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.

"9781118241257c01" — 2013/9/10 — 16:19 — page 2 — #2

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 $\Omega = \{$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. $P(\omega) = 1/5$, 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 $0 \le p \le 1$. 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 $A = \{$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^cB^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

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

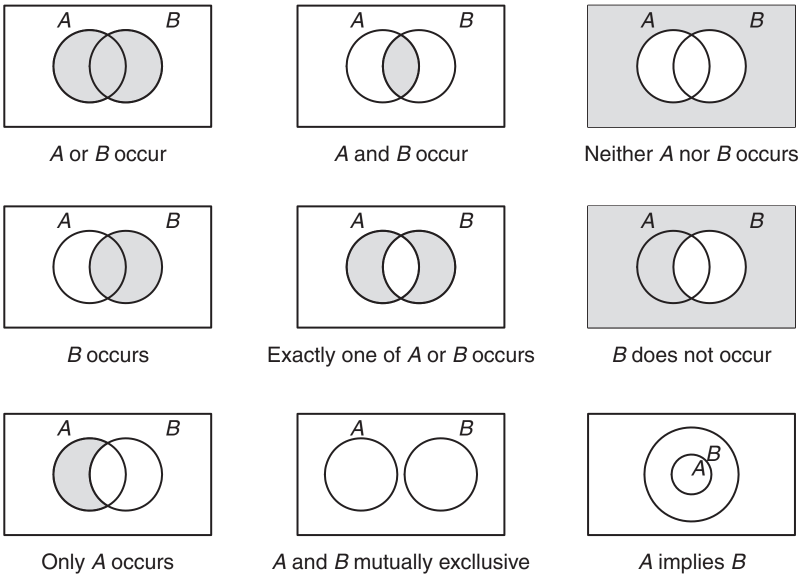

FIGURE 1.1: Venn diagrams.

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

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

Here we highlight other key properties that are consequences of the defining properties of a probability function and the addition rule for disjoint events.

PROPERTIES OF PROBABILITIES

1. If A implies B, that is, if $A \subseteq B$, then $P(A) \leq P(B)$.
2. $P(A \text{ does not occur}) = P(A^c) = 1 - P(A)$.
3. For all events A and B,

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(AB). \tag{1.3}$$

Each property is derived next.

1. As $A \subseteq B$, write B as the disjoint union of A and BA^c. By the addition rule for disjoint events,

$$P(B) = P(A \cup BA^c) = P(A) + P(BA^c) \geq P(A),$$

since probabilities are nonnegative.
2. The sample space Ω can be written as the disjoint union of any event A and its complement A^c. Thus,

$$1 = P(\Omega) = P\left(A \cup A^c\right) = P(A) + P(A^c).$$

Rearranging gives the result.
3. Write $A \cup B$ as the disjoint union of A and A^cB. Also write B as the disjoint union of AB and A^cB. Then $P(B) = P(AB) + P(A^cB)$ and thus

$$P(A \cup B) = P(A) + P(BA^c) = P(A) + P(B) - P(AB).$$

Observe that the addition rule for mutually exclusive events follows from Property 3 since if A and B are disjoint, then $P(AB) = 0$.

Example 1.8 In a city, 75% of the population have brown hair, 40% have brown eyes, and 25% have both brown hair and brown eyes. A person is chosen at random. What is the probability that they

1. have brown eyes or brown hair?
2. have neither brown eyes nor brown hair?

To gain intuition, draw a Venn diagram, as in Figure 1.2. Let H be the event of having brown hair; let E denote brown eyes.

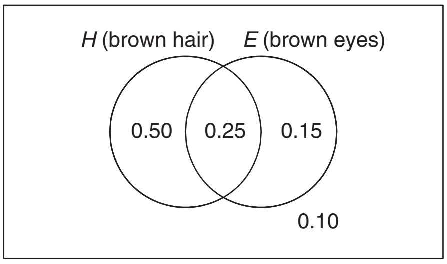

FIGURE 1.2: Venn diagram.

1. The probability of having brown eyes or brown hair is

$$P(E \text{ or } H) = P(E) + P(H) - P(EH) = 0.75 + 0.40 - 0.25 = 0.90.$$

Notice that E and H are not mutually exclusive. If we made a mistake and used the simple addition rule $P(E \text{ or } H) = P(E) + P(H)$, we would mistakenly get $0.70 + 0.40 = 1.10 > 1$.

2. The complement of having neither brown eyes nor brown hair is having brown eyes or brown hair. Thus

$$P(E^c H^c) = P((E \text{ or } H)^c) = 1 - P(E \text{ or } H) = 1 - 0.90 = 0.10. \quad \blacksquare$$

1.5 EQUALLY LIKELY OUTCOMES

The simplest probability model for a finite sample space is that all outcomes are equally likely. If Ω has k elements, then the probability of each outcome is $1/k$, since probabilities sum to 1. That is, $P(\omega) = 1/k$, for all $\omega \in \Omega$.

Computing probabilities for equally likely outcomes takes a fairly simple form. Suppose A is an event with s elements, with $s \leq k$. Since $P(A)$ is the sum of the probabilities of all the outcomes contained in A,

$$P(A) = \sum_{\omega \in A} P(\omega) = \sum_{\omega \in A} \frac{1}{k} = \frac{s}{k} = \frac{\text{Number of elements of } A}{\text{Number of elements of } \Omega}.$$

Probability with equally likely outcomes reduces to *counting*.

Example 1.9 A *palindrome* is a word that reads the same forward or backward. Examples include mom, civic, and rotator. Pick a three-letter "word" at random choosing from D, O, or G for each letter. What is the probability that the resulting word is a palindrome?

There are 27 possible words (three possibilities for each letter). List them and count the palindromes: DDD, OOO, GGG, DOD, DGD, ODO, OGO, GDG, and GGG. The probability of getting a palindrome is $9/27 = 1/3$. ■

Example 1.10 A bowl has r red balls and b blue balls. What is the probability of selecting a red ball?

The sample space consists of $r+b$ balls. The event $A = \{\text{Red ball}\}$ has r elements. Therefore, $P(A) = r/(r + b)$. ■

Example 1.11 A field has 36 trees planted in six equally spaced rows and columns as in Figure 1.3. A tree is chosen at random so that each tree is equally likely to be chosen. What is the probability that the tree that is picked is within the wedge at the lower left-hand corner of the field?

There are eight points in the wedge, so the desired probability is $8/36 = 2/9$. ■

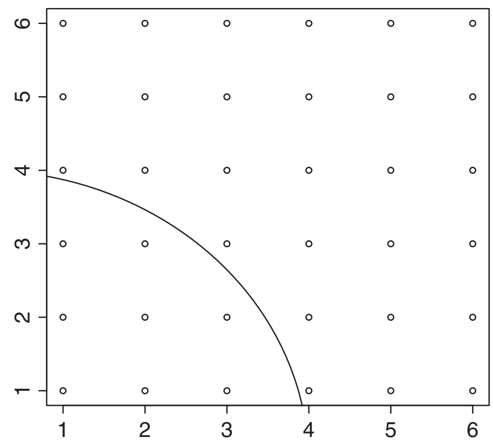

FIGURE 1.3: Trees in a field.

A model for equally likely outcomes assumes a finite sample space. Interestingly, it is impossible to have a probability model of equally likely outcomes on an infinite sample space. To see why, suppose $\Omega = \{\omega_1, \omega_2, \ldots\}$ and $P(\omega_i) = c$ for all i, where c is a nonzero constant. Then summing the probabilities gives

$$\sum_{i=1}^{\infty} P(\omega_i) = \sum_{i=1}^{\infty} c = \infty \neq 1.$$

While equally likely outcomes are not possible in the infinite case, there are many ways to assign probabilities for an infinite sample space where outcomes are not equally likely. For instance, let $\Omega = \{\omega_1, \omega_2, \ldots\}$ with $P(\omega_i) = (1/2)^i$, for $i = 1, 2, \ldots$ Then

$$\sum_{i=1}^{\infty} P(\omega_i) = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^i = \left(\frac{1}{2}\right) \frac{1}{1-(1/2)} = 1.$$

Since counting plays a fundamental role in probability when outcomes are equally likely, we introduce some basic counting principles.

1.6 COUNTING I

Counting sets is sometimes not as easy as 1, 2, 3. . . . But a basic counting principle known as the *multiplication principle* allows for tackling a wide range of problems.

MULTIPLICATION PRINCIPLE

If there are m ways for one thing to happen, and n ways for a second thing to happen, there are $m \times n$ ways for both things to happen.

More generally—and more formally—consider an n-element sequence $(a_1, a_2, \ldots, a_n)$. If there are k_1 possible values for the first element, k_2 possible values for the second element, . . ., and k_n possible values for the nth element, there are $k_1 \times k_2 \times \cdots \times k_n$ possible sequences.

For instance, in tossing a coin three times, there are $2 \times 2 \times 2 = 2^3 = 8$ possible outcomes. Rolling a die four times gives $6 \times 6 \times 6 \times 6 = 6^4 = 1296$ possible rolls.

Example 1.12 License plates in Minnesota are issued with three letters from A to Z followed by three digits from 0 to 9. If each license plate is equally likely, what is the probability that a random license plate starts with G-Z-N?

The solution will be equal to the number of license plates that start with G-Z-N divided by the total number of license plates. By the multiplication principle, there are $26 \times 26 \times 26 \times 10 \times 10 \times 10 = 17{,}576{,}000$ possible license plates.

For the number of plates that start with G-Z-N, think of a 6-element plate of the form G-Z-N-_-_-_. For the three blanks, there are $10 \times 10 \times 10$ possibilities. Thus the desired probability is $10^3/(26^3 \times 10^3) = 1/26^3 = 0.0000569$.

Aside: The author has the sense that everywhere he goes he sees license plates with the same initial three letters as his own car. This can be explained either by an interesting psychological phenomenon involving coincidences or by the fact that license plates in Minnesota are not, in fact, equally likely. ■

Example 1.13 A DNA strand is a long polymer string made up of four nucleotides—adenine, cytosine, guanine, and thymine. It can be thought of as a sequence of As, Cs, Gs, and Ts. DNA is structured as a double helix with two paired strands running in opposite directions on the chromosome. Nucleotides always pair the same way: A with T and C with G. A *palindromic sequence* is equal to its "reverse complement." For instance, the sequences ACGT and GTTAGCTAAC are palindromic sequences, but ACCA is not. Such sequences play a significant role in molecular biology.

Suppose the nucleotides on a DNA strand of length six are generated in such a way so that all strands are equally likely. What is the probability that the DNA sequence is a palindromic sequence?

By the multiplication principle, the number of DNA strands is 4^6 since there are four possibilities for each site. A palindromic sequence of length six is completely determined by the first three sites. There are 4^3 palindromic sequences. The desired probability is $4^3/4^6 = 1/64$. ■

Example 1.14 Mark is taking four final exams next week. His studying was erratic and all scores A, B, C, D, and F are equally likely for each exam. What is the probability that Mark will get at least one A?

Take complements. The complementary event of getting at least one A is getting no As. Since outcomes are equally likely, by the multiplication principle there are 4^4 exam outcomes with no As (four grade choices for each of four exams). And there are 5^4 possible outcomes in all. The desired probability is $1 - 4^4/5^4 = 0.5904$. ■

Given a set of distinct objects, a *permutation* is an ordering of the elements of the set. For the set $\{a, b, c\}$, there are six permutations:

$$(a, b, c), (a, c, b), (b, a, c), (b, c, a), (c, a, b), \text{ and } (c, b, a).$$

How many permutations are there of an n-element set? There are n possibilities for the first element of the permutation, $n - 1$ for the second, and so on. The result follows by the multiplication principle.

COUNTING PERMUTATIONS

There are $n \times (n-1) \times \cdots \times 1 = n!$ permutations of an n-element set.

The factorial function $n!$ grows very large very fast. In a classroom of 10 people with 10 chairs, there are $10! = 3{,}628{,}800$ ways to seat the students. There are $52! \approx 8 \times 10^{67}$ orderings of a standard deck of cards, which is "almost" as big as the number of atoms in the observable universe, which is estimated to be about 10^{80}.

Functions of the form c^n, where c is a constant, are said to exhibit *exponential* growth. The factorial function $n!$ grows like n^n, which is sometimes called *super-exponential* growth.

Example 1.15 Bob has three bookshelves in his office and 15 books—5 are math books and 10 are novels. If each shelf holds exactly five books and books are placed randomly on the shelves (all orderings are equally likely), what is the probability that the bottom shelf contains all the math books?

There are $15!$ ways to permute all the books on the shelves. There are $5!$ ways to put the math books on the bottom shelf and $10!$ ways to put the remaining novels on the other two shelves. Thus by the multiplication principle, the desired probability is $(5!10!)/15! = 1/3003 = 0.000333$. ■

Example 1.16 A bag contains six Scrabble tiles with the letters A-D-M-N-O-R. You reach into the bag and take out tiles one at a time. What is the probability that you will spell the word R-A-N-D-O-M?

How many possible words can be formed? All the letters are distinct and a "word" is a permutation of the set of six letters. There are $6! = 720$ possible words. Only one of them spells R-A-N-D-O-M, so the desired probability is $1/720 = 0.001389$. ■

Example 1.17 **Scrabble continued.** Change the previous problem. After you pick a tile from the bag, write down that letter and then return the tile to the bag. So every time you reach into the bag it contains the six original letters.

Now there are $6 \times \cdots \times 6 = 6^6 = 46,656$ possible words, and the desired probability is $1/46,656 = 0.0000214$. ■

SAMPLING WITH AND WITHOUT REPLACEMENT

The last examples highlight two different sampling methods called *sampling without replacement* and *sampling with replacement.* When sampling with replacement, a unit that is selected from a population is returned to the population before another unit is selected. When sampling without replacement, the unit is not returned to the population after being selected. When solving a probability problem involving sampling (such as selecting cards or picking balls from urns), make sure you know the sampling method before computing the related probability.

Example 1.18 When national polling organizations conduct nationwide surveys, they often select about 1000 people sampling without replacement. If N is the number of people in a target population, then by the multiplication principle there are $N \times (N-1) \times (N-2) \times \cdots \times (N-999)$ possible ordered samples. For national polls in the United States, where N, the number of people age 18 or over, is about 250 million, that gives about $(250{,}000{,}000)^{1000}$ possible ordered samples, which is a mind-boggling 2.5 with 8000 zeros after it. ■

1.7 PROBLEM-SOLVING STRATEGIES: COMPLEMENTS, INCLUSION–EXCLUSION

Consider a sequence of events $A_1, A_2, \ldots$. In this section, we consider strategies to find the probability that *at least* one of the events occurs, which is the probability of the union $\cup_i A_i$.

Sometimes the complement of an event can be easier to work with than the event itself. The complement of the event that at least one of the A_is occurs is the event that none of the A_is occur, which is the intersection $\cap_i A_i^c$.

Check with a Venn diagram (and if you are comfortable working with sets prove it yourself) that

$$(A \cup B)^c = A^c B^c \quad \text{and} \quad (AB)^c = A^c \cup B^c.$$

Complements turn unions into intersections, and vice versa. These set-theoretic results are known as DeMorgan's laws. The results extend to infinite sequences. Given events $A_1, A_2, \ldots$,

$$\left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_1^c \quad \text{and} \quad \left(\bigcap_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c.$$

Example 1.19 Four dice are rolled. Find the probabilty of getting at least one 6.

The sample space is the set of all outcomes of four dice rolls

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

By the multiplication principle, there are $6^4 = 1296$ elements. If the dice are fair, each of these outcomes is equally likely. It is not obvious, without some new tools, how to count the number of outcomes that have at least one 6.

Let A be the event of getting at least one 6. Then the complement A^c is the event of getting no sixes in four rolls. An outcome has no sixes if the dice rolls a 1, 2, 3, 4, or 5 on every roll. By the multiplication principle, there are $5^4 = 625$ possibilities. Thus $P(A^c) = 5^4/6^4 = 625/1296$ and

$$P(A) = 1 - P(A^c) = 1 - \frac{625}{1296} = 0.5177. \qquad \blacksquare$$

Recall the formula in Equation 1.3 for the probability of a union of two events. We generalize for three or more events using the principle of inclusion–exclusion.

Proposition 1.2. For events A, B, and C,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC).$$

Since we first *include* the sets, then *exclude* the pairwise intersections, then *include* the triple intersection, this is called the *inclusion–exclusion principle.* The proof is intuitive with the help of a Venn diagram. Write

$$A \cup B \cup C = \left[A \cup B\right] \cup \left[C(AC \cup BC)^c\right].$$

The bracketed sets $A \cup B$ and $C(AC \cup BC)^c$ are disjoint. Thus,

$$\begin{aligned} P(A \cup B \cup C) &= P(A \cup B) + P\left(C(AC \cup BC)^c\right) \\ &= P(A) + P(B) - P(AB) + P\left(C(AC \cup BC)^c\right). \end{aligned} \tag{1.4}$$

Write C as the disjoint union

$$C = [C(AC \cup BC)] \cup [C(AC \cup BC)^c] = [AC \cup BC] \cup [C(AC \cup BC)^c].$$

This gives

$$P(C(AC \cup BC)^c) = P(C) - P(AC \cup BC).$$

Together with Equation 1.4,

$$\begin{aligned} P(A \cup B \cup C) = P(A) &+ P(B) - P(AB) \\ &+ P(C) - [P(AC) + P(BC) - P(ABC)]. \end{aligned}$$

Extending further to more than three events gives the general principle of inclusion–exclusion. We will not prove it, but if you know how to use mathematical induction, give it a try.

INCLUSION–EXCLUSION

Given events $A_1, \ldots, A_n$, the probability that at least one event occurs is

$$\begin{aligned} P(A_1 \cup \cdots \cup A_n) &= \sum_i P(A_i) - \sum_{i<j} P(A_i A_j) \\ &+ \sum_{i<j<k} P(A_i A_j A_k) - \cdots + (-1)^{n+1} P(A_1 \cdots A_n). \end{aligned}$$

Example 1.20 An integer is drawn uniformly at random from $\{1, \ldots, 1000\}$ such that each number is equally likely. What is the probability that the number drawn is divisible by 3, 5, or 6?

Let D_3, D_5, and D_6 denote the events that the number drawn is divisible by 3, 5, and 6, respectively. The problem asks for $P(D_3 \cup D_5 \cup D_6)$. By inclusion–exclusion,

$$\begin{aligned} P(D_3 \cup D_5 \cup D_6) = P(D_3) + P(D_5) + P(D_6) - P(D_3 D_5) - P(D_3 D_6) \\ - P(D_5 D_6) + P(D_3 D_5 D_6). \end{aligned}$$

Let $\lfloor x \rfloor$ denote the integer part of x. There are $\lfloor 1000/x \rfloor$ numbers from 1 to 1000 that are divisible by x. Since all selections are equally likely,

$$P(D_3) = \lfloor 1000/3 \rfloor / 1000 = 0.333.$$

$$P(D_5) = \lfloor 1000/5 \rfloor / 1000 = 0.20.$$

$$P(D_6) = \lfloor 1000/6 \rfloor / 1000 = 0.166.$$

A number is divisible by 3 and 5 if and only if it is divisible by 15. Thus, $D_3D_5 = D_{15}$. If a number is divisible by 6, it is also divisible by 3, so $D_3D_6 = D_6$. Also, $D_5D_6 = D_{30}$. And $D_3D_5D_6 = D_{30}$. This gives

$$P(D_3D_5) = \lfloor 1000/15 \rfloor / 1000 = 0.066.$$

$$P(D_3D_6) = 0.166.$$

$$P(D_5D_6) = \lfloor 1000/30 \rfloor / 1000 = 0.033.$$

$$P(D_3D_5D_6) = 0.033.$$

Putting it all together gives

$$P(D_3 \cup D_5 \cup D_6)$$
$$= 0.333 + 0.2 + 0.166 - 0.066 - 0.166 - 0.033 + 0.033 = 0.467. \quad \blacksquare$$

We have presented two different ways of computing the probability that at least one of several events occurs: (i) a "back-door" approach of taking complements and working with the resulting "and" probabilities; and (ii) a direct "frontal-attack" by inclusion-exclusion. Here is a third way, which illustrates decomposing an event into a union of mutually exclusive subsets.

Example 1.21 Consider a random experiment that has k equally likely outcomes, one of which we call *success*. Repeat the experiment n times. Let A be the event that at least one of the n outcomes is a success. For instance, consider rolling a die 10 times, where success means rolling a three. Here $n = 10$, $k = 6$, and A is the event of rolling at least one 3.

Define a sequence of events $A_1, \ldots, A_n$, where A_i is the event that the ith trial is a success. Then $A = A_1 \cup \cdots \cup A_n$ and $P(A) = P(A_1 \cup \cdots \cup A_n)$. We cannot use the addition rule on this probability since the A_is are not mutually exclusive.

To define a sequence of mutually exclusive events, let B_i be the event that the *first* success occurs on the ith trial. Then the B_is are mutually exclusive. Furthermore,

$$B_1 \cup \cdots \cup B_n = A_1 \cup \cdots \cup A_n = A.$$

Thus,

$$P(A) = P(B_1 \cup \cdots \cup B_n) = P(B_1) + \cdots + P(B_n).$$

To find $P(B_i)$, observe that if the first success occurs on the ith trial, then the first $i-1$ trials are necessarily not successes and the ith trial is a success. There are $k-1$ possible outcomes for each of the first $i-1$ trials, one outcome for the ith trial, and k possible outcomes for each of the remaining $n-i$ trials. By the multiplication principle, there are $(k-1)^{i-1}k^{n-i}$ outcomes where the first success occurs on the ith trial. And there are k^n possible outcomes in all. Thus,

$$P(B_i) = \frac{(k-1)^{i-1}k^{n-i}}{k^n} = \frac{1}{k}\left(\frac{k-1}{k}\right)^{i-1} = \frac{1}{k}\left(1 - \frac{1}{k}\right)^{i-1},$$

for $i = 1, \ldots, n$. The desired probability is

$$\begin{aligned} P(A) = P(B_1) + \cdots + P(B_n) &= \sum_{i=1}^{n} \frac{1}{k}\left(1 - \frac{1}{k}\right)^{i-1} \\ &= \frac{1}{k}\left(\frac{1-(1-1/k)^n}{1-(1-1/k)}\right) \\ &= 1 - \left(1 - \frac{1}{k}\right)^n. \end{aligned}$$

For instance, the probability of rolling at least one 3 in 10 rolls of a die is

$$1 - \left(1 - \frac{1}{6}\right)^{10} = 1 - \left(\frac{5}{6}\right)^{10} = 0.8385.$$ ■

1.8 RANDOM VARIABLES

Often the outcomes of a random experiment take on numerical values. For instance, we might be interested in how many heads occur in three coin tosses. Let X be the number of heads. Then X is equal to 0, 1, 2, or 3, depending on the outcome of the coin tosses. The object X is called a *random variable.* The possible *values* of X are 0, 1, 2, and 3.

RANDOM VARIABLE

A random variable assigns numerical values to the outcomes of a random experiment.

Random variables are enormously useful and allow us to use algebraic expressions, equalities, and inequalities when manipulating events. In many of the previous examples, we have been working with random variables without using the name, for example, the number of threes in rolls of a die, the number of votes received, the number of palindromes, the number of heads in repeated coin tosses.

Example 1.22 In tossing three coins, let X be the number of heads. Then the event of getting two heads can be written as $\{X = 2\}$. The probability of getting two heads is thus

$$P(X = 2) = P(\{\text{HHT, HTH, THH}\}) = \frac{3}{8}. \qquad \blacksquare$$

Example 1.23 If we throw two dice, what is the probability that the sum of the dice is greater than four?

We can, of course, find the probability by direct counting. But we will use random variables. Let Y be the sum of two dice rolls. Then Y is a random variable whose possible values are $2, 3, \ldots, 12$. The event that the sum is greater than 4 can be written as $\{Y > 4\}$. Observe that the complementary event is $\{Y \le 3\}$. By taking complements,

$$\begin{aligned} P(Y > 4) &= 1 - P(Y \le 3) \\ &= P(Y = 2 \text{ or } Y = 3) = P(Y = 2) + P(Y = 3) \\ &= P(\{(1, 1)\}) + P(\{(1, 2), (2, 1)\}) = \frac{1}{36} + \frac{2}{36} = \frac{1}{12}. \qquad \blacksquare \end{aligned}$$

Example 1.24 Recall Example 1.3. One thousand students are voting. Suppose the number of votes that Yolanda receives is equally likely to be any number from 0 to 1000. What is the probability that Yolanda beats Zach by at least 100 votes?

We approach the problem using random variables. Let Y be the number of votes for Yolanda. Let Z be the number of votes for Zach. Then the total number of votes is $Y + Z = 1000$. Thus, $Z = 1000 - Y$. The event that Yolanda beats Zach by at least 100 votes is $\{Y - Z \ge 100\} = \{Y - (1000 - Y) \ge 100\} = \{2Y \ge 1100\} = \{Y \ge 550\}$. The desired probability is

$$P(Y - Z \ge 100) = P(Y \ge 550) = 451/1001,$$

since there are 1001 possible votes for Yolanda and 451 of them are greater than 550.

$\blacksquare$

If a random variable X takes values in a finite set all of whose elements are equally likely, we say that X is *uniformly distributed* on that set.

UNIFORM RANDOM VARIABLE

Let $S = \{s_1, \ldots, s_k\}$ be a finite set. A random variable X is *uniformly distributed on* S if

$$P(X = s_i) = \frac{1}{k}, \quad \text{for } i = 1, \ldots, k.$$

We write $X \sim \text{Unif}(S)$.

Example 1.25 Rachel picks an integer "at random" between 1 and 50. (i) Find the probability that she picks 13. (ii) Find the probability that her number is between 10 and 20. (iii) Find the probability that her number is prime.

Let X be Rachel's number. Then X is uniformly distributed on $\{1, \ldots, 50\}$.

(i) The probability that Rachel picks 13 is $P(X = 13) = 1/50$.

(ii) There are 11 numbers between 10 and 20 (including 10 and 20). The desired probability is

$$P(10 \leq X \leq 20) = \frac{11}{50} = 0.22.$$

(iii) One counts 15 prime numbers between 1 and 50. Thus,

$$P(X \text{ is prime}) = \frac{15}{50} = 0.3. \qquad \blacksquare$$

We write $\{X = 2\}$ for the event that the random variable takes the value 2. More generally, we write $\{X = x\}$ for the event that the random variable X takes the value x, where x is a specific number. The difference between the uppercase X (a random variable) and the lowercase x (a number) can be confusing but is extremely important to clarify.

Example 1.26 Roll a pair of dice. What is the probability of getting a 4? Of getting each number between 2 and 12?

Assuming the dice fair, each die number is equally likely. There are six possibilities for the first roll, six possibilities for the second roll, so $6 \times 6 = 36$ possible rolls. We thus assign the probability of 1/36 to each dice pair. Let X be the sum of the two dice. Then

$$\begin{aligned} P(X = 4) &= P(\{(1, 3), (3, 1), (2, 2)\}) \\ &= P((1, 3)) + P((3, 1)) + P((2, 2)) = 3\left(\frac{1}{36}\right) = \frac{1}{12}. \end{aligned}$$

Consider $P(X = x)$ for $x = 2, 3, \ldots, 12$. By counting all the possible combinations, verify the probabilities in Table 1.3.

TABLE 1.3: Probability distribution for the sum of two dice.

x	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Observe that while the outcomes of each individual die are equally likely, the values of the *sum* of two dice are not. ■

1.9 A CLOSER LOOK AT RANDOM VARIABLES

> While writing my book [Stochastic Processes] I had an argument with Feller. He asserted that everyone said "random variable" and I asserted that everyone said "chance variable." We obviously had to use the same name in our books, so we decided the issue by a [random] procedure. That is, we tossed for it and he won.
>
> —Joe Doob, quoted in *Statistical Science*

Random variables are central objects in probability. The name can be confusing since they are really neither "random" nor a "variable" in the way that that word is used in algebra or calculus. A random variable is actually a *function*, a function whose domain is the sample space.

A random variable assigns every outcome of the sample space a real number. Consider the three coins example, letting X be the number of heads in three coin tosses. Depending upon the outcome of the experiment, X takes on different values. To emphasize the dependency of X on the outcome ω, we can write $X(\omega)$, rather than just X. In particular,

$$X(\omega) = \begin{cases} 0, & \text{if } \omega = \text{TTT} \\ 1, & \text{if } \omega = \text{HTT, THT, or TTH} \\ 2, & \text{if } \omega = \text{HHT, HTH, or THH} \\ 3, & \text{if } \omega = \text{HHH}. \end{cases}$$

The probability of getting exactly two heads is written as $P(X = 2)$, which is shorthand for $P(\{\omega : X(\omega) = 2\})$.

You may be unfamiliar with this last notation, used for describing sets. The notation $\{\omega : \text{Property}\}$ describes the set of all ω that satisfies some property. So $\{\omega : X(\omega) = 2\}$ is the set of all ω with the property that $X(\omega) = 2$. That is, the set of all outcomes that result in exactly two heads, which is $\{HHT, HTH, THH\}$.

Similarly, the probability of getting at most one head in three coin tosses is

$$P(X \leq 1) = P(\{\omega : X(\omega) \leq 1\}) = P(\{\text{TTT, HTT, THT, TTH}\}).$$

Because of simplicity and ease of notation, authors typically use the shorthand X in writing random variables instead of the more verbose $X(\omega)$.

1.10 A FIRST LOOK AT SIMULATION

Using random numbers on a computer to simulate probabilities is called the Monte Carlo method. Today, Monte Carlo tools are used extensively in statistics, physics, engineering, and across many disciplines. The name was coined in the 1940s by mathematicians John von Neumann and Stanislaw Ulam working on the Manhattan Project. It was named after the famous Monte Carlo casino in Monaco.

Ulam's description of his inspiration to use random numbers to simulate complicated problems in physics is quoted in Eckhardt (1987):

> The first thoughts and attempts I made to practice [the Monte Carlo method] were suggested by a question which occurred to me in 1946 as I was convalescing from an illness and playing solitaires.
>
> The question was what are the chances that a Canfield solitaire laid out with 52 cards will come out successfully? After spending a lot of time trying to estimate them by pure combinatorial calculations, I wondered whether a more practical method than "abstract thinking" might not be to lay it out say one hundred times and simply observe and count the number of successful plays. This was already possible to envisage with the beginning of the new era of fast computers, and I immediately thought of problems of neutron diffusion and other questions of mathematical physics, and more generally how to change processes described by certain differential equations into an equivalent form interpretable as a succession of random operations. Later [in 1946], I described the idea to John von Neumann, and we began to plan actual calculations.

The Monte Carlo simulation approach is based on the relative frequency model for probabilities. Given a random experiment and some event A, the probability $P(A)$ is estimated by repeating the random experiment many times and computing the proportion of times that A occurs.

More formally, define a sequence $X_1, X_2, \ldots$, where

$$X_k = \begin{cases} 1, & \text{if } A \text{ occurs on the } k\text{th trial} \\ 0, & \text{if } A \text{ does not occur on the } k\text{th trial,} \end{cases}$$

for $k = 1, 2, \ldots$. Then

$$\frac{X_1 + \cdots + X_n}{n}$$

is the proportion of times in which A occurs in n trials. For large n, the Monte Carlo method estimates $P(A)$ by

$$P(A) \approx \frac{X_1 + \cdots + X_n}{n}. \tag{1.5}$$

MONTE CARLO SIMULATION

Implementing a Monte Carlo simulation of $P(A)$ requires three steps:

1. **Simulate a trial.** Model, or translate, the random experiment using random numbers on the computer. One iteration of the experiment is called a "trial."
2. **Determine success.** Based on the outcome of the trial, determine whether or not the event A occurrs. If yes, call that a "success."
3. **Replication.** Repeat the aforementioned two steps many times. The proportion of successful trials is the simulated estimate of $P(A)$.

Monte Carlo simulation is intuitive and matches up with our sense of how probabilities "should" behave. We give a theoretical justification for the method and Equation 1.5 in Chapter 9, where we study limit theorems and the law of large numbers.

Here is a most simple, even trivial, starting example. Consider simulating the probability that an ideal fair coin comes up heads. One could do a *physical* simulation by just flipping a coin many times and taking the proportion of heads to estimate $P(\text{Heads})$.

Using a computer, choose the number of trials n (the larger the better) and type the R command

```
> sample(0:1,n,replace=T)
```

The command samples with replacement from $\{0, 1\}$ n times such that outcomes are equally likely. Let 0 represent tails and 1 represent heads. The output is a sequence of n ones and zeros corresponding to heads and tails. The average, or mean, of the list is precisely the proportion of ones. To simulate $P(\text{Heads})$, type

```
> mean(sample(0:1,n,replace=T))
```

Repeat the command several times (use the up arrow key). These give repeated Monte Carlo estimates of the desired probability. Observe the accuracy in the estimate with one million trials:

```
> mean(sample(0:1,1000000,replace=T))
[1] 0.500376
> mean(sample(0:1,1000000,replace=T))
[1] 0.499869
> mean(sample(0:1,1000000,replace=T))
```

```
[1] 0.498946
> mean(sample(0:1,1000000,replace=T))
[1] 0.500115
```

The R script **CoinFlip.R** simulates a familiar probability—the probability of getting three heads in three coin tosses.

R : SIMULATING THE PROBABILITY OF THREE HEADS IN THREE COIN TOSSES

```
# CoinFlip.R

# Trial
> trial <- sample(0:1,3,replace=TRUE)

# Success
> if (sum(trial)==3) 1 else 0

# Replication
> n <- 10000   # Number of repetitions
> simlist <- numeric(n) # Initialize vector
> for (i in 1:n) {
    trial <- sample(0:1, 3, replace=TRUE)
    success <- if (sum(trial)==3) 1 else 0
    simlist[i] <- success }
> mean(simlist) # Proportion of trials with 3 heads
[1] 0.1231
```

The script is divided into three parts to illustrate (i) coding the trial, (ii) determining success, and (iii) implementing the replication.

To simulate three coin flips, use the `sample` command. Again letting 1 represent heads and 0 represent tails, the command

```
> trial <- sample(0:1, 3, replace=TRUE)
```

chooses a head or tails three times. The three results are stored as a three-element list (called a vector in R) in the variable `trial`.

After flipping three coins, the routine must decide whether or not they are all heads. This is done by summing the outcomes. The sum will equal three if and only if all flips are heads. This is checked with the command

```
> if (sum(trial)==3) 1 else 0
```

which returns a 1 for success, and 0, otherwise.

For the actual simulation, the commands are repeated n times in a loop. The output from each trial is stored in the vector `simlist`. This vector will consist of n ones and zeros corresponding to success or failure for each trial, where success is flipping three heads.

Finally, after repeating n trials, we find the proportion of successes in all the trials, which is the proportion of ones in `simlist`. Given a list of zeros and ones, the average, or mean, of the list is precisely the proportion of ones in the list. The command `mean(simlist)` finds this average giving the simulated probability of getting three heads.

Run the script file and see that the resulting estimate is fairly close to the exact solution $1/8 = 0.125$. Increase n to 100,000 or even a million to get more precise estimates.

The script **Divisible356.R** simulates the divisibility problem in Example 1.20 that a random integer from $\{1, \ldots, 1000\}$ is divisible by 3, 5, or 6. The problem, and the resulting code, is more complex.

The function `simdivis()` simulates one trial. Inside the function, the expression `num%%x==0` checks whether `num` is divisible by x. The `if` statement checks whether `num` is divisible by 3, 5, or 6, returning 1 if it is, and 0, otherwise.

After defining the function, typing `simdivis()` will simulate one trial. By repeatedly typing `simdivis()` on your computer, you get a feel for how this random experiment behaves over repeated trials.

In this script, instead of writing a loop, we use the `replicate` command. This powerful R command is an alternative to writing loops for simple expressions. The syntax is `replicate(n,expr)`. The expression `expr` is repeated n times creating an n-element vector. Thus the result of typing

```
> simlist <- replicate(1000, simdivis())
```

is a vector of 1000 ones and zeros stored in the variable `simlist` corresponding to success or failure in the divisibility experiment. The average `mean(simlist)` gives the simulated probability.

Play with this script. Based on 1000 trials, you might guess that the true probability is between 0.45 and 0.49. Increase the number of trials to 10,000 and the estimates are roughly between 0.46 and 0.48. At 100,000, the estimates become even more precise, between 0.465 and 0.468.

We can actually quantify this increase in precision in Monte Carlo simulation as n gets large. But that is a topic that will to have to wait until Chapter 9.

R: SIMULATING THE DIVISIBILITY PROBABILITY

```
# Divisible356.R
# simdivis() simulates one trial
> simdivis <- function()  {
   num <- sample(1:1000,1)
   if (num%%3==0 || num%%5==0 || num%%6==0) 1 else 0
   }
> simlist <- replicate(1000, simdivis())
mean(simlist)
```

1.11 SUMMARY

In this chapter, the first principles of probability were introduced: from random experiment and sample space to the properties of probability functions. We start with discrete sample spaces—sets are either finite or countably infinite. The simplest probability model is when outcomes in a finite sample space are equally likely. In that case, probability reduces to "counting." Basic counting principles are presented. General properties of probabilities are derived from the three defining properties of a probability function. Random variables are introduced. The chapter ends with a first look at simulation.

- **Random experiment:** An activity, process, or experiment in which the outcome is uncertain.
- **Sample space Ω:** Set of all possible outcomes of a random experiment.
- **Outcome ω:** The elements of a sample space.
- **Event:** A subset of the sample space; a collection of outcomes.
- **Random variable X:** Assigns numbers to the outcomes of a random experiment. A real-valued function defined on the sample space.
- **Probability function:** A function P that assigns numbers to the elements $\omega \in \Omega$ such that
  1. $P(\omega) \geq 0$
  2. $\sum_\omega P(\omega) = 1$
  3. For events A, $P(A) = \sum_{\omega \in A} P(\omega)$
- **Equally likely outcomes:** Probability model for a finite sample space in which all elements have the same probability.
- **Uniform distribution:** Let $S = \{s_1, \ldots, s_k\}$. Then X is uniformly distributed on S if $P(X = s_i) = 1/k$, for $i = 1 \ldots, k$.
- **Counting**
  1. **Multiplication principle:** If there are m ways for one thing to happen, and n ways for a second thing to happen, then there are mn ways for both things to happen.
  2. **Permutations:** A permutation of $\{1, \ldots, n\}$ is an n-element ordering of the n numbers. There are $n!$ permutations of an n-element set.
- **Sampling:** When sampling from a population, *sampling with replacement* is when objects are returned to the population after they are sampled; *sampling without replacement* is when objects are not returned to the population after they are sampled.
- **Properties of probabilities**
  1. **Simple addition rule:** If A and B are mutually exclusive, that is, disjoint, then $P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$.

2. **Implication:** If A implies B, that is, if $A \subseteq B$, then $P(A) \leq P(B)$.
3. **Complement:** The probability that A does not occur $P(A^c) = 1 - P(A)$.
4. **General addition rule:** For all events A and B, $P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(AB)$.

- **Monte Carlo simulation** is based on the relative frequency interpretation of probability. Given a random experiment and an event A, $P(A)$ is approximately the fraction of times in which A occurs in n repetitions of the random experiment. A Monte Carlo simulation of $P(A)$ is based on three principles:
  1. Trials—Simulate the random experiment, typically on a computer using the computer's random numbers.
  2. Success—Based on the outcome of each trial, determine whether or not A occurs.
  3. Replication—Repeat the aforementioned steps n times. The proportion of successful trials is the simulated estimate of $P(A)$.
- **Problem–solving strategies**
  1. **Taking complements:** Finding $P(A^c)$, the probability of the complement of an event, might be easier in some cases than finding $P(A)$, the probability of the event. This arises in "at least" problems. For instance, the complement of the event that "at least one of several things occur" is the event that "none of those things occur." In the former case, the event involves a union. In the latter case, the event involves an intersection.
  2. **Inclusion–exclusion:** This is another method for tackling "at least" problems. For three events, inclusion–exclusion gives

$$P(A \cup B \cup C) \\ = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC).$$

EXERCISES

Sample space, event, random variable

1.1 Your friend was sick and unable to make today's class. Explain to your friend, using your own words, the meaning of the terms (i) random experiment, (ii) sample space, (iii) event, and (iv) random variable.

For the following problems 1.2–1.5, identify (i) the random experiment, (ii) sample space, (iii) event, and (iv) random variable. Express the probability in question in terms of the defined random variable, but do not compute the probability.

1.2 Roll four dice. Consider the probability of getting all fives.

1.3 A pizza shop offers three toppings: pineapple, peppers, and pepperoni. A pizza can have 0, 1, 2, or 3 toppings. Consider the probability that a random customer asks for two toppings.

1.4 Bored one day, you decide to play the video game Angry Birds until you win. Every time you lose, you start over. Consider the probability that you win in less than 1000 tries.

1.5 In Angel’s garden, there is a 3% chance that a tomato will be bad. Angel harvests 100 tomatoes and wants to know the probability that at most five tomatoes are bad.

1.6 In two dice rolls, let X be the outcome of the first die, and Y the outcome of the second die. Then $X+Y$ is the sum of the two dice. Describe the following events in terms of simple outcomes of the random experiment:

(a) $\{X+Y=4\}$. (Solution: $\{13, 22, 31\}$.)

(b) $\{X+Y=9\}$.

(c) $\{Y=3\}$.

(d) $\{X=Y\}$.

(e) $\{X>2Y\}$.

1.7 A bag contains r red and b blue balls. You reach into the bag and take k balls. Let R be the number of red balls you take. Let B be the number of blue balls. Express the following events in terms of the random variables R and B:

(a) You pick no red balls. (Solution: $\{R=0\}$.)

(b) You pick one red and two blue balls.

(c) You pick four balls.

(d) You pick twice as many red balls as blue balls.

1.8 A couple plans to continue having children until they have a girl or until they have six children, whichever comes first. Describe a sample space and a reasonable random variable for this random experiment.

Probability functions

1.9 A sample space has four elements $\omega_1, \dots, \omega_4$ such that ω_1 is twice as likely as ω_2, which is three times as likely as ω_3, which is four times as likely as ω_4. Find the probability function.

1.10 A random experiment has three possible outcomes a, b, and c, with

$$P(a)=p, \quad P(b)=p^2, \quad \text{and} \quad P(c)=p.$$

What choice(s) of p makes this a valid probability model?

1.11 Let P_1 and P_2 be two probability functions on Ω. Define a new function P such that $P(A)=(P_1(A)+P_2(A))/2$. Show the P is a probability function.

"9781118241257c01" — 2013/9/10 — 16:19 — page 29 — #29

1.12 Suppose $P_1, \dots, P_k$ are probability functions on Ω. Let $a_1, \dots, a_k$ be a sequence of numbers. Under what conditions on the a_is will

$$P = a_1 P_1 + \cdots + a_k P_k$$

be a probability function?

1.13 Let P be a probability function on $\Omega = \{a, b\}$ such that $P(a) = p$ and $P(b) = 1 - p$ for $0 \leq p \leq 1$. Let Q be a function on Ω defined by $Q(\omega) = [P(\omega)]^2$. For what value(s) of p will Q be a valid probability function?

Equally likely outcomes and counting

1.14 A club has 10 members and is choosing a president, vice-president, and treasurer. All selections are equally likely.

(a) What is the probability that Tom is selected president?

(b) What is the probability that Brenda is chosen president and Liz is chosen treasurer?

1.15 A fair coin is flipped six times. What is the probability that the first two flips are heads and the last two flips are tails? Use the multiplication principle.

1.16 Suppose that license plates can be two, three, four, or five letters long, taken from the alphabets A to Z. All letters are possible, including repeats. A license plate is chosen at random in such a way so that all plates are equally likely.

(a) What is the probability that the plate is "A-R-R?"

(b) What is the probability that the plate is four letters long?

(c) What is the probability that the plate is a palindrome?

(d) What is the probability that the plate has at least one "R?"

1.17 Suppose you throw five dice and all outcomes are equally likely.

(a) What is the probability that all dice are the same? (In the game of Yahtzee, this is known as a *yahtzee*.)

(b) What is the probability of getting at least one 4?

(c) What is the probability that all the dice are different?

1.18 Amy is picking her fall term classes. She needs to fill three time slots, and there are 20 distinct courses to choose from, including probability 101, 102, and 103. She will pick her classes at random so that all outcomes are equally likely.

(a) What is the probability that she will get probability 101?

(b) What is the probability that she will get probability 101 and Probability 102?

(c) What is the probability she will get all three probability courses?

1.19 Suppose k numbers are chosen from $\{1,\ldots,n\}$, where $k < n$, sampling without replacement. All outcomes are equally likely. What is the probability that the numbers chosen are in increasing order?

Properties of probabilities

1.20 Suppose $P(A) = 0.40$, $P(B) = 0.60$, and $P(A \text{ or } B) = 0.80$. Find
(a) $P(\text{neither } A \text{ nor } B \text{ occur})$.
(b) $P(AB)$.
(c) $P(\text{one of the two events occurs, and the other does not})$.

1.21 Suppose A and B are mutually exclusive, with $P(A) = 0.30$ and $P(B) = 0.60$. Find the probability that
(a) At least one of the two events occurs
(b) Both of the events occur
(c) Neither event occurs
(d) Exactly one of the two events occur

1.22 Suppose $P(A \cup B) = 0.6$ and $P(A \cup B^c) = 0.8$. Find $P(A)$.

1.23 Suppose X is a random variable that takes values on all positive integers. Let $A = \{2 \le X \le 4\}$ and $B = \{X \ge 4\}$. Describe the events (i) A^c; (ii) B^c; (iii) AB; and (iv) $A \cup B$.

1.24 Suppose X is a random variable that takes values on $\{0, 0.01, 0.02, \ldots, 0.99, 1\}$. If each outcome is equally likely, find
(a) $P(X \le 0.33)$.
(b) $P(0.55 \le X \le .66)$.

1.25 Let A, B, C, be three events. At least one event always occurs. But it never happens that exactly one event occurs. Nor does it ever happen that all three events occur. If $P(AB) = 0.10$ and $P(AC) = 0.20$, find $P(B)$.

1.26 See the assignment of probabilities to the Venn diagram in Figure 1.4. Find the following:
(a) P(No events occur).
(b) P(Exactly one event occurs).
(c) P(Exactly two events occur).
(d) P(Exactly three events occur).
(e) P(At least one event occurs).
(f) P(At least two events occur).
(g) P(At most one event occurs).
(h) P(At most two events occur).

1.27 Four coins are tossed. Let A be the event that the first two coins comes up heads. Let B be the event that the number of heads is odd. Assume that all

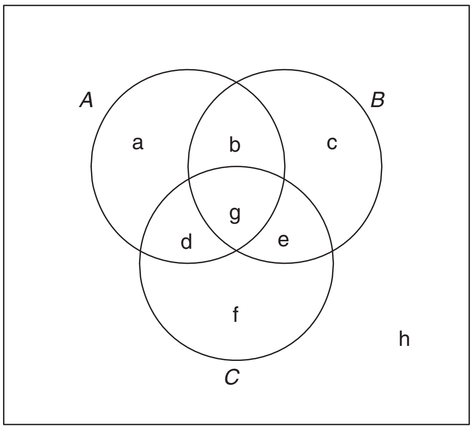

FIGURE 1.4: Venn diagram.

16 elements of the sample space are equally likely. Describe and find the probabilities of (i) AB, (ii) $A \cup B$, and (iii) AB^c.

1.28 Two dice are rolled. Let X be the maximum number obtained. (Thus, if 1 and 2 are rolled, $X = 2$; if 5 and 5 are rolled, $X = 5$.) Assume that all 36 elements of the sample space are equally likely. Find the probability function for X. That is, find $P(X = x)$, for $x = 1, 2, 3, 4, 5, 6$.

1.29 Judith has a penny, nickel, dime, and quarter in her pocket. So does Joe. They both reach into their pockets and choose a coin. Let X be the greater (in cents) of the two.

(a) Construct a sample space and describe the events $\{X = k\}$ for $k = 1, 5, 10, 25$.

(b) Assume that coin selections are equally likely. Find the probabilities for each of the aforementioned four events.

(c) What is the probability that Judith's coin is worth more than Joe's? (It is not $1/2$.)

1.30 A tetrahedron dice is four-sided and labeled with 1, 2, 3, and 4. When rolled it lands on the base of a pyramid and the number rolled is the number on the base. In five rolls, what is the probability of rolling at least one 2?

1.31 Let

$$Q(k) = \frac{2}{3^{k+1}}, \text{ for } k = 0, 1, 2, \ldots.$$

(a) Show that Q is a probability function. That is, show that the terms are nonnegative and sum to 1.

(b) Let X be a random variable such that $P(X = k) = Q(k)$, for $k = 0, 1, 2, \ldots$. Find $P(X > 2)$ without summing an infinite series.

1.32 The function

$$P(k) = c\frac{3^k}{k!}, \text{ for } k = 0, 1, 2, \ldots,$$

is a probability function for some choice of c. Find c.

1.33 Let A, B, C be three events. Find expressions for the events:

(a) At least one of the events occurs.

(b) Only B occurs.

(c) At most one of the events occurs.

(d) All of the events occur.

(e) None of the events occur.

1.34 The *odds in favor* of an event is the ratio of the probability that the event occurs to the probability that it will not occur. For example, the odds that you were born on a Friday, assuming birth days are equally likely, is 1 to 6, often written 16 or 1 to 6.

(a) In Texas Hold'em Poker, the odds of being dealt a pair (two cards of the same denomination) is 116. What is the chance of not being dealt a pair?

(b) For sporting events, bookies usually quote odds as odds against, as opposed to odds in favor. In the Kentucky Derby horse race, our horse Daddy Long Legs was given 2–9 odds. What is the chance that Daddy Long Legs wins the race?

1.35 An exam had three questions. One-fifth of the students answered the first question correctly; one-fourth answered the second question correctly; and one-third answered the third question correctly. For each pair of questions, one-tenth of the students got that pair correct. No one got all three questions right. Find the probability that a randomly chosen student did not get any of the questions correct.

1.36 Suppose $P(ABC) = 0.05$, $P(AB) = 0.15$, $P(AC) = 0.2$, $P(BC) = 0.25$, $P(A) = P(B) = P(C) = 0.5$. For each of the events given next, write the event using set notation in terms of A, B, and C, and compute the corresponding probability.

(a) At least one of the three events A, B, C occur.

(b) At most one of the three events occurs.

(c) All of the three events occurs.

(d) None of the three events occurs.

(e) At least two of the three events occurs.

(f) At most two of the three events occurs.

1.37 Find the probability that a random integer between 1 and 5000 is divisible by 4, 7 or 10.

1.38

(a) Each of the four squares of a two-by-two checkerboard is randomly colored red or black. Find the probability that at least one of the two columns of the checkerboard is all red.

(b) Each of the six squares of a two-by-three checkerboard is randomly colored red or black. Find the probability that at least one of the three columns of the checkerboard is all red.

1.39 Given events A and B, show that the probability that exactly one of the events occurs equals

$$P(A) + P(B) - 2P(AB).$$

1.40 Given events A, B, C, show that the probability that exactly one of the events occurs equals

$$P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + 3P(ABC).$$

Simulation and R

1.41 Modify the code in the R script **CoinFlip.R** to simulate the probability of getting exactly one head in four coin tosses.

1.42 Modify the code in the R script **Divisible356.R** to simulate the probability that a random integer between 1 and 5000 is divisible by 4, 7, or 10. Compare with your answer in Exercise 1.37.

1.43 Use R to simulate the probability of getting at least one 8 in the sum of two dice rolls.

1.44 Use R to simulate the probability in Exercise 1.30.

1.45 See the help file for the `sample` command (type `?sample`). Let X be a random variable taking values 1, 4, 8, and 16 with respective probabilities $0.1, 0.2, 0.3, 0.4$. Show how to simulate X.

1.46 Write a function `dice(`k`)` for generating k throws of a fair die. Use your function and R's `sum` function to generate the sum of two dice throws.

1.47 Make up your own random experiment and write an R script to simulate it.