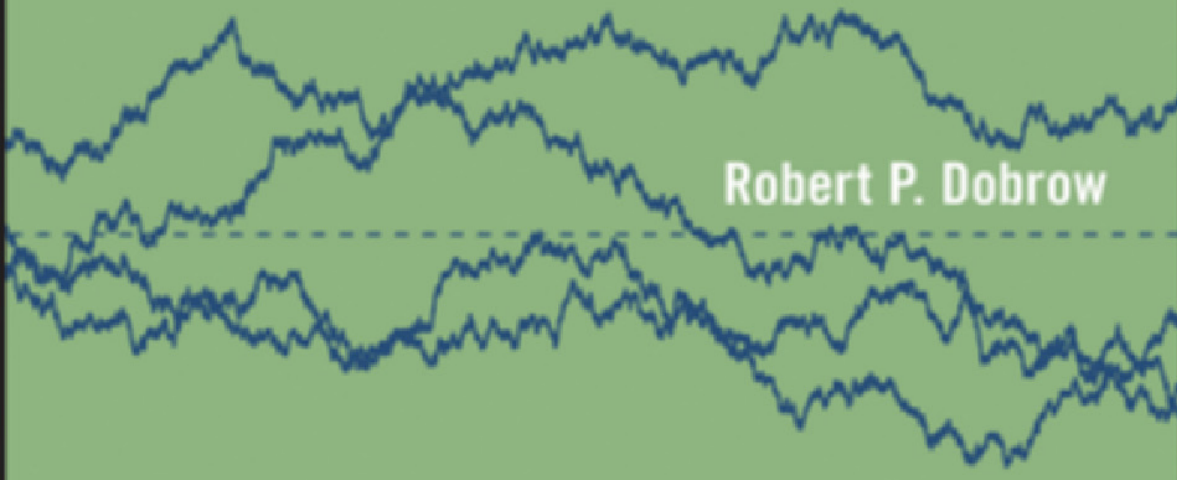
Probability

WITH APPLICATIONS AND R



Robert P. Dobrow

WILEY

PROBABILITY

PROBABILITY

With Applications and R

ROBERT P. DOBROW
Department of Mathematics
Carleton College
Northfield, MN

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

Dobrow, Robert P.
Probability: with applications and R / Robert P. Dobrow, Department of Mathematics, Carleton College.
pages cm
Includes bibliographical references and index.
ISBN 978-1-118-24125-7 (hardback)
1. Probabilities–Data processing. 2. R (Computer program language) I. Title.
QA276.45.R3D63 2013
519.20285′5133–dc23

2013013972

Printed in the United States of America

ISBN: 9781118241257

10 9 8 7 6 5 4 3 2 1

To my wonderful family
Angel, Joe, Danny, Tom

CONTENTS

PREFACE

Probability: With Applications and R is a probability textbook for undergraduates. It assumes knowledge of multivariable calculus. While the material in this book stands on its own as a "terminal" course, it also prepares students planning to take upper level courses in statistics, stochastic processes, and actuarial sciences.

There are several excellent probability textbooks available at the undergraduate level, and we are indebted to many, starting with the classic *Introduction to Probability Theory and Its Applications* by William Feller.

Our approach is tailored to our students and based on the experience of teaching probability at a liberal arts college. Our students are not only math majors but come from disciplines throughout the natural and social sciences, especially biology, physics, computer science, and economics. Sometimes we will even get a philosophy, English, or arts history major. They tend to be sophomores and juniors. These students love to see connections with "real-life" problems, with applications that are "cool" and compelling. They are fairly computer literate. Their mathematical coursework may not be extensive, but they like problem solving and they respond well to the many games, simulations, paradoxes, and challenges that the subject offers.

Several features of our textbook set it apart from others. First is the emphasis on simulation. We find that the use of simulation, both with "hands-on" activities in the classroom and with the computer, is an invaluable tool for teaching probability. We use the free software R and provide an appendix for getting students up to speed in using and understanding the language. We recommend that students work through this appendix and the accompanying worksheet. The book is not meant to be an instruction manual in R; we do not teach programming. But the book does have numerous examples where a theoretical concept or exact calculation is reinforced by a computer simulation. The R language offers simple commands for generating

samples from probability distributions. The book includes pointers to numerous R script files, that are available for download, as well as many short R "one-liners" that are easily shown in the classroom and that students can quickly and easily duplicate on their computer. Throughout the book are numerous "R:" display boxes that contain these code and scripts.

In addition to simulation, another emphasis of the book is on applications. We try to motivate the use of probability throughout the sciences and find examples from subjects as diverse as homelessness, genetics, meteorology, and cryptography. At the same time, the book does not forget its roots, and there are many classical chestnuts like the problem of points, Buffon's needle, coupon collecting, and Montmort's problem of coincidences.

Following is a synopsis of the book's 10 chapters.

Chapter 1 begins with basics and general principles: random experiment, sample space, and event. Probability functions are defined and important properties derived. Counting, including the multiplication principle and permutations, are introduced in the context of equally likely outcomes. Random variables are introduced in this chapter. A first look at simulation gives accessible examples of simulating several of the probability calculations from the chapter.

Chapter 2 emphasizes conditional probability, along with the law of total probability and Bayes formula. There is substantial discussion of the birthday problem.

Independence and sequences of independent random variables are the focus of Chapter 3. The most important discrete distributions—binomial, Poisson, and uniform—are introduced early and serve as a regular source of examples for the concepts to come. Binomial coefficients are introduced in the context of deriving the binomial distribution.

Chapter 4 contains extensive material on discrete random variables, including expectation, functions of random variables, and variance. Joint discrete distributions are introduced. Properties of expectation, such as linearity, are presented, as well as the method of indicator functions. Covariance and correlation are first introduced here.

Chapter 5 highlights several families of discrete distributions: geometric, negative binomial, hypergeometric, multinomial, and Benford's law.

Continuous probability begins with Chapter 6. The chapter introduces the uniform and exponential distributions. Functions of random variables are emphasized, including maximums, minimums, and sums of independent random variables. There is material on geometric probability.

Chapter 7 highlights several important continuous distributions starting with the normal distribution. There is substantial material on the Poisson process, constructing the process by means of probabilistic arguments from i.i.d. exponential inter-arrival times. The gamma and beta distributions are presented. There is also a section on the Pareto distribution with discussion of power law and scale invariant distributions.

Chapter 8 is devoted to conditional distributions, both in the discrete and continuous settings. Conditional expectation and variance are emphasized as well as computing probabilities by conditioning.

The important limit theorems of probability—law of large numbers and central limit theorem—are the topics of Chapter 9. There is a section on moment-generating functions, which are used to prove the central limit theorem.

Chapter 10 has optional material for supplementary discussion and/or projects. The first section covers the bivariate normal distribution. Transformations of two or more random variables are presented next. Another section introduces the method of moments. The last three sections center on random walk on graphs and Markov chains, culminating in an introduction to Markov chain Monte Carlo. The treatment does not assume linear algebra and is meant as a broad strokes introduction.

There is more than enough material in this book for a one-semester course. The range of topics allows much latitude for the instructor. We feel that essential material for a first course would include Chapters 1–4, 6, and parts of 8 and 9.

Additional features of the book include the following:

- Over 200 examples throughout the text and some 800 end-of-chapter exercises. Includes short numerical solutions for most odd-numbered exercises.
- End-of-chapter reviews and summaries highlight the main ideas and results from each chapter for easy access.
- Starred subsections are optional and contain more challenging material, assuming a higher mathematical level.
- Appendix A: *Getting up to speed in R* introduces students to the basics of R. Includes worksheet for practice.
- Appendix E: *More problems to practice* contains a representative sample of 25 exercises with fully worked out solutions. The problems offer a good source of additional problems for students preparing for midterm and/or final exams.
- A website containing relevant files and errata has been established. All the R code and script files used throughout the book, including the code for generating the book's graphics, are available at this site. The URL is

  `http://www.people.carleton.edu/~rdobrow/Probability`
- An instructor's solutions manual with detailed solutions to all the exercises is available for instructors who teach from this book.

ACKNOWLEDGMENTS

We are indebted to friends and colleagues who encouraged and supported this project. The students of my Fall 2012 Probability class were real troopers for using an early manuscript that had an embarrassing number of typos and mistakes and offering a volume of excellent advice. We also thank Marty Erickson, Jack Goldfeather, Matthew Rathkey, Wenli Rui, and Zach Wood-Doughty. Professor Laura Chihara field-tested an early version of the text in her class and has made many helpful suggestions. Thank you to Jack O'Brien at Bowdoin College for a detailed reading of the manuscript and for many suggestions that led to numerous improvements.

Carleton College and the Department of Mathematics were enormously supportive, and I am grateful for a college grant and additional funding that supported this work. Thank you to Mike Tie, the Department's Technical Director, and Sue Jandro, the Department's Administrative Assistant, for help throughout the past year.

The staff at Wiley, including Steve Quigley, Amy Hendrickson, and Sari Friedman, provided encouragement and valuable assistance in preparing this book.

INTRODUCTION

All theory, dear friend, is gray, but the golden tree of life springs ever green.
—Johann Wolfgang von Goethe

Probability began by first considering games of chance. But today it is the practical applications in areas as diverse as astronomy, economics, social networks, and zoology that enrich the theory and give the subject its unique appeal.

In this book, we will flip coins, roll dice, and pick balls from urns, all the standard fare of a probability course. But we have also tried to make connections with real-life applications and illustrate the theory with examples that are current and engaging.

You will see some of the following case studies again throughout the text. They are meant to whet your appetite for what is to come.

I.1 WALKING THE WEB

There are about one trillion websites on the Internet. When you google a phrase like "Can Chuck Norris divide by zero?," a remarkable algorithm called PageRank searches these sites and returns a list ranked by importance and relevance, all in the blink of an eye. PageRank is the heart of the Google search engine. The algorithm assigns an "importance value" to each web page and gives it a rank to determine how useful it is.

PageRank is a significant accomplishment of mathematics and linear algebra. It can be understood using probability, particularly a branch of probability called Markov chains and random walk. Imagine a web surfer who starts at some web page and clicks on a link at random to find a new site. At each page, the surfer chooses

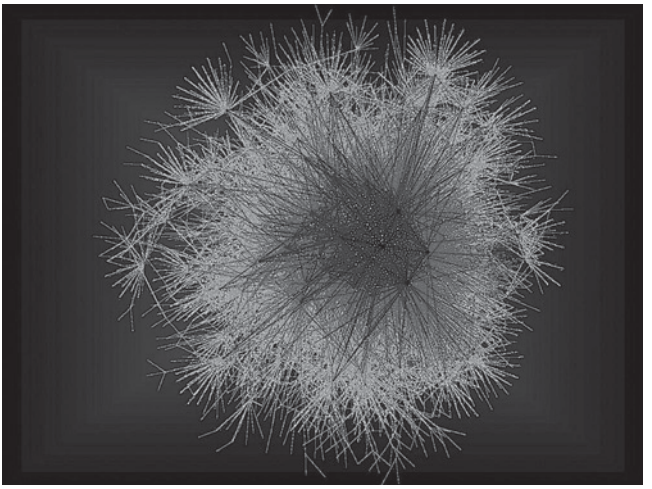


FIGURE I.1: A subgraph of the BGP (Gateway Protocol) web graph consisting of major Internet routers. It has about 6400 vertices and 13,000 edges. Image produced by Ross Richardson and rendered by Fan Chung Graham.

from one of the available hypertext links equally at random. If there are two links, it is a coin toss, heads or tails, to decide which one to pick. If there are 100 links, each one has a 1% chance of being chosen. As the web surfer moves from page to random page, they are performing a random walk on the web (Fig. I.1).

What is the PageRank of site x? Suppose the web surfer has been randomly walking the web for a very long time (infinitely long in theory). The probability that they visit site x is precisely the PageRank of that site. Sites that have lots of incoming links will have a higher PageRank value than sites with fewer links.

The PageRank algorithm is actually best understood as an assignment of *probabilities* to each site on the web. Such a list of numbers is called a *probability distribution.* And since it comes as the result of a theoretically infinitely long random walk, it is known as the *limiting distribution* of the random walk. Remarkably, the PageRank values for billions of websites can be computed quickly and in real time.

I.2 BENFORD'S LAW

Turn to a random page in this book. Look in the middle of the page and point to the first number you see. Write down the first digit of that number.

You might think that such first digits are equally likely to be any integer from 1 to 9. But a remarkable probability rule known as Benford's law predicts that most of your first digits will be 1 or 2; the chances are almost 50%. The probabilities go down as the numbers get bigger, with the chance that the first digit is 9 being less than 5% (Fig. I.2).

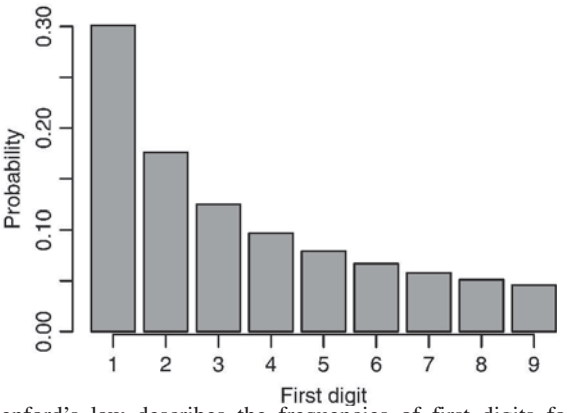


FIGURE I.2: Benford's law describes the frequencies of first digits for many real-life datasets.

Benford's law, also known as the "first-digit phenomenon," was discovered over 100 years ago, but it has generated new interest in the past 10 years. There are a huge number of datasets that exhibit Benford's law, including street addresses, populations of cities, stock prices, mathematical constants, birth rates, heights of mountains, and line items on tax returns. The last example, in particular, caught the eye of business Professor Mark Nigrini (2012), who showed that Benford's law can be used in forensic accounting and auditing as an indicator of fraud.

Durtschi et al. (2004) describe an investigation of a large medical center in the western United States. The distribution of first digits of check amounts differed significantly from Benford's law. A subsequent investigation uncovered that the financial officer had created bogus shell insurance companies in her own name and was writing large refund checks to those companies.

I.3 SEARCHING THE GENOME

Few areas of modern science employ probability more than biology and genetics. A strand of DNA, with its four nucleotide bases adenine, cytosine, guanine, and thymine, abbreviated by their first letters, presents itself as a sequence of outcomes of a four-sided die. The enormity of the data—about three billion "letters" per strand of human DNA—makes randomized methods relevant and viable.

Restriction sites are locations on the DNA that contain a specific sequence of nucleotides, such as G-A-A-T-T-C. Such sites are important to identify because they are locations where the DNA can be cut and studied. Finding all these locations is akin to finding patterns of heads and tails in a long sequence of coin tosses. Theoretical limit theorems for idealized sequences of coin tosses become practically relevant for exploring the genome. The locations for such restriction sites are well described by the Poisson process, a fundamental class of random processes that model locations

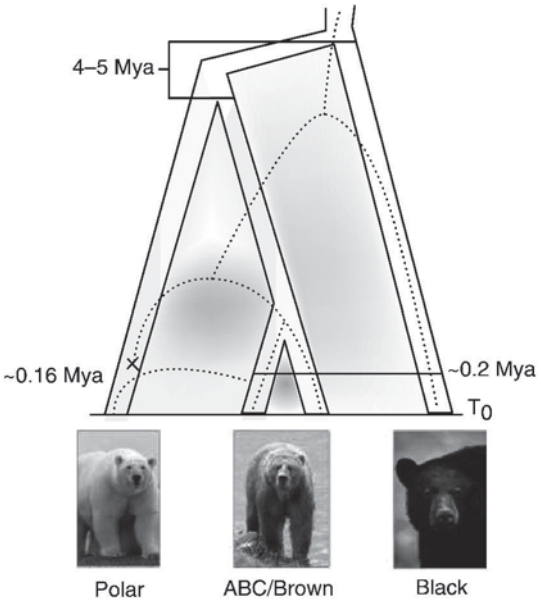


FIGURE I.3: Reconstructed evolutionary tree for bears. Polar bears may have evolved from brown bears four to five million years ago with occasional exchange of genes between the two species (shaded gray areas) fluctuating with key climactic events. Credit: Penn State University.

of restriction sites on a chromosome, as well as car accidents on the highway, service times at a fast food chain, and when you get your text messages.

On the macro level, random processes are used to study the evolution of DNA over time in order to construct evolutionary trees showing the divergence of species. DNA sequences change over time as a result of mutation and natural selection. Models for sequence evolution, called Markov processes, are continuous time analogues of the type of random walk models introduced earlier.

Miller et al. (2012) analyze the newly sequenced polar bear genome and give evidence that the size of the bear population fluctuated with key climactic events over the past million years, growing in periods of cooling and shrinking in periods of warming. Their paper, published in the *Proceedings of the National Academy of Sciences*, is all biology and genetics. But the appendix of supporting information is all probability and statistics (Fig. I.3).

I.4 BIG DATA

The search for the Higgs boson, the so-called "God particle," at the Large Hadron Collider in Geneva, Switzerland, generated 200 petabytes of data (1 petabyte = 10^{15} bytes). That is as much data as the total amount of printed material in the world! In physics, genomics, climate science, marketing, even online gaming and film, the

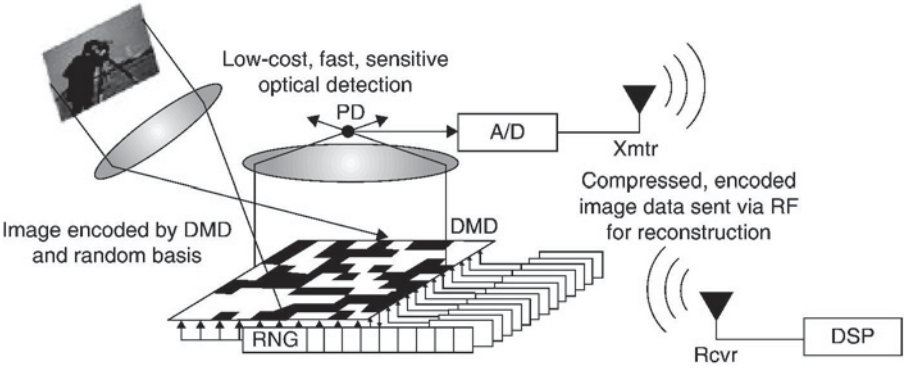


FIGURE I.4: Researchers at Rice University have developed a one-pixel camera based on compressed sensing that randomly alters where the light hitting the single pixel originates from within the camera's field of view as it builds the image. Image courtesy: Rice University.

sizes of datasets are staggering. How to store, transmit, visualize, and process such data is one the great challenges of science.

Probability is being used in a central way for such problems in a new methodology called *compressed sensing*.

In the average hospital, many terabytes (1 terabyte = 10^{12} bytes) of digital magnetic resonance imaging (MRI) data are generated each year. A half-hour MRI scan might collect 100 Mb of data. These data are then compressed to a smaller image, say 5 Mb, with little loss of clarity or detail. Medical and most natural images are compressible since lots of pixels have similar values. Compression algorithms work by essentially representing the image as a sum of simple functions (such as sine waves) and then discarding those terms that have low information content. This is a fundamental idea in signal processing, and essentially what is done when you take a picture on your cell phone and then convert it to a JPEG file for sending to a friend or downloading to the web.

Compressed sensing asks: If the data are ultimately compressible, is it really necessary to acquire all the data in the first place? Can just the final compressed data be what is initially gathered? And the startling answer is that by *randomly* sampling the object of interest, the final image can be reconstructed with similar results as if the object had been fully sampled. Random sampling of MRI scans produces an image of similar quality as when the entire object is scanned. The new technique has reduced MRI scan time to one-seventh the original time, from about half an hour to less than 5 minutes, and shows enormous promise for many other applied areas (Fig. I.4).

I.5 FROM APPLICATION TO THEORY

Having sung the praises of applications and case studies, we come back to the importance of theory.

Probability has been called the science of uncertainty. "Mathematical probability" may seem an oxymoron like jumbo shrimp or civil war. If any discipline can profess a claim of "certainty," surely it is mathematics with its adherence to rigorous proof and timeless results.

One of the great achievements of modern mathematics was putting the study of probability on a solid scientific foundation. This was done in the 1930s, when the Russian mathematician Andrey Nikolaevich Kolmogorov built up probability theory in a rigorous way similarly to how Euclid built up geometry. Much of his work is the material of a graduate-level course, but the basic framework of axiom, definition, theory, and proof sets the framework for the modern treatment of the subject.

One of the joys of learning probability is the compelling imagery we can exploit. Geometers draw circles and squares; probabilists toss coins and roll dice. There is no perfect circle in the physical universe. And the "fair coin" is an idealized model. Yet when you take *real* pennies and toss them repeatedly, the results conform so beautifully to the theory.

In this book, we use the computer package R. R is free software and an interactive computing environment available for download at `http://www.r-project.org/`. If you have never used R before, work through Appendix A: *Getting Started with R*.

Simulation plays a significant role in this book. Simulation is the use of random numbers to generate samples from a random experiment. Today it is a bedrock tool in the sciences and data analysis. Many problems that were for all practical purposes impossible to solve before the computer age are now easily handled with simulation.

There are many compelling reasons for including simulation in a probability course. Simulation helps build invaluable intuition for how random phenomena behave. It will also give you a flexible platform to test how changes in assumptions and parameters can affect outcomes. And the exercise of translating theoretical models into usable simulation code (easy to do in R) will make the subject more concrete and hopefully easier to understand.

And, most importantly, it is fun! Students enjoy the hands-on approach to the subject that simulation offers. This author still gets a thrill from seeing some complex theoretical calculation "magically" verified by a simulation.

To succeed in this subject, read carefully, work through the examples, and do as many problems as you can. But most of all, enjoy the ride!

> The results concerning fluctuations in coin tossing show that widely held beliefs . . . are fallacious. They are so amazing and so at variance with common intuition that even sophisticated colleagues doubted that coins actually misbehave as theory predicts. The record of a simulated experiment is therefore included. . . .
>
> —William Feller, *An Introduction to Probability Theory and Its Applications*

1

FIRST PRINCIPLES

First principles, Clarice. Read Marcus Aurelius. Of each particular thing ask: what is it in itself? What is its nature?

—Hannibal Lecter, *Silence of the Lambs*

1.1 RANDOM EXPERIMENT, SAMPLE SPACE, EVENT

Probability begins with some activity, process, or experiment whose outcome is uncertain. This can be as simple as throwing dice or as complicated as tomorrow's weather.

Given such a "random experiment," the set of all possible outcomes is called the *sample space*. We will use the Greek capital letter Ω (omega) to represent the sample space.

Perhaps the quintessential random experiment is flipping a coin. Suppose a coin is tossed three times. Let H represent heads and T represent tails. The sample space is

$$\Omega = \{\text{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}\},$$

consisting of eight outcomes. The Greek lowercase omega ω will be used to denote these outcomes, the elements of Ω.

Probability: With Applications and R, First Edition. Robert P. Dobrow.

An *event* is a set of outcomes. The event of getting all heads in three coin tosses can be written as

$$A = \{\text{Three heads}\} = \{HHH\}.$$

The event of getting at least two tails is

$$B = \{\text{At least two tails}\} = \{\text{HTT, THT, TTH, TTT}\}.$$

We take probabilities of events. But before learning how to find probabilities, first learn to identify the sample space and relevant event for a given problem.

Example 1.1 The weather forecast for tomorrow says rain. The amount of rainfall can be considered a random experiment. If at most 24 inches of rain will fall, then the sample space is the interval $\Omega = [0, 24]$. The event that we get between 2 and 4 inches of rain is $A = [2, 4]$. ■

Example 1.2 Roll a pair of dice. Find the sample space and identify the event that the sum of the two dice is equal to 7.

The random experiment is rolling two dice. Keeping track of the roll of each die gives the sample space

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

The event is $A = \{\text{Sum is 7}\} = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$. ■

Example 1.3 Yolanda and Zach are running for president of the student association. One thousand students will be voting. We will eventually ask questions like, What is the probability that Yolanda beats Zach by at least 100 votes? But before actually finding this probability, first identify (i) the sample space and (ii) the event that Yolanda beats Zach by at least 100 votes.

(i) The outcome of the vote can be denoted as $(x, 1000-x)$, where x is the number of votes for Yolanda, and $1000 - x$ is the number of votes for Zach. Then the sample space of all voting outcomes is

$$\Omega = \{(0,1000),(1,999),(2,998),\ldots,(999,1),(1000,0)\}.$$

(ii) Let A be the event that Yolanda beats Zach by at least 100 votes. The event A consists of all outcomes in which $x - (1000 - x) \geq 100$, or $550 \leq x \leq 1000$. That is, $A = \{(550,450),(551,449),\ldots,(999,1),(1000,0)\}$. ■

Example 1.4 Joe will continue to flip a coin until heads appears. Identify the sample space and the event that it will take Joe at least three coin flips to get a head.

The sample space is the set of all sequences of coin flips with one head preceded by some number of tails. That is,

$$\Omega = \{\text{H, TH, TTH, TTTH, TTTTH, TTTTTH}, \ldots\}.$$

The desired event is $A = \{\text{TTH, TTTH, TTTTH}, \ldots\}$. Note that in this case both the sample space and the event A are infinite. ■

1.2 WHAT IS A PROBABILITY?

What does it mean to say that *the probability that A occurs* is equal to x?

From a formal, purely mathematical point of view, a probability is a number between 0 and 1 that satisfies certain properties, which we will describe later. From a practical, empirical point of view, a probability matches up with our intuition of the likelihood or "chance" that an event occurs. An event that has probability 0 "never" happens. An event that has probability 1 is "certain" to happen. In repeated coin flips, a coin comes up heads about half the time, and the probability of heads is equal to one-half.

Let A be an event associated with some random experiment. One way to understand the probability of A is to perform the following thought exercise: imagine conducting the experiment over and over, infinitely often, keeping track of how often A occurs. Each experiment is called a *trial.* If the event A occurs when the experiment is performed, that is a *success.* The proportion of successes is the probability of A, written $P(A)$.

This is the *relative frequency* interpretation of probability, which says that the probability of an event is equal to its relative frequency in a large number of trials.

When the weather forecaster tells us that tomorrow there is a 20% chance of rain, we understand that to mean that if we could repeat today's conditions—the air pressure, temperature, wind speed, etc.—over and over again, then 20% of the resulting "tomorrows" will result in rain. Closer to what weather forecasters actually do in coming up with that 20% number, together with using satellite and radar information along with sophisticated computational models, is to go back in the historical record and find other days that match up closely with today's conditions and see what proportion of those days resulted in rain on the following day.

There are definite limitations to constructing a rigorous mathematical theory out of this intuitive and empirical view of probability. One cannot actually repeat an experiment infinitely many times. To define probability carefully, we need to take a formal, axiomatic, mathematical approach. Nevertheless, the relative frequency viewpoint will still be useful in order to gain intuitive understanding. And by the end of the book, we will actually derive the relative frequency viewpoint as a consequence of the mathematical theory.

1.3 PROBABILITY FUNCTION

We assume for the next several chapters that the sample space is *discrete*. This means that the sample space is either finite or countably infinite.

A set is *countably infinite* if the elements of the set can be arranged as a sequence. The natural numbers $1, 2, 3, \ldots$ is the classic example of a countably infinite set. And all countably infinite sets can be put in one-to-one correspondence with the natural numbers.

If the sample space is finite, it can be written as $\Omega = \{\omega_1, \ldots, \omega_k\}$. If the sample space is countably infinite, it can be written as $\Omega = \{\omega_1, \omega_2, \ldots\}$.

The set of all real numbers is an infinite set that is not countably infinite. It is called *uncountable.* An interval of real numbers, such as (0,1), the numbers between 0 and 1, is also uncountable. Probability on uncountable spaces will require differential and integral calculus, and will be discussed in the second half of this book.

A *probability function* assigns numbers between 0 and 1 to events according to three defining properties.

PROBABILITY FUNCTION

Definition 1.1. Given a random experiment with discrete sample space Ω, a *probability function* P is a function on Ω with the following properties:

1. $$P(\omega) \geq 0, \text{ for all } \omega \in \Omega$$

2. $$\sum_{\omega \in \Omega} P(\omega) = 1 \tag{1.1}$$

3. For all events $A \subseteq \Omega$,

$$P(A) = \sum_{\omega \in A} P(\omega). \tag{1.2}$$

You may not be familiar with some of the notations in this definition. The symbol $\in$ means "is an element of." So $\omega \in \Omega$ means ω is an element of Ω. We are also using a generalized Σ-notation in Equation 1.1 and Equation 1.2, writing a condition under the Σ to specify the summation. The notation $\sum_{\omega \in \Omega}$ means that the sum is over all ω that are elements of the sample space, that is, all outcomes in the sample space.

In the case of a finite sample space $\Omega = \{\omega_1, \ldots, \omega_k\}$, Equation 1.1 becomes

$$\sum_{\omega \in \Omega} P(\omega) = P(\omega_1) + \cdots + P(\omega_k) = 1.$$

And in the case of a countably infinite sample space $\Omega = \{\omega_1, \omega_2, \ldots\}$, this gives

$$\sum_{\omega \in \Omega} P(\omega) = P(\omega_1) + P(\omega_2) + \cdots = \sum_{i=1}^{\infty} P(\omega_i) = 1.$$

In simple language, probabilities sum to 1.

The third defining property of a probability function says that the probability of an event is the sum of the probabilities of all the outcomes contained in that event.

We might describe a probability function with a table, function, graph, or qualitative description.

Example 1.5 A type of candy comes in red, yellow, orange, green, and purple colors. Choose a candy at random. The sample space is $\Omega = \{\text{R, Y, O, G, P}\}$. Here are three equivalent ways of describing the probability function corresponding to equally likely outcomes:

1.

R	Y	O	G	P
0.20	0.20	0.20	0.20	0.20

2. $P(\omega) = 1/5, \text{ for all } \omega \in \Omega.$
3. The five colors are equally likely. ■

In the discrete setting, we will often use *probability model* and *probability distribution* interchangeably with probability function. In all cases, to specify a probability function requires identifying (i) the outcomes of the sample space and (ii) the probabilities associated with those outcomes.

Letting H denote heads and T denote tails, an obvious model for a simple coin toss is

$$P(\text{H}) = P(\text{T}) = 0.50.$$

Actually, there is some extremely small, but nonzero, probability that a coin will land on its side. So perhaps a better model would be

$$P(\text{H}) = P(\text{T}) = 0.49999999995 \quad \text{and} \quad P(\text{Side}) = 0.0000000001.$$

Ignoring the possibility of the coin landing on its side, a more general model is

$$P(\text{H}) = p \quad \text{and} \quad P(\text{T}) = 1 - p,$$

where $0 \leq p \leq 1$. If $p = 1/2$, we say the coin is *fair*. If $p \neq 1/2$, we say that the coin is *biased*.

In a mathematical sense, all of these coin tossing models are "correct" in that they are consistent with the definition of what a probability is. However, we might debate

which model most accurately reflects reality and which is most useful for modeling actual coin tosses.

Example 1.6 A college has six majors: biology, geology, physics, dance, art, and music. The proportion of students taking these majors are 20, 20, 5, 10, 10, and 35, respectively. Choose a random student. What is the probability they are a science major?

The random experiment is choosing a student. The sample space is

$$\Omega = \{\text{Bio, Geo, Phy, Dan, Art, Mus}\}.$$

The probability function is given in Table 1.1. The event in question is

$$A = \{\text{Science major}\} = \{\text{Bio, Geo, Phy}\}.$$

Finally,

$$\begin{aligned} P(A) &= P(\{\text{Bio, Geo, Phy}\}) = P(\text{Bio}) + P(\text{Geo}) + P(\text{Phy}) \\ &= 0.20 + 0.20 + 0.05 = 0.45. \end{aligned}$$

■

TABLE 1.1: Probabilities for majors.

Bio	Geo	Phy	Dan	Art	Mus
0.20	0.20	0.05	0.10	0.10	0.35

This example is probably fairly clear and may seem like a lot of work for a simple result. However, when starting out, it is good preparation for the more complicated problems to come to clearly identify the sample space, event and probability model before actually computing the final probability.

Example 1.7 In three coin tosses, what is the probability of getting at least two tails?

Although the probability model here is not explicitly stated, the simplest and most intuitive model for fair coin tosses is that every outcome is equally likely. Since the sample space

$$\Omega = \{\text{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}\}$$

has eight outcomes, the model assigns to each outcome the probability $1/8$.

The event of getting at least two tails can be written as $A = \{\text{HTT, THT, TTH, TTT}\}$. This gives

$$\begin{aligned} P(A) &= P(\{\text{HTT, THT, TTH, TTTT}\}) \\ &= P(\text{HTT}) + P(\text{THT}) + P(\text{TTH}) + P(\text{TTT}) \\ &= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2}. \end{aligned}$$

■

1.4 PROPERTIES OF PROBABILITIES

Events can be combined together to create new events using the connectives "or," "and," and "not." These correspond to the set operations union, intersection, and complement.

For sets $A, B \subseteq \Omega$, the *union* $A \cup B$ is the set of all elements of Ω that are in either A or B or both. The *intersection* AB is the set of all elements of Ω that are in both A and B. (Another common notation for the intersection of two events is $A \cap B$.) The *complement* A^c is the set of all elements of Ω that are not in A.

In probability word problems, descriptive phrases are typically used rather than set notation. See Table 1.2 for some equivalences.

TABLE 1.2: Events and sets.

Description	Set notation
Either A or B or both occur	$A \cup B$
A and B	AB
Not A	A^c
A implies B	$A \subseteq B$
A but not B	AB^c
Neither A nor B	$A^c B^c$
At least one of the two events occurs	$A \cup B$
At most one of the two events occurs	$(AB)^c = A^c \cup B^c$

A Venn diagram is a useful tool for working with events and subsets. A rectangular box denotes the sample space Ω, and circles are used to denote events. See Figure 1.1 for examples of Venn diagrams for the most common combined events obtained from two events A and B.

One of the most basic, and important, properties of a probability function is the simple addition rule for mutually exclusive events. We say that two events are *mutually exclusive,* or *disjoint,* if they have no outcomes in common. That is, A and B are mutually exclusive if $AB = \emptyset$, the empty set.

ADDITION RULE FOR MUTUALLY EXCLUSIVE EVENTS

If A and B are mutually exclusive events, then

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B).$$

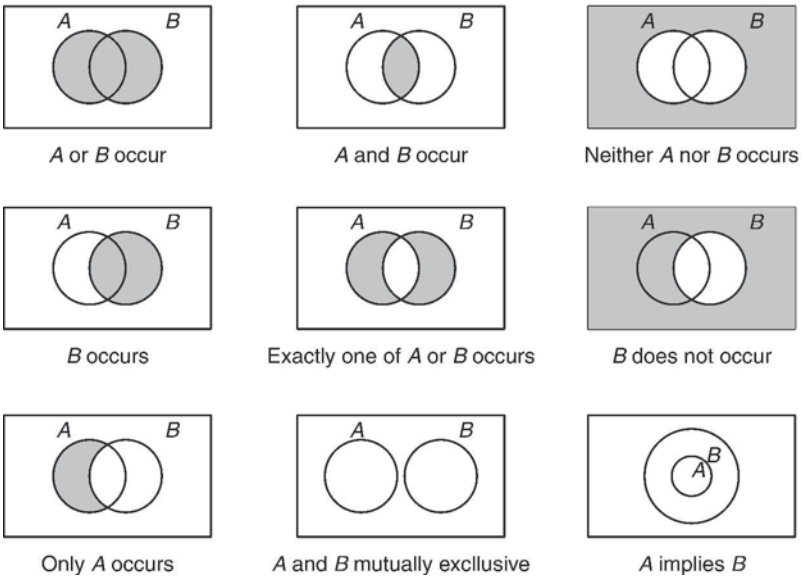


FIGURE 1.1: Venn diagrams.

The addition rule is a consequence of the third defining property of a probability function. We have that

$$\begin{aligned} P(A \text{ or } B) = P(A \cup B) &= \sum_{\omega \in A \cup B} P(\omega) \\ &= \sum_{\omega \in A} P(\omega) + \sum_{\omega \in B} P(\omega) \\ &= P(A) + P(B), \end{aligned}$$

where the third equality follows since the events are disjoint. The addition rule for mutually exclusive events extends to more than two events.

ADDITION RULE FOR MUTUALLY EXCLUSIVE EVENTS

Suppose $A_1, A_2, \ldots$ is a sequence of pairwise mutually exclusive events. That is, A_i and A_j are mutually exclusive for all $i \neq j$. Then

$$P(\text{at least one of the } A_i\text{'s occurs}) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$