use of vectorization along with other notational simplifications such as recycling. Vectorized scripts
look more like the mathematical expressions found in the text, making the code easier to understand. Vectorized code often runs much faster than the corresponding code containing loops.

The scripts are not intended to be a tutorial on how to do mathematical programming in R. A description of the algorithm used in the scripts is in each section. The description is usually in full sentences rather than the more formal symbolic representation found in computer science pseudo-code. Given the description and some basic awareness of programming ideas, the scripts provide multiple examples for study. The scripts provide a starting point for investigating, testing and comparing language features from the documentation or from other sources. The scripts use good programming style whenever possible, but clarity, simplicity and similarity to the mathematics are primary considerations.

Connections to MAA CUPM guidelines. The nature of the text as an interdisciplinary capstone text intentionally addresses each of the cognitive and content recommendations from the Mathematical Association of America’s Committee on the Undergraduate Curriculum for courses and programs in mathematical sciences.

Cognitive Recommendations.

(1) Students should develop effective thinking and communication skills.

An emphasis in the text is on development, solution and subsequent critical analysis of mathematical models in economics and finance. Exercises in most sections ask students to write comparisons and critical analyses of the ideas, theories, and concepts.

(2) Students should learn to link applications and theory.

The entire text is committed to linking methods and theories of probability and stochastic processes and difference and differential equations to modern applications in economics and finance. Each chapter has a specific application of the methods to a model in economics and finance.

(3) Students should learn to use technological tools.

Computing examples in the modern programming language R appear throughout and many exercises encourage further adaptation and experimentation with the scripts.

(4) Students should develop mathematical independence and experience open ended inquiry.

Many exercises encourage further experimentation, data exploration, and up-to-date comparisons with new or extended data.

Content Recommendations.

(1) Mathematical sciences major programs should include concepts and methods from calculus and linear algebra.

The text makes extensive use of calculus for continuous probability and through differential equations. The text extends some of the ideas of calculus to the domain of stochastic calculus. Linear algebra appears throughout, from the theory of solutions of linear non-homogeneous equations in linear spaces to operators and data structures in the program scripts.
(2) Students majoring in the mathematical sciences should learn to read, understand, analyze, and produce proofs, at increasing depth as they progress through a major.

Mathematical proofs are not emphasized, because rigorous methods for stochastic processes need measure-theoretic tools from probability. However, when elementary tools from analysis are familiar to students, then the text provides proofs or proof sketches. Derivation of the solution of specific difference equations and differential equations appears in detail but without reference to general solution methods.

(3) Mathematical sciences major programs should include concepts and methods from data analysis, computing, and mathematical modeling.

Mathematical modeling in economics and finance is the reason for this book. Collection and analysis of economic and financial data from public sources is emphasized throughout, and the exercises extend and renew the data. Providing extensive simulation of concepts through computing in modern scripting languages is provided throughout. Exercises encourage the extension and adaptation of the scripts for more simulation and data analysis.

(4) Mathematical sciences major programs should present key ideas and concepts from a variety of perspectives to demonstrate the breadth of mathematics.

As a text for a capstone course in mathematics, the text uses the multiple perspectives of mathematical modeling, ideas from calculus, probability and statistics, difference equations, and differential equations, all for the purposes of a deeper understanding of economics and finance. The book emphasizes mathematical modeling as a motivation for new mathematical ideas, and the application of known mathematical ideas as a framework for mathematical models. The book emphasizes difference and differential equations to analyze stochastic processes. The analogies between fundamental ideas of calculus and ways to analyze stochastic processes is also emphasized.

(5) All students majoring in the mathematical sciences should experience mathematics from the perspective of another discipline.

The goal of the text focuses on mathematics modeling as a tool for understanding economics and finance. Collection and analysis of economic and financial data from public sources using the mathematical tools is emphasized throughout.

(6) Mathematical sciences major programs should present key ideas from complementary points of view: continuous and discrete; algebraic and geometric; deterministic and stochastic; exact and approximate.

The text consistently moves from discrete models in finance to continuous models in finance by developing discrete methods in probability into continuous time stochastic process ideas. The text emphasizes the differences between exact mathematics and approximate models.

(7) Mathematical sciences major programs should require the study of at least one mathematical area in depth, with a sequence of upper-level courses.
The text is for a capstone course combining significant mathematical modeling using probability theory, stochastic processes, difference equations, differential equations to understand economic and finance at a level beyond the usual undergraduate approach to economic and finance using only calculus ideas.

(8) Students majoring in the mathematical sciences should work, independently or in a small group, on a substantial mathematical project that involves techniques and concepts beyond the typical content of a single course.

Many of the exercises, especially those that extend the scripts or that call for more data and data analysis are suitable for projects done either independently or in small groups.

(9) Mathematical sciences major programs should offer their students an orientation to careers in mathematics.

Financial services, banking, insurance and risk, financial regulation, and data analysis combined with some knowledge of computing are all growth areas for careers for students from the mathematical sciences. This text is an introduction to all of those areas.

Thanks. Finally, I thank Stan Seltzer and the reviewers for many helpful suggestions and corrections.

Steven R. Dunbar
CHAPTER 1

Background

1.1. Brief History of Mathematical Finance

Section Starter Question. Name as many financial instruments as you can, and name or describe the market where you would buy them. Also describe the instrument as high risk or low risk.

Introduction. Two common sayings are *compound interest is the eighth wonder of the world* and *the stock market is just a big casino*. These are colorful sayings but each focuses on only one aspect of one financial instrument. Combined, the time value of money and uncertainty are the central elements influencing the value of financial instruments. Considering only the time aspect of finance, the tools of calculus and differential equations are adequate. When considering only the uncertainty, the tools of probability theory compute the possible outcomes. Considering time and uncertainty together, we begin the study of advanced mathematical finance.

Finance theory is the study of economic agents’ behavior allocating financial resources and risks across alternative financial instruments over time in an uncertain environment. Familiar examples of financial instruments are bank accounts, loans, stocks, government bonds and corporate bonds. Many less familiar examples abound. Economic agents are units who buy and sell financial resources in a market. Typical economic agents are individual investors, banks, businesses, mutual funds and hedge funds. Each agent has many choices of where to buy, sell, invest and consume assets. Each choice comes with advantages and disadvantages. An agent distributes resources among the many possible investments with a goal in mind, often maximum return or minimum risk.

Mathematical finance is the study of more sophisticated financial instruments. A *derivative* is a financial agreement between two parties that depends on the future price or performance of an underlying asset. Derivatives are so called not because they involve a rate of change, but because their value is *derived* from the underlying asset. The underlying asset could be a stock, a bond, a currency or a commodity. Derivatives have become one of the financial world’s most important risk-management tools. Finance is about allocating risk and derivatives are especially efficient for that purpose [44].

Two common derivatives are futures and options. Futures trading, a key practice in modern finance, probably originated in seventeenth century Japan, but the idea goes as far back as ancient Greece. Options were a feature of the “tulip mania” in seventeenth century Holland.

Derivatives come in many types. The most common examples are futures, agreements to trade something at a set price at a given date; options, the right but not the obligation to buy or sell at a given price; forwards, like futures but
traded directly between two parties instead of on exchanges; and swaps, exchanging flows of income from different investments to manage different risk exposure. For example, one party in a deal may want the potential of rising income from a loan with a floating interest rate, while the other might prefer the predictable payments ensured by a fixed interest rate. The name of this elementary swap is a *plain vanilla swap*. More complex swaps mix the performance of multiple income streams with varieties of risk [44]. Another more complex swap is a *credit-default swap* in which a seller receives a regular fee from the buyer in exchange for agreeing to cover losses arising from defaults on the underlying loans. These swaps are somewhat like insurance [44]. These more complex swaps are the source of controversy since many people believe that they are responsible for the collapse or near-collapse of several large financial firms in late 2008. As long as two parties are willing to trade risks and can agree on a price they can craft a corresponding derivative from any financial instrument. Businesses use derivatives to shift risks to other firms, chiefly banks. About 95% of the world’s 500 biggest companies use derivatives. Markets called exchanges are the usual place to buy and sell derivatives with standardized terms. Derivatives tailored for specific purposes or risks are bought and sold “over the counter” from big banks. The “over the counter” market dwarfs the exchange trading. In November 2009, the Bank for International Settlements put the face value of over the counter derivatives at $604.6 trillion. Using face value is misleading, after stripping out off-setting claims the residual value is $3.7 trillion, still a large figure [62].

Mathematical models in modern finance contain beautiful applications of differential equations and probability theory. Additionally, mathematical models of modern financial instruments have had a direct and significant influence on finance practice.

**Early History.** Louis Bachelier originated mathematical financial modelling in his doctoral thesis on the theory of speculation in the Paris markets completed at the Sorbonne in 1900. His dissertation considered both continuous time stochastic processes and the continuous time economics of option pricing. Bachelier provided the first mathematical analysis of what is now called the *Wiener process* or *Brownian motion*. After Bachelier, mathematical modeling in finance laid mostly dormant until economists and mathematicians renewed study of it in the late 1960s. Jarrow and Protter [25] speculate that this may have been because the Paris mathematical elite scorned economics as an application of mathematics.

Bachelier’s work was 5 years before Albert Einstein’s famous mathematical theory of Brownian motion in 1905. Einstein proposed a model for the motion of small particles with diameters on the order of 0.001 mm suspended in a liquid. He predicted that the particles would undergo microscopically observable and statistically predictable motion. The English botanist Robert Brown had already reported such motion in 1827 while observing pollen grains in water with a microscope. The physical motion is now called *Brownian motion* in honor of Brown’s description.

The paper was Einstein’s justification of the molecular and atomic nature of matter. Even in 1905 the scientific community did not completely accept the atomic theory of matter. In 1908, the experimental physicist Jean-Baptiste Perrin conducted a series of experiments that empirically verified Einstein’s theory. Perrin thereby determined the physical constant known as Avogadro’s number for which he won the Nobel prize in 1926. Nevertheless, Einstein’s theory was difficult to
rigorously justify mathematically. In a series of papers from 1918 to 1923, the mathematician Norbert Wiener constructed a mathematical model of Brownian motion. Wiener and others proved many surprising facts about his mathematical model of Brownian motion, research that continues today. In recognition of his work, his mathematical construction is often called the Wiener process.

**Growth of Mathematical Finance.** Modern mathematical finance theory begins in 1965 when the economist Paul Samuelson published two papers that argue that stock prices fluctuate randomly. One explained the Samuelson and Fama efficient markets hypothesis that in a well-functioning and informed market the best estimate of an asset’s future price is the current price, possibly adjusted for a fair expected rate of return. Under this hypothesis, past price data or publicly available forecasts about economic fundamentals do not help to predict security prices. In the other paper with mathematician Henry McKean, Samuelson shows that a good model for stock price movements is geometric Brownian motion. Samuelson noted that Bachelier’s model failed to ensure that stock prices would always be positive, whereas geometric Brownian motion avoids this possibility.

The most important development was the 1973 Black-Scholes model for option pricing. The two economists Fischer Black and Myron Scholes (and simultaneously, and somewhat independently, the economist Robert Merton) created a mathematical model for calculating the prices of options. Their key insight is modeling the random variation of the underlying asset in order to remove it by hedging. The formal press release from the Royal Swedish Academy of Sciences announcing the 1997 Nobel Prize in Economics states that they gave the honor “for a new method to determine the value of derivatives. Robert C. Merton and Myron S. Scholes have, in collaboration with the late Fischer Black developed a pioneering formula for the valuation of stock options. Their methodology has paved the way for economic valuations in many areas. It has also generated new types of financial instruments and facilitated more efficient risk management in society.”

The Chicago Board Options Exchange (CBOE) began publicly trading options in the United States in April 1973, a month before the official publication of the Black-Scholes model. By 1975, traders on the CBOE were using the model to both price and hedge their options positions. In fact, Texas Instruments even created a special hand-held calculator programmed to produce Black-Scholes option prices and hedge ratios.

The basic insight underlying the Black-Scholes model is that a dynamic portfolio trading strategy in the stock can replicate the returns from an option on that stock. This is hedging an option and it is the most important idea underlying the Black-Scholes-Merton approach. Much of the rest of the book will explain what that insight means and how to apply it to calculate option values.

The history of the Black-Scholes-Merton option pricing model is that Black started working on this problem by himself in the late 1960s. He found the option value satisfies a partial differential equation. Black then teamed up with Myron Scholes. Together they solved the partial differential equation using a combination of economic intuition and earlier pricing formulas.

At this time, Myron Scholes was at MIT as was Robert Merton, who had a PhD in economics under Paul Samuelson and a background in engineering mathematics. Merton was the first to call the solution the Black-Scholes option pricing formula. Merton provided an alternative derivation of the formula using a perfectly hedged
portfolio of the stock and the call option together with the notion that no arbitrage opportunities exist. This is the approach we will take. In the late 1970s and early 1980s mathematicians J. Harrison, D. Kreps and S. Pliska showed that a more abstract formulation of the solution as a mathematical model called a martingale provides greater generality.

In the 1980s, the development of sophisticated mathematical models and their adoption into financial practice accelerated. A wave of de-regulation in the financial sector was an important element driving innovation.

Conceptual breakthroughs in finance theory in the 1980s were fewer and less fundamental than in the 1960s and 1970s. The personal computer and increases in computer speed and memory enabled new financial markets and expansions in the size of existing ones. These same technologies made the numerical solution of complex models possible. Faster computers also speeded up the solution of existing models to allow virtually real-time calculations of prices and hedge ratios.

**Ethical considerations.** Prior to the 1970s, mathematical models had a limited influence on finance theory and practice. Since the introduction in 1973 of the Black-Scholes-Merton ideas these models have become central in all financial markets. In the future, mathematical models will remain central in the functioning of the global financial system. The ideas will also be a cornerstone for creating regulations and accounting principles to govern the system.

The ideas and models introduced in the 1970s and 1980s sparked a huge expansion in the size, scope and influence of finance in the economy. In 1995, the sector composed of finance, insurance, and real estate overtook the manufacturing sector in America’s gross domestic product. By the year 2000 this sector led manufacturing in profits [49]. The growth, driven in part by new financial products including complex and exotic options, was largely unregulated.

The application of mathematical models in finance practice can be taken to an extreme. At times, the mathematics of the models becomes so interesting that we lose sight of the models’ ultimate purpose. The mathematics is precise, but the models are not, being only approximations to a complicated world. The practitioner should apply the models only after assessing their limitations carefully. We always need to seriously question the assumptions that make models of derivatives work: the assumptions that the market follows known probability models and the assumptions underneath the mathematical relations. What if unprecedented events do occur? Will they affect markets in ways that no mathematical model can predict? What if the regularity that all mathematical models assume ignores social and cultural variables that are not subject to mathematical analysis?

Financial events since late 2008 show that the concerns of the previous paragraphs have occurred. Complex derivatives called credit default swaps appear to have used faulty assumptions that did not account for unprecedented events including social and cultural variables that encouraged unsustainable borrowing and debt. Extremely large positions in derivatives which failed to account for unlikely events caused bankruptcy for financial firms such as Lehman Brothers and the bail out of insurance giants. The causes are complex, but critics fix some of the blame on the complex mathematical models and the people who created them. This blame results from distrust of that which is not understood. Understanding the models and their limitations is a prerequisite for creating a future with appropriate risk management.
Section Ending Answer. A few financial instruments would be bank accounts, loans, mortgages, stocks, bonds and contracts for agricultural or mineral commodities. There are many other financial instruments. Bank accounts, loans and mortgages are all available at local financial institutions and may be considered low to medium risk. Stocks and bonds are typically purchased through a brokerage firm and may be low to high risk. Contracts for commodities are purchased in special markets or exchanges or from brokers and are usually high risk.

Problems.

Exercise 1.1. Write a short summary of the “tulip mania” in seventeenth century Holland.

Exercise 1.2. Write a short summary of the “South Sea Island” bubble in eighteenth century England.

Exercise 1.3. Pick a commodity and find current futures prices for that commodity.

Exercise 1.4. Pick a stock and find current options prices on that stock.

1.2. Options and Derivatives

Section Starter Question. Suppose your rich neighbor offered an agreement to you today to sell his classic Jaguar sports-car to you (and only you) a year from today at a reasonable price agreed upon today. You and your neighbor will exchange cash and car a year from today. What would be the advantages and disadvantages to you of such an agreement? Would that agreement be valuable? How would you determine how valuable that agreement is?

Definitions. A call option is the right to buy an asset at an established price at a certain time. A put option is the right to sell an asset at an established price at a certain time. Another slightly simpler financial instrument is a future which is a contract to buy or sell an asset at an established price at a certain time.

More fully, a call option is an agreement or contract by which at a definite time in the future, known as the expiry date, the holder of the option may purchase from the option writer an asset known as the underlying asset for a definite amount known as the exercise price or strike price. A put option is an agreement or contract by which at a definite time in the future, known as the expiry date, the holder of the option may sell to the option writer an asset known as the underlying asset for a definite amount known as the exercise price or strike price. The holder of a European option may only exercise it at the end of its life on the expiry date. The holder of an American option may exercise it at any time during its life up to the expiry date. For comparison, in a futures contract the writer must buy (or sell) the asset to the holder at the agreed price at the prescribed time. The underlying assets commonly traded on options exchanges include stocks, foreign currencies, and stock indices. For futures, in addition to these kinds of assets the common assets are commodities such as minerals and agricultural products. In this text we will usually refer to options based on stocks, since stock options are easily described, commonly traded and prices are easily found.
Jarrow and Protter [25, page 7] tell a story about the origin of the names European options and American options. While writing his important 1965 article on modeling stock price movements as a geometric Brownian motion, Paul Samuelson went to Wall Street to discuss options with financial professionals. Samuelson’s Wall Street contact informed him that there were two kinds of options, one more complex that could be exercised at any time, the other more simple that could be exercised only at the maturity date. The contact said that only the more sophisticated European mind (as opposed to the American mind) could understand the former more complex option. In response, when Samuelson wrote his paper, he used these prefixes and reversed the ordering! Now in a further play on words, financial markets offer many more kinds of options with geographic labels but no relation to that place name. For example; two common types are Asian options and Bermuda options.

The Markets for Options. In the United States, some exchanges trading options are the Chicago Board Options Exchange (CBOE), the American Stock Exchange (AMEX), and the New York Stock Exchange (NYSE) among others. Exchanges are not the only place to trade options. Over-the-counter markets allow financial institutions and corporations to trade directly in options on stocks, foreign exchange rates and interest rates. In the over-the-counter markets, a financial institution can customize an option for a customer. For example, the strike price and maturity do not have to conform to exchange standards or the option can include special exercise features. A disadvantage of over-the-counter options is that the terms of the contract need not be open to inspection by others and the contract may be so different from standard derivatives that it is hard to evaluate in terms of risk and value.

A European put option allows the holder to sell the asset on a certain date for a set amount. The put option writer is obligated to buy the asset from the option holder. At exercise, if the underlying asset price is below the strike price, the holder makes a profit because the holder can buy the asset at the current low price and sell it at the set higher price instead of the current price. If the underlying asset price goes above the strike price, the holder exercises the right not to sell to the writer. The put option has payoff properties that are the opposite to those of a call. The holder of a call option wants the asset price to rise, the higher the asset price, the higher the immediate profit. The holder of a put option wants the asset price to fall as low as possible. The further below the strike price, the more valuable is the put option.

The expiry date specifies the month in which the European option ends. Exchange traded options expire on the Saturday following the third Friday of the expiration month. The last day for an option trade is that third Friday of the expiration month. Exchange traded options are typically offered with lifetimes of 1, 2, 3, and 6 months.

An important parameter of an option is the strike price, the buying or selling price of the underlying asset. For exchange traded options on stocks, the exchange typically chooses strike prices spaced $1, $2.50, $5, or $10 apart around a current value of the stock. Exchanges generally use a $2.50 spacing if the stock price is below $25, $5 spacing when it is between $25 and $200, and $10 spacing when it is above $200. For example, if Corporation XYZ has a current stock price of $12.25, options traded on it may have strike prices of $10, $12.50, $15, $17.50 and $20. A
stock trading at $144.88 has options with strike prices at $5 intervals from $110 to $185 as well as a few others near the current price with closer spacing.

**Characteristics of Options.** Options can be in the money, at the money or out of the money. An in-the-money option has positive value for the holder if it were exercised now. Similarly, an at-the-money option has zero value if exercised now, and an out-of-the-money option has negative value if it were exercised now. Symbolically, let $S$ be the stock price and $K$ the strike price. A call option is in the money if $S > K$, at the money if $S = K$ and out of the money if $S < K$. A holder will exercise an option only when it is in-the-money.

More precisely than in- or out-of-the-money, the intrinsic value of an option is the maximum of zero and the value it would have if exercised now. For a call option, the intrinsic value is $\max(S - K, 0)$ and for a put option, $\max(K - S, 0)$. Note that the intrinsic value does not consider the transaction costs or fees associated with buying or selling an asset.

The word “may” in the description of options, and the name “option” itself implies that for the holder, the contract is a right, and not an obligation. The other party of the contract, known as the writer does have a potential obligation, since the writer must sell (or buy) the asset if the holder chooses to buy (or sell) it. Since the writer confers on the holder a right with no obligation an option has some value. The holder must pay for the right at the time of opening the contract. The writer of the option must be compensated for the obligation taken on. Our main goal is to answer the following questions:

- How much should one pay for that right? That is, what is the value of an option?
- How does that value vary in time?
- How does that value depend on the underlying asset?

The value of the option contract also depends on the characteristics of the underlying asset. If the asset has relatively large variations in price, then we might believe that the option contract would be relatively high-priced since with some probability the option will be in-the-money. The option contract value is derived from the asset price, and so we call it a derivative.

Two other factors affect an option value. We need to compare owning an option with the time value of money measured by the interest rate of a risk-free bond like a
Treasury note. Finally, the underlying asset may pay dividends, affecting its value. For simplicity this text will ignore the case that an asset makes dividend payments.

Summarizing, six factors affect the price of a stock option:

- the current stock price $S$;
- the strike price $K$;
- the time to expiration $T - t$ where $T$ is the expiration time and $t$ is the current time;
- the volatility of the stock price;
- the risk-free interest rate; and
- the dividends expected during the life of the option.

Consider what happens to option prices when one of these factors changes while all the others remain fixed. Table 1 summarizes the results. The changes regarding the stock price, the strike price, the time to expiration and the volatility are easy to explain.

Upon exercising it at some time in the future, the payoff from a call option will be the amount by which the stock price exceeds the strike price. Call options therefore become more valuable as the stock price increases and less valuable as the strike price increases. For a put option, the payoff on exercise is the amount by which the strike price exceeds the stock price. Put options therefore behave in the opposite way to call options. Put options become less valuable as stock price increases and more valuable as strike price increases.

Consider next the effect of the expiration date. European put and call options do not necessarily become more valuable as the time to expiration increases. The owner of a long-life European option can only exercise at the maturity of the option.

Roughly speaking, the volatility of a stock price is a measure of how much future stock price movements may vary relative to the current price. As volatility increases, the chance that the stock price will either increase or decrease greatly relative to the present price also increases. For the owner of a stock, these two outcomes tend to offset each other. However, this is not so for the owner of a put or call option. The owner of a call benefits from price increases, but has limited downside risk in the event of price decrease since the most that he or she can lose is the price of the option. Similarly, the owner of a put benefits from price decreases but has limited upside risk in the event of price increases. The values of puts and calls therefore increase as volatility increases.

The reader will observe that the language about option prices in this section has been qualitative and imprecise.
• an option is “a contract to buy or sell an asset at an established price” without specifying how the price is obtained;
• “...the option contract would be relatively high-priced...”;
• “Call options therefore become more valuable as the stock price increases ...” without specifying the rate of change; and
• “As volatility increases, the chance that the stock price will either increase or decrease greatly ... increases”.

The goal in following sections is to develop a mathematical model which gives quantitative and precise statements about options prices and to judge the validity and reliability of the model.

Section Ending Answer. Your rich neighbor is essentially offering you a futures contract on his car. The advantage would be that if the car value appreciates over a year, you have the opportunity to buy a valuable car at a favorable price. If the car depreciates over a year, perhaps it develops mechanical problems or gets in an accident, it is a disadvantage since you will be obligated to pay more than the car is worth at that time. This agreement is singular and depends on factors such as the driving and maintenance habits of your neighbor, so it is difficult to value.

Problems.

Exercise 1.5. (1) Find and write the definition of a future, also called a futures contract. Graph the intrinsic value of a futures contract at its contract date, or expiration date, as in Figure 1.

(2) Explain why holding a call option and writing a put option with the same strike price \( K \) on the same asset is the same as having a futures contract on the asset with strike price \( K \). Draw a graph of the intrinsic value of the option combination and the value of the futures contract on the same axes.

Exercise 1.6. Puts and calls are not the only option contracts available, just the most fundamental and the simplest. Puts and calls eliminate the risk of up or down price movements in the underlying asset. Some other option contracts designed to eliminate other risks are combinations of puts and calls.

(1) Draw the graph of the intrinsic value of the option contract composed of holding a put option with strike price \( K_1 \) and holding a call option with strike price \( K_2 \) where \( K_1 < K_2 \). (Assume both the put and the call have the same expiration date.) The holder profits only if the underlier moves dramatically in either direction. This is known as a long strangle.

(2) Draw the graph of the intrinsic value of an option contract composed of holding a put option with strike price \( K \) and holding a call option with the same strike price \( K \). (Assume both the put and the call have the same expiration date.) This is called an long straddle, and also called a bull straddle.

(3) Draw the graph of the intrinsic value of an option contract composed of holding one call option with strike price \( K_1 \) and the simultaneous writing of a call option with strike price \( K_2 \) with \( K_1 < K_2 \). (Assume both the options have the same expiration date.) This is known as a bull call spread.
(4) Draw the graph of the intrinsic value of an option contract created by simultaneously holding one call option with strike price \( K_1 \), holding another call option with strike price \( K_2 \) where \( K_1 < K_2 \), and writing two call options at strike price \( (K_1 + K_2)/2 \). This is known as a **butterfly spread**.

(5) Draw the graph of the intrinsic value of an option contract created by holding one put option with strike price \( K \) and holding two call options on the same underlying security, strike price, and maturity date. This is known as a **triple option** or strap.

### 1.3. Speculation and Hedging

**Section Starter Question.** Discuss examples of speculation in your experience. (Example: think of scalping tickets.) A hedge is a transaction or investment that is taken out specifically to reduce or cancel out risk. Discuss examples of hedges in your experience.

**Definitions.** Two primary uses of options are **speculation** and **hedging**. **Speculation** is to assume a financial risk in anticipation of a gain, especially to buy or sell to profit from market fluctuations. The market fluctuations are random financial variations with a known (or assumed) probability distribution.

**Risk and Uncertainty.** Risk, first articulated by the economist F. Knight in 1921, is a variability that you know in advance. That is, risk is random financial variation that has a known (or assumed) probability distribution. Suppose you place a $1 bet on red in American roulette; that is, you bet that the ball will fall in a numbered bin colored red on the edge of the wheel. You will lose $1 if the ball lands in a black or green bin, you will get your bet back and win an additional $1 if the ball lands on red. Your finances will vary but the probability distribution of outcomes is well understood in advance.

**Uncertainty** is chance variability due to unknown and unmeasured factors. You might have some awareness (or not) of the variability out there. You may have no idea of how many such factors exist, or when any one may strike, or how big the effects will be. Uncertainty is *unknown unknowns*.

Risk sparks a free-market economy with the impulse to make a gain. Uncertainty halts an economy with fear.

**Example: Speculation on a stock with calls.** An investor who believes that a particular stock, say XYZ, is going to rise may purchase some shares in the company. If she is correct, she makes money, if she is wrong she loses money. The investor is *speculating*. Suppose the price of the stock goes from $2.50 to $2.70, then the investor makes $0.20 on each $2.50 investment, or a gain of 8%. If the price falls to $2.30, then the investor loses $0.20 on each $2.50 share, for a loss of 8%. These are standard calculations.

Alternatively, suppose the investor thinks that the share price is going to rise within the next couple of months, and that the investor buys a call option with exercise price of $2.50 and expiry date in three months.

Now assume that it costs $0.10 to purchase a European call option on stock XYZ with expiration date in three months and strike price $2.50. That means in three months time, the investor could, if the investor chooses to, purchase a share of XYZ at price $2.50 per share *no matter what the current price of XYZ stock*
Note that the price of $0.10 for this option may not be an proper price for the option, but we use $0.10 simply because it is easy to calculate with. However, 3-month option prices are often about 5% of the stock price, so $0.10 is reasonable. In three months time if the XYZ stock price is $2.70, then the holder of the option may purchase the stock for $2.50. This action is called exercising the option. It yields an immediate profit of $0.20. That is, the option holder can buy the share for $2.50 and immediately sell it in the market for $2.70. On the other hand if in three months time, the XYZ share price is only $2.30, then it would not be sensible to exercise the option. The holder lets the option expire. Now observe carefully:

By purchasing an option for $0.10, the holder can derive a net profit of $0.10 ($0.20 revenue less $0.10 cost) or a loss of $0.10 (no revenue less $0.10 cost). The profit or loss is magnified to 100%. Investors usually buy options in quantities of hundreds, thousands, even tens of thousands so the absolute dollar amounts can be large. Compared with stocks, options offer a great deal of leverage, that is, large relative changes in value for the same investment. Options expose a portfolio to a large amount of risk cheaply. Sometimes a large degree of risk is desirable. This is the use of options and derivatives for speculation.

Example: Speculation on a stock with calls. Consider the profit and loss of a investor who buys 100 call options on XYZ stock with a strike price of $140. Suppose the current stock price is $138, the expiration date of the option is two months, and the option price is $5. Since the options are European, the investor can exercise only on the expiration date. If the stock price on this date is less than $140, the investor will choose not to exercise the option since buying a stock at $140 that has a market value less than $140 is not sensible. In these circumstances the investor loses the whole of the initial investment of $500. If the stock price is above $140 on the expiration date, the holder will exercise the options. Suppose for example, the stock price is $155. By exercising the options, the investor is able to buy 100 shares for $140 per share. By selling the shares immediately, the investor makes a gain of $15 per share, or $1500 ignoring transaction costs. Taking the initial cost of the option into account, the net profit to the investor is $10 per option, or $1000 on an initial investment of $500. Note that this calculation ignores any time value of money.

Example: Speculation on a stock with puts. Consider an investor who buys 100 European put options on XYZ with a strike price of $90. Suppose the current stock price is $86, the expiration date of the option is in 3 months and the option price is $7. Since the options are European, the holder will exercise only if the stock price is below $90 at the expiration date. Suppose the stock price is $65 on this date. The investor can buy 100 shares for $65 per share, and under the terms of the put option, sell the same stock for $90 to realize a gain of $25 per share, or $2500. Again, this simple example ignores transaction costs. Taking the initial cost of the option into account, the investor’s net profit is $18 per option, or $1800. This is a profit of 257% even though the stock has only changed price $25 from an initial of $90, or 28%. Of course, if the final price is above $90, the put option expires worthless, and the investor loses $7 per option, or $700.

Example: Hedging with a portfolio with calls. Since the value of a call option rises when an asset price rises, what happens to the value of a portfolio containing both shares of stock of XYZ and a negative position in call options on
XYZ stock? If the stock price is rising, the call option value will also rise, the negative position in calls will become greater, and the net portfolio should remain approximately constant if the positions are in the right ratio. If the stock price is falling then the call option value price is also falling. The negative position in calls will become smaller. If held in the proper amounts, the total value of the portfolio should remain constant! The risk (or more precisely, the variation) in the portfolio is reduced! The reduction of risk by taking advantage of such correlations is called hedging. Used carefully, options are an indispensable tool of risk management.

Consider a stock currently selling at $100 and having a standard deviation in its price fluctuations of $10, for a proportion variation of 10%. With some additional information (the risk-free interest rate), the Black-Scholes formula derived later shows that a call option with a strike price of $100 and a time to expiration of one year would sell for $11.84. A 1 percent rise in the stock from $100 to $101 would drive the option price to $12.73. Consider the total effects in Table 2.

Suppose a trader has an original portfolio comprised of 8,944 shares of stock selling at $100 per share. The unusual number of 8,944 shares comes from the Black-Scholes formula as a hedge ratio. Assume also that a trader short sells call options on 10,000 shares at the current price of $11.84. That is, the short seller borrows the options from another trader and therefore must later return the options at the option price at the return time. The obligation to return the borrowed options creates a negative position in the option value. The transaction is called short selling because the trader sells a good he or she does not actually own and must later pay it back. In Table 2 this debt or short position in the option is indicated by a minus sign. The entire portfolio of shares and options has a net value of $776,000.

Now consider the effect of a 1 percent change in the price of the stock. If the stock increases 1 percent, the shares will be worth $903,344. The option price will increase from $11.84 to $12.73. But since the portfolio also involves a short position in 10,000 options, this creates a loss of $8,900. This is the additional value of what the borrowed options are now worth, so the borrower must additionally pay this amount back! Taking these two effects into account, the value of the portfolio will be $776,044. This is nearly the same as the original value. The slight discrepancy of $44 is rounding error due to the fact that the number of stock shares calculated from the hedge ratio is rounded to an integer number of shares for simplicity in the example, and the change in option value is rounded to the nearest penny, also for simplicity. In actual practice, financial institutions take great care to avoid round-off differences.

On the other hand of the stock price falls by 1 percent, there will be a loss in the stock of $8,944. The price on this option will fall from $11.84 to $10.95 and this means that the entire drop in the price of the 10,000 options will be $8900. Taking both of these effects into account, the portfolio will then be worth $776,956. The overall value of the portfolio will not change (to within $44 due to round-off effects) regardless of what happens to the stock price. If the stock price increases, there is an offsetting loss on the option; if the stock price falls, there is an offsetting gain on the option.

This example is not intended to illustrate a prudent investment strategy. If an investor desired to maintain a constant amount of money, putting the sum of money invested in shares into the bank or in Treasury bills instead would safeguard the sum and even pay a modest amount of interest. If the investor wished to maximize
the investment, then investing in stocks solely and enduring a probable 10% loss in value would still leave a larger total investment.

This example is a first example of short selling. It is also an illustration of how holding an asset and short selling a related asset in carefully calibrated ratios can hold a total investment constant. The technique of holding and short-selling to keep a portfolio constant will later be an important idea in deriving the Black-Scholes formula.

**Section Ending Answer.** Common experience examples of speculation are buying tickets to a concert or sporting event with the expectation of selling them at a higher price. The risk is that the event may not be as popular as expected and the seller cannot sell or only sell below cost. Another example is “flipping houses”, buying low-cost property, fixing it up and selling for a profit. The most common form of hedging is buying insurance, paying a relatively small amount to guard against the possibility of a large or even catastrophic expense.

**Problems.**

**Exercise 1.7.** You would like to speculate on a rise in the price of a certain stock. The current stock price is $50 and a 3-month call with strike of $52 costs $2.50. You have $5,000 to invest. Consider two alternate strategies, one involving investment exclusively in the stock, and the other involving investment exclusively in the option. What are the potential gains or losses from each due to a rise to $51 in three months? What are the potential gains or losses from each due to a fall to $48 in three months?

**Exercise 1.8.** The current price of a stock is $94 and 3-month call options with a strike price of $95 currently sell for $4.70. An investor who feels that the price of the stock will increase is trying to decide between buying 100 shares and buying 2,000 call options. Both strategies involve an investment of $9,400. Write and solve an inequality to find how high the stock price must rise for the option strategy to be the more profitable.
1.4. Arbitrage

**Section Starter Question.** It’s the day of the big game. You know that your rich neighbor really wants to buy tickets, in fact you know he’s willing to pay $50 a ticket. While on campus, you see a hand lettered sign offering “Two general-admission tickets at $25 each, inquire immediately at the mathematics department.” You have your phone with you, what should you do? Discuss whether this is a frequent event, and why or why not? Is this market efficient? Is there any risk in this market?

**Definition of Arbitrage.** The notion of arbitrage is crucial in the modern theory of finance. It is a cornerstone of the Black, Scholes and Merton option pricing theory, developed in 1973, for which Scholes and Merton received the Nobel Prize in 1997 (Fisher Black died in 1995).

An arbitrage opportunity is a circumstance where the simultaneous purchase and sale of related securities is guaranteed to produce a riskless profit. Arbitrage opportunities should be rare, but in a world-wide basis some do occur.

The best way to understand arbitrage is to consider simple examples. The following examples use some realistic data mixed with some values highlighting the arbitrage opportunity. The examples use markets in the U.S. and Europe; separated by time zones and geography but still connected in the global economy, the two markets might occasionally offer an opportunity to find riskless profits.

**An arbitrage opportunity in exchange rates.** Consider a stock that is traded on both the New York Stock Exchange and the Frankfurt (Germany) Stock Exchange. Suppose that the stock price is $145 in New York and 125 Euros in Frankfurt at a time when the exchange rate is 0.8757 dollar per Euro. An arbitrageur in New York could simultaneously buy 100 shares of the stock in Frankfurt and sell them in New York to obtain a risk-free profit of

\[
-100 \text{ shares} \times \frac{125 \text{ Euros}}{\text{share}} \div 0.8757 \frac{\text{S}}{\text{Euros}} + 100 \text{ shares} \times 145 \frac{\$}{\text{share}} = 225.71
\]

in the absence of transaction costs. Costs associated with buying and selling stocks and exchanging currency would reduce or eliminate the profit on a small transaction like this. However, large investment houses face low transaction costs in both the stock market and the foreign exchange market. Trading firms would find this arbitrage opportunity very attractive and would try to take advantage of it in quantities of many thousands of shares.

The shares in Frankfurt are underpriced relative to the shares in New York with the exchange rate taken into consideration. However, note that the demand for the purchase of many shares in Frankfurt would soon drive the price up. The sale of many shares in New York would soon drive the price down. The market would soon reach a point where the arbitrage opportunity disappears.

**An arbitrage opportunity in exchange rates.** In October 2007, the exchange rate between the U.S. Dollar and the Euro was 1.4280, that is, it cost $1.4280 to buy one Euro. At that time, the 1-year Fed Funds rate, (the bank-to-bank lending rate), in the United States was 4.7500% (assume it is compounded continuously). The forward rate (the exchange rate in a forward contract that allows you to buy Euros in a year) for purchasing Euros in 1 year was 1.4312. Suppose
1.4. ARBITRAGE

that a bank in Europe offered a 5% interest rate on Euros (assume it too is compounded continuously). This set of economic circumstances creates an arbitrage opportunity for a large bank.

One dollar invested in October 2007 in the U.S. will be worth \(1 \cdot e^{0.0475} = 1.048646201\) in one year. One dollar in October 2007 will buy \(1/(1.4280) = 0.70028\) Euros. Those Euros will be worth \(0.70028 \cdot e^{0.05}\) in one year. With the forward rate, those Euros would be worth \(1.4312 \cdot 0.70028 \cdot e^{0.05} = 1.05363\) dollars in a year. So an arbitrageur could borrow $1,000,000 from a bank in the U.S. and buy 700,280 Euros and put them in the European bank. Simultaneously, the arbitrageur would create a contract to sell Euros in an year at the forward rate. After a year, converting the investment back to dollars, the arbitrageur would have $1,053,630 but has to pay back the loan with $1,048,646. This gives a risk-free profit of $4984.

This example ignores transaction costs and assumes interests are paid at the end of the lending period.

**Discussion about arbitrage.** Arbitrage opportunities as just described cannot last for long. In the first example, as arbitrageurs sell the stock in New York, the forces of supply and demand will cause the New York dollar price to fall. Similarly as the arbitrageurs buy the stock in Frankfurt, they drive up the Frankfurt price. The two stock prices will quickly become equal at the current exchange rate. Indeed the existence of profit-hungry arbitrageurs (usually pictured as frenzied traders carrying on several conversations at once!) makes it unlikely that a major disparity between the Euro price and the dollar price could ever exist in the first place. In the second example, once arbitrageurs start to deposit money in Europe, the interest rate will drop. The demand for the loans in the U.S. will cause the interest rates to rise. Although arbitrage opportunities can arise in financial markets, they cannot last long.

Generalizing, the existence of arbitrageurs means that in practice, only tiny arbitrage opportunities are observed only for short times in most financial markets. As soon as sufficiently many observant investors find the arbitrage, the prices quickly change as the investors buy and sell to take advantage of such an opportunity. As a consequence, the arbitrage opportunity disappears. The principle can stated as follows: in an efficient market there are no arbitrage opportunities. In this text many arguments depend on the assumption that arbitrage opportunities do not exist, or equivalently, that we are operating in an efficient market.

A joke illustrates this principle well: A mathematical economist and a financial analyst are walking down the street together. Suddenly each spots a $100 bill lying in the street at the curb! The financial analyst yells “Wow, a $100 bill, grab it quick!” The mathematical economist says “Don’t bother, if it were a real $100 bill, somebody would have picked it up already.” Arbitrage opportunities are like $100 bills on the ground, they do exist in real life, but one has to be quick and observant. For purposes of mathematical modeling, we can treat arbitrage opportunities as non-existent as $100 bills lying in the street. It might happen, but we don’t base our financial models on the expectation of finding them.

The principle of arbitrage pricing is that any two investments with identical payout streams must have the same price. If this were not so, we could simultaneously sell the more expensive instrument and buy the cheaper one: the payment from our sale exceeds the payment for our purchase. We can make an immediate profit.
Before the 1970s most economists approached the valuation of a security by considering the probability of the stock going up or down. Economists now find the price of a security by arbitrage without the consideration of probabilities. We will use the principle of arbitrage pricing extensively in this text.

Section Ending Answer. You could purchase the two tickets at $25 each, call your rich neighbor, and offer to sell him the two tickets at $50 each, making a quick profit. This is probably an infrequent event, lucky for you to be in the right place at the right time. There is a risk that you will not reach your rich neighbor to sell the tickets. This is not an efficient market, since not everyone has all information about all available tickets and prices.

Problems.

Exercise 1.9. Consider the hypothetical country of Mathtopia, where the government has declared a policy requiring the exchange rate between the domestic currency, the Math Buck, denoted by MB, and the U.S. Dollar to stay in a prescribed range, namely:

$$0.90\text{USD} \leq MB \leq 1.10\text{USD}$$

for at least one year. Suppose also that the Mathtopian government has issued 1-year notes denominated in the MB that pay a continuously compounded interest rate of 28%. Assuming that the corresponding continuously compounded interest rate for US deposits is 6%, show that an arbitrage opportunity exists.

Exercise 1.10. (1) At a certain time, the exchange rate between the U.S. Dollar and the Euro was 1.4280, that is, it cost $1.4280 to buy one Euro. At that time, the 1-year Fed Funds rate, (the bank-to-bank lending rate), in the United States was 4.7500% (assume it is compounded continuously). The forward rate (the exchange rate in a forward contract that allows you to buy Euros in a year) for purchasing Euros 1 year from today was 1.4312. What was the corresponding bank-to-bank lending rate in Europe (assume it is compounded continuously), and what principle allows you to claim that value?

(2) Find the current exchange rate between the U.S. Dollar and the Euro, the current 1-year Fed Funds rate, and the current forward rate for exchange of Euros to Dollars. Use those values to compute the bank-to-bank lending rate in Europe.

Exercise 1.11. According to the article “Bullion bulls” on page 81 in the October 8, 2009 issue of The Economist, gold rose from about $510 per ounce in January 2006 to about $1050 per ounce in October 2009, 46 months later.

(1) What was the continuously compounded annual rate of increase of the price of gold over this period?

(2) In October 2009, one could borrow or lend money at 5% interest, again assume it was compounded continuously. In view of this, describe a strategy that would have made a profit in October 2010, involving borrowing or lending money, assuming that the rate of increase in the price of gold stayed constant over this time.
(3) The article suggests that the rate of increase for gold would stay constant. In view of this, what do you expect to happen to interest rates and what principle allows you to conclude that?

Exercise 1.12. Consider a market that has a security and a bond so that money can be borrowed or loaned at an annual interest rate of $r$ compounded continuously. At the end of a time period $T$, the security will have increased in value by a factor $U$ to $SU$, or decreased in value by a factor $D$ to value $SD$. Show that a forward contract with strike price $k$ that, is, a contract to buy the security which has potential payoffs $SU - k$ and $SD - k$ should have the strike price set at $S \exp(rT)$ to avoid an arbitrage opportunity.

1.5. Mathematical Modeling

Section Starter Question. Do you believe in the ideal gas law? Does it make sense to believe in an equation?

Mathematical Modeling. Remember the following proverb: All mathematical models are wrong, but some mathematical models are useful. [9]

Mathematical modeling involves two equally important activities:

- building a mathematical structure, a model, based on hypotheses about relations among the quantities that describe the real-world situation, and then deriving new relations;
- evaluating the model, comparing the new relations with the real world and making predictions from the model.

Good mathematical modeling explains the hypotheses, the development of the model and its solutions, and then supports the findings by comparing them mathematically with the actual circumstances. A successful model must allow a user to consider the effects of different hypotheses.

Successful modeling requires a balance between so much complexity that making predictions from the model may be intractable and so little complexity that the predictions are unrealistic and useless. Complex models often give more precise, but not necessarily more accurate, answers and so can fool the modeler into believing that the model is better at prediction that it actually is. On the other hand, simple models may be useful for understanding, but are probably too blunt to make useful predictions. Nate Silver quotes economist Arnold Zellner advising to “Keep it sophisticatedly simple.” [61, page 225]

At a more detailed level, mathematical modeling involves 4 successive phases in the cycle of modeling:

1. A real-world situation,
2. a mathematical model,
3. a new relation among the quantities,
4. predictions and verifications.

Consider the diagram in Figure 2 which illustrates the cycle of modeling. Connecting phase 1 to 2 in the more detailed cycle builds the mathematical structure and connecting phase 3 to phase 4 evaluates the model.

Modeling: Connecting Phase 1 to Phase 2. A good description of the model will begin with an organized and complete description of important factors and observations. The description will often use data gathered from observations of the problem. It will also include the statement of scientific laws and relations
that apply to the important factors. From there, the model must summarize and condense the observations into a small set of hypotheses that capture the essence of the observations. The small set of hypotheses is a restatement of the problem, changing the problem from a descriptive, even colloquial, question into a precise formulation that moves the question from the general to the specific. Here the modeler demonstrates a clear link between the listed assumptions and the building of the model.

The hypotheses translate into a mathematical structure that becomes the heart of the mathematical model. Many mathematical models, particularly those from physics and engineering, become a single equation but mathematical models need not be a single concise equation. Mathematical models may be a regression relation, either a linear regression, an exponential regression or a polynomial regression. The choice of regression model should explicitly follow from the hypotheses since the growth rate is an important consequence of the observations. The mathematical model may be a linear or nonlinear optimization model, consisting of an objective
function and a set of constraints. Again the choice of linear or nonlinear functions for the objective and constraints should explicitly follow from the nature of the factors and the observations. For dynamic situations, the observations often involve some quantity and its rates of change. The hypotheses express some connection between these quantities and the mathematical model then becomes a differential equation, either linear or nonlinear depending on the explicit details of the scientific laws relating the factors considered. For discretely sampled data instead of continuous time expressions the model may become a difference equation. If an important feature of the observations and factors is noise or randomness, then the model may be a probability distribution or a stochastic process. The classical models from science and engineering usually take one of these equation-like forms but not all mathematical models need to follow this format. Models may be a connectivity graph, or a group of transformations.

If the number of variables is more than a few, or the relations are too complicated to write in a concise mathematical expression then the model can be a computer program. Programs written in either high-level languages such as C, FORTRAN or Basic and very-high-level languages such as R, Octave, Scientific Python, or a computer algebra system are mathematical models. Spreadsheets combining the data and the calculations are a popular and efficient way to construct a mathematical model. The collection of calculations in the spreadsheet expresses the laws connecting the factors that the data in the rows and columns of the spreadsheet represent. Some mathematical models use more elaborate software specifically designed for modeling. Some software allows the user to graphically describe the connections among factors to create and alter a model.

Although this set of examples of mathematical models varies in theoretical sophistication and the equipment used, the core of each is to connect the data and the relations into a mechanism that allows the user to vary elements of the model. Creating a model, whether a single equation, a complicated mathematical structure, a quick spreadsheet, or a large program is the essence of the first step connecting the phases labeled 1 and 2 in Figure 2.

First models need not be sophisticated or detailed. For beginning analysis back of the envelope calculations and dimensional analysis will be as effective as spending time setting up an elaborate model or solving equations with advanced mathematics. Unit analysis to check consistency and outcomes of relations is important to check the harmony of the modeling assumptions. A good model pays attention to units, the quantities should be sensible and match. Even more important, a non-dimensionalized model reveals significant relationships, and major influences. Unit analysis is an important part of modeling, and goes far beyond simple checking to make sure units cancel. [37, 39]

Mathematical Solution: Connecting Phase 2 to Phase 3. Once the modelers create the model, then they should derive some new relations among the important quantities selected to describe the real-world situation. This is the step connecting the phases labeled 2 and 3 in the diagram. If the model is an equation, for instance the Ideal Gas Law, then one can solve the equation for one of the variables in terms of the others. In the Ideal Gas Law, solving for one of the gas parameters is easy. A regression model may need no additional mathematical solution, although it might be useful to find auxiliary quantities such as rates of growth or maxima or minima. For an optimization problem the solution is the set of optima or the rates of change.
of optima with respect to the constraints. If the model is a differential equation or a difference equation, then the solution may have some mathematical substance. For instance, for a ballistics problem, the model may be a differential equation and the solution by calculus methods yields the equation of motion. For a problem with randomness, the derivation may find the mean or the variance. For a connectivity graph, some interesting quantities are the number of cycles, components or the diameter of the graph. If the model is a computer program, then this step usually involves running the program to obtain the output.

It is easy for students to focus most attention on the solution stage of the process, since the methods are the core of the typical mathematical curriculum. This step usually requires no interpretation, and the model dictates the methods to use. This step is often the easiest in the sense that it is the clearest on how to proceed, although the mathematical procedures may be daunting.

Testing and Sensitivity: Connecting Phase 3 to Phase 4. Once this step is done, the model is ready for testing and sensitivity analysis. This is the step that connects the phases labeled 3 and 4. At the least, the modelers should try to verify, even with common sense, the results of the solution. Typically for a mathematical model, the previous step allows the modelers to produce some important or valuable quantity of the model. Modelers compare the results of the model with standard or common inputs with known quantities for the data or statement of the problem. This may be as easy as substituting into the derived equation, regression expression, or equation of motion. When running a computer model or program, this may involve sequences of program runs and related analysis. With any model, the results will probably not be exactly the same as the known data so interpretation or error analysis will be necessary. The interpretation requires judgment on the relative magnitudes of the quantities produced in light of the confidence in the exactness or applicability of the hypotheses.

Another important activity at this stage in the modeling process is the sensitivity analysis. The modelers should choose some critical feature of the model and then vary the parameter value that quantifies that feature. The results should be carefully compared to the real world and to the predicted values. If the results do not vary substantially, then perhaps the feature or parameter is not as critical as believed. This is important new information for the model. On the other hand, if a predicted or modeled value varies substantially in comparison to the parameter as it is slightly varied, then the accuracy of measurement of the critical parameter assumes new importance. In sensitivity analysis, just as in all modeling, we measure varying substantially with significant digits, relative magnitudes, and rates of change. Here is another area where expressing parameters in dimensionless groups is important [39]. In some areas of applied mathematics such as linear optimization and statistics, a side effect of the solution method is that it automatically produces sensitivity parameters. In linear optimization, these are sometimes called the shadow prices and these additional solution values should be used whenever possible.

Interpretation and Refinement: Connecting Phase 4 to Phase 1. Finally the modelers must take the results from the previous steps and use them to refine the interpretation and understanding of the real-world situation. In the diagram the interpretation step connects the phases labeled 4 and 1, completing the cycle of modeling. For example, if the situation is modeling motion, then examining results...
may show that the predicted motion is faster than measured, or that the object
does not travel as far as the model predicts. Then it may be that the model does
not include the effects of friction, and so friction should be incorporated into a
new model. At this step, the modeler has to be open and honest in assessing the
strengths and weaknesses of the model. It also requires an improved understanding
of the real-world situation to include the correct new elements and hypotheses to
correct the discrepancies in the results. A good model can be useful even when it
fails. When a model fails to match the reality of the situation then the modeler
must understand how it is wrong, and what to do when it is wrong, and minimizing
the cost when it is wrong.

The step between stages 4 and 1 may suggest new processes, or experimental
conditions to alter the model. If the problem suggests changes then the modeler
should implement those changes should and test them in another cycle in the mod-
eling process.

Summary. A good summary of the modeling process is that it is an intense
and structured application of critical thinking. Sophistication of mathematical
techniques is not always necessary, the mathematics connecting steps 2 and 3 or
potentially steps 3 and 4 may only be arithmetic. The key to good modeling is the
critical thinking that occurs between steps 1 and 2, steps 3 and 4, and 4 and 1. If
a model does not fit into this paradigm, it probably does not meet the criteria for
a good model.

Good mathematical modeling, like good critical thinking, does not arise au-
tomatically or naturally. Scientists, engineers, mathematicians and students must
practice and develop the craft of creating, solving, using, and interpreting a mathe-
atical model. The structured approach to modeling helps distinguish the distinct
steps, each requiring separate intellectual skills. It also provides a framework for
developing and explaining a mathematical model.

A classification of mathematical models. The goal of mathematical mod-
eling is to represent a real-world situation in a way that is a close approximation,
even ideally exact, with a mathematical structure that is still solvable. This repre-
sents the interplay between the mathematics occurring between steps 1 and 2 and
steps 2 and 3. Then we can classify models based on this interplay:

- exact models with exact solutions;
- exact models with approximate solutions;
- approximate models with exact solutions;
- approximate models with approximate solutions.

Exact models are uncommon because only rarely do we know physical represen-
tations well enough to claim to represent situations exactly. Exact models typically
occur in discrete situations where we can enumerate all cases and represent them
mathematically.

Example 1.13. The average value of the 1-cent, 5-cent, 10-cent and 25-cent
coins in Paula’s purse is 20 cents (the average value is the total money value divided
by the number of coins.) If she had one more 25-cent coin, the average value would
be 21 cents. What coins does Paula have in her purse?
The mathematical model is to let \( V \) be the total value of the coins and \( c \) be the number of coins. Write two equations in two unknowns

\[
\frac{V}{c} = 20 \\
\frac{(V + 25)}{(c + 1)} = 21.
\]

The solution is 3 25-cent coins and 1 5-cent coin. Because of the discrete nature and fixed value of the coins, and the simplicity of the equations, an exact model has an exact solution.

Approximate models are much more typical.

**Example 1.14.** Modeling the motion of a pendulum is a classic problem in applied mathematics. Galileo first modeled the pendulum mathematically and mathematicians and physicists still investigate more sophisticated models. Most books on applied mathematics and differential equations consider some form of the pendulum model, for example [37]. An overview of modeling the pendulum appears in [56].

As a physically inclusive model of the pendulum, consider the general elastic pendulum shown in Figure 3. The pendulum is a heavy bob \( B \) attached to a pivot point \( P \) by an elastic spring. The pendulum has at least 5 degrees of freedom about its rest position. The pendulum can swing in the \( x-y-z \) spatial dimensions. The pendulum can rotate torsionally around the axis of the spring \( PB \). Finally, the pendulum can lengthen and shorten along the axis \( PB \). Although we usually think of a pendulum as the short rigid rod in a clock, some pendula move in all these degrees of freedom, for example weights dangling from a huge construction crane.

We usually simplify the modeling with assumptions:

1. The pendulum is a rigid rod so that it does not lengthen or shorten, and it does not rotate torsionally.
2. All the mass is concentrated in the bob \( B \), that is, the pendulum rod is massless.
3. The motion occurs only in one plane, so the pendulum has only one degree of freedom, measured by the angle \( \theta \).
4. There is no air resistance or friction in the bearing at \( P \).
5. The only forces on the pendulum are at the contact point \( P \) and the force of gravity due to the earth.
6. The pendulum is small relative to the earth so that gravity acts in the down direction only and is constant.

Then with standard physical modeling with Newton’s laws of motion, we obtain the differential equation for the pendulum angle \( \theta \) with respect to vertical,

\[
ml\theta'' + mg\sin \theta = 0, \quad \theta(0) = \theta_0, \theta'(0) = v_0.
\]

Note the application of the math modeling process moving from phase 1 to phase 2 with the identification of important variables, the hypotheses that simplify, and the application of physical laws. It is clear that already the differential equation describing the pendulum is an approximate model.

Nevertheless, this nonlinear differential equation cannot be solved in terms of elementary functions. It can be solved in terms of a class of higher transcendental functions known as elliptic integrals, but in practical terms that is equivalent to not being able to solve the equation analytically. There are two alternatives, either
Figure 3. Schematic diagram of a pendulum.

to solve this equation numerically or to reduce the equation further to a simpler form.

For small angles $\sin \theta \approx \theta$, so if the pendulum does not have large oscillations the model becomes

$$ml\theta'' + mg\theta = 0, \quad \theta(0) = \theta_0, \theta'(0) = v_0.$$  

This differential equation is analytically solvable with standard methods yielding

$$\theta(t) = A \cos(\omega t + \phi)$$

for appropriate constants $A$, $\omega$, and $\phi$ derived from the constants of the model. This is definitely an exact solution to an approximate model. We use the approximation of $\sin \theta$ by $\theta$ specifically so that the differential equation is explicitly solvable in elementary functions. The solution to the differential equation is the process that leads from step 2 to step 3 in the modeling process. The analysis necessary to measure the accuracy of the exact solution compared to the physical motion is
difficult. The measurement of the accuracy would be the process of moving from step 2 to step 3 in the cycle of modeling.

For a clock maker the accuracy of the solution is important. A day has 86,400 seconds, so an error on the order of even 1 part in 10,000 for a physical pendulum oscillating once per second accumulates unacceptably. This error concerned physicists such as C. Huygens who created advanced pendula with a period independent of the amplitude. This illustrates step 3 to step 4 and the next modeling cycle.

Models of oscillation, including the pendulum, have been a central paradigm in applied mathematics for centuries. Here we use the pendulum model to illustrate the modeling process leading to an approximate model with an exact solution.

**An example from physical chemistry.** This section illustrates the cycle of mathematical modeling with a simple example from physical chemistry. This simple example provides a useful analogy to the role of mathematical modeling in mathematical finance. The historical order of discovery is slightly modified to illustrate the idealized modeling cycle. Scientific progress rarely proceeds in simple order.

Scientists observed that diverse gases such as air, water vapor, hydrogen, and carbon dioxide all behave predictably and similarly. After many observations, scientists derived empirical relations such as Boyle’s law, and the law of Charles and Gay-Lussac about the gas. These laws express relations among the volume $V$, the pressure $P$, the amount $n$, and the temperature $T$ of the gas.

In classical theoretical physics, we can define an *ideal gas* by making the following assumptions [19]:

1. A gas consists of particles called molecules which have mass, but essentially have no volume, so the molecules occupy a negligibly small fraction of the volume occupied by the gas.
2. The molecules can move in any direction with any speed.
3. The number of molecules is large.
4. No appreciable forces act on the molecules except during a collision.
5. The collisions between molecules and the container are elastic, and of negligible duration so both kinetic energy and momentum are conserved.
6. All molecules in the gas are identical.

From this limited set of assumptions about theoretical entities called molecules physicists derive the equation of state for an ideal gas using the 4 quantifiable elements of volume, pressure, amount, and temperature. The *equation of state* or *ideal gas law* is

$$PV = nRT,$$

where $R$ is a measured constant, called the universal gas constant. The ideal gas law is a simple algebraic equation relating the 4 quantifiable elements describing a gas. The equation of state or ideal gas law predicts very well the properties of gases under the wide range of pressures, temperatures, masses and volumes commonly experienced in everyday life. The ideal gas law predicts with accuracy necessary for safety engineering the pressure and temperature in car tires and commercial gas cylinders. This level of prediction works even for gases we know do not satisfy the assumptions, such as air, which chemistry tells us is composed of several kinds of molecules which have volume and do not experience completely elastic collisions because of intermolecular forces. We know the mathematical model is wrong, but it is still useful.
Nevertheless, scientists soon discovered that the assumptions of an ideal gas predict that the difference in the constant-volume specific heat and the constant-pressure specific heat of gases should be the same for all gases, a prediction that scientists observe to be false. The simple ideal gas theory works well for monatomic gases, such as helium, but does not predict so well for more complex gases. This scientific observation now requires additional assumptions, specifically about the shape of the molecules in the gas. The derivation of the relationship for the observables in a gas is now more complex, requiring more mathematical techniques.

Moreover, under extreme conditions of low temperatures or high pressures, scientists observe new behaviors of gases. The gases condense into liquids, pressure on the walls drops and the gases no longer behave according to the relationship predicted by the ideal gas law. We cannot neglect these deviations from ideal behavior in accurate scientific or engineering work. We now have to admit that under these extreme circumstances we can no longer ignore the size of the molecules, which do occupy some appreciable volume. We also must admit that we cannot ignore intermolecular forces. The two effects just described can be incorporated into a modified equation of state proposed by J.D. van der Waals in 1873. Van der Waals’ equation of state is:

\[
(P + \frac{n^2 a}{V^2})(V - b) = nRT.
\]

The added constants \(a\) and \(b\) represent the new elements of intermolecular attraction and volume effects respectively. If \(a\) and \(b\) are small because we are considering a monatomic gas under ordinary conditions, the Van der Waals equation of state can be well approximated by the ideal gas law. Otherwise we must use this more complicated relation for engineering our needs with gases.

Physicists now realize that because of complicated intermolecular forces, a real gas cannot be rigorously described by any simple equation of state. We can honestly say that the assumptions of the ideal gas are not correct, yet are sometimes useful. Likewise, the predictions of the van der Waals equation of state describe quite accurately the behavior of carbon dioxide gas in appropriate conditions. Yet for very low temperatures, carbon dioxide deviates from even these modified predictions because we know that the van der Waals model of the gas is wrong. Even this improved mathematical model is wrong, but it still is useful.

**An example from mathematical finance.** For modeling mathematical finance, we make a limited number of idealized assumptions about securities markets. We start from empirical observations of economists about supply and demand and the role of prices as a quantifiable element relating them. We ideally assume that

1. The market has a very large number of identical, rational traders.
2. All traders always have complete information about all assets they are trading.
3. Prices vary randomly with some continuous probability distribution.
4. Trading transactions take negligible time.
5. Trading transactions can be in any amounts.

These assumptions are similar to the assumptions about an ideal gas. From the assumptions we can make some standard economic arguments to derive some interesting relationships about option prices. These relationships can help us manage risk, and speculate intelligently in typical markets. However, caution is necessary.
In discussing the economic collapse of 2008-2009, blamed in part on the overuse or even abuse of mathematical models of risk, Valencia [65] says “Trying ever harder to capture risk in mathematical formulae can be counterproductive if such a degree of accuracy is intrinsically unobtainable.” If the dollar amounts get very large (so that rationality no longer holds!), or the market has only a few traders, or sharp jumps in prices occur, or the trades come too rapidly for information to spread effectively, we must proceed with caution. The observed financial outcomes may deviate from predicted ideal behavior in accurate scientific or economic work, or financial engineering.

We must then alter our assumptions, re-derive the quantitative relationships, perhaps with more sophisticated mathematics or introducing more quantities and begin the cycle of modeling again.

Section Ending Answer. One should believe in an equation to the extent that the mathematical model it represents conforms to real world principles and the relations derived from it provide meaningful and useful predictions. At commonly encountered temperatures and pressures, the ideal gas law meets both of these requirements.

Problems.

Exercise 1.15. How many jelly beans fill a cubical box 10 cm on a side?

(1) Find an upper bound for number of jelly beans in the box. Create and solve a mathematical model for this situation, enumerating explicitly the simplifying assumptions that you make to create and solve the model.

(2) Find an lower bound for number of jelly beans in the box. Create and solve a mathematical model for this situation, enumerating explicitly the simplifying assumptions that you make to create and solve the model.

(3) Can you use these bounds to find a better estimate for the number of jelly beans in the box?

(4) Suppose the jelly beans are in a 1-liter jar instead of a cubical box. (Note that a cubical box 10 cm on a side has a volume of 1 liter.) What would change about your estimates? What would stay the same? How does the container being a jar change the problem?

A good way to do this problem is to divide into small groups, and have each group work the problem separately. Then gather all the groups and compare answers.

Exercise 1.16. How many ping-pong balls fit into the room where you are reading this? Create and solve a mathematical model for this situation, enumerating explicitly the simplifying assumptions that you make to create and solve the model.

Compare this problem to the previous jelly bean problem. How is it different and how is it similar? How is it easier and how is it more difficult?

Exercise 1.17. How many flat toothpicks would fit on the surface of a sheet of poster board?

Create and solve a mathematical model for this situation, enumerating explicitly the simplifying assumptions that you make to create and solve the model.
Exercise 1.18. If your life earnings were doled out to you at a certain rate per hour for every hour of your life, how much is your time worth?
Create and solve a mathematical model for this situation, enumerating explicitly the simplifying assumptions that you make to create and solve the model.

Exercise 1.19. A ladder stands with one end on the floor and the other against a wall. The ladder slides along the floor and down the wall. A cat is sitting at the middle of the ladder. What curve does the cat trace out as the ladder slides?
Create and solve a mathematical model for this situation, enumerating explicitly the simplifying assumptions that you make to create and solve the model. How is this problem similar to, and different from, the previous problems about jelly beans in box, ping-pong balls in a room, toothpicks on a poster board, and life-earnings?

Exercise 1.20. Glenn Ledder defines the process of mathematical modeling as
“A mathematical model is a mathematical construction based on a real setting and created in the hope that its mathematical behavior will resemble the real behavior enough to be useful.”

He uses the diagram in Figure 4 as a conceptual model of the modeling process.

![Figure 4. The process of mathematical modeling according to Glenn Ledder.](image)

Compare and contrast this description of mathematical modeling with the cycle of mathematical modeling of this section, explicitly noting similarities and differences.

Exercise 1.21. While co-teaching a combined mathematics and physics course, the author emphasized a learning and problem solving strategy abbreviated as EDPIC. EDPIC stood for
- E: Explore, Experience, Experiment, Encounter;
- D: Discuss, Diagram, Draw, Describe;
- P: Plan, Principles, Physics;
- I: Implement, Investigate, Integrate;
- C: Check, Compare, Contrast.

Compare and contrast this strategy for learning and problem solving in physics with the cycle of mathematical modeling of this section, explicitly noting similarities and differences.

1.6. Randomness

Section Starter Question. What do we mean when we say something is random? What is the dictionary definition of random?
**Coin Flips and Randomness.** The simplest, most common, and in some ways most basic example of a random process is a coin flip. We flip a coin, and it lands one side up. We assign the probability 1/2 to the event that the coin will land heads and probability 1/2 to the event that the coin will land tails. But what does that assignment of probabilities really express?

To assign the probability 1/2 to the event that the coin will land heads and probability 1/2 to the event that the coin will land tails is a mathematical model that summarizes our experience with coins. We have flipped many coins many times, and we see that about half the time the coin comes up heads, and about half the time the coin comes up tails. So we abstract this observation to a mathematical model containing only one parameter, the probability of a heads.

From this simple model of the outcome of a coin flip we can derive some mathematical consequences. We will do this extensively in the chapter on limit theorems for coin-flipping. One of the first consequences we can derive is a theorem called the Weak Law of Large Numbers. This consequence reassures us that if we make the probability assignment then long term observations with the model will match our expectations. The mathematical model shows its worth by making definite predictions of future outcomes. We will prove other more sophisticated theorems, some with reasonable consequences, others are surprising. Observations show the predictions generally match experience with real coins, and so this simple mathematical model has value in explaining and predicting coin flip behavior. In this way, the simple mathematical model is satisfactory.

In other ways the probability approach is unsatisfactory. A coin flip is a physical process, subject to the physical laws of motion. The renowned applied mathematician J. B. Keller investigated coin flips in this way. He assumed a circular coin with negligible thickness flipped from a given height $y_0 = a > 0$, and considered its motion both in the vertical direction under the influence of gravity, and its rotational motion imparted by the flip until the coin lands on the surface $y = 0$. The initial conditions imparted to the coin flip are the initial upward velocity and the initial rotational velocity. With additional simplifying assumptions Keller shows that the fraction of flips which land heads approaches 1/2 if the initial vertical and rotational velocities are high enough. Keller shows more, that for high initial velocities narrow bands of initial conditions determine the outcome of heads or tails. From Keller’s analysis we see the randomness comes from the choice of initial conditions. Because of the narrowness of the bands of initial conditions, slight variations of initial upward velocity and rotational velocity lead to different outcomes. The assignment of probabilities 1/2 to heads and tails is actually a statement of the measure of the initial conditions that determine the outcome precisely.

The assignment of probabilities 1/2 to heads and tails is actually a statement of our inability to measure the initial conditions and the dynamics precisely. The heads or tails outcomes alternate in adjacent narrow initial conditions regions, so we cannot accurately predict individual outcomes. We instead measure the whole proportion of initial conditions leading to each outcome.

If the coin lands on a hard surface and bounces, the physical prediction of outcomes is now almost impossible because we know even less about the dynamics of the bounce, let alone the new initial conditions imparted by the bounce.

Another mathematician who often collaborated with J. B. Keller, Persi Diaconis, has exploited this determinism. Diaconis, an accomplished magician, is...
mathematicians Diaconis, Susan Holmes and Richard Montgomery have done an even more detailed analysis of the physics of coin flips, [14]. The coin-flipping machines pictured there help to show that flipping physical coins is actually slightly biased. Coins have a slight physical bias favoring the coin’s initial position 51% of the time. The bias results from the rotation of the coin around three axes of rotation at once. Their more complete dynamical description of coin flipping needs even more initial information.

If the coin bounces or rolls the physics becomes more complicated. This is particularly true if the coin rolls on one edge upon landing. The edges of coins are often milled with a slight taper, so the coin is really more conical than cylindrical. When landing on edge or spinning, the coin will tip in the tapered direction.

The assignment of a reasonable probability to a coin toss both summarizes and hides our inability to measure the initial conditions precisely and to compute the physical dynamics easily. The probability assignment is usually a good enough model, even if wrong. Except in circumstances of extreme experimental care with millions of measurements, using 1/2 for the proportion of heads is sensible.

Randomness and the Markets. A branch of financial analysis, generally called technical analysis, claims to predict security prices with the assumption that market data, such as price, volume, and patterns of past behavior predict future (usually short-term) market trends. Technical analysis also usually assumes that market psychology influences trading in a way that enables predicting when a stock will rise or fall.
In contrast is random walk theory. This theory claims that market prices follow a random path without influence by past price movements. The randomness makes it impossible to predict which direction the market will move at any point, especially in the short term. More refined versions of the random walk theory postulate a probability distribution for the market price movements. In this way, the random walk theory mimics the mathematical model of a coin flip, substituting a probability distribution of outcomes for the ability to predict what will really happen.

If a coin flip, although deterministic and ultimately simple in execution cannot be practically predicted with well-understood physical principles, then it should be even harder to believe that some technical forecasters predict market dynamics. Market dynamics depend on the interactions of thousands of variables and the actions of millions of people. The economic principles at work on the variables are incompletely understood compared with physical principles. Much less understood are the psychological principles that motivate people to buy or sell at a specific price and time. Even allowing that economic principles which might be mathematically expressed as unambiguously as the Lagrangian dynamics of the coin flip determine market prices, that still leaves the precise determination of the initial conditions and the parameters.

It is more practical to admit our inability to predict using basic principles and to instead use a probability distribution to describe what we see. In this text, we use the random walk theory with minor modifications and qualifications. We will see that random walk theory leads to predictions we can test against evidence, just as a coin-flip sequence can be tested against the classic limit theorems of probability. In certain cases, with extreme care, special tools and many measurements of data we may be able to discern biases, even predictability in markets. This does not invalidate the utility of the less precise first-order models that we build and investigate. All models are wrong, but some models are useful.

True Randomness. The outcome of a coin flip is physically determined. The numbers generated by an random-number-generator algorithm are deterministic, and are more properly known as pseudo-random numbers. The movements of prices in a market are governed by the hopes and fears of presumably rational human beings, and so might in principle be predicted. For each of these, we substitute a probability distribution of outcomes as a sufficient summary of what we have experienced in the past but are unable to predict precisely. Does true randomness exist anywhere? Yes, in two deeper theories, algorithmic complexity theory and quantum mechanics.

In algorithmic complexity theory, a number is not random if it is computable, that is, if a computer program will generate it, [18]. Roughly, a computable number has an algorithm that will generate its decimal digit expression. For example, for a rational number the division of the denominator into the numerator determines the repeating digit blocks of the decimal expression. Therefore rational numbers are not random, as one would expect. Irrational square roots are not random since a simple algorithm determines the digits of the nonterminating, nonrepeating decimal expression. Even the mathematical constant \( \pi \) is not random since a short formula can generate the digits of \( \pi \).

In the 1960s mathematicians A. Kolmogorov and G. Chaitin were looking for a true mathematical definition of randomness. They found one in the theory of
information: they noted that if a mathematician could produce a sequence of numbers with a computer program significantly shorter than the sequence, then the mathematician would know the digits were not random. In the algorithm, the mathematician has a simple theory that accounts for a large set of facts and allows for prediction of digits still to come, [18]. Remarkably, Kolmogorov and Chaitin showed that many real numbers do not fit this definition and therefore are random. One way to describe such non-computable or random numbers is that they are not predictable, containing nothing but one surprise after another.

This definition helps explain a paradox in probability theory. Suppose we roll a fair die 20 times. One possible result is 11111111111111111111 and another possible result is 66234441536125563152. Which result is more probable to occur? Each sequence of numbers is equally likely to occur, with probability $1/6^{20}$. However, our intuition of algorithmic complexity tells us the short program “repeat 1 20 times” gives 11111111111111111111, so it seems to be not random. A description of 66234441536125563152 requires 20 separate specifications, just as long as the number sequence itself. We then believe the first monotonous sequence is not random, while the second unpredictable sequence is random. Neither sequence is long enough to properly apply the theory of algorithmic complexity, so the intuition remains vague. The paradox results from an inappropriate application of a definition of randomness. Furthermore, the second sequence has $20!/(3! \cdot 3! \cdot 3! \cdot 4! \cdot 4!) = 3,259,095,840,000$ permutations but there is only one permutation of the first. Instead of thinking of the precise sequence we may confuse it with the more than $3 \times 10^{12}$ other permutations and believe it is therefore more likely. The confusion of the precise sequence with the set of permutations contributes to the paradox.

In the quantum world the time until the radioactive disintegration of a specific N-13 atom to a C-13 isotope is apparently truly random. It seems we fundamentally cannot determine when it will occur by calculating some physical process underlying the disintegration. Scientists must use probability theory to describe the physical processes associated with true quantum randomness.

Einstein found this quantum theory hard to accept. His famous remark is that “God does not play at dice with the universe.” Nevertheless, experiments have confirmed the true randomness of quantum processes. Some results combining quantum theory and cosmology imply even more profound and bizarre results.

**Section Ending Answer.** In ordinary conversation, the word *random* is often used as a synonym for arbitrary. It is often also used to mean accidental or unguided. Dictionaries list these meanings first while the more mathematical definition appears second.

### 1.7. Stochastic Processes

**Section Starter Question.** Name something that is both random and varies over time. Does the randomness depend on the history of the process or only on its current state?

**Definition and Notations.** A sequence or interval of random outcomes, that is, random outcomes dependent on time is a *stochastic process*. Stochastic is a synonym for random. The word is of Greek origin and means “pertaining to chance”
1. BACKGROUND

(Greek *stokhastikos*, skillful in aiming; from *stokhasts*, diviner; from *stokhazesthai*, to guess at, to aim at; and from *stochos* target, aim, guess). The modifier stochastic indicates that a subject is random in some aspect. Stochastic is often used in contrast to deterministic, which means that random phenomena are not involved.

More formally, let $J$ be subset of the non-negative real numbers. Usually $J$ is the nonnegative integers $0, 1, 2, \ldots$ or the nonnegative reals $\{t : t \geq 0\}$. $J$ is the index set of the process, and we usually refer to $t \in J$ as the time variable. Let $\Omega$ be a set, usually called the *sample space* or *probability space*. An element $\omega$ of $\Omega$ is a *sample point* or *sample path*. Let $S$ be a set of values, often the real numbers, called the *state space*. A stochastic process is a function $X : (J, \Omega) \to S$, a function of both time and the sample point to the state space.

Because we are usually interested in the probability of sets of sample points that lead to a set of outcomes in the state space and not the individual sample points, the common practice is to suppress the dependence on the sample point. That is, we usually write $X(t)$ instead of the more complete $X(t, \omega)$. Furthermore, especially if the time set is discrete, say the nonnegative integers, then we usually write the index variable or time variable as a subscript. Thus $X_n$ would be the usual notation for a stochastic process indexed by the nonnegative integers and $X_t$ or $X(t)$ is a stochastic process indexed by the non-negative reals. Because of the randomness, we can think of a stochastic process as a random sequence if the index set is the nonnegative integers and a random function if the time variable is the nonnegative reals.

**Examples.** The most fundamental example of a stochastic process is a coin flip sequence. The index set is the set of positive integers, counting the number of the flip. The sample space is the set of all possible infinite coin flip sequences $\Omega = \{HHTHTTTHT, \ldots, THTHTTHHT, \ldots\}$. We take the state space to be the set $1, 0$ so that $X_n = 1$ if flip $n$ comes up heads, and $X_n = 0$ if the flip comes up tails. Then the coin flip stochastic process can be viewed as the set of all random sequences of 1’s and 0’s. An associated random process is to take $X_0 = 0$ and $S_n = \sum_{j=0}^{n} X_j$ for $n \geq 0$. Now the state space is the set of nonnegative integers. The stochastic process $S_n$ counts the number of heads encountered in the flipping sequence up to flip number $n$.

Alternatively, we can take the same index set, the same probability space of coin flip sequences and define $Y_n = 1$ if flip $n$ comes up heads, and $Y_n = -1$ if the flip comes up tails. This is just another way to encode the coin flips now as random sequences of 1’s and -1’s. A more interesting associated random process is to take $Y_0 = 0$ and $T_n = \sum_{j=0}^{n} Y_j$ for $n \geq 0$. Now the state space is the set of integers. The stochastic process $T_n$ gives the position in the integer number line after taking a step to the right for a head, and a step to the left for a tail. This particular stochastic process is usually called a *simple random walk*. We can generalize random walk by allowing the state space to be the set of points with integer coordinates in two-, three- or higher-dimensional space, called the integer lattice, and using some random device to select the direction at each step.

**Markov Chains.** A Markov chain is sequence of random variables $X_j$ where the index $j$ runs through $0, 1, 2, \ldots$. The sample space is not specified explicitly, but it involves a sequence of random selections detailed by the effect in the state space. The state space may be either a finite or infinite set of discrete states. The
defining property of a Markov chain is that
\[ P[X_j = l \mid X_0 = k_0, X_1 = k_1, \ldots, X_{j-1} = k_{j-1}] = P[X_j = l \mid X_{j-1} = k_{j-1}] . \]
In more detail, the probability of transition from state \( k_{j-1} \) at time \( j-1 \) to state \( l \) at time \( j \) depends only on \( k_{j-1} \) and \( l \), not on the history \( X_0 = k_0, X_1 = k_1, \ldots, X_{j-2} = k_{j-2} \) of how the process got to \( k_{j-1} \).

A simple random walk is an example of a Markov chain. The states are the integers and the transition probabilities are
\[ P[X_j = l \mid X_{j-1} = k] = \begin{cases} 1/2 & \text{if } l = k - 1 \text{ or } l = k + 1, \\ 0 & \text{otherwise.} \end{cases} \]

Another example would be the position of a game piece in the board game Monopoly. The index set is the nonnegative integers listing the plays of the game, with \( X_0 \) denoting the starting position at the “Go” corner. The sample space is the set of infinite sequences of rolls of a pair of dice. The state space is the set of 40 real-estate properties and other positions around the board.

Markov chains are an important and useful class of stochastic processes. Markov chains extended to making optimal decisions under uncertainty are Markov decision processes. Another extension to signal processing and bioinformatics is the hidden Markov model. Mathematicians have extensively studied and classified Markov chains and their extensions but we will not examine them carefully in this text.

A generalization of a Markov chain is a Markov process. In a Markov process, we allow the index set to be either a discrete set of times as the integers or an interval, such as the nonnegative reals. Likewise the state space may be either a set of discrete values or an interval, even the whole real line. In mathematical notation a stochastic process \( X(t) \) is called Markov if for every \( n \) and \( t_1 < t_2 < \cdots < t_n \) and real number \( x_n \), we have
\[ P[X(t_n) \leq x_n \mid X(t_{n-1}), \ldots, X(t_1)] = P[X(t_n) \leq x_n \mid X(t_{n-1})] . \]
Many of the models in this text will naturally be Markov processes because of the intuitive modeling appeal of this memory-less property.

Many stochastic processes are naturally expressed as taking place in a discrete state space with a continuous time index. For example, consider radioactive decay, counting the number of atomic decays that have occurred up to time \( t \) by using a Geiger counter. The discrete state variable is the number of clicks heard. The mathematical Poisson process is an excellent model of this physical process. More generally, instead of radioactive events giving a single daughter particle, imagine a birth event with a random number (distributed according to some probability law) of offspring born at random times. Then the stochastic process measures the population in time. These are birth processes and make excellent models in population biology and the physics of cosmic rays. Continue to generalize and imagine that each individual in the population has a random life-span distributed according to some law, then dies. This gives a birth-and-death process. In another variation, imagine a disease with a random number of susceptible individuals getting infected, in turn infecting a random number of other individuals in the population, then recovering and becoming immune. The stochastic process counts the number of susceptible, infected and recovered individuals at any time, an SIR epidemic process.
In another variation, consider customers arriving at a service counter at random intervals with some specified distribution, often taken to be an exponential probability distribution with parameter $\lambda$. The customers get served one-by-one, each taking a random service time, again often taken to be exponentially distributed. The state space is the number of customers waiting for service, the queue length at any time. These are called queuing processes. Mathematically, these processes can be studied with \textit{compound Poisson processes}.

Continuous space processes usually take the state space to be the real numbers or some interval of the reals. One example is the magnitude of noise on top of a signal, say a radio message. In practice the magnitude of the noise can be taken to be a random variable taking values in the real numbers, and changing in time. Then subtracting off the known signal leaves a continuous-time, continuous state-space stochastic process. To mitigate the noise’s effect engineers model the characteristics of the process. To model noise means to specify the probability distribution of the random magnitude. A simple model is to take the distribution of values to be normally distributed, leading to the class of \textit{Gaussian processes} including \textit{white noise}.

Another continuous space and continuous time stochastic process is a model of the motion of particles suspended in a liquid or a gas. The random thermal perturbations in a liquid are responsible for a random walk phenomenon known as \textit{Brownian motion} and also as the \textit{Wiener process}, and the collisions of molecules in a gas create a random walk responsible for diffusion. In this process, we measure the position of the particle over time so that is a stochastic process from the nonnegative real numbers to either one-, two- or three-dimensional real space. Random walks have fascinating mathematical properties. Scientists make the model more realistic by including the effects of friction leading to a more refined form of Brownian motion called the \textit{Ornstein-Uhlenbeck process}.

Extending this idea to economics, we will model market prices of financial assets such as stocks as a continuous time, continuous space process. Random market forces create small but constantly occurring price changes. This results in a stochastic process from a continuous time variable representing time to the reals or nonnegative reals representing prices. By refining the model so that prices are nonnegative leads to the stochastic process known as \textit{geometric Brownian motion}.

\textbf{Family of Stochastic Processes.} A sequence or interval of random outcomes, that is to say, a string of random outcomes dependent on time as well as the randomness is a \textit{stochastic process}. With the inclusion of a time variable, the rich range of random outcome distributions becomes a huge variety of stochastic processes. Nevertheless, the most commonly studied types of random processes have connections. A diagram of relationships is in Figure 6, along with an indication of the stochastic process types studied in this text.

\textbf{Ways to Interpret Stochastic Processes.} Stochastic processes are functions of two variables, the time index and the sample point. As a consequence, stochastic processes are interpreted in several ways. The simplest is to look at the stochastic process at a fixed value of time. The result is a random variable with a probability distribution, just as studied in elementary probability.

Another way to look at a stochastic process is to consider the stochastic process as a function of the sample point $\omega$. Each $\omega$ maps to an associated function
Figure 6. A tree of some stochastic processes, from most general at the top to more specific at the end leaves. Stochastic processes studied in this text have thick borders, others are thin.

\(X(t)\). This means that one can look at a stochastic process as a mapping from the sample space \(\Omega\) to a set of functions. In this interpretation, stochastic processes are a generalization from the random variables of elementary probability theory. In elementary probability theory, random variables are a mapping from a sample space to the real numbers, for stochastic processes the mapping is from a sample space to a space of functions. Now we ask questions like:

- What is the probability of the set of functions that exceed a fixed value on a fixed time interval?
- What is the probability of the set of functions having a certain limit at infinity?
- What is the probability of the set of functions that are differentiable everywhere?
This is a fruitful way to consider stochastic processes, but it requires sophisticated mathematical tools and careful analysis.

Another way to look at stochastic processes is to ask what happens at special times. For example, consider the time it takes until the function takes on one of two certain values, say $a$ and $b$. Then ask “What is the probability that the stochastic process assumes the value $a$ before it assumes the value $b$?” Note that the time that each function assumes the value $a$ is different, it is a random time. This provides an interaction between the time variable and the sample point through the values of the function. This too is a fruitful way to think about stochastic processes.

In this text, we will consider each of these approaches with the corresponding questions.

**Section Ending Answer.** The cumulative number of clicks on a Geiger counter is both random, due to the random nature of radioactive disintegration, and accumulates over time, so it increases. In the next unit of time, the number of clicks will go up by none, one or few, but the cumulative number of clicks will depend only on the current number, not the history of how it was obtained.

In a temperate climate, the high temperature each day of the year will have some probability distribution, so it is random, and it certainly varies over the year. In addition to seasonal variations, during a heat wave or a cold snap the probability distribution is shifted somewhat, so the randomness does depend on the recent history.

### 1.8. A Model of Collateralized Debt Obligations

**Section Starter Question.** How can you evaluate cumulative binomial probabilities

$$P \left[ S_N \leq n \right] = \sum_{j=0}^{n} \binom{N}{j} p^j (1 - p)^{N-j}$$

when the value of $N$ is large, say $N = 100$ and the value of $p$ is small, say $p = 0.05$?

**A binomial model of mortgages.** To illustrate the previous sections altogether we will make a simple binomial probability model of a financial instrument called a CDO, standing for Collateralized Debt Obligation. The market in this derivative financial instrument is large, in 2007 amounting to at least $1.3$ trillion dollars, of which 56% came from derivatives based on residential mortgages. Heavy reliance on these financial derivatives contributed to the end of old-line brokerage firms Bear Stearns and Merrill Lynch as independent companies in the autumn of 2008. The quick loss in value of these derivatives sparked a lack of economic confidence which led to the sharp economic downturn in the fall of 2008 and the subsequent recession. We will build a “stick-figure” model of these instruments, and even this simple model will demonstrate that the CDOs were far more sensitive to mortgage failure rates than was commonly understood. While this model does not fully describe CDOs, it does provide an interesting and accessible example of the modeling process in mathematical finance.

Consider the following financial situation. A lending company has made 100 mortgage loans to home-buyers. We make two modeling assumptions about the loans.
(1) For simplicity, each loan will have precisely one of 2 outcomes. Either the home-buyer will pay off the loan resulting in a profit of 1 unit of money to the lender, or the home-buyer will default on the loan, resulting in a payoff or profit to the company of 0. For further simplicity we will say that the unit profit is $1. (The payoff is typically in the thousands of dollars.)

(2) We assume that the probability of default on a loan is \( p \) and we will assume that the probability of default on each loan is independent of default on all the other loans.

Let \( S_{100} \) be the number of loans that default, resulting in a total profit of \( 100 - S_{100} \). The probability of \( n \) or fewer of these 100 mortgage loans defaulting is

\[
P[S_{100} \leq n] = \sum_{j=0}^{n} \binom{100}{j} p^j (1 - p)^{100 - j}.
\]

We can evaluate this expression in several ways including direct calculation and approximation methods. For our purposes here, one can use a binomial probability table, or more easily a computer program which has a cumulative binomial probability function. The expected number of defaults is \( 100p \), the resulting expected loss is \( 100p \) and the expected profit is \( 100(1 - p) \).

But instead of simply making the loans and waiting for them to be paid off the loan company wishes to bundle these debt obligations differently and sell them as a financial derivative contract to an investor. Specifically, the loan company will create a collection of 101 contracts also called tranches. Contract 1 will pay 1 dollar if 0 of the loans default. Contract 2 will pay 1 dollar if 0 or 1 of the loans defaults, and in general contract \( n \) will pay 1 dollar if \( n - 1 \) or fewer of the loans defaults. (This construction is a much simplified model of mortgage backed securities. In actual practice banks combine many mortgages with various levels of risk and then package them into derivative securities with differing levels of risk called tranches. Each tranche pays out a revenue stream, not a single unit payment. A tranche is usually backed by thousands of mortgages.)

Suppose to be explicit that 5 of the 100 loans default. Then the seller will have to pay off contracts 6 through 101. The loan company who creates the contracts will receive 95 dollars from the 95 loans that do not default and will pay out 95 dollars. If the lender prices the contracts appropriately, then the lender will have enough money to cover the payout and will have some profit from selling the contracts.

For the contract buyer, the contract will either pay off with a value of 1 or will default. The probability of payoff on contract \( i \) will be the sum of the probabilities that \( i - 1 \) or fewer mortgages default:

\[
\sum_{j=0}^{i-1} \binom{100}{j} p^j (1 - p)^{100 - j},
\]

that is, a binomial cumulative distribution function. The probability of default on contract \( i \) will be a binomial complementary distribution function, which we will denote by

\[
p_T(i) = 1 - \sum_{j=0}^{i-1} \binom{100}{j} p^j (1 - p)^{100 - j}.
\]
We should calculate a few default probabilities: The probability of default on contract 1 is the probability of 0 defaults among the 100 loans,

$$p_T(1) = 1 - \binom{100}{0}p^0 (1 - p)^{100} = 1 - (1 - p)^{100}.$$ 

If $p = 0.05$, then the probability of default is 0.99408. But for the contract 10, the probability of default is 0.028188. By the 10th contract, this financial construct has created an instrument that is safer than owning one of the original mortgages! Because the newly derived security combines the risks of several individual loans, under the assumptions of the model it is less exposed to the potential problems of any one borrower.

The expected payout from the collection of contracts will be

$$\mathbb{E}[U] = \sum_{n=1}^{101} \sum_{j=0}^{n-1} \binom{100}{j} p^j (1 - p)^{100-j}$$ 

$$= \sum_{j=0}^{100} (100 - j) \binom{100}{j} p^j (1 - p)^{100-j} = 100 - 100p.$$ 

That is, the expected payout from the collection of contracts is exactly the same as the expected payout from the original collection of mortgages. However, the lender will also receive the profit of the contracts sold. Moreover, since the lender is now only selling the possibility of a payout derived from mortgages and not the mortgages themselves, the lender can even sell the same contract several times to several different buyers if the profits outweigh the risk of multiple payouts.

Why rebundle and sell mortgages as tranches? The reason is that for many of the tranches the risk exposure is less, but the payout is the same as owning a mortgage loan. Reduction of risk with the same payout is very desirable for many investors. Those investors may even pay a premium for low risk investments. In fact, some investors like pension funds are required by law, regulation or charter to invest in securities that have a low risk. Some investors may not have direct access to the mortgage market, again by law, regulation or charter, but in a rising (or bubble) market they desire to get into that market. These derivative instruments look like a good investment to them.

**Collateralized Debt Obligations.** If rebundling mortgages once is good, then doing it again should be better! So now assume that the loan company has 10,000 loans, and that it divides these into 100 groups of 100 each, and creates contracts as above. Label the groups Group 1, Group 2, and so on to Group 100. Now for each group, the bank makes 101 new contracts. Contract 1.1 will pay off $1 if 0 mortgages in Group 1 default. Contract 1.2 will pay $1 if 0 or 1 mortgages in Group 1 default. Continue, so for example, Contract 1.10 will pay $1 if 0, 1, 2, ..., 9 mortgages default in Group 1. Do this for each group, so for example, in the last group, Contract 100.10 will pay off $1 if 0, 1, 2, ..., 9 mortgages in Group 100 default. Now for example, the lender gathers up the 100 contracts $j$. 10, one from each group, into a secondary group and bundles them just as before, paying off 1 dollar if $i - 1$ or fewer of these contracts $j$. 10 defaults. These new derivative contracts are now called collateralized debt obligations or CDOs. Again, this is a much simplified “stick-figure” model of a real CDO, see [27]. Sometimes,
these second-level constructs are called a “CDO squared” [17]. Just as before, the probability of payout for the contracts $j.10$ is

$$\sum_{j=0}^{9} \binom{100}{j} \left( p_{T}(10) \right)^{j} \left( 1 - p_{T}(10) \right)^{100-j}$$

and the probability of default is

$$p_{CDO}(10) = 1 - \sum_{j=0}^{9} \binom{100}{j} \left( p_{T}(10) \right)^{j} \left( 1 - p_{T}(10) \right)^{100-j}.$$  

For example, with $p = 0.05$ and $p_{T}(10) = 0.028188$ then $p_{CDO}(10) = 0.00054385$. Roughly, the CDO has only $1/100$ of the default probability of the original mortgages, by virtue of re-distributing the risk.

The construction of the contracts $j.10$ is a convenient example to illustrate the relative values of the risk of the original mortgage, the tranche and the second-order tranche or “CDO-squared”. The number 10 for the contracts $j.10$ is not special. In fact, the bank could make a “super-CDO $j.M.N$” where it pays $1$ if 0, 1, 2, 3, ..., $M-1$ of the $N$-tranches in group $j$ fail, even without a natural relationship between $M$ and $N$. Even more is possible, the bank could make a contract that would pay some amount if some arbitrary finite sequence of contracts composed from some arbitrary sequence of tranches from some set of groups fails. We could still calculate the probability of default or non-payment, it’s just a mathematical problem. The only questions would be what contracts would be risky, what the bank could sell and how to price everything to be profitable to the bank.

The possibility of creating this “super-CDO $j.M.N$” illustrates one problem with the whole idea of CDOs that led to the collapse and the recession of 2008. These contracts quickly become confusing and hard to understand. These contracts are now so far removed from the reality of a homeowner paying a mortgage to a bank that they become their own gambling game. But the next section analyzes a more serious problem with these second-order contracts.

**Sensitivity to the parameters.** Now we investigate the robustness of the model. We do this by varying the probability of mortgage default to see how it affects the risk of the tranches and the CDOs.

Assume that the underlying mortgages actually have a default probability of 0.06, a 20% increase in the risk although it is only a 0.01 increase in the actual rates. This change in the default rate may be due to several factors. One may be the inherent inability to measure a fairly subjective parameter such as “mortgage default rate” accurately. Finding the probability of a home-owner defaulting is not the same as calculating a losing bet in a dice game. Another may be a faulty evaluation (usually overconfident or optimistic) of the default rates themselves by the agencies who provide the service of evaluating the risk on these kinds of instruments. Some economic commentators allege that before the 2008 economic crisis the rating agencies were under intense competitive pressure to provide good ratings in order to get the business of the firms who create derivative instruments. The agencies may have shaded their ratings to the favorable side in order to keep the business. Finally, the underlying economic climate may be changing and the earlier estimate, while reasonable for the prior conditions, is no longer valid. If the
economy deteriorates or the jobless rate increases, weak mortgages called sub-prime mortgages may default at increased rates.

Now with $p = 0.06$ we calculate that each contract $j.10$ has a default probability of 0.078, a 275% increase from the previous probability of 0.028. Worse, the 10th CDO-squared made of the contracts $j.10$ will have a default probability of 0.247, an increase of over 45,400%! The financial derivatives amplify any error in measuring the default rate to a completely unacceptable risk. The model shows that the financial instruments are not robust to errors in the assumptions!

But shouldn’t the companies either buying or selling the derivatives recognize this? A human tendency is to blame failures, including the failures of the Wall Street giants, on ignorance, incompetence or wrongful behavior. In this case, the traders and “rocket scientists” who created the CDOs were probably neither ignorant nor incompetent. Because they ruined a profitable endeavor for themselves, we can probably rule out misconduct too. But distraction resulting from an intense competitive environment allowing no time for rational reflection along with over-confidence during a bubble can make us willfully ignorant of the risks. A failure to properly complete the modeling cycle leads the users to ignore the real risks.

**Criticism of the model.** This model is far too simple to base any investment strategy or more serious economic analysis on it. First, a binary outcome of either pay-off or default is too simple. Lenders will restructure shaky loans or they will sell them to other financial institutions so that the lenders will probably get some return, even if less than originally intended.

The assumption of a uniform probability of default is too simple by far. Lenders make some loans to safe and reliable home-owners who dutifully pay off the mortgage in good order. Lenders also make some questionable loans to people with
poor credit ratings, these are called sub-prime loans or sub-prime mortgages. The probability of default is not the same. In fact, rating agencies grade mortgages and loans according to risk. There are 20 grades ranging from AAA with a 1-year default probability of less than 0.001 through BBB with a 1-year default probability of slightly less than 0.01 to CC with a 1-year default probability of more than 0.35. The mortgages may also change their rating over time as economic conditions change, and that will affect the derived securities. Also too simple is the assumption of an equal unit payoff for each loan.

The assumption of independence is clearly incorrect. The similarity of the mortgages increases the likelihood that most will prosper or suffer together and potentially default together. Due to external economic conditions, such as an increase in the unemployment rate or a downturn in the economy, default on one loan may indicate greater probability of default on other, even geographically separate loans, especially sub-prime loans. This is the most serious objection to the model, since it invalidates the use of binomial probabilities.

However, relaxing any assumptions make the calculations much more difficult. The non-uniform probabilities and the lack of independence means that elementary theoretical tools from probability are not enough to analyze the model. Instead, simulation models will be the next means of analysis.

Nevertheless, the sensitivity of the simple model should make us very wary of optimistic claims about the more complicated model.

Section Ending Answer. One way to evaluate cumulative binomial probabilities is to use a modern statistical computing program such as R that includes functions for precisely this calculation. An alternative is to use printed tables of the probabilities. Finally, approximations using either the normal or Poisson distribution calculate the values to satisfactory accuracy with relative simplicity.

Problems.

Exercise 1.22. Suppose a 20% decrease in the default probability from 0.05 to 0.04 occurs. By what factor do the default rates of the 10-tranches and the derived 10th CDO change?

Exercise 1.23. For the tranches create a table of probabilities of default for contracts $i = 5$ to $i = 15$ for probabilities of default $p = 0.03, 0.04, 0.05, 0.06$ and 0.07 and determine where the contracts become safer investments than the individual mortgages on which they are based.

Exercise 1.24. For a base mortgage default rate of 5%, draw the graph of the default rate of the contracts as a function of the contract number.

Exercise 1.25. The text asserts that the expected payout from the collection of contracts will be

$$E[U] = \sum_{n=1}^{101} \sum_{j=0}^{n-1} \binom{100}{j} p^j (1-p)^{100-j}$$

$$= \sum_{j=0}^{100} (100 - j) \binom{100}{j} p^j (1-p)^{100-j} = 100(1-p).$$
That is, the expected payout from the collection of contracts is exactly the same the expected payout from the original collection of mortgages. More generally, show that

\[
\sum_{n=1}^{N+1} \sum_{j=0}^{n-1} a_j = \sum_{j=0}^{N} (N - j) \cdot a_j.
\]

**Exercise 1.26.** Write the general expressions for the probabilities of payout and default for the \(i\)th contract from the CDO-squared.

**Exercise 1.27.** The following problem does not have anything to do with money, mortgages, tranches, or finance. It is instead a problem that creates and investigates a mathematical model using binomial probabilities, so it naturally belongs in this section. This problem is adapted from the classic 1943 paper by Robert Dorfman on group blood testing for syphilis among US military draftees.

Suppose that you have a large population that you wish to test for a certain characteristic in their blood or urine (for example, testing athletes for steroid use or military personnel for a particular disease). Each test will be either positive or negative. Since the number of individuals to be tested is quite large, we can expect that the cost of testing will also be large. How can we reduce the number of tests needed and thereby reduce the costs? If the blood could be pooled by putting a portion of, say, 10 samples together and then testing the pooled sample, the number of tests might be reduced. If the pooled sample is negative, then all the individuals in the pool are negative, and we have checked 10 people with one test. If, however, the pooled sample is positive, we only know that at least one of the individuals in the sample will test positive. Each member of the sample must then be retested individually and a total of 11 tests will be necessary to do the job. The larger the group size, the more we can eliminate with one test, but the more likely the group is to test positive. If the blood could be pooled by putting \(G\) samples together and then testing the pooled sample, the number of tests required might be minimized. Create a model for the blood testing cost involving the probability of an individual testing positive (\(p\)) and the group size (\(G\)) and use the model to minimize the total number of tests required. Investigate the sensitivity of the cost to the probability \(p\).

**Exercise 1.28.** The following problem does not have anything to do with money, mortgages, tranches, or finance. It is instead a problem that creates and investigates a mathematical model using binomial probabilities, so it naturally belongs in this section.

Suppose you are taking a test with 25 multiple-choice questions. Each question has 5 choices. Each problem is scored so that a correct answer is worth 6 points, an incorrect answer is worth 0 points, and an unanswered question is worth 1.5 points. You wish to score at least 100 points out of the possible 150 points. The goal is to create a model of random guessing that optimizes your chances of achieving a score of at least 100.

1. How many questions must you answer correctly, leaving all other questions blank, to score your goal of 100 points? For reference, let this number of questions be \(N\).

2. Discuss the form of your mathematical model for the number of questions required for success. What mathematics did you use for the solution?
3. Suppose you can only answer $N - 1$ questions. Create and analyze a model of guessing on unanswered questions to determine the optimal number of questions to guess.

4. Discuss the form of your mathematical model for success with guessing. What mathematics did you use for the solution?

5. Now begin some testing and sensitivity analysis. Suppose you can only answer $N - i$ questions, where $1 \leq i \leq N$. Adjust and analyze your model of guessing on unanswered questions to determine the optimal number of questions to guess.

6. Test the sensitivity of the model to changes in the probability of guessing a correct answer.

7. Critically analyze the model, interpreting it in view of the sensitivity analysis. Change the model appropriately.
CHAPTER 2

Binomial Models

2.1. Single Period Binomial Models

Section Starter Question. If you have two unknown values, how many equations will you need to derive to find the two unknown values? What must you know about the equations?

Single Period Binomial Model. The single period binomial model is the simplest possible financial model, yet it has the elements of all future models. It is strong enough to be a conceptual model of financial markets. It is simple enough to permit pencil-and-paper calculation. It can be comprehended as a whole. It is also structured enough to point to natural generalization. The single period binomial model is an excellent place to start studying mathematical finance.

Begin by reviewing some definitions:

1. A security is a promise to pay, or an evidence of a debt or property, typically a stock or a bond. A security is also referred to as an asset.
2. A bond is an interest bearing security that can either make regular interest payments or a lump sum payment at maturity or both.
3. A stock is a security representing partial ownership of a company, varying in value with the value of the company. Stocks are also known as shares or equities.
4. A derivative is a security whose value depends on or is derived from the future price or performance of another security. Derivatives are also known as financial derivatives, derivative securities, derivative products, and contingent claims.

The quantifiable elements of the single period binomial financial model are:

1. A single interval of time, from $t = 0$ to $t = T$.
2. A single stock of initial value $S$ and final value $S_T$. In the time interval $[0, T]$ it may either increase by a factor $U$ to value $S_T = SU$ with probability $p$, or it may decrease in value by factor $D$ to value $S_T = SD$ with probability $q = 1 - p$. The reason for multiplication by a factor instead of adding or subtracting an incremental amount is that changes are often in terms of percentages or relative amounts.
3. A single bond with a continuously compounded interest rate $r$ over the interval $[0, T]$. If the initial value of the bond is $B$, then the final value of the bond will be $B \exp(rT)$.
4. A market for derivatives (such as options) dependent on the value of the stock at the end of the period. The payoff of the derivative to the investor would be $f(SU)$ and $f(SD)$. For example, a futures contract with strike
price $K$ would have value $f(S_T) = S_T - K$. A call option with strike price $K$ would have value $f(S_T) = \max(S_T - K, 0)$.

A realistic financial assumption would be $D < \exp(rT) < U$. Then investment in the risky security may pay better than investment in a risk free bond, but it may also pay less! The mathematics only requires that $U \neq D$, see below.

We can try to find the value of the derivative by creating a portfolio of the stock and the bond that will have the same value as the derivative itself in any circumstance, called a replicating portfolio. Consider a portfolio consisting of $\phi$ units of the stock worth $\phi S$ and $\psi$ units of the bond worth $\psi B$. Note the assumption that the stock and bond are divisible, so we can buy them in any amounts including negative amounts that are short positions. If we were to buy the this portfolio at time zero, it would cost

$$\phi S + \psi B.$$
One time period of length $T$ on the trading clock later, the portfolio would be worth
\[ \phi SD + \psi B \exp(rT) \]
after a down move and
\[ \phi SU + \psi B \exp(rT) \]
after an up move. You should find this mathematically meaningful: there are two unknown quantities $\phi$ and $\psi$ to buy for the portfolio, and we have two expressions to match with the two values of the derivative! That is, the portfolio will have the same value as the derivative if
\[ \phi SD + \psi B \exp(rT) = f(SD) \]
\[ \phi SU + \psi B \exp(rT) = f(SU). \]
The solution is
\[ \phi = \frac{f(SU) - f(SD)}{SU - SD} \]
and
\[ \psi = \frac{f(SU) - f(SD) \cdot SU}{B \exp(\tau T) - f(SU) - f(SD) \cdot B \exp(\tau T)}. \]
Note that the solution requires $SU \neq SD$, but we have already assumed this natural requirement. Without this requirement there would be no risk in the stock, and we would not be asking the question in the first place. The value (or price) of the portfolio, and therefore the derivative should then be
\[ V = \phi S + \psi B \]
\[ = \frac{f(SU) - f(SD)}{SU - SD} \cdot S + \left[ \frac{f(SD)}{B \exp(\tau T)} - \frac{(f(SU) - f(SD)) \cdot SD}{(SU - SD) B \exp(\tau T)} \right] \cdot B \]
\[ = \frac{f(SU) - f(SD)}{U - D} + \frac{1}{\exp(\tau T)} \cdot \frac{f(SD) U - f(SU) D}{(U - D)}. \]
We can make one final simplification that will be useful in the next section. Define
\[ \pi = \frac{\exp(\tau T) - D}{U - D} \]
so then
\[ 1 - \pi = \frac{U - \exp(\tau T)}{U - D} \]
so that we write the value of the derivative as
\[ V = \exp(-rT)[\pi f(SU) + (1 - \pi) f(SD)]. \]
Here $\pi$ is not used as the mathematical constant giving the ratio of the circumference of a circle to its diameter. Instead the Greek letter for $p$ suggests a similarity to the probability $p$.

Now consider some other trader offering to sell this derivative with payoff function $f$ for a price $P$ less than $V$. Anyone could buy the derivative in arbitrary quantity, and short the $(\phi, \psi)$ stock-bond portfolio in exactly the same quantity. At the end of the period, the value of the derivative would be exactly the same as the portfolio. So selling each derivative would repay the short with a profit of $V - P$ and the trade carries no risk! So $P$ would not have been a rational price for the trader to quote and the market would have quickly mobilized to take advantage of the “free” money on offer in arbitrary quantity. This ignores transaction costs.
For an individual, transaction costs might eliminate the profit. However, for large firms trading in large quantities, transaction costs can be minimal.

Similarly, if a seller quoted the derivative at a price $P$ greater than $V$, anyone could short sell the derivative and buy the $(\phi, \psi)$ portfolio to lock in a risk-free profit of $P - V$ per unit trade. Again, the market would take advantage of the opportunity. Hence, $V$ is the only rational price for the derivative. We have determined the price of the derivative through arbitrage.

**How not to price the derivative and a hint of a better way.** Note that we did not determine the price of the derivative in terms of the expected value of the stock or the derivative. A seemingly logical thing to do would be to say that the derivative will have value $f(SU)$ with probability $p$ and will have value $f(SD)$ with probability $1 - p$. Therefore, the expected value of the derivative at time $T$ is

$$E[f] = pf(SU) + (1 - p)f(SD).$$

The present value of the expectation of the derivative value is

$$\exp(-rT)E[f] = \exp(-rT)[pf(SU) + (1 - p)f(SD)].$$

Except in the peculiar case that the expected value just happened to match the value $V$ of the replicating portfolio, arbitrage drives pricing by expectation out of the market! The problem is that the probability distribution $(p, 1 - p)$ only takes into account the movements of the security price. The expected value is the value of the derivative over many identical iterations or replications of that distribution, but there will be only one trial of this particular experiment, so expected value is not a reasonable way to weight the outcomes. Also, the expected value does not take into account the rest of the market. In particular, the expected value does not take into account that an investor has the opportunity to simultaneously invest in alternate combinations of the risky stock and the risk-free bond. A special combination of the risky stock and risk-free bond replicates the derivative. As such the movement probabilities alone do not completely assess the risk associated with the transaction.

Nevertheless, we still might have with a nagging feeling that pricing by arbitrage as done above ignores the probability associated with security price changes. One may legitimately ask if there is a way to value derivatives by taking some kind of expected value. The answer is yes, a different probability distribution associated with the binomial model correctly takes into account the rest of the market. In fact, the quantities $\pi$ and $1 - \pi$ define this probability distribution. This is called the risk-neutral measure or more completely the risk-neutral martingale measure. Economically speaking, the market assigns a fairer set of probabilities $\pi$ and $1 - \pi$ that give a value for the option compatible with the no-arbitrage principle. Another way to say this is that the market changes the odds to make option pricing fairer. The risk-neutral measure approach is the modern, sophisticated, and general way to approach derivative pricing.

**Section Ending Answer.** It takes two linearly independent equations to determine two unknown values. This algebraic principle applies here by setting equations for each of the two derivative values in terms of the two unknowns for the security and bond amounts in the replicating portfolio.

**Algorithms, Scripts, Simulations.**
**Algorithm.** The aim is to set up and solve for the replicating portfolio and the value of the corresponding derivative security in a single period binomial model. The scripts output the portfolio amounts and the derivative security value. First set values of $S$, $U$, $D$, $r$, $T$, and $K$. Define the derivative security payoff function. Set the matrix of coefficients in the linear system. Set the right-hand-side payoff vector. Solve for the portfolio using a linear solver. Finally, take the dot product of the portfolio with the security and bond values to get the derivative security value.

```r
S <- 50
down <- 0.10
up <- 0.03
B <- 1
r <- 0.06
T <- 1
K <- 50

f <- function(x, strike) {
  max(x - strike, 0)
}

m <- rbind(c(S * (1-down), B*exp(r*T)), c(S * (1+up), B*exp(r*T)))
payoff <- c(f(S * (1-down), K), f(S * (1+up), K))
portfolio <- solve(m, payoff)
value <- portfolio %*% c(S,B)

cat("portfolio: phi = ", portfolio[1], " psi = ", portfolio[2], 
     "value = ", value, 
"
"
"
"
"
```

**Key Concepts.**

1. The simplest model for pricing an option is based on a market having a single period, a single security having two uncertain outcomes, and a single bond.
2. Replication of the option payouts with the single security and the single bond leads to pricing the derivative by arbitrage.

**Vocabulary.**

1. A **security** is a promise to pay, or an evidence of a debt or property, typically a stock or a bond. A security is also referred to as an **asset**.
2. A **bond** is an interest bearing security that can either make regular interest payments or a lump sum payment at maturity or both.
3. A **stock** is a security representing partial ownership of a company, varying in value with the value of the company. Stocks are also known as **shares** or **equities**.
4. A **derivative** is a security whose value depends on or is derived from the future price or performance of another security. Derivatives are also known as **financial derivatives**, **derivative securities**, **derivative products**, and **contingent claims**.
50 2. BINOMIAL MODELS

(5) A portfolio of the stock and the bond that will have the same value as the derivative itself in any circumstance is a replicating portfolio.

Problems.

Exercise 2.1. Consider a stock whose price today is $50. Suppose that over the next year, the stock price can either go up by 10%, or down by 3%, so the stock price at the end of the year is either $55 or $48.50. The continuously compounded interest rate on a $1 bond is 6%. If there also exists a call on the stock with an exercise price of $50, then what is the price of the call option? Also, what is the replicating portfolio?

Exercise 2.2. A stock price is currently $50. It is known that at the end of 6 months, it will either be $60 or $42. The risk-free rate of interest with continuous compounding on a $1 bond is 12% per annum. Calculate the value of a 6-month European call option on the stock with strike price $48 and find the replicating portfolio.

Exercise 2.3. A stock price is currently $40. It is known that at the end of 3 months, it will either $45 or $34. The risk-free rate of interest with quarterly compounding on a $1 bond is 8% per annum. Calculate the value of a 3-month European put option on the stock with a strike price of $40, and find the replicating portfolio.

Exercise 2.4. Your friend, the financial analyst, comes to you, the mathematical economist, with a proposal: “The single period binomial pricing is all right as far as it goes, but it is certainly is simplistic. Why not modify it slightly to make it a little more realistic? Specifically, assume the stock can assume three values at time T, say it goes up by a factor U with probability p_U, it goes down by a factor D with probability p_D, where D < 1 < U and the stock stays somewhere in between, changing by a factor M with probability p_M where D < M < U and p_D + p_M + p_U = 1.” The market contains only this stock, a bond with a continuously compounded risk-free rate r and an option on the stock with payoff function f(S_T). Make a mathematical model based on your friend’s suggestion and provide a critique of the model based on the classical applied mathematics criteria of existence of solutions to the model and uniqueness of solutions to the model.

Exercise 2.5. Modify the script to accept interactive inputs of important parameters and to output results nicely formatted and rounded to cents.

2.2. Multiperiod Binomial Tree Models

Section Starter Question. Suppose that you owned a 3-month option, and that you tracked the value of the underlying security at the end of each month. Suppose you were forced to sell the option at the end of two months. How would you find a fair price for the option at that time? What simple modeling assumptions would you make?

The Binomial Tree model. The multiperiod binomial model has N time intervals created by N+1 trading times t_0 = 0, t_1, ..., t_N = T. The spacing between time intervals is Δt_i = t_i - t_{i-1}, and typically the spacing is equal, although it is not necessary. The time intervals can be any convenient time length appropriate
for the model, e.g. months, days, minutes, even seconds. Later, we will take them to be relatively short compared to $T$.

We model a limited market where a trader can buy or short-sell a risky security (for instance a stock) and lend or borrow money at a riskless rate $r$. For simplicity we assume $r$ is constant over $[0, T]$. This assumption of constant $r$ is not necessary, taking $r$ to be $r_i$ on $[t_{i-1}, t_i]$ only makes calculations messier.

$S_n$ denotes the price of the risky security at time $t_n$ for $n = 0, 1, \ldots, N$. This price changes according to the rule $S_{n+1} = S_n H_{n+1}$, $0 \leq n \leq N - 1$ where $H_{n+1}$ is a Bernoulli (two-valued) random variable such that

$H_{n+1} = \begin{cases} U, & \text{with probability } p \\ D, & \text{with probability } q = 1 - p. \end{cases}$

Again for simplicity we assume $U$ and $D$ are constant over $[0, T]$. This assumption of constant $U$, $D$ is not necessary, for example, taking $U$ to be $U_i$ for $i = 0, 1, \ldots, N$ only makes calculations messier. A binomial tree is a way to visualize the multiperiod binomial model, as in Figure 2.

A pair of integers $(n, j)$, with $n = 0, \ldots, N$ and $j = 0, \ldots, n$ identifies each node in the tree. We use the convention that node $(n, j)$ leads to nodes $(n+1, j)$ and $(n+1, j+1)$ at the next trading time, with the “up” change corresponding to $(n+1, j+1)$ and the “down” change corresponding to $(n+1, j)$. The index $j$ counts the number of up changes to that time, so $n-j$ is the number of down changes. Several paths lead to node $(n, j)$, in fact $\binom{n}{j}$ of them. The price of the risky underlying asset at trading time $t_n$ is then $SU^j D^{n-j}$. The probability of going from price $S$ to price $SU^j D^{n-j}$ is

$p_{n,j} = \binom{n}{j} p^j (1-p)^{n-j}.$

To value a derivative with payout $f(S_N)$, the key idea is that of dynamic programming — extending the replicating portfolio and corresponding portfolio values back one period at a time from the claim values to the starting time.

An example will make this clear. Consider a binomial tree on the times $t_0$, $t_1$, $t_2$. Assume $U = 1.05$, $D = 0.95$, and $\exp(r \Delta t_i) = 1.02$, so the effective interest rate on each time interval is 2%. We take $S_0 = 100$. We value a European call option with strike price $K = 100$, see Figure 3. Using the formula derived in the previous section

$\pi = \frac{1.02 - 0.95}{1.05 - 0.95} = 0.7$

and $1 - \pi = 0.3$. Then concentrating on the single period binomial branch in the large square box, the value of the option at node $(1, 1)$ is $7.03$ (rounded to cents). Likewise, the value of the option at node $(1, 0)$ is $0$. Then we work back one step and value a derivative with potential payouts $7.03$ and $0$ on the single period binomial branch at $(0, 0)$. This uses the same arithmetic to obtain the value $4.83$ (rounded to cents) at time 0. In the figure, the values of the security at each node are in the circles, the value of the option at each node is in the small box beside the circle.

As another example, consider a European put on the same security, see Figure 4. The strike price is again 100. All of the other parameters are the same. We work
backward again through the tree to obtain the value at time 0 as $0.944$. In the figure, the values of the security at each node are in the circles, the value of the option at each node is in the small box beside the circle.

The multiperiod binomial model for pricing derivatives of a risky security is also called the Cox-Ross-Rubenstein model or CRR model for short, after those who introduced it in 1979.

Advantages and Disadvantages of the model. The disadvantages of the binomial model are:

1. Trading times are not really at discrete times, trading goes on continuously.
2. Securities do not change value according to a Bernoulli (two-valued) distribution on a single time step, or a binomial distribution on multiple time periods, they change over a range of values with a continuous distribution.
3. The calculations are tedious.
(4) Developing a continuous theory will take detailed limit-taking considerations.

The advantages of the model are:

(1) It clearly reveals the construction of the replicating portfolio.
(2) It clearly reveals that the probability distribution is not centrally involved, since expectations of outcomes aren’t used to value the derivatives.
(3) It is simple to calculate, although it can get tedious.
(4) It reveals that we need more probability theory to get a complete understanding of path dependent probabilities of security prices.

It is possible, with considerable attention to detail, to make a limiting argument and pass from the binomial tree model of Cox, Ross and Rubenstein to the Black-Scholes pricing formula. However, this approach is not the most instructive. Instead, we will back up from derivative pricing models, and consider simpler
models with only risk, that is, gambling, to get a more complete understanding of stochastic processes before returning to pricing derivatives.

Some caution is also needed when reading from other sources about the Cox-Ross-Rubenstein or Binomial Option Pricing Model. Many other sources derive the Binomial Option Pricing Model by discretizing the Black-Scholes Option Pricing Model. The discretization is different from building the model from scratch because the parameters have special and more restricted interpretations than the simple model. More sophisticated discretization procedures from the numerical analysis of partial differential equations also lead to additional discrete option pricing models that are hard to justify by building them from scratch. The discrete models derived from the Black-Scholes model are used for simple and rapid numerical evaluation of option prices rather than for motivation.

**Section Ending Answer.** To find the no-arbitrage price of a 3-month option after two months, a trader could make a 3-period binomial model of one month
2.2. MULTIPERIOD BINOMIAL TREE MODELS

Each period. Assuming the interest rates and the up and down percentage changes of the security are constant each month, the dynamic programming method of this section allows a trader to calculate back one period back from the end to determine the price.

Algorithm, Scripts, Simulations.

Algorithm. The goal is to set up and solve for the value of the European call option in a two period binomial model. The script will output the derivative security value. First set values of $S$, $U$, $D$, $r$, $T$, and $K$. Define the derivative security payoff function. In the script, it is for a European call option. Define the risk neutral measure $\pi$. Solve for derivative values at $(1,1)$ and $(1,0)$ with the risk neutral measure formula. Solve for the derivative value with the risk neutral measure formula linear solver. Finally, print the derivative value.

```r
S <- 100
factorUp <- 1.05
factorDown <- 0.95
B <- 1
effR <- 1.02
deltati <- 1
K <- 100

f <- function (x, strike) {
  # European call option
  max (x - strike, 0)
}

riskNeutralMeas <- function (fUp , fDown , exprdt ) {
  # risk neutral measure $\pi$
  (exprdt - fDown)/(fUp - fDown)
}

piRNM <- riskNeutralMeas (factorUp , factorDown , effR)

v11 <- (1/ effR ) * (piRNM * f(S * factorUp * factorUp , K) + (1 - piRNM) * f(S * factorUp * factorDown, K))

v10 <- (1/effR) * (piRNM * f(S * factorUp * factorDown, K) + (1 - piRNM) * f(S * factorDown * factorDown, K))

value <- (1/ effR) * (piRNM * v11 + (1 - piRNM) * v10)

cat("value:", value, 
```

Key Concepts.
(1) A multiperiod binomial derivative model can be valued by dynamic pro-
gramming — computing the replicating portfolio and corresponding port-
folio values back one period at a time from the claim values to the starting
time.

Vocabulary.
(1) The multiperiod binomial model for pricing derivatives of a risky security
is also called the Cox-Ross-Rubenstein model or CRR model for
short, after those who introduced it in 1979.

Problems.

Exercise 2.6. Consider a two-time-stage example. Each time stage is a year.
A stock starts at 50. In each year, the stock can go up by 10% or down by 3%. The
continuously compounded interest rate on a $1 bond is constant at 6% each year.
Find the price of a call option with exercise price 50, with exercise date at the end
of the second year. Also, find the replicating portfolio at each node.

Exercise 2.7. Consider a three-time-stage example. The first time interval
is a month, then the second time interval is two months, finally, the third time
interval is a month again. A stock starts at 50. In the first interval, the stock can
go up by 10% or down by 3%, in the second interval the stock can go up by 5% or
down by 5%, finally in the third time interval, the stock can go up by 6% or down
by 3%. The continuously compounded interest rate on a $1 bond is 2% in the first
period, 3% in the second period, and 4% in the third period. Find the price of a
call option with exercise price 50, with exercise date at the end of the 4 months.
Also, find the replicating portfolio at each node.

Exercise 2.8. A European cash-or-nothing binary option pays a fixed amount
of money if it expires with the underlying stock value above the strike price. The
binary option pays nothing if it expires with the underlying stock value equal to or
less than the strike price. A stock currently has price $100 and goes up or down
by 20% in each time period. What is the value of such a cash-or-nothing binary
option with payoff $100 at expiration 2 time units in the future and strike price
$100? Assume a simple interest rate of 10% in each time period.

Exercise 2.9. A long strangle option pays \( \max(K_1 - S, 0, S - K_2) \) if it expires
when the underlying stock value is \( S \). The parameters \( K_1 \) and \( K_2 \) are the lower
strike price and the upper strike price, and \( K_1 < K_2 \). A stock currently has price
$100 and goes up or down by 20% in each time period. What is the value of such a
long strangle option with lower strike 90 and upper strike 110 at expiration 2 time
units in the future? Assume a simple interest rate of 10% in each time period.

Exercise 2.10. A long straddle option pays \( |S - K| \) if it expires when the
underlying stock value is \( S \). The option is a portfolio composed of a call and a put
on the same security with \( K \) as the strike price for both. A stock currently has
price $100 and goes up or down by 10% in each time period. What is the value of
such a long straddle option with strike price \( K = 110 \) at expiration 2 time units in
the future? Assume a simple interest rate of 5% in each time period.
CHAPTER 3

First Step Analysis

3.1. A Coin Tossing Experiment

Section Starter Question. Suppose you start with a fortune of $10, and you want to gamble to achieve $20 before you go broke. Your gambling game is to flip a fair coin successively, and gain $1 if the coin comes up heads and lose $1 if the coin comes up tails. What do you estimate is the probability of achieving $20 before going broke? How long do you estimate it takes before one of the two outcomes occurs? How do you estimate each of these quantities?

Pure Risk Modeling. We need a better understanding of the paths that risky securities take. We shall make and investigate a greatly simplified model of randomness and risk. For our model, we assume:

1. Time is discrete, occurring at \( t_0 = 0, t_1, t_2, \ldots \).
2. No risk-free investments are available, (i.e. no interest-bearing bonds.)
3. No options, and no financial derivatives are available.
4. The only investments are risk-only, that is, our net fortune at the \( n \)th time is a random variable:

\[
T_{n+1} = T_n + Y_{n+1}
\]

where \( T_0 = 0 \) is our given initial fortune and set \( Y_0 = T_0 \) for convenience.

For simplicity, take

\[
Y_n = \begin{cases} +1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2. \end{cases}
\]

Our model is commonly called gambling and we will investigate the probability of making a fortune by gambling.

An Experiment. The reader should perform the following experiment as a gambler to gain intuition about the coin-tossing game. We call victory the outcome of reaching a fortune goal by chance before going broke. We call ruin the outcome of going broke by chance before reaching a fortune goal.

1. Each gambler has a chart for recording the outcomes of each game (see below) and a sheet of graph paper.
2. Each gambler has a fair coin to flip, say a penny.
3. Each gambler flips the coin, and records a +1 (gains $1) if the coin comes up heads and records -1 (loses $1) if the coin comes up tails. On the chart, the player records the outcome of each flip with the flip number, the outcome as “H” or “T” and keeps track of the cumulative fortune of the gambler so far. Keep these records in a neat chart, since some problems refer to them later.
Each gambler should record 100 flips, which takes about 10 to 20 minutes.

<table>
<thead>
<tr>
<th>Toss n</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>H or T</td>
<td>Y_n = +1, -1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T_n</td>
<td>$\sum_{i=1}^{n} Y_i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toss n</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>H or T</td>
<td>Y_n = +1, -1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T_n</td>
<td>$\sum_{i=1}^{n} Y_i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toss n</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>H or T</td>
<td>Y_n = +1, -1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T_n</td>
<td>$\sum_{i=1}^{n} Y_i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toss n</td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>34</td>
<td>35</td>
<td>36</td>
<td>37</td>
<td>38</td>
<td>39</td>
<td>40</td>
</tr>
<tr>
<td>H or T</td>
<td>Y_n = +1, -1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T_n</td>
<td>$\sum_{i=1}^{n} Y_i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toss n</td>
<td>41</td>
<td>42</td>
<td>43</td>
<td>44</td>
<td>45</td>
<td>46</td>
<td>47</td>
<td>48</td>
<td>49</td>
<td>50</td>
</tr>
<tr>
<td>H or T</td>
<td>Y_n = +1, -1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T_n</td>
<td>$\sum_{i=1}^{n} Y_i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toss n</td>
<td>51</td>
<td>52</td>
<td>53</td>
<td>54</td>
<td>55</td>
<td>56</td>
<td>57</td>
<td>58</td>
<td>59</td>
<td>60</td>
</tr>
<tr>
<td>H or T</td>
<td>Y_n = +1, -1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T_n</td>
<td>$\sum_{i=1}^{n} Y_i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toss n</td>
<td>61</td>
<td>62</td>
<td>63</td>
<td>64</td>
<td>65</td>
<td>66</td>
<td>67</td>
<td>68</td>
<td>69</td>
<td>70</td>
</tr>
<tr>
<td>H or T</td>
<td>Y_n = +1, -1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T_n</td>
<td>$\sum_{i=1}^{n} Y_i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toss n</td>
<td>71</td>
<td>72</td>
<td>73</td>
<td>74</td>
<td>75</td>
<td>76</td>
<td>77</td>
<td>78</td>
<td>79</td>
<td>80</td>
</tr>
<tr>
<td>H or T</td>
<td>Y_n = +1, -1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T_n</td>
<td>$\sum_{i=1}^{n} Y_i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toss n</td>
<td>81</td>
<td>82</td>
<td>83</td>
<td>84</td>
<td>85</td>
<td>86</td>
<td>87</td>
<td>88</td>
<td>89</td>
<td>90</td>
</tr>
<tr>
<td>H or T</td>
<td>Y_n = +1, -1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T_n</td>
<td>$\sum_{i=1}^{n} Y_i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toss n</td>
<td>91</td>
<td>92</td>
<td>93</td>
<td>94</td>
<td>95</td>
<td>96</td>
<td>97</td>
<td>98</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>H or T</td>
<td>Y_n = +1, -1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T_n</td>
<td>$\sum_{i=1}^{n} Y_i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Some Outcomes. A simulation with 30 gamblers obtained the following results:

1. 14 gamblers reached a net loss of −10 before reaching a net gain of +10, that is, were ruined.
3.1. A COIN TOSsing EXPERIMENT

(2) 9 gamblers reached a net gain of +10 before reaching a net loss of −10, that is, achieved victory.

(3) 7 gamblers still had not reached a net loss of +10 or −10 yet.

This roughly matches the predicted outcomes of 1/2 the gamblers being ruined, not as close for the predicted proportion of 1/2 for victory. The mean duration until victory or ruin was 41.6. Compare this to your results.

Section Ending Answer. Was your estimation of the probability of achieving $20 before going broke close to the results from your simulation? Was it close to the average of 30 outcomes cited above? Was your estimation of the time it takes to reach one of these outcomes close to your experimental results? How did it compare to the mean duration for 30 experiments? Does the experiment give you a better feeling for estimating these quantities?

Algorithms, Scripts, Simulations.

Algorithm. The probability of heads is $p$. In matrix-oriented languages, use $k$ columns of the matrix for the iterations of an experiment each with $n$ trials. This is more efficient than using for-loops or the equivalent.

Then each $n \times 1$ column vector contains the outcomes of the experiment. Use the random number generator to create a random number in (0, 1). To simulate the flip of a coin use logical functions to score a 1, also counted as a Head, if the value is less than $p$; a 0, also counted as a Tail, if the value is greater than $p$. Column sums then give a $1 \times k$ vector of total numbers of heads. Statistical measures can then be applied to this vector. For wins and losses, linearly transform the coin flips to 1 if heads, a −1 if tails. Cumulatively sum the entries to create the running total. Apply statistical measures to the last row of the totals.

Finally, set values for victory and ruin. Check when each column first hits victory or ruin.

```r
1 p <- 0.5
2 n <- 100
3 k <- 30
4 coinFlips <- array( 0+(runif(n*k) <= p), dim=c(n,k)) # 0+ coerces Boolean to numeric
5 headsTotal <- colSums(coinFlips) # 0..n binomial rv sample, size k
6 muHeads <- mean(headsTotal) # Expected value is n/2
7 sigmaSquaredHeads <- var(headsTotal) # Theoretical value is np(1-p)
8 cat(sprintf("Empirical Mean of Heads : %f \n", muHeads ))
9 cat(sprintf("Empirical Variance of Heads : %f \n", sigmaSquaredHeads ))
10
11 winLose <- 2*coinFlips - 1 # -1 for Tails, 1 for Heads
12 totals <- apply( winLose, 2, cumsum)
13 # the second argument ‘2’ means column-wise
14 muWinLose <- mean( totals[,]) # Expected value is 0
15 sigmaSquaredWinLose <- var( totals[,]) # Theoretical value is 4np(1-p)
16 cat(sprintf("Empirical Mean of Wins minus Losses : %f \n", muWinLose ))
```
19 cat(sprintf("Empirical Variance of Wins minus Losses: %f \n", 
    sigmaSquaredWinLose ))

1 victory <- 10  # top boundary for random walk
2 ruin <- -10    # bottom boundary for random walk
3
4 victoryOrRuin <- array(0, dim = c(k))
5
6 hitVictory <- apply( (totals >= victory), 2, which)
7 hitRuin <- apply( (totals <= ruin), 2, which)
8
9 for (j in 1:k) {
  10    if ( length(hitVictory[[j]]) == 0 && length(hitRuin[[j]]) == 0 ) {
  11      # no victory, no ruin
  12      # do nothing
  13    }
  14    else if ( length(hitVictory[[j]]) > 0 && length(hitRuin[[j]]) == 0 ) {
  15      # victory, no ruin
  16      victoryOrRuin[j] <- min(hitVictory[[j]])
  17    }
  18    else if ( length(hitVictory[[j]]) == 0 && length(hitRuin[[j]]) > 0 ) {
  19      # no victory, ruin
  20      victoryOrRuin[j] <- -min(hitRuin[[j]])
  21    }
  22    else # ( length(hitVictory[[j]]) > 0 && length(hitRuin[[j]]) > 0 )
  23      # victory and ruin
  24      if ( min(hitVictory[[j]]) < min(hitRuin[[j]]) ) { # victory first
  25        victoryOrRuin[j] <- min(hitVictory[[j]]) # code hitting victory
  26      } else { # ruin first
  27        victoryOrRuin[j] <- -min(hitRuin[[j]]) # code hitting ruin as negative
  28    }
  29  }
 30 }

31 victoryBeforeRuin <- sum(0*(victoryOrRuin > 0))  # count exits through top
32 ruinBeforeVictory <- sum(0*(victoryOrRuin < 0))  # count exits through bottom
33 noRuinOrVictory <- sum(0*(victoryOrRuin == 0))
34
35 cat(sprintf("Victories: %i Ruins: %i No Ruin or Victory: %i \n", 
    victoryBeforeRuin, ruinBeforeVictory, noRuinOrVictory))
36
37 avgTimeVictoryOrRuin <- mean( abs(victoryOrRuin) )
38 varTimeVictoryOrRuin <- var( abs(victoryOrRuin) )
39
40 cat(sprintf("Average Time to Victory or Ruin: %f \n", 
    avgTimeVictoryOrRuin))
3.2. Ruin Probabilities

Section Starter Question. What is the solution of the recurrence equation
\[ x_n = ax_{n-1} \] where \( a \) is a constant? What kind of a function is the solution? What more, if anything, needs to be known to obtain a complete solution?

Understanding a Stochastic Process. We consider a sequence of Bernoulli random variables, \( Y_1, Y_2, Y_3, \ldots \) where \( Y_i = +1 \) with probability \( p \) and \( Y_i = -1 \) with probability \( q \). We start with an initial value \( T_0 > 0 \) and set \( Y_0 = T_0 \) for convenience. We define the sequence of sums \( T_n = \sum_{i=0}^{n} Y_i \). We are interested in the stochastic process \( T_1, T_2, T_3, \ldots \). It turns out this is a complicated sequence to understand in full, so we single out particular simpler features to understand first. For example, we can look at the probability that the process will achieve the value

Key Concepts.
(1) Performing an experiment to gain intuition about coin-tossing games.

Vocabulary.
(1) We call victory the outcome of reaching a fortune goal by chance before going broke.
(2) We call ruin the outcome of going broke by chance before reaching a fortune goal.

Problems.

Exercise 3.1. How many heads were obtained in your sequence of 100 flips? What is the class average of the number of heads in 100 flips? What is the variance of the number of heads obtained by the class members?

Exercise 3.2. What is the net win-loss total obtained in your sequence of 100 flips? What is the class average of the net win-loss total in 100 flips? What is the variance of the net win-loss total obtained by the class members?

Exercise 3.3. How many of the class reached a net gain of +10 (call it victory) before reaching a net loss of −10 (call it ruin)?

Exercise 3.4. How many flips did it take before you reached a net gain of +10 (victory) or a net loss of −10 (ruin)? What is the class average of the number of flips before reaching victory before ruin at +10 or a net loss of −10?

Exercise 3.5. What is the maximum net value achieved in your sequence of flips? What is the class distribution of maximum values achieved in the sequence of flips?

Exercise 3.6. Perform some simulations of the coin-flipping game, varying \( p \). How does the value of \( p \) affect the experimental probability of victory and ruin?
0 before it achieves the value \( a \). This is a special case of a larger class of probability problems called \textit{first passage probabilities}.

**Theorems about Ruin Probabilities.** Consider a gambler who wins or loses a dollar on each turn of a game with probabilities \( p \) and \( q = 1 - p \) respectively. Let his initial capital be \( T_0 > 0 \). The game continues until the gambler’s capital either is reduced to 0 or has increased to \( a \). Recall that we call \textit{victory} the outcome of reaching a fortune goal by chance before going broke and \textit{ruin} the outcome of going broke by chance before reaching a fortune goal. Let \( q_{T_0} \) be the probability of the gambler’s ultimate ruin and \( p_{T_0} \) the probability of his winning. We shall show later that (see also the section on Duration of the Game Until Ruin)

\[
p_{T_0} + q_{T_0} = 1
\]

so that we need not consider the positive probability of an unending game.

**THEOREM 3.7.** The probability of the gambler’s ruin is

\[
q_{T_0} = \frac{(q/p)^a - (q/p)^{T_0}}{(q/p)^a - 1}
\]

if \( p \neq q \) and

\[
q_{T_0} = 1 - T_0/a
\]

if \( p = q = 1/2 \).

**PROOF.** The proof uses a \textit{first step analysis} considering how the probabilities change after one step or trial. After the first trial the gambler’s fortune is either \( T_0 - 1 \) or \( T_0 + 1 \) and therefore we must have

\[
q_{T_0} = pq_{T_0+1} + qq_{T_0-1}
\]

provided \( 1 < T_0 < a - 1 \). For \( T_0 = 1 \), the first trial may lead to ruin, and (3.1) is replaced by

\[
q_1 = pq_2 + q.
\]

Similarly, for \( T_0 = a - 1 \) the first trial may result in victory, and therefore

\[
q_{a-1} = qq_{a-2}.
\]

To unify our equations, we define as a natural convention that \( q_0 = 1 \), and \( q_a = 0 \). Then the probability of ruin satisfies (3.1) for \( T_0 = 1, 2, \ldots, a - 1 \). This defines a set of \( a - 1 \) difference equations, with boundary conditions at 0 and \( a \). If we solve the system of difference equations, then we will have the desired probability \( q_{T_0} \) for any value of \( T_0 \).

Note that we can rewrite the difference equations as

\[
pq_{T_0} + qq_{T_0} = pq_{T_0+1} + qq_{T_0-1}.
\]

Then we can rearrange and factor to obtain

\[
\frac{q_{T_0+1} - q_{T_0}}{q_{T_0} - q_{T_0-1}} = \frac{q}{p}.
\]

This says the ratio of successive differences of \( q_{T_0} \) is constant. This suggests that \( q_{T_0} \) is a power function,

\[
q_{T_0} = \lambda^{T_0}
\]

since power functions have this property.
3.2. RUIN PROBABILITIES

We first take the case when \( p \neq q \). Then based on the guess above (or also on standard theory for linear difference equations), we try a solution of the form \( q_{T_0} = \lambda^{T_0} \). That is
\[
\lambda^{T_0} = p\lambda^{T_0+1} + q\lambda^{T_0-1}.
\]
This reduces to
\[
p\lambda^2 - \lambda + q = 0.
\]
Since \( p + q = 1 \), this factors as
\[(p\lambda - q)(\lambda - 1) = 0,
\]
so the solutions are \( \lambda = q/p \), and \( \lambda = 1 \). (One could also use the quadratic formula to obtain the same values.) Again by the standard theory of linear difference equations, the general solution is
\[
(3.2)
q_{T_0} = A \cdot 1 + B \cdot (q/p)^{T_0}
\]
for some constants \( A \), and \( B \).

Now we determine the constants by using the boundary conditions:
\[
q_0 = A + B = 1
\]
\[
q_a = A + B(q/p)^a = 0.
\]
Solving, substituting, and simplifying:
\[
q_{T_0} = \frac{(q/p)^a - (q/p)^{T_0}}{(q/p)^a - 1}.
\]
(Check for yourself that with this expression \( 0 \leq q_{T_0} \leq 1 \) as it should be for a probability.)

We should show that the solution is unique. So suppose \( r_{T_0} \) is another solution of the difference equations. Given the arbitrary solution of (3.1), the two constants \( A \) and \( B \) can be determined so that (3.2) agrees with \( r_{T_0} \) at \( T_0 = 0 \) and \( T_0 = a \). (The reader should be able to explain why by reference to a theorem in Linear Algebra.) From these two values, all other values for both solution forms can be found by substituting in (3.1) successively \( T_0 = 1, 2, 3, \ldots \). Therefore, two solutions that agree for \( T_0 = 0 \) and \( T_0 = a \) are identical, hence every solution is of the form (3.2).

The solution breaks down if \( p = q = 1/2 \), since then we do not get two linearly independent solutions of the difference equation, we get the solution 1 repeated twice. Instead, we need to borrow a result from differential equations (from the variation-of-parameters/reduction-of-order set of ideas used to derive a complete linearly independent set of solutions.) Certainly, 1 is still a solution of the difference equation (3.1). A second linearly independent solution is \( T_0 \), (check it out!) and the general solution is \( q_{T_0} = A + BT_0 \). To satisfy the boundary conditions, we put \( A = 1 \), and \( A + Ba = 0 \), hence \( q_{T_0} = 1 - T_0/a \).

We can consider a symmetric interpretation of this gambling game. Instead of a single gambler playing at a casino, trying to make a goal \( a \) before being ruined, consider two gamblers Alice and Bill playing against each other. Let Alice’s initial capital be \( T_0 \) and let her play against adversary Bill with initial capital \( a - T_0 \) so that their combined capital is \( a \). The game continues until one gambler’s capital either is reduced to zero or has increased to \( a \), that is, until one of the two players is ruined.
Corollary 3.8. \( p_{T_0} + q_{T_0} = 1 \)

Proof. The probability \( p_{T_0} \) of Alice’s winning the game equals the probability of Bill’s ruin. Bill’s ruin (and Alice’s victory) is therefore obtained from our ruin formulas on replacing \( p, q, \) and \( T_0 \) by \( q, p, \) and \( a - T_0 \) respectively. That is, from our formula (for \( p \neq q \)) the probability of Alice’s ruin is

\[
q_{T_0} = \frac{(q/p)^a - (q/p)^{T_0}}{(q/p)^a - 1}
\]

and the probability of Bill’s ruin is

\[
p_{T_0} = \frac{(p/q)^a - (p/q)^{a-T_0}}{(p/q)^a - 1}.
\]

Then add these together, and after some algebra, the total is 1. (Check it out!)

For \( p = 1/2 = q \), the proof is simpler, since then \( p_{T_0} = 1 - (a - T_0)/a \), and \( q_{T_0} = 1 - T_0/a \), and \( p_{T_0} + q_{T_0} = 1 \) easily. \( \square \)

Corollary 3.9. The expected gain is \( E[G] = (1 - q_{T_0})a - T_0 \).

Proof. In the game, the gambler’s ultimate gain (or loss) is a Bernoulli (two-valued) random variable, \( G \), where \( G \) assumes the value \(-T_0\) with probability \( q_{T_0} \), and assumes the value \( a - T_0 \) with probability \( p_{T_0} \). Thus the expected value is

\[
E[G] = (a - T_0)p_{T_0} + (-T_0)q_{T_0}
= p_{T_0}a - T_0
= (1 - q_{T_0})a - T_0.
\]

Now notice that if \( q = 1/2 = p \), so that we are dealing with a fair game, then \( E[G] = (1 - (1 - T_0/a)) \cdot a - T_0 = 0 \). That is, a fair game in the short run (one trial) is a fair game in the long run (expected value). However, if \( p < 1/2 < q \), so the game is not fair then our expectation formula says

\[
E[G] = \left(1 - \frac{(q/p)^a - (q/p)^{T_0}}{(q/p)^a - 1}\right) a - T_0
= \frac{(q/p)^{T_0} - 1}{(q/p)^a - 1} a - T_0
= \left(\frac{[(q/p)^{T_0} - 1]a}{[(q/p)^a - 1]T_0 - 1}\right) T_0.
\]

The sequence \([(q/p)^a - 1]/n\) is an increasing sequence, so

\[
\left(\frac{[(q/p)^{T_0} - 1]a}{[(q/p)^a - 1]T_0 - 1}\right) < 0.
\]

This shows that an unfair game in the short run (one trial) is an unfair game in the long run.

Corollary 3.10. The probability of ultimate ruin of a gambler with initial capital \( T_0 \) playing against an infinitely rich adversary is

\( q_{T_0} = 1, \quad p \leq q \)

and

\( q_{T_0} = (q/p)^{T_0}, \quad p > q \).
3.2. RUIN PROBABILITIES

<table>
<thead>
<tr>
<th>$p$</th>
<th>$q$</th>
<th>$T_0$</th>
<th>$a$</th>
<th>Probability of Ruin</th>
<th>Probability of Victory</th>
<th>Expected Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>9</td>
<td>10</td>
<td>0.1000</td>
<td>0.9000</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>90</td>
<td>100</td>
<td>0.1000</td>
<td>0.9000</td>
<td>0</td>
</tr>
<tr>
<td>0.49</td>
<td>0.51</td>
<td>9</td>
<td>10</td>
<td>0.1189</td>
<td>0.8811</td>
<td>-0.189</td>
</tr>
<tr>
<td>0.49</td>
<td>0.51</td>
<td>90</td>
<td>100</td>
<td>0.3359</td>
<td>0.6641</td>
<td>-23.59</td>
</tr>
<tr>
<td>0.45</td>
<td>0.55</td>
<td>9</td>
<td>10</td>
<td>0.2101</td>
<td>0.7899</td>
<td>-1.1</td>
</tr>
<tr>
<td>0.45</td>
<td>0.55</td>
<td>90</td>
<td>100</td>
<td>0.8656</td>
<td>0.1344</td>
<td>-76.56</td>
</tr>
<tr>
<td>0.4</td>
<td>0.6</td>
<td>9</td>
<td>10</td>
<td>0.3392</td>
<td>0.6608</td>
<td>-2.392</td>
</tr>
<tr>
<td>0.4</td>
<td>0.6</td>
<td>90</td>
<td>100</td>
<td>0.9827</td>
<td>0.0173</td>
<td>-88.27</td>
</tr>
</tbody>
</table>

**Table 1.** Probabilities of ruin and victory, and expected gains.

**Proof.** Let $a \to \infty$ in the formulas. (Check it out!)

**Remark 3.11.** This corollary says that the probability of “breaking the bank at Monte Carlo” as in the movies is zero, at least for the simple games we are considering.

**Why do we hear about people who actually win?** We occasionally hear from people who claim to consistently win at the casino. How can this be in the face of the theorems above?

A simple illustration makes clear how this is possible. Assume for convenience a gambler who repeatedly visits the casino, each time with a certain amount of capital. His goal is to make a 10% gain on his starting capital. That is, in units of his initial capital $T_0 = 10$, and $a = 11$. For simplicity and as a best case, assume that the casino is fair so that $p = 1/2 = q$. Then the probability of ruin in any one visit is:

$$q_{T_0} = 1 - \frac{10}{11} \approx 0.091.$$  
That is, if the working capital is much greater than the amount required for victory, then the probability of ruin is reasonably small.

Then the probability of an unbroken string of successes in 7 visits is:

$$(1 - 1/11)^7 \approx 0.513,$$

slightly better than half. This much success is reasonable, but the gambler might boast about his skill instead of crediting it to luck. Moreover, simple psychology suggests the gambler would also blame one failure on oversight, momentary distraction or it might even be forgotten.

**Another Interpretation as a Random Walk.** Another common interpretation of this probability game is to imagine it as a random walk. That is, we imagine an individual on a number line, starting at some position $T_0 > 0$. The person takes a step to the right to $T_0 + 1$ with probability $p$ and takes a step to the left to $T_0 - 1$ with probability $q$ and continues this random process. Then instead of the total fortune at any time, we consider the geometric position on the line at any time. Instead of reaching financial ruin or attaining a financial goal, we talk instead about reaching or passing a certain position. For example, Corollary 3 says that if $p \leq q$, then the probability of visiting the origin before going to infinity is 1. The two interpretations are equivalent and either can be used depending on which
is more useful. The problems below use the random walk interpretation, because they are more naturally posed in terms of reaching or passing certain points on the number line.

**The interpretation as Markov Processes and Martingales.** The fortune in the coin-tossing game is the first and simplest encounter with two of the most important ideas in modern probability theory.

We can interpret the fortune in our gambler’s coin-tossing game as a Markov process. That is, at successive times the process is in various states. In our case, the states are the values of the fortune. The probability of passing from one state at the current time $t$ to another state at time $t + 1$ is completely determined by the present state. That is, for our coin-tossing game

$$
\begin{align*}
P[T_{t+1} = x + 1 | T_t = x] &= p \\
P[T_{t+1} = x - 1 | T_t = x] &= q \\
P[T_{t+1} = y | T_t = x] &= 0 \text{ for all } y \neq x + 1, x - 1.
\end{align*}
$$

The most important property of a Markov process is that the probability of being in the next state is completely determined by the current state and not the history of how the process arrived at the current state. In that sense, we often say that a Markov process is memory-less.

We can also note the fair coin-tossing game with $p = 1/2 = q$ is a martingale. That is, the expected value of the process at the next step is the current value. Using expectation for estimation, the best estimate we have of the gambler’s fortune at the next step is the current fortune:

$$
E[T_{n+1} | T_n = x] = (x + 1)(1/2) + (x - 1)(1/2) = x.
$$

This characterizes a fair game, after the next step, one can neither expect to be richer or poorer. Note that the coin-tossing games with $p \neq q$ do not have this property.

In later sections we have more occasions to study the properties of martingales, and to a lesser degree Markov processes.

**Section Ending Answer.** The solution to $x_n = ax_{n-1}$ is $x_n = x_0a^n$, a geometric sequence. The initial value $x_0$ of the sequence must be known to completely specify the solution. This is the simple first-order version of the second-order difference equations considered in this section.

**Algorithms, Scripts, Simulations.**

*Algorithm.* The goal is to simulate the probability function for ruin with a given starting value. First set the probability $p$, number of Bernoulli trials $n$, and number of experiments $k$. Set the ruin and victory values $r$ and $v$, the boundaries for the random walk. For each starting value from ruin to victory, fill an $n \times k$ matrix with the Bernoulli random variables. For languages with multi-dimensional arrays each the data is kept in a three-dimensional array of size $n \times k \times (v - r + 1)$. Cumulatively sum the Bernoulli random variables to create the fortune or random walk. For each starting value, for each random walk or fortune path, find the step where ruin or victory is encountered. For each starting value, find the proportion of fortunes encountering ruin. Finally, find a least squares linear fit of the ruin probabilities as a function of the starting value.
1 p <- 0.5
2 n <- 150
3 k <- 60
4
5 victory <- 10
6 # top boundary for random walk
7 ruin <- -10
8 # bottom boundary for random walk
9 interval <- victory - ruin + 1
10
11 winLose <- 2 * (array( 0+(runif(n*k*interval) <= p), dim=c(n,k, interval))) - 1
12 # 0+ coerces Boolean to numeric
13 totals <- apply( winLose, 2:3, cumsum)
14 # the second argument '2:3' means column-wise in each panel
15 start <- outer( array(1, dim=c(n+1,k)), ruin:victory, "*")
16
17 paths <- array( 0 , dim=c(n+1, k, interval ) )
18 paths[2:(n+1), 1:k, 1:interval] <- totals
19 paths <- paths + start
20
21 hitVictory <- apply( paths , 2:3 , ( function (x) match ( victory ,x, nomatch =n +2) ));
22 hitRuin <- apply( paths , 2:3 , ( function (x) match (ruin, x, nomatch =n +2) ));
23 # the second argument '2:3' means column-wise in each panel
24 # If no ruin or victory on a walk, nomatch=n+2 sets the hitting
25 # time to be two more than the number of steps, one more than
26 # the column length. Without the nomatch option, get NA which
27 # works poorly with the comparison hitRuin < hitVictory next.
28
29 probRuinBeforeVictory <-
30 apply( (hitRuin < hitVictory), 2,
31 (function(x)length((which(x,arr.ind=FALSE)))) )/k
32
33 startValues <- (ruin:victory);
34 ruinFunction <- lm(probRuinBeforeVictory ~ startValues)
35 # lm is the R function for linear models, a more general view of
36 # least squares linear fitting for response ~ terms
37 cat(sprintf("Ruin function Intercept: %f \n", coefficients( ruinFunction)[1] ))
38 cat(sprintf("Ruin function Slope: %f \n", coefficients(ruinFunction) [2] ))
39
40 plot(startValues, probRuinBeforeVictory);
41 abline(ruinFunction)
42

Key Concepts.

(1) The probabilities and the interpretation of the “gambler’s ruin”.
Vocabulary.

(1) We call **victory** the outcome of reaching a fortune goal by chance before going broke.
(2) We call **ruin** the outcome of going broke by chance before reaching a fortune goal.

Problems.

**Exercise 3.12.** Consider the ruin probabilities \( q_{T_0} \) as a function of \( T_0 \). What is the domain of \( q_{T_0} \)? What is the range of \( q_{T_0} \)? Explain heuristically why \( q_{T_0} \) is decreasing as a function of \( T_0 \).

**Exercise 3.13.** Show that power functions have the property that the ratio of successive differences is constant.

**Exercise 3.14.** Show the sequence \([ (q/p)^n - 1 ] / n \) is an increasing sequence for \( 0 < p < 1/2 < q < 1 \).

**Exercise 3.15.** In a random walk starting at the origin find the probability that the point \( a > 0 \) will be reached before the point \(-b < 0\).

**Exercise 3.16.** A famous spy wants to ruin the casino at Monte Carlo by consistently betting 1 Euro on Red at the roulette wheel. A European roulette wheel has 18 red, 1 green and 18 black pockets on the wheel so the probability of the spy winning at one turn in this game is \( 18/37 \approx 0.48649 \). The spy is backed by the full financial might of his powerful government, and so can be considered to have unlimited funds. Approximately how much money should the casino have to start with so that the spy has only a “one-in-a-million” chance of ruining the casino?

**Exercise 3.17.** A gambler starts with $2 and wants to win $2 more to get to a total of $4 before being ruined by losing all his money. He plays a coin-flipping game, with a coin that changes with his fortune.

(1) If the gambler has $2 he plays with a coin that gives probability \( p = 1/2 \) of winning a dollar and probability \( q = 1/2 \) of losing a dollar.
(2) If the gambler has $3 he plays with a coin that gives probability \( p = 1/4 \) of winning a dollar and probability \( q = 3/4 \) of losing a dollar.
(3) If the gambler has $1 he plays with a coin that gives probability \( p = 3/4 \) of winning a dollar and probability \( q = 1/4 \) of losing a dollar.

Use “first step analysis” to write three equations in three unknowns (with two additional boundary conditions) that give the probability that the gambler will be ruined. Solve the equations to find the ruin probability.

**Exercise 3.18.** A gambler plays a coin flipping game in which the probability of winning on a flip is \( p = 0.4 \) and the probability of losing on a flip is \( q = 1 - p = 0.6 \). The gambler wants to reach the victory level of $16 before being ruined with a fortune of $0. The gambler starts with $8, bets $2 on each flip when the fortune is $6, $8, $10 and bets $4 when the fortune is $4 or $12 Compute the probability of ruin in this game.

**Exercise 3.19.** Prove: In a random walk starting at the origin the probability to reach the point \( a > 0 \) before returning to the origin equals \( p(1 - q_1) \).
Exercise 3.20. Prove: In a random walk starting at \( a > 0 \) the probability to reach the origin before returning to the starting point equals \( qq^{a-1} \).

Exercise 3.21. In the simple case \( p = 1/2 = q \), conclude from the preceding problem: In a random walk starting at the origin, the number of visits to the point \( a > 0 \) that take place before the first return to the origin has a geometric distribution with ratio \( 1 - qq^{a-1} \). (Why is the condition \( q \geq p \) necessary?)

Exercise 3.22. (1) Draw a sample path of a random walk (with \( p = 1/2 = q \)) starting from the origin where the walk visits the position 5 twice before returning to the origin.

(2) Using the results from the previous problems, it can be shown with careful but elementary reasoning that the number of times \( N \) that a random walk \( (p = 1/2 = q) \) reaches the value \( a \) a total of \( n \) times before returning to the origin is a geometric random variable with probability

\[
P[N = n] = \left( \frac{1}{2a} \right)^n \left( 1 - \frac{1}{2a} \right).
\]

Compute the expected number of visits \( E[N] \) to level \( a \).

(3) Compare the expected number of visits of a random walk (with \( p = 1/2 = q \)) to the value 1,000,000 before returning to the origin and to the level 10 before returning to the origin.

Exercise 3.23. Modify the ruin probability scripts to perform simulations of the ruin calculations in Table 1 and compare the results.

Exercise 3.24. Perform some simulations of the coin-flipping game, varying \( p \) and the start value. How does the value of \( p \) affect the experimental probability of victory and ruin?

Exercise 3.25. Modify the simulations by changing the value of \( p \) and comparing the experimental results for each starting value to the theoretical ruin function.

3.3. Duration of the Gambler’s Ruin

Section Starter Question. Consider a gambler who wins or loses a dollar on each turn of a fair game with probabilities \( p = 1/2 \) and \( q = 1/2 \) respectively. Let his initial capital be $10. The game continues until the gambler’s capital either is reduced to 0 or has increased to $20. What is the length of the shortest possible game the gambler could play? What are the chances of this shortest possible game? What is the length of the second shortest possible game? How would you find the probability of this second shortest possible game occurring?

Understanding a Stochastic Process. We start with a sequence of Bernoulli random variables, \( Y_1, Y_2, Y_3, \ldots \) where \( Y_i = +1 \) with probability \( p \) and \( Y_i = -1 \) with probability \( q \). We start with an initial value \( T_0 \) and set \( Y_0 = T_0 \) for convenience. We define the sequence of sums \( T_n = \sum_{i=0}^{n} Y_i \). We want to understand the stochastic process \( T_1, T_2, T_3, \ldots \). It turns out this is a complicated sequence to understand in full, so we single out particular simpler features to understand first. For example, we can look at how many trials the process will experience until it achieves the value 0 or \( a \). In symbols, consider \( N = \min \{ n : T_n = 0, \text{ or } T_n = a \} \). It is possible to consider the probability distribution of this newly defined random variable. Even this turns out to be complicated, so we look at the expected value of
the number of trials, \( D = E[N] \). This is a special case of a larger class of probability problems called \textit{first passage distributions} for \textit{first passage times}.

The principle of first step analysis, also known as conditional expectations, provides equations for important properties of coin-flipping games and random walks. The important properties include ruin probabilities and the duration of the game until ruin. Difference equations derived from first step analysis or conditional expectations provide the way to deduce the expected length of the game in the gambler’s ruin, just as for the probability of ruin or victory. **Expectation by conditioning** is the process of deriving a difference equation for the expectation by conditioning the outcome over an exhaustive, mutually exclusive set of events, each of which leads to a simpler probability calculation, then weighting by the probability of each outcome of the conditioning events. **First step analysis** refers to the simple expectation by conditioning process that we use to analyze the ruin probabilities and expected duration. It is a more specific description for coin-tossing games of the more general technique of expectation by conditioning.

**Expected length of the game.** Note that in the following we implicitly assume that the expected duration of the game is finite. This fact is true, see below for a proof.

**Theorem 3.26.** The expected duration of the game in the classical ruin problem is
\[
D_{T_0} = \frac{T_0}{q - p} - \frac{a}{q - p} \frac{1 - (q/p)^{T_0}}{1 - (q/p)^a} \quad \text{for} \quad p \neq q
\]

and
\[
T_0(a - T_0) \quad \text{for} \quad p = 1/2 = q.
\]

**Proof.** If the first trial results in success, the game continues as if the initial position had been \( T_0 + 1 \). The conditional expectation of the duration conditioned on success at the first trial is therefore \( D_{T_0+1} + 1 \). Likewise if the first trial results in a loss, the duration conditioned on the loss at the first trial is \( D_{T_0-1} + 1 \).

This argument shows that the expected duration satisfies the difference equation, obtained by expectation by conditioning
\[
D_{T_0} = pD_{T_0+1} + qD_{T_0-1} + 1
\]
with the boundary conditions
\[
D_0 = 0, \quad D_a = 0.
\]

The appearance of the term 1 makes the difference equation non-homogeneous. Taking a cue from linear algebra, or more specifically the theory of linear non-homogeneous differential equations, we need to find the general solution to the homogeneous equation
\[
D^h_{T_0} = pD^h_{T_0+1} + qD^h_{T_0-1}
\]
and a particular solution to the non-homogeneous equation. We already know the general solution to the homogeneous equation is \( D^h_{T_0} = A + B(q/p)^{T_0} \). The best way to find the particular solution is inspired guessing, based on good experience. We can re-write the non-homogeneous equation for the particular solution as
\[
-1 = pD_{T_0+1} - D_{T_0} + qD_{T_0-1}.
\]

The right side is a weighted second difference, a difference equation analog of the second derivative. Functions whose second derivative is a constant are quadratic
functions. Therefore, it make sense to try a function of the form $D^p_{T_0} = k + lT_0 + mT^2_0$. In the exercises, we show that the particular solution is actually $D_{T_0} = T_0/(q - p)$ if $p \neq q$.

It follows that the general solution of the duration equation is:

$$D_{T_0} = T_0/(q - p) + A + B(q/p)T_0.$$

The boundary conditions require that

$$A + B = 0$$

and

$$A + B(q/p)^a + a/(q - p) = 0.$$ 

Solving for $A$ and $B$, we find

$$D_{T_0} = \frac{T_0}{q - p} - \frac{a}{q - p} \frac{1 - (q/p)T_0}{1 - (q/p)^a}.$$ 

The calculations are not valid if $p = 1/2 = q$. In this case, the particular solution $T_0/(q - p)$ no longer makes sense for the equation

$$D_{T_0} = \frac{1}{2}D_{T_0+1} + \frac{1}{2}D_{T_0-1} + 1.$$ 

The reasoning about the particular solution remains the same however, and we can show that the particular solution is $-T_0^2$. It follows that the general solution is of the form $D_{T_0} = -T_0^2 + A + BT_0$. The required solution satisfying the boundary conditions is

$$D_{T_0} = T_0(a - T_0).$$ 

\[\square\]

**Corollary 3.27.** Playing until ruin with no upper goal for victory against an infinitely rich adversary, the expected duration of the game until ruin is

$$T_0/(q - p) \quad \text{for} \quad p \neq q$$

and

$$\infty \quad \text{for} \quad p = 1/2 = q.$$ 

**Proof.** Pass to the limit $a \to \infty$ in the preceding formulas. \[\square\]

**Example 3.28.** The duration can be considerably longer than we expect. For instance in a fair game with two players, each with $500, flipping a coin until one is ruined, the average duration of the game is 250,000 trials. If a gambler has only $1 and his adversary $1000, with a fair coin toss, the average duration of the game is 999 trials, although some games will be quite short! Very long games can occur with sufficient probability to give a long average.
### Proof that the duration is finite.

When we check the arguments for the probability of ruin or the duration of the game, we find a logical gap. We have assumed that the duration $D_{T_0}$ of the game is finite. How do we know for sure that the gambler’s net winnings will eventually reach $a$ or $0$ in a finite time? This important fact requires proof.

The proof uses a common argument in probability, an *extreme case argument*. We identify an extreme event with a small but positive probability of occurring. We are interested in the complementary “good” event which at least avoids the extreme event. Therefore the complementary event must happen with probability not quite 1. The avoidance must happen infinitely many independent times, but the probability of such a run of good events must go to zero.

For the gambler’s ruin, we are interested in the event of the game continuing forever. Consider the extreme event that the gambler wins $a$ times in a row. If the gambler is not already ruined (at 0), then such a streak of $a$ wins in a row is guaranteed to boost his fortune above $a$ and end the game in victory for the gambler. Such a run of luck is unlikely, but it has positive probability, in fact, probability $P = p^a$. We let $E_k$ denote the event that the gambler wins on each turn in the time interval $[ka, (k+1)a-1]$, so the $E_k$ are independent events. Hence the complementary events $E^C_k = \Omega - E_k$ are also independent. Then $D_{T_0} > na$ at least implies that all of the $E_k$ with $0 \leq k \leq n$ fail to occur. Thus, we find

$$
\mathbb{P}[D_{T_0} > na] \leq \mathbb{P}\left[\bigcap_{k=0}^{n} E^C_k\right] = (1 - P)^n.
$$

Note that

$$
\mathbb{P}[D_{T_0} = \infty | T_0 = z] \leq \mathbb{P}[D_{T_0} > na | T_0 = z]
$$

for all $n$. Hence, $\mathbb{P}[D_{T_0} = \infty] = 0$, justifying our earlier assumption.

### Section Ending Question.

The shortest possible game would be 10 successive tosses of heads or tails, occurring with probability $2 \cdot (\frac{1}{2})^{10}$. The second shortest game would be 12 tosses long with 10 heads or tails and two of the opposite occurring before the last flip. Calculating this probability uses a binomial probability for 11 trials multiplied by the probability of the concluding flip. This calculation shows that direct counting of the sequences leading to ruin or victory gets complicated. The first step analysis is a better way to calculate the expected duration.
Algorithms, Scripts, Simulations.

Algorithm. The goal is to simulate the duration until ruin or victory as a function of starting value. First set the probability \( p \), number of Bernoulli trials \( n \), and number of experiments \( k \). Set the ruin and victory values \( r \) and \( v \), also interpreted as the boundaries for the random walk. For each starting value from ruin to victory, fill an \( n \times k \) matrix with the Bernoulli random variables. Languages with multi-dimensional arrays keep the data in a three-dimensional array of size \( n \times k \times (v − r + 1) \). Cumulatively sum the Bernoulli random variables to create the fortune or random walk. For each starting value, for each random walk or fortune path, find the duration until ruin or victory. For each starting value, find the mean of the duration until ruin or victory. Finally, find a least squares polynomial fit for the duration as a function of the starting value.

```r
1 p <- 0.5
2 n <- 300
3 k <- 200
4
5 victory <- 10
6 # top boundary for random walk
7 ruin <- 0
8 # bottom boundary for random walk
9 interval <- victory - ruin + 1
10
11 winLose <- 2 * (array( 0+(runif(n*k*interval) <= p), dim=c(n,k,
12 interval))) - 1
13 # 0+ coerces Boolean to numeric
14 totals <- apply( winLose, 2:3, cumsum)
15 # the second argument "2:3" means column-wise in each panel
16 start <- outer( array(1, dim=c(n+1,k)), ruin:victory, "*")
17
18 paths <- array( 0 , dim=c(n+1, k, interval ) )
19 paths[2:(n+1), 1:k, 1:interval] <- totals
20 paths <- paths + start
21
22 hitVictory <- apply(paths, 2:3, (function(x) match(victory,x, nomatch=n +2) ));
23 hitRuin <- apply(paths, 2:3, (function(x) match(ruin, x, nomatch=n +2)));
24 # the second argument "2:3" means column-wise in each panel
25 # If no ruin or victory on a walk, nomatch=n+2 sets the hitting time to be two more than the number of steps, one more than the column length. Without the nomatch option, get NA which works poorly with the comparison hitRuin < hitVictory next.
26
27 duration <- pmin(hitVictory, hitRuin) - 1
28 # Subtract 1 since R arrays are 1-based, so duration is 1 less than index
29 is.na(duration) = duration > n
30 # Remove durations greater than length of trials
31 meanDuration = colMeans( duration, na.rm=TRUE)
32
33 startValues <- (ruin:victory);
```
durationFunction <- lm( meanDuration ~ poly(startValues,2,raw=TRUE) )

# lm is the R function for linear models, a more general view of
# least squares linear fitting for response ~ terms

plot(startValues, meanDuration, col = "blue");
lines(startValues, predict(durationFunction, data=startValues), col = "red")

cat(sprintf("Duration function is: %f + %f x + %f x^2 
", coefficients(durationFunction)[1], coefficients(durationFunction)[2], coefficients(durationFunction)[3]))

Key Concepts.
(1) The principle of first step analysis, also known as conditional expectations, provides equations for important properties of coin-flipping games and random walks. The important properties include ruin probabilities and the duration of the game until ruin.
(2) Difference equations derived from first step analysis or conditional expectations provide the way to deduce the expected length of the game in the gambler’s ruin, just as for the probability of ruin or victory.

Vocabulary.
(1) Expectation by conditioning is the process of deriving a difference equation for the expectation by conditioning the outcome over an exhaustive, mutually exclusive set of events, each of which leads to a simpler probability calculation, then weighting by the probability of each outcome of the conditioning events.
(2) First step analysis refers to the simple expectation by conditioning process that we use to analyze the ruin probabilities and expected duration. It is a more specific description for coin-tossing games of the more general technique of expectation by conditioning.

Problems.
Exercise 3.29. (1) Using a trial function of the form $D_{T_0}^p = k + lT_0 + mT_0^2$, show that a particular solution of the non-homogeneous equation

$$D_{T_0} = pD_{T_0+1} + qD_{T_0-1} + 1$$

is $T_0/(q - p)$.
(2) Using a trial function of the form $D_{T_0}^p = k + lT_0 + mT_0^2$, show that a particular solution of the non-homogeneous equation

$$D_{T_0} = \frac{1}{2}D_{T_0+1} + \frac{1}{2}D_{T_0-1} + 1$$

is $-T_0^2$.

Exercise 3.30. A gambler starts with $2 and wants to win $2 more to get to a total of $4 before being ruined by losing all his money. He plays a coin-flipping game, with a coin that changes with his fortune.
(1) If the gambler has $2 he plays with a coin that gives probability \( p = \frac{1}{2} \) of winning a dollar and probability \( q = \frac{1}{2} \) of losing a dollar.

(2) If the gambler has $3 he plays with a coin that gives probability \( p = \frac{1}{4} \) of winning a dollar and probability \( q = \frac{3}{4} \) of losing a dollar.

(3) If the gambler has $1 he plays with a coin that gives probability \( p = \frac{3}{4} \) of winning a dollar and probability \( q = \frac{1}{4} \) of losing a dollar.

Use “first step analysis” to write three equations in three unknowns (with two additional boundary conditions) that give the expected duration of the game that the gambler plays. Solve the equations to find the expected duration.

Exercise 3.31. A gambler plays a coin flipping game in which the probability of winning on a flip is \( p = 0.4 \) and the probability of losing on a flip is \( q = 1 - p = 0.6 \). The gambler wants to reach the victory level of $16 before being ruined with a fortune of $0. The gambler starts with $8, bets $2 on each flip when the fortune is $6, $8, $10 and bets $4 when the fortune is $4 or $12. Compute the duration of this game.

Exercise 3.32. Several North American professional sports leagues have a “best-of-seven” format for their seasonal championship (the World Series for Major League Baseball, the NBA Finals for the National Basketball Association and the Stanley Cup Finals for the National Hockey League.) A best-of-seven playoff is a sequence of games between two teams in which one team must win four games to win the series. If one team has won four games before all seven games have been played, remaining games are omitted.

(1) Explain why or why not the first step analysis for the expected duration model for victory-or-ruin is sensible for estimating the expected length of a best-of-seven championship series.

(2) How many games would we expect to be needed to complete a best-of-seven series if each team has a 50 percent chance of winning each individual game? What modeling assumptions are you making?

(3) Using the same assumptions, how many games are expected to complete a best-of-seven series if one team has a 60 percent chance of winning each individual game? A 70 percent chance?

(4) Using the same assumptions, make a graph of the expected number of games as a function of \( p \), the probability of one team winning an individual game.

Exercise 3.33. Perform some simulations of the coin-flipping game, varying \( p \) and the start value. How does the value of \( p \) affect the experimental duration of victory and ruin?

Exercise 3.34. Modify the simulations by changing the value of \( p \) and comparing the experimental results for each starting value to the theoretical duration function.

Exercise 3.35. Modify the duration scripts to perform simulations of the duration calculations in the table in this section.

3.4. A Stochastic Process Model of Cash Management

Section Starter Question. Suppose that you have a stock of 5 units of a product. It costs you \( r \) dollars per unit of product to hold the product for a week.
You sell and eliminate from the stock one unit of product per week. What is the total cost of holding the product? Now suppose that the amount of product sold is 0 or 1 each week, determined by a coin-tossing game. How would you calculate the expected cost of holding the product?

**Background.** The **reserve requirement** is a bank regulation that sets the minimum reserves of cash a bank must hold on hand for customer withdrawals. This is also called the **Federal Reserve requirement** or the **reserve ratio**. These reserves exist so banks can satisfy cash withdrawal demands. The reserves also help regulate the national money supply. Specifically in 2016 the Federal Reserve regulations require that the first $15.2 million in deposits are exempt from reserve requirements. A 3 percent reserve ratio applies to net transaction accounts (a net total of certain account types defined by the Federal Reserve) over $15.2 million up to and including $110.2 million. A 10 percent reserve ratio applies to net transaction accounts in excess of $110.2 million. The Federal Reserve adjusts these amounts from time to time.

Of course, bank customers are frequently depositing and withdrawing money so the amount of money for the reserve requirement is constantly changing. If customers deposit more money, the cash on hand exceeds the reserve requirement. The bank would like to put the excess cash to work, perhaps by buying Treasury bills. If customers withdraw cash, the available cash can fall below the required amount to cover the reserve requirement so the bank gets more cash, perhaps by selling Treasury bills.

The bank has a dilemma: buying and selling the Treasury bills has a transaction cost, so the bank does not want to buy and sell too often. On the other hand, the bank could put excess cash to use by loaning it out, and so the bank does not want to have too much cash idle. What is the optimal level of cash that signals a time to buy or sell, and how much should the bank buy or sell?

**Modeling.** We assume for a simple model that a bank’s cash level fluctuates randomly as a result of many small deposits and withdrawals. We model this by dividing time into successive, equal length periods, each of short duration. The periods might be weekly, the reporting period the Federal Reserve Bank requires for some banks. In each time period, assume the reserve randomly increases or decreases one unit of cash, perhaps measured in units of $100,000, each with probability $1/2$. That is, in period $n$, the change in the bank’s reserves is

$$Y_n = \begin{cases} +1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2. \end{cases}$$

The equal probability assumption simplifies calculations for this model. It is possible to relax the assumption to the case where the probability of change in reserves is not $1/2$ but we will not do this here.

Let $T_0 = s$ be the initial cash on hand. Then $T_n = T_0 + \sum_{j=1}^{n} Y_j$ is the total cash on hand at period $n$.

The bank will intervene if the reserve gets too small or too large. Again for simple modeling, if the reserve level drops to zero, the bank sells assets such as Treasury bonds to replenish the reserve back up to $s$. If the cash level ever increases to $S$, the bank buys Treasury bonds to reduce the reserves to $s$. What we have modeled here is a version of the Gambler’s Ruin, except that when this game reaches
the ruin or victory boundaries, 0 or $S$ respectively, the game immediately restarts again at $s$.

Now the cash level fluctuates in a sequence of cycles or games. Each cycle begins with $s$ units of cash on hand and ends with either a replenishment of cash, or a reduction of cash. See Figure 1 for a simple example with $s = 2$ and $S = 5$.

**Mean number of visits to a particular state.** Now let $k$ be one of the possible reserve states with $0 < k < S$ and let $W_{sk}$ be the expected number of visits to the level $k$ up to the ending time of the cycle starting from $s$. A formal mathematical expression for this expression is

$$W_{sk} = \mathbb{E} \left[ \sum_{j=1}^{N-1} 1_{\{T_j = k\}} \right]$$

where $1_{\{T_j = k\}}$ is the indicator random variable where

$$1_{\{T_j = k\}} = \begin{cases} 1 & T_j = k \\ 0 & T_j \neq k. \end{cases}$$

Note that the inner sum is a random sum, since it depends on the length of the cycle $N$, which is cycle dependent.

Then using first step analysis $W_{sk}$ satisfies the equations

$$W_{sk} = \delta_{sk} + \frac{1}{2} W_{s-1,k} + \frac{1}{2} W_{s+1,k}$$

with boundary conditions $W_{0k} = W_{Sk} = 0$. The term $\delta_{sk}$ is the Kronecker delta

$$\delta_{sk} = \begin{cases} 1 & \text{if } k = s \\ 0 & \text{if } k \neq s. \end{cases}$$

The explanation of this equation is very similar to the derivation of the equation for the expected duration of the coin-tossing game. The terms $\frac{1}{2} W_{s-1,k} + \frac{1}{2} W_{s+1,k}$ arise from the standard first step analysis or expectation-by-conditioning argument for $W_{sk}$. The non-homogeneous term in the prior expected duration equation (which is +1) arises because the game will always be at least 1 step longer after the first step. In the current equation, the $\delta_{sk}$ non-homogeneous term arises because the
number of visits to level $k$ after the first step will be 1 more if $k = s$ but the number of visits to level $k$ after the first step will be 0 more if $k \neq s$.

For the ruin probabilities, the difference equation was homogeneous, and we only needed to find the general solution. For the expected duration, the difference equation was non-homogeneous with a non-homogeneous term which was the constant 1, making the particular solution reasonably easy to find. Now the non-homogeneous term depends on the independent variable, so solving for the particular solution will be more involved.

First we find the general solution $W_{sk}^h$ to the homogeneous linear difference equation

$$W_{sk}^h = \frac{1}{2} W_{s-1,k}^h + \frac{1}{2} W_{s+1,k}^h.$$  

This is easy, we already know that it is $W_{sk}^h = A + Bs$.

Then we must find a particular solution $W_{sk}^p$ to the non-homogeneous equation

$$W_{sk}^p = \delta_{sk} + \frac{1}{2} W_{s-1,k}^p + \frac{1}{2} W_{s+1,k}^p.$$  

For purposes of guessing a plausible particular solution, temporarily re-write the equation as

$$-2\delta_{sk} = W_{s-1,k}^p - 2W_{sk}^p + W_{s+1,k}^p.$$  

The expression on the right is a centered second difference. For the prior expected duration equation, we looked for a particular solution with a constant centered second difference. Based on our experience with functions it made sense to guess a particular solution of the form $C + Ds + Es^2$ and then substitute to find the coefficients. Here we seek a function whose centered second difference is 0 except at $k$ where the second difference jumps to 1. This suggests the particular solution is piecewise linear, say

$$W_{sk}^p = \begin{cases} C + Ds & \text{if } s \leq k \\ E + Fs & \text{if } s > k. \end{cases}$$

In the exercises, we verify that the coefficients of the function are

$$W_{sk}^p = \begin{cases} 0 & \text{if } s \leq k \\ 2(k - s) & \text{if } s > k. \end{cases}$$

We can write this as $W_{sk}^p = -2\max(s - k, 0)$

Then solving for the boundary conditions, the full solution is

$$W_{sk} = 2 \left[ s(1 - k/S) - \max(s - k, 0) \right].$$

**Expected Duration and Expected Total Cash in a Cycle.** Consider the **first passage time** $N$ when the reserves first reach 0 or $S$, so that cycle ends and the bank intervenes to change the cash reserves. The value of $N$ is a random variable, it depends on the sample path. We are first interested in $D_s = \mathbb{E}[N]$, the expected duration of a cycle. From the previous section we already know $D_s = s(S - s)$.

Next, we need the mean cost of holding cash on hand during a cycle $i$, starting from amount $s$. Call this mean $W_s$. Let $r$ be the cost per unit of cash, per unit of time. We then obtain the cost by weighting $W_{sk}$, the mean number of times the
3.4. A STOCHASTIC PROCESS MODEL OF CASH MANAGEMENT

Cash is at number of units $k$ starting from $s$, multiplying by $k$, multiplying by the factor $r$ and summing over all the available amounts of cash:

$$W_s = \sum_{k=1}^{S-1} rkW_{sk}$$

$$= 2 \left[ \frac{s}{S} \sum_{k=1}^{S-1} rk(S-k) - \sum_{k=1}^{S-1} rk(s-k) \right]$$

$$= 2 \left[ \frac{s}{S} \left( \frac{rS(S-1)(S+1)}{6} \right) - \frac{rs(s-1)(s+1)}{6} \right]$$

$$= rs \left[ S^2 - s^2 \right].$$

These results are interesting and useful in their own right as estimates of the length of a cycle and the expected cost of cash on hand during a cycle. Now we use these results to evaluate the long run behavior of the cycles. Upon resetting the cash at hand to $s$ when the amount of cash reaches 0 or $S$, the cycles are independent of each of the other cycles because of the assumption of independence of each step. Let $K$ be the fixed cost of the buying or selling of the treasury bonds to start the cycle, let $N_i$ be the random length of the cycle $i$, and let $R_i$ be the total cost of holding cash on hand during cycle $i$. Then the cost over $n$ cycles is $nK + R_1 + \cdots + R_n$. Divide by $n$ to find the average cost

Expected total cost in cycle $i = K + \mathbb{E}[R_i]$,

but we have another expression for the expectation $\mathbb{E}[R_i]$,

Expected cost $= \mathbb{E}[R_i] = rs \left[ S^2 - s^2 \right].$

Likewise the total length of $n$ cycles is $N_1 + \cdots + N_n$. Divide by $n$ to find the average length,

Expected length $= \frac{N_1 + \cdots + N_n}{n} = s(S-s)$.

These expected values allow us to calculate the average costs

Long run average cost, dollars per week $= \frac{K + \mathbb{E}[R_i]}{\mathbb{E}[N_i]}$.

Then $\mathbb{E}[R_i] = W_s$ and $\mathbb{E}[N_i] = s(S-s)$. Therefore

Long run average cost, dollars per week $= \frac{K + (1/3)rs(S^2 - s^2)}{s(S-s)}$.

Simplify the analysis by setting $x = s/S$ so that the expression of interest is

Long run average cost $= \frac{K + (1/3)rs^3x(1-x^2)}{S^2x(1-x)}$.

Remark 3.36. Aside from being a good thing to non-dimensionalize the model as much as possible, it also appears that optimizing the original long run cost average in the original variables $S$ and $s$ is messy and difficult. This of course would not be known until you had tried it. However, knowing the optimization is difficult in variables $s$ and $S$ additionally motivates making the transformation to the non-dimensional ratio $x = s/S$. 
Now we have a function in two variables that we wish to optimize. Take the partial derivatives with respect to \( x \) and \( S \) and set them equal to 0, then solve, to find the critical points.

The results are that

\[
x_{\text{opt}} = \frac{1}{3}
\]

\[
S_{\text{opt}} = 3 \left( \frac{3K}{4r} \right)^{\frac{1}{3}}
\]

That is, the optimal value of the maximum amount of cash to keep varies as the cube root of the cost ratios, and the reset amount of cash is 1/3 of that amount.

**Criticism of the model.** The first test of the model would be to look at the amounts \( S \) and \( s \) for well-managed banks and determine if the banks are using optimal values. That is, one could do a statistical survey of well-managed banks and determine if the values of \( S \) vary as the cube root of the cost ratio, and if the restart value is 1/3 of that amount. Of course, this assumes that the model is valid and that banks are following the predictions of the model, either consciously or not.

This model is too simple and we could modify in a number of ways. One change might be to change the reserve requirements to vary with the level of deposits, just as the 2010 Federal Reserve requirements vary. Adding additional reserve requirement levels to the current model adds a level of complexity, but does not substantially change the level of mathematics involved.

The most important change would be to allow the changes in deposits to have a continuous distribution instead of jumping up or down by one unit in each time interval. Modification to continuous time would make the model more realistic instead of changing the cash at discrete time intervals. The assumption of statistical independence from time step to time step is questionable, and so could also be relaxed. All these changes require deeper analysis and more sophisticated stochastic processes.

**Section Ending Answer.** The cost to hold the products would be the number of units present each week times the holding rate \( r \), so \( (5 + 4 + 3 + 2 + 1)r = 15r \). If the number sold in week \( i \) is randomly \( X_i = 0 \) or \( X_i = 1 \), the number of units present in week \( n \) is \( 5 - \sum_{i=1}^{n} X_i \) and the cost of holding the units is \( (5 - \sum_{i=1}^{n} X_i) r \).

Over 5 weeks, the cost would be \( 25 - \sum_{n=1}^{5} \sum_{i=1}^{n} X_i \). The expected cost would be \( (25 - \sum_{n=1}^{5} n/2)r = (25 - 15/2)r = 17.5r \). Both of these are simplified versions of the cost of holding money over a cash cycle considered above.

**Algorithms, Scripts, Simulations.**

**Algorithm.** Set the top boundary state value and the start and reset state value. Set the probability of a transition of an interior state to the next higher state. Set the number of steps in the Markov chain. Create the Markov chain transition probability matrix. For this Markov chain, the transition probability matrix is tridiagonal, with 0 on the main diagonal, \( p \) on the upper diagonal and \( 1 - p \) on the lower diagonal. For the boundary states, the transition probability is 1 to the start or reset state.

Initialize the vector holding the number of visits to each state, the number of cycles from start to reset, and the length of each cycle. For each transition,
choose a random value $u$ chosen from a uniform distribution on $(0, 1)$. Starting
from an initialized cumulative probability of 0, compute the cumulative probability
of the current state’s transition probability distribution. Compare the cumulative
probability to $u$ until the cumulative probability exceeds $u$. This is the new state.
Update the number of visits to each state, the number of cycles from start to reset,
and the length of each cycle and at the end of a cycle compute the average number
of visit and the average length of the cycle. Print the average number of visits, and
compare to the theoretical number of visits. For cash management, compute the
actual cash management cost and the theoretical cash management value.

```r
n <- 10  # Top boundary, number of states 0..n is n+1
s <- 5   # Start and Reset state number 1 <= s <= n-1
p <- 1/2
steps <- 1000

# vector to hold the count of visits to each state during a cycle
count <- mat.or.vec(1, n + 1)

dataFrame <- as.data.frame(count)

dataFrame$Header <- c("state", "count")
dataFrame$state <- as.factor(dataFrame$state)
dataFrame$count <- as.numeric(dataFrame$count)
dataFrame$mean <- NULL
dataFrame$mean[1:n] <- count[1:n] / steps

dataFrame$Upper <- c("upper", "upper")
dataFrame$Upper[1:n] <- 1
dataFrame$Upper[1] <- 0

dataFrame$Lower <- c("lower", "lower")
dataFrame$Lower[1:n] <- 1

dataFrame$Difference <- NULL
dataFrame$Period <- NULL
dataFrame$Period[1:n] <- 1

dataFrame$Value <- NULL

dataFrame$Cost <- NULL

dataFrame$Cost <- dataFrame$Cost[1:n]
dataFrame$Cost <- dataFrame$Cost %*% c(1, 0)
dataFrame$Cost <- dataFrame$Cost[1]
dataFrame$Cost <- dataFrame$Cost %*% c(1, 0)
dataFrame$Cost <- dataFrame$Cost[1]

dataFrame$Value <- dataFrame$Value[1:n]
dataFrame$Value <- dataFrame$Value %*% c(1, 0)
dataFrame$Value <- dataFrame$Value[1]
dataFrame$Value <- dataFrame$Value %*% c(1, 0)
dataFrame$Value <- dataFrame$Value[1]
```

```r
1 n <- 10  # Top boundary, number of states 0..n is n+1
2 s <- 5   # Start and Reset state number 1 <= s <= n-1
3 p <- 1/2
4 steps <- 1000
5
6 diag.num <- outer(seq(1,n+1),seq(1,n+1), "+=")
7 # diag.num is a matrix whose ith lower diagonal equals i, opposite for
8 # upper diagonal
9 T <- mat.or.vec(n+1,n+1)
10 # mat.or.vec creates an nxn zero matrix if nc is greater than 1
11 # Also remember that matrices in R are 1-based, so need n+1 states,
12 n
13 T[diag.num == -1] <- p
14 T[diag.num == 1] <- 1-p
15 T[1,2] <- 0; T[1,s+1] <- 1;
16 T[n+1,n] <- 0; T[n+1,s+1] <- 1;
17
18 # vector to hold the count of visits to each state during a cycle
19 count <- mat.or.vec(1, n + 1)
20 # Initialize the number of cycles
21 numberCycles <- 0
22 # Initialize the length of the cycle
23 cycleLength <- 0
24
25 # Start in the state s
26 state = s+1
27
28 # Make steps through the markov chain
29 for (i in 1:steps)
30 {
31     x = 0;
32     u = runif(1, 0, 1);
33     for (j in 1:ncol(T))
34     {
35         x = x + T[state, j];
36         if (x >= u)
37         {
38             newState = j;
39             break;
40         }
41     }
42 }
```
3. FIRST STEP ANALYSIS

```r
## newState <- sample(1:ncol(T), 1, prob=T[state,])
state = newState;
count[state] <- count[state] + 1
cycleLength <- cycleLength + 1
if (state == n+1 || state == 1) {
  numberCycles <- numberCycles + 1
  avgVisits <- count/numberCycles
  avgCycleLength <- i/numberCycles
}
Wsk <- avgVisits
theoreticalWsk <- 2*(s*(1-(0:n)/n) - pmax(s - (0:n),0));
cat(sprintf("Average number of visits to each state in a cycle: \n "))
cat(Wsk)
cat("\n\n")
cat(sprintf("Theoretical number of visits to each state in a cycle: \n "))
cat(theoreticalWsk)
cat("\n")
```

```
1 n <- 10  # Top boundary, number of states 0..n is n+1
2 s <- 5   # Start and Reset state number 1 <= s <= n-1
3 p <- 1/2
4 steps <- 1000
5 rate <- 1.0
6 K <- 2   # Charge or cost to reset the Markov chain
7
diag.num <- outer(seq(1,n+1),seq(1,n+1), "-")
# diag.num is a matrix whose ith lower diagonal equals i, opposite for upper diagonal
8 T <- mat.or.vec(n+1,n+1)
# mat.or.vec creates an nr by nc zero matrix if nc is greater than 1
# Also remember that matrices in R are 1-based, so need n+1 states, 0..n
9 T[diag.num == -1] <- p
10 T[diag.num == 1] <- 1-p
11 T[1,2] <- 0; T[1,s+1] <- 1;
12 T[n+1,n] <- 0; T[n+1,s+1] <- 1;
13
count <- mat.or.vec(1, n+1)
14 numberCycles <- 0
15 totalCost <- 0
16 state = s+1
17
data <- sample(1:n, 1, prob=diag.num)
```

18 # vector to hold the count of visits to each state during a cycle
```
Key Concepts.

(1) The **reserve requirement** is a bank regulation that sets the minimum reserves of cash a bank must hold on hand for customer withdrawals. An important question for the bank is: What is the optimal level of cash for the bank to hold?

(2) We model the cash level with a sequence of cycles or games. Each cycle begins with $s$ units of cash on hand and ends with either a replenishment of cash, or a reduction of cash. In between these levels, the cash level is a stochastic process, specifically a coin-tossing game or random walk.

(3) By solving a non-homogeneous difference equation we can determine the expected number of visits to an intermediate level in the random process.
(4) Using the expected number of visits to a level we can model the expected costs of the reserve requirement as a function of the maximum amount to hold and the starting level after a buy or sell. Then we can minimize the expected costs with calculus to find the optimal values of the maximum amount and the starting level.

**Vocabulary.**

(1) The **reserve requirement** is a bank regulation that sets the minimum reserves of cash a bank must hold for customer withdrawals.

(2) The mathematical expression $\delta_{sk}$ is the **Kronecker delta**

$$\delta_{sk} = \begin{cases} 1 & \text{if } k = s \\ 0 & \text{if } k \neq s \end{cases}.$$

(3) If $X$ is a random variable assuming some values including $k$, the **indicator random variable** is

$$1_{\{X=k\}} = \begin{cases} 1 & X = k \\ 0 & X \neq k. \end{cases}$$

The indicator random variable indicates whether a random variable assumes a value, or is in a set. The expected value of the indicator random variable is the probability of the event.

**Problems.**

**Exercise 3.37.** Find a particular solution $W^p_{sk}$ to the non-homogeneous equation

$$W^p_{sk} = \delta_{sk} + \frac{1}{2} W^p_{s-1,k} + \frac{1}{2} W^p_{s+1,k^*},$$

using the trial function

$$W^p_{sk} = \begin{cases} C + Ds & \text{if } s \leq k \\ E + Fs & \text{if } s > k. \end{cases}$$

**Exercise 3.38.** Show that

$$W_s = \sum_{k=1}^{S-1} kW_{sk}$$

$$= 2 \left[ \frac{s}{3} \sum_{k=1}^{S-1} k(S-k) - \sum_{k=1}^{s-1} k(s-k) \right]$$

$$= 2 \left[ \frac{s}{3} \left( \frac{S(S-1)(S+1)}{6} - \frac{s(s-1)(s+1)}{6} \right) \right]$$

$$= \frac{s}{3} \left[ S^2 - s^2 \right].$$

You will need formulas for $\sum_{k=1}^{N} k$ and $\sum_{k=1}^{N} k^2$ or alternatively for $\sum_{k=1}^{N} k(M-k)$. These are easily found or derived.

**Exercise 3.39.** (1) For the long run average cost

$$C = \frac{K + (1/3)rS^3x(1-x^2)}{S^2x(S-x)},$$

find $\partial C/\partial x$. 
(2) For the long run average cost

\[ C = K + \frac{(1/3)rS^3x(1 - x^2)}{S^2x(1 - x)}. \]

find \( \frac{\partial C}{\partial S} \).

(3) Find the optimum values of \( x \) and \( S \).
CHAPTER 4

Limit Theorems for Coin Tossing

4.1. Laws of Large Numbers

Section Starter Question. Consider a fair \( p = 1/2 = q \) coin tossing game carried out for 1000 tosses. Explain in a sentence what the "law of averages" says about the outcomes of this game.

The Weak Law of Large Numbers.

Lemma 4.1 (Markov’s Inequality). If \( X \) is a random variable that takes only nonnegative values, then for any \( a > 0 \):

\[
P[X \geq a] \leq \frac{E[X]}{a}.
\]

Proof. Here is a proof for the case where \( X \) is a continuous random variable with probability density \( f \):

\[
E[X] = \int_{0}^{\infty} x f(x) \, dx
\]

\[
= \int_{0}^{a} x f(x) \, dx + \int_{a}^{\infty} x f(x) \, dx
\]

\[
\geq \int_{a}^{\infty} x f(x) \, dx
\]

\[
\geq \int_{a}^{\infty} a f(x) \, dx
\]

\[
= a \int_{a}^{\infty} f(x) \, dx
\]

\[
= a P[X \geq a].
\]

The proof for the case where \( X \) is a purely discrete random variable is similar with summations replacing integrals. The proof for the general case is exactly as given with \( dF(x) \) replacing \( f(x) \, dx \) and interpreting the integrals as Riemann-Stieltjes integrals. \( \square \)

Lemma 4.2 (Chebyshev’s Inequality). If \( X \) is a random variable with finite mean \( \mu \) and variance \( \sigma^2 \), then for any value \( k > 0 \):

\[
P[|X - \mu| \geq k] \leq \frac{\sigma^2}{k^2}.
\]

Proof. Since \((X - \mu)^2\) is a nonnegative random variable, apply Markov’s inequality (with \( a = k^2 \)) to obtain

\[
P[(X - \mu)^2 \geq k^2] \leq \frac{E[(X - \mu)^2]}{k^2}.
\]
But since \((X - \mu)^2 \geq k^2\) if and only if \(|X - \mu| \geq k\), the inequality above is equivalent to:

\[
P \left[ |X - \mu| \geq k \right] \leq \frac{\sigma^2}{k^2}
\]

and the proof is complete. \(\square\)

**Theorem 4.3 (Weak Law of Large Numbers).** Let \(X_1, X_2, X_3, \ldots\) be independent, identically distributed random variables each with mean \(\mu\) and variance \(\sigma^2\). Let \(S_n = X_1 + \cdots + X_n\). Then \(S_n/n\) converges in probability to \(\mu\). That is:

\[
\lim_{n \to \infty} P_n \left[ \left| \frac{S_n}{n} - \mu \right| > \epsilon \right] = 0.
\]

**Proof.** Since the mean of a sum of random variables is the sum of the means, and scalars factor out of expectations:

\[
E \left[ \frac{S_n}{n} \right] = \frac{1}{n} \sum_{i=1}^{n} E [X_i] = \frac{1}{n}(n\mu) = \mu.
\]

Since the variance of a sum of independent random variables is the sum of the variances, and scalars factor out of variances as squares:

\[
Var \left[ \frac{S_n}{n} \right] = \frac{1}{n^2} \sum_{i=1}^{n} Var [X_i] = \frac{1}{n^2}(n\sigma^2) = \sigma^2/n.
\]

Fix a value \(\epsilon > 0\). Then using elementary definitions for probability measure and Chebyshev’s Inequality:

\[
0 \leq P_n \left[ \left| \frac{S_n}{n} - \mu \right| > \epsilon \right] \leq P_n \left[ \left| \frac{S_n}{n} - \mu \right| \geq \epsilon \right] \leq \frac{\sigma^2}{(n\epsilon^2)}.
\]

Then by the squeeze theorem for limits

\[
\lim_{n \to \infty} P_n \left[ \left| \frac{S_n}{n} - \mu \right| > \epsilon \right] = 0.
\]

\(\square\)

Jacob Bernoulli originally proved the Weak Law of Large Numbers in 1713 for the special case when the \(X_i\) are binomial random variables. Bernoulli had to create an ingenious proof to establish the result, since Chebyshev’s inequality was not known at the time. The theorem then became known as Bernoulli’s Theorem. Simeon Poisson proved a generalization of Bernoulli’s binomial Weak Law and first called it the Law of Large Numbers. In 1929 the Russian mathematician Aleksandr Khinchin proved the general form of the Weak Law of Large Numbers presented here. Many other versions of the Weak Law are known, with hypotheses that do not require such stringent requirements as being identically distributed, and having finite variance.

**The Strong Law of Large Numbers.**

**Theorem 4.4 (Strong Law of Large Numbers).** Let \(X_1, X_2, X_3, \ldots\) be independent, identically distributed random variables each with mean \(\mu\) and variance \(E \left[ X_j^2 \right] < \infty\). Let \(S_n = X_1 + \cdots + X_n\). Then \(S_n/n\) converges with probability 1 to \(\mu\),

\[
P \left[ \lim_{n \to \infty} \frac{S_n}{n} = \mu \right] = 1.
\]
The proof of this theorem is beautiful and deep, but would take us too far afield to prove it. The Russian mathematician Andrey Kolmogorov proved the Strong Law in the generality stated here, culminating a long series of investigations through the first half of the 20th century.

**Discussion of the Weak and Strong Laws of Large Numbers.** In probability theory a theorem that tells us how a sequence of probabilities converges is called a **weak law**. For coin tossing, the sequence of probabilities is the sequence of binomial probabilities associated with the first $n$ tosses. The Weak Law of Large Numbers says that if we take $n$ large enough, then the binomial probability of the mean from the first $n$ tosses differing “much” from the theoretical mean should be small. However, this is a limit statement and the Weak Law of Large Numbers above does not indicate the rate of convergence, nor the dependence of the rate of convergence on the difference. Note furthermore that the Weak Law of Large Numbers in no way justifies the false notion called the “Gambler’s Fallacy”, namely that a long string of successive Heads indicates a Tail “is due to occur soon.” The independence of the random variables completely eliminates that sort of influence.

A **strong law** tells how the sequence of random variables as a sample path behaves in the limit. That is, among the infinitely many sequences (or paths) of coin tosses we select one “at random” and then evaluate the sequence of means along that path. The Strong Law of Large Numbers says that with probability 1 that sequence of means along that path will converge to the theoretical mean. The formulation of the notion of probability on an infinite (in fact an uncountably infinite) sample space requires mathematics beyond the scope of the course, partially accounting for the lack of a proof for the Strong Law here.

Note carefully the difference between the Weak Law of Large Numbers and the Strong Law. We do not simply move the limit inside the probability. These two results express different limits. The Weak Law is a statement that the group of finite-length experiments whose sample mean is close to the population mean approaches all of the possible experiments as the length increases. The Strong Law is an experiment-by-experiment statement, it says (almost every) sequence has a sample mean that approaches the population mean. This is reflected in the subtle difference in notation here. In the Weak Law the probabilities are written with a subscript: $P_n [\cdot]$ indicating this is a binomial probability distribution with parameter $n$ (and $p$). In the Strong Law, the probability is written without a subscript, indicating this is a probability measure on a sample space. Weak laws are usually much easier to prove than strong laws.

**Section Ending Answer.** There is no mathematical statement of a “law of averages” so it is not possible to explain it. Often the law of averages is the mistake that in 1000 tosses of a fair coin, there should be about the same number of heads and tails. The Weak Law of Large Numbers does say that the probability that the relative proportion of heads approaches $1/2$ as the number of tosses increases but gives no information about how many tosses are required for a given closeness. Even so, the absolute difference of heads from tails can be large while the relative proportion is close to $1/2$. These questions are examined in later sections of this chapter.

**Algorithms, Scripts, Simulations.**
Algorithm. The experiment is flipping a coin \( n \) times, and repeat the experiment \( k \) times. Then compute the proportion for which the sample mean deviates from \( p \) by more than \( \epsilon \).

```r
1 p <- 0.5
2 n <- 10000
3 k <- 1000
4 coinFlips <- array( (runif(n*k) <= p), dim=c(n,k))
5 # 0+ coerces Boolean to numeric
6 headsTotal <- colSums(coinFlips)
7 # 0..n binomial rv sample, size k
8
9 epsilon <- 0.01
10 mu <- p
11 prob <- sum( abs( headsTotal/n - mu ) > epsilon )/k
12 cat(sprintf("Empirical probability: %f 
", prob))
```

Key Concepts.
(1) The precise statement, meaning and proof of the Weak Law of Large Numbers.
(2) The precise statement and meaning of the Strong Law of Large Numbers.

Vocabulary.
(1) The **Weak Law of Large Numbers** is a precise mathematical statement of what is usually loosely referred to as the “law of averages”. Precisely, let \( X_1, \ldots, X_n \) be independent, identically distributed random variables each with mean \( \mu \) and variance \( \sigma^2 \). Let \( S_n = X_1 + \cdots + X_n \) and consider the sample mean \( \bar{X}_n \). Then the Weak Law of Large Numbers says that the sample mean \( \bar{X}_n \) converges in probability to the population mean \( \mu \). That is:

\[
\lim_{n \to \infty} P_n \left[ |\bar{X}_n - \mu| > \epsilon \right] = 0.
\]

In words, the proportion of those samples whose sample mean differs significantly from the population mean diminishes to zero as the sample size increases.

(2) The **Strong Law of Large Numbers** says that \( \bar{X}_n \) converges to \( \mu \) with probability 1. That is:

\[
P \left[ \lim_{n \to \infty} \bar{X}_n = \mu \right] = 1.
\]

In words, the Strong Law of Large Numbers “almost every” sample mean approaches the population mean as the sample size increases.

Problems.

**Exercise 4.5.** Suppose \( X \) is an exponentially distributed random variable with mean \( \mathbb{E}[X] = 1 \). For \( x = 0.5, 1, \) and \( 2 \), compare \( P[X \geq x] \) with the Markov Inequality bound.
Exercise 4.6. Suppose \( X \) is a Bernoulli random variable with \( \Pr[X = 1] = p \) and \( \Pr[X = 0] = 1 - p = q \). Compare \( \Pr[X \geq 1] \) with the Markov Inequality bound.

Exercise 4.7. Make a sequence of 100 coin tosses and keep a record as in the Experiment section. How “typical” was your coin flip sequence? All \( 2^{100} \) coin flip sequences are equally likely of course, so yours is neither more nor less typical than any other in that way. However, some sets or events of coin flip sequences are more or less “typical” when measured by the probability of a corresponding event. What is your value of \( S_{100} \), and the number of heads and the number of tails in your record? Using your value of \( S_{100} \) let \( \epsilon = |S_{100}| \) and use Chebyshev’s Inequality as in the proof of the Weak Law of Large Numbers to provide an upper bound on the probability that for all possible records \( |S_{100}| > \epsilon \).

Exercise 4.8. Write a proof of Markov’s Inequality for a random variable taking positive integer values.

Exercise 4.9. Modify the scripts to compute the proportion of sample means which deviate from \( p \) by more than \( \epsilon \) for increasing values of \( n \). Do the proportions decrease with increasing values of \( n \)? If so, at what rate do the proportions decrease?

### 4.2. Moment Generating Functions

**Section Starter Question.** Give some examples of transform methods in mathematics, science or engineering that you have seen or used and explain why transform methods are useful.

**Transform Methods.** We need some tools to aid in proving theorems about random variables. In this section we develop a tool called the **moment generating function** which converts problems about probabilities and expectations into problems from calculus about function values and derivatives. Moment generating functions are one of the large class of transforms in mathematics that turn a difficult problem in one domain into a manageable problem in another domain. Other examples are Laplace transforms, Fourier transforms, Z-transforms, generating functions, and even logarithms.

The general method can be expressed schematically in Figure 1.

**Expectation of Independent Random Variables.**

**Lemma 4.10.** If \( X \) and \( Y \) are independent random variables, then for any functions \( g \) and \( h \):
\[
\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)] \mathbb{E}[h(Y)]
\]

**Proof.** To make the proof definite suppose that \( X \) and \( Y \) are jointly continuous, with joint probability density function \( f(x, y) \). Then:
\[
\mathbb{E}[g(X)h(Y)] = \int \int_{(x,y)} g(x)h(y)f(x, y) \, dx \, dy
= \int_{x} \int_{y} g(x)h(y)f_{X}(x)f_{Y}(y) \, dx \, dy
= \int_{x} g(x)f_{X}(x) \, dx \int_{y} h(y)f_{Y}(y) \, dy
= \mathbb{E}[g(X)] \mathbb{E}[h(Y)].
\]
\[\square\]
Remark 4.11. In words, the expectation of the product of independent random variables is the product of the expectations.

The Moment Generating Function. The moment generating function \( \phi_X(t) \) is defined by

\[
\phi_X(t) = \mathbb{E} [e^{tX}] = \begin{cases} 
\sum_i e^{tx_i}p(x_i) & \text{if } X \text{ is discrete} \\
\int e^{tx}f(x) \, dx & \text{if } X \text{ is continuous}
\end{cases}
\]

for all values \( t \) for which the integral converges.

Example 4.12. The degenerate probability distribution has all the probability concentrated at a single point. That is, the degenerate random variable is a discrete random variable exhibiting certainty of outcome. If \( X \) is a degenerate random variable, then \( X = \mu \) with probability 1 and \( X \) is any other value with probability 0. The moment generating function of the degenerate random variable
4.2. MOMENT GENERATING FUNCTIONS

is particularly simple:

\[ \sum_{x_i = \mu} e^{tx_i} = e^{\mu t}. \]

If the moments of order \( k \) exist for \( 0 \leq k \leq k_0 \), then the moment generating function is continuously differentiable up to order \( k_0 \) at \( t = 0 \). Assuming that all operations can interchanged, the moments of \( X \) can be generated from \( \phi_X(t) \) by repeated differentiation:

\[
\begin{align*}
\phi_X'(t) &= \frac{d}{dt} \mathbb{E} \left[ e^{tX} \right] \\
&= \frac{d}{dx} \int x e^{tx} f_X(x) \, dx \\
&= \int x \frac{d}{dt} e^{tx} f_X(x) \, dx \\
&= \int x e^{tx} f_X(x) \, dx \\
&= \mathbb{E} \left[ X e^{tX} \right].
\end{align*}
\]

Then

\( \phi_X'(0) = \mathbb{E} [X] \).

Likewise

\[
\phi_X''(t) = \frac{d}{dt} \phi_X'(t)
\]

\[
= \frac{d}{dx} \int x e^{tx} f_X(x) \, dx \\
= \int x^2 \frac{d}{dt} e^{tx} f_X(x) \, dx \\
= \int x^2 e^{tx} f_X(x) \, dx \\
= \mathbb{E} \left[ X^2 e^{tX} \right].
\]

Then

\( \phi_X''(0) = \mathbb{E} [X^2] \).

Continuing in this way:

\( \phi_X^{(n)}(0) = \mathbb{E} [X^n] \).

**Remark 4.13.** In words, the value of the \( n \)th derivative of the moment generating function evaluated at 0 is the value of the \( n \)th moment of \( X \).

**Theorem 4.14.** If \( X \) and \( Y \) are independent random variables with moment generating functions \( \phi_X(t) \) and \( \phi_Y(t) \) respectively, then \( \phi_{X+Y}(t) \), the moment generating function of \( X + Y \) is given by \( \phi_X(t) \phi_Y(t) \). In words, the moment generating function of a sum of independent random variables is the product of the individual moment generating functions.
PROOF. Using the lemma on independence above:

\[ \phi_{X+Y}(t) = \mathbb{E}\left[e^{t(X+Y)}\right] = \mathbb{E}\left[e^{tX}e^{tY}\right] = \mathbb{E}\left[e^{tX}\right]\mathbb{E}\left[e^{tY}\right] = \phi_X(t)\phi_Y(t). \]

\[ \square \]

**Theorem 4.15.** If the moment generating function is defined in a neighborhood of \( t = 0 \) then the moment generating function uniquely determines the probability distribution. That is, there is a one-to-one correspondence between the moment generating function and the distribution function of a random variable, when the moment-generating function is defined and finite.

**Proof.** This proof is too sophisticated for the mathematical level we have now. \[ \square \]

**The moment generating function of a normal random variable.**

**Theorem 4.16.** If \( Z \sim \mathcal{N}(\mu, \sigma^2) \), then \( \phi_Z(t) = \exp(\mu t + \sigma^2 t^2/2) \).

**Proof.**

\[
\phi_Z(t) = \mathbb{E} [e^{tX}] = \frac{1}{\sqrt{2\pi \sigma^2}} \int_{-\infty}^{\infty} e^{tx} e^{-\left(x-\mu\right)^2/(2\sigma^2)} \, dx \\
= \frac{1}{\sqrt{2\pi \sigma^2}} \int_{-\infty}^{\infty} \exp \left( -\frac{(x^2 - 2\mu x + \mu^2 - 2\sigma^2 tx)}{2\sigma^2} \right) \, dx.
\]

By completing the square:

\[
x^2 - 2\mu x + \mu^2 - 2\sigma^2 tx = (x - (\mu + \sigma^2 t))^2 - (\mu + \sigma^2 t)^2 + \mu^2 \\
= (x - (\mu + \sigma^2 t))^2 - \sigma^4 t^2 - 2\mu \sigma^2 t.
\]

So returning to the calculation of the m.g.f.

\[
\phi_Z(t) = \frac{1}{\sqrt{2\pi \sigma^2}} \int_{-\infty}^{\infty} \exp \left( -\frac{(x - (\mu + \sigma^2 t))^2 - \sigma^4 t^2 - 2\mu \sigma^2 t}{2\sigma^2} \right) \, dx \\
= \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left( \frac{\sigma^4 t^2 + 2\mu \sigma^2 t}{2\sigma^2} \right) \int_{-\infty}^{\infty} \exp \left( -\frac{(x - (\mu + \sigma^2 t))^2}{2\sigma^2} \right) \, dx \\
= \exp \left( \frac{\sigma^4 t^2 + 2\mu \sigma^2 t}{2\sigma^2} \right) \\
= \exp \left( \mu t + \frac{\sigma^2 t^2}{2} \right). 
\]

\[ \square \]

**Theorem 4.17.** If \( Z_1 \sim \mathcal{N}(\mu_1, \sigma_1^2) \), and \( Z_2 \sim \mathcal{N}(\mu_2, \sigma_2^2) \) and \( Z_1 \) and \( Z_2 \) are independent, then \( Z_1 + Z_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) \). In words, the sum of independent normal random variables is again a normal random variable whose mean is the sum of the means, and whose variance is the sum of the variances.
4.3. THE CENTRAL LIMIT THEOREM

Proof. We compute the moment generating function of the sum using our theorem about sums of independent random variables. Then we recognize the result as the moment generating function of the appropriate normal random variable.

\[
\phi_{Z_1+Z_2}(t) = \phi_{Z_1}(t)\phi_{Z_2}(t) \\
= \exp(\mu_1 t + \sigma_1^2 t^2 / 2) \exp(\mu_2 t + \sigma_2^2 t^2 / 2) \\
= \exp((\mu_1 + \mu_2) t + (\sigma_1^2 + \sigma_2^2) t^2 / 2).
\]

\[\square\]

Section Ending Answer. The most common transform methods in mathematics, science and engineering are the Fourier Transform and the Laplace Transform but there are also many other transforms. These transforms are useful because they make finding solutions to differential equations easier and they expose important features of functions.

Problems.

Exercise 4.18. Calculate the moment generating function of a random variable \(X\) having a uniform distribution on \([0, 1]\). Use this to obtain \(E[X]\) and \(\text{Var}[X]\).

Exercise 4.19. Calculate the moment generating function of a discrete random variable \(X\) having a geometric distribution. Use this to obtain \(E[X]\) and \(\text{Var}[X]\).

Exercise 4.20. Calculate the moment generating function of a Bernoulli random variable having value 1 with probability \(p\) and value 0 with probability \(1 - p\). From this, derive the moment generating function of a binomial random variable with parameters \(n\) and \(p\).

4.3. The Central Limit Theorem

Section Starter Question. What is the most important probability distribution? Why do you choose that distribution as most important?

Convergence in Distribution.

Lemma 4.21. Let \(X_1, X_2, \ldots\) be a sequence of random variables having cumulative distribution functions \(F_{X_n}\) and moment generating functions \(\phi_{X_n}\). Let \(X\) be a random variable having cumulative distribution function \(F_X\) and moment generating function \(\phi_X\). If \(\phi_{X_n}(t) \to \phi_X(t)\), for all \(t\), then \(F_{X_n}(t) \to F_X(t)\) for all \(t\) at which \(F_X(t)\) is continuous.

We say that the sequence \(X_i\) converges in distribution to \(X\) and we write

\[X_i \xrightarrow{D} X.\]

Notice that \(P[a < X_i \leq b] = F_{X_i}(b) - F_{X_i}(a) \to F_X(b) - F_X(a) = P[a < X \leq b]\), so convergence in distribution implies convergence of probabilities of events. Likewise, convergence of probabilities of events implies convergence in distribution.

This lemma is useful because it is routine to determine the pointwise limit of a sequence of functions using ideas from calculus. It is usually much easier to check the pointwise convergence of the moment generating functions than it is to check the convergence in distribution of the corresponding sequence of random variables.
We won’t prove this lemma, since it would take us too far afield into the theory of moment generating functions and corresponding distribution theorems. However, the proof is a routine application of ideas from the mathematical theory of real analysis.

**Application: Weak Law of Large Numbers.** Here’s a simple representative example of using the convergence of the moment generating function to prove a useful result. We will prove a version of the Weak Law of Large Numbers that does not require the finite variance of the sequence of independent, identically distributed random variables.

**Theorem 4.22 (Weak Law of Large Numbers).** Let \( X_1, \ldots, X_n \) be independent, identically distributed random variables each with mean \( \mu \) and such that \( \mathbb{E} [ |X_i|] \) is finite. Let \( S_n = X_1 + \cdots + X_n \). Then \( S_n/n \) converges in probability to \( \mu \). That is:

\[
\lim_{n \to \infty} \mathbb{P}_n [ |S_n/n - \mu| > \epsilon] = 0.
\]

**Proof.** If we denote the common moment generating function of \( X_i \) by \( \phi(t) \), then the moment generating function of \( S_n/n = \sum_{i=1}^n X_i/n \) is \((\phi(t/n))^n\). The existence of the first moment assures us that \( \phi(t) \) is differentiable at 0 with a derivative equal to \( \mu \). Therefore, by tangent-line approximation (first-degree Taylor polynomial approximation)

\[
\phi \left( \frac{t}{n} \right) = 1 + \mu \frac{t}{n} + R_1(t/n)
\]

where \( r_2(t/n) \) is an error term such that

\[
\lim_{n \to \infty} \frac{r_2(t/n)}{(t/n)} = 0.
\]

This is equivalent to \((1/t) \lim_{n \to \infty} nR_1(t/n) = 0 \) or \( \lim_{n \to \infty} nR_1(t/n) = 0 \), needed for taking the limit in (4.1). Then we need to consider

(4.1) \[
\phi \left( \frac{t}{n} \right)^n = \left( 1 + \mu \frac{t}{n} + R_1(t/n) \right)^n.
\]

Taking the logarithm of \((1 + \mu(t/n) + R_1(t/n))^n\) and using L’Hospital’s Rule, we see that

\[
\phi(t/n)^n \to \exp(\mu t).
\]

But this last expression is the moment generating function of the degenerate point mass distribution concentrated at \( \mu \). Hence,

\[
\lim_{n \to \infty} \mathbb{P}_n [ |S_n/n - \mu| > \epsilon] = 0.
\]

\( \square \)
The Central Limit Theorem.

**Theorem 4.23 (Central Limit Theorem).** Let random variables $X_1, \ldots, X_n$

- be independent and identically distributed;
- have common mean $E[X_i] = \mu$ and common variance $\text{Var}[X_i] = \sigma^2$; and
- the common moment generating function $\phi_{X_i}(t) = E[e^{tX_i}]$ exists and is finite in a neighborhood of $t = 0$.

Consider $S_n = \sum_{i=1}^{n} X_i$. Let

- $Z_n = (S_n - n\mu)/(\sigma\sqrt{n}) = (1/\sigma)(S_n/n - \mu)/\sqrt{n}$; and
- $Z$ be the standard normally distributed random variable with mean 0 and variance 1.

Then $Z_n$ converges in distribution to $Z$, that is:

$$
\lim_{n \to \infty} \mathbb{P}_n[Z_n \leq a] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} \exp(-u^2/2) \, du.
$$

**Remark 4.24.** The Central Limit Theorem is true even under the slightly weaker assumptions that $X_1, \ldots, X_n$ only are independent and identically distributed with finite mean $\mu$ and finite variance $\sigma^2$ without the assumption that the moment generating function exists. However, the proof below using moment generating functions is simple and direct enough to justify using the additional hypothesis.

**Proof.** Assume at first that $\mu = 0$ and $\sigma^2 = 1$. Assume also that the moment generating function of the $X_i$ (which are identically distributed, so there is only one m.g.f.) is $\phi_X(t)$, exists and is everywhere finite. Then the m.g.f. of $X_i/\sqrt{n}$ is

$$
\phi_{X/\sqrt{n}}(t) = E \left[ \exp(\frac{tX_i}{\sqrt{n}}) \right] = \phi_X(t/\sqrt{n}).
$$

Recall that the m.g.f. of a sum of independent r.v.s is the product of the m.g.f.s. Thus the m.g.f of $S_n/\sqrt{n}$ is (note that here we used $\mu = 0$ and $\sigma^2 = 1)$

$$
\phi_{S_n/\sqrt{n}}(t) = [\phi_X(t/\sqrt{n})]^n
$$

The quadratic approximation (second-degree Taylor polynomial expansion) of $\phi_X(t)$ at 0 is by calculus:

$$
\phi_X(t) = \phi_X(0) + \phi_X'(0)t + (\phi_X''(0)/2)t^2 + r_3(t) = 1 + t^2/2 + r_3(t)
$$

again since the hypotheses assume $E[X] = \phi'(0) = 0$ and $\text{Var}[X] = E[X^2] - (E[X])^2 = \phi''(0) - (\phi'(0))^2 = \phi''(0) = 1$. Here $r_3(t)$ is an error term such that $\lim_{t \to 0} r_3(t)/t^2 = 0$. Thus,

$$
\phi(t/\sqrt{n}) = 1 + t^2/(2n) + r_3(t/\sqrt{n})
$$

implying that

$$
\phi_{S_n/\sqrt{n}} = [1 + t^2/(2n) + r_3(t/\sqrt{n})]^n.
$$

Now by some standard results from calculus,

$$
[1 + t^2/(2n) + r_3(t/\sqrt{n})]^n \to \exp(t^2/2)
$$

as $n \to \infty$. (If the reader needs convincing, it’s easy to show that

$$
n \log(1 + t^2/(2n) + r_3(t/\sqrt{n})) \to t^2/2,
$$

using L’Hospital’s Rule to account for the $r_3(t)$ term.)
4. LIMIT THEOREMS FOR COIN TOSING

To handle the general case, consider the standardized random variables \( (X_i - \mu)/\sigma \), each of which now has mean 0 and variance 1 and apply the result.

Abraham de Moivre proved the first version of the Central Limit Theorem around 1733 in the special case when the \( X_i \) are binomial random variables with \( p = 1/2 = q \). Pierre-Simon Laplace subsequently extended the proof to the case of arbitrary \( p \neq q \). Laplace also discovered the more general form of the Central Limit Theorem presented here. His proof however was not completely rigorous, and in fact, cannot be made completely rigorous. A truly rigorous proof of the Central Limit Theorem was first presented by the Russian mathematician Aleksandr Lyapunov in 1901-1902. As a result, the Central Limit Theorem (or a slightly stronger version of the Central Limit Theorem) is occasionally referred to as Lyapunov’s theorem. A theorem with weaker hypotheses but with equally strong conclusion is Lindeberg’s Theorem of 1922. It says that the sequence of random variables need not be identically distributed, but instead need only have zero means, and the individual variances are small compared to their sum.

Accuracy of the Approximation by the Central Limit Theorem. The statement of the Central Limit Theorem does not say how good the approximation is. One rule of thumb is that the approximation given by the Central Limit Theorem applied to a sequence of Bernoulli random trials or equivalently to a binomial random variable is acceptable when \( np(1-p) > 18 \) [36, page 34], [53, page 134]. The normal approximation to a binomial deteriorates as the interval \((a,b)\) over which the probability is computed moves away from the binomial’s mean value \( np \). Another rule of thumb is that the normal approximation is acceptable when \( n \geq 30 \) for all reasonable probability distributions.

The Berry-Esséen Theorem gives an explicit bound: For independent, identically distributed random variables \( X_i \) with \( \mu = \mathbb{E}[X_i] = 0 \), \( \sigma^2 = \mathbb{E}[X_i^2] \), and \( \rho = \mathbb{E}[|X_i^3]| \), then

\[
\left| \mathbb{P} \left[ \frac{S_n}{\sigma \sqrt{n}} < a \right] - \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \, du \right| \leq \frac{33}{4} \frac{\rho}{\sigma^3} \frac{1}{\sqrt{n}}.
\]

Illustration 1. Figure 2 is a graphical illustration of the Central Limit Theorem. More precisely, this is an illustration of the de Moivre-Laplace version, the approximation of the binomial distribution with the normal distribution.

The figure is actually a non-centered and unscaled illustration since the binomial random variable \( S_n \) is not shifted by the mean, nor normalized to unit variance. Therefore, the binomial and the corresponding approximating normal are both centered at \( \mathbb{E}[S_n] = np \). The variance of the approximating normal is \( \sigma^2 = \sqrt{npq} \) and the widths of the bars denoting the binomial probabilities are all unit width, and the heights of the bars are the actual binomial probabilities.

Illustration 2. We expect the normal distribution to apply whenever the numerical description of a state of a system results from numerous small additive effects, with no single or small group of effects dominant. Here is another illustration of that principle.

We can use the Central Limit Theorem to assess risk. Two large banks compete for customers to take out loans. The banks have comparable offerings. Assume that each bank has a certain amount of funds available for loans to customers. Any
customers seeking a loan beyond the available funds will cost the bank, either as a lost opportunity cost, or because the bank itself has to borrow to secure the funds to lend to the customer. If too few customers take out loans then that also costs the bank since the bank then has unused funds.

We create a simple mathematical model of this situation. We suppose that the loans are all of equal size and for definiteness each bank has funds available for a certain number (to be determined) of these loans. Then suppose $n$ customers select a bank independently and at random. Let $X_i = 1$ if customer $i$ selects bank H with probability $1/2$ and $X_i = 0$ if customers select bank T, also with probability $1/2$. Then $S_n = \sum_{i=1}^n X_i$ is the number of loans from bank H to customers. Now there is some positive probability that more customers will turn up than the bank can accommodate. We can approximate this probability with the Central Limit Theorem:

$$
P[S_n > s] = P_n \left[ \left( S_n - n/2 \right) / \left( (1/2) \sqrt{n} \right) > \left( s - n/2 \right) / \left( (1/2) \sqrt{n} \right) \right]

\approx P \left[ Z > \left( s - n/2 \right) / \left( (1/2) \sqrt{n} \right) \right]

= P \left[ Z > \left( 2s - n \right) / \sqrt{n} \right].$$

Now if $n$ is large enough that this probability is less than (say) 0.01, then the number of loans will be sufficient in 99 of 100 cases. Looking up the value in a normal probability table,

$$\frac{2s - n}{\sqrt{n}} > 2.33$$

so if $n = 1000$, then $s = 537$ will suffice. If both banks assume the same risk of sellout at 0.01, then each will have 537 for a total of 1074 loans, of which 74 will be unused.
Now the possibilities for generalization and extension are apparent. A first generalization would be to allow the loan amounts to be random with some distribution. Still we could apply the Central Limit Theorem to approximate the demand on available funds. Second, the cost of either unused funds or lost business could be multiplied by the chance of occurring. The total of the products would be an expected cost, which could then be minimized.

**Section Ending Answer.** The standard normal distribution is the most important because it is central to probability theory, applied probability, and statistics.

**Algorithms, Scripts, Simulations.**

**Algorithm.** The experiment is flipping a coin \( n \) times, and repeat the experiment \( k \) times. Then compute the proportion for which the deviation of the sample sum from \( np \) by more than \( \sqrt{p(1-p)n} \) is less than \( a \). Compare this to the theoretical probability from the standard normal distribution.

**Script.**

```r
1 p <- 0.5
2 n <- 10000
3 k <- 1000
4 coinFlips <- array( 0+(runif(n*k) <= p), dim=c(n,k))
5 # 0+ coerces Boolean to numeric
6 headsTotal <- colSums(coinFlips)
7 # 0..n binomial rv sample, size k
8
9 mu <- p
10 sigma <- sqrt(p*(1-p))
11 a <- 0.5
12 Zn <- (headsTotal - n*mu)/(sigma * sqrt(n))
13 prob <- sum( 0+(Zn < a) )/k
14 theoretical <- pnorm(a, mean=0 , sd=1)
15 cat(sprintf("Empirical probability: %f \n", prob ))
16 cat(sprintf("Central Limit Theorem estimate: %f \n", theoretical))
```

**Problems.**

**Exercise 4.25.** Suppose you bought a stock at a price \( b+c \), where \( c > 0 \) and the present price is \( b \). (Too bad!) You have decided to sell the stock after 30 more trading days have passed. Assume that the daily change of the company’s stock on the stock market is a random variable with mean 0 and variance \( \sigma^2 \). That is, if \( S_n \) represents the price of the stock on day \( n \) with \( S_0 \) given, then

\[
S_n = S_{n-1} + X_n, n \geq 1
\]

where \( X_1, X_2, \ldots \) are independent, identically distributed continuous random variables with mean 0 and variance \( \sigma^2 \). Write an expression for the probability that you do not recover your purchase price.
Exercise 4.26. If you buy a lottery ticket in 50 independent lotteries, and in each lottery your chance of winning a prize is 1/100, write down and evaluate the probability of winning and also approximate the probability using the Central Limit Theorem of

1. exactly one prize,
2. at least one prize,
3. at least two prizes.

Explain with a reason whether or not you expect the approximation to be a good approximation.

Exercise 4.27. Find a number $k$ such that the probability is about 0.6 that the number of heads obtained in 1000 tossings of a fair coin will be between 440 and $k$.

Exercise 4.28. Find the moment generating function $\phi_X(t) = E[\exp(tX)]$ of the random variable $X$ which takes values 1 with probability 1/2 and $-1$ with probability 1/2. Show directly (that is, without using Taylor polynomial approximations) that $\phi_X(t/\sqrt{n})^n \to \exp(t^2/2)$. (Hint: Use L’Hospital’s Theorem to evaluate the limit, after taking logarithms of both sides.)

Exercise 4.29. A bank has $1,000,000 available to make for car loans. The loans are in random amounts uniformly distributed from $5,000 to $20,000. How many loans can the bank make with 99% confidence that it will have enough money available?

Exercise 4.30. An insurance company is concerned about health insurance claims. Through an extensive audit, the company has determined that overstatements (claims for more health insurance money than is justified by the medical procedures performed) vary randomly with an exponential distribution $X$ with a parameter 1/100 which implies that $E[X] = 100$ and $\text{Var}[X] = 100^2$. The company can afford some overstatements simply because it is cheaper to pay than it is to investigate and counter-claim to recover the overstatement. Given 100 claims in a month, the company wants to know what amount of reserve will give 95% certainty that the overstatements do not exceed the reserve. (All units are in dollars.) What assumptions are you using?

Exercise 4.31. Modify the scripts to vary the upper bounds $a$ and lower bound $b$ (with the other parameters fixed) and observe the difference of the empirical probability and the theoretical probability.

Exercise 4.32. Modify the scripts to vary the probability $p$ (with the other parameters fixed) and observe the difference of the empirical probability and the theoretical probability. Make a conjecture about the difference as a function of $p$ (i.e. where is the difference increasing, decreasing.)

Exercise 4.33. Modify the scripts to vary the number of trials $n$ (with the other parameters fixed) and observe the difference of the empirical probability and the theoretical probability. Test the rate of decrease of the deviation with increasing $n$. Does it follow the predictions of the Berry-Esséen Theorem?
CHAPTER 5

Brownian Motion

5.1. Intuitive Introduction to Diffusions

Section Starter Question. Suppose you wanted to display the function $y = \sqrt{x}$ with a computer plotting program or a graphing calculator. Describe the process to choose a proper window to display the graph.

Visualizing Limits of Random Walks. How should we set up the limiting process so we can make a continuous time limit of the discrete time random walk? First we consider a discovery approach to this question by asking what do we require so that we can visualize the limiting process. Next we take a probabilistic view using the Central Limit Theorem to justify the limiting process to pass from a discrete probability distribution to a probability density function. Finally, we consider the limiting process to a differential equation derived from the difference equation that is the result of first-step analysis.

The Random Walk. Consider a random walk starting at the origin. Starting from 0, the $n$th step takes the walker to the position $T_n = Y_1 + \cdots + Y_n$, the sum of $n$ independent, identically distributed Bernoulli random variables $Y_i$ assuming the values +1 and −1 with probabilities $p$ and $q = 1 - p$ respectively. Then recall that the mean of a sum of random variables is the sum of the means:

$$E[T_n] = (p - q)n$$

and the variance of a sum of independent random variables is the sum of the variances:

$$\text{Var}[T_n] = 4pqn.$$

Trying to use the mean to derive the limit. Now suppose we want to display a video of the random walk moving left and right along the $x$-axis. This would be a video of the phase line diagram of the random walk. Suppose we want the video to display 1 million steps and be a reasonable length of time, say 1000 seconds, between 16 and 17 minutes. This fixes the time scale at a rate of one step per millisecond. What should be the window in the screen to get a good sense of the random walk? For this question, we use a fixed unit of measurement, say

![Figure 1](image-url)

**Figure 1.** Image of a possible random walk in phase line after an odd number of steps.
Let \( \delta \) be the length of the steps. To find the window to display the random walk on the axis, we then need to know the size of \( \delta \cdot T_n \). Now

\[
\mathbb{E}[\delta \cdot T_n] = \delta \cdot (p-q)n
\]

and

\[
\text{Var}[\delta \cdot T_n] = \delta^2 \cdot 4pqn.
\]

We want \( n \) to be large (about 1 million) and to see the walk on the screen we want the expected end place to be comparable to the screen size, say 30 cm. That is,

\[
\mathbb{E}[\delta \cdot T_n] = \delta \cdot (p-q)n < \delta \cdot n \approx 30 \text{ cm}
\]

so \( \delta \) must be \( 3 \times 10^{-5} \) cm to get the end point on the screen. But then the movement of the walk measured by the standard deviation

\[
\sqrt{\text{Var}[\delta \cdot T_n]} \leq \delta \cdot \sqrt{n} = 3 \times 10^{-2} \text{ cm}
\]

will be so small as to be indistinguishable. We will not see any random variations!

**Trying to use the variance to derive the limit.** Let us turn the question around: We want to see the variations in many-step random walks, so the standard deviations must be a reasonable fraction \( D \) of the screen size

\[
\sqrt{\text{Var}[\delta \cdot T_n]} \leq \delta \cdot \sqrt{n} = D \cdot 30 \text{ cm}.
\]

For \( n = 10^6 \) this is possible if \( \delta = D \cdot 3 \times 10^{-2} \) cm. We still want to be able to see the expected ending position which will be

\[
\mathbb{E}[\delta \cdot T_n] = \delta \cdot (p-q)n = (p-q) \cdot D \cdot 3 \times 10^4 \text{ cm}.
\]

To be consistent with the requirement that the ending position is on the screen this will only be possible if \( (p-q) \approx 10^{-3} \). That is, \( p-q \) must be at most comparable in magnitude to \( \delta = 3 \times 10^{-2} \).

**The limiting process.** Now generalize these results to visualize longer and longer walks in a fixed amount of time. Since \( \delta \to 0 \) as \( n \to \infty \), then likewise \( (p-q) \to 0 \), while \( p+q = 1 \), so \( p \to 1/2 \). The analytic formulation of the problem is as follows. Let \( \delta \) be the size of the individual steps, let \( r \) be the number of steps per unit time. We ask what happens to the random walk in the limit where \( \delta \to 0 \), \( r \to \infty \), and \( p \to 1/2 \) in such a manner that:

\[
(p-q) \cdot \delta \cdot r \to c
\]

and

\[
4pq \cdot \delta^2 \cdot r \to D.
\]

Each of these says that we should consider symmetric \( (p = 1/2 = q) \) random walks with step size inversely proportional to the square root of the stepping rate.

The limiting process taking the discrete time random walk to a continuous time process is delicate. It is delicate because we are attempting to scale in two variables, the step size or space variable and the stepping rate or time variable, simultaneously. The variables are not independent. Two relationships connect them, one for the expected value and one for the variance. Therefore we expect that the scaling is only possible when the step size and stepping rate have a special relationship, namely the step size inversely proportional to the square root of the stepping rate.
Probabilistic Solution of the Limit Question. In our accelerated random walk, consider the nth step at time \( t = n/r \) and consider the position on the line \( x = k \cdot \delta \). Let 

\[
v_{k,n} = \mathbb{P} [\delta \cdot T_n = k \delta]
\]

be the probability that the nth step is at position \( k \). We are interested in the probability of finding the walk at given instant \( t \) and in the neighborhood of a given point \( x \), so we investigate the limit of \( v_{k,n} \) as \( n/r \to t \), and \( k \cdot \delta \to x \) with the additional conditions that \((p - q) \cdot \delta \cdot r \to c\) and \(4pq \cdot \delta^2 \cdot r \to D\).

Remember that the random walk can only reach an even-numbered position after an even number of steps, and an odd-numbered position after an odd number of steps. Therefore in all cases \( n + k \) is even and \((n + k)/2\) is an integer. Likewise \( n - k \) is even and \((n - k)/2\) is an integer. We reach position \( k \) at time step \( n \) if the walker takes \((n + k)/2\) steps to the right and \((n - k)/2\) steps to the left. The mix of steps to the right and the left can be in any order. So the walk \( \delta \cdot T_n \) reaches position \( k \delta \) at step \( n = rt \) with binomial probability

\[
v_{k,n} = \binom{n}{(n + k)/2} p^{(n + k)/2} q^{(n - k)/2}.
\]

From the Central Limit Theorem

\[
\mathbb{P} [\delta \cdot T_n = k \cdot \delta] = \mathbb{P} [(k - 1) \cdot \delta < \delta \cdot T_n < (k + 1) \cdot \delta]
\]

\[
= \mathbb{P} \left[ \frac{(k - 1)\delta - (p - q)\delta n}{\sqrt{4pq\delta^2 n}} < \frac{\delta T_n - (p - q)\delta n}{\sqrt{4pq\delta^2 n}} < \frac{(k + 1)\delta - (p - q)\delta n}{\sqrt{4pq\delta^2 n}} \right]
\]

\[
\approx \int_{(k-1)\delta}^{(k+1)\delta} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \, du
\]

\[
= \int_{(k-1)\delta}^{(k+1)\delta} \frac{1}{\sqrt{2\pi} \cdot 4pq\delta^2 n} e^{-(z-(p-q)\delta n)^2/(2\cdot 4pq\delta^2 n)} \, dz
\]

\[
= \frac{2\delta}{\sqrt{2\pi} \cdot 4pq\delta^2 n} e^{-(k\delta-(p-q)\delta n)^2/(2\cdot 4pq\delta^2 n)}
\]

\[
\approx \frac{2\delta}{\sqrt{2\pi} \cdot 4pq\delta^2 rt} e^{-(k\delta-(p-q)\delta r)^2/(2\cdot 4pq\delta^2 rt)}
\]

\[
= \frac{2\delta}{\sqrt{2\pi} DT} e^{-(x-ct)^2/(2\cdot Dt)}
\]

Similarly

\[
\mathbb{P} [a \cdot \delta < \delta \cdot T_n \cdot \delta < b\delta] \to \frac{1}{\sqrt{2\pi DT}} \int_a^b \exp \left( \frac{-(x-ct)^2}{2Dt} \right) \, dt.
\]

The integral on the right may be expressed in terms of the standard normal cumulative distribution function.

Note that we derived the limiting approximation of the binomial distribution

\[
v_{k,n} \sim \left( \frac{2\delta}{\sqrt{2\pi} DT} \right) \exp \left( \frac{-(x-ct)^2}{2Dt} \right)
\]

by applying the general form of the Central Limit Theorem. However, it is possible to derive this limit directly through careful analysis. The direct derivation is the de
Moivre-Laplace Limit Theorem and it is the most basic form of the Central Limit Theorem.

**Differential Equation Solution of the Limit Question.** Another method is to start from the difference equations governing the random walk, and then pass to a differential equation in the limit. Later we can generalize the differential equation and find that the generalized equations govern new continuous-time stochastic processes. Since differential equations have a well-developed theory and many tools to manipulate, transform and solve them, this method turns out to be useful.

Consider the position of the walker in the random walk at the \( n \)th and \((n + 1)\)st trial. Through a first-step analysis the probabilities \( v_{k,n} \) satisfy the difference equations:

\[
v_{k,n+1} = p \cdot v_{k-1,n} + q \cdot v_{k+1,n}.
\]

In the limit as \( k \to \infty \) and \( n \to \infty \), \( v_{k,n} \) will be the sampling of the function \( v(t,x) \) at time intervals \( r^{-1} \), so that \( n = rt \), and space intervals so that \( k\delta = x \). That is, the function \( v(t,x) \) should be an approximate solution of the difference equation:

\[
v(t + r^{-1}, x) = pv(t, x - \delta) + qv(t, x + \delta).
\]

We assume \( v(t,x) \) is a smooth function so that we can expand \( v(t,x) \) in a Taylor series at any point. Using the first order approximation in the time variable on the left, and the second-order approximation on the right in the space variable, we get (after canceling the leading terms \( v(t,x) \))

\[
\frac{\partial v(t,x)}{\partial t} = (q - p) \cdot \delta r \frac{\partial v(t,x)}{\partial x} + \frac{1}{2} \delta^2 r \frac{\partial^2 v(t,x)}{\partial x^2}.
\]

In our passage to limit, the omitted terms of higher order tend to zero, so we neglect them. The remaining coefficients are already accounted for in our limits and so the equation becomes:

\[
\frac{\partial v(t,x)}{\partial t} = -c \frac{\partial v(t,x)}{\partial x} + \frac{1}{2} D \frac{\partial^2 v(t,x)}{\partial x^2}.
\]

This is a special diffusion equation, more specifically, a diffusion equation with convective or drift terms, also known as the Fokker-Planck equation for diffusion. It is a standard problem to solve the differential equation for \( v(t,x) \) and therefore, we can find the probability of being at a certain position at a certain time. One can verify that

\[
v(t,x) = \frac{1}{\sqrt{2\pi Dt}} \exp \left( \frac{-(x - ct)^2}{2Dt} \right)
\]

is a solution of the diffusion equation, so we reach the same probability distribution for \( v(t,x) \).

The diffusion equation can be immediately generalized by permitting the coefficients \( c \) and \( D \) to depend on \( x \), and \( t \). Furthermore, the equation possesses obvious analogues in higher dimensions and all these generalization can be derived from general probabilistic postulates. We will ultimately describe stochastic processes related to these equations as **diffusions**.
Section Ending Answer. First choose a fixed domain \(0 \leq x \leq a\) for the graph, then an appropriate window for the graph of the function over this domain would have a horizontal range from 0 to \(a\) and a vertical range of 0 to \(\sqrt{a}\), perhaps adding some small margins for context. These ranges are just enough for viewing the entire graph on the chosen domain. The same motivations of seeing all the relevant features of a phase line diagram of a random walk guide the choice of limit process.

Problems.

Exercise 5.1. Consider a random walk with a step to the right having probability \(p\) and a step to the left having probability \(q\). The step length is \(\delta\). The walk is taking \(r\) steps per minute. What is the rate of change of the expected final position and the rate of change of the variance? What must we require on the quantities \(p\), \(q\), \(r\) and \(\delta\) in order to see the entire random walk with more and more steps at a fixed size in a fixed amount of time?

Exercise 5.2. Verify the limit taking to show that
\[
v_{k,n} \sim \frac{1}{\sqrt{2\pi Dt}} \exp\left(\frac{-(x - ct)^2}{2Dt}\right).
\]

Exercise 5.3. Show that
\[
v(t, x) = \frac{1}{\sqrt{2\pi Dt}} \exp\left(\frac{-(x - ct)^2}{2Dt}\right)
\]
is a solution of
\[
\frac{\partial v(t, x)}{\partial t} = -c \frac{\partial v(t, x)}{\partial x} + \frac{1}{2} D \frac{\partial^2 v(t, x)}{\partial x^2}
\]
by substitution.

5.2. The Definition of Brownian Motion and the Wiener Process

Section Starter Question. Some mathematical objects are defined by a formula or an expression. Some other mathematical objects are defined by their properties, not explicitly by an expression. That is, the objects are defined by how they act, not by what they are. Can you name a mathematical object defined by its properties?

Definition of Wiener Process. Brownian motion is the physical phenomenon named after the English botanist Robert Brown who discovered it in 1827. Brownian motion is the zig-zagging motion exhibited by a small particle, such as a grain of pollen, immersed in a liquid or a gas. Albert Einstein gave the first explanation of this phenomenon in 1905. He explained Brownian motion by assuming the immersed particle was constantly buffeted by the molecules of the surrounding medium. Since then the abstracted process has been used for modeling the stock market and in quantum mechanics. The Wiener process is the mathematical definition and abstraction of the physical process as a stochastic process. The American mathematician Norbert Wiener gave the definition and properties in a series of papers starting in 1918. Generally, the terms Brownian motion and Wiener process are the same, although Brownian motion emphasizes the physical aspects and Wiener process emphasizes the mathematical aspects. Bachelier
process means the same thing as Brownian motion and Wiener process. In 1900, Louis Bachelier introduced the limit of random walk as a model for prices on the Paris stock exchange, and so is the originator of the mathematical idea now called Brownian motion. This term is occasionally found in financial literature.

Previously we considered a discrete time random process. That is, at discrete times \( n = 1, 2, 3, \ldots \) corresponding to coin flips, we considered a sequence of random variables \( T_n \). We are now going to consider a continuous time random process, a function \( W(t) \) that is a random variable at each time \( t \geq 0 \). To say \( W(t) \) is a random variable at each time is too general so we must put more restrictions on our process to have something interesting to study.

**Definition 5.4 (Wiener process).** The standard Wiener process is a stochastic process \( W(t) \), for \( t \geq 0 \), with the following properties:

1. Every increment \( W(t) - W(s) \) over an interval of length \( t - s \) is normally distributed with mean 0 and variance \( t - s \), that is

\[
W(t) - W(s) \sim N(0, t - s).
\]

2. For every pair of time intervals \([t_1, t_2] \) and \([t_3, t_4] \), with \( t_1 < t_2 \leq t_3 < t_4 \), the increments \( W(t_4) - W(t_3) \) and \( W(t_2) - W(t_1) \) are independent random variables with distributions given as in part 1, and similarly for \( n \) time intervals where \( n \) is an arbitrary positive integer.

3. \( W(0) = 0 \).

4. \( W(t) \) is continuous for all \( t \).

Note that property 2 says that if we know \( W(s) = x_0 \), then the independence (and \( W(0) = 0 \)) tells us that further knowledge of the values of \( W(\tau) \) for \( \tau < s \) give no added knowledge of the probability law governing \( W(t) - W(s) \) with \( t > s \). More formally, this says that if \( 0 \leq t_0 < t_1 < \cdots < t_n < t \), then

\[
\mathbb{P}[W(t) \geq x \mid W(t_0) = x_0, W(t_1) = x_1, \ldots, W(t_n) = x_n] = \mathbb{P}[W(t) \geq x \mid W(t_n) = x_n].
\]

This is a statement of the Markov property of the Wiener process.

Recall that the sum of independent random variables that are respectively normally distributed with mean \( \mu_1 \) and \( \mu_2 \) and variances \( \sigma^2_1 \) and \( \sigma^2_2 \) is a normally distributed random variable with mean \( \mu_1 + \mu_2 \) and variance \( \sigma^2_1 + \sigma^2_2 \). Therefore for increments \( W(t_3) - W(t_2) \) and \( W(t_2) - W(t_1) \) the sum \( W(t_3) - W(t_2) + W(t_2) - W(t_1) = W(t_3) - W(t_1) \) is normally distributed with mean 0 and variance \( t_3 - t_1 \) as we expect. Property 2 of the definition is consistent with properties of normal random variables.

Let

\[
p(x, t) = \frac{1}{\sqrt{2\pi t}} \exp(-x^2/(2t))
\]

denote the probability density for a \( N(0, t) \) random variable. Then to derive the joint density of the event

\[
W(t_1) = x_1, W(t_2) = x_2, \ldots, W(t_n) = x_n
\]

with \( t_1 < t_2 < \cdots < t_n \), it is equivalent to know the joint probability density of the equivalent event

\[
W(t_1) - W(0) = x_1, W(t_2) - W(t_1) = x_2 - x_1, \ldots, W(t_n) - W(t_{n-1}) = x_n - x_{n-1}.
\]
Then by part 2, we immediately get the expression for the joint probability density function:

\[ f(x_1, t_1; x_2, t_2; \ldots; x_n, t_n) = p(x_1, t) p(x_2 - x_1, t_2 - t_1) \ldots p(x_n - x_{n-1}, t_n - t_{n-1}). \]

**Comments on Modeling Security Prices with the Wiener Process.** A plot of security prices over time and a plot of one-dimensional Brownian motion versus time has at least a superficial resemblance.

If we were to use Brownian motion to model security prices (ignoring for the moment that security prices are better modeled with the more sophisticated geometric Brownian motion rather than simple Brownian motion) we would need to verify that security prices have the four defining properties of Brownian motion.
Figure 3. A standardized density histogram of 1000 daily close-to-close returns on the S & P 500 Index, from February 29, 2012 to March 1, 2012, up to February 21, 2016, to February 22, 2016.

(1) The assumption of normal distribution of stock price changes seems to be a reasonable first assumption. Figure 3 illustrates this reasonable agreement. The Central Limit Theorem provides a reason to believe the agreement, assuming the requirements of the Central Limit Theorem are met, including independence. (Unfortunately, although the figure shows what appears to be reasonable agreement a more rigorous statistical analysis shows that the data distribution does not match normality.)

Another good reason for still using the assumption of normality for the increments is that the normal distribution is easy to work with. The normal probability density uses simple functions familiar from calculus, the normal cumulative probability distribution is tabulated, the moment-generating function of the normal distribution is easy to use, and the sum of independent normal distributions is again normal. A substitution of
another distribution is possible but the resulting stochastic process models are difficult to analyze, beyond the scope of this model.

However, the assumption of a normal distribution ignores the small possibility that negative stock prices could result from a large negative change. This is not reasonable. (The log normal distribution from geometric Brownian motion that avoids this possibility is a better model.)

Moreover, the assumption of a constant variance on different intervals of the same length is not a good assumption since stock volatility itself seems to be volatile. That is, the variance of a stock price changes and need not be proportional to the length of the time interval.

(2) The assumption of independent increments seems to be a reasonable assumption, at least on a long enough term. From second to second, price increments are probably correlated. From day to day, price increments are probably independent. Of course, the assumption of independent increments in stock prices is the essence of what economists call the Efficient Market Hypothesis, or the Random Walk Hypothesis, which we take as a given in order to apply elementary probability theory.

(3) The assumption of \( W(0) = 0 \) is simply a normalizing assumption.

(4) The assumption of continuity is a mathematical abstraction of the collected data, but it makes sense. Securities trade second by second or minute by minute so prices jump discretely by small amounts. Examined on a scale of day by day or week by week, then the short-time changes are tiny and in comparison prices seem to change continuously.

At least as a first assumption, we will try to use Brownian motion as a model of stock price movements. Remember the mathematical modeling proverb quoted earlier: All mathematical models are wrong, some mathematical models are useful. The Brownian motion model of stock prices is at least moderately useful.

Section Ending Answer. Two examples of mathematical objects defined by properties would be a linear operator and a group. A linear operator is a function on a vector space defined by the two properties that the value of a sum is the sum of the values, and the value of a scalar times a vector is the scalar times the function value. A group is defined by a set with an closed operation that is associative and has an identity and inverses.

Algorithms, Scripts, Simulations.

Algorithm. Simulate a sample path of the Wiener process as follows, see [23]. Divide the interval \([0, T]\) into a grid \(0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = T\) with \(t_{i+1} - t_i = \Delta t\). Set \(i = 1\) and \(W(0) = W(t_0) = 0\) and iterate the following algorithm.

1. Generate a new random number \(z\) from the standard normal distribution.
2. Set \(i\) to \(i + 1\).
3. Set \(W(t_i) = W(t_{i-1}) + z\sqrt{\Delta t}\).
4. If \(i < N\), iterate from step 1.

This method of approximation is valid only on the points of the grid. In between any two points \(t_i\) and \(t_{i-1}\), the Wiener process is approximated by linear interpolation.
5. BROWNIAN MOTION

```r
1 p <- 0.5
2 N <- 400
3 T <- 1
4 S <- array(0, c(N+1))
5 rw <- cumsum(2 * (runif(N) <= p) - 1)
6 S[2:(N+1)] <- rw
7 WcaretN <- function(x) {
8 Delta <- T/N
9   prior = floor(x/Delta) + 1
10  subsequent = ceiling(x/Delta) + 1
11  retval <- sqrt(Delta)*(S[prior] + ((x/Delta + 1) - prior)*(S[subsequent] - S[prior]))
12 }
13 plot(WcaretN, 0, 1, n=400)
```

Key Concepts.

(1) We define Brownian motion in terms of the normal distribution of the increments, the independence of the increments, the value at 0, and its continuity.

(2) The joint density function for the value of Brownian motion at several times is a multivariate normal distribution.

Vocabulary.

(1) **Brownian motion** is the physical phenomenon named after the English botanist Robert Brown who discovered it in 1827. Brownian motion is the zig-zagging motion exhibited by a small particle, such as a grain of pollen, immersed in a liquid or a gas. Albert Einstein gave the first explanation of this phenomenon in 1905. He explained Brownian motion by assuming the immersed particle was constantly buffeted by the molecules of the surrounding medium. Since then the abstracted process has been used for modeling the stock market and in quantum mechanics.

(2) The **Wiener process** is the mathematical definition and abstraction of the physical process as a stochastic process. The American mathematician Norbert Wiener gave the definition and properties in a series of papers starting in 1918. Generally, the terms **Brownian motion** and **Wiener process** are the same, although Brownian motion emphasizes the physical aspects and Wiener process emphasizes the mathematical aspects.

(3) **Bachelier process** means the same as Brownian motion and Wiener process. In 1900, Louis Bachelier introduced the limit of random walk as a model for prices on the Paris stock exchange, and so is the originator of the
mathematical idea now called Brownian motion. This term is occasionally found in financial literature.

Problems.

Exercise 5.5. Let \( W(t) \) be standard Brownian motion.

(1) Find the probability that \( 0 < W(1) < 1 \).
(2) Find the probability that \( 0 < W(1) < 1 \) and \( 1 < W(2) - W(1) < 3 \).
(3) Find the probability that \( 0 < W(1) < 1 \) and \( 1 < W(2) - W(1) < 3 \) and \( 0 < W(3) - W(2) < 1/2 \).

Exercise 5.6. Let \( W(t) \) be standard Brownian motion.

(1) Find the probability that \( 0 < W(1) < 1 \).
(2) Find the probability that \( 0 < W(1) < 1 \) and \( 1 < W(2) < 3 \).
(3) Find the probability that \( 0 < W(1) < 1 \) and \( 1 < W(2) < 3 \) and \( 0 < W(3) < 1/2 \).
(4) Explain why this problem is different from the previous problem, and also explain how to numerically evaluate to the probabilities.

Exercise 5.7. Write the joint probability density function for \( W(t_1) = x_1 \) and \( W(t_2) = x_2 \) explicitly.

Exercise 5.8. Let \( W(t) \) be standard Brownian motion.

(1) Find the probability that \( W(5) \leq 3 \) given that \( W(1) = 1 \).
(2) Find the number \( c \) such that \( P[W(9) > c | W(1) = 1] = 0.10 \).

Exercise 5.9. Let \( Z \) be a normally distributed random variable, with mean 0 and variance 1, \( Z \sim N(0,1) \). Then consider the continuous time stochastic process \( X(t) = \sqrt{t}Z \). Show that the distribution of \( X(t) \) is normal with mean 0 and variance \( t \). Is \( X(t) \) a Brownian motion?

Exercise 5.10. Let \( W_1(t) \) be a Brownian motion and \( W_2(t) \) be another independent Brownian motion, and \( \rho \) is a constant between \(-1\) and \(1\). Then consider the process \( X(t) = \rho W_1(t) + \sqrt{1-\rho^2} W_2(t) \). Is this \( X(t) \) a Brownian motion?

Exercise 5.11. What is the distribution of \( W(s) + W(t) \), for \( 0 \leq s \leq t \)? (Hint: Note that \( W(s) \) and \( W(t) \) are not independent. But you can write \( W(s) + W(t) \) as a sum of independent variables. Done properly, this problem requires almost no calculation.)

Exercise 5.12. For two random variables \( X \) and \( Y \), statisticians call

\[
\text{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]
\]

the covariance of \( X \) and \( Y \). If \( X \) and \( Y \) are independent, then \( \text{Cov}[X,Y] = 0 \). A positive value of \( \text{Cov}[X,Y] \) indicates that \( Y \) tends to increases as \( X \) does, while a negative value indicates that \( Y \) tends to decrease when \( X \) increases. Thus, \( \text{Cov}[X,Y] \) is an indication of the mutual dependence of \( X \) and \( Y \). Show that

\[
\text{Cov}[W(s),W(t)] = E[W(s)W(t)] = \min(t,s)
\]

Exercise 5.13. Show that the probability density function

\[
p(t;x,y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(x-y)^2}{2t}\right)
\]
satisfies the partial differential equation for heat flow (the heat equation)
\[
\frac{\partial p}{\partial t} = \frac{1}{2} \frac{\partial^2 p}{\partial x^2}
\]

**Exercise 5.14.** Change the scripts to simulate the Wiener process
(1) over intervals different from [0, 1], both longer and shorter,
(2) with more grid points, that is, smaller increments,
(3) with several simulations on the same plot.
Discuss how changes in the parameters of the simulation change the Wiener process.

**Exercise 5.15.** Choose a stock index such as the S & P 500, the Wilshire 5000, etc., and obtain closing values of that index for a year-long (or longer) interval of trading days. Find the variance of the closing values and create a random walk on the same interval with the same initial value and variance. Plot both sets of data on the same axes, as in Figure 2. Discuss the similarities and differences.

**Exercise 5.16.** Choose an individual stock or a stock index such as the S & P 500, the Wilshire 5000, etc., and obtain values of that index at regular intervals such as daily or hourly for a long interval of trading. Find the regular differences, and normalize by subtracting the mean and dividing by the standard deviation. Simultaneously plot a histogram of the differences and the standard normal probability density function. Discuss the similarities and differences.

### 5.3. Approximation of Brownian Motion by Coin-Flipping Sums

**Section Starter Question.** Suppose you know the graph \( y = f(x) \) of the function \( f(x) \). What is the effect on the graph of the transformation \( f(ax) \) where \( a > 1 \)? What is the effect on the graph of the transformation \( (1/b)f(x) \) where \( b > 1 \)? What about the transformation \( f(ax)/b \) where \( a > 1 \) and \( b > 1 \) ?

**Approximation of Brownian Motion by Fortunes.** As we have now assumed many times, for \( i \geq 1 \) let
\[
Y_i = \begin{cases} 
+1 & \text{with probability } 1/2 \\
-1 & \text{with probability } 1/2 
\end{cases}
\]
be a sequence of independent, identically distributed Bernoulli random variables. Note that \( \text{Var}[Y_i] = 1 \), which we will need to use in a moment. Let \( Y_0 = 0 \) for convenience and let
\[
T_n = \sum_{i=0}^{n} Y_i
\]
be the sequence of sums which represent the successive net fortunes of our notorious gambler. Sketch the random fortune \( T_n \) versus time using linear interpolation between the points \((n - 1, T_{n-1})\) and \((n, T_n)\) to obtain a continuous, piecewise linear function. The interpolation defines a function \( \hat{W}(t) \) defined on \([0, \infty)\) with \( \hat{W}(n) = T_n \). This function is piecewise linear with segments of length \( \sqrt{2} \). The notation \( \hat{W}(t) \) reminds us of the piecewise linear nature of the function.

We will compress time, and rescale the space in a special way. Let \( N \) be a large integer, and consider the rescaled function
\[
\hat{W}_N(t) = \left( \frac{1}{\sqrt{N}} \right) \hat{W}(Nt).
\]
This has the effect of taking a step of size ±1/√N in 1/N time unit. For example,

\[ \hat{W}_N(1/N) = \left( \frac{1}{\sqrt{N}} \right) W(N \cdot 1/N) = \frac{T_1}{\sqrt{N}} = \frac{Y_1}{\sqrt{N}}. \]

Now consider

\[ \hat{W}_N(1) = \frac{\hat{W}(N \cdot 1)}{\sqrt{N}} = \frac{\hat{W}(N)}{\sqrt{N}} = \frac{T_N}{\sqrt{N}}. \]

According to the Central Limit Theorem, this quantity is approximately normally distributed, with mean zero, and variance 1. More generally,

\[ \hat{W}_N(t) = \frac{\hat{W}(Nt)}{\sqrt{N}} = \sqrt{t} \frac{\hat{W}(Nt)}{\sqrt{Nt}}. \]

If Nt is an integer, \( \hat{W}_N(t) \) is approximately normally distributed with mean 0 and variance \( t \). Furthermore, \( \hat{W}_N(0) = 0 \) and \( \hat{W}_N(t) \) is a continuous function, and so is continuous at 0. At times \( t_j = j/N, t_k = k/N, t_\ell = \ell/N, \) and \( t_m = m/N \) with \( t_j < t_k \leq t_\ell < t_m \) (so \( j < k \leq \ell < m \)) the function differences \( \hat{W}_N(t_k) - \hat{W}_N(t_j) \) and \( \hat{W}_N(t_m) - \hat{W}_N(t_\ell) \) are the differences \( (T_k - T_j)/\sqrt{N} \) and \( (T_m - T_\ell)/\sqrt{N} \), hence independent.

Altogether, this should be a strong suggestion that \( \hat{W}_N(t) \) is an approximation to standard Brownian motion. We will define the very jagged piecewise linear function \( \hat{W}_N(t) \) as approximate Brownian motion.

**Theorem 5.17.** The limit of the rescaled random walk defining approximate Brownian motion is Brownian motion in the following sense:

1. \[ \mathbb{P} \left[ \hat{W}_N(t) < x \right] \to \mathbb{P} \left[ W(t) < x \right] \text{ as } N \to \infty. \]

2. More generally, the limit of the rescaled random walk defining approximate Brownian motion is Brownian motion in the following sense:

\[ \mathbb{P} \left[ \hat{W}_N(t_1) < x_1, \hat{W}_N(t_2) < x_2, \ldots, \hat{W}_N(t_n) < x_n \right] \to \mathbb{P} \left[ W(t_1) < x_1, W(t_2) < x_2, \ldots, W(t_n) < x_n \right] \]

as \( N \to \infty \) where \( t_1 < t_2 < \cdots < t_n \). That is, the joint distributions of \( \hat{W}_N(t) \) converges to the joint normal distribution

\[ f(x_1, t_1; x_2, t_2; \ldots; x_n, t_n) = p(x_1, t_1)p(x_2 - x_1, t_2 - t_1) \ldots p(x_n - x_{n-1}, t_n - t_{n-1}) \]

of the standard Brownian motion.

With some additional foundational work, a mathematical theorem establishes that the rescaled fortune processes actually converge to the mathematical object called the standard Brownian motion as defined in the previous section. The proof of this mathematical theorem is beyond the scope of a text of this level, but the theorem above should strongly suggest how this can happen, and give some intuitive feel for the approximation of Brownian motion through the rescaled coin-flip process.

Using a scaled random walk is not the only way to approximate Brownian motion. Other approximations of the Wiener process use “global” approximations such as Fourier series (or more generally \( L^2[0, T] \) expansions in an orthogonal basis)
or Bernstein polynomials. The Fourier series representation is also known as the Karhunen-Loève expansion of the Wiener process; for elementary details see [23]. For the representation as a limit of Bernstein polynomials, see [35]. Both of these approximations use ideas from probability theory and analysis which are beyond the scope of this book. When one only needs to simulate the position of a sample path of Brownian motion at one or even several time points, then the scaled random walk approximation is simple and the accuracy can be estimated with the Central Limit Theorem or the Berry-Esseen Theorem. If one needs information on a whole sample path of Brownian motion, then the “global” methods are more appropriate approximations. The “global” methods are not only more mathematically sophisticated, they also are more expensive in terms of processing and rely on the efficiency of the underlying implementations of the mathematical functions used.

**Section Ending Answer.** The effect on the graph of the transformation \( f(ax) \) where \( a > 1 \) is to rescale the horizontal axis, compressing the graph horizontally. The effect on the graph of the transformation \( (1/b)f(x) \) where \( b > 1 \) is to rescale the vertical axis, compressing the graph vertically. The transformation \( f(ax)/b \) where \( a > 1 \) and \( b > 1 \) makes both compressions simultaneously. This dual effect rescales the random walk or gambler’s fortune into an approximation of Brownian motion.

**Algorithms, Scripts, Simulations.**

**Algorithm.** Simulate a sample path of the Wiener process as follows, see [23]. Divide the interval \([0, T]\) into a grid of \( N + 1 \) nodes \( 0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = T \) with \( t_{i+1} - t_i = \Delta \). The nodes may be indexed from either 0 to \( N \) or from 1 to \( N + 1 \) depending on the language. Create a Bernoulli random walk \( T_0, T_1, T_2, \ldots, T_N \) of \( N + 1 \) steps with \( T_0 = 0 \). Given a value \( x \in [0, T] \) the prior node with \( t_k \leq x \) is \( \lfloor x/\Delta \rfloor \) for 0-based arrays (or \( \lceil x/\Delta \rceil \) for 1-based arrays.). The subsequent node with \( x \leq t_{k+1} \) is \( \lceil x/\Delta \rceil \) for 0-based arrays (or \( \lfloor x/\Delta \rfloor + 1 \) for 1-based arrays.). Then define the value of the approximation function \( \hat{W}_N(x) \) by linear interpolation between the values of the random walk at \( T_k \) and \( T_{k+1} \).

A feature of this \( N + 1 \)-step random walk scaling approximation algorithm is that it creates the approximation as a function on \([0, T]\). This function can then be plotted with a function plotting routine on any time grid on the interval \([0, T]\). If the time grid used for plotting on \([0, T]\) has less than \( N \) points, then some of the information in the \( N \)-step scaling approximation is ignored, and the plotted function will be less representative of the approximation than it could be. If the time grid on \([0, T]\) is greater than \( N \) points, then the plotted function will just represent the linear interpolation between the random-walk points at \( t_j = jT/N \) and no new information is represented.

Depending on the internal plotting routines used by the language, plotting the approximation function \( \hat{W}_N(t) \) can result in plot artifacts. One simple artifact may be horizontal segments in the plot. If the plotting algorithms attempt to use adaptive point selection to densely position a greater portion of a fixed number of plotting points in a region of rapid variation, then other regions will have fewer plotting points. Those regions with fewer plotting points will miss some of the information in that region. Depending on the language, the plotting routine may use smoothing or some other nonlinear interpolation between plotting points which
will result in curved segments instead of a piecewise linear function. If the intention is to plot an approximate Brownian motion, then there are more direct and efficient ways to create and plot the \( N + 1 \) coordinate pairs \((jT/N, \sqrt{T/N}S_j)\) defining the vertices of the piecewise linear scaled random walk approximation with an appropriate amount of information. Here the intention is to first to demonstrate the creation of the approximation function as a piecewise linear function, then second to use the function to plot a graph.

```r
1 p <- 0.5
2 N <- 400
3
4 T <- 1
5
6 S <- array(0, c(N+1))
7 rw <- cumsum( 2 * (runif(N) <= p) -1 )
8 S[2:(N+1)] <- rw
9
10 WcaretN <- function(x) {
11   Delta <- T/N
12   # add 1 since arrays are 1-based
13   prior = floor(x/Delta) + 1
14   subsequent = ceiling(x/Delta) + 1
15   retval <- sqrt(Delta)*(S[prior] + ((x/Delta+1) - prior)*(S[ subsequent] - S[prior]))
16 }
17
18 plot(WcaretN, 0,1, n =400)
```

**Key Concepts.**

1. A properly scaled “random fortune” process (i.e. random walk) approximates Brownian motion.
2. Brownian motion is the limit of “random fortune” discrete time processes (i.e. random walks), properly scaled. The study of Brownian motion is therefore an extension of the study of random fortunes.

**Vocabulary.**

1. We define approximate Brownian motion \( \hat{W}_N(t) \) to be the rescaled random walk with steps of size \( 1/\sqrt{N} \) taken every \( 1/N \) time units where \( N \) is a large integer.

**Problems.**

**Exercise 5.18.** Flip a coin 25 times, recording whether it comes up Heads or Tails each time. Scoring \( Y_i = +1 \) for each Heads and \( Y_i = -1 \) for each flip, also keep track of the accumulated sum \( T_n = \sum_{i=1}^{n} T_i \) for \( i = 1 \ldots 25 \) representing the net fortune at any time. Plot the resulting \( T_n \) versus \( n \) on the interval \([0, 25]\).
Finally, using $N = 5$, plot the rescaled approximation $\hat{W}_5(t) = (1/\sqrt{5})T(5t)$ on the interval $[0, 5]$ on the same graph.

**Exercise 5.19.**
1. What are the slopes of the linear segments of $\hat{W}_N(t)$?
2. Is the slope necessarily undefined at the nodes defining $\hat{W}_N(t)$?
3. Explain how a plotting routine could create a horizontal segment in the plot of $\hat{W}_N(t)$.

**Exercise 5.20.** Modify the script in the following ways:
1. Plot more than one Wiener process sample path on the same set of axes.
2. Change $p$ to a value greater than 0.5 and describe the effect on the sample paths. Change $p$ to a value less than 0.5 and describe the effect on the sample paths.
3. Plot the approximation functions for increasing values of $N$ and describe the effects on the sample paths.
4. For a fixed, large value of $N$, plot the approximation at increasing values of the number of plotting points which are not divisors of $N$ and describe the effects on the sample paths.
5. Change the value of $T$ and plot the resulting sample paths, both without increasing the value of $N$ and increasing the value of $N$ proportionally to $T$ and describe the resulting sample paths.

**Exercise 5.21.** Iacus [23] uses a different random walk step function approximation to Brownian motion:

\[
\bar{W}_N(t) = \frac{T\lfloor Nt \rfloor}{\sqrt{N}}.
\]

Create scripts that plot this random walk approximation for $N = 10$, $N = 100$ and $N = 1000$.

**Exercise 5.22.** Under the assumption of the theorem that the joint distributions of $\hat{W}_N(t)$ converges to the joint normal distribution is true, show then that

\[
P \left[ \frac{T\lfloor Nt \rfloor}{\sqrt{N}} < x \right] \to P [W(t) < x].
\]

### 5.4. Transformations of the Wiener Process

**Section Starter Question.** Suppose you know the graph $y = f(x)$ of the function $f(x)$. What is the effect on the graph of the transformation $f(x+h) - f(h)$? What is the effect on the graph of the transformation $f(1/x)$? Consider the function $f(x) = \sin(x)$ as an example.

**Transformations of the Wiener Process.** A set of transformations of the Wiener process produce the Wiener process again. Since these transformations result in the Wiener process, each tells us something about the “shape” and “characteristics” of the Wiener process. These results are especially helpful when studying the properties of the Wiener process sample paths. The first of these transformations is a time homogeneity that says the Wiener process can be re-started anywhere. The second says that the Wiener process may be rescaled in time and space. The third is an inversion. Roughly, each of these says the Wiener process is self-similar in various ways. See the comments after the proof for more detail.
Theorem 5.23. Let

(1) \( W_{\text{shift}}(t) = W(t + h) - W(h) \), for fixed \( h > 0 \).
(2) \( W_{\text{scale}}(t) = cW(t/c^2) \), for fixed \( c > 0 \).

Then each of \( W_{\text{shift}}(t) \) and \( W_{\text{scale}}(t) \) are a version of the standard Wiener process.

Proof. We have to systematically check each of the defining properties of the Wiener process in turn for each of the transformed processes.

(1) \( W_{\text{shift}}(t) = W(t + h) - W(h) \).

(a) The increment is
\[
W_{\text{shift}}(t + s) - W_{\text{shift}}(s) = [W(t + s + h) - W(h)] - [W(s + h) - W(h)] = W(t + s + h) - W(s + h)
\]
which is by definition normally distributed with mean 0 and variance \( t \).

(b) The increment
\[
W_{\text{shift}}(t_4) - W_{\text{shift}}(t_3) = W(t_4 + h) - W(t_3 + h)
\]
is independent from
\[
W_{\text{shift}}(t_2) - W_{\text{shift}}(t_1) = W(t_2 + h) - W(t_1 + h)
\]
by the property of independence of disjoint increments of \( W(t) \).

(c) \( W_{\text{shift}}(0) = W(0 + h) - W(h) = 0 \).

(d) As the composition and difference of continuous functions, \( W_{\text{shift}} \) is continuous.

(2) \( W_{\text{scale}}(t) = cW(t/c^2) \)

(a) The increment
\[
W_{\text{scale}}(t) - W_{\text{scale}}(s) = cW(t/c^2) - cW(s/c^2) = c(W(t/c^2) - W(s/c^2))
\]
is normally distributed because it is a multiple of a normally distributed random variable. Since the increment \( W(t/c^2) - W(s/c^2) \) has mean zero, then
\[
W_{\text{scale}}(t) - W_{\text{scale}}(s) = c(W(t/c^2) - W(s/c^2))
\]
must have mean zero. The variance is
\[
E \left[ (W_{\text{scale}}(t) - W(s))^2 \right] = E \left[ (cW(t/c^2) - cW(s/c^2))^2 \right] = c^2 E \left[ (W(t/c^2) - W(s/c^2))^2 \right] = c^2(t/c^2 - s/c^2) = t - s.
\]

(b) Note that if \( t_1 < t_2 < t_3 < t_4 \), then \( t_1/c^2 < t_2/c^2 < t_3/c^2 < t_4/c^2 \), and the corresponding increments \( W(t_4/c^2) - W(t_3/c^2) \) and \( W(t_2/c^2) - W(t_1/c^2) \) are independent. Then the multiples of each by \( c \) are independent and so \( W_{\text{scale}}(t_4) - W_{\text{scale}}(t_3) \) and \( W_{\text{scale}}(t_2) - W_{\text{scale}}(t_1) \) are independent.
(c) \( W_{\text{scale}}(0) = cW(0/c^2) = cW(0) = 0 \).

(d) As the composition of continuous functions, \( W_{\text{scale}} \) is continuous.

\[ \square \]

**Theorem 5.24.** Suppose \( W(t) \) is a standard Wiener process. Then the transformed processes \( W_{\text{inv}}(t) = tW(1/t) \) for \( t > 0 \), \( W_{\text{inv}}(t) = 0 \) for \( t = 0 \) is a version of the standard Wiener process.

**Proof.** To show that \( W_{\text{inv}}(t) = tW(1/t) \) is a Wiener process by the four defining properties requires another fact that is outside the scope of the text. The fact is that any Gaussian process \( X(t) \) with mean 0 and Cov \( [X(s), X(t)] = \min(s, t) \) must be the Wiener process. See the references for more information. Using this information, a partial proof follows:

(1) \( W_{\text{inv}}(t) - W_{\text{inv}}(s) = tW(1/t) - sW(1/s) \)

is the difference of normally distributed random variables each with mean 0, so the difference will be normal with mean 0. It remains to check that the normal random variable has the correct variance.

\[
E \left[ (W_{\text{inv}}(t) - W_{\text{inv}}(s))^2 \right] = E \left[ (sW(1/s) - tW(1/t))^2 \right] = E[(sW(1/s) - sW(1/t)] + sW(1/t) - tW(1/t) - (s - t)W(0))^2] = s^2E \left[ (W(1/s) - W(1/t))^2 \right] + 2s(s - t)E \left[(W(1/s) - W(1/t))(W(1/t) - W(0))] + (s - t)^2E \left[(W(1/t) - W(0))^2 \right] = s^2E \left[ (W(1/s) - W(1/t))^2 \right] + (s - t)^2E \left[(W(1/t) - W(0))^2 \right] = s^2(1/s - 1/t) + (s - t)^2(1/t) = t - s.
\]

Note the use of independence of \( W(1/s) - W(1/t) \) from \( W(1/t) - W(0) \) at the third equality.

(2) It is hard to show the independence of increments directly. Instead rely on the fact that a Gaussian process with mean 0 and covariance function \( \min(s, t) \) is a Wiener process, and thus prove it indirectly.

Note that

\[
\text{Cov} \left[ W_{\text{inv}}(s), W_{\text{inv}}(t) \right] = st \cdot \min(1/s, 1/t) = \min(s, t).
\]

(3) By definition, \( W_{\text{inv}}(0) = 0 \).

(4) The argument that \( \lim_{t \to 0} W_{\text{inv}}(t) = 0 \) is equivalent to showing that \( \lim_{t \to \infty} W(t)/t = 0 \). To show this requires use of Kolmogorov’s inequality for the Wiener process and clever use of the Borel-Cantelli lemma and is beyond the scope of this course. Use the translation property in the third statement of this theorem to prove continuity at every value of \( t \).

\[ \square \]
The following comments are adapted from *Stochastic Calculus and Financial Applications* by J. Michael Steele. Springer, New York, 2001, page 40. These laws tie the Wiener process to three important groups of transformations on \([0, \infty)\), and a basic lesson from the theory of differential equations is that such symmetries can be extremely useful. On a second level, the laws also capture the somewhat magical fractal nature of the Wiener process. The scaling law tells us that if we had even one-billionth of a second of a Wiener process path, we could expand it to a billions years’ worth of an equally valid Wiener process path! The translation symmetry is not quite so startling, it merely says that Wiener process can be restarted anywhere. That is, any part of a Wiener process captures the same behavior as at the origin. The inversion law is perhaps most impressive, it tells us that the first second of the life of a Wiener process path is rich enough to capture the behavior of a Wiener process path from the end of the first second until the end of time.

**Section Ending Answer.** The transformation \(f(x + h) - f(h)\) translates the graph of \(f(x)\) so that the point \((h, f(h))\) is at the origin. The transformation \(f(1/x)\) inverts the graph so that values on the interval \([1, \infty)\) are compressed to the interval \((0, 1]\) and values on the interval \((0, 1]\) are expanded to the interval \([1, \infty)\). The inverted function \(\sin(1/x)\) is frequently used as the basis for counterexamples in analysis.

**Algorithms, Scripts, Simulations.**

**Algorithm.** Let constants \(h\) and \(c\) be given. Create an approximation \(\hat{W}(t)\) of Wiener process \(W(t)\) on a domain \([0, T]\) large enough to accommodate \(\hat{W}(1 + h)\) and \(\hat{W}(1/c^2)\). Using the approximation \(\hat{W}(t)\) create approximations to \(W_{\text{shift}}(t)\) and \(W_{\text{scale}}(t)\) with functions \(\hat{W}(t + h) - \hat{W}(h)\) and \(c\hat{W}(t/c^2)\). Plot all functions on the same set of axes.

```r
1 p <- 0.5
2 N <- 400
3
4 T <- 2
5 h <- 0.25
6 c <- 2.0
7
8 S <- array(0, c(N+1))
9 rw <- cumsum( 2 * ( runif(N) <= p ) - 1 )
10 S[2:(N+1)] <- rw
11
12 WcaretN <- function(x) {
13   Delta <- T/N
14   # add 1 since arrays are 1-based
15   prior = floor(x/Delta) + 1
16   subsequent = ceiling(x/Delta) + 1
17   retval <- sqrt(Delta) * (S[prior] + ((x/Delta + 1) - prior) * (S[subsequent] - S[prior]))
18 }
19```

5. BROWNIAN MOTION

```r
Wshift <- function(x) {
  retval <- WcaretN(x+h) - WcaretN(h)
}
Wscale <- function(x) {
  retval <- c* WcaretN(x/c ^2)
}
curve ( WcaretN , 0,1, n =400 , col = "black ")
curve ( Wshift , 0,1, n =400 , add = TRUE , col = "blue ")
curve ( Wscale , 0,1, n= 400 , add = TRUE , col = "red ")
```

### Problems.

**Exercise 5.25.** Explain why there is no script for simulation or approximation of the inversion transformation of the Wiener process, or if possible provide such a script.

**Exercise 5.26.** Given the piecewise linear approximation $\hat{W}_N(t)$, what are the slopes of the piecewise linear segments of the scaling transformation $c\hat{W}_N(t/c^2)$?

**Exercise 5.27.** Modify the scripts to plot an approximation of $W_{scale}(t)$ on $[0,1]$ with the same degree of approximation as $\hat{W}_N(t)$ for some $N$. Plot both on the same set of axes.

**Exercise 5.28.** Show that $st \min(1/s,1/t) = \min(s,t)$

### 5.5. Hitting Times and Ruin Probabilities

**Section Starter Question.** What is the probability that a simple random walk with $p = 1/2 = q$ starting at the origin will hit value $a > 0$ before it hits value $-b < 0$, where $b > 0$? What do you expect in analogy for the standard Wiener process and why?

**Hitting Times.** Consider the standard Wiener process $W(t)$, which starts at $W(0) = 0$. Let $a > 0$. The hitting time $T_a$ is the first time the Wiener process hits $a$. In notation from analysis

$$T_a = \inf\{t > 0 : W(t) = a\}.$$

Note the very strong analogy with the duration of the game in the gambler’s ruin.

Some Wiener process sample paths will hit $a > 0$ fairly directly. Others will make an excursion to negative values and take a long time to finally reach $a$. Thus $T_a$ will have a probability distribution. We determine that probability distribution by a heuristic procedure similar to the first step analysis we made for coin-flipping fortunes.

Specifically, we consider a probability by conditioning, that is, conditioning on whether or not $T_a \leq t$, for some given value of $t$

$$P[W(t) \geq a] = P[W(t) \geq a | T_a \leq t]P[T_a \leq t] + P[W(t) \geq a | T_a > t]P[T_a > t].$$

Now note that the second conditional probability is 0 because it is an empty event. Therefore:

$$P[W(t) \geq a] = P[W(t) \geq a | T_a \leq t]P[T_a \leq t].$$
Now, consider Wiener process “started over” again at the time $T_a$ when it hits $a$. By the shifting transformation from the previous section, the “started-over” process has the distribution of Wiener process again, so

$$
\mathbb{P}[W(t) \geq a \mid T_a \leq t] = \mathbb{P}[W(t) \geq a \mid W(T_a) = a, T_a \leq t]
= \mathbb{P}[W(t) - W(T_a) \geq 0 \mid T_a \leq t]
= 1/2.
$$

This argument is a specific example of the Reflection Principle for the Wiener process. It says that the Wiener process reflected about a first passage has the same distribution as the original motion. Thus

$$
\mathbb{P}[W(t) \geq a] = (1/2)\mathbb{P}[T_a \leq t].
$$

or

$$
\mathbb{P}[T_a \leq t] = 2\mathbb{P}[W(t) \geq a]
= \frac{2}{\sqrt{2\pi t}} \int_0^{\infty} \exp(-u^2/(2t)) \, du
= \frac{2}{\sqrt{2\pi}} \int_{a/\sqrt{t}}^{\infty} \exp(-v^2/2) \, dv
$$

(note the change of variables $v = u/\sqrt{t}$ in the second integral) and so we have derived the c.d.f. of the hitting time random variable. One can easily differentiate to obtain the p.d.f

$$
f_{T_a}(t) = \frac{a}{\sqrt{2\pi}} t^{-3/2} \exp(-a^2/(2t)).
$$

Actually, this argument contains a serious logical gap, since $T_a$ is a random time, not a fixed time. That is, the value of $T_a$ is different for each sample path, it varies with $\omega$. On the other hand, the shifting transformation defined in the prior section depends on having a fixed time, called $h$ in that section. To fix this logical gap, we must make sure that “random times” act like fixed times. Under special conditions, random times can act like fixed times. Specifically, this proof can be fixed and made completely rigorous by showing that the standard Wiener process has the strong Markov property and that $T_a$ is a Markov time corresponding to the event of first passage from 0 to $a$.

Note that deriving the p.d.f. of the hitting time is much stronger than the analogous result for the duration of the game until ruin in the coin-flipping game. There we were only able to derive an expression for the expected value of the hitting time, not the probability distribution of the hitting time. Now we are able to derive the probability distribution of the hitting time fairly intuitively (although strictly speaking there is a gap). Here is a place where it is simpler to derive a quantity for Wiener process than it is to derive the corresponding quantity for random walk.

Let us now consider the probability that the Wiener process hits $a > 0$, before hitting $-b < 0$ where $b > 0$. To compute this we will make use of the interpretation of the standard Wiener process as being the limit of the symmetric random walk. Recall from the exercises following the section on the gambler’s ruin in the fair ($p = 1/2 = q$) coin-flipping game that the probability that the random walk goes
up to value $a$ before going down to value $b$ when the step size is $\Delta x$ is

$$P \left[ \text{to } a \text{ before } -b \right] = \frac{b \Delta x}{(a + b) \Delta x} = \frac{b}{a + b}$$

Thus, the probability of hitting $a > 0$ before hitting $-b < 0$ does not depend on the step size, and also does not depend on the time interval. Therefore, passing to the limit in the scaling process for random walks, the probabilities should remain the same. Here is a place where it is easier to derive the result from the coin-flipping game and pass to the limit than to derive the result directly from Wiener process principles.

**The Distribution of the Maximum.** Let $t$ be a given time, let $a > 0$ be a given value, then

$$P \left[ \max_{0 \leq u \leq t} W(u) \geq a \right] = P [T_a \leq t]$$

$$= \frac{2}{\sqrt{2\pi}} \int_{a/\sqrt{t}}^{\infty} \exp(-y^2/2) \, dy$$

**Algorithms, Scripts, Simulations.**

**Algorithm.** Set a time interval length $T$ sufficiently long to have a good chance of hitting the fixed value $a$ before fixed time $t < T$. Set a large value $n$ for the length of the random walk used to create the Wiener process, then fill an $n \times k$ matrix with Bernoulli random variables. Cumulatively sum the Bernoulli random variables to create a scaled random walk approximating the Wiener process. For each random walk, find when the hitting time is encountered. Also find the maximum value of the scaled random walk on the interval $[0, t]$. Since the approximation is piecewise linear, only the nodes need to be examined. Compute the fraction of the $k$ walks which have a hitting time less than $t$ or a maximum greater than $a$ on the interval $[0, t]$. Compare the fraction to the theoretical value.

**Technical Note:** The careful reader will note that the hitting time $T_a = \inf\{t > 0 : W(t) = a\}$ and the events $[T_a \leq t]$ and $[\max_{0 \leq u \leq t} W(u) \geq a]$ are “global” events on the set of all Wiener process paths. However, the definition of Wiener process only has prescribed probability distributions on the values at specified times. Implicit in the definition of the Wiener process is a probability distribution on the “global” set of Wiener processes, but the proof of the existence of the probability distribution is beyond the scope of this text. Moreover, there is no easy distribution function to compute the probability of such events as there is with the normal probability distribution.

The algorithm approximates the probability by counting how many of $k$ scaled binomial random variables hit a value greater than or equal to $a$ at a node which corresponds to a time less than or equal to $t$. Convergence of the counting probability to the global Wiener process probability requires justification with Donsker’s Invariance Principle. The principle says that the piecewise linear processes $\hat{W}_N(t)$ converge in distribution to the Wiener process.

To apply Donsker’s Principle, consider the functional

$$\phi(f) = h(\max_{0 \leq u \leq t} f(u))$$
where \( h(\cdot) \) is a bounded continuous function. The convergence in distribution from the Invariance Principle implies that

\[
E[\phi(W_N(t))] \rightarrow E[\phi W].
\]

Now taking \( h(\cdot) \) to be an approximate indicator function on the interval \([a, \infty)\) shows that the counting probability converges to the global Wiener process probability. The details require careful analysis.

```r
1 T <- 10
2 a <- 1
3 time <- 2
4
5 p <- 0.5
6 n <- 10000
7 k <- 1000
8
9 Delta = T/n
10
11 winLose <- 2 * (array( 0+(runif(n*k) <= p), dim=c(n,k))) - 1
12 # 0+ coerces Boolean to numeric
13 totals <- apply( winLose, 2, cumsum)
14
15 paths <- array( 0 , dim=c(n+1, k) )
16 paths[2:(n+1), 1:k] <- sqrt(Delta)*totals
17
18 hitIndex <- apply( 0+(paths <= a), 2, (function (x) match (0 , x, nomatch =n+2) ))
19 # If no hitting on a walk , nomatch=n+2 sets the hitting
20 # time to be two more than the number of steps, one more than
21 # the column length. Without the nomatch option, get NA which
22 # works poorly with the comparison
23
24 hittingTime = Delta*(hitIndex-1)
25 ## subtract 1 since vectors are 1-based
26 # subtract 1 since vectors are 1-based
27 probHitlessTa <- sum( 0+(hittingTime <= time))/k
28 probMax = sum( 0+( apply(paths[1:((time/Delta)+1)], 2, max ) >= a ) )/k
29 theoreticalProb = 2*pnorm(a/sqrt(time), lower=FALSE)
30
31 cat(sprintf("Empirical probability Wiener process paths hit %f before
32 %f: \%f \n", a, time, probHitlessTa ))
33 cat(sprintf("Empirical probability Wiener process paths greater than %
34 f before %f: \%f \n", a, time, probMax ))
35 cat(sprintf("Theoretical probability: %f \n", theoreticalProb ))
```

**Key Concepts.**

1. With the Reflection Principle, we can derive the p.d.f of the hitting time \( T_a \).
(2) With the hitting time, we can derive the c.d.f. of the maximum of the Wiener Process on the interval $0 \leq u \leq t$.

**Vocabulary.**

(1) The **Reflection Principle** says the Wiener process reflected about a first passage has the same distribution as the original motion.

(2) The **hitting time** $T_a$ is the first time the Wiener process assumes the value $a$. In notation from analysis

$$T_a = \inf\{t > 0 : W(t) = a\}.$$  

**Problems.**

(1) Differentiate the c.d.f. of $T_a$ to obtain the expression for the p.d.f of $T_a$.

(2) Show that $E[T_a] = \infty$ for $a > 0$.

(3) Suppose that the fluctuations of a share of stock of a certain company are well described by a Wiener process. Suppose that the company is bankrupt if ever the share price drops to zero. If the starting share price is $A(0) = 10$, what is the probability that the company is bankrupt by $t = 30$? What is the probability that the share price is above 20 at $t = 30$?

(4) Modify the scripts by setting $p > 0.5$ or $p < 0.5$. What happens to the hitting time?

(5) (a) Modify the scripts to plot the probability that the hitting time is less than or equal to $a$ as a function of $a$.

(b) Modify the scripts to plot the probability that the hitting time is less than or equal to $a$ as a function of $t$. On the same set of axes plot the theoretical probability as a function of $t$.

5.6. Path Properties of Brownian Motion

**Section Starter Question.** Provide an example of a continuous function that is not differentiable at some point. Why does the function fail to have a derivative at that point? What are the possible reasons that a derivative could fail to exist at some point?

**Non-differentiability of Brownian Motion paths.** Probability theory uses the term **almost surely** to indicate an event that occurs with probability 1. The complementary events occurring with probability 0 are sometimes called **negligible events**. In infinite sample spaces, it is possible to have meaningful events with probability zero. So to say an event occurs “almost surely” or is a negligible event is not an empty phrase.

**Theorem 5.29.** With probability 1 (i.e. almost surely) Brownian motion paths are continuous functions.

To state this as a theorem may seem strange in view of property 4 of the definition of Brownian motion. Property 4 requires that Brownian motion is continuous. However, some authors weaken property 4 in the definition to only require that Brownian motion be continuous at $t = 0$. Then this theorem shows that the weaker definition implies the stronger definition used in this text. This theorem is difficult to prove, and well beyond the scope of this course. In fact, even the statement above is imprecise. Specifically, there is an explicit representation of the defining properties of Brownian motion as a random variable $W(t, \omega)$ which is a continuous
function of $t$ with probability 1. We need the continuity for much of what we do later, and so this theorem is stated here as a fact without proof.

**Theorem 5.30.** With probability 1 (i.e. almost surely) a Brownian motion is nowhere (except possibly on set of Lebesgue measure 0) differentiable.

This property is even deeper and requires more understanding of analysis to prove than does the continuity theorem, so we will not prove it here. Rather, we use this fact as another piece of evidence of the strangeness of Brownian motion.

In spite of one’s intuition from calculus, Theorem 5.30 shows that continuous, nowhere differentiable functions are actually common. Indeed, continuous, nowhere differentiable functions are useful for stochastic processes. One can imagine non-differentiability by considering the function $f(t) = |t|$ which is continuous but not differentiable at $t = 0$. Because of the corner at $t = 0$, the left and right limits of the difference quotient exist but are not equal. Even more to the point, the function $t^{2/3}$ is continuous but not differentiable at $t = 0$ because of a sharp cusp there. The left and right limits of the difference quotient do not exist (more precisely, the left limit is $-\infty$ and the right limit is $+\infty$) at $x = 0$. One can imagine Brownian motion as being spiky with tiny cusps and corners at every point. This becomes somewhat easier to imagine by thinking of the limiting approximation of Brownian motion by scaled random walks. The re-scaled coin-flipping fortune graphs look spiky with many corners. The approximating graphs suggest why the theorem is true, although this is not sufficient for the proof.

**Properties of the Path of Brownian Motion.**

**Theorem 5.31.** With probability 1 (i.e. almost surely) a Brownian motion path has no intervals of monotonicity. That is, there is no interval $[a, b]$ with $W(t_2) - W(t_1) > 0$ (or $W(t_2) - W(t_1) < 0$) for all $t_2, t_1 \in [a, b]$ with $t_2 > t_1$.

**Theorem 5.32.** With probability 1 (i.e. almost surely) Brownian motion $W(t)$ has

$$\limsup_{n \to \infty} \frac{W(n)}{\sqrt{n}} = +\infty,$$

$$\liminf_{n \to \infty} \frac{W(n)}{\sqrt{n}} = -\infty.$$

From Theorem 5.32 and the continuity we can deduce that (with probability 1) for arbitrarily large $t_1$, there is a $t_2 > t_1$ such that $W(t_2) = 0$. That is, Brownian motion paths always cross the time-axis at some time greater than any arbitrarily large value of $t$. Equivalently, almost surely, Brownian motion never eventually stays in the upper half-plane (or lower half-plane).

**Theorem 5.33.** With probability 1 (i.e. almost surely), 0 is an accumulation point of the zeros of $W(t)$.

From Theorem 5.32 and the inversion $tW(1/t)$ also being a standard Brownian motion, we heuristically deduce that 0 is an accumulation point of the zeros of $W(t)$. That is, standard Brownian motion crosses the time axis arbitrarily often near 0.
Theorem 5.34. With probability 1 (i.e. almost surely) the zero set of Brownian motion
\[ \{ t \in [0, \infty) : W(t) = 0 \} \]
is an uncountable closed set with no isolated points.

Theorem 5.35. With probability 1 (i.e. almost surely) the graph of a Brownian motion path has Hausdorff dimension 3/2.

Roughly, this means that the graph of a Brownian motion path is “fuzzier” or “thicker” than the graph of, for example, a continuously differentiable function which would have Hausdorff dimension 1. In popular language, this theorem says that Brownian motion is a fractal.

Altogether these theorems indicate the bewildering strangeness of Brownian motion. Nevertheless, Brownian motion is still a useful mathematical object for modeling physical and economic processes.

Section Ending Answer. A continuous function that is not differentiable at some point is \( f(x) = |x| \) because the left and right limits of the difference quotient at 0 do not agree. Other possible reasons that a derivative could fail to exist at a point are that the difference quotients have unbounded limits, or the limit of the difference quotient does not exist because it oscillates.

Algorithms, Scripts, Simulations.
Algorithm. For a given value of \( p \) and number of steps \( N \) on a time interval \([0, T]\) create a scaled random walk \( \hat{W}_N(t) \) on \([0, T]\). Then for a given minimum increment \( h_0 \) up to a maximum increment \( h_1 \) create a sequence of equally spaced increments. Then at a fixed base-point, calculate the difference quotient for each of the increments. Plot the difference quotients versus the increments on a semi-logarithmic set of axes.

Because the difference quotients are computed using the scaled random walk approximation of the Wiener process, the largest possible slope is
\[ \sqrt{T/N}/(T/N) = \sqrt{N/T}. \]

So the plotted difference quotients will “max out” once the increment is less than the scaled random walk step size.

```r
1 p <- 0.5
2 N <- 1000
3
4 T <- 1
5
6 S <- array(0, c(N+1))
7 rw <- cumsum( 2 * ( runif(N) <= p)-1 )
8 S[2:(N+1)] <- rw
9
10 WcaretN <- function(x) {
11     Delta <- T/N
12
13     # add 1 since arrays are 1-based
14     prior = floor(x/Delta) + 1
15     subsequent = ceiling(x/Delta) + 1
16 ...```
Key Concepts.

(1) With probability 1 a Brownian motion path is continuous but nowhere differentiable.

(2) Although a Brownian motion path is continuous, it has many counterintuitive properties not usually associated with continuous functions.

Vocabulary.

(1) Probability theory uses the term *almost surely* to indicate an event that occurs with probability 1. The complementary events occurring with probability 0 are sometimes called *negligible events*. In infinite sample spaces, it is possible to have meaningful events with probability zero. So to say an event occurs “almost surely” or is a negligible event is not an empty phrase.

Problems.

**Exercise 5.36.** In an infinite sequence of fair coin flips, consider the event that there are only finitely many tails. What is the probability of this event? Is this event empty? Is this event impossible?

**Exercise 5.37.** Provide a more complete heuristic argument based on Theorem 5.32 that almost surely there is a sequence \( t_n \) with \( \lim_{t \to \infty} t_n = \infty \) such that \( W(t) = 0 \)

**Exercise 5.38.** Provide a heuristic argument based on Theorem 5.33 and the shifting property that the zero set of Brownian motion

\[ \{ t \in [0, \infty) : W(t) = 0 \} \]

has no isolated points.

**Exercise 5.39.** Looking in more advanced references, find another property of Brownian motion which illustrates strange path properties.
5.7. Quadratic Variation of the Wiener Process

Section Starter Question. What is an example of a function that varies a lot? What is an example of a function that does not vary a lot? How would you measure the variation of a function?

Variation.

Definition 5.40. A function \( f(x) \) is said to have bounded variation on the closed interval \([a, b]\) if there exists an \( M \) such that

\[
|f(t_1) - f(a)| + |f(t_2) - f(t_1)| + \cdots + |f(b) - f(t_n)| \leq M
\]

for all partitions \( a = t_0 < t_1 < t_2 < \cdots < t_n < t_{n+1} = b \) of the interval.

The idea is that we measure the total (hence the absolute value) up-and-down movement of a function. This definition is similar to other partition-based definitions such as the Riemann integral and the arc-length of the graph of the function. A monotone increasing or decreasing function has bounded variation. A function with a continuous derivative has bounded variation. Some functions, for instance the Wiener process, do not have bounded variation.

Definition 5.41. A function \( f(t) \) is said to have quadratic variation on the closed interval \([a, b]\) if there exists an \( M \) such that

\[
(f(t_1) - f(a))^2 + (f(t_2) - f(t_1))^2 + \cdots + (f(b) - f(t_n))^2 \leq M
\]

for all partitions \( a = t_0 < t_1 < t_2 < \cdots < t_n < t_{n+1} = b \) of the interval.

Again, the idea is that we measure the total (hence the positive terms created by squaring) up-and-down movement of a function. However, the squaring will make small ups-and-downs smaller, so that a function without bounded variation might have quadratic variation. In fact, this is the case for the Wiener process.

Definition 5.42. The total quadratic variation \( Q \) of a function \( f \) on an interval \([a, b]\) is

\[
Q = \sup_P \sum_{i=0}^{n} (f(t_{i+1}) - f(t_i))^2
\]

where the supremum is taken over all partitions \( P \) with \( a = t_0 < t_1 < \cdots < t_n < t_{n+1} = b \), with mesh size going to zero as the number of partition points \( n \) goes to infinity.

Quadratic Variation of the Wiener Process. We can guess that the Wiener process might have quadratic variation by considering the quadratic variation of the approximation using a coin-flipping fortune. Consider the piecewise linear function \( \hat{W}_N(t) \) on \([0, 1]\) defined by the sequence of sums \((1/\sqrt{N})T_n = (1/\sqrt{N})Y_1 + \cdots + (1/\sqrt{N})Y_n\) from the Bernoulli random variables \( Y_i = +1 \) with probability \( p = 1/2 \) and \( Y_i = -1 \) with probability \( q = 1 - p = 1/2 \). With some analysis, it is possible to show that we need only consider the quadratic variation at the node points. Then each term \((\hat{W}_N((i+1)/N) - \hat{W}_N(i/N))^2 = (1/N)Y_{i+1}^2 = 1/N\). Therefore, the quadratic variation using the total number of steps is \( Q = (1/N) \cdot N = 1 \). Now remembering the Wiener process is approximated by \( \hat{W}_N(t) \) suggests that quadratic variation of the Wiener process on \([0, 1]\) is 1.
5.7. QUADRATIC VARIATION OF THE WIENER PROCESS

We will not rigorously prove that the total quadratic variation of the Wiener process is \( t \) with probability 1 because the proof requires deeper analytic tools. We will instead prove a pair of theorems close to the general definition of quadratic variation. First is a weak convergence version of the quadratic variation of the Wiener process, see [5].

**Theorem 5.43.** Let \( W(t) \) be the standard Wiener process. For every fixed \( t > 0 \)

\[
\lim_{n \to \infty} E \left[ \sum_{k=1}^{n} \left( W \left( \frac{kt}{n} \right) - W \left( \frac{(k-1)t}{n} \right) \right)^2 \right] = t.
\]

**Proof.** Consider

\[
\sum_{k=1}^{n} \left( W \left( \frac{kt}{n} \right) - W \left( \frac{(k-1)t}{n} \right) \right)^2
\]

Let

\[
Z_{nk} = \frac{(W \left( \frac{kt}{n} \right) - W \left( \frac{(k-1)t}{n} \right))}{\sqrt{t/n}}.
\]

Then for each \( n \), the sequence \( Z_{nk} \) is a sequence of independent, identically distributed \( N(0,1) \) standard normal random variables. We can write the quadratic variation on the regularly spaced partition \( 0 < 1/n < 2/n < \cdots < (n-1)/n < 1 \) as

\[
\sum_{k=1}^{n} \left( W \left( \frac{kt}{n} \right) - W \left( \frac{(k-1)t}{n} \right) \right)^2 = \sum_{k=1}^{n} \frac{t}{n} Z_{nk}^2
\]

\[
= t \left( \frac{1}{n} \sum_{k=1}^{n} Z_{nk}^2 \right).
\]

But notice that the expectation \( E \left[ Z_{nk}^2 \right] \) of each term is the same as calculating the variance of a standard normal \( N(0,1) \) which is 1. Then

\[
E \left[ \frac{1}{n} \sum_{k=1}^{n} Z_{nk}^2 \right]
\]

converges to 1 by the Weak Law of Large Numbers. Therefore

\[
\lim_{n \to \infty} E \left[ \sum_{k=1}^{n} \left( W \left( \frac{kt}{n} \right) - W \left( \frac{(k-1)t}{n} \right) \right)^2 \right] = t.
\]

**Remark 5.44.** This proof is in itself not sufficient to prove the almost sure theorem above because it relies on the Weak Law of Large Numbers. Hence the theorem establishes convergence in distribution only, while for the strong theorem we want convergence almost surely. This is another example showing that is easier to prove a weak convergence theorem in contrast to an almost sure convergence theorem.

**Theorem 5.45.** Let \( W(t) \) be the standard Wiener process. For every fixed \( t > 0 \)

\[
\lim_{n \to \infty} \sum_{n=1}^{2^n} \left( W \left( \frac{k}{2^n} t \right) - W \left( \frac{k-1}{2^n} t \right) \right)^2 = t
\]

with probability 1 (that is, almost surely).
Proof. To introduce some briefer notation for the proof let

\[ \Delta_{nk} = W\left(\frac{k}{2^n}t\right) - W\left(\frac{k-1}{2^n}t\right) \quad k = 1, \ldots, 2^n \]

and

\[ W_{nk} = \Delta_{nk}^2 - t/2^n \quad k = 1, \ldots, 2^n. \]

We want to show that \( \sum_{k=1}^{2^n} \Delta_{nk}^2 \to t \) or equivalently:

\[ \sum_{k=1}^{2^n} W_{nk} \to 0. \]

For each \( n \), the random variables \( W_{nk}, k = 1, \ldots, 2^n \) are independent and identically distributed by properties 1 and 2 of the definition of the standard Wiener process. Furthermore,

\[ \mathbb{E}[W_{nk}] = \mathbb{E}[\Delta_{nk}^2] - t/2^n = 0 \]

by property 1 of the definition of the standard Wiener process.

A routine (but omitted) computation of the fourth moment of the normal distribution shows that

\[ \mathbb{E}[W_{nk}^2] = 2t^2/4^n. \]

Finally, by property 2 of the definition of the standard Wiener process

\[ \mathbb{E}[W_{nk}W_{nj}] = 0, k \neq j. \]

Expanding the square of the sum, and applying all of these computations

\[ \mathbb{E}\left[ \left( \sum_{k=1}^{2^n} W_{nk} \right)^2 \right] = \sum_{k=1}^{2^n} \mathbb{E}[W_{nk}^2] = 2^{n+1}t^2/4^n = 2t^2/2^n. \]

Apply Chebyshev’s Inequality to see that

\[ \mathbb{P}\left[ \left| \sum_{k=1}^{2^n} W_{nk} \right| > \epsilon \right] \leq \frac{2t^2}{\epsilon^2} \left( \frac{1}{2} \right)^n. \]

Since \( \sum (1/2)^n \) is a convergent series, the Borel-Cantelli lemma implies that the event

\[ \left| \sum_{k=1}^{2^n} W_{nk} \right| > \epsilon \]

can occur for only finitely many \( n \). That is, for any \( \epsilon > 0 \), there is an \( N \), such that for \( n > N \)

\[ \left| \sum_{k=1}^{2^n} W_{nk} \right| < \epsilon. \]

Therefore we must have that \( \lim_{n \to \infty} \sum_{k=1}^{2^n} W_{nk} = 0 \), and we have established what we wished to show. \( \square \)

Remark 5.46. Starting from

\[ \lim_{n \to \infty} \sum_{n=1}^{2^n} \left[ W\left(\frac{k}{2^n}t\right) - W\left(\frac{k-1}{2^n}t\right)\right]^2 = t \]

and without thinking too carefully about what it might mean, we can imagine an elementary calculus limit of the left side and write the formula:

\[ \int_0^t [dW(\tau)]^2 = t = \int_0^t d\tau. \]
In fact, more advanced mathematics makes this sensible and mathematically sound. Now from this relation, we could write the integral equality in differential form:

\[ dW(\tau)^2 = d\tau. \]

The important thing to remember here is that the formula suggests that the Wiener process has differentials that cannot be ignored in second (or squared, or quadratic) order.

**Remark 5.47.** This theorem can be nicely summarized in the following way: Let \( dW(t) = W(t + dt) - W(t) \). Then (although mathematically not rigorously) it is helpful to remember

\[ dW(t) \sim N(0, dt) \]

\[ (dW(t))^2 \sim N(dt, 0). \]

**Theorem 5.48.** Let \( W(t) \) be the standard Wiener process. For every fixed \( t > 0 \)

\[ \lim_{n \to \infty} 2^n \sum_{n=1}^{2^n} \left| W \left( \frac{k}{2^n} t \right) - W \left( \frac{k-1}{2^n} t \right) \right| = \infty. \]

In other words, the total variation of a Wiener process path is infinite, with probability 1.

**Proof.**

\[ \sum_{n=1}^{2^n} \left| W \left( \frac{k}{2^n} t \right) - W \left( \frac{k-1}{2^n} t \right) \right| \geq \frac{\sum_{n=1}^{2^n} \left| W \left( \frac{k}{2^n} t \right) - W \left( \frac{k-1}{2^n} t \right) \right|^2}{\max_{j=1,\ldots,2^n} \left| W \left( \frac{k}{2^n} t \right) - W \left( \frac{k-1}{2^n} t \right) \right|}. \]

The numerator on the right converges to \( t \), while the denominator goes to 0 because Wiener process are continuous, therefore uniformly continuous on bounded intervals. Therefore the fraction on the right goes to infinity.

**Section Ending Question.** A function that we intuitively feel varies a lot is \( f(t) = \sin(t) \) because of the oscillations. We can strengthen this intuition by using the idea of inversion from a previous section. We intuitively feel that \( f(t) = \sin(1/t) \) (with \( f(0) = 0 \) varies a lot on the interval \([0,1]\)). On the other hand, a constant function does not vary at all. These examples suggest that totaling the oscillations of a function should be a measure of the variation. The definitions and theorems of this section formalize that intuition.

**Algorithms, Scripts, Simulations.**

**Algorithm.** The simulation of the quadratic variation of the Wiener process demonstrates how much the properties of the Wiener process depend on the full limiting function, not an approximation. For simulation of quadratic variation of the Wiener process using an approximation \( W_N(x) \) of the Wiener process, the mesh size of the partition should be approximately the same size as \( 1/N \). This is easiest to illustrate when the number of partition points is a divisor or multiple of \( N \) and evenly spaced. We already showed above that if the mesh points coincide with the \( N \) steps of the scaled random walk, then the quadratic variation is 1.

Consider the case when the number of evenly spaced partition points is a multiple of \( N \), say \( m = kN \). Then on each subinterval \([i/m, (i+1)/m]\) = \([i/(kN), (i+1)/(kN)]\),
1) \((\pm \sqrt{\frac{T}{N}}) \cdot (1/(kN))\) the quadratic variation is \(((\pm \sqrt{\frac{T}{N}}) \cdot (1/(kN)))^2 = 1/(k^2NT)\) and the total of all \(kN\) steps is \(1/(kT)\) which approaches 0 as \(k\) increases. This is not what is predicted by the theorem about the quadratic variation of the Wiener process, but it is consistent with the approximation function \(\hat{W}_N(t)\), see the discussion below.

Now consider the case when \(m\) is a divisor of \(N\), say \(km = N\). Then the quadratic variation of \(\hat{W}_N(t)\) on a partition interval will be the sum of \(k\) scaled random walk steps, squared. An example will show what happens. With \(T = 1\), let \(N = 1000\) and let the partition have 200 equally spaced points, so \(k = 5\). Then each partition interval will have a quadratic variation of:

1. \((1/\sqrt{1000})^2\) with probability \(2 \cdot \left(\frac{5}{1}\right) \cdot (1/32)\);
2. \((3/\sqrt{1000})^2\) with probability \(2 \cdot \left(\frac{5}{3}\right) \cdot (1/32)\); and
3. \((5/\sqrt{1000})^2\) with probability \(2 \cdot \left(\frac{5}{5}\right) \cdot (1/32)\).

Each partition interval quadratic variation has a mean of \((1/1000) \cdot (160/32) = 1/200\) and a variance of \((1/1000)^2 \cdot (1280/32) = 40/(1000)^2\). Add 200 partition interval quadratic variations to get the total quadratic variation of \(\hat{W}_N(t)\) on \([0,1]\). By the Central Limit Theorem the total quadratic variation will be approximately normally distributed with mean 1 and standard deviation \(\sqrt{200 \cdot (\sqrt{40}/1000)} = \sqrt{5}/25 \approx 0.089\).

Note that although \(\hat{W}_N(t)\) does not have a continuous first derivative, it fails to have a derivative at only finitely many points, so in that way it is almost differentiable. In fact, \(\hat{W}_N(t)\) satisfies a uniform Lipschitz condition (with Lipschitz constant \(\sqrt{N/T}\)), which is enough to show that it has bounded variation. As such the quadratic variation of \(\hat{W}_N(t)\) is 0.

```r
1 p <- 0.5
2 N <- 1000
3 T <- 1
4 S <- array(0, c(N+1))
5 rw <- cumsum(2 * (runif(N) <= p) - 1)
6 S[2:(N+1)] <- rw
7 WcaretN <- function(x) {
8   Delta <- T/N
9   # add 1 since arrays are 1-based
10  prior = floor(x/Delta) + 1
11  subsequent = ceiling(x/Delta) + 1
12  retval <- sqrt(Delta) * (S[prior] + ((x/Delta + 1) - prior) * (S[ subsequent] - S[prior]))
13 }
14 m1 <- N/5
15 partition1 <- seq(0, T, 1/m1)
16 m2 <- N
17 partition2 <- seq(0, T, 1/m2)
18 m3 <- 3*N
19 partition3 <- seq(0, T, 1/m3)
```
5.7. QUADRATIC VARIATION OF THE WIENER PROCESS

26
qv1 <- sum((WcaretN(partition1[-1]) - WcaretN(partition1[-length(partition1)]))^2)
27 qv2 <- sum((WcaretN(partition2[-1]) - WcaretN(partition2[-length(partition2)]))^2)
28 qv3 <- sum((WcaretN(partition3[-1]) - WcaretN(partition3[-length(partition3)]))^2)
29
30 cat(sprintf("Quadratic variation of approximation of Wiener process 
paths with %d scaled random steps with %d partition intervals is: 
%f \n", N, m1, qv1))
31 cat(sprintf("Quadratic variation of approximation of Wiener process 
paths with %d scaled random steps with %d partition intervals is: 
%f \n", N, m2, qv2))
32 cat(sprintf("Quadratic variation of approximation of Wiener process 
paths with %d scaled random steps with %d partition intervals is: 
%f \n", N, m3, qv3))

Key Concepts.

(1) The total quadratic variation of the Wiener process on \([0, t]\) is \(t\).
(2) This fact has profound consequences for dealing with the Wiener process 
analytically and ultimately will lead to Itô’s formula.

Vocabulary.

(1) A function \(f(t)\) is said to have **bounded variation** on \([a, b]\) if there exists an \(M\) such that

\[
|f(t_1) - f(a)| + |f(t_2) - f(t_1)| + \ldots + |f(b) - f(t_n)| \leq M
\]

for all partitions \(a = t_0 < t_1 < t_2 < \ldots < t_n < t_{n+1} = b\) of the interval.

(2) A function \(f(t)\) is said to have **quadratic variation** on \([a, b]\) if there exists an \(M\) such that

\[
(f(t_1) - f(a))^2 + (f(t_2) - f(t_1))^2 + \ldots + (f(b) - f(t_n))^2 \leq M
\]

for all partitions \(a = t_0 < t_1 < t_2 < \ldots < t_n < t_{n+1} = b\) of the interval.

(3) The **mesh size** of a partition \(P\) with \(a = t_0 < t_1 < \ldots < t_n < t_{n+1} = b\) is \(\max_{j=0, \ldots, n}(t_{j+1} - t_j)\).

(4) The **total quadratic variation** of a function \(f\) on an interval \([a, b]\) is

\[
\sup_{P} \sum_{j=0}^{n} (f(t_{j+1}) - f(t_j))^2
\]

where the supremum is taken over all partitions \(P\) with \(a = t_0 < t_1 < \ldots < t_n < t_{n+1} = b\), with mesh size going to zero as the number of partition points \(n\) goes to infinity.
Problems.

Exercise 5.49. Show that a monotone increasing function has bounded variation.

Exercise 5.50. Show that a function with continuous derivative has bounded variation.

Exercise 5.51. Show that the function
\[ f(t) = \begin{cases} t^2 \sin(1/t) & 0 < t \leq 1 \\ 0 & t = 0 \end{cases} \]
is of bounded variation, while the function
\[ f(t) = \begin{cases} t \sin(1/t) & 0 < t \leq 1 \\ 0 & t = 0 \end{cases} \]
is not of bounded variation.

Exercise 5.52. Show that a continuous function of bounded variation is also of quadratic variation.

Exercise 5.53. Show that the fourth moment \( \mathbb{E}[Z^4] = 3 \) where \( Z \sim N(0,1) \). Then show that
\[ \mathbb{E}[W_{nk}^2] = 2t^2/4^n \]

Exercise 5.54. Modifying one of the scripts, find the quadratic variation of \( \hat{W}_N(t) \) with a partition with \( m \) partition intervals whose endpoints are randomly selected values in \([0,T]\). One way to approach this is to create a list of \( m - 1 \) points uniformly distributed in \([0,T]\), append the values 0 and \( T \) to the list, then sort into ascending order to create the partition points. Find the mesh of this random partition and print its value along with the quadratic variation. What happens when \( m \) is a multiple of \( N \)? What happens when \( m \) is a divisor of \( N \)? What could possibly go wrong in this calculation?

Exercise 5.55. Generalize the example with \( N = 1000 \) and \( m = 200 \) of the quadratic variation of \( \hat{W}_N(t) \) on \([0,1]\) to the case when the number of partition intervals \( m \) is a divisor of some \( N \), say \( km = N \).
Stochastic Calculus

6.1. Stochastic Differential Equations

Section Starter Question. Explain how to use a slope-field diagram to solve
the ordinary differential equation
\[
\frac{dx}{dt} = x, \quad x(t_0) = x_0.
\]
How would you turn that process into an algorithm to numerically compute an
approximate solution without a diagram?

Stochastic Differential Equations: Symbolically. The straight-line seg-
ment is the building block of differential calculus. The basic idea behind differential
calculus is that differentiable functions, no matter how difficult their global behav-
ior, are locally approximated by straight-line segments. In particular, this is the
idea behind Euler’s method for approximating differentiable functions defined by
differential equations.

We know that rescaling (zooming in on) Brownian motion does not produce
a straight-line segment, it produces another image of Brownian motion. This self-
similarity is ideal for an infinitesimal building block; for instance, we could build
global Brownian motion out of lots of local “chunks” of Brownian motion. This
suggests we could build other stochastic processes out of suitably scaled Brownian
motion. In addition, if we include straight-line segments we can overlay the behavior
of differentiable functions onto the stochastic processes as well. Thus, straight-line
segments and chunks of Brownian motion are the building blocks of stochastic
calculus.

With stochastic differential calculus, we can build new stochastic processes. We
do this by specifying how to build the new stochastic processes locally as a com-
bination of basic elements: the deterministic straight line and standard Brownian
motion. We write the local change in value of the stochastic process over a time
interval of (infinitesimal) length \(dt\) as
\[
(6.1) \quad dX = G(X(t)) \, dt + H(X(t)) \, dW(t), \quad X(t_0) = X_0.
\]
Note that we are not allowed to write
\[
\frac{dX}{dt} = G(X(t)) + H(X(t)) \frac{dW}{dt}, \quad X(t_0) = X_0
\]
since standard Brownian motion is nowhere differentiable with probability 1. Actu-
ally, the informal stochastic differential equation (6.1) is a compact way of writing
a rigorously defined, equivalent implicit Itô integral equation. Since we do not have
the required rigor, we will approach the stochastic differential equation intuitively.

The stochastic differential equation says the initial point \((t_0, X_0)\) is specified,
perhaps with \(X_0\) a random variable with a given distribution. A deterministic
component at each point has a slope determined through $G$ at that point. In addition, some random perturbation affects the evolution of the process. The random perturbation is normally distributed with mean 0. The variance of the random perturbation is $(H(X(t)))^2$ at $(t, X(t))$. This is a simple expression of a Stochastic Differential Equation (SDE) that determines a stochastic process, just as an Ordinary Differential Equation (ODE) determines a differentiable function. We extend the process with the incremental change information and repeat. This is an expression in words of the Euler-Maruyama method for numerically simulating the stochastic differential expression.

**Example 6.1.** A very simple stochastic differential equation is

$$dX = r\, dt + dW, \quad X(0) = b$$

with $r$ a constant. Take a deterministic initial condition to be $X(0) = b$. The new process is the stochastic extension of the differential equation expression of a straight line. The new stochastic process $X$ is drifting or trending at constant rate $r$ with a random variation due to Brownian motion perturbations around that trend. We will later show explicitly that the solution of this SDE is $X(t) = b + rt + W(t)$ although it is seems intuitively clear that this should be the process. We will call this Brownian motion with drift.

**Example 6.2.** Another very simple stochastic differential equation is

$$dX = \sigma\, dW, \quad X(0) = b$$

This stochastic differential equation says that the process is evolving as a multiple of standard Brownian motion. The solution may be easily guessed as $X(t) = \sigma W(t)$ which has variance $\sigma^2 t$ on increments of length $t$. Sometimes the new process is called Brownian motion (in contrast to standard Brownian motion which has variance $t$ on increments of length $t$).

We combine the previous two examples to consider

$$dX = r\, dt + \sigma\, dW, \quad X(0) = b,$$

which has solution $X(t) = b + rt + \sigma W(t)$, a multiple of Brownian motion with drift $r$ started at $b$. Sometimes this extension of standard Brownian motion is called Brownian motion. Some authors consider this process directly instead of the more special case we considered in the previous chapter.

**Example 6.3.** The next simplest and first non-trivial differential equation is

$$dX = X \, dW.$$

Here the differential equation says that process is evolving like Brownian motion with a variance which is the square of the process value. When the process is small, the variance is small, when the process is large, the variance is large. Expressing the stochastic differential equation as $dX / X = dW$ we may say that the relative change acts like standard Brownian motion. The resulting stochastic process is called geometric Brownian motion and it will figure extensively later as a model of security prices.
Example 6.4. The next simplest differential equation is
\[ dX = rX \, dt + \sigma X \, dW, \quad X(0) = b. \]

Here the stochastic differential equation says that the growth of the process at a point is proportional to the process value, with a random perturbation proportional to the process value. Again looking ahead, we could write the differential equation as \( dX/X = r \, dt + \sigma \, dW \) and interpret it to say the relative rate of increase is proportional to the time observed together with a random perturbation like a Brownian increment corresponding to the length of time. We will show later that the analytic expression for the stochastic process defined by this SDE is \( b \exp((r - \frac{1}{2} \sigma^2)t + \sigma W(t)). \)

Stochastic Differential Equations: Numerically. The sample path that the Euler-Maruyama method produces numerically is the analog of using the Euler method.

The formula for the Euler-Maruyama (EM) method is based on the definition of the Itô stochastic integral:
\[ X_j = X_{j-1} + G(X_{j-1}) \, dt + H(X_{j-1})(W(t_{j-1} + \Delta t) - W(t_{j-1})), \]
\[ t_j = t_{j-1} + \Delta t. \]

Note that the initial conditions \( X_0 \) and \( t_0 \) set the starting point.

In this text, we use coin-flipping sequences of an appropriate length scaled to create an approximation to \( W(t) \) just as in the section Approximation of Brownian Motion. The coin-flipping sequences emphasize the discrete nature of the simulations with an easily constructed random process. This is consistent with the approach of this text which always uses coin-flipping sequences to create random processes. Note that since the increments \( W(t_{j-1} + \Delta t) - W(t_{j-1}) \) are independent and identically distributed, we will use independent coin-flip sequences to generate the approximation of the increments. The EM method could use independent normal random variates directly to obtain the increments \( W(t_{j-1} + \Delta t) - W(t_{j-1}) \). Using independent normal random variates directly would be easier and more efficient. The exercises modify the example scripts to use independent normal random variates directly.

Then
\[ dW = W(t_{j-1} + \Delta t) - W(t_{j-1}) = W(\Delta t) \]
\[ \approx \hat{W}(\Delta t) = \frac{\hat{W}(N \, \Delta t)}{\sqrt{N}} = \sqrt{\frac{\Delta t}{N} \hat{W}(N \, \Delta t)}. \]

The first equality above is the definition of an increment, the second equality means the random variables \( W(t_{j-1} + \Delta t) - W(t_{j-1}) \) and \( W(\Delta t) \) have the same distribution because of the definition of standard Brownian motion which specifies that increments with equal length are normally distributed with variance equal to the increment length. The approximate equality occurs because of the approximation of Brownian motion by coin-flipping sequences. We generate the approximations using a random number generator, but we could as well use actual coin-flipping. In Table 1 the generation of the sequences is not recorded, only the summed and scaled (independently sampled) outcomes. For convenience, take \( \Delta t = 1/10, N = 100, \) so we need \( \hat{W}(100 \cdot (1/10))/\sqrt{100} = T_{10}/10. \) Then to obtain the entries in the column
labeled $dW$ in the table we flip a coin 10 times and record $T_{10}/10$. Take $r = 2$, $b = 1$, and $\sigma = 1$, so we simulate the solution of
\[ dX = 2X \, dt + X \, dW, \quad X(0) = 1. \]
A computer program can produce such a table with the step size made much smaller, presumably resulting in better approximation properties. In fact, it is possible to consider kinds of convergence for the EM method comparable to the Strong Law of Large Numbers and the Weak Law of Large Numbers. See the Problems for examples.

Discussion. The numerical approximation procedure using coin flips makes it clear that the Euler-Maruyama method generates a random process. The value of the process depends on the time value and the coin-flip sequence. Each generation of an approximation will be different because the coin-flip sequence is different. The Euler-Maruyama method generates a stochastic process path approximation. To derive distributions and statistics about the process requires generating multiple paths; see the Problems for examples.

This shows that stochastic differential equations provide a way to define new stochastic processes. This is analogous to the notion that ordinary differential equations define new functions to study and use. In fact, one approach to developing calculus and the analysis of functions is to start with differential equations, use the Euler method to define approximations of solutions, and then to develop a theory to handle the passage to continuous variables. This approach is especially useful for a mathematical modeling viewpoint since the model often uses differential equations.

This text follows the approach of starting with stochastic differential equations to describe a situation and numerically defining new stochastic processes to model the situation. At certain points, we appeal to more rigorous mathematical theory to justify the modeling and approximation. One important justification asserts that if we write a stochastic differential equation, then solutions exist and the stochastic differential equation always yields the same process under equivalent conditions.
The Existence-Uniqueness Theorem shows that under reasonable modeling conditions stochastic differential equations do indeed satisfy this requirement.

**Theorem 6.5 (Existence-Uniqueness).** For the stochastic differential equation
\[ dX = G(t, X(t)) \, dt + H(t, X(t)) \, dW(t), \quad X(t_0) = X_0 \]
assume
(1) Both $G(t, x)$ and $H(t, x)$ are continuous on $(t, x) \in [t_0, T] \times \mathbb{R}$.
(2) The coefficient functions $G$ and $H$ satisfy a Lipschitz condition:
\[ |G(t, x) - G(t, y)| + |H(t, x) - H(t, y)| \leq K|x - y|. \]
(3) The coefficient functions $G$ and $H$ satisfy a growth condition in the second variable
\[ |G(t, x)|^2 + |H(t, x)|^2 \leq K(1 + |x|^2) \]
for all $t \in [t_0, T]$ and $x \in \mathbb{R}$.

Then the stochastic differential equation has a strong solution on $[t_0, T]$ that is continuous with probability 1 and
\[ \sup_{t \in [t_0, T]} \mathbb{E}[X^2(t)] < \infty \]
and for each given Wiener process $W(t)$, the corresponding strong solutions are pathwise unique which means that if $X$ and $Y$ are two strong solutions, then
\[ \mathbb{P}\left[ \sup_{t \in [t_0, T]} |X(t) - Y(t)| = 0 \right] = 1. \]

See [33] for a precise definition of “strong solution” but essentially it means that for each given Wiener process $W(t)$ we can generate a solution to the SDE. Note that the coefficient functions here are two-variable functions of both time $t$ and location $x$, which is more general than the functions considered in equation (6.1). The restrictions on the functions $G(t, x)$ and $H(t, x)$, especially the continuity condition, can be considerably relaxed and the theorem will still remain true.

**Section Ending Answer.** On a portion of the $t-x$ plane containing $(t_0, x_0)$, on a grid of points including $(t_0, x_0)$, with reasonably small horizontal spacing $\Delta t$, draw a field of short arrows with slope equal to the $x$-coordinate. From $(t_0, x_0)$, follow the short arrows, interpolating a polygonal path through the grid. Numerically, this becomes
\[ x(t_{j+1}) \approx x_{j+1} = x_j + x_j \cdot \Delta t, \quad x_0 = x(t_0). \]
This is the graphical and numerical expression of the Euler method, the intuitive basis for the Euler-Maruyama method.

*Algorithms, Scripts, Simulations.*
Algorithm. The scripts apply the EM method to simulate the solution of
\[ dX = rX \, dt + \sigma X \, dW, \quad X(0) = b. \]
The parameters \( N \) for the number of steps in the EM method, \( T \) for the ending
time, and stochastic differential equation parameters \( r \) and \( \sigma \) are set. Find the time
step and initialize the arrays holding the time steps and the solution simulation.

Using \( M = 30N \) create a piecewise linear function \( W_M(t) \) using the approxi-
mation scripts in the section Approximation of Brownian Motion. The parameter
30 is chosen according to the rule of thumb in the DeMoivre-Laplace Central Limit
Theorem. Then each time increment will have 30 coin flips, sufficient according
to the rule of thumb to guarantee a scaled sum which is appropriately normally
distributed.

Then loop over the number of steps using the EM algorithm
\[ X_j = X_{j-1} + G(X_{j-1}) \, dt + H(X_{j-1})(W(t_{j-1} + dt) - W(t_{j-1})) \]
and plot the resulting simulation.

```r
r <- -1  # growth/decay rate
sigma <- 0.5  # relative standard deviation
b <- 3  # initial condition

M <- 100  # number of steps for EM method to take
T <- 1  # maximum time
h <- T/M  # time step
t <- seq(length=M+1, from=0, by=h)  # t is the vector [0 1h 2h 3h ...
N <- 30*(M+1)  # number of steps for the Brownian motion approx

X[1] <- b  # place to store locations

for (i in 1:M) {
  X[i+1] <- X[i] + r*X[i]*h + sigma*X[i]*(WcaretN(t[i]+h)-WcaretN(t[i]))
}
```
plot(t,X,"l", xlim=c(0 , T), ylim=c(X[1] - exp (abs (r)*T +1) , X[1]+ exp (abs (r)*T +1)) )
title (main = paste ("r = ", r, " sigma = ", sigma , " steps =", M))

Key Concepts.
(1) We can numerically simulate the solution to stochastic differential equations with an analog to Euler’s method, called the Euler-Maruyama (EM) method.

Vocabulary.
(1) A stochastic differential equation is a mathematical equation relating a stochastic process to its local deterministic and random components. The goal is to extend the relation to find the stochastic process. Under mild conditions on the relationship, and with a specifying initial condition, solutions of stochastic differential equations exist and are unique.
(2) The Euler-Maruyama (EM) method is a numerical method for simulating the solutions of a stochastic differential equation based on the definition of the Itô stochastic integral: Given
\[ dX(t) = G(X(t)) \, dt + H(X(t)) \, dW(t), \quad X(t_0) = X_0, \]
and a step size \( dt \), we approximate and simulate with
\[ X_j = X_{j-1} + G(X_{j-1}) \, dt + H(X_{j-1})(W(t_{j-1} + dt) - W(t_{j-1})). \]
(3) Extensions and variants of standard Brownian motion defined through stochastic differential equations are Brownian motion with drift, scaled Brownian motion, and geometric Brownian motion.

Problems.

Exercise 6.6. Graph an approximation of a multiple of Brownian motion with drift with parameters \( b = 2, r = 1/2 \) and \( \sigma = 2 \) on the interval \([0, 5]\) in two ways.
(1) Flip a coin 25 times, recording whether it comes up Heads or Tails each time, Scoring \( Y_i = +1 \) for each Heads and \( Y_i = -1 \) for each flip, also keep track of the accumulated sum \( T_n = \sum_{i=1}^{n} T_i \) for \( i = 1 \ldots 25 \). Using \( N = 5 \) compute the rescaled approximation \( \tilde{W}_5(t) = (1/\sqrt{5})T_{5t} \) at the values \( t = 0, 1/5, 2/5, 3/5, \ldots 24/5, 5 \) on \([0, 5]\). Finally compute and graph the value of \( X(t) = b + rt + \sigma \tilde{W}_5(t) \).
(2) Using the same values of \( \tilde{W}_5(t) \) as approximations for \( W(dt) \) compute the values of the solution of the stochastic differential equation \( dX = r \, dt + \sigma \, dW, \quad X(0) = b \) on the interval \([0, 5]\).

Exercise 6.7. Repeat the previous problem with parameters \( b = 2, r = -1/2 \) and \( \sigma = 2 \).

Exercise 6.8. Repeat the previous problem with parameters \( b = 2, r = 1/2 \) and \( \sigma = -2 \).
Exercise 6.9. Modify the scripts to use normally distributed random deviates and then simulate the solution of the stochastic differential equation
\[ dX(t) = X(t) \, dt + 2X(t) \, dW \]
on the interval \([0, 1]\) with initial condition \(X(0) = 1\) and step size \(\Delta t = 1/10\).

Exercise 6.10. Modify the scripts to use normally distributed random deviates and then simulate the solution of the stochastic differential equation
\[ dX(t) = tX(t) \, dt + 2X(t) \, dW \]
on the interval \([0, 1]\) with initial condition \(X(0) = 1\) and step size \(\Delta t = 1/10\). Note the difference with the previous problem, now the multiplier of the \(dt\) term is a function of time.

Exercise 6.11. Modify the scripts to use general functions \(G(X)\) and \(H(x)\) and then apply it to the SDE for the
1. Ornstein-Uhlenbeck process:
\[ dX(t) = \theta (\mu - X(t)) \, dt + \sigma dW(t), \quad X(t_0) = X_0. \]
2. Cox-Ingersoll-Ross process:
\[ dX(t) = \theta (\mu - X(t)) \, dt + \sigma \sqrt{X(t)} \, dW(t), \quad X(t_0) = X_0. \]
3. the modified Cox-Ingersoll-Ross process:
\[ dX(t) = -\theta X(t) \, dt + \theta \sqrt{X(t)} \, dW(t), \quad X(t_0) = X_0. \]

Exercise 6.12. Write a program with parameters \(r, \sigma, b, T\) and \(N\) (so \(dt = T/N\)) that computes and graphs the approximation of the solution of the stochastic differential equation
\[ dX(t) = rX(t) \, dt + \sigma X(t) \, dW \]
with \(X(0) = b\) on the interval \([0, T]\). Apply the program to the stochastic differential equation with \(r = 2, \sigma = 1, b = 1,\) and \(N = 2^6, 2^7, 2^8\) on the interval \([0, 1]\).

Exercise 6.13. Generalize the program from the previous problem to include a parameter \(M\) for the number of sample paths computed. Then using this program on the interval \([0, 1]\) with \(M = 1000,\) and \(N = 2^8\) compute \(\mathbb{E}[|X_N - X(1)|]\), where \(X(1) = be^{r - \frac{1}{2}\sigma^2 + \sigma W_{2^8}(1)}\).

Exercise 6.14. Using the program from the previous problem with \(M = 1000\) and \(N = 2^5, 2^6, 2^7, 2^8, 2^9\) compute \(\mathbb{E}[|X_N - X(1)|]\), where \(X(1) = be^{r - \frac{1}{2}\sigma^2 + \sigma W_{2^9}(1)}\). Then for the 5 values of \(N\), make a log-log plot of \(\mathbb{E}[|X_N - X(1)|]\) on the vertical axis against \(\Delta t = 1/N\) on the horizontal axis. Using the slope of the resulting best-fit line experimentally determine the order of convergence \(\gamma\) so that
\[ \mathbb{E}[|X_N - X(1)|] \leq C(\Delta t)^\gamma. \]

6.2. Itô’s Formula

Section Starter Question. State the Taylor expansion about \(x_0\) of a differentiable function \(f(x)\) up to order 1. What is the relation of this expansion to the Mean Value Theorem of calculus? What is the relation of this expansion to the Fundamental Theorem of calculus?
Motivation, Examples and Counterexamples. We need some operational rules that allow us to manipulate stochastic processes with stochastic calculus. The important thing to know about traditional differential calculus is that it is:

- the Fundamental Theorem of Calculus;
- the chain rule; and
- Taylor polynomials and Taylor series

that enable us to calculate with functions. A deeper understanding of calculus recognizes that these three calculus theorems are all aspects of the same fundamental idea. Likewise we need similar rules and formulas for stochastic processes. Itô’s formula will perform that function for us. However, Itô’s formula acts in the capacity of all three of the calculus theorems, and we have only one such theorem for stochastic calculus.

The next example will show us that we will need some new rules for stochastic calculus; the old rules from calculus will no longer make sense.

Example 6.15. Consider the process that is the square of the Wiener process:

\[ Y(t) = W(t)^2. \]

We notice that this process is always non-negative, \( Y(0) = 0 \), \( Y \) has infinitely many zeroes on \( t > 0 \) and \( \mathbb{E}[Y(t)] = \mathbb{E}[W(t)^2] = t \). What more can we say about this process? For example, what is the stochastic differential of \( Y(t) \) and what would that tell us about \( Y(t) \)?

Using naive calculus, we might conjecture using the ordinary chain rule

\[ dY = 2W(t) \, dW(t). \]

If that were true then the Fundamental Theorem of Calculus would imply

\[ Y(t) = \int_0^t dY = \int_0^t 2W(t) \, dW(t) \]

should also be true. But consider \( \int_0^t 2W(t) \, dW(t) \). It ought to correspond to a limit of a summation (for instance a Riemann-Stieltjes left sum):

\[ \int_0^t 2W(t) \, dW(t) \approx \sum_{i=1}^n 2W((i-1)t/n)[W(it/n) - W((i-1)t/n)]. \]

But look at this carefully: \( W((i-1)t/n) = W((i-1)t/n) - W(0) \) is independent of \( W(it/n) - W((i-1)t/n) \) by property 2 of the definition of the Wiener process. Therefore, if what we conjecture is true, the expected value of the summation will be zero:

\[
\mathbb{E}[Y(t)] = \mathbb{E}
\left[
\int_0^t 2W(t) \, dW(t)
\right]
\]

\[
= \mathbb{E}
\left[
\lim_{n \to \infty} \sum_{i=1}^n 2W((i-1)t/n)(W(it/n) - W((i-1)t/n))
\right]
\]

\[
= \lim_{n \to \infty} \sum_{i=1}^n 2\mathbb{E}[(W((i-1)t/n) - W(0))[W(it/n) - W((i-1)t/n)]
\]

\[ = 0. \] (Note the assumption that the limit and the expectation can be interchanged!)
But the mean of $Y(t) = W(t)^2$ is $t$ which is definitely not zero! The two stochastic processes don’t agree even in the mean, so something is not right! If we agree that the integral definition and limit processes should be preserved, then the rules of calculus will have to change.

We can see how the rules of calculus must change by rearranging the summation. Use the simple algebraic identity 
\[ 2b(a - b) = (a^2 - b^2 - (a - b)^2) \]
to re-write
\[
\int_0^t 2W(t) \, dW(t) = \lim_{n \to \infty} \sum_{i=1}^n 2W((i - 1)t/n)[W(it/n) - W((i - 1)t/n)]
\]
\[
= \lim_{n \to \infty} \sum_{i=1}^n \left( W(it/n)^2 - W((i - 1)t/n)^2 - (W(it/n) - W((i - 1)t/n))^2 \right)
\]
\[
= \lim_{n \to \infty} \left( W(t)^2 - W(0)^2 - \sum_{i=1}^n (W(it/n) - W((i - 1)t/n))^2 \right)
\]
\[
= W(t)^2 - \lim_{n \to \infty} \sum_{i=1}^n (W(it/n) - W((i - 1)t/n))^2.
\]

We recognize the second term in the last expression as being the quadratic variation of Wiener process that we have already evaluated; and so
\[
\int_0^t 2W(t) \, dW(t) = W(t)^2 - t.
\]

**Itô’s Formula and Itô calculus.** Itô’s formula is an expansion expressing a stochastic process in terms of the deterministic differential and the Wiener process differential, that is, the stochastic differential equation for the process.

**Theorem 6.16 (Itô’s formula).** If $Y(t)$ is scaled Wiener process with drift, satisfying $dY = r \, dt + \sigma \, dW$ and $f$ is a twice continuously differentiable function, then $Z(t) = f(Y(t))$ is also a stochastic process satisfying the stochastic differential equation
\[
dZ = (rf'(Y) + (\sigma^2/2)f''(Y)) \, dt + (\sigma f'(Y)) \, dW.
\]

In words, Itô’s formula in this form tells us how to expand (in analogy with the chain rule or Taylor’s formula) the differential of a process which is defined as an elementary function of scaled Brownian motion with drift.

Itô’s formula is often also called Itô’s lemma by other authors and texts. Most authors believe that this result is more important than a mere lemma, and so this text adopts the alternative name of formula. Formula also emphasizes the analogy with the chain rule and the Taylor expansion.

**Example 6.17.** Consider $Z(t) = W(t)^2$. Here the stochastic process is standard Brownian motion, so $r = 0$ and $\sigma = 1$ so $dY = dW$. The twice continuously differentiable function $f$ is the squaring function, $f(x) = x^2$, $f'(x) = 2x$ and $f''(x) = 2$. Then according to Itô’s formula:
\[
d(W^2) = (0 \cdot (2W(t)) + (1/2)(2)) \, dt + (1 \cdot 2W(t)) \, dW = dt + 2W(t) \, dW.
\]
Notice the additional $dt$ term! Note also that if we repeated the integration steps above in the example, we would obtain $W(t)^2$ as expected!
Example 6.18. Consider geometric Brownian motion

$$\text{exp}(rt + \sigma W(t)).$$

What SDE does geometric Brownian motion follow? Take $Y(t) = rt + \sigma W(t)$, so that $dY = r dt + \sigma dW$. Then geometric Brownian motion can be written as

$$Z(t) = \exp(Y(t)),$$

so $f$ is the exponential function. Itô’s formula is

$$dZ = (rf'(Y(t)) + (1/2)\sigma^2 f''(Y(t))) dt + \sigma f'(Y) dW.$$

Computing the derivative of the exponential function and evaluating, $f'(Y(t)) = \exp(Y(t)) = Z(t)$ and likewise for the second derivative. Hence

$$dZ = (r + (1/2)\sigma^2)Z(t) dt + \sigma Z(t) dW.$$

The case where $dY = dW$, that is the base process is standard Brownian motion so $Z = f(W)$, occurs commonly enough that we record Itô’s formula for this special case:

Corollary 6.19 (Itô’s Formula applied to functions of standard Brownian motion). If $f$ is a twice continuously differentiable function, then $Z(t) = f(W(t))$ is also a stochastic process satisfying the stochastic differential equation

$$dZ = df(W) = (1/2)f''(W) dt + f'(W) dW.$$

Guessing Processes from SDEs with Itô’s Formula. One of the key needs we will have is to go in the opposite direction and convert SDEs to processes, in other words to solve SDEs. We take guidance from ordinary differential equations, where finding solutions to differential equations comes from judicious guessing based on a thorough understanding and familiarity with the chain rule. For SDEs the solution depends on inspired guesses based on a thorough understanding of the formulas of stochastic calculus. Following the guess we require a proof that the proposed solution is an actual solution, again using the formulas of stochastic calculus.

A few rare examples of SDEs can be solved with explicit familiar functions. This is just like ODEs in that the solutions of many simple differential equations cannot be solved in terms of elementary functions. The solutions of the differential equations define new functions which are useful in applications. Likewise, the solution of an SDE gives us a way of defining new processes which are useful.

Example 6.20. Suppose we are asked to solve the SDE

$$dZ(t) = \sigma Z(t) dW.$$

We need an inspired guess, so we try

$$\text{exp}(rt + \sigma W(t))$$

where $r$ is a constant to be determined while the $\sigma$ term is given in the SDE. Itô’s formula for the guess is

$$dZ = (r + (1/2)\sigma^2)Z(t) dt + \sigma Z(t) dW.$$

We notice that the stochastic term (or Wiener process differential term) is the same as the SDE. We need to choose the constant $r$ appropriately to eliminate the deterministic or drift differential term. If we choose $r$ to be $-(1/2)\sigma^2$ then the drift term in the differential equation would match the SDE we have to solve as well. We therefore guess

$$Y(t) = \exp(\sigma W(t) - (1/2)\sigma^2 t).$$
We should double check by applying Itô’s formula.

Soluble SDEs are scarce, and this one is special enough to give a name. It is the Doléan’s exponential of Brownian motion.

**Section Ending Answer.** The Taylor expansion about $x_0$ of a differentiable function up to order 1 is

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + R_1(x)$$

where $R_1(x)/(x - x_0) \to 0$ as $x \to x_0$. The Mean Value Theorem of calculus says that there is a point $x^*$ with $x_0 \leq x^* \leq x$ such that

$$\frac{f(x) - f(x_0)}{x - x_0} = f'(x^*).$$

The Fundamental Theorem of Calculus says that

$$f(x) - f(x_0) = \int_{x_0}^{x} f'(t) \, dt.$$ 

If $x - x_0$ is small, then $f'(x) \approx f'(x_0)$ and all three expressions are the same, up to a small error term.

**Problems.**

**Exercise 6.21.** Find the solution of the stochastic differential equation

$$dY(t) = Y(t) \, dt + 2Y(t) \, dW.$$ 

**Exercise 6.22.** Find the solution of the stochastic differential equation

$$dY(t) = tY(t) \, dt + 2Y(t) \, dW.$$ 

Note the difference with the previous problem, now the multiplier of the $dt$ term is a function of time.

**Exercise 6.23.** Find the solution of the stochastic differential equation

$$dY(t) = \mu Y(t) \, dt + \sigma Y(t) \, dW.$$ 

**Exercise 6.24.** Find the solution of the stochastic differential equation

$$dY(t) = \mu t Y(t) \, dt + \sigma Y(t) \, dW.$$ 

Note the difference with the previous problem, now the multiplier of the $dt$ term is a function of time.

**Exercise 6.25.** Find the solution of the stochastic differential equation

$$dY(t) = \mu(t) Y(t) \, dt + \sigma Y(t) \, dW.$$ 

Note the difference with the previous problem, now the multiplier of the $dt$ term is a general (technically, a locally bounded integrable) function of time.

**6.3. Properties of Geometric Brownian Motion**

**Section Starter Question.** What is the relative rate of change of a function?

For the function defined by the ordinary differential equation

$$\frac{dx}{dt} = rx \quad x(0) = x_0$$

what is the relative rate of growth? What is the function?
Geometric Brownian Motion. Geometric Brownian motion is the continuous time stochastic process \( X(t) = z_0 \exp(\mu t + \sigma W(t)) \) where \( W(t) \) is standard Brownian motion. Most economists prefer geometric Brownian motion as a simple model for market prices because it is everywhere positive (with probability 1), in contrast to Brownian motion, even Brownian motion with drift. Furthermore, as we have seen from the stochastic differential equation for geometric Brownian motion, the relative change is a combination of a deterministic proportional growth term similar to inflation or interest rate growth plus a normally distributed random change

\[
\frac{dX}{X} = r \, dt + \sigma \, dW,
\]

see Itô’s Formula and Stochastic Calculus. On a short time scale this is a sensible economic model.

A random variable \( X \) is said to have the lognormal distribution (with parameters \( \mu \) and \( \sigma \)) if \( \log(X) \) is normally distributed \( (\log(X) \sim N(\mu, \sigma^2)) \). The p.d.f. for \( X \) is

\[
f_X(x) = \frac{1}{\sqrt{2\pi}x\sqrt{t}} \exp\left(-\frac{1}{2}\left(\frac{\log(x) - \log(z_0) - \mu t}{\sigma \sqrt{t}}\right)^2\right).
\]

**Theorem 6.26.** At fixed time \( t \), geometric Brownian motion \( z_0 \exp(\mu t + \sigma W(t)) \) has a lognormal distribution with parameters \( (\log(z_0) + \mu t) \) and \( \sigma \sqrt{t} \).

**Proof.**

\[
F_X(x) = \mathbb{P}[X \leq x]
= \mathbb{P}[z_0 \exp(\mu t + \sigma W(t)) \leq x]
= \mathbb{P}[\mu t + \sigma W(t) \leq \log(x/z_0)]
= \mathbb{P}[W(t) \leq (\log(x/z_0) - \mu t)/\sigma]
= \mathbb{P}\left[W(t)/\sqrt{t} \leq (\log(x/z_0) - \mu t)/(\sigma \sqrt{t})\right]
= \int_{-\infty}^{(\log(x/z_0) - \mu t)/(\sigma \sqrt{t})} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \, dy
\]

Now differentiating with respect to \( x \), we obtain that

\[
f_X(x) = \frac{1}{\sqrt{2\pi}x\sqrt{t}} \exp\left(-\frac{1}{2}\left(\frac{\log(x) - \log(z_0) - \mu t}{\sigma \sqrt{t}}\right)^2\right).
\]

\[\square\]

**Calculation of the Mean.** We can calculate the mean of geometric Brownian motion by using the m.g.f. for the normal distribution.

**Theorem 6.27.** \( \mathbb{E}[z_0 \exp(\mu t + \sigma W(t))] = z_0 \exp(\mu t + (1/2)\sigma^2 t) \).
Figure 2. The p.d.f. (thin line) and c.d.f. (thick line) for a log-normal random variable with $m = 1$, $s = 1.5$.

**Proof.**

$$E[X(t)] = E[z_0 \exp(\mu t + \sigma W(t))]$$
$$= z_0 \exp(\mu t) E[\exp(\sigma W(t))]$$
$$= z_0 \exp(\mu t) E[\exp(\sigma W(t)u) | u = 1]$$
$$= z_0 \exp(\mu t) \exp(\sigma^2 tu^2 / 2) | u = 1$$
$$= z_0 \exp(\mu t + (1/2)\sigma^2 t)$$

since $\sigma W(t) \sim N(0, \sigma^2 t)$ and $E[\exp(Yu)] = \exp(\sigma^2 tu^2 / 2)$ when $Y \sim N(0, \sigma^2 t)$. See Moment Generating Functions.

**Calculation of the Variance.** We can calculate the variance of geometric Brownian motion by using the m.g.f. for the normal distribution, together with the
common formula
\[ \text{Var} [X] = \mathbb{E} [(X - \mathbb{E}[X])^2] = \mathbb{E} [X^2] - (\mathbb{E}[X])^2 \]
and the previously obtained formula for \( \mathbb{E}[X] \).

**Theorem 6.28.** \( \text{Var}[z_0 \exp(\mu t + \sigma W(t))] = z_0^2 \exp(2\mu t + \sigma^2 t)[\exp(\sigma^2 t) - 1]. \)

**Proof.** First compute:
\[
\begin{align*}
\mathbb{E} [X(t)^2] &= \mathbb{E} [z_0^2 \exp(\mu t + \sigma W(t))^2] \\
&= z_0^2 \mathbb{E} \left[ \exp(2\mu t + 2\sigma W(t)) \right] \\
&= z_0^2 \exp(2\mu t) \mathbb{E} \left[ \exp(2\sigma W(t)) \right] \\
&= z_0^2 \exp(2\mu t) \mathbb{E} \left[ \exp(2\sigma W(t)u) \right] |_{u=1} \\
&= z_0^2 \exp(2\mu t) \exp(4\sigma^2 tu^2/2) |_{u=1} \\
&= z_0^2 \exp(2\mu t + 2\sigma^2 t).
\end{align*}
\]

Therefore,
\[
\begin{align*}
\text{Var}[z_0 \exp(\mu t + \sigma W(t))] &= z_0^2 \exp(2\mu t + 2\sigma^2 t) - z_0^2 \exp(2\mu t + \sigma^2 t) \\
&= z_0^2 \exp(2\mu t + 2\sigma^2 t)[\exp(\sigma^2 t) - 1].
\end{align*}
\]
\[\square\]

Note that this has the consequence that the variance starts at 0 and then grows exponentially. The variation of geometric Brownian motion starts small, and then increases, so that the motion generally makes larger and larger swings as time increases.

**Stochastic Differential Equation and Parameter Summary.** If a geometric Brownian motion is defined by the stochastic differential equation
\[
dX = rX \, dt + \sigma X \, dW \quad X(0) = z_0
\]
then the geometric Brownian motion is
\[
X(t) = z_0 \exp((r - (1/2)\sigma^2)t + \sigma W(t)).
\]
At each time the geometric Brownian motion has lognormal distribution with parameters \( \log(z_0) + rt - (1/2)\sigma^2 t \) and \( \sigma \sqrt{t} \). The mean of the geometric Brownian motion is \( \mathbb{E}[X(t)] = z_0 \exp(rt) \). The variance of the geometric Brownian motion is
\[
\text{Var}[X(t)] = z_0^2 \exp(2rt)[\exp(\sigma^2 t) - 1].
\]

If the primary object is the geometric Brownian motion
\[
X(t) = z_0 \exp(\mu t + \sigma W(t)).
\]
then by Itô’s formula the SDE satisfied by this stochastic process is
\[
dX = (\mu + (1/2)\sigma^2)X(t) \, dt + \sigma X(t) \, dW \quad X(0) = z_0.
\]
At each time the geometric Brownian motion has lognormal distribution with parameters \( \log(z_0) + \mu t \) and \( \sigma \sqrt{t} \). The mean of the geometric Brownian motion is \( \mathbb{E}[X(t)] = z_0 \exp(\mu t + (1/2)\sigma^2 t) \). The variance of the geometric Brownian motion is
\[
z_0^2 \exp(2\mu t + \sigma^2 t)[\exp(\sigma^2 t) - 1].
\]
Ruin and Victory Probabilities for Geometric Brownian Motion. Because of the exponential-logarithmic connection between geometric Brownian motion and Brownian motion, many results for Brownian motion can be immediately translated into results for geometric Brownian motion. Here is a result on the probability of victory, now interpreted as the condition of reaching a certain multiple of the initial value. For \( A < 1 < B \) define the duration to ruin or victory, or the hitting time, as

\[
T_{A,B} = \min \{ t \geq 0 : \frac{z_0 \exp(\mu t + \sigma W(t))}{z_0} = A, \frac{z_0 \exp(\mu t + \sigma W(t))}{z_0} = B \}.
\]

**Theorem 6.29.** For a geometric Brownian motion with parameters \( \mu \) and \( \sigma \), and \( A < 1 < B \),

\[
P \left[ \frac{z_0 \exp(\mu T_{A,B} + \sigma W(T_{A,B}))}{z_0} = B \right] = 1 - \frac{A^{1-(2\mu - \sigma^2)/\sigma^2}}{B^{1-(2\mu - \sigma^2)/\sigma^2} - A^{1-(2\mu - \sigma^2)/\sigma^2}}
\]

**Quadratic Variation of Geometric Brownian Motion.** The quadratic variation of geometric Brownian motion may be deduced from Itô’s formula:

\[
dX = (\mu - \sigma^2/2)X \ dt + \sigma X \ dW
\]

so that

\[
(dX)^2 = (\mu - \sigma^2/2)^2X^2 \ dt^2 + (\mu - \sigma^2/2)\sigma X^2 \ dt \ dW + \sigma^2 X^2 (dW)^2.
\]

Guided by the heuristic principle that terms of order \((dt)^3/2\) are small and may be ignored, and that \((dW)^2 = dt\), we obtain:

\[
(dX)^2 \approx \sigma^2 X^2 \ dt.
\]

Continuing heuristically, the expected quadratic variation is

\[
E \left[ \int_0^T (dX)^2 \right] = E \left[ \int_0^T \sigma^2 X^2 \ dt \right]
\]

\[
= \sigma^2 \int_0^T E[X^2] \ dt
\]

\[
= \sigma^2 \int_0^T z_0^2 \exp(2\mu t + 2\sigma^2 t) \ dt
\]

\[
= \frac{\sigma^2 z_0^2}{2\mu + 2\sigma^2} (\exp((2\mu + 2\sigma^2)T) - 1).
\]

Note the assumption that the order of the integration and the expectation can be interchanged.

**Section Ending Answer.** The relative rate of change of a function is

\[
f'(x) \quad \frac{f(x)}{f(x)}.
\]

For the exponential function defined by the ordinary differential equation

\[
\frac{dx}{dt} = rx \quad x(0) = x_0
\]
the relative rate of growth is the constant \( r \). The analog to these common calculus facts in stochastic calculus is geometric Brownian motion.

**Algorithms, Scripts, Simulations.**

*Algorithm.* Given values for \( \mu, \sigma \) and an interval \([0, T]\), the script creates \( \text{trials} \) sample paths of geometric Brownian motion, sampled at \( N \) equally-spaced values on \([0, T]\). The scripts do this by creating \( \text{trials} \) Brownian motion sample paths sampled at \( N \) equally-spaced values on \([0, T]\) using the definition of Brownian motion having normally distributed increments. Adding the drift term and then exponentiating the sample paths creates \( \text{trials} \) geometric Brownian motion sample paths sampled at \( N \) equally-spaced values on \([0, T]\). Then the scripts use semi-logarithmic least-squares statistical fitting to calculate the relative growth rate of the mean of the sample paths. The scripts also compute the predicted relative growth rate to compare it to the calculated relative growth rate. The problems at the end of the section explore plotting the sample paths, comparing the sample paths to the predicted mean with standard deviation bounds, and comparing the mean quadratic variation of the sample paths to the theoretical quadratic variation of geometric Brownian motion.

```
1 mu <- 1
2 sigma <- 0.5
3 T <- 1
4 # length of the interval [0, T] in time units
5
6 trials <- 200
7 N <- 200
8 # number of end-points of the grid including T
9 Delta <- T/N
10 # time increment
11
12 t <- t(seq(0, T, length=N+1) * t(matrix(1, trials, N+1)) )
13 # Note use of the R matrix recycling rules, by columns, so transposes
14 W <- cbind(0, t(apply(sqrt(Delta) * matrix(rnorm(trials*N), trials, N), 1, cumsum)) )
15 # Wiener process, Note the transpose after the apply, (side effect of
16 # apply is the result matches the length of individual calls to FUN,
17 # then the MARGIN dimension/s come next. So it’s not so much
18 # “transposed” as that being a consequence of apply in 2D.) Note
19 # use of recycling with cbind to start at 0
20
21 GBM <- exp(mu*t + sigma*W)
22
23 meanGBM <- colMeans(GBM)
24
25 meanGBM_rate <- lm(log(meanGBM) ~ seq(0, T, length=N+1))
26 predicted_mean_rate = mu + (1/2)*sigma^2
27
28 cat(sprintf("Observed meanGBM relative rate: %.f \n", coefficients(meanGBM_rate)[2])
29 cat(sprintf("Predicted mean relative rate: %.f \n", predicted_mean_rate
```
Key Concepts.

(1) Geometric Brownian motion is the continuous time stochastic process $z_0 \exp(\mu t + \sigma W(t))$ where $W(t)$ is standard Brownian motion.

(2) The mean of geometric Brownian motion is $z_0 \exp(\mu t + (1/2)\sigma^2 t)$.

(3) The variance of geometric Brownian motion is $z_0^2 \exp(2\mu t + \sigma^2 t)(\exp(\sigma^2 t) - 1)$.

Vocabulary.

(1) Geometric Brownian motion is the continuous time stochastic process $z_0 \exp(\mu t + \sigma W(t))$ where $W(t)$ is standard Brownian motion.

(2) A random variable $X$ is said to have the lognormal distribution (with parameters $\mu$ and $\sigma$) if $\log(X)$ is normally distributed ($\log(X) \sim N(\mu, \sigma^2)$). The p.d.f. for $X$ is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp((-1/2)[(\log(x) - \mu)/\sigma]^2).$$

Problems.

Exercise 6.30. Differentiate

$$\int_{-\infty}^{(\log(x/z_0) - \mu t)/\sigma} \frac{1}{\sqrt{2\pi}} \exp(-y^2/2) \, dy$$

to obtain the p.d.f. of geometric Brownian motion.

Exercise 6.31. What is the probability that geometric Brownian motion with parameters $\mu = -\sigma^2/2$ and $\sigma$ (so that the mean is constant) ever rises to more than twice its original value? In economic terms, if you buy a stock or index fund whose fluctuations are described by this geometric Brownian motion, what are your chances to double your money?

Exercise 6.32. What is the probability that geometric Brownian motion with parameters $\mu = 0$ and $\sigma$ ever rises to more than twice its original value? In economic terms, if you buy a stock or index fund whose fluctuations are described by this geometric Brownian motion, what are your chances to double your money?

Exercise 6.33. Derive the probability of ruin (the probability of geometric Brownian motion hitting $A < 1$ before hitting $B > 1$).

Exercise 6.34. Modify the scripts to plot several sample paths of geometric Brownian motion all on the same set of axes.

Exercise 6.35. Modify the scripts to plot several sample paths of geometric Brownian motion and the mean function of geometric Brownian motion and the mean function plus and minus one standard deviation function, all on the same set of axes.

Exercise 6.36. Modify the scripts to measure the quadratic variation of each of many sample paths of geometric Brownian motion, find the mean quadratic variation and compare to the theoretical quadratic variation of geometric Brownian motion.
6.4. Models of Stock Market Prices

**Section Starter Question.** What would be some desirable characteristics for a stochastic process model of a security price?

**Returns, Log-Returns, Compound Returns.** Let \( s_t \) be a sequence of prices on a stock or a portfolio of stocks, measured at regular intervals, say day to day. A natural question is how best to measure the variation of the prices on the time scale of the regular measurement. A first natural definition of variation is the **proportional return** \( r_t \) at time \( t \)

\[
 r_t = \frac{s_t - s_{t-1}}{s_{t-1}}.
\]

The proportional return is usually just called the return, and often it is expressed as a percentage. A benefit of using returns versus prices is **normalization**: measuring all variables in a comparable metric, essentially percentage variation. Using proportional returns allows consistent comparison among two or more securities even though their price sequences may differ by orders of magnitude. Having comparable variations is a requirement for many multidimensional statistical analyses. For example, interpreting a covariance is meaningful when the variables are measured in percentages.

Define the **log-return**

\[
 \rho_t = \log \left( \frac{s_t}{s_{t-1}} \right)
\]

as another measure of variation on the time scale of the sequence of prices. For small returns, the difference between returns and log-returns is small. Notice that

\[
 1 + r_t = \frac{s_t}{s_{t-1}} = \exp \left( \log \left( \frac{s_t}{s_{t-1}} \right) \right) = \exp(\rho_t).
\]

Therefore

\[
 \rho_t = \log(1 + r_t) \approx r_t, \quad \text{for } r_t \ll 1.
\]

More generally, a statistic calculated from a sequence of prices is the **compounding return** at time \( t \) over \( n \) periods, defined as

\[
 \frac{s_t}{s_{t-n}} = (1 + r_t)(1 + r_{t-1}) \cdots (1 + r_{t-n+1}).
\]

Taking logarithms, a simplification arises,

\[
 \log \left( \frac{s_t}{s_{t-n}} \right) = \log(s_t) - \log(s_{t-n}),
\]

the logarithm of the compounding return over \( n \) periods is just the difference between the logarithm of the price at the final and initial periods. Furthermore,

\[
 \log \left( \frac{s_t}{s_{t-n}} \right) = \log ((1 + r_t)(1 + r_{t-1}) \cdots (1 + r_{t-n+1})) \\
 = \rho_1 + \rho_2 + \cdots + \rho_n.
\]

So, the advantage of using log-returns is that they are additive. Recall that the sum of independent normally-distributed random variables is normal. If we *assume* that log-returns are independent and normally distributed then the logarithm of the compounding return is normally distributed. However the product of normally-distributed variables has no easy distribution; in particular, it is not normal. So even if we make the simplifying assumption that the returns are normally distributed, there is no corresponding result for the compounded return.
Modeling from Stochastic Differential Equations. Using Brownian motion for modeling stock prices varying over continuous time has two obvious problems:

1. Even if started from a positive value $X_0 > 0$; at each time there is a positive probability that the process attains negative values; this is unrealistic for stock prices.
2. Stocks selling at small prices tend to have small increments in price over a given time interval, while stocks selling at high prices tend to have much larger increments in price on the same interval. Brownian motion has a variance that depends on a time interval but not on the process value, so this too is unrealistic for stock prices.

In 1965 Nobel prize-winning economist Paul Samuelson proposed a solution to both problems by modeling stock prices as a geometric Brownian motion.

Let $S(t)$ be the continuous-time stock process. The following assumptions about price increments are the foundation for a model of stock prices.

1. Stock price increments have a deterministic component. In a short time, changes in price are proportional to the stock price itself with constant proportionality rate $r$.
2. The stock price increments have a random component. In a short time, changes in price are jointly proportional to the stock price, a standard normal random variable, and the time increment with constant proportionality rate $\sigma$.

The first assumption is based on the observation that stock prices have a general overall growth (or decay if $r < 0$) rate due to economic conditions. For mathematical simplicity, we take $r$ to be constant, because we know that in the absence of randomness, this leads to the exponential function, a simple mathematical function. The second observation is based on the observation that stock prices vary both up and down on short times. The change is apparently random due to a variety of influences, and that the distribution of changes appears to be normally distributed. The normal distribution has many mathematically desirable features, so for simplicity the randomness is taken to be normal. The proportionality constants are taken to be constant for mathematical simplicity.

These assumptions can be mathematically modeled with a stochastic differential equation

$$dS(t) = rS \, dt + \sigma S \, dW(t), \quad S(0) = S_0.$$ 

We already have the solution of this stochastic differential equation as geometric Brownian motion:

$$S(t) = S_0 \exp((r - (1/2)\sigma^2)t + \sigma W(t)).$$

At each time the geometric Brownian motion has a lognormal distribution with parameters $(\log(S_0) + rt - (1/2)\sigma^2t)$ and $\sigma\sqrt{t}$. The mean stock price at any time is $E[X(t)] = S_0 \exp(rt)$. The variance of the stock price at any time is

$$\text{Var}[X(t)] = S_0^2 \exp(2rt)[\exp(\sigma^2t) - 1].$$

Note that with this model, the log-return over a period from $t - n$ to $t$ is $(r - \sigma^2/2)n + \sigma[W(t) - W(t - n)]$. The log-return is normally distributed with mean and variance characterized by the parameters associated with the security.
This is consistent with the assumptions about the distribution of log-returns of regular sequences of security processes.

The constant $r$ is often called the **drift** and $\sigma$ is called the **volatility**. Drifts and volatility are usually reported on an annual basis. Therefore, some care and attention is necessary when converting this annual rate to a daily rate. An “annual basis” in finance often means 252 days per year because there are that many trading days, not counting weekends and holidays in the 365-day calendar. In order to be consistent in this text with most other applied mathematics we will use 365 days for annual rates. If necessary, the conversion of annual rates to financial-year annual rates is a straightforward matter of dividing annual rates by 365 and then multiplying by 252.

**Testing the Assumptions on Data.** The **Wilshire 5000 Total Market Index**, or more simply the Wilshire 5000, is an index of the market value of all stocks actively traded in the United States. The index is intended to measure the performance of publicly traded companies headquartered in the United States. Stocks of extremely small companies are excluded.

In spite of the name, the Wilshire 5000 does not have exactly 5,000 stocks. Developed in the summer of 1974, the index had just shy of the 5,000 issues at that time. The membership count has ranged from 3,069 to 7,562. The member count was 3,818 as of September 30, 2014.

The index is computed as

$$W = \alpha \sum_{i=1}^{M} N_i P_i$$

where $P_i$ is the price of one share of issue $i$ included in the index, $N_i$ is the number of shares of issue $i$, $M$ is the number of member companies included in the index, and $\alpha$ is a fixed scaling factor. The base value for the index was 1404.60 points on base date December 31, 1980, when it had a total market capitalization of $1,404.596$ billion. On that date, each one-index-point change in the index was equal to $1$ billion. However, index divisor adjustments due to index composition changes have changed the relationship over time, so that by 2005 each index point reflected a change of about $1.2$ billion in the total market capitalization of the index.

The index was renamed the “Dow Jones Wilshire 5000” in April 2004, after Dow Jones & Company assumed responsibility for its calculation and maintenance. On March 31, 2009, the partnership with Dow Jones ended and the index returned to Wilshire Associates.

The Wilshire 5000 is the weighted sum of many stock values, each of which we may reasonably assume is a random variable presumably with a finite variance. If the random variables are independent, then the Central Limit Theorem would suggest that the index should be normally distributed. Therefore, a reasonable hypothesis is that the Wilshire 5000 is a normal random variable, although we do not know the mean or variance in advance. (The assumption of independence is probably too strong, since general economic conditions affect most stocks similarly.)

Data for the Wilshire 5000 is easy to obtain. For example, the Yahoo Finance page for W5000 provides a download with the Date, Open, Close, High, Low, Volume and Adjusted Close values of the index in reverse order from today to April 1, 2009, the day Wilshire Associates resumed calculation of the index. (The Adjusted Close is an adjusted price for dividends and splits that does not affect
this analysis.) The data come in the form of a comma-separated-value text file. This file format is well-suited as input for many programs, especially spreadsheets and data analysis programs such as R. This analysis uses R.

The data from December 31, 2014, back to April 1, 2009, provide 1449 records with seven fields each. This analysis uses the logarithm of each of the Close prices. Reversing them and then taking the differences gives 1448 daily log-returns. The mean of the 1448 daily log-returns is 0.0006675644 and the variance of 1448 daily log-returns is 0.0001178775. Assume that the log-returns are normally distributed so that the mean change over a year of 252 trading days is 0.1682262 and the variance over a year of 252 trading days is 0.02970512. Then the annual standard deviation is 0.1722922. The initial value on April 1, 2009, is 8242.38.

Use the values \( r = 0.1682262 \) and \( \sigma = 0.1722922 \) with initial value \( S_0 = 8242.38 \) in the stochastic differential equation

\[
dS(t) = rS \, dt + \sigma S \, dW(t), \quad S(0) = S_0.
\]

The resulting geometric Brownian motion is a model for the Wilshire 5000 index. Figure 3 is a plot of the actual data for the Wilshire 5000 over the period April 1, 2009, to December 31, 2014, along with a simulation using a geometric Brownian motion with these parameters.

Testing the Normality Hypothesis. Comparison of the graphs of the actual Wilshire 5000 data with the geometric Brownian motion suggests that geometric Brownian motion gives plausible results. However, deeper investigation to find if the fundamental modeling hypotheses are actually satisfied is important.

Consider again the 1448 daily log-returns. The log-returns are normalized by subtracting the mean and dividing by the standard deviation of the 1448 changes. The maximum of the 1448 normalized changes is 4.58 and the minimum is −6.81. Already we have a hint that the distribution of the data is not normally distributed, since the likelihood of seeing normally distributed data varying 4 to 6 standard deviations from the mean is negligible.

In R, the \texttt{hist} command on the normalized data gives an empirical density histogram. For simplicity, here the histogram is taken over the 14 one-standard-deviation intervals from −7 to 7. For this data, the density histogram of the normalized data has the values in Table 1.

<table>
<thead>
<tr>
<th>Interval</th>
<th>Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>(−7, −6]</td>
<td>0.00069</td>
</tr>
<tr>
<td>(−6, −5]</td>
<td>0.00000</td>
</tr>
<tr>
<td>(−5, −4]</td>
<td>0.00276</td>
</tr>
<tr>
<td>(−4, −3]</td>
<td>0.00622</td>
</tr>
<tr>
<td>(−3, −2]</td>
<td>0.02693</td>
</tr>
<tr>
<td>(−2, −1]</td>
<td>0.08771</td>
</tr>
<tr>
<td>(−1, 0]</td>
<td>0.35428</td>
</tr>
</tbody>
</table>

Table 1. Relative frequency distribution by unit intervals for the normalized Wilshire 5000 data.

This means, for example, that 0.00069 of the 1448 log-returns, that is 1, occurred in the interval (−7, −6] and a fraction 0.4088397790, or 592 points, fall in the interval (0, 1]. The normal distribution gives the expected density on the same intervals. The ratio between the empirical density and normal density gives an indication of the deviation from normality. For data within 4 standard deviations, the ratios are about what we would expect. This is reassuring that for reasonably small changes, the log-returns are approximately normal. However, the ratio on the
Figure 3. The Wilshire 5000 Index from April 1, 2009, to December 31, 2014, plotted (thin line) along with a geometric Brownian motion having the same mean, variance and starting value (thick line).

interval (−7, 6] is approximately 70,000, and the ratio on the interval (4, 5] is approximately 66. Each of these is much greater than we expect; that is, the extreme tails of the empirical density have much greater probability than expected.

Quantile-Quantile Plots. A quantile-quantile (q-q) plot is a graphical technique for determining if two data sets come from populations with a common distribution. A q-q plot is a plot of the quantiles of the first data set against the quantiles of the second data set. By a quantile, we mean the fraction (or percent) of points below the given value. That is, the 0.3 (or 30%) quantile is the data point at which 30% percent of the data fall below and 70% fall above. As another example, the median is the 0.5 quantile.

A q-q plot is formed by plotting estimated quantiles from data set 2 on the vertical axis against estimated quantiles from data set 1 on the horizontal axis.
Both axes are in units of their respective data sets. For a given point on the q-q plot, we know the quantile level is the same for both points, but not what the quantile level actually is.

If the data sets have the same size, the q-q plot is essentially a plot of sorted data set one against sorted data set two. If the data sets are not of equal size, the quantiles are usually picked to correspond to the sorted values from the smaller data set and then the quantiles for the larger data set are interpolated from the data.

The quantiles from the normal distribution are the values from the inverse cumulative distribution function. These quantiles are tabulated, or they may be obtained from statistical software. For example, in R the function `qnorm` gives the quantiles, e.g. `qnorm(0.25) = -0.6744898`, and `qnorm(0.90) = 1.281552`. Estimating the quantiles for the normalized Wilshire 5000 data is more laborious but is not difficult in principle. Sorting the data and then finding the value for which $100k\%$ of the data is less and $100(1-k)\%$ of the data is greater gives the $k$ percentile for $0 \leq k \leq 1$. For example, the 0.25-quantile for the daily log-changes is $-0.4342129$ and the 0.90-quantile for the daily daily log-changes is $1.072526$. Then for the q-q plot, plot the values $(-0.6744898, -0.4342129)$ and $(1.281552, 1.072526)$. Using many quantiles gives the full q-q plot.

In R, to create a q-q plot of the normalized daily log-changes for the Wilshire 5000 data with a reference line against the normal distribution use the two commands `qqnorm()` and `qqline()` on the normalized data of changes in the logarithm of the close values. In the q-q plot in Figure 4, the “twist” of the plot above and below the reference line indicates that the tails of the normalized Wilshire 5000 data are more dispersed than standard normal data. The low quantiles of the normalized Wilshire 5000 quantiles occur at more negative values than the standard normal distribution. The high quantiles occur at values greater than the standard normal distribution. However, for quantiles near the median, the data does seem to follow the normal distribution. The plot is a graphical representation of the fact that extreme events are more likely to occur than would be predicted by a normal distribution.

**Summary.** Using log-returns for stock price has the simplifying advantage of additivity. If we assume that log-returns are normally distributed, then the logarithm of the compounding return is normally distributed.

Two problems with using Brownian motion for modeling stock prices varying over continuous time lead to modeling stock prices with the stochastic differential equation for geometric Brownian motion. The log-return over a period from $t - n$ to $t$ is $(r - \sigma^2/2)n + \sigma[W(t) - W(t-n)]$. The log-return is normally distributed with mean and variance characterized by the parameters associated with the security.

The modeling assumptions can be tested against data for the Wilshire 5000 Index. The geometric Brownian motion models using the parameters for the Wilshire 5000 Index resemble the actual data. However, deeper statistical investigation of the log-returns shows that while log-returns within 4 standard deviations from the mean are normally distributed, extreme events are more likely to occur than would be predicted by a normal distribution.

**Section Ending Answer.** Desirable characteristics for a stochastic process model for a security price are that it should be a positive process. At each time, the
process should have a probability distribution that is reasonable to work with, either mathematically or computationally. It should also mimic available security data as well as possible. Geometric Brownian motion satisfies the first two characteristics, but it only mimics the data for the Wilshire 5000 for small return values.

**Algorithms, Scripts, Simulations.**

*Algorithm.* This script simulates the Wilshire 5000 Index over the period April 1, 2009, to December 31, 2014, using geometric Brownian motion with drift and standard deviation parameters calculated from the data over that same period. The resulting simulation is plotted over that 5.75 year time period. Set the drift $r$, variance $\sigma$, time interval $T$, and starting value $S_0$. Set the number of time divisions $n$ and compute the time increment $\Delta$. Compute a Wiener process simulation, then apply the exponential to create a geometric Brownian motion with the parameters.
Plot the simulation on the time interval, or output the data to a file for later plotting or use.

```r
mu <- 0.1682262
sigma <- 0.1722922
T <- 5.75
# length of the interval [0, T] in time units of years
S0 <- 8242.38

N <- 1448
# number of end-points of the grid including T
Delta <- T/N
# time increment,

t <- seq(0, T, length = N + 1)
# Note use of the R matrix recycling rules, by columns, so transposes
W <- c(0, cumsum( sqrt(Delta) * rnorm(N)))
# Wiener process,
GBM <- S0 * exp(mu * t + sigma * W)

plot(t, GBM, type = "l", xaxt = "n", ylab = "Simulated Wilshire 5000 Index")
axis(1, at = c(0.75, 1.75, 2.75, 3.75, 4.75, 5.75), label = c("2010", "2011", "2012", "2013", "2014", "2015"))
```

```r
wil5000Data <- read.csv("table.csv", stringsAsFactors = FALSE)
closingValue <- wil5000Data$Close
closingValue <- rev(closingValue)
Time <- as.Date(rev(wil5000Data$Date))
y <- cbind(closingValue, GBM)
```
### Key Concepts.

1. A natural definition of variation of a stock price $s_t$ is the **proportional return** $r_t$ at time $t$

   $$r_t = \left( \frac{s_t - s_{t-1}}{s_{t-1}} \right).$$

2. The **log-return**

   $$\rho_t = \log\left( \frac{s_t}{s_{t-1}} \right)$$

   is another measure of variation on the time scale of the sequence of prices.

3. For small returns, the difference between returns and log-returns is small.

4. The advantage of using log-returns is that they are additive.

5. Using Brownian motion for modeling stock prices varying over continuous time has two obvious problems:
   - Brownian motion can attain negative values.
   - Increments in Brownian motion have certain variance on a given time interval, so do not reflect proportional changes.

6. Modeling security price changes with a stochastic differential equation leads to a geometric Brownian motion model.

7. Deeper statistical investigation of the log-returns shows that while log-returns within 4 standard deviations from the mean are normally distributed, extreme events are more likely to occur than would be predicted by a normal distribution.

### Vocabulary.

1. A natural definition of variation of a stock price $s_t$ is the **proportional return** $r_t$ at time $t$

   $$r_t = \left( \frac{s_t - s_{t-1}}{s_{t-1}} \right).$$

2. The **log-return**

   $$\rho_t = \log\left( \frac{s_t}{s_{t-1}} \right)$$

   is another measure of variation on the time scale of the sequence of prices.

3. The **compounding return** at time $t$ over $n$ periods is

   $$\frac{s_t}{s_{t-n}} = (1 + r_t)(1 + r_{t-1}) \cdots (1 + r_{t-n+1})$$

4. The **Wilshire 5000 Total Market Index**, or more simply the Wilshire 5000, is an index of the market value of all stocks actively traded in the United States.

5. A **quantile-quantile (q-q) plot** is a graphical technique for determining if two data sets come from populations with a common distribution.
Problems.

Exercise 6.37. Show that
\[ \rho_t = \log(1 + r_t) \approx r_t, \text{ for } r \ll 1. \]

Exercise 6.38. If standard Brownian motion is started from a positive value \( X_0 > 0 \), write the expression for the positive probability that the process can attain negative values at time \( t \). Write the expression for the probability that Brownian motion can ever become negative.

Exercise 6.39. Choose a stock index such as the S & P 500, the Dow Jones Industrial Average etc., and obtain closing values of that index for a year-long (or longer) interval of trading days. Find the growth rate and variance of the closing values and create a geometric Brownian motion on the same interval with the same initial value, growth rate, and variance. Plot both sets of data on the same axes, as in Figure 3. Discuss the similarities and differences.

Exercise 6.40. Choose an individual stock or a stock index such as the S & P 500, the Dow Jones Industrial Average, etc., and obtain values of that index at regular intervals such as daily or hourly for a long interval of trading. Find the log-changes, and normalize by subtracting the mean and dividing by the standard deviation. Create a q-q plot for the data as in Figure 4. Discuss the similarities and differences.
The Black-Scholes Equation

7.1. Derivation of the Black-Scholes Equation

Section Starter Question. What is the most important idea in the derivation of the binomial option pricing model?

Explicit Assumptions Made for Modeling and Derivation. For mathematical modeling of a market for a risky security we will ideally assume:

1. Many identical, rational traders always have complete information about all assets each is trading.
2. Changes in prices are given by a continuous random variable with some probability distribution.
3. Trading transactions take negligible time.
4. Purchases and sales can be in any amounts; that is, the stocks and bonds are divisible, and we can buy them in any amounts including negative amounts, which are short positions.
5. The risky security issues no dividends.

The first assumption is the essence of what economists call the efficient market hypothesis. The efficient market hypothesis leads to the second assumption as a conclusion, called the random walk hypothesis. Another version of the random walk hypothesis says that traders cannot predict the direction of the market or the magnitude of the change in a stock so the best predictor of the market value of a stock is the current price. We will make the second assumption stronger and more precise by specifying the probability distribution of the changes with a stochastic differential equation. The remaining hypotheses are simplifying assumptions that can be relaxed at the expense of more difficult mathematical modeling.

We wish to find the value $V$ of a derivative instrument based on an underlying security that has value $S$. Mathematically, we assume:

1. The price of the underlying security follows the stochastic differential equation
   \[ dS = rS \, dt + \sigma S \, dW \]
   or equivalently that $S(t)$ is a geometric Brownian motion with parameters $r - \sigma^2/2$ and $\sigma$.
2. The risk free interest rate $r$ and the volatility $\sigma$ are constants.
3. The value $V$ of the derivative depends only on the current value of the underlying security $S$ and the time $t$, so we can write $V(S, t)$.
4. All variables are real-valued, and all functions are sufficiently smooth to justify necessary calculus operations.

The first assumption is a mathematical translation of a strong form of the efficient market hypothesis from economics. It is a reasonable modeling assumption.
but finer analysis strongly suggests that security prices have a higher probability of large price swings than geometric Brownian motion predicts. Therefore the first assumption is not supported by data. However, it is useful since we have good analytic understanding of geometric Brownian motion.

The second assumption is a reasonable assumption for a modeling attempt although good evidence indicates neither interest rates nor the volatility are constant. On reasonably short times scales, say a period of three months for the lifetime of most options, the interest rate and the volatility are approximately constant. The third and fourth assumptions are mathematical translations of the assumptions that securities can be bought and sold in any amount and that trading times are negligible, so that standard tools of mathematical analysis apply. Both assumptions are reasonable for modern security trading.

**Derivation of the Black-Scholes equation.** We consider a simple derivative instrument, an option written on an underlying asset, say a stock that trades in the market at price \( S(t) \). A payoff function \( \Lambda(S) \) determines the value of the option at expiration time \( T \). For \( t < T \), the option value should depend on the underlying price \( S \) and the time \( t \). We denote the price as \( V(S, t) \). So far all we know is the value at the final time \( V(S, T) = \Lambda(S) \). We would like to know the value \( V(S, 0) \) so that we know an appropriate buying or selling price of the option.

As time passes, the value of the option changes, both because the contract approaches its expiration date and because the stock price changes. We assume the option price changes smoothly in both the security price and the time. Across a short time interval \( \delta t \) we can write the Taylor series expansion of \( V \):

\[
\delta V = V_t \delta t + V_S \delta S + \frac{1}{2} V_{SS} (\delta S)^2 + \cdots
\]

The neglected terms are of order \((\delta t)^2\), \(\delta S \delta t\), and \((\delta S)^3\) and higher. We rely on our intuition from random walks and Brownian motion to explain why we keep the terms of order \((\delta S)^2\) but neglect the other terms. More about this later.

To determine the price, we construct a replicating portfolio. This will be a specific investment strategy involving only the stock and a cash account that will yield exactly the same eventual payoff as the option in all possible future scenarios. Its present value must therefore be the same as the present value of the option and if we can determine one we can determine the other. We thus define a portfolio \( \Pi \) consisting of \( \phi(t) \) shares of stock and \( \psi(t) \) units of the cash account \( B(t) \). The portfolio constantly changes in value as the security price changes randomly and the cash account accumulates interest.

In a short time interval, we can take the changes in the portfolio to be

\[
\delta \Pi = \phi(t) \delta S + \psi(t) r B(t) \delta t
\]

since \( \delta B(t) \approx r B(t) \delta t \), where \( r \) is the interest rate. This says that short-time changes in the portfolio value are due only to changes in the security price, and the interest growth of the cash account, and not to additions or subtraction of the portfolio amounts. Any additions or subtractions are due to subsequent reallocations financed through the changes in the components themselves.

The difference in value between the two portfolios changes by

\[
\delta (V - \Pi) = (V_t - \psi(t) r B(t)) \delta t + (V_S - \phi(t)) \delta S + \frac{1}{2} V_{SS} (\delta S)^2 + \cdots
\]
This represents changes in a three-part portfolio consisting of an option and short-selling $\phi$ units of the security and $\psi$ units of bonds.

Next come a couple of linked insights: As an initial insight we will eliminate the first-order dependence on $S$ by taking $\phi(t) = V_S$. Note that this means the rate of change of the derivative value with respect to the security value determines a policy for $\phi(t)$. Looking carefully, we see that this policy removes the randomness from the equation for the difference in values! (What looks like a little trick right here hides a world of probability theory. This is really a Radon-Nikodym derivative that defines a change of measure that transforms a diffusion, i.e. a transformed Brownian motion with drift, to a standard Wiener measure.)

Second, since the difference portfolio is now non-risky, it must grow in value at exactly the same rate as any risk-free bank account:

$$\delta(V - \Pi) = r(V - \Pi)\delta t.$$  

Recall the quadratic variation of geometric Brownian motion is deterministic, namely $(\delta S)^2 = \sigma^2 S(t)^2 \delta t$,

$$r(V - \Pi)\delta t = (V_t - \psi(t)rB(t))\delta t + \frac{1}{2}V_{SS}(\delta S)^2.$$  

Divide out the $\delta t$ factor, and recall that $V - \Pi = V - \phi(t)S - \psi(t)B(t)$, and $\phi(t) = V_S$, so that on the left $r(V - \Pi) = rV - rV_S S - r\psi(t)B(t)$. The terms $-r\psi(t)B(t)$ on left and right cancel, and we are left with the Black-Scholes equation:

$$V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + rSV_S - rV = 0.$$  

Note that under the assumptions made for the purposes of the modeling the partial differential equation depends only on the constant volatility $\sigma$ and the constant risk-free interest rate $r$. This partial differential equation (PDE) must be satisfied by the value of any derivative security depending on the asset $S$.

There are two things to immediately notice about the PDE.

(1) The PDE is linear: Since the solution of the PDE is the worth of the option, then two options are worth twice as much as one option, and a portfolio consisting two different options has value equal to the sum of the individual options. A linear PDE is a reasonable modeling outcome.

(2) The PDE is backwards parabolic because it is a partial differential equation of the form $V_t + DV_{xx} + \cdots = 0$ with highest derivative terms in $t$ of order 1 and highest derivative terms $x$ of order 2 respectively. Thus, a terminal value $V(S,T)$ at an end time $t = T$ must be specified.
in contrast to the initial values at $t = 0$ required by many problems in physics and engineering. The value of a European option at expiration is known as a function of the security price $S$, so we have a terminal value. The PDE determines the value of the option at times before the expiration date.

**Comment on the derivation:** The derivation above generally follows the original derivation of Black, Scholes and Merton. Option prices can also be calculated and the Black-Scholes equation derived by more advanced probabilistic methods. In this equivalent formulation, the discounted price process $\exp(-rt)S(t)$ is shifted into a risk-free measure using the Girsanov Theorem, so that it becomes a martingale. The option price $V(S, t)$ is then the discounted expected value of the payoff $\Lambda(t)$ in this measure, and the PDE is obtained as the backward evolution equation for the expectation. The probabilistic view is more modern and can be more easily extended to general market models.

The derivation of the Black-Scholes equation above uses the fairly intuitive partial derivative equation approach because of the simplicity of the derivation. This derivation:

- is easily motivated and related to similar derivations of partial differential equations in physics and engineering;
- avoids the burden of developing additional probability theory and measure theory machinery, including filtrations, sigma-fields, previsibility, and changes of measure including Radon-Nikodym derivatives and the Girsanov Theorem;
- also avoids, or at least hides, martingale theory that we have not yet developed or exploited; and
- does depend on the stochastic process knowledge that we have gained already, but not more than that knowledge!

The disadvantages are that:

- we must skim certain details, relying on motivation instead of strict mathematical rigor;
- we still have to solve the partial differential equation to get the price on the derivative, whereas the probabilistic methods deliver the solution almost automatically and give the partial differential equation as an auxiliary by-product; and
- the probabilistic view is more modern and can be more easily extended to general market models.

**Section Ending Answer.** The most important idea in the derivation of the binomial option pricing model is that an option on a security can be replicated by a portfolio consisting of the security and a risk-free bond. That idea appears again here as the instantaneously adjusted hedging strategy.

**Problems.**

**Exercise 7.1.** Show by substitution that two exact solutions of the Black-Scholes equations are

1. $V(S, t) = AS$, $A$ some constant.
2. $V(S, t) = A \exp(rt)$
Explain in financial terms what each of these solutions represents. That is, describe a simple “claim”, “derivative” or “option” for which this solution to the Black Scholes equation gives the value of the claim at any time.

**Exercise 7.2.** Draw the expiry diagrams (that is, a graph of terminal condition of portfolio value versus security price $S$) at the expiration time for the portfolio which is

1. Short one share, long two calls both with exercise price $K$. (This is called a straddle.)
2. Long one call, and one put both with exercise price $K$. (This is also called a straddle.)
3. Long one call, and two puts, all with exercise price $K$. (This is called a strip.)
4. Long one put, and two calls, all with exercise price $K$. (This is called a strap.)
5. Long one call with exercise price $K_1$ and one put with exercise price $K_2$. Compare the three cases when $K_1 > K_2$, (known as a strangle), $K_1 = K_2$, and $K_1 < K_2$.
6. As before, but also short one call and one put with exercise price $K$. (When $K_1 < K < K_2$, this is called a butterfly spread.)

### 7.2. Solution of the Black-Scholes Equation

**Section Starter Question.** What is the solution method for the Euler equidi-mensional (or Cauchy-Euler) type of ordinary differential equation:

$$x^2 \frac{d^2 v}{dx^2} + ax \frac{dv}{dx} + bv = 0？$$

**Conditions for Solution of the Black-Scholes Equation.** We have to start somewhere, and to avoid the problem of deriving everything back to calculus, we will assert that the initial value problem for the heat equation on the real line is well-posed. That is, consider the solution to the partial differential equation

$$\frac{\partial u}{\partial \tau} = \frac{\partial^2 u}{\partial x^2} - \infty < x < \infty, \; \tau > 0$$

with the initial condition

$$u(x,0) = u_0(x).$$

Assume the initial condition and the solution satisfy the following technical requirements:

1. $u_0(x)$ has at most a finite number of discontinuities of the jump kind.
2. $\lim_{|x| \to \infty} u_0(x)e^{-ax^2} = 0$ for any $a > 0$.
3. $\lim_{|x| \to \infty} u(x,\tau)e^{-ax^2} = 0$ for any $a > 0$.

Under these mild assumptions, the solution exists for all time and is unique. Most importantly, the solution is represented as

$$u(x, \tau) = \frac{1}{2\sqrt{\pi\tau}} \int_{-\infty}^{\infty} u_0(s)e^{-\frac{(x-s)^2}{4\tau}} ds.$$

**Remark 7.3.** This solution can be derived in several different ways, the easiest being to use Fourier transforms. The derivation of this solution representation is standard in any course or book on partial differential equations.
Remark 7.4. Mathematically, the conditions above are unnecessarily restrictive, and can be considerably weakened. However, they will be more than sufficient for all practical situations we encounter in mathematical finance.

Remark 7.5. The use of $\tau$ for the time variable (instead of the more natural $t$) is to avoid a conflict of notation in the several changes of variables we will soon have to make.

The Black-Scholes terminal value problem for the value $V(S,t)$ of a European call option on a security with price $S$ at time $t$ is

$$
\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} - rV = 0
$$

with $V(0,t) = 0$, $V(S,t) \sim S$ as $S \to \infty$ and

$$
V(S,T) = \max(S - K, 0).
$$

Note that this looks a little like the heat equation on the infinite interval in that it has a first derivative of the unknown with respect to time and the second derivative of the unknown with respect to the other (space) variable. On the other hand, notice:

1. Each time the unknown is differentiated with respect to $S$, it also multiplied by the independent variable $S$, so the equation is not a constant coefficient equation.
2. There is a first derivative of $V$ with respect to $S$ in the equation.
3. The sign on the second derivative is the opposite of the heat equation form, so the equation is of backward parabolic form.
4. The data of the problem is given at the final time $T$ instead of the initial time 0, consistent with the backward parabolic form of the equation.
5. There is a boundary condition $V(0,t) = 0$ specifying the value of the solution at one sensible boundary of the problem. The boundary is sensible since security values must only be zero or positive. This boundary condition says that any time the security value is 0, then the call value (with strike price $K$) is also worth 0.
6. There is another boundary condition $V(S,t) \sim S$, as $S \to \infty$, but although this is financially sensible (it says that for very large security prices, the call value with strike price $K$ is approximately $S$), it is more in the nature of a technical condition, and we will ignore it without consequence.

We eliminate each objection with a suitable change of variables. The plan is to change variables to reduce the Black-Scholes terminal value problem to the heat equation, then to use the known solution of the heat equation to represent the solution, and finally to change variables back. This is a standard solution technique in partial differential equations. All the transformations are standard and well-motivated.

Solution of the Black-Scholes Equation. First we take $t = T - \frac{\tau}{(1/2)\sigma^2}$ and $S = Ke^\tau$, and we set

$$
V(S,t) = Kv(x, \tau).
$$

Remember, $\sigma$ is the volatility, $r$ is the interest rate on a risk-free bond, and $K$ is the strike price. In the changes of variables above, the choice for $t$ reverses the
7.2. SOLUTION OF THE BLACK-SCHOLES EQUATION

sense of time, changing the problem from backward parabolic to forward parabolic. The choice for $S$ is a well-known transformation based on experience with the Euler equidimensional equation in differential equations. In addition, the variables have been carefully scaled so as to make the transformed equation expressed in dimensionless quantities. All of these techniques are standard and are covered in most courses and books on partial differential equations and applied mathematics.

Some extremely wise advice from Stochastic Calculus and Financial Applications by J. Michael Steele, [63, page 186], is appropriate here.

There is nothing particularly difficult about changing variables and transforming one equation to another, but there is an element of tedium and complexity that slows us down. There is no universal remedy for this molasses effect, but the calculations do seem to go more quickly if one follows a well-defined plan. If we know that $V(S, t)$ satisfies an equation (like the Black-Scholes equation) we are guaranteed that we can make good use of the equation in the derivation of the equation for a new function $v(x, \tau)$ defined in terms of the old if we write the old $V$ as a function of the new $v$ and write the new $\tau$ and $x$ as functions of the old $t$ and $S$. This order of things puts everything in the direct line of fire of the chain rule; the partial derivatives $V_t$, $V_S$ and $V_{SS}$ are easy to compute and at the end, the original equation stands ready for immediate use.

Following the advice, write

$$\tau = \frac{\sigma^2}{2} (T - t)$$

and

$$x = \log \left( \frac{S}{K} \right).$$

The first derivatives are

$$\frac{\partial V}{\partial t} = K \frac{\partial v}{\partial t} \cdot \frac{d\tau}{dt} = K \frac{\partial v}{\partial \tau} \cdot -\frac{\sigma^2}{2}$$

and

$$\frac{\partial V}{\partial S} = K \frac{\partial v}{\partial x} \cdot \frac{dx}{dS} = K \frac{\partial v}{\partial x} \cdot \frac{1}{S}.$$

The second derivative is

$$\frac{\partial^2 V}{\partial S^2} = \frac{\partial}{\partial S} \left( \frac{\partial V}{\partial S} \right)
= \frac{\partial}{\partial S} \left( K \frac{1}{x S} \right)
= K \frac{\partial v}{\partial x} \cdot -\frac{1}{S^2} + K \frac{\partial}{\partial S} \left( \frac{\partial v}{\partial x} \right) \cdot \frac{1}{S}
= K \frac{\partial v}{\partial x} \cdot -\frac{1}{S^2} + K \frac{\partial^2 v}{\partial x^2} \cdot \frac{dx}{dS} \cdot \frac{1}{S}
= K \frac{\partial v}{\partial x} \cdot -\frac{1}{S^2} + K \frac{\partial^2 v}{\partial x^2} \cdot \frac{1}{S^2}.$$
The terminal condition is

\[ V(S, T) = \max(S - K, 0) = \max(Ke^x - K, 0) \]

but \( V(S, T) = K v(x, 0) \) so \( v(x, 0) = \max(e^x - 1, 0) \).

Now substitute all of the derivatives into the Black-Scholes equation to obtain:

\[
K \frac{\partial v}{\partial \tau} - \frac{\sigma^2}{2} S^2 \left( K \frac{\partial v}{\partial x} + \frac{1}{S^2} \frac{\partial^2 v}{\partial x^2} \right) + rS \left( K \frac{\partial v}{\partial x} \right) - rKv = 0.
\]

Now begin the simplification:

1. Divide out the common factor \( K \).
2. Transpose the \( \tau \)-derivative to the other side, and divide through by \( \sigma^2 \).
3. Rename the remaining constant \( \frac{\sigma^2}{2} \) as \( k \) which measures the ratio between the risk-free interest rate and the volatility.
4. Cancel the \( S^2 \) terms in the second derivative.
5. Cancel the \( S \) terms in the first derivative.
6. Gather up like-order terms.

What remains is the rescaled, constant coefficient equation now expressed with subscript derivatives for compactness

\[ u_{\tau} = u_{xx} + (k - 1)u_x - ku. \]

(1) Now there is only one dimensionless parameter \( k \) measuring the risk-free interest rate as a multiple of the volatility and a rescaled time to expiry \( \frac{\sigma^2}{2} T \), not the original four dimensioned quantities \( K, T, \sigma^2 \) and \( r \).

(2) The equation is defined on the interval \(-\infty < x < \infty\), since this \( x \)-interval defines \( 0 < S < \infty \) through the change of variables \( S = Ke^x \).

(3) The equation now has constant coefficients.

In principle, we could now solve the equation directly.

Instead, we will simplify further by changing the dependent variable scale yet again, by

\[ v = e^{\alpha x + \beta \tau} u(x, \tau) \]

where \( \alpha \) and \( \beta \) are yet to be determined. Using the product rule:

\[ v_\tau = \beta e^{\alpha x + \beta \tau} u + e^{\alpha x + \beta \tau} u_\tau \]

and

\[ v_x = \alpha e^{\alpha x + \beta \tau} u + e^{\alpha x + \beta \tau} u_x \]

and

\[ v_{xx} = \alpha^2 e^{\alpha x + \beta \tau} u + 2\alpha e^{\alpha x + \beta \tau} u_x + e^{\alpha x + \beta \tau} u_{xx}. \]

Put these into our constant coefficient partial differential equation, divide the common factor of \( e^{\alpha x + \beta \tau} \) throughout and obtain:

\[ \beta u + u_\tau = \alpha^2 u + 2\alpha u_x + u_{xx} + (k - 1)(\alpha u + u_x) - ku. \]

Gather like terms:

\[ u_\tau = u_{xx} + [2\alpha + (k - 1)]u_x + [\alpha^2 + (k - 1)\alpha - k - \beta]u. \]

Choose \( \alpha = -\frac{k - 1}{2} \) so that the \( u_x \) coefficient is 0, and then choose \( \beta = \alpha^2 + (k - 1)\alpha - k = -\frac{(k+1)^2}{4} \) so the \( u \) coefficient is likewise 0. With this choice, the equation is reduced to

\[ u_\tau = u_{xx}. \]
We need to transform the initial condition too. This transformation is
\[ u(x, 0) = e^{-\frac{(k+1)}{2}x} v(x, 0) = e^{\frac{(k+1)}{2}x} \max(e^x - 1, 0) = \max(e^{\frac{(k+1)}{2}x} - e^{\frac{(k-1)}{2}x}, 0). \]

For future reference, we notice that this function is strictly positive when the argument \( x \) is strictly positive; that is \( u_0(x) > 0 \) when \( x > 0 \), otherwise, \( u_0(x) = 0 \) for \( x \leq 0 \).

We are in the final stage since we are ready to apply the heat-equation solution representation formula:
\[ u(x, \tau) = \frac{1}{\sqrt{2\pi \tau}} \int_{-\infty}^{\infty} u_0(s)e^{-\frac{(x-s)^2}{4\tau}} \, ds. \]

However first we want to make a change of variable in the integration by taking \( z = \frac{(x-s)}{\sqrt{2\tau}} \); (and thereby \( dz = \frac{-1}{\sqrt{2\tau}} \, dx \)), so that the integration becomes:
\[ u(x, \tau) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u_0(z\sqrt{2\tau} + x) e^{-\frac{z^2}{2}} \, dz. \]

We may as well only integrate over the domain where \( u_0 > 0 \), that is for \( z > -\frac{x}{\sqrt{2\tau}} \).

On that domain, \( u_0 = e^{\frac{k+1}{2}(x+z\sqrt{2\tau})} - e^{\frac{k-1}{2}(x+z\sqrt{2\tau})} \) so we are down to:
\[ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{k+1}{2}(x+z\sqrt{2\tau})} e^{-\frac{z^2}{2}} \, dz - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{k-1}{2}(x+z\sqrt{2\tau})} e^{-\frac{z^2}{2}} \, dz. \]

Call the two integrals \( I_1 \) and \( I_2 \) respectively.

We will evaluate \( I_1 \) (the integral with the \( k+1 \) term) first. This is easy since completing the square in the exponent yields a standard, tabulated integral. The exponent is
\[
\frac{k+1}{2} \left( x + z\sqrt{2\tau} \right) - \frac{z^2}{2} = \frac{-1}{2} \left( z^2 - \sqrt{2\tau} (k+1) z + \left( \frac{k+1}{2} \right)^2 x \right) + \left( \frac{k+1}{2} \right)^2 x.
\]

Therefore
\[ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{k+1}{2}(x+z\sqrt{2\tau})} e^{-\frac{z^2}{2}} \, dz = \frac{e^{\frac{(k+1)x}{2} + \frac{(k+1)^2}{4}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \left( z - \sqrt{\tau/2}(k+1) \right)^2} \, dz. \]
Now, change variables again in the integral, choosing \( y = z - \sqrt{\tau/2} (k + 1) \) so \( dy = dz \), and all we need to change are the limits of integration:

\[
\frac{e^{(k+1)x + \frac{(k+1)^2}{4}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} \, dz.
\]

The integral can be represented in terms of the cumulative distribution function of a normal random variable, usually denoted \( \Phi \). That is,

\[
\Phi(d) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d} e^{-\frac{y^2}{2}} \, dy
\]

so

\[
I_1 = e^{(k+1)x + \frac{(k+1)^2}{4}} \Phi(d_1)
\]

where \( d_1 = \frac{x}{\sqrt{2\tau}} + \sqrt{\frac{\tau}{2}} (k + 1) \). Note the use of the symmetry of the integral! The calculation of \( I_2 \) is identical, except that \((k + 1)\) is replaced by \((k - 1)\) throughout.

The solution of the transformed heat equation initial value problem is

\[
u(x, \tau) = e^{(k+1)x + \frac{(k+1)^2}{4}} \Phi(d_1) - e^{(k-1)x + \frac{(k-1)^2}{4}} \Phi(d_2)
\]

where \( d_1 = \frac{x}{\sqrt{2\tau}} + \sqrt{\frac{\tau}{2}} (k + 1) \) and \( d_2 = \frac{x}{\sqrt{2\tau}} + \sqrt{\frac{\tau}{2}} (k - 1) \).

Now we must systematically unwind each of the changes of variables, starting from \( u \). First, \( v(x, \tau) = e^{-\frac{(k-1)x^2}{2}} u(x, \tau) \). Notice how many of the exponentials neatly combine and cancel! Next put \( x = \log (S/K), \tau = (\frac{1}{2}) \sigma^2 (T - t) \) and \( V(S, t) = K v(x, \tau) \).

The ultimate result is the Black-Scholes formula for the value of a European call option at time \( T \) with strike price \( K \), if the current time is \( t \) and the underlying security price is \( S \), the risk-free interest rate is \( r \) and the volatility is \( \sigma \):

\[
V(S, t) = S \Phi \left( \frac{\log(S/K) + (r + \sigma^2/2)(T - t)}{\sigma \sqrt{T - t}} \right)
- Ke^{-r(T-t)} \Phi \left( \frac{\log(S/K) + (r - \sigma^2/2)(T - t)}{\sigma \sqrt{T - t}} \right).
\]

Usually one doesn’t see the solution as this full closed form solution. Most versions of the solution write intermediate steps in small pieces, and then present the solution as an algorithm putting the pieces together to obtain the final answer. Specifically, let

\[
d_1 = \frac{\log(S/K) + (r + \sigma^2/2)(T - t)}{\sigma \sqrt{T - t}}
\]
\[
d_2 = \frac{\log(S/K) + (r - \sigma^2/2)(T - t)}{\sigma \sqrt{T - t}}
\]

and writing \( V_C(S, t) \) to remind ourselves this is the value of a call option,

\[
V_C(S, t) = S \cdot \Phi (d_1) - Ke^{-r(T-t)} \cdot \Phi (d_2).
\]
Solution of the Black-Scholes Equation Graphically. Consider for purposes of graphical illustration the value of a call option with strike price $K = 100$. The risk-free interest rate per year, continuously compounded is 12%, so $r = 0.12$, the time to expiration is $T = 1$ measured in years, and the standard deviation per year on the return of the stock, or the volatility is $\sigma = 0.10$. The value of the call option at maturity plotted over a range of stock prices $70 \leq S \leq 130$ surrounding the strike price is illustrated in Figure 1.

We use the Black-Scholes formula above to compute the value of the option before expiration. With the same parameters as above the value of the call option is plotted over a range of stock prices $70 \leq S \leq 130$ at time remaining to expiration $t = 1, t = 0.8, t = 0.6, t = 0.4, t = 0.2$ and at expiration $t = 0$.

Using this graph, notice two trends in the option value:

1. For a fixed time, as the stock price increases the option value increases.
2. As the time to expiration decreases, for a fixed stock value the value of the option decreases to the value at expiration.
We predicted both trends from our intuitive analysis of options; see the table in the section on Options in the Background chapter. The Black-Scholes option pricing formula makes the intuition precise.

We can also plot the solution of the Black-Scholes equation as a function of security price and the time to expiration as a value surface.

This value surface shows both trends.

Section Ending Answer. The solution method for the Euler equidimensional equation (or Cauchy-Euler) type of differential equation

\[ x^2 \frac{d^2 v}{dx^2} + ax \frac{dv}{dx} + bv = 0 \]

is to make the change of independent variable \( t = \log x \) of equivalently \( x = e^t \). Then

\[ \frac{dv}{dx} = \frac{1}{x} \frac{dv}{dt} \]
\[ \frac{d^2 v}{dx^2} = \frac{1}{x^2} \frac{d^2 v}{dt^2} - \frac{1}{x^3} \frac{dv}{dt} \]

so the equation becomes

\[ \frac{d^2 v}{dt^2} + (a - 1) \frac{dv}{dt} + bv = 0, \]

a second-order constant coefficient equation, which is easy to solve. This change of variables is applied to the Black-Scholes equation, which resembles the Cauchy-Euler equation in the \( S \) variable.
7.2. SOLUTION OF THE BLACK-SCHOLES EQUATION

**Algorithms, Scripts, Simulations.**

*Algorithm.* For given parameter values, the Black-Scholes-Merton solution formula is sampled at a specified $m \times 1$ array of times and at a specified $1 \times n$ array of security prices using vectorization. The result can be plotted as functions of the security price as done in the text. The calculation is vectorized for an array of $S$ values and an array of $t$ values, but it is not vectorized for arrays in the parameters $K$, $r$, $T$, and $\sigma$. This approach is taken to illustrate the use of vectorization for efficient evaluation of an array of solution values from a complicated formula. In particular, the calculation of $d_1$ and $d_2$ then uses recycling.

The calculation relies on using the rules for handling of infinity and NaN (Not a Number) which come from divisions by 0, taking logarithms of 0, and negative numbers and calculating the normal cdf at infinity and negative infinity. The plotting routines will not plot a NaN which accounts for the gap at $S = 0$ in the graph line for $t = 1$.

```r
m <- 6;
n <- 61;
S0 <- 70;
S1 <- 130;
K <- 100;
r <- 0.12;
T <- 1.0;
sigma <- 0.10;
time <- seq(T,0, length=m);
S <- seq(S0,S1, length=n);
numerd1 <- outer( ( (r + sigma^2/2)*(T-time)), log(S/K), "+");
umerd2 <- outer( ( (r - sigma^2/2)*(T-time)), log(S/K), "+");
d1 <- numerd1/(sigma*sqrt(T-time));
d2 <- numerd2/(sigma*sqrt(T-time));
part1 <- t(t(pnorm(d1))*S);
part2 <- K*exp(-r*(T-time))*pnorm(d2);
VC <- part1 - part2;
matplot(S, t(VC), type = "l");
```

**Key Concepts.**

1. We solve the Black-Scholes equation for the value of a European call option on a security by judicious changes of variables that reduce the equation to the heat equation. The heat equation has a solution formula. Using the solution formula with the changes of variables gives the solution to the Black-Scholes equation.

2. Solving the Black-Scholes equation is an example of how to choose and execute changes of variables to solve a partial differential equation.
Vocabulary.

(1) A differential equation with auxiliary initial conditions and boundary conditions, that is an initial value problem, is said to be well-posed if the solution exists, is unique, and small changes in the equation parameters, the initial conditions or the boundary conditions produce only small changes in the solutions.

Problems.

Exercise 7.6. Explicitly evaluate the integral $I_2$ in terms of the c.d.f. $\Phi$ and other elementary functions as was done for the integral $I_1$.

Exercise 7.7. What is the price of a European call option on a non-dividend-paying stock when the stock price is $50, the strike price is $48, the risk-free interest rate is 5% per annum (compounded continuously), the volatility is 30% per annum, and the time to maturity is 3 months?

Exercise 7.8. What is the price of a European call option on a non-dividend paying stock when the stock price is $25, the exercise price is $28, the risk-free interest rate is 4%, the volatility is 25% per annum, and the time to maturity is 3 months?

Exercise 7.9. Show that the Black-Scholes formula for the price of a call option tends to $\max(S - K, 0)$ as $t \to T$.

Exercise 7.10. For a particular scripting language of your choice, modify the scripts to create a function within that language that will evaluate the Black-Scholes formula for a call option at a time and security value for given parameters.

Exercise 7.11. For a particular scripting language of your choice, modify the scripts to create a script within that language that will evaluate the Black-Scholes formula for a call option at a specified time for given parameters, and return a function of the security prices that can plotted by calling the function over an interval of security prices.

Exercise 7.12. Plot the price of a European call option on a non-dividend paying stock over the stock prices $20$ to $40$, given that the exercise price is $29$, the risk-free interest rate is 5%, the volatility is 25% per annum, and the time to maturity is 4 months.

Exercise 7.13. For a particular scripting language of your choice, modify the scripts to create a script within that language that will plot the Black-Scholes solution for $V_C(S, t)$ as a surface over the two variables $S$ and $t$.

7.3. Put-Call Parity

Section Starter Question. What does it mean to say that a differential equation is a linear differential equation?

Put-Call Parity by Linearity. The Black-Scholes equation is

$$V_t + \frac{1}{2} \sigma^2 S^2 V_{SS} + rSV_S - rV = 0.$$ 

With the additional terminal condition $V(S, T)$ given, a solution exists and is unique. We observe that the Black-Scholes is a linear equation, so the linear combination of any two solutions is again a solution.
From the problems in the previous section (or by easy verification right now) we know that $S$ is a solution of the Black-Scholes equation and $Ke^{-r(T-t)}$ is also a solution, so $S - Ke^{-r(T-t)}$ is a solution. At the expiration time $T$, the solution has value $S - K$.

Now if $C(S,t)$ is the value of a call option at security value $S$ and time $t < T$, then $C(S,t)$ satisfies the Black-Scholes equation, and has terminal value $\max(S - K, 0)$. If $P(S,t)$ is the value of a put option at security value $S$ and time $t < T$, then $P(S,t)$ also satisfies the Black-Scholes equation, and has terminal value $\max(K - S, 0)$. Therefore by linearity, $C(S,t) - P(S,t)$ is a solution and has terminal value $C(S,T) - P(S,T) = S - K$. By uniqueness, the solutions must be the same, and so

$$C - P = S - Ke^{-r(T-t)}.$$  

This relationship is known as the put-call parity principle between the price $C$ of a European call option and the price $P$ of a European put option, each with strike price $K$ and underlying security value $S$.

This same principle of linearity and the composition of more exotic options in terms of puts and calls allows us to write closed form formulas for the values of exotic options such as straddles, strangles, and butterfly options.

**Put-Call Parity by Reasoning about Arbitrage.** Assume that an underlying security satisfies the assumptions of the previous sections. Assume further that:

- the security price is currently $S = 100$;
- the strike price is $K = 100$;
- the expiration time is one year, $T = 1$;
- the risk-free interest rate is $r = 0.12$; and
- the volatility is $\sigma = 0.10$.

One can then calculate that the price of a call option with these assumptions is 11.84.

Consider an investor with the following portfolio:

- buy one share of stock at price $S = 100$;
- sell one call option at $C = V(100,0) = 11.84$;
- buy one put option at unknown price.

At expiration, the stock price could have many different values, and those would determine the values of each of the derivatives. See Table 1 for some representative values.

At expiration this portfolio always has a value which is the strike price. Holding this portfolio will give a risk-free investment that will pay $100 in any circumstance.

<table>
<thead>
<tr>
<th>Security</th>
<th>Call</th>
<th>Put</th>
<th>Portfolio</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>0</td>
<td>20</td>
<td>100</td>
</tr>
<tr>
<td>90</td>
<td>0</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>110</td>
<td>-10</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>120</td>
<td>-20</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

**Table 1.** Security, call and put option values at expiration.
This example portfolio has total value 100. Therefore the value of the whole portfolio must equal the present value of a riskless investment that will pay off $100 in one year. This is an illustration of the use of options for hedging an investment, in this case the extremely conservative purpose of hedging to preserve value.

The parameter values chosen above are not special and we can reason with general $S$, $C$ and $P$ with parameters $K$, $r$, $σ$, and $T$. Consider buying a put and selling a call, each with the same strike price $K$. We will find at expiration $T$ that:

- if the stock price $S$ is below $K$ we will realize a profit of $K - S$ from the put option that we own; and
- if the stock price is above $K$, we will realize a loss of $S - K$ from fulfilling the call option that we sold.

But this payout is exactly what we would get from a futures contract to sell the stock at price $K$. The price set by arbitrage of such a futures contract must be $Ke^{-r(T-t)} - S$. Specifically, one could sell (short) the stock right now for $S$, and lend $Ke^{-r(T-t)}$ dollars right now for a net cash outlay of $Ke^{-r(T-t)} - S$, then at time $T$ collect the loan at $K$ dollars and actually deliver the stock. This replicates the futures contract, so the future must have the same price as the initial outlay. Therefore we obtain the put-call parity principle:

$$-C + P = Ke^{-r(T-t)} - S$$

or more naturally

$$S - C + P = Ke^{-r(T-t)}.$$

**Synthetic Portfolios.** Another way to view this formula is that it instructs us how to create synthetic portfolio. A *synthetic portfolio* is a combination of securities, bonds, and options that has the same payout at expiration as another financial instrument. Since

$$S + P - Ke^{-r(T-t)} = C,$$

a portfolio “long in the underlying security, long in a put, short $Ke^{-r(T-t)}$ in bonds” replicates a call. This same principle of linearity and the composition of more exotic options in terms of puts and calls allows us to create synthetic portfolios for exotic options such as straddles, strangles, and so on. As noted above, we can easily write their values in closed form solutions.

**Explicit Formulas for the Put Option.** Knowing any two of $S$, $C$ or $P$ allows us to calculate the third. Of course, the immediate use of this formula will be to combine the security price and the value of the call option from the solution of the Black-Scholes equation to obtain the value of the put option:

$$P = C - S + Ke^{-r(T-t)}.$$
For the sake of mathematical completeness we can write the value of a European put option explicitly as:

\[
V_P(S, t) = S\Phi\left(\frac{\log(S/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}\right) - Ke^{-r(T-t)}\Phi\left(\frac{\log(S/K) + (r - \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}\right) 
- S + Ke^{-r(T-t)}.
\]

Usually one doesn’t see the solution as this full closed form solution. Instead, most versions of the solution write intermediate steps in small pieces, and then present the solution as an algorithm putting the pieces together to obtain the final answer. Specifically, let

\[
d_1 = \frac{\log(S/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}
\]

\[
d_2 = \frac{\log(S/K) + (r - \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}
\]

so that

\[
V_P(S, t) = S(\Phi(d_1) - 1) - Ke^{-r(T-t)}(\Phi(d_2) - 1).
\]

Using the symmetry properties of the c.d.f. \( \Phi \), we obtain

\[
V_P(S, t) = Ke^{-r(T-t)}\Phi(-d_2) - S\Phi(-d_1).
\]

**Graphical Views of the Put Option Value.** For graphical illustration let \( P \) be the value of a put option with strike price \( K = 100 \). The risk-free interest rate per year, continuously compounded is 12\%, so \( r = 0.12 \), the time to expiration is \( T = 1 \) measured in years, and the standard deviation per year on the return of the stock, or the volatility is \( \sigma = 0.10 \). The value of the put option at maturity plotted over a range of stock prices \( 0 \leq S \leq 150 \) surrounding the strike price is illustrated below:
Now we use the Black-Scholes formula to compute the value of the option before expiration. With the same parameters as above the value of the put option is plotted over a range of stock prices $0 \leq S \leq 150$ at time remaining to expiration $t = 1$, $t = 0.8$, $t = 0.6$, $t = 0.4$, $t = 0.2$ and at expiration $t = 0$.

Notice two trends in the value from this graph:

1. As the stock price increases, for a fixed time the option value decreases.
2. As the time to expiration decreases, for a fixed stock value price less than the strike price the value of the option increases to the value at expiration.

We can also plot the value of the put option as a function of security price and the time to expiration as a value surface.

This value surface shows both trends.

**Section Ending Answer.** A differential equation is linear if the linear combination of solutions is again a solution. The Black-Scholes equation is linear and this fact is used to derive the put-call parity relation.

**Algorithms, Scripts, Simulations.**

*Algorithm.* For given parameter values for $K$, $r$, $T$, and $\sigma$, the Black-Scholes-Merton solution formula for a put option is sampled at a specified $m \times 1$ array of times and at a specified $1 \times n$ array of security prices using vectorization. The result can be plotted as functions of the security price as done in the text. The calculation is vectorized for an array of $S$ values and an array of $t$ values, but it is not vectorized for arrays in the parameters $K$, $r$, $T$, and $\sigma$. This approach is taken to illustrate the use of vectorization and broadcasting for efficient evaluation of an array of solution values from a complicated formula. In particular, the calculation of $d_1$ and $d_2$ uses recycling.

The calculation relies on using the rules for calculation and handling of infinity and NaN (Not a Number) which come from divisions by 0, taking logarithms of 0, and negative numbers and calculating the normal cdf at infinity and negative infinity. The plotting routines will not plot a NaN which accounts for the gap at $S = 0$ in the graph line for $t = 1$. 

**Figure 5.** Value of the call option at various times.
Figure 6. Value surface from the Black-Scholes formula.

```r
m <- 6;
n <- 61;
S0 <- 70;
S1 <- 130;
K <- 100;
r <- 0.12;
T <- 1.0;
sigma <- 0.10;
time <- seq(T, 0, length=m);
S <- seq(S0, S1, length=n);
numerd1 <- outer((r + sigma^2/2)*(T-time)), log(S/K), "+");
umerd2 <- outer((r - sigma^2/2)*(T-time)), log(S/K), "+");
d1 <- numerd1/(sigma*sqrt(T-time));
d2 <- numerd2/(sigma*sqrt(T-time));
part1 <- K*exp(-r*(T-time))*pnorm(-d2);
part2 <- t(t(pnorm(-d1))*S);
VC <- part1 - part2;
matplot(S, t(VC), type = "l");
```
Key Concepts.
(1) The put-call parity principle links the price of a put option, a call option and the underlying security price.
(2) The put-call parity principle can be used to price European put options without having to solve the Black-Scholes equation.
(3) The put-call parity principle is a consequence of the linearity of the Black-Scholes equation.

Vocabulary.
(1) The put-call parity principle is the relationship
\[ C - P = S - Ke^{-r(T-t)} \]
between the price \( C \) of a European call option and the price \( P \) of a European put option, each with strike price \( K \) and underlying security value \( S \).
(2) A synthetic portfolio is a combination of securities, bonds, and options that has the same payout at expiration as another financial instrument.

Problems.
Exercise 7.14. Calculate the price of a 3-month European put option on a non-dividend-paying stock with a strike price of $48 when the current stock price is $50, the risk-free interest rate is 5% per year (compounded continuously) and the volatility is 30% per year.

Exercise 7.15. What is the price of a European put option on a non-dividend paying stock when the stock price is $70, the strike price is $68, the risk-free interest rate is 5% per year (compounded continuously), the volatility is 25% per year, and the time to maturity is 6 months?

Exercise 7.16. Show that the Black-Scholes formula for the price of a put option tends to \( \max(K-S, 0) \) as \( t \to T \).

Exercise 7.17. For a particular scripting language of your choice, modify the scripts to create a function within that language that will evaluate the Black-Scholes formula for a put option at a time and security value for given parameters.

Exercise 7.18. For a particular scripting language of your choice, modify the scripts to create a script within that language that will evaluate the Black-Scholes formula for a put option at a specified time for given parameters, and return a function of the security prices that can plotted by calling the function over an interval of security prices.

Exercise 7.19. Plot the price of a European put option on a non-dividend paying stock over the stock prices $20 to $40, given that the exercise price is $29, the risk-free interest rate is 5%, the volatility is 25% per annum, and the time to maturity is 4 months.

Exercise 7.20. For a particular scripting language of your choice, modify the scripts to create a script within that language that will plot the Black-Scholes solution for \( V_P(S,t) \) as a surface over the two variables \( S \) and \( t \).
7.4. IMPLIED VOLATILITY

Section Starter Question. What are some methods you could use to find the solution of \( f(x) = c \) for \( x \) where \( f \) is a function that is so complicated that you cannot use elementary functions and algebraic operations to isolate \( x \)?

Historical volatility. Estimates of historical volatility of security prices use statistical estimators, usually one of the estimators of variance. A main problem for historical volatility is to select the sample size, or window of observations, used to estimate \( \sigma^2 \). Different time-windows usually give different volatility estimates. Furthermore, for some customized “over the counter” derivatives the necessary price data may not exist.

Another problem with historical volatility is that it assumes future market performance is the same as past market data. Although this is a natural scientific assumption, it does not take into account historical anomalies such as the October 1987 stock market drop that may be unusual. That is, computing historical volatility has the usual statistical difficulty of how to handle outliers. The assumption that future market performance is the same as past performance also does not take into account underlying changes in the market such as new economic conditions.

To estimate the volatility of a security price empirically observe the security price at regular intervals, such as every day, every week, or every month. Define:

- the number of observations \( n + 1 \);
- the security price at the end of the \( i \)th interval \( S_i, i = 0, 1, 2, 3, \ldots, n \);
- the length of each of the time intervals (say in years) \( \tau \); and
- the increment of the logarithms of the security prices

\[
    u_i = \log(S_i) - \log(S_{i-1}) = \log\left(\frac{S_i}{S_{i-1}}\right)
\]

for \( i = 1, 2, 3, \ldots \).

We are modeling the security price as a geometric Brownian motion, so that \( \log(S_i) - \log(S_{i-1}) \sim N(\tau r, \sigma^2 \tau) \).

Inverting \( u_i = \log(S_i/S_{i-1}) \) to obtain \( S_i = S_{i-1} e^{u_i} \), we see that \( u_i \) is the continuously compounded return (not annualized) in the \( i \)th interval. Then the usual estimate \( s \) of the standard deviation of the \( u_i \)s is

\[
    s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (u_i - \bar{u})^2}
\]

where \( \bar{u} \) is the mean of the \( u_i \)’s. Sometimes it is more convenient to use the equivalent formula

\[
    s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} u_i^2 - \frac{1}{n} \left( \sum_{i=1}^{n} u_i \right)^2}.
\]

We assume the security price varies as a geometric Brownian motion. That means that the logarithm of the security price is a Wiener process with some drift and on the period of time \( \tau \), would have a variance \( \sigma^2 \tau \). Therefore, \( s \) is an estimate of \( \sigma \sqrt{\tau} \). It follows that \( \sigma \) can be estimated as

\[
    \sigma \approx \frac{s}{\sqrt{\tau}}.
\]
Choosing an appropriate value for $n$ is not obvious. Remember the variance expression for geometric Brownian motion is an increasing function of time. If we model security prices with geometric Brownian motion, then $\sigma$ does change over time, and data that are too old may not be relevant for the present or the future. A compromise that seems to work reasonably well is to use closing prices from daily data over the most recent 90 to 180 days. Empirical research indicates that only trading days should be used, so days when the exchange is closed should be ignored for the purposes of the volatility calculation [21, page 215].

Economists and financial analysts often estimate historical volatility with more sophisticated statistical time series methods.

**Implied Volatility.** The implied volatility is the parameter $\sigma$ in the Black-Scholes formula that would give the option price that is observed in the market, all other parameters being known.

The Black-Scholes formula is too complicated to invert to explicitly express $\sigma$ as a function of the other parameters. Therefore, we use numerical techniques to implicitly solve for $\sigma$. A simple idea is to use the method of bisection search to find $\sigma$.

**Example 7.21.** Suppose the value $C$ of a call on a non-dividend-paying security is 1.85 when $S = 21$, $K = 20$, $r = 0.10$, and $T-t = 0.25$ and $\sigma$ is unknown. We start by arbitrarily guessing $\sigma = 0.20$. The Black-Scholes formula gives $C = 1.7647$, which is too low. Since $C$ is an increasing function of $\sigma$, this suggests we try a value of $\sigma = 0.30$. This gives $C = 2.1010$, too high, so we bisect the interval $[0.20, 0.30]$ and try $\sigma = 0.25$. This value of $\sigma$ gives a value of $C = 1.9268$, still too high. Bisect the interval $[0.20, 0.25]$ and try a value of $\sigma = 0.225$, which yields $C = 1.8438$, slightly too low. Try $\sigma = 0.2375$, giving $C = 1.8849$. Finally try $\sigma = 0.23125$ giving $C = 1.8642$. To 2 significant digits, the significance of the data, $\sigma = 0.23$, with a predicted value of $C = 1.86$. 

---

**Figure 7.** Schematic diagram of using Newton’s Method to solve for implied volatility. The current call option value is the horizontal line. The value of the call option as a function of $\sigma$ is the thin curve. The tangent line is the thicker line.
Another procedure is to use Newton’s method which is also iterative. Essentially we are trying to solve

\[ f(\sigma, S, K, r, T - t) - C = 0, \]

so from an initial guess \( \sigma_0 \), we form the Newton iterates

\[ \sigma_{i+1} = \sigma_i - \frac{f(\sigma_i)}{df(\sigma_i)/d\sigma). \]

See Figure 7. Using Newton’s method means one has to differentiate the Black-Scholes formula with respect to \( \sigma \). This derivative is one of the “greeks” known as \( \text{vega} \) which we will look at more extensively in the next section. A formula for vega for a European call option is

\[ \frac{df}{d\sigma} = S \sqrt{T - t} \Phi'(d_1) \exp(-r(T - t)). \]

A natural way to do the iteration is with a computer program rather than by hand.

Implied volatility is a forward-looking estimation technique, in contrast to the backward-looking historical volatility. That is, it incorporates the market’s expectations about the prices of securities and their derivatives, or more concisely, market expectations about risk. More sophisticated combinations and weighted averages combining estimates from several different derivative claims can be developed.

**Section Ending Answer.** To find the solution of a complicated equation \( f(x) = c \) use numerical methods such as repeated bisection, Newton’s method, or more specialized numerical techniques. This section illustrates solving the Black-Scholes formula for the volatility with these techniques.

**Algorithms, Scripts, Simulations.**

*Algorithm.* First define the function \( f = V_C - C \) as a function of \( \sigma \) and parameters \( S, K, r, T - t, \) and \( C \). Next define the derivative of \( f \) with respect to \( \sigma \). For given numerical values for \( \sigma_0 \), the guess for the volatility; \( S \), the current security price; \( K \), the strike price; \( r \), the risk-free interest rate; \( T - t \), the time to expiration; and \( C \), the current call option price, the script uses Newton’s method to find the implied volatility with error tolerance \( \epsilon \). The implied volatility is the value of \( \sigma \) which makes \( f \approx 0 \) to within \( \epsilon \). The Newton’s method iteration uses a repeat–until loop construction which means that at least one iteration of the loop is computed.

```r
1 f <- function(sigma, S, K, r, Tminust, C) {
2  d1 <- (log(S/K) + ( (r + sigma^2/2)*(Tminust)))/(sigma*sqrt(Tminust));
3  d2 <- (log(S/K) + ( (r - sigma^2/2)*(Tminust)))/(sigma*sqrt(Tminust));
4  part1 <- pnorm(d1) * S;
5  part2 <- K*exp(-r*(Tminust)) * pnorm(d2);
6  VC <- part1 - part2;
7  f <- VC - C;
8  f
9 }
10
11 fprime <- function(sigma, S, K, r, Tminust, C) {
12  d1 <- (log(S/K) + ( (r + sigma^2/2)*(Tminust)))/(sigma*sqrt(Tminust));
13  d2 <- (log(S/K) + ( (r - sigma^2/2)*(Tminust)))/(sigma*sqrt(Tminust));
14  return(-part1/(2*sigma) + part2/(2*sigma) + d2)*dnorm(d1)
15  }
```

Section Ending Answer. To find the solution of a complicated equation \( f(x) = c \) use numerical methods such as repeated bisection, Newton’s method, or more specialized numerical techniques. This section illustrates solving the Black-Scholes formula for the volatility with these techniques.
THE BLACK-SCHOLES EQUATION

Key Concepts.

(1) We estimate historical volatility by applying the standard deviation estimator from statistics to the observations \( \log(S_t/S_{t-1}) \).

(2) We deduce implied volatility by numerically solving the Black-Scholes formula for \( \sigma \).

Vocabulary.

(1) Historical volatility of a security is the variance of the changes in the logarithm of the price of the underlying asset, obtained from past data.

(2) Implied volatility of a security is the numerical value of the volatility parameter that makes the market price of an option equal to the value from the Black-Scholes formula.

Problems.

Exercise 7.22. Suppose that the observations on a security price (in dollars) at the end of each of 15 consecutive weeks are as follows: 30.25, 32, 31.125, 30.25, 30.375, 30.625, 33, 32.875, 33, 33.5, 33.5, 33.75, 33.5, 33.25. Estimate the security price volatility.

Exercise 7.23. Pick a publicly traded security, obtain a record of the last 90 days of that security’s prices, and compute the historical volatility of the security.

Exercise 7.24. A call option on a non-dividend paying security has a market price of $2.50. The security price is $15, the exercise price is $13, the time to
maturity is 3 months, and the risk-free interest rate is 5% per year. Using repeated bisection, what is the implied volatility?

Exercise 7.25. For a particular scripting language of your choice, create a script within that language that will compute implied volatility by repeated bisection.

Exercise 7.26. A call option on a non-dividend paying security has a market price of $2.50. The security price is $15, the exercise price is $13, the time to maturity is 3 months, and the risk-free interest rate is 5% per year. Using Newton’s method, what is the implied volatility?

Exercise 7.27. Sources of financial data on options typically provide information on the current security price on a quote date, a variety of strike prices, the current quote date and the expiration date, the option price and the implied volatility. (Some even provide additional data on the values of the greeks.) However, the risk-free interest rate is not provided. For a particular scripting language of your choice, create a script within that language that will compute the risk-free interest rate by repeated bisection. Compare that computed rate to historical interest rate on U.S. Treasury bonds. For example, on October 1, 2010, a call option on Google stock valued at $525.62 with a strike price of $620 on the expiration date of January 22, 2011 had a price of $4.75 and an implied volatility of 0.2529. What was the risk-free interest rate on that date? What was the interest rate on U.S. Treasury bonds on that date?

7.5. Sensitivity, Hedging and the Greeks

Section Starter Question. Recall when we first considered options in the Options section of the first chapter.

(1) What did we intuitively predict would happen to the value of a call option if the underlying security value increased?

(2) What did we intuitively predict would happen to the value of a call option as the time increased to the expiration date?

(3) What did we intuitively predict would happen to the value of a call option if the volatility of the underlying security value increased?

Sensitivity of the Black-Scholes Formula. To start the examination of each of the sensitivities, restate the Black-Scholes formula for the value of a European call option:

\[
\begin{align*}
    d_1 &= \frac{\log(S/K) + (r + \sigma^2/2)(T - t)}{\sigma \sqrt{T - t}} \\
    d_2 &= \frac{\log(S/K) + (r - \sigma^2/2)(T - t)}{\sigma \sqrt{T - t}}
\end{align*}
\]

and then

\[
V_C(S, t) = S \Phi (d_1) - K e^{-r(T-t)} \Phi (d_2).
\]

Note that \( d_2 = d_1 - \sigma \sqrt{T - t} \).
Delta. The Delta of a European call option is the rate of change of its value with respect to the underlying security price:

\[
\Delta = \frac{\partial V_C}{\partial S} = \Phi(d_1) + S \Phi'(d_1) \frac{\partial d_1}{\partial S} - Ke^{-r(T-t)}\Phi'(d_2) \frac{\partial d_2}{\partial S}
\]

\[
= \Phi(d_1) + S \frac{1}{\sqrt{2\pi}} \frac{e^{-d_1^2/2}}{S\sigma\sqrt{T-t}} - Ke^{-r(T-t)} \frac{1}{\sqrt{2\pi}} \frac{e^{-d_2^2/2}}{S\sigma\sqrt{T-t}}
\]

\[
= \Phi(d_1) + \frac{1}{\sqrt{2\pi}} \frac{e^{-d_1^2/2}}{S\sigma\sqrt{T-t}} \left[ 1 - Ke^{-r(T-t)} \frac{e^{d_1\sigma\sqrt{T-t} - \sigma^2(T-t)/2}}{S} \right]
\]

\[
= \Phi(d_1) + \frac{e^{-d_1^2/2}}{\sqrt{2\pi}S\sigma\sqrt{T-t}} \left[ 1 - Ke^{-r(T-t)} S e^{d_1\sigma\sqrt{T-t} - \sigma^2(T-t)/2} \right]
\]

\[
= \Phi(d_1) + \frac{e^{-d_1^2/2}}{\sqrt{2\pi}S\sigma\sqrt{T-t}} \left[ 1 - Ke^{-r(T-t)} e^{\log(S/K) + (r+\sigma^2/2)(T-t) - \sigma^2(T-t)/2} \right]
\]

\[
= \Phi(d_1) + \frac{e^{-d_1^2/2}}{\sqrt{2\pi}S\sigma\sqrt{T-t}} \left[ 1 - Ke^{-r(T-t)} e^{\log(S/K) + r(T-t)} \right]
\]

\[
= \Phi(d_1)
\]

Figure 9 is a graph of Delta as a function of \( S \) for several values of \( t \). Note that since \( 0 < \Phi(d_1) < 1 \) (for all reasonable values of \( d_1 \)), \( \Delta > 0 \), and so the value of a European call option is always increasing as the underlying security value increases. This is precisely as we intuitively predicted when we first considered
Delta Hedging. Notice that for any sufficiently differentiable function $F(S)$

$$F(S_1) - F(S_2) \approx \frac{dF}{dS} \cdot (S_1 - S_2).$$

Therefore, for the Black-Scholes formula for a European call option, using our current notation $\Delta = \frac{\partial V}{\partial S}$,

$$V(S_1) - V(S_2) \approx \Delta \cdot (S_1 - S_2).$$

Equivalently for small changes in security price from $S_1$ to $S_2$,

$$V(S_1) - \Delta \cdot S_1 \approx V(S_2) - \Delta \cdot S_2.$$

In financial language, we express this relationship as:

Being long 1 derivative and short $\Delta$ units of the underlying asset is approximately market neutral for small changes in the asset value.

We say that the sensitivity of the financial derivative value with respect to the asset value, denoted $\Delta$, gives the hedge ratio. The hedge ratio is the number of short units of the underlying asset which combined with a call option will offset immediate market risk. After a change in the asset value, $\Delta(S)$ will also change, and so we will need to dynamically adjust the hedge ratio to keep pace with the changing asset value. Thus $\Delta(S)$ as a function of $S$ provides a dynamic strategy for hedging against risk.

We have seen this strategy before. In the derivation of Black-Scholes equation, we required that the amount of security in our portfolio, namely $\phi(t)$, be chosen so that $\phi(t) = V_S$. (See the section Derivation of the Black-Scholes Equation.) The choice $\phi(t) = V_S$ gave us a risk-free portfolio that must change in the same way as a risk-free asset.

**Gamma: The convexity factor.** The Gamma ($\Gamma$) of a derivative is the sensitivity of $\Delta$ with respect to $S$:

$$\Gamma = \frac{\partial^2 V}{\partial S^2}.$$
The concept of Gamma is important when the hedged portfolio cannot be adjusted continuously in time according to $\Delta(S(t))$. If Gamma is small then Delta changes very little with $S$. This means the portfolio requires only infrequent adjustments in the hedge ratio. However, if Gamma is large, then the hedge ratio Delta is sensitive to changes in the price of the underlying security.

According to the Black-Scholes formula,

$$
\Gamma = \frac{1}{S\sqrt{2\pi}\sigma\sqrt{T-t}}e^{-d_2^2/2}
$$

Notice that $\Gamma > 0$, so the call option value is always concave-up with respect to $S$. See this in Figure 8.

**Theta:** The time decay factor. The Theta $(\Theta)$ of a European claim with value function $V(S,t)$ is

$$
\Theta = \frac{\partial V}{\partial t}
$$

Note that this definition is the rate of change with respect to the real (or calendar) time, some other authors define the rate of change with respect to the time-to-expiration $T-t$, so be careful when reading. The Black-Scholes partial differential equation can now be written as

$$
\Theta + rS\Delta + \frac{1}{2}\sigma^2S^2\Gamma = rV.
$$

The analytic expression of Theta now follows easily, the calculus to do it directly is tedious. More generally, given the parameters $r$, and $\sigma^2$, and any 4 of $\Theta$, $\Delta$, $\Gamma$, $S$ and $V$ the remaining quantity is implicitly determined.

The Theta of a claim is sometimes referred to as the time decay of the claim. For a European call option on a non-dividend-paying stock,

$$
\Theta = -\frac{S\cdot\sigma}{2\sqrt{T-t}} \cdot \frac{e^{-d_2^2/2}}{\sqrt{2\pi}} - rKe^{-r(T-t)}\Phi(d_2).
$$

Note that $\Theta$ for a European call option is negative, so the value of a European call option is decreasing See this in Figure 8. This is new information, previously we could not intuitively predict the change with respect to time to expiration.

Theta does not act like a hedging parameter as do Delta and Gamma. Although there is uncertainty about the future stock price, there is no uncertainty about the passage of time. It does not make sense to hedge against the passage of time on an option.

**Rho:** The interest rate factor. The rho $(\rho)$ of a derivative security is the rate of change of the value of the derivative security with respect to the interest rate. It measures the sensitivity of the value of the derivative security to interest rates. For a European call option on a non-dividend paying stock,

$$
\rho = K(T-t)e^{-r(T-t)}\Phi(d_2)
$$

so $\rho$ is always positive. An increase in the risk-free interest rate means a corresponding increase in the derivative value.
Vega: The volatility factor. The Vega ($\Lambda$) of a derivative security is the rate of change of value of the derivative with respect to the volatility of the underlying asset. (Some authors denote Vega by variously $\lambda$, $\kappa$, and $\sigma$; referring to Vega by the corresponding Greek letter name.) For a European call option on a non-dividend-paying stock,

$$\Lambda = S \sqrt{T - t} \frac{e^{-d_1^2/2}}{\sqrt{2\pi}}$$

so the Vega is always positive. An increase in the volatility will lead to a corresponding increase in the call option value. These formulas implicitly assume that the price of an option with variable volatility (which we have not derived, we explicitly assumed volatility was a constant!) is the same as the price of an option with constant volatility. To a reasonable approximation this seems to be the case, for more details and references, see [21, page 316].

Hedging in Practice. It would be wrong to give the impression that traders continuously balance their portfolios to maintain Delta neutrality, Gamma neutrality, Vega neutrality, and so on as would be suggested by the continuous mathematical formulas presented above. In practice, transaction costs make frequent balancing expensive. Rather than trying to eliminate all risks, an option trader usually concentrates on assessing risks and deciding whether they are acceptable. Traders tend to use Delta, Gamma, and Vega measures to quantify the different aspects of risk in their portfolios.

Section Ending Answer. In the Options section of the first chapter, we made several intuitive predictions.

1. The prediction was that the value of a call option would increase if the underlying security value increased.
2. There was no prediction about the value of a call option as the time increased to the expiration date.
3. The prediction was that the value of a call option would increase if volatility of the underlying security value increased.

The greeks derived in this section measure the sensitivity of the value of a call option to the parameters and all verify the intuitive conclusions and provide analytic expressions.
The plotting routines will not plot a NaN which accounts for the gaps or omissions.

The scripts plot Delta and Gamma as functions of $S$ for the specified $m \times 1$ array of times in two side-by-side subplots in a single plotting frame.

```
1 m <- 6
2 n <- 61
3 S0 <- 70
4 S1 <- 130
5 K <- 100
6 r <- 0.12
7 T <- 1
8 sigma <- 0.1
9
time <- seq(T, 0, length = m)
10 S <- seq(S0, S1, length = n)
11
12 numerd1 <- outer(((r + sigma^2/2) * (T - time)), log(S/K), "+")
13 d1 <- numerd1/(sigma * sqrt(T - time))
14 Delta <- pnorm(d1)
15
16 factor1 <- 1/(sqrt(2 * pi) * sigma * outer(sqrt(T - time), S, "*"))
17 factor2 <- exp(-d1^2/2)
18 Gamma <- factor1 * factor2
19
do.old.par <- par(mfrow = c(1, 2))
21 matplot(S, t(Delta), type = "l")
22 matplot(S, t(Gamma), type = "l")
23 par(do.old.par)
```

**Key Concepts.**

1. The sensitivity of the Black-Scholes formula to each of the variables and parameters is named, is fairly easily expressed, and has important consequences for hedging investments.

2. The sensitivity of the Black-Scholes formula (or any mathematical model) to its parameters is important for understanding the model and its utility.

**Vocabulary.**

1. The **Delta** ($\Delta$) of a financial derivative is the rate of change of the value with respect to the value of the underlying security, in symbols
   $$\Delta = \frac{\partial V}{\partial S}.$$  

2. The **Gamma** ($\Gamma$) of a derivative is the sensitivity of $\Delta$ with respect to $S$, in symbols
   $$\Gamma = \frac{\partial^2 V}{\partial S^2}. $$
(3) The **Theta** ($\Theta$) of a European claim with value function $V(S,t)$ is

$$\Theta = \frac{\partial V}{\partial t}.$$ 

(4) The **rho** ($\rho$) of a derivative security is the rate of change of the value of the derivative security with respect to the interest rate, in symbols

$$\rho = \frac{\partial V}{\partial r}.$$ 

(5) The **Vega** ($\Lambda$) of derivative security is the rate of change of value of the derivative with respect to the volatility of the underlying asset, in symbols

$$\Lambda = \frac{\partial V}{\partial \sigma}.$$ 

(6) **Hedging** is the attempt to make a portfolio value immune to small changes in the underlying asset value (or its parameters).

**Problems.**

**Exercise 7.28.** How can a short position in 500 call options be made Delta neutral when the Delta of each option is 0.75?

**Exercise 7.29.** Calculate the Delta of an at-the-money 3-month European call option on a non-dividend paying stock, when the risk-free interest rate is 6% per year (compounded continuously) and the stock price volatility is 20% per year.

**Exercise 7.30.** Use the put-call parity relationship to derive the relationship between

1. the Delta of a European call option and the Delta of a European put option,
2. the Gamma of a European call option and the Gamma of a European put option,
3. the Vega of a European call option and a European put option, and
4. the Theta of a European call option and a European put option.

**Exercise 7.31.** (1) Derive the expression for $\Gamma$ for a European call option.

(2) For a particular scripting language of your choice, modify the script to draw a graph of $\Gamma$ versus $S$ for $K = 50$, $r = 0.10$, $\sigma = 0.25$, $T - t = 0.25$.

(3) For a particular scripting language of your choice, modify the script to draw a graph of $\Gamma$ versus $t$ for a call option on an at-the-money stock, with $K = 50$, $r = 0.10$, $\sigma = 0.25$, $T - t = 0.25$.

(4) For a particular scripting language of your choice, modify the script to draw the graph of $\Gamma$ versus $S$ and $t$ for a European call option with $K = 50$, $r = 0.10$, $\sigma = 0.25$, $T - t = 0.25$.

(5) Comparing the graph of $\Gamma$ versus $S$ and $t$ with the graph of $V_C$ versus $S$ and $t$ in of Solution the Black Scholes Equation, explain the shape and values of the $\Gamma$ graph. This only requires an understanding of calculus, not financial concepts.

**Exercise 7.32.** (1) Derive the expression for $\Theta$ for a European call option, as given in the notes.
(2) For a particular scripting language of your choice, modify the script to draw a graph of $\Theta$ versus $S$ for $K = 50$, $r = 0.10$, $\sigma = 0.25$, $T - t = 0.25$.

(3) For a particular scripting language of your choice, modify the script to draw a graph of $\Theta$ versus $t$ for an at-the-money stock, with $K = 50$, $r = 0.10$, $\sigma = 0.25$, $T = 0.25$.

**Exercise 7.33.** (1) Derive the expression for $\rho$ for a European call option as given in this section.

(2) For a particular scripting language of your choice, modify the script to draw a graph of $\rho$ versus $S$ for $K = 50$, $r = 0.10$, $\sigma = 0.25$, $T - t = 0.25$.

**Exercise 7.34.** (1) Derive the expression for $\Lambda$ for a European call option as given in this section.

(2) For a particular scripting language of your choice, modify the script to draw a graph of $\Lambda$ versus $S$ for $K = 50$, $r = 0.10$, $\sigma = 0.25$, $T - t = 0.25$.

**Exercise 7.35.** For a particular scripting language of your choice, modify the script to create a function within that language that will evaluate the call option greeks Delta and Gamma at a time and security value for given parameters.

### 7.6. Limitations of the Black-Scholes Model

**Section Starter Question.** We have derived and solved the Black-Scholes equation. We have derived parameter dependence and sensitivity of the solution. Are we done? What’s next? How should we go about implementing and analyzing that next step, if any?

**Validity of Black-Scholes.** Recall that the Black-Scholes Model is based on several assumptions:

1. The price of the underlying security for which we are considering a derivative financial instrument follows the stochastic differential equation
   \[ dS = rS \, dt + \sigma S \, dW \]
   or equivalently that $S(t)$ is a geometric Brownian motion
   \[ S(t) = z_0 \exp((r - \sigma^2/2)t + \sigma W(t)). \]
   At each time the geometric Brownian motion has a lognormal distribution with parameters $(\log(z_0) + rt - \sigma^2/2t)$ and $\sigma \sqrt{t}$. The mean value of the geometric Brownian motion is $E[S(t)] = z_0 \exp(rt)$.

2. The risk free interest rate $r$ and volatility $\sigma$ are constants.

3. The value $V$ of the derivative depends only on the current value of the underlying security $S$ and the time $t$, so we can write $V(S, t)$.

4. All variables are real-valued, and all functions are sufficiently smooth to justify necessary calculus operations.

See the section on Derivation of the Black-Scholes Equation for the context of these assumptions.

One judgment on the validity of these assumptions statistically compares the predictions of the Black-Scholes model with the market prices of call options. This is the observation or validation phase of the cycle of mathematical modeling, see Brief Remarks on Math Models for the cycle and diagram. A detailed examination (the financial and statistical details of this examination are outside the scope of these notes) shows that the Black-Scholes formula misprices options. In fact, the
Black-Scholes model overprices at the money options, that is with $S \approx K$, and underprices options at the ends, either deep in the money, $S \gg K$, or deep out of the money, $S \ll K$. When the actual option price is higher than the price from the Black-Scholes formula, this indicates that the market assigns greater risk to events that are far from the center than the model would predict.

Another test of some of these assumptions is to gather data on the actual market price of call options on a security, all with the same expiration date, with a variety of strike prices. Then one can compute the implied volatility from the data. The implied volatility is not constant, in spite of our assumption of constant volatility! The computed implied volatility is higher for either deep in the money, $S_0 \gg K$, or deep out of the money, $S_0 \ll K$, than at the money, $S_0 \approx K$, where $S_0$ is the current asset price and $K$ is the strike price. The resulting concave up shape is called the volatility smile so called because it looks like a cartoon happy-face smile. The greater implied volatility far from the center indicates that the market assigns greater risk to events that are far from the center. Figure 10 shows actual data on implied volatility. The data are from a call option on the S & P 500 index price on July 1, 2013, with an expiration date of July 5, 2013, just four calendar days later. The S & P 500 price $S_0$ on July 1, 2013 was 1614.96, shown with the vertical dashed line. The implied volatility in this example is large at the ends since the S & P 500 index would have to be very volatile to change by 200 to 400 points in just three trading days. (July 4 is a holiday.) This example clearly illustrates that the market does not value options consistently with the constant volatility assumptions of the Black-Scholes model.

We have already seen evidence in the section Stock Market Model that returns from the Wilshire 5000 are more dispersed than standard normal data. The low quantiles of the normalized Wilshire 5000 quantiles occur at more negative values than the standard normal distribution. The high quantiles occur at values greater than the standard normal distribution. That is, probabilities of extreme changes in daily returns are greater than expected from the hypothesized normal probability model.

Some studies have shown that the event of downward jumps three standard deviations below the mean is three times more likely than a normal distribution
would predict. This means that if we used geometric Brownian motion to compute the historical volatility of the S&P 500, we would find that the normal theory seriously underestimates the likelihood of large downward jumps. Jackwerth and Rubinstein [24] note that with the geometric Brownian motion model, the crash of 1987 is an impossibly unlikely event:

Take for example the stock market crash of 1987. Following the standard paradigm, assume that the stock market returns are log-normally distributed with an annualized volatility of 20%. On October 19, 1987, the two-month S&P 500 futures price fell 29%. Under the log-normal hypothesis, this has a probability of $10^{-160}$, which is virtually impossible.

The popular term for such extreme changes is a black swan, reflecting the rarity of spotting a black swan among white swans. In financial markets black swans occur much more often than the standard probability models predict [41, 65].

All these empirical tests indicate that the price process has fat tails, i.e., a distribution with the probability of large changes in price $S$ larger than the lognormal distribution predicts, [4, page 9]. The assumption that the underlying security has a price modeled by geometric Brownian motion, or equivalently that at any time the security price has a lognormal distribution, is incorrect. Large changes are more frequent than the model expects.

An example of a fat-tailed distribution is the Cauchy distribution with density function $f_C(x) = \frac{1}{\pi (1 + x^2)}$. The rate of decay to 0 of the density as $x \to \infty$ is the inverse power law $x^{-2}$. This is much slower than the rate of decay of the normal distribution density function $f_N(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. For large values, the probability of events governed by the Cauchy distribution is greater than the probability of events governed by the normal distribution.

An obstacle to using the Cauchy distribution is that its variance is undefined (or more loosely stated, the variance is infinite.) The assumption of finite variance is useful for theoretical analysis of probabilities. Consider how many times this text uses the assumption that the variance of a distribution was finite.
Another example of a fat-tailed distribution is the Student’s t-distribution with density function

\[ f_S(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu \pi} \Gamma(\frac{\nu}{2})} \left( 1 + \frac{t^2}{\nu} \right)^{-\frac{\nu+1}{2}}. \]

This distribution is intermediate in the sense that for \( \nu = 1 \) it is the Cauchy distribution, but as \( \nu \to \infty \), the t-distribution approaches the normal distribution. For \( \nu > 1 \) the variance of the distribution is finite, but now with another parameter \( \nu \) to fit. Lions [38] finds that a Student’s t-distribution with \( \nu = 3.4 \) provides a good fit to the distribution of the daily returns of the S & P 500.

More fundamentally, one can look at whether general market prices and security price changes fit the hypotheses of following the stochastic differential equation for geometric Brownian motion. Studies of security market returns reveal an important fact: As in Lions [38], large changes in security prices are more likely than normally distributed random effects would predict. Put another way, the stochastic differential equation model predicts that large price changes are much less likely than is actually the case.

The origin of the difference between the stochastic differential equation model for geometric Brownian motion and real financial markets may be a fundamental misapplication of probability modeling. The mathematician Benoit Mandelbrot argues that finance is prone to a wild randomness not usually seen in traditional applications of statistics [66]. Mandelbrot says that rare big changes can be more significant than the sum of many small changes. That is, Mandelbrot calls into question the applicability of the Central Limit Theorem in finance. It may be that the mathematical hypotheses required for the application of the Central Limit Theorem do not hold in finance.

Recall that we explicitly assumed that many of the parameters were constant; in particular, volatility is assumed constant. Actually, we might wish to relax that idea somewhat, and allow volatility to change in time. Nonconstant volatility introduces another dimension of uncertainty and also of variability into the problem. Still, changing volatility is an area of active research, both practically and academically.

We also assumed that trading was continuous in time, and that security prices moved continuously. Of course, continuous change is an idealizing assumption. In fact, in October 1987, the markets dropped suddenly, almost discontinuously, and market strategies based on continuous trading were not able to keep with the selling panic. The October 1987 drop is yet another illustration that the markets do not behave exactly as the assumptions made. Another relaxation of the Black-Scholes model is to assume that the price process can make sudden discontinuous jumps, leading to what are called jump-diffusion models. The mathematics associated with these processes is necessarily more complicated. This is another area of active research [12].

**Black-Scholes as an Approximate Model with an Exact Solution.** In the classification of mathematical modeling in the section Brief Remarks on Mathematical Models, the Black-Scholes model is an approximate model with an exact solution. The previous paragraphs show that each of the assumptions made to derive the Black-Scholes equation is an approximation made expressly to derive a tractable model. The assumption that security prices act as a geometric Brownian motion means that we can use the techniques of Itô stochastic calculus. Even more
so, the resulting lognormal distribution is based on the normal distribution which has simple mathematical properties. The assumptions that the risk free interest rate $r$ and volatility $\sigma$ are constants simplify the derivation and reduce the size of the resulting model. The assumption that the value $V$ of the derivative depends only on the current value of the underlying security $S$ and the time $t$ likewise reduces the size of the resulting model. The assumption that all variables are real-valued and all functions are sufficiently smooth allows necessary calculus operations.

Once the equation is derived we are able to solve it with standard mathematical techniques. In fact, the resulting Black-Scholes formula fits on a single line as a combination of well-known functions. The formula is simple enough that by 1975 Texas Instruments created a hand-held calculator specially programmed to produce Black-Scholes option prices and hedge ratios. The simple exact solution puts the Black-Scholes equation in the same position as the pendulum equation and the Ideal Gas Law, each an approximate model with an exact solution. The trust we have in solutions to mathematical equations creates a sanitizing effect when we use mathematical models. The mere fact that sophisticated mathematical methods created a solution makes a conclusion, faulty or not, seem inevitable and unimpeachable, [45, page 4]. Although an exact solution is satisfying, even beautiful, exact solutions hide the approximate origins. We often interpret the results of mathematical analysis as “objective” when they are only as objective as the underlying assumptions. The deep mathematical analysis relies in complex ways on the assumptions, so the result is an apparently strong, objective result that is actually neither strong nor objective. In the worst cases, such analysis is an example of the “garbage in, garbage out” problem, [45, page 6]. It is tempting to assume you have the perfect answer when at best you have a noisy approximation, [45, page 2].

The elegant one-line solution may actually encourage some of the misuses of mathematical modeling detailed below. By programming the elegant solution into a simple calculator everyone can use it. Had the Black-Scholes equation been nonlinear or based on more sophisticated stochastic processes, it probably would not have a closed-form solution. Solutions would have relied on numerical calculation, more advanced mathematical techniques or simulation; that is, solutions would not have been easily obtained. Those end-users who persevered in applying the theory might have sought additional approximations more conscious of the compromises with reality.

**Misuses of Mathematical Modeling.** By 2005, about 5% of jobs in the finance industry were in mathematical finance. The heavy use of flawed mathematical models contributed to the failure and near-failure of some Wall Street firms in 2009. Some critics blamed the mathematics and the models for the general economic troubles that resulted. In spite of the flaws, mathematical modeling in finance is not going away. Consequently, modelers and users have to be honest and aware of the limitations in mathematical modeling, [66]. Mathematical models in finance do not have the same experimental basis and long experience as do mathematical models in physical sciences. For the time being, we should cautiously use mathematical models in finance as good indicators that point to the values of financial instruments, but do not predict with high precision.

Actually, the problem goes deeper than just realizing that the precise distribution of security price movements is slightly different from the assumed lognormal
distribution. Even after specifying the probability distribution, giving a mathematical description of the risk, we still would have the uncertainty of not knowing the precise parameters of the distribution to specify it totally. From a scientific point of view, the way to estimate the parameters is to statistically evaluate the outcomes from the past to determine the parameters. We looked at one case of this when we described historical volatility as a way to determine $\sigma$ for the lognormal distribution, see Implied Volatility. However, this implicitly assumes that the past is a reasonable predictor of the future. While this faith is justified in the physical world where physical parameters do not change, such a faith in constancy is suspect in the human world of the markets. Consumers, producers, and investors all change habits suddenly in response to fads, bubbles, rumors, news, and real changes in the economic environment. Their change in economic behavior changes the parameters.

Even within finance, the models may vary in applicability. Analysis of the 2008-2009 market collapse indicates that the markets for interest rates and foreign exchange may have followed the models, but the markets for debt obligations may have failed to take account of low-probability extreme events such as the fall in house prices [66]. The models for debt obligations may have also assumed independence of events that were actually connected and correlated.

Models can have other problems that are more social than mathematical. Sometimes the use of the models changes the market priced by the model. This feedback process in economics has been noted with the Black-Scholes model, [66]. Sometimes special financial instruments can be so complex that modeling them requires too many assumptions, yet the temptation to make an approximate model with an exact solution overtakes the solution step in the modeling process. Special debt derivatives called collateralized debt obligations or CDOs implicated in the economic collapse of 2008 are an example. Each CDO was a unique mix of assets, but CDO modeling used general assumptions that were not associated with the specific mix. Additionally, the CDO models used assumptions which underestimated the correlation of movements of the parts of the mix [66]. Valencia [66] says that “The CDO fiasco was an egregious and relatively rare case of an instrument getting way ahead of the ability to map it mathematically.”

It is important to remember to apply mathematical models only under circumstances where the assumptions apply [66]. For example “Value At Risk” or VAR models use volatility to statistically estimate the likelihood that a given portfolio’s losses will exceed a certain amount. However, VAR works only for liquid securities over short periods in normal markets. VAR cannot predict losses under sharp unexpected drops that are known to occur more frequently than expected under simple hypotheses. Mathematical economists, especially Taleb, have heavily criticized the misuse of VAR models.

Alternatives to Black-Scholes. Financial economists and mathematicians have suggested several alternatives to the Black-Scholes model. These alternatives include:

- **stochastic volatility** models where the future volatility of a security price is uncertain; and
- **jump-diffusion models** where the security price experiences occasional jumps rather than continuous change.
The difficulty in mathematically analyzing these models and the typical lack of closed-form solutions means that these models are not as widely known or celebrated as the Black-Scholes-Merton theory.

In spite of these flaws, the Black-Scholes model does an adequate job of generally predicting market prices. Generally, the empirical research is supportive of the Black-Scholes model. Observed differences have been small (but real!) compared with transaction costs. Even more importantly for mathematical economics, the Black-Scholes model shows how to assign prices to risky assets by using the principle of no-arbitrage applied to a replicating portfolio and reducing the pricing to applying standard mathematical tools.

Section Ending Answer. After solving the Black-Scholes equation and deriving parameter dependence and sensitivity of the solution is the stage with evaluation of the relations by prediction and verification. This is done with actual financial data. Then comes comparison and refinement of the model, beginning the cycle again.

Problems.

Exercise 7.36. A pharmaceutical company has a stock that is currently $25. Early tomorrow morning the Food and Drug Administration will announce that it has either approved or disapproved for consumer use the company’s cure for the common cold. This announcement will either immediately increase the stock price by $10 or decrease the price by $10. Discuss the merits of using the Black-Scholes formula to value options on the stock.

Exercise 7.37. Consider
1. A standard normal random variable $Z$ with pdf $f_Z(x) = e^{-x^2/2}/\sqrt{2\pi}$,
2. a “double exponential random variable” $X$ with pdf $f_X(x) = (1/2)e^{-|x|}$,
3. a standard uniform random variable $U$ with pdf $f_U(x) = 1$ for $x = [-1/2, 1/2]$ and 0 elsewhere, and
4. a Cauchy random variable $Y$ with pdf $f_Y(x) = 1/(\pi(1 + x^2))$.
(1) On the same set of axes with a horizontal axis $-5 \leq x \leq 5$, graph the pdf of each of these random variables.
(2) For each of the random variables $Z$, $X$, $U$ and $Y$ calculate the probability that the r.v. is greater than 1, 2, 3, 4. Make a well-organized and labeled table to display your results.
(3) On the basis of the answers from the previous two questions, rank the random variables $Z$, $X$, $U$ and $Y$ in terms of the fatness of the tails.

Exercise 7.38. Show that the variance of the Cauchy distribution is undefined.

Exercise 7.39. Find a source of historical data on options. These sources usually list the implied volatility along with the call and put option prices. (If the implied volatility is not listed, calculate the implied volatility.) Plot the implied volatility as a function of the strike price to illustrate the volatility smile.
Notes

0. 1. (page xiii) The Mathematical Association of America Committee on the Undergraduate Program recommendations are from the 2015 CUPM Curriculum Guide to Majors in the Mathematical Sciences, [57], used with permission, copyright Mathematical Association of America, 2015, all rights reserved.

1. 1. (page 5) The history of mathematical finance is complete and detailed because many of the important figures are still alive to tell their story directly. It is also colorful and exciting because of fortunes made and lost and because it now directly affects both national and personal economies. This section draws from [25, 26, 42]. An excellent discussion of the ethical issues is [49].
2. (page 9) The story about Samuelson and the names for options is from R. Jarrow and P. Protter [25]. More discussion about options is in [21, 70].
3. (page 13) N. Silver [61] has an extended essay about uncertainty and risk. Additional examples of speculation and hedging with options are in [21, 70]. The example of hedging is expanded from [34].
4. (page 16) Arbitrage is a central concept in finance, and the principle of no arbitrage is critical. Additional discussion and examples are in [2, 21]. A somewhat more mathematical view is in [63] and [13].
5. (page 26) Some of the ideas here about mathematical modeling are motivated by [65] and [40]. Silver [61] is an excellent popular survey of models and their predictions, especially large models developed from big data sets. The classification of mathematical models and solutions is adapted from [56]. Glenn Ledder reviewed this section and suggested several of the problems.
6. (page 31) The nature of randomness is still a topic for philosophic debate, but this text takes a practical modeling approach which is motivated by [31]. For a popular summary of coin flipping, see also the article [1]. Algorithmic complexity is discussed in a popular way at greater length in James Gleick’s book [18].
7. (page 36) Stochastic processes are a central topic in advanced probability. Many texts treat them in varying depth and emphasis, see for instance [15, 53, 55, 64].
8. (page 41) This section is developed from a presentation by Jonathan Kaplan of D.E. Shaw and Co. in summer 2010. Allan Baktoft Jakobsen suggested some improvements and clarifications. The definition of CDO squared is noted in [17, page 166]. Some facts and figures are derived from the graphics in [50] and [27].

2. 1. (page 48) This section follows the general idea presented in [5]. Additional ideas are developed from [2, 6, 13, 70].
2. (page 55) This section is developed from ideas in [5]. The comments about discretization procedures result from ideas presented in [2]. Additional ideas are developed from [6, 13, 70].
8. Notes

3. 1. (page 59) This section is inspired by ideas in William Feller’s classic text [15].
2. (page 66) This section is inspired by W. Feller, in [15]. Some ideas are adapted from [63] and [30]. Steele has an excellent discussion at about the same level as here, but with a slightly more rigorous approach to solving the difference equations. He also gives more information about the fact that the duration is almost surely finite, showing that all moments of the duration are finite. Karlin and Taylor give a treatment of the ruin problem by direct application of Markov chain analysis, which is not essentially different, but points to greater generality.
3. (page 72) This section is inspired by ideas from [63] with additional background information from [15].
4. (page 80) This model in this section is an expanded and simplified version of the model in [64].

4. 1. (page 89) The Weak and Strong Laws are standard topics in probability theory. For example, see [53, 54, 36].
2. (page 95) Moment generating functions and similar transforms are a standard topic in probability theory. See for example [55] and [67].
3. (page 100) The Central Limit Theorem is a standard topic in probability theory. The proofs in this section are adapted from [53]. Further examples and considerations come from [36]. Illustration 2 is adapted from [15].

5. 1. (page 107) This section is inspired by the brief treatment in W. Feller, [15, page 354].
2. (page 111) Brownian motion and the Wiener process are standard topics in probability theory, see for instance [30, 54].
3. (page 116) Approximating Brownian motion with a random walk is a standard idea, dating to the earliest formulation of the process. See [11, page 251]. This section also benefits from ideas in W. Feller [15, Chapter III], and [64].
4. (page 121) The transformations of Brownian motion are the beginning of many marvelous properties, see for example [30, pages 351–353] and [22, pages 23-24].
5. (page 124) This section is adapted from [30] and [54]. The technical note about the algorithm combines discussion about Donkers’s principle from [10, 16, 32, 47].
6. (page 128) Brownian motion as an object of mathematical study fills whole books, the theorems here are a sample from [48] as well as [8, 16, 28, 47, 52].
7. (page 133) Total variation is a standard topic in analysis, quadratic variation is less common, except for stochastic processes. The heuristic proof of the weak law is adapted from [5]. The almost sure theorem is expanded from [29]. The mnemonic statement of the quadratic variation in differential form is derived from Steele’s text [63].

6. 1. (page 141) Stochastic differential equations now has a large literature, most of it very advanced. The material in this section is developed from the more elementary references [20] and [5]. The Existence-Uniqueness Theorem is adapted from [3] and [33].
2. (page 148) Itô’s formula is the fundamental basis for stochastic integration and stochastic differential equations, so it appears in many advanced texts. Elementary treatments are in [20] and [5].
3. (page 161) Although the mathematical model in this section is now over 50 years old and is still used in spite of its flaws, the modeling is not often discussed in this
detail. This section draws ideas from [46, 58]. Information about the Wilshire 5000 comes from [68]. The explanation of q-q plots is adapted from the NIST Engineering Statistics Handbook, [43].

7. 1. (page 168) This derivation of the Black-Scholes equation follows the intuitive derivation in [4].
2. (page 176) The solution of the initial value problem for the heat equation on the infinite real line using Fourier transforms is in [69] or [71] as well as many other texts. Another detailed solution of the Black-Scholes equation is in [70] and [63].
3. (page 182) Using linearity of the Black-Scholes equation to establish put-call parity is from [63]. Deriving put-call parity from arbitrage is the usual way to establish the relation, see [34].
4. (page 187) A more detailed discussion of finding implied volatility is in [2]. A more practical discussion of implied volatility is in [21].
5. (page 193) The greeks are not often discussed in mathematically oriented texts, but they are frequently encountered in financial engineering texts. For example, consider [2] and [21].
6. (page 202) This section is inspired by a number of sources, including [4, 7, 12, 21, 40, 45, 49]. The data in Figure 10 is from HistoricalOptionData.com, http://www.historicaloptiondata.com/; accessed on April 9, 2015.


American Stock Exchange, 6
AMEX, 6
approximate Brownian motion, 115
arbitrage, 4, 14, 15
arbitrage pricing, 15
asset, 5
at the money, 197
Bachelier process, 108, 113
Bachelier, Louis, 2, 108, 112
Bank for International Settlements, 2
Bernoulli, Jacob, 88
Bernstein polynomials, 116
Berry-Esseen Theorem, 98
binomial model
- single period, 45
- binomial probability
  - cumulative distribution, 37
  - binomial tree, 50, 51
birth processes, 33
birth-and-death process, 33
bisection search, 186
black swans, 198
Black, Fisher, 3
Black-Scholes
- option pricing model, 3
Black-Scholes equation, 170
- boundary condition, 170
- terminal value problem, 170
Black-Scholes model, 3
Black-Scholes pricing formula
- call option, 174
- put option, 181
bond, 45, 49
bounded variation, 130
Boyle’s Law, 24
Brown, Robert, 107, 112
Brownian motion, 2, 34, 107, 112
- with drift, 138
- approximate, 115
- continuity, 126
- geometric, 138
- Hausdorff dimension, 128
- non-monotonicity, 127
nowhere differentiable, 127
oscillation, 127
scaled, 138
zero as accumulation point, 127
calculator, 3, 200
Cauchy-Euler equation, see also Euler equation
CBOE, 3, 6
CDO, 36, 38
Central Limit Theorem, 97, 157
- Liapunov’s Theorem, 98
- Lindeberg’s Theorem, 98
Chaïtin, G., 31
Chebyshev’s Inequality, 87
Chicago Board Options Exchange, 3, 6
coin flip, 28, 32
collateralized debt obligation, 38
compound Poisson process, 34
convergence
- in distribution, 95
- in probability, 95
- pointwise, 95
Cox-Ross-Rubenstein, 52, 56
credit-default swap, 2
cycle of modeling, 17
de Moivre, Abraham, 98
degenerate probability distribution, 92
Delta, 190
derivative, 2, 45, 49
difference equations, 62
- general solution, 63
- non-homogeneous, 70
diffusion equation, 106
diffusions, 106
Dolčan’s exponential, 148
Donsker’s Invariance Principle, 124
drift, 157
dynamic hedging, 191
dynamic portfolio, 3
economic agents, 1
efficient market hypothesis, 3, 165, 166
pendulum, 22
Perrin, Jean-Baptiste, 2
Poisson process, 33
Poisson, Simeon, 88
probability by conditioning, 122
probability space, see also sample space
put option, 181
put-call parity, 179
put-call parity principle
  arbitrage, 180
q-q plot, 159, 160
quadratic variation, 130
quantile, 159, 160
quantum randomness, 31
queuing processes, 34
random time, 123
random walk, 33, 34, 65, 103
  Central Limit Theorem, 106
  limiting process, 104
  mean, 104
  phase line, 103
  probability of position, 105
  variance, 104
random walk hypothesis, 165
random walk theory, 30
Reflection Principle, 123
replicating portfolio, 46, 50, 166
reserve ratio, see also reserve requirement
  reserve requirement, 76, 83, 84
return, log, 155
returns, 155
Rho, 192
risk, 10, 201
risk free interest rate, 165
risk management, 12
sample path, 32
sample point, 32
sample space, 32
Samuelson, Paul, 3
Scholes, Myron, 3
security, 45, 49
sensitivity analysis, 20
short selling, 12
simple random walk, 32
speculation, 10, 11
spreadsheets, 19
state space, 32
stochastic, 31
stochastic differential equation, 137
stochastic process, 31, 32
stochastic volatility, 201
stock, 45, 49
strike price, 5, 6
Strong Law of Large Numbers, 89
swaps, 2
symmetries, 121
synthetic portfolios, 180
technical analysis, 29
Texas Instruments, 3, 200
Theta, 192
total quadratic variation, 130
tranche, 37
transform methods, 91
transformations, 121
uncertainty, 10, 201
Value At Risk, 201
van der Waals, J.D., 25
Vega, 193
volatility, 157, 165
  historical, 185
  implied, 186
volatility smile, 197
weak law, 89
Weak Law of Large Numbers, 28, 88, 96
well-posed, 178
white noise, 34
Wiener process, 2, 3, 34, 107, 108, 112
  hitting time, 122
  independent increments, 108
  inversion, 118
  joint probability density, 109
  Markov property, 108
  maximum, 124
  quadratic variation, 132
  reflection principle, 123
  ruin problem, 124
  scale, 118
  security prices, 109
  shift, 118
  total variation, 133
Wiener, Norbert, 3, 107, 112
Wilshire 5000, 157