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. In his doctoral thesis completed at the Sorbonne in 1900, Louis Bachelier originated mathematical financial modeling using the theory of speculation in the Paris markets. 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 their 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 five 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 [**25**].

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 [**25**]. 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 [**25**].

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 handheld 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 that 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 deregulation 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, as they are only approximations to a complicated world. The practitioner should apply the models only after carefully assessing their limitations. 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 bailout 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.[1]

Key Concepts.

(1) **Finance theory** is the study of economic agents' behavior allocating financial resources and risks across alternative financial instruments over time in an uncertain environment. Mathematics provides tools to model and analyze that behavior in allocation and time, taking into account uncertainty.

(2) Louis Bachelier originated mathematical financial modeling 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.

(3) The most important theoretical development was the Black-Scholes model for option pricing published in 1973.

(4) In the 1980s, the development of sophisticated mathematical models and their adoption into financial practice accelerated. 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.

Vocabulary.

(1) **Finance theory** is the study of economic agents' behavior allocating financial resources and risks across alternative financial instruments over time in an uncertain environment.

(2) A **derivative** is a financial agreement between two parties that depends on the future price or performance of an underlying asset. The underlying asset could be a stock, a bond, a currency, or a commodity.

(3) **Types of derivatives**: 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 one lot of obligations for another. Derivatives can be based on pretty much anything as long as two parties are willing to trade risks and can agree on a price.

(4) The **efficient markets hypothesis** says 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.

Problems.

Exercise 1.1. Write a short summary of "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**, p. 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 of those of a call. The holder of a call option wants the asset price to rise, and 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

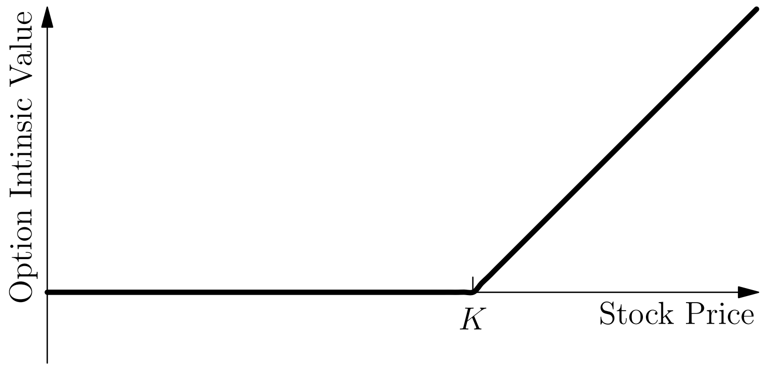

Figure 1.1. Intrinsic value of a call option

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 let K be 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 (see Figure 1.1). 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 imply 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*.

Table 1.1. Effect on price of increases in the variables influencing option prices.

Increase in Variable	European Call	European Put
Stock Price	Increase	Decrease
Strike Price	Decrease	Increase
Time to Expiration	Not obvious	Not obvious
Volatility	Increase	Increase
Risk-free Rate	Not obvious	Not obvious

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 such as 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 1.1 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 of 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 the option 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 specifiying the rate of change; and
- "as volatility increases, the chance that the stock price will either increase or decrease greatly ... increases".

The goal in the 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.[2]

Key Concepts.

(1) A *call option* is the right to buy an asset at an established price at a certain time.

(2) A *put option* is the right to sell an asset at an established price at a certain time.

(3) A European option may only be exercised at the end of its life on the expiry date, an American option may be exercised at any time during its life up to the expiry date.

(4) 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 σ;
    - the risk-free interest rate r; and
    - the dividends expected during the life of the option.

Vocabulary.

(1) A **call option** is the right to buy an asset at an established price at a certain time.

(2) A **put option** is the right to sell an asset at an established price at a certain time.

(3) A **future** is a contract to buy (or sell) an asset at an established price at a certain time.

(4) The **expiry date** specifies the month in which the European option ends.

(5) **Volatility** is a measure of the variability and therefore the risk of a price, usually the price of a security.

(6) The **holder** of the option *may* purchase from (or sell to) the **option writer** an asset known as the **underlying asset** for a definite amount known as the **exercise price** or **strike price**.

(7) 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.

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 1.1.

(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 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 a **long straddle** and is 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, where $K_1 < K_2$. (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 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 she chooses to, purchase a share of XYZ at price \$2.50 per share *no matter what the current price of XYZ stock is!* 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. (Three-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: More speculation with calls. Consider the profit and loss of an 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 three 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% rise in the stock from $100 to $101 would drive the option price to $12.73. Consider the total effects in Table 1.2.

Suppose a trader has an original portfolio composed 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 1.2 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% change in the price of the stock. If the stock increases 1%, 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, if the stock price falls by 1%, 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

Table 1.2. Hedging to keep a portfolio constant.

Original Portfolio	$S = 100, C = \$11.84$
8,944 shares of stock	\$894, 400
Short position on 10,000 options	−\$118,400
Net value	\$776,000
Stock Price rises 1%	$S = 101, C = \$12.73$
8,944 shares of stock	\$903,344
Short position on 10, 000 options	−\$127,300
Net value	\$776,044
Stock price falls 1%	$S = 99, C = \$10.95$
8,944 shares of stock	\$885, 456
Short position on options	−\$109,500
Net value	\$775,956

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", or buying low-cost properties, fixing them 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.[3]

Key Concepts.

(1) Options have two primary uses, **speculation** and **hedging**.

(2) Options can be a cheap way of exposing a portfolio to a large amount of risk. Sometimes a large amount of risk is desirable. This is the use of options and derivatives for **speculation**.

(3) Options allow the investor to insure against adverse security value movements while still benefiting from favorable movements. This is use of options for **hedging**. This insurance comes at the cost of buying the option.

Vocabulary.

(1) **Risk** is random financial variation that has a known (or assumed) probability distribution. **Uncertainty** is chance variability that is due to unknown and unmeasured factors.

(2) **Speculation** is to assume a financial risk in anticipation of a gain, especially to buy or sell to profit from market fluctuations.

(3) **Hedging** is to protect oneself financially against loss by a counterbalancing transaction, especially to buy or sell assets as a protection against loss because of price fluctuation.

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 three-month call with strike of \$52 costs \$2.50. You have \$5,000 to invest. Consider two alternative 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 three-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–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, on 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 United States 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 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

$$-100 \text{ shares } \times 125 \frac{€}{\text{share}} \div 0.8757 \frac{\$}{€} + 100 \text{ shares} \times 145 \frac{\$}{\text{share}} = \$225.71$$

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 such as 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 one year was 1.4312. Suppose 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 United States 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 United States and buy 700, 280 Euros and put them in the European bank. Simultaneously, the arbitrageur would create a contract to sell Euros in a 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 United States 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 nonexistent 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.[4]

Key Concepts.

(1) 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 market they can occur.

(2) Prices change as the investors move to take advantage of such an opportunity. As a consequence, the arbitrage opportunity disappears. This becomes an economic principle: *in an efficient market there are no arbitrage opportunities.*

(3) The principle of *arbitrage pricing* is that any two investments with identical payout streams must have the same price.

Vocabulary.

(1) **Arbitrage** is locking in a riskless profit by simultaneously entering into transactions in two or more markets, exploiting mismatches in pricing.

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,

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

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 U.S. 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 the modeling 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 thus 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, who advises "Keep it sophisticatedly simple" [**61**, page 225].

At a more detailed level, mathematical modeling involves four 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 1.2 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, which changes 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.

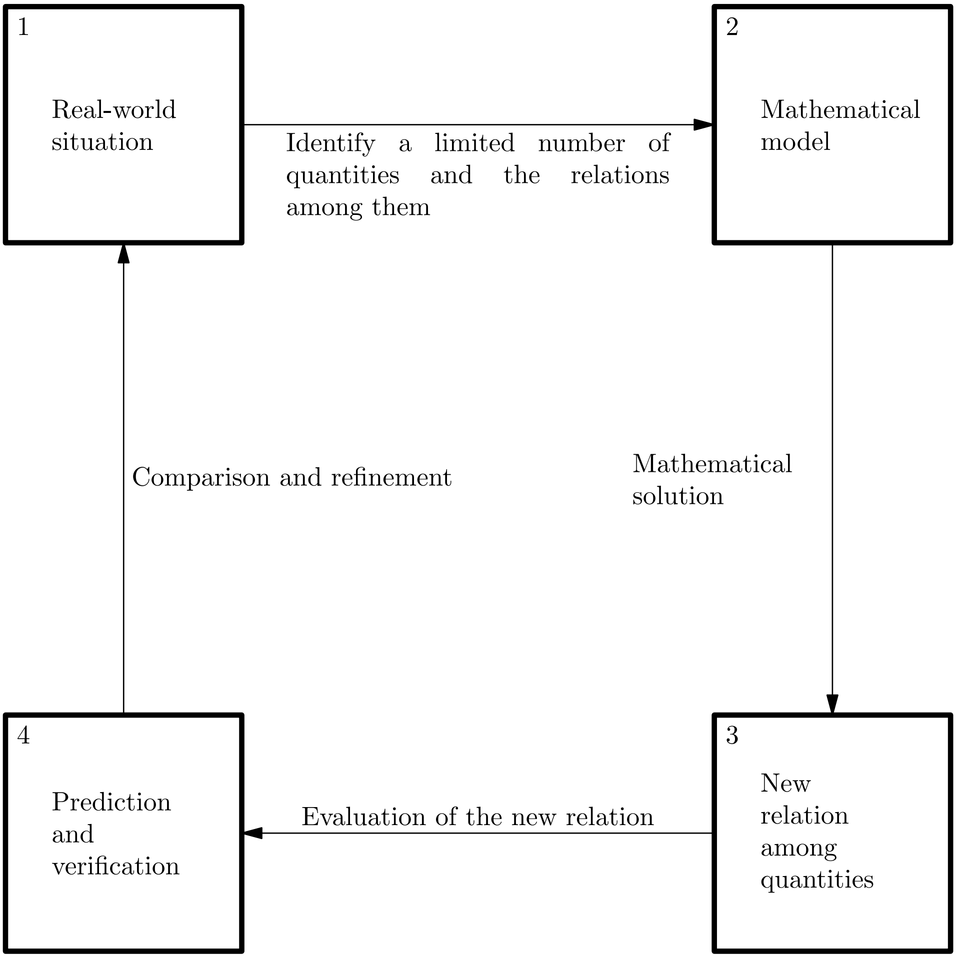

Figure 1.2. The cycle of modeling.

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 function and its 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, or 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 1.2.

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 nondimensionalized model reveals significant relationships and major influences. Unit analysis is an important part of modeling, and it 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, they should then 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, what to do when it is wrong, and how to minimize 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 and test them in another cycle in the modeling 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 steps 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 automatically or naturally. Scientists, engineers, mathematicians, and students must practice and develop the craft of creating, solving, using, and interpreting a mathematical 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 modeling is to represent a real-world situation in a way that is a close approximation, even ideally exact, to a mathematical structure that is still solvable. This represents 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 representations 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 let c be the number of coins. Write two equations in two unknowns:

$$V/c = 20,$$
$$(V + 25)/(c + 1) = 21.$$

The solution is three 25-cent coins and one 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.

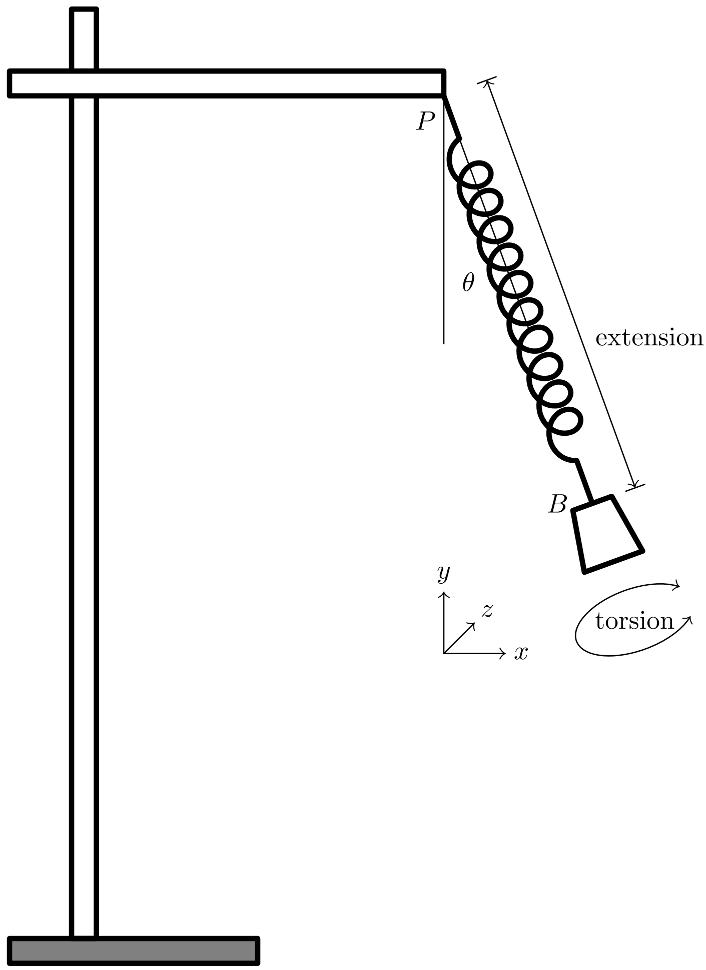

Figure 1.3. Schematic diagram of a pendulum.

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 1.3. The pendulum is a heavy bob B attached to a pivot point P by an elastic spring. The pendulum has at least five 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 θ.

(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 using Newton's laws of motion, we obtain the differential equation for the pendulum angle θ with respect to vertical:

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

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 solve this equation numerically or 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

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

This differential equation is analytically solvable with standard methods yielding

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

for appropriate constants A, ω, and ϕ 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 θ 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 four 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 four 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

$$\left(P + \frac{n^2 a}{V^2}\right)(V - b) = nRT.$$

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 the following.

(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 in financial engineering.

We must then alter our assumptions, rederive 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.[5]

Key Concepts.

(1) "All mathematical models are wrong, but some mathematical models are useful." [**9**]

(2) 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.

(3) 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.

(4) A good description of the model will begin with an organized and complete description of important factors and observations. From there, the model must summarize and condense the observations into a small set of hypotheses that capture the essence of the observations. The hypotheses translate into a mathematical structure that becomes the heart of the mathematical model.

(5) Once the modelers create the model, then they should derive some new relations among the important quantities selected to describe the real-world situation.

(6) A model requires testing and sensitivity analysis.

(7) Modelers must take the results from the previous steps and use them to refine the interpretation and understanding of the real-world situation.

(8) The modeling process is an intense and structured application of *critical thinking*.

(9) Models can be classified as

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

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

(10) When observed outcomes deviate unacceptably from predicted behavior in honest scientific or engineering work, we must alter our assumptions, re-derive the quantitative relationships, perhaps with more sophisticated mathematics or introducing more quantities and begin the cycle of modeling again.

Vocabulary.

(1) A **mathematical model** is a mathematical structure (often an equation) expressing a relationship among a limited number of quantifiable elements from the real world or some isolated portion of it.

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 a 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 one-liter jar instead of a cubical box. (Note that a cubical box 10 cm on a side has a volume of one 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:

> 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 1.4 as a conceptual model of the modeling process.

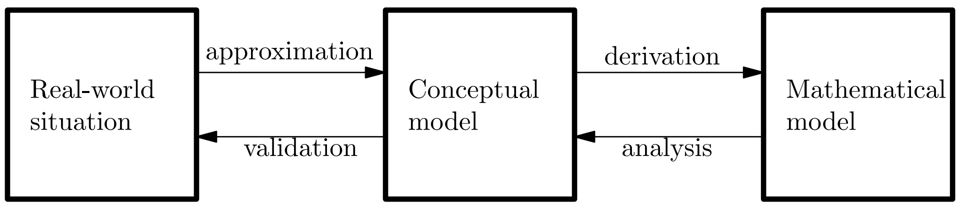

Figure 1.4. The process of mathematical modeling according to Glenn Ledder.

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 coteaching 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 Tosses and Randomness. The simplest, most common and, in some ways, most basic example of a random process is a coin toss. We toss 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 tossed 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 toss, we can derive some mathematical consequences. We will do this extensively in the chapter on limit theorems for coin tossing. 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 toss behavior. In this way, the simple mathematical model is satisfactory.

In other ways the probability approach is unsatisfactory. A coin toss is a physical process, subject to the physical laws of motion. The renowned applied mathematician J. B. Keller investigated coin tosses in this way. He assumed a circular coin with negligible thickness tossed 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 toss until the coin lands on the surface $y = 0$. The initial conditions imparted to the coin toss are the initial upward velocity and the initial rotational velocity. With additional simplifying assumptions, Keller shows that the fraction of tosses that 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; see Figure 1.5. 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.

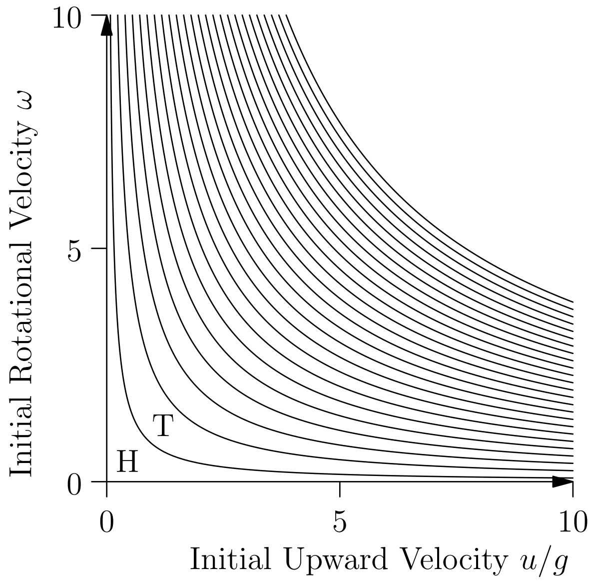

Figure 1.5. Initial conditions for a coin toss, following Keller.

Another mathematician who often collaborated with J. B. Keller, Persi Diaconis, has exploited this determinism. Diaconis, an accomplished magician, is reportedly able to toss many heads in a row using his manual skill. Moreover, he has worked with mechanical engineers to build a precise coin tossing machine that can toss many heads in a row by controlling the initial conditions precisely.

Mathematicians Diaconis, Susan Holmes, and Richard Montgomery have done an even more detailed analysis of the physics of coin tosses in [**14**]. The coin tossing machines pictured there help to show that tossing 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 tossing 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 one 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 random walk theory postulate a probability distribution for the market price movements. In this way, random walk theory mimics the mathematical model of a coin toss, substituting a probability distribution of outcomes for the ability to predict what will really happen.

If a coin toss, 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 toss 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 toss 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 toss is physically determined. The numbers generated by a 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 π is not random since a short formula can generate the digits of π.

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 the 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 twenty 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 $20!/(3!\cdot 3!\cdot 3!\cdot 3!\cdot 4!\cdot 4!) = 3{,}259{,}095{,}840{,}000$ 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 3×10^{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.[6]

Key Concepts.

(1) Assigning probability 1/2 to the event that a coin will land heads and probability 1/2 to the event that a coin will land tails is a mathematical model that summarizes our experience with many coins.

(2) A coin flip is a deterministic physical process, subject to the physical laws of motion. Extremely narrow bands of initial conditions determine the outcome of heads

or tails. The assignment of probabilities 1/2 to heads and tails is a summary measure of all initial conditions that determine the outcome precisely.

(3) The **random walk theory** of asset prices claims that market prices follow a random path without any influence by past price movements. This theory says it is 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.

Vocabulary.

(1) **Technical analysis** claims to predict security prices by relying on the assumption that market data, such as price, volume, and patterns of past behavior can help predict future (usually short-term) market trends.

(2) The **random walk theory** of the market claims that market prices follow a random path up and down according to some probability distribution without any influence by past price movements. This assumption means that it is not possible to predict which direction the market will move at any point, although the probability of movement in a given direction can be calculated.

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 Notation. 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" (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 nonnegative real numbers. Usually J represents the nonnegative integers $0, 1, 2, \dots$ 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 Ω be a set, usually called the **sample space** or **probability space**. An element ω of Ω 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 nonnegative 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 toss sequence. The index set is the set of positive integers, counting the number of the toss. The sample space is the set of all possible infinite coin toss sequences

$$\Omega = \{HHTHTTTHT, \dots, THTHTTHHT, \dots, \dots\}.$$

We take the state space to be the set $1, 0$ so that $X_n = 1$ if toss n comes up heads, and $X_n = 0$ if the toss comes up tails. Then the coin toss 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 tossing sequence up to toss number n.

Alternatively, we can take the same index set, the same probability space of coin toss sequences, and define $Y_n = 1$ if toss n comes up heads and $Y_n = -1$ if the toss comes up tails. This is just another way to encode the coin tosses 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 a 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 $0, 1, 2, \dots$. 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

$$\mathbb{P}\left[X_j = l \,|\, X_0 = k_0, X_1 = k_1, \dots, X_{j-1} = k_{j-1}\right] = \mathbb{P}\left[X_j = l \,|\, X_{j-1} = k_{j-1}\right].$$

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 $X_0 = k_0, X_1 = k_1, \dots, X_{j-2} = k_{j-2}$ 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

$$\begin{aligned} &\mathbb{P}\left[X_j = l \,|\, X_{j-1} = k\right] = 1/2 \quad \text{if } l = k-1 \text{ or } l = k+1, \\ &\mathbb{P}\left[X_j = l \,|\, X_{j-1} = k\right] = 0 \quad \text{otherwise.} \end{aligned}$$

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 the 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

$$\mathbb{P}\left[X(t_n) \le x_n \,|\, X(t_{n-1}), \dots, X(t_1)\right] = \mathbb{P}\left[X(t_n) \le x_n \,|\, X(t_{n-1})\right].$$

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 lifespan 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 λ. 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 *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 *Gaussian processes*, including *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 *Brownian motion* and also as the *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 *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. Refining the model so that prices are nonnegative leads to the stochastic process known as *geometric Brownian motion*.

Family of Stochastic Processes. A sequence or interval of random outcomes, that is, a string of random outcomes dependent on time as well as the randomness, is a **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 1.6, along with an indication of the stochastic process types studied in this text.

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 ω. Each ω maps to an associated function $X(t)$. This means that one can look at a stochastic process as a mapping from the sample space Ω 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 several questions:

- 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

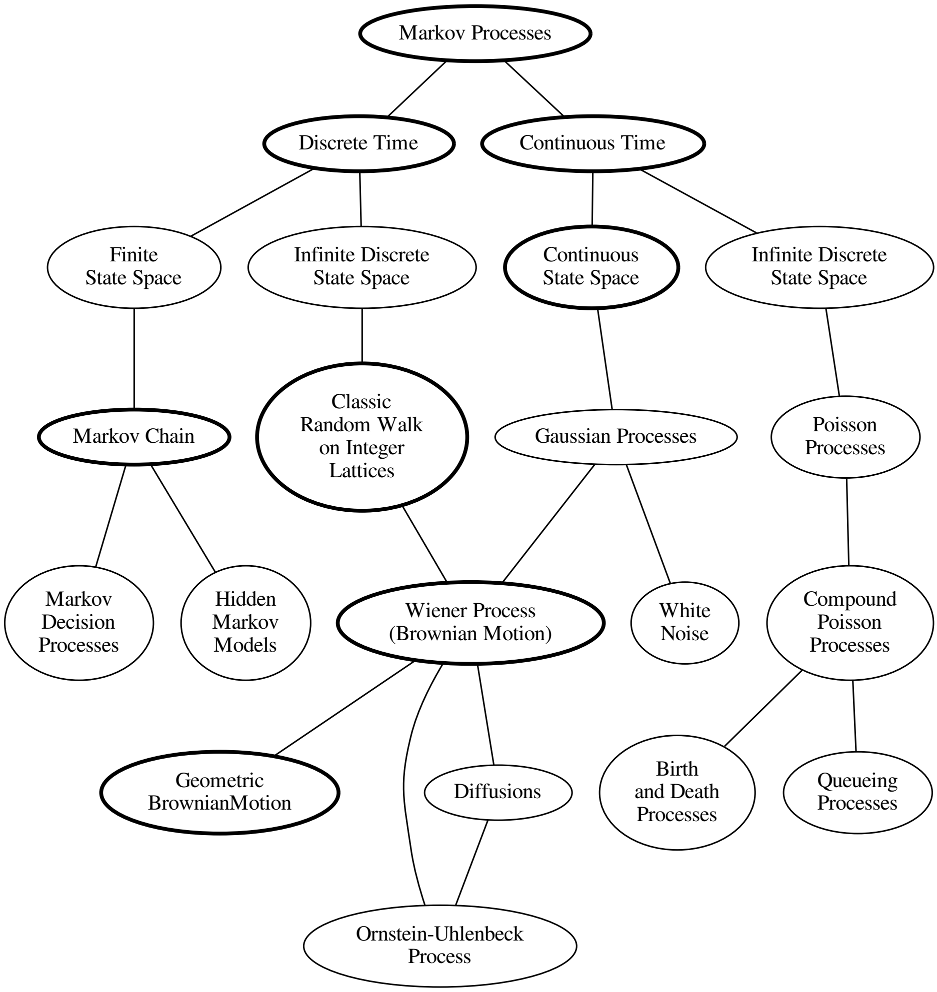

Figure 1.6. A tree of some stochastic processes, from most general at the top to more specific at the end leaves. Stochastic pqrocesses studied in this text have thick borders, others are thin.

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.[7]

Key Concepts.

(1) A sequence or interval of random outcomes, that is, a string of random outcomes dependent on time as well as the randomness is a **stochastic process**. With the inclusion of a time variable, the rich range of random outcome distributions becomes a great variety of stochastic processes. Nevertheless, the most commonly studied types of random processes have connections.

(2) 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.

(3) Another way to look at a stochastic process is to consider the stochastic process as a function of the sample point ω. Each ω maps to an associated function of time $X(t)$. This means that one may look at a stochastic process as a mapping from the sample space Ω to a set of functions. In this interpretation, stochastic processes are a generalization from the random variables of elementary probability theory.

Vocabulary.

(1) A sequence or interval of random outcomes, that is, random outcomes dependent on time is a **stochastic process**.

(2) Let J be a subset of the non-negative real numbers. Let Ω be a set, usually called the **sample space** or **probability space**. An element ω of Ω 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 $X : (J, \Omega) \to S$, that is a function of both time and the sample point to the state space.

(3) The particular stochastic process usually called a **simple random walk** T_n gives the position in the integers after taking a step to the right for a head, and a step to the left for a tail.

(4) A **Markov chain** is sequence of random variables X_j where the index j runs through $0, 1, 2, \ldots$. The state space may be either a finite or infinite set of discrete states. The defining property of a Markov chain is that

$$\mathbb{P}\left[X_j = l \,|\, X_0 = k_0, X_1 = k_1, \ldots, X_{j-1} = k_{j-1}\right] = \mathbb{P}\left[X_j = l \,|\, X_{j-1} = k_{j-1}\right].$$

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}

(5) 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 non-negative 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 $t_1 < t_2 < \dots < t_n$ and real number x_n, we have

$$\mathbb{P}\left[X(t_n) \le x_n \,|\, X(t_{n-1}), \dots, X(t_1)\right] = \mathbb{P}\left[X(t_n) \le x_n \,|\, X(t_{n-1})\right].$$

Many of the models in this text will naturally be Markov processes because of the intuitive appeal of this memory-less property.

1.8 A Model of Collateralized Debt Obligations

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

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

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 it amounted 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 two 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

$$\mathbb{P}\left[S_{100} \le n\right] = \sum_{j=0}^{n} \binom{100}{j} p^j (1-p)^{100-j}.$$

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.)

To be explicit, suppose 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, such as 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 it 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 $0, 1, 2, \dots, 9$ 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 $0, 1, 2, \dots, 9$ 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

$$\sum_{j=0}^{9} \binom{100}{j} p_T(10)^j (1 - p_T(10))^{100-j},$$

and the probability of default is

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

For example, with $p = 0.05$ and $p_T(10) = 0.028188$, then $p_{\text{CDO}}(10) = 0.00054385$. Roughly, the CDO has only 1/100 of the default probability of the original mortgages, by virtue of redistributing 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 $0, 1, 2, 3 \dots, M-1$ 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 nonpayment—it is 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. 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 risk although it is only a 0.01 increase in 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 subprime 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! (See Figure 1.7 for a graphical image of the sensitivity of the probability of default.)

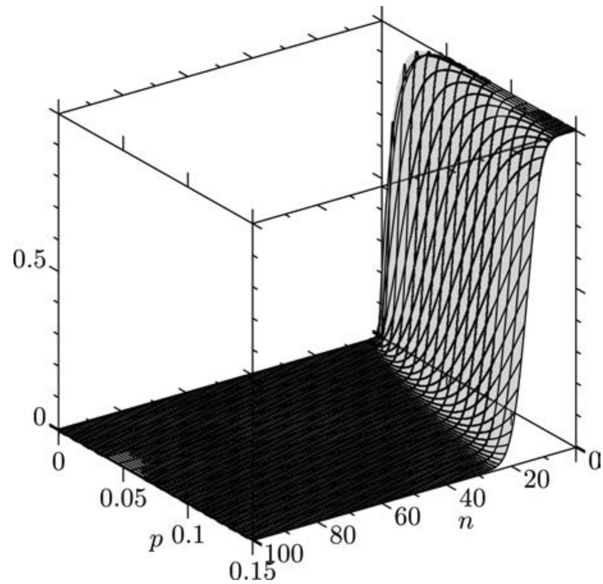

Figure 1.7. Default probabilities as a function of both the tranche number 0 to 100 and the base mortgage default probability 0.01 to 0.15.

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 overconfidence 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 form the basis of any investment strategy or more serious economic analysis. 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 subprime loans or subprime 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 subprime 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 nonuniform 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.[8]

Key Concepts.

(1) We can make a simple mathematical model of a type of financial derivative using only the idea of a binomial probability.

(2) We must investigate the sensitivity of the model to the parameter values to completely understand the model.

(3) This simple model provides our first illustration of the modeling cycle in mathematical finance, but even so, it yields valuable insights.

Vocabulary.

(1) A **tranche** is a portion or slice of a set of securities.

(2) A **collateralized debt obligation** or **CDO** is a derivative security backed by a pool or slice of other securities. CDOs can be made of any kind of debt and do not necessarily derive from mortgages. CDOs contain multiple slices or tranches, each slice made of debt with an associated risk.

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

$$\begin{aligned}\mathbb{E}\left[U\right] &= \sum_{n=1}^{101}\sum_{j=0}^{n-1}\binom{100}{j}p^j(1-p)^{100-j} \\ &= \sum_{j=0}^{100}(100-j)\binom{100}{j}p^j(1-p)^{100-j} = 100(1-p).\end{aligned}$$

That is, the expected payout from the collection of contracts is exactly the same as 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 ith 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 U.S. 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 the 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.