
1

BASIC PROBABILITY THEORY

1.1 INTRODUCTION

Probability theory is the mathematics of randomness. This statement immediately invites the question “What is randomness?” This is a deep question that we cannot attempt to answer without invoking the disciplines of philosophy, psychology, mathematical complexity theory, and quantum physics, and still there would most likely be no completely satisfactory answer. For our purposes, an informal definition of randomness as “what happens in a situation where we cannot predict the outcome with certainty” is sufficient. In many cases, this might simply mean lack of information. For example, if we flip a coin, we might think of the outcome as random. It will be either heads or tails, but we cannot say which, and if the coin is fair, we believe that both outcomes are equally likely. However, if we knew the force from the fingers at the flip, weight and shape of the coin, material and shape of the table surface, and several other parameters, we would be able to predict the outcome with certainty, according to the laws of physics. In this case we use randomness as a way to describe uncertainty due to lack of information.¹

Next question: “What is probability?” There are two main interpretations of probability, one that could be termed “objective” and the other “subjective.” The first is

¹To quote the French mathematician Pierre-Simon Laplace, one of the first to develop a mathematical theory of probability: “Probability is composed partly of our ignorance, partly of our knowledge.”

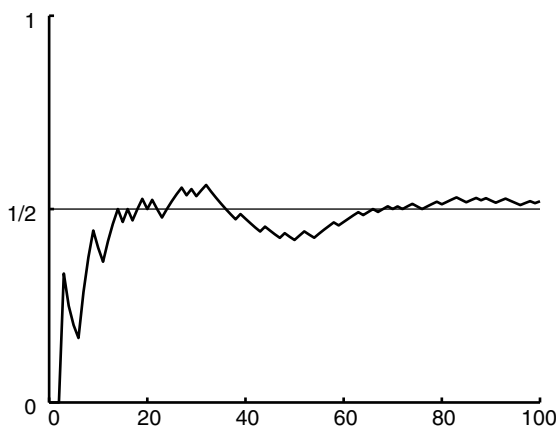


FIGURE 1.1 Consecutive relative frequencies of heads in 100 coin flips.

the interpretation of a probability as a *limit of relative frequencies*; the second, as a *degree of belief*. Let us briefly describe each of these.

For the first interpretation, suppose that we have an experiment where we are interested in a particular outcome. We can repeat the experiment over and over and each time record whether we got the outcome of interest. As we proceed, we count the number of times that we got our outcome and divide this number by the number of times that we performed the experiment. The resulting ratio is the *relative frequency* of our outcome. As it can be observed empirically that such relative frequencies tend to stabilize as the number of repetitions of the experiment grows, we might think of the limit of the relative frequencies as the probability of the outcome. In mathematical notation, if we consider n repetitions of the experiment and if S_n of these gave our outcome, then the relative frequency would be $f_n = S_n/n$, and we might say that the probability equals $\lim_{n \rightarrow \infty} f_n$. Figure 1.1 shows a plot of the relative frequency of heads in a computer simulation of 100 hundred coin flips. Notice how there is significant variation in the beginning but how the relative frequency settles in toward $\frac{1}{2}$ quickly.

The second interpretation, probability as a degree of belief, is not as easily quantified but has obvious intuitive appeal. In many cases, it overlaps with the previous interpretation, for example, the coin flip. If we are asked to quantify our degree of belief that a coin flip gives heads, where 0 means “impossible” and 1 means “with certainty,” we would probably settle for $\frac{1}{2}$ unless we have some specific reason to believe that the coin is not fair. In some cases it is not possible to repeat the experiment in practice, but we can still imagine a sequence of repetitions. For example, in a weather forecast you will often hear statements like “there is a 30% chance of rain tomorrow.” Of course, we cannot repeat the experiment; either it rains tomorrow or it does not. The 30% is the meteorologist’s measure of the chance of rain. There is still a connection to the relative frequency approach; we can imagine a sequence of days

with similar weather conditions, same time of year, and so on, and that in roughly 30% of the cases, it rains the following day.

The “degree of belief” approach becomes less clear for statements such as “the Riemann hypothesis is true” or “there is life on other planets.” Obviously, these are statements that are either true or false, but we do not know which, and it is not unreasonable to use probabilities to express how strongly we believe in their truth. It is also obvious that different individuals may assign completely different probabilities.

How, then, do we actually *define* a probability? Instead of trying to use any of these interpretations, we will state a strict mathematical definition of probability. The interpretations are still valid to develop intuition for the situation at hand, but instead of, for example, *assuming* that relative frequencies stabilize, we will be able to *prove* that they do, within our theory.

1.2 SAMPLE SPACES AND EVENTS

As mentioned in the introduction, probability theory is a mathematical theory to describe and analyze situations where randomness or uncertainty are present. Any specific such situation will be referred to as a *random experiment*. We use the term “experiment” in a wide sense here; it could mean an actual physical experiment such as flipping a coin or rolling a die, but it could also be a situation where we simply observe something, such as the price of a stock at a given time, the amount of rain in Houston in September, or the number of spam emails we receive in a day. After the experiment is over, we call the result an *outcome*. For any given experiment, there is a set of possible outcomes, and we state the following definition.

Definition 1.1. The set of all possible outcomes in a random experiment is called the *sample space*, denoted S .

Here are some examples of random experiments and their associated sample spaces.

Example 1.1. Roll a die and observe the number.

Here we can get the numbers 1 through 6, and hence the sample space is

$$S = \{1, 2, 3, 4, 5, 6\}$$

□

Example 1.2. Roll a die repeatedly and count the number of rolls it takes until the first 6 appears.

Since the first 6 may come in the first roll, 1 is a possible outcome. Also, we may fail to get 6 in the first roll and then get 6 in the second, so 2 is also a possible outcome. If

we continue this argument we realize that any positive integer is a possible outcome and the sample space is

$$S = \{1, 2, \dots\}$$

the set of positive integers. □

Example 1.3. Turn on a lightbulb and measure its lifetime, that is, the time until it fails.

Here it is not immediately clear what the sample space should be since it depends on how accurately we can measure time. The most convenient approach is to note that the lifetime, at least in theory, can assume any nonnegative real number and choose as the sample space

$$S = [0, \infty)$$

where the outcome 0 means that the lightbulb is broken to start with. □

In these three examples, we have sample spaces of three different kinds. The first is *finite*, meaning that it has a finite number of outcomes, whereas the second and third are infinite. Although they are both infinite, they are different in the sense that one has its points separated, $\{1, 2, \dots\}$ and the other is an entire continuum of points. We call the first type *countable infinity* and the second *uncountable infinity*. We will return to these concepts later as they turn out to form an important distinction.

In the examples above, the outcomes are always numbers and hence the sample spaces are subsets of the real line. Here are some examples of other types of sample spaces.

Example 1.4. Flip a coin twice and observe the sequence of heads and tails.

With H denoting heads and T denoting tails, one possible outcome is HT , which means that we get heads in the first flip and tails in the second. Arguing like this, there are four possible outcomes and the sample space is

$$S = \{HH, HT, TH, TT\}$$

□

Example 1.5. Throw a dart at random on a dartboard of radius r .

If we think of the board as a disk in the plane with center at the origin, an outcome is an ordered pair of real numbers (x, y) , and we can describe the sample space as

$$S = \{(x, y) : x^2 + y^2 \leq r^2\}$$

□

Once we have described an experiment and its sample space, we want to be able to compute probabilities of the various things that may happen. What is the probability that we get 6 when we roll a die? That the first 6 does not come before the fifth roll? That the lightbulb works for at least 1500 h? That our dart hits the bull's eye? Certainly, we need to make further assumptions to be able to answer these questions, but before that, we realize that all these questions have something in common. They all ask for probabilities of either single outcomes or groups of outcomes. Mathematically, we can describe these as subsets of the sample space.

Definition 1.2. A subset of S , $A \subseteq S$, is called an *event*.

Note the choice of words here. The terms “outcome” and “event” reflect the fact that we are describing things that may happen in real life. Mathematically, these are described as elements and subsets of the sample space. This duality is typical for probability theory; there is a verbal description and a mathematical description of the same situation. The verbal description is natural when real-world phenomena are described and the mathematical formulation is necessary to develop a consistent theory. See Table 1.1 for a list of set operations and their verbal description.

Example 1.6. If we roll a die and observe the number, two possible events are that we get an odd outcome and that we get at least 4. If we view these as subsets of the sample space, we get

$$A = \{1, 3, 5\} \quad \text{and} \quad B = \{4, 5, 6\}$$

If we want to use the verbal description, we might write this as

$$A = \{\text{odd outcome}\} \quad \text{and} \quad B = \{\text{at least 4}\}$$

□

We always use “or” in its nonexclusive meaning; thus, “ A or B occurs” includes the possibility that both occur. Note that there are different ways to express combinations of events; for example, $A \setminus B = A \cap B^c$ and $(A \cup B)^c = A^c \cap B^c$. The latter is known as one of *De Morgan's laws*, and we state these without proof together with some other basic set theoretic rules.

TABLE 1.1 Basic Set Operations and Their Verbal Description

Notation	Mathematical Description	Verbal Description
$A \cup B$	The union of A and B	A or B (or both) occurs
$A \cap B$	The intersection of A and B	Both A and B occur
A^c	The complement of A	A does not occur
$A \setminus B$	The difference between A and B	A occurs but not B
$\emptyset$	The empty set	Impossible event

Proposition 1.1. Let A , B , and C be events. Then

$$(a) \text{ (Distributive Laws) } (A \cap B) \cup C = (A \cup C) \cap (B \cup C) \\ (A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$

$$(b) \text{ (De Morgan's Laws) } (A \cup B)^c = A^c \cap B^c \\ (A \cap B)^c = A^c \cup B^c$$

As usual when dealing with set theory, *Venn diagrams* are useful. See Figure 1.2 for an illustration of some of the set operations introduced above. We will later return to how Venn diagrams can be used to calculate probabilities. If A and B are such that $A \cap B = \emptyset$, they are said to be *disjoint* or *mutually exclusive*. In words, this means that they cannot both occur simultaneously in the experiment.

As we will often deal with unions of more than two or three events, we need more general versions of the results given above. Let us first introduce some notation. If $A_1, A_2, \dots, A_n$ is a sequence of events, we denote

$$\bigcup_{k=1}^n A_k = A_1 \cup A_2 \cup \dots \cup A_n$$

the union of all the A_k and

$$\bigcap_{k=1}^n A_k = A_1 \cap A_2 \cap \dots \cap A_n$$

the intersection of all the A_k . In words, these are the events that *at least one* of the A_k occurs and that *all* the A_k occur, respectively. The distributive and De Morgan's laws extend in the obvious way, for example

$$\left(\bigcup_{k=1}^n A_k \right)^c = \bigcap_{k=1}^n A_k^c$$

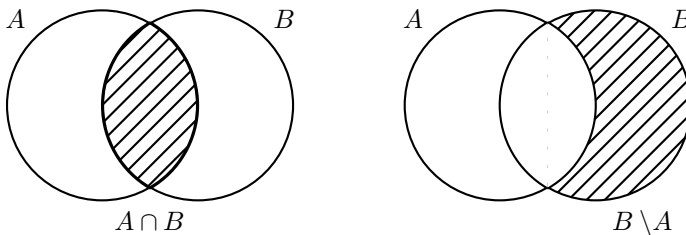


FIGURE 1.2 Venn diagrams of the intersection and the difference between events.

It is also natural to consider infinite unions and intersections. For example, in Example 1.2, the event that the first 6 comes in an odd roll is the infinite union $\{1\} \cup \{3\} \cup \{5\} \cup \dots$ and we can use the same type of notation as for finite unions and write

$$\{\text{first 6 in odd roll}\} = \bigcup_{k=1}^{\infty} \{2k - 1\}$$

For infinite unions and intersections, distributive and De Morgan's laws still extend in the obvious way.

1.3 THE AXIOMS OF PROBABILITY

In the previous section, we laid the basis for a theory of probability by describing random experiments in terms of the sample space, outcomes, and events. As mentioned, we want to be able to compute probabilities of events. In the introduction, we mentioned two different interpretations of probability: as a limit of relative frequencies and as a degree of belief. Since our aim is to build a consistent mathematical theory, as widely applicable as possible, our definition of probability should not depend on any particular interpretation. For example, it makes intuitive sense to require a probability to always be less than or equal to one (or equivalently, less than or equal to 100%). You cannot flip a coin 10 times and get 12 heads. Also, a statement such as "I am 150% sure that it will rain tomorrow" may be used to express extreme pessimism regarding an upcoming picnic but is certainly not sensible from a logical point of view. Also, a probability should be equal to one (or 100%), when there is absolute certainty, regardless of any particular interpretation.

Other properties must hold as well. For example, if you think there is a 20% chance that Bob is in his house, a 30% chance that he is in his backyard, and a 50% chance that he is at work, then the chance that he is at home is 50%, the sum of 20% and 30%. Relative frequencies are also *additive* in this sense, and it is natural to demand that the same rule apply for probabilities.

We now give a mathematical definition of probability, where it is defined as a real-valued function of the events, satisfying three properties, which we refer to as the *axioms of probability*. In the light of the discussion above, they should be intuitively reasonable.

Definition 1.3 (Axioms of Probability). A *probability measure* is a function P , which assigns to each event A a number $P(A)$ satisfying

- (a) $0 \leq P(A) \leq 1$
- (b) $P(S) = 1$

- (c) If $A_1, A_2, \dots$ is a sequence of *pairwise disjoint* events, that is, if $i \neq j$, then $A_i \cap A_j = \emptyset$, then

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$$

We read $P(A)$ as “the probability of A .” Note that a probability in this sense is a real number between 0 and 1 but we will occasionally also use percentages so that, for example, the phrases “The probability is 0.2” and “There is a 20% chance” mean the same thing.²

The third axiom is the most powerful assumption when it comes to deducing properties and further results. Some texts prefer to state the third axiom for finite unions only, but since infinite unions naturally arise even in simple examples, we choose this more general version of the axioms. As it turns out, the finite case follows as a consequence of the infinite. We next state this in a proposition and also that the empty set has probability zero. Although intuitively obvious, we must prove that it follows from the axioms. We leave this as an exercise.

Proposition 1.2. Let P be a probability measure. Then

- (a) $P(\emptyset) = 0$
 (b) If $A_1, \dots, A_n$ are pairwise disjoint events, then

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n P(A_k)$$

In particular, if A and B are disjoint, then $P(A \cup B) = P(A) + P(B)$. In general, unions need not be disjoint and we next show how to compute the probability of a union in general, as well as prove some other basic properties of the probability measure.

Proposition 1.3. Let P be a probability measure on some sample space S and let A and B be events. Then

- (a) $P(A^c) = 1 - P(A)$
 (b) $P(A \setminus B) = P(A) - P(A \cap B)$
 (c) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
 (d) If $A \subseteq B$, then $P(A) \leq P(B)$

²If the sample space is very large, it may be impossible to assign probabilities to *all* events. The class of events then needs to be restricted to what is called a σ -field. For a more advanced treatment of probability theory, this is a necessary restriction, but we can safely disregard this problem.

Proof. We prove (b) and (c), and leave (a) and (d) as exercises. For (b), note that $A = (A \cap B) \cup (A \setminus B)$, which is a disjoint union, and Proposition 1.2 gives

$$P(A) = P(A \cap B) + P(A \setminus B)$$

which proves the assertion. For (c), we write $A \cup B = A \cup (B \setminus A)$, which is a disjoint union, and we get

$$P(A \cup B) = P(A) + P(B \setminus A) = P(A) + P(B) - P(A \cap B)$$

by part (b). ■

Note how we repeatedly used Proposition 1.2(b), the finite version of the third axiom. In Proposition 1.3(c), for example, the events A and B are not necessarily disjoint but we can represent their union as a union of other events that are disjoint, thus allowing us to apply the third axiom.

Example 1.7. Mrs Boudreaux and Mrs Thibodeaux are chatting over their fence when the new neighbor walks by. He is a man in his sixties with shabby clothes and a distinct smell of cheap whiskey. Mrs B, who has seen him before, tells Mrs T that he is a former Louisiana state senator. Mrs T finds this very hard to believe. “Yes,” says Mrs B, “he is a former state senator who got into a scandal long ago, had to resign and started drinking.” “Oh,” says Mrs T, “that sounds more probable.” “No,” says Mrs B, “I think you mean less probable.”

Actually, Mrs B is right. Consider the following two statements about the shabby man: “He is a former state senator” and “He is a former state senator who got into a scandal long ago, had to resign, and started drinking.” It is tempting to think that the second is more probable because it gives a more exhaustive explanation of the situation at hand. However, this is precisely why it is a *less* probable statement. To explain this with probabilities, consider the experiment of observing a person and the two events

$$A = \{\text{he is a former state senator}\}$$

$$B = \{\text{he got into a scandal long ago, had to resign, and started drinking}\}$$

The first statement then corresponds to the event A and the second to the event $A \cap B$, and since $A \cap B \subseteq A$, we get $P(A \cap B) \leq P(A)$. Of course, what Mrs T meant was that it was easier to believe that the man was a former state senator once she knew more about his background.

In their book *Judgment under Uncertainty*, Kahneman et al. [5], show empirically how people often make similar mistakes when asked to choose the most probable among a set of statements. With a strict application of the rules of probability, we get it right. □

Example 1.8. Consider the following statement: “I heard on the news that there is a 50% chance of rain on Saturday and a 50% chance of rain on Sunday. Then there must be a 100% chance of rain during the weekend.”

This is, of course, not true. However, it may be harder to point out precisely where the error lies, but we can address it with probability theory. The events of interest are

$$A = \{\text{rain on Saturday}\} \quad \text{and} \quad B = \{\text{rain on Sunday}\}$$

and the event of rain during the weekend is then $A \cup B$. The percentages are reformulated as probabilities so that $P(A) = P(B) = 0.5$ and we get

$$\begin{aligned} P(\text{rain during the weekend}) &= P(A \cup B) \\ &= P(A) + P(B) - P(A \cap B) \\ &= 1 - P(A \cap B) \end{aligned}$$

which is less than 1, that is, the chance of rain during the weekend is less than 100%. The error in the statement lies in that we can add probabilities only when the events are disjoint. In general, we need to subtract the probability of the intersection, which in this case is the probability that it rains both Saturday and Sunday. $\square$

Example 1.9. A dartboard has an area of 143 in^2 (square inches). In the center of the board, there is the “bull’s eye,” which is a disk of area 1 in^2 . The rest of the board is divided into 20 sectors numbered $1, 2, \dots, 20$. There is also a triple ring that has an area of 10 in^2 and a double ring of area 15 in^2 (everything rounded to nearest integers). Suppose that you throw a dart at random on the board. What is the probability that you get (a) double 14, (b) 14 but not double, (c) triple or the bull’s eye, and (d) an even number or a double?

Introduce the events $F = \{14\}$, $D = \{\text{double}\}$, $T = \{\text{triple}\}$, $B = \{\text{bull’s eye}\}$, and $E = \{\text{even}\}$. We interpret “throw a dart at random” to mean that any region is hit with a probability that equals the fraction of the total area of the board that region occupies. For example, each number has area $(143 - 1)/20 = 7.1 \text{ in}^2$ so the corresponding probability is $7.1/143$. We get

$$P(\text{double 14}) = P(D \cap F) = \frac{0.75}{143} \approx 0.005$$

$$\begin{aligned} P(14 \text{ but not double}) &= P(F \setminus D) = P(F) - P(F \cap D) \\ &= \frac{7.1}{143} - \frac{0.75}{143} \approx 0.044 \end{aligned}$$

$$\begin{aligned}
 P(\text{triple or bulls eye}) &= P(T \cup B) = P(T) + P(B) \\
 &= \frac{10}{143} + \frac{1}{143} \approx 0.077
 \end{aligned}$$

$$\begin{aligned}
 P(\text{even or double}) &= P(E \cup D) = P(E) + P(D) - P(E \cap D) \\
 &= \frac{71}{143} + \frac{15}{143} - \frac{7.5}{143} \approx 0.55
 \end{aligned}$$

□

Let us say a word here about the interplay between logical statements and events. In the previous example, consider the events $E = \{\text{even}\}$ and $F = \{14\}$. Clearly, if we get 14, we also get an even number. As a logical relation between statements, we would express this as

the number is 14 $\Rightarrow$ the number is even

and in terms of events, we would say “If F occurs, then E must also occur.” But this means that $F \subseteq E$ and hence

$\{\text{the number is 14}\} \subseteq \{\text{the number is even}\}$

and thus the set-theoretic analog of “ $\Rightarrow$ ” is “ $\subseteq$ ” that is useful to keep in mind.

Venn diagrams turn out to provide a nice and useful interpretation of probabilities. If we imagine the sample space S to be a rectangle of area 1, we can interpret the probability of an event A as the area of A (see Figure 1.3). For example, Proposition 1.3(c) says that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. With the interpretation of probabilities as areas, we thus have

$$\begin{aligned}
 P(A \cup B) &= \text{area of } A \cup B \\
 &= \text{area of } A + \text{area of } B - \text{area of } A \cap B \\
 &= P(A) + P(B) - P(A \cap B)
 \end{aligned}$$

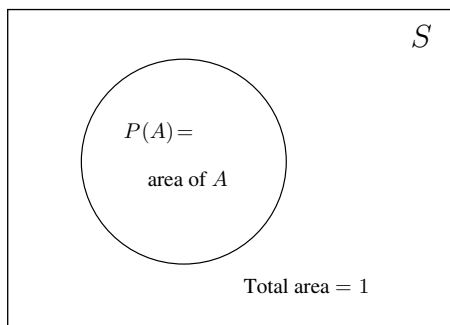


FIGURE 1.3 Probabilities with Venn diagrams.

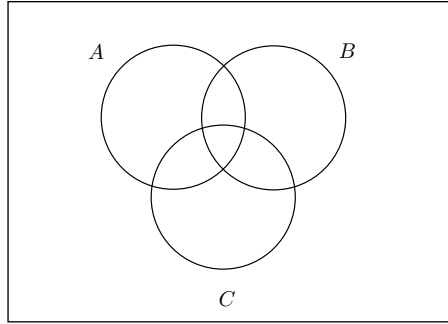


FIGURE 1.4 Venn diagram of three events.

since when we add the areas of A and B , we count the area of $A \cap B$ twice and must subtract it (think of A and B as overlapping pancakes where we are interested only in how much area they cover). Strictly speaking, this is not a proof but the method can be helpful to find formulas that can then be proved formally. In the case of three events, consider Figure 1.4 to argue that

$$\begin{aligned}
 \text{Area of } A \cup B \cup C &= \text{area of } A + \text{area of } B + \text{area of } C \\
 &\quad - \text{area of } A \cap B - \text{area of } A \cap C - \text{area of } B \cap C \\
 &\quad + \text{area of } A \cap B \cap C
 \end{aligned}$$

since the piece in the middle was first added three times and then removed three times, so in the end we have to add it again. Note that we must draw the diagram so that we get all possible combinations of intersections between the events. We have argued for the following proposition, which we state and prove formally.

Proposition 1.4. Let A , B , and C be three events. Then

$$\begin{aligned}
 P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\
 &\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) \\
 &\quad + P(A \cap B \cap C)
 \end{aligned}$$

Proof. By applying Proposition 1.3(c) twice—first to the two events $A \cup B$ and C and second to the events A and B —we obtain

$$\begin{aligned}
 P(A \cup B \cup C) &= P(A \cup B) + P(C) - P((A \cup B) \cap C) \\
 &= P(A) + P(B) - P(A \cap B) + P(C) - P((A \cup B) \cap C)
 \end{aligned}$$

The first four terms are what they should be. To deal with the last term, note that by the distributive laws for set operations, we obtain

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$

and yet another application of Proposition 1.3(c) gives

$$\begin{aligned} P((A \cup B) \cap C) &= P((A \cap C) \cup (B \cap C)) \\ &= P(A \cap C) + P(B \cap C) - P(A \cap B \cap C) \end{aligned}$$

which gives the desired result. ■

Example 1.10. Choose a number at random from the numbers $1, \dots, 100$. What is the probability that the chosen number is divisible by either 2, 3, or 5?

Introduce the events

$$A_k = \{\text{divisible by } k\} \text{ for } k = 1, 2, \dots$$

We interpret “at random” to mean that any set of numbers has a probability that is equal to its relative size, that is, the number of elements divided by 100. We then get

$$P(A_2) = 0.5, \quad P(A_3) = 0.33, \quad \text{and} \quad P(A_5) = 0.2$$

For the intersection, first note that, for example, $A_2 \cap A_3$ is the event that the number is divisible by both 2 and 3, which is the same as saying it is divisible by 6. Hence $A_2 \cap A_3 = A_6$ and

$$P(A_2 \cap A_3) = P(A_6) = 0.16$$

Similarly, we get

$$P(A_2 \cap A_5) = P(A_{10}) = 0.1, \quad P(A_3 \cap A_5) = P(A_{15}) = 0.06$$

and

$$P(A_2 \cap A_3 \cap A_5) = P(A_{30}) = 0.03$$

The event of interest is $A_2 \cup A_3 \cup A_5$, and Proposition 1.4 yields

$$P(A_2 \cup A_3 \cup A_5) = 0.5 + 0.33 + 0.2 - (0.16 + 0.1 + 0.06) + 0.03 = 0.74$$

□

It is now easy to believe that the general formula for a union of n events starts by adding the probabilities of the events, then subtracting the probabilities of the pairwise intersections, adding the probabilities of intersections of triples, and so on, finishing with either adding or subtracting the intersection of all the n events, depending on whether n is odd or even. We state this in a proposition that is sometimes referred to

as the *inclusion–exclusion formula*. It can, for example, be proved by induction, but we leave the proof as an exercise.

Proposition 1.5. Let $A_1, A_2, \dots, A_n$ be a sequence of n events. Then

$$\begin{aligned} P\left(\bigcup_{k=1}^n A_k\right) &= \sum_{k=1}^n P(A_k) \\ &\quad - \sum_{i < j} P(A_i \cap A_j) \\ &\quad + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) \\ &\quad \vdots \\ &\quad + (-1)^{n+1} P(A_1 \cap A_2 \cap \dots \cap A_n) \end{aligned}$$

We finish this section with a theoretical result that will be useful from time to time. A sequence of events is said to be *increasing* if

$$A_1 \subseteq A_2 \subseteq \dots$$

and *decreasing* if

$$A_1 \supseteq A_2 \supseteq \dots$$

In each case we can define the *limit* of the sequence. If the sequence is increasing, we define

$$\lim_{n \rightarrow \infty} A_n = \bigcup_{k=1}^{\infty} A_k$$

and if the sequence is decreasing

$$\lim_{n \rightarrow \infty} A_n = \bigcap_{k=1}^{\infty} A_k$$

Note how this is similar to limits of sequences of numbers, with $\subseteq$ and $\supseteq$ corresponding to $\leq$ and $\geq$, respectively, and union and intersection corresponding to supremum and infimum. The following proposition states that the probability measure is a *continuous set function*. The proof is outlined in Problem 18.

Proposition 1.6. If $A_1, A_2, \dots$ is either increasing or decreasing, then

$$P(\lim_{n \rightarrow \infty} A_n) = \lim_{n \rightarrow \infty} P(A_n)$$

1.4 FINITE SAMPLE SPACES AND COMBINATORICS

The results in the previous section hold for an arbitrary sample space S . In this section, we will assume that S is finite, $S = \{s_1, \dots, s_n\}$, say. In this case, we can always define the probability measure by assigning probabilities to the individual outcomes.

Proposition 1.7. Suppose that $p_1, \dots, p_n$ are numbers such that

(a) $p_k \geq 0, \quad k = 1, \dots, n$

(b) $\sum_{k=1}^n p_k = 1$

and for any event $A \subseteq S$, define

$$P(A) = \sum_{k: s_k \in A} p_k$$

Then P is a probability measure.

Proof. Clearly, the first two axioms of probability are satisfied. For the third, note that in a finite sample space, we cannot have infinitely many disjoint events, so we only have to check this for a disjoint union of two events A and B . We get

$$P(A \cup B) = \sum_{k: s_k \in A \cup B} p_k = \sum_{k: s_k \in A} p_k + \sum_{k: s_k \in B} p_k = P(A) + P(B)$$

and we are done. (Why are two events enough?) ■

Hence, when dealing with finite sample spaces, we do not need to explicitly give the probability of every event, only for each outcome. We refer to the numbers $p_1, \dots, p_n$ as a *probability distribution* on S .

Example 1.11. Consider the experiment of flipping a fair coin twice and counting the number of heads. We can take the sample space

$$S = \{HH, HT, TH, TT\}$$

and let $p_1 = \cdots = p_4 = \frac{1}{4}$. Alternatively, since all we are interested in is the number of heads and this can be 0, 1, or 2, we can use the sample space

$$S = \{0, 1, 2\}$$

and let $p_0 = \frac{1}{4}$, $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{4}$. □

Of particular interest is the case when all outcomes are equally likely. If S has n equally likely outcomes, then $p_1 = p_2 = \cdots = p_n = \frac{1}{n}$, which is called a *uniform distribution* on S . The formula for the probability of an event A now simplifies to

$$P(A) = \sum_{k: s_k \in A} \frac{1}{n} = \frac{\#A}{n}$$

where $\#A$ denotes the number of elements in A . This formula is often referred to as the *classical definition of probability* since historically this was the first context in which probabilities were studied. The outcomes in the event A can be described as *favorable* to A and we get the following formulation.

Corollary 1.1. *In a finite sample space with uniform probability distribution*

$$P(A) = \frac{\# \text{favorable outcomes}}{\# \text{possible outcomes}}$$

In daily language, the term “at random” is often used for something that has a uniform distribution. Although our concept of randomness is more general, this colloquial notion is so common that we will also use it (and already have). Thus, if we say “pick a number at random from 1, . . . , 10,” we mean “pick a number according to a uniform probability distribution on the sample space $\{1, 2, \dots, 10\}$.”

Example 1.12. Roll a fair die three times. What is the probability that all numbers are the same?

The sample space is the set of the 216 ordered triples (i, j, k) , and since the die is fair, these are all equally probable and we have a uniform probability distribution. The event of interest is

$$A = \{(1, 1, 1), (2, 2, 2), \dots, (6, 6, 6)\}$$

which has six outcomes and probability

$$P(A) = \frac{\# \text{favorable outcomes}}{\# \text{possible outcomes}} = \frac{6}{216} = \frac{1}{36}$$

□

Example 1.13. Consider a randomly chosen family with three children. What is the probability that they have exactly one daughter?

There are eight possible sequences of boys and girls (in order of birth), and we get the sample space

$$S = \{bbb, bbg, bgb, bgg, gbb, gbg, ggb, ggg\}$$

where, for example, *bbg* means that the oldest child is a boy, the middle child a boy, and the youngest child a girl. If we assume that all outcomes are equally likely, we get a uniform probability distribution on S , and since there are three outcomes with one girl, we get

$$P(\text{one daughter}) = \frac{3}{8}$$

□

Example 1.14. Consider a randomly chosen girl who has two siblings. What is the probability that she has no sisters?

Although this seems like the same problem as in the previous example, it is not. If, for example, the family has three girls, the chosen girl can be any of these three, so there are three different outcomes and the sample space needs to take this into account. Let g^* denote the chosen girl to get the sample space

$$S = \{g^*gg, gg^*g, ggg^*, g^*gb, gg^*b, g^*bg, bg^*g, bgg^*, g^*bb, bg^*b, bbg^*\}$$

and since 3 out of 12 equally likely outcomes have no sisters we get

$$P(\text{no sisters}) = \frac{1}{4}$$

which is smaller than the $\frac{3}{8}$ we got above. On average, 37.5% of families with three children have a single daughter and 25% of girls in three-children families are single daughters. □

1.4.1 Combinatorics

Combinatorics, “the mathematics of counting,” gives rise to a wealth of probability problems. The typical situation is that we have a set of objects from which we draw repeatedly in such a way that all objects are equally likely to be drawn. It is often tedious to list the sample space explicitly, but by counting combinations we can find the total number of cases and the number of favorable cases and apply the methods from the previous section.

The first problem is to find general expressions for the total number of combinations when we draw k times from a set of n distinguishable objects. There are different ways to interpret this. For example, we can draw *with* or *without replacement*,

depending on whether the same object can be drawn more than once. We can also draw *with* or *without regard to order*, depending on whether it matters in which order the objects are drawn. With these distinctions, there are four different cases, illustrated in the following simple example.

Example 1.15. Choose two numbers from the set $\{1, 2, 3\}$ and list the possible outcomes.

Let us first choose with regard to order. If we choose with replacement, the possible outcomes are

$$(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)$$

and if we choose without replacement

$$(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)$$

Next, let us choose without regard to order. This means that, for example, the outcomes $(1, 2)$ and $(2, 1)$ are regarded as the same and we denote it by $\{1, 2\}$ to stress that this is the *set* of 1 and 2, *not* the ordered pair. If we choose with replacement, the possible cases are

$$\{1, 1\}, \{1, 2\}, \{1, 3\}, \{2, 2\}, \{2, 3\}, \{3, 3\}$$

and if we choose without replacement

$$\{1, 2\}, \{1, 3\}, \{2, 3\}$$

□

To find expressions in the four cases for arbitrary values of n and k , we first need the following result. It is intuitively quite clear, and we state it without proof.

Proposition 1.8. If we are to perform r experiments in order, such that there are n_1 possible outcomes of the first experiment, n_2 possible outcomes of the second experiment, $\dots$, n_r possible outcomes of the r th experiment, then there is a total of $n_1 n_2 \cdots n_r$ outcomes of the sequence of the r experiments.

This is called the *fundamental principle of counting* or the *multiplication principle*. Let us illustrate it by a simple example.

Example 1.16. A Swedish license plate consists of three letters followed by three digits. How many possible license plates are there?

Although there are 28 letters in the Swedish alphabet, only 23 are used for license plates. Hence we have $r = 6$, $n_1 = n_2 = n_3 = 23$, and $n_4 = n_5 = n_6 = 10$. This gives a total of $23^3 \times 10^3 \approx 12.2$ million different license plates. □

We can now address the problem of drawing k times from a set of n objects. It turns out that choosing with regard to order is the simplest, so let us start with this and first consider the case of choosing with replacement. The first object can be chosen in n ways, and for each such choice, we have n ways to choose also the second object, n ways to choose the third, and so on. The fundamental principle of counting gives

$$n \times n \times \cdots \times n = n^k$$

ways to choose with replacement and with regard to order.

If we instead choose without replacement, the first object can be chosen in n ways, the second in $n - 1$ ways, since the first object has been removed, the third in $n - 2$ ways, and so on. The fundamental principle of counting gives

$$n(n - 1) \cdots (n - k + 1)$$

ways to choose without replacement and with regard to order. Sometimes, the notation

$$(n)_k = n(n - 1) \cdots (n - k + 1)$$

will be used for convenience, but this is not standard.

Example 1.17. From a group of 20 students, half of whom are female, a student council president and vice president are chosen at random. What is the probability of getting a female president and a male vice president?

The set of objects is the 20 students. Assuming that the president is drawn first, we need to take order into account since, for example, (Brenda, Bruce) is a favorable outcome but (Bruce, Brenda) is not. Also, drawing is done without replacement. Thus, we have $k = 2$ and $n = 20$ and there are $20 \times 19 = 380$ equally likely different ways to choose a president and a vice president. The sample space is the set of these 380 combinations and to find the probability, we need the number of favorable cases. By the fundamental principle of counting, this is $10 \times 10 = 100$. The probability of getting a female president and male vice president is $\frac{100}{380} \approx 0.26$. $\square$

Example 1.18. A human gene consists of nucleotide base pairs of four different kinds, A , C , G , and T . If a particular region of interest of a gene has 20 base pairs, what is the probability that a randomly chosen individual has no base pairs in common with a particular reference sequence in a database?

The set of objects is $\{A, C, G, T\}$, and we draw 20 times with replacement and with regard to order. Thus $k = 20$ and $n = 4$, so there are 4^{20} possible outcomes, and let us, for the sake of this example, assume that they are equally likely (which would not be true in reality). For the number of favorable outcomes, $n = 3$ instead of 4 since we need to avoid one particular letter in each choice. Hence, the probability is $3^{20}/4^{20} \approx 0.003$. $\square$

Example 1.19 (The Birthday Problem). This problem is a favorite in the probability literature. In a group of 100 people, what is the probability that at least two have the same birthday?

To simplify the solution, we disregard leap years and assume a uniform distribution of birthdays over the 365 days of the year. To assign birthdays to 100 people, we choose 100 out of 365 with replacement and get 365^{100} different combinations. The sample space is the set of those combinations, and the event of interest is

$$A = \{\text{at least two birthdays are equal}\}$$

and as it turns out, it is easier to deal with its complement

$$A^c = \{\text{all 100 birthdays are different}\}$$

To find the probability of A^c , note that the number of cases favorable to A^c is obtained by choosing 100 days out of 365 *without* replacement and hence

$$P(A) = 1 - P(A^c) = 1 - \frac{365 \times 364 \times \cdots \times 266}{365^{100}} \approx 0.9999997$$

Yes, that is a sequence of six 9s followed by a 7! Hence, we can be almost certain that any group of 100 people has at least two people sharing birthdays. A similar calculation reveals the probability of a shared birthday already exceeds $\frac{1}{2}$ at 23 people, a quite surprising result. About 50% of school classes thus ought to have kids who share birthdays, something that those with idle time on their hands can check empirically. $\square$

A check of real-life birthday distributions will reveal that the assumption of birthdays being uniformly distributed over the year is not true. However, the already high probability of shared birthdays only gets higher with a nonuniform distribution. Intuitively, this is because the less uniform the distribution, the more difficult it becomes to avoid birthdays already taken. For an extreme example, suppose that everybody was born in January, in which case there would be only 31 days to choose from instead of 365. Thus, in a group of 100 people, there would be absolute certainty of shared birthdays. Generally, it can be shown that the uniform distribution minimizes the probability of shared birthdays (we return to this in Problems 55 and 56).

Example 1.20 (The Birthday Problem Continued). A while ago I was in a group of exactly 100 people and asked for their birthdays. It turned out that nobody had the same birthday as I do. In the light of the previous problem, would this not be a very unlikely coincidence?

No, because here we are only considering the case of avoiding one particular birthday. Hence, with

$$B = \{\text{at least 1 out of 99 birthdays is the same as mine}\}$$

we get

$$B^c = \{99 \text{ birthdays are different from mine}\}$$

and the number of cases favorable to B^c is obtained by choosing with replacement from the 364 days that do not match my birthday. We get

$$P(B) = 1 - P(B^c) = 1 - \frac{364^{99}}{365^{99}} \approx 0.24$$

Thus, it is actually quite likely that nobody shares my birthday, and it is at the same time almost certain that at least somebody shares somebody else's birthday. $\square$

Next, we turn to the case of choosing without regard to order. First, suppose that we choose without replacement and let x be the number of possible ways, in which this can be done. Now, there are $n(n-1) \cdots (n-k+1)$ ways to choose with regard to order and each such ordered set can be obtained by first choosing the objects and then order them. Since there are x ways to choose the unordered objects and $k!$ ways to order them, we get the relation

$$n(n-1) \cdots (n-k+1) = x \times k!$$

and hence there are

$$x = \frac{n(n-1) \cdots (n-k+1)}{k!} \quad (1.1)$$

ways to choose without replacement, without regard to order. In other words, this is the number of subsets of size k of a set of size n , called the *binomial coefficient*, read “ n choose k ” and usually denoted and defined as

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

but we use the expression in Equation (1.1) for computations. By convention,

$$\binom{n}{0} = 1$$

and from the definition it follows immediately that

$$\binom{n}{k} = \binom{n}{n-k}$$

which is useful for computations. For some further properties, see Problem 24.

Example 1.21. In Texas Lotto, you choose five of the numbers $1, \dots, 44$ and one bonus ball number, also from $1, \dots, 44$. Winning numbers are chosen randomly. Which is more likely: that you match the first five numbers but not the bonus ball or that you match four of the first five numbers and the bonus ball?

Since we have to match five of our six numbers in each case, are the two not equally likely? Let us compute the probabilities and see. The set of objects is $\{1, 2, \dots, 44\}$ and the first five numbers are drawn without replacement and without regard to order. Hence, there are $\binom{44}{5}$ combinations and for each of these there are then 44 possible choices of the bonus ball. Thus, there is a total of $\binom{44}{5} \times 44 = 47,784,352$ different combinations. Introduce the events

$$\begin{aligned} A &= \{\text{match the first five numbers but not the bonus ball}\} \\ B &= \{\text{match four of the first five numbers and the bonus ball}\} \end{aligned}$$

For A , the number of favorable cases is 1×43 (only one way to match the first five numbers, 43 ways to avoid the winning bonus ball). Hence

$$P(A) = \frac{1 \times 43}{\binom{44}{5} \times 44} \approx 9 \times 10^{-7}$$

To find the number of cases favorable to B , note that there are $\binom{5}{4} = 5$ ways to match four out of five winning numbers and then $\binom{39}{1} = 39$ ways to avoid the fifth winning number. There is only one choice for the bonus ball and we get

$$P(B) = \frac{5 \times 39 \times 1}{\binom{44}{5} \times 44} \approx 4 \times 10^{-6}$$

so B is more than four times as likely as A . □

Example 1.22. You are dealt a poker hand (5 cards out of 52 without replacement).
(a) What is the probability that you get no hearts? **(b)** What is the probability that you get exactly k hearts? **(c)** What is the most likely number of hearts?

We will solve this by disregarding order. The number of possible cases is the number of ways in which we can choose 5 out of 52 cards, which equals $\binom{52}{5}$. In (a), to get a favorable case, we need to choose all 5 cards from the 39 that are not hearts. Since this can be done in $\binom{39}{5}$ ways, we get

$$P(\text{no hearts}) = \frac{\binom{39}{5}}{\binom{52}{5}} \approx 0.22$$

In (b), we need to choose k cards among the 13 hearts, and for each such choice, the remaining $5 - k$ cards are chosen among the remaining 39 that are not hearts. This gives

$$P(k \text{ hearts}) = \frac{\binom{13}{k} \binom{39}{5-k}}{\binom{52}{5}}, \quad k = 0, 1, \dots, 5$$

and for (c), direct computation gives the most likely number as 1, which has probability 0.41. $\square$

The problem in the previous example can also be solved by taking order into account. Hence, we imagine that we get the cards one by one and list them in order and note that there are $(52)_5$ different cases. There are $(13)_k(39)_{5-k}$ ways to choose so that we get k hearts and $5 - k$ nonhearts in a particular order. Since there are $\binom{5}{k}$ ways to choose position for the k hearts, we get

$$P(k \text{ hearts}) = \frac{\binom{5}{k} (13)_k (39)_{5-k}}{(52)_5}$$

which is the same as we got when we disregarded order above. It does not matter to the solution of the problem whether we take order into account, but we must be consistent and count the same way for the total and the favorable number of cases. In this particular example, it is probably easier to disregard order.

Example 1.23. An urn contains 10 white balls, 10 red balls, and 10 black balls. You draw five balls at random without replacement. What is the probability that you do not get all colors?

Introduce the events

$$R = \{\text{no red balls}\}, \quad W = \{\text{no white balls}\}, \quad B = \{\text{no black balls}\}$$

The event of interest is then $R \cup W \cup B$, and we will apply Proposition 1.4. First note that by symmetry, $P(R) = P(W) = P(B)$. Also, each intersection of any two events has the same probability and finally $R \cap W \cap B = \emptyset$. We get

$$P(\text{not all colors}) = 3P(R) - 3P(R \cap W)$$

In order to get no red balls, the 5 balls must be chosen among the 20 balls that are not red and hence

$$P(R) = \binom{20}{5} / \binom{30}{5}$$

Similarly, to get neither red nor white balls, the five balls must be chosen among the black balls and

$$P(R \cap W) = \binom{10}{5} / \binom{30}{5}$$

We get

$$P(\text{not all colors}) = 3 \left(\binom{20}{5} - \binom{10}{5} \right) / \binom{30}{5} \approx 0.32$$

□

Example 1.24. The final case, choosing with replacement and without regard to order, turns out to be the trickiest. As we noted above, when we choose without replacement, each unordered set of k objects corresponds to exactly $k!$ ordered sets. The relation is not so simple when we choose with replacement. For example, the unordered set $\{1, 1\}$ corresponds to one ordered set $(1, 1)$, whereas the unordered set $\{1, 2\}$ corresponds to two ordered sets $(1, 2)$ and $(2, 1)$. To find the general expression, we need to take a less direct route.

Imagine a row of n slots, numbered from 1 to n and separated by single walls where slot number j represents the j th object. Whenever object j is drawn, a ball is put in slot number j . After k draws, we will thus have k balls distributed over the n slots (and slots corresponding to objects never drawn are empty). The question now reduces to how many ways there are to distribute k balls over n slots. This is equivalent to rearranging the $n - 1$ inner walls and the k balls, which in turn is equivalent to choosing positions for the k balls from a total of $n - 1 + k$ positions. But this can be done in $\binom{n-1+k}{k}$ ways, and hence this is the number of ways to choose with replacement and without regard to order. □

Example 1.25. The Texas Lottery game “Pick 3” is played by picking three numbers with replacement from the numbers $0, 1, \dots, 9$. You can play “exact order” or “any order.” With the “exact order” option, you win when your numbers match the winning numbers in the exact order they are drawn. With the “any order” option, you win whenever your numbers match the winning numbers in any order. How many possible winning combinations are there with the “any order” option?

We have $n = 10$, $k = 3$, and the winning numbers are chosen with replacement and without regard to order and hence there are

$$\binom{10 - 1 + 3}{3} = \binom{12}{3} = 220$$

possible winning combinations. □

Example 1.26. Draw twice from the set $\{1, \dots, 9\}$ at random with replacement. What is the probability that the two drawn numbers are equal?

We have $n = 9$ and $k = 2$. Taking order into account, there are $9 \times 9 = 81$ possible cases, 9 of which are favorable. Hence the probability is $\frac{9}{81} = \frac{1}{9}$. If we disregard order, we have $\binom{9-1+2}{2} = 45$ possible cases and still 9 favorable and the probability is $\frac{9}{45} = \frac{1}{5}$. Since whether we draw with or without regard to order does not seem to matter to the question, why do we get different results?

The problem is that in the second case, when we draw without regard to order, the *distribution is not uniform*. For example, the outcome $\{1, 2\}$ corresponds to the two equally likely ordered outcomes $(1, 2)$ and $(2, 1)$ and is thus twice as likely as the outcome $\{1, 1\}$, which corresponds to only one ordered outcome $(1, 1)$. Thus, the first solution $\frac{1}{9}$ is correct. $\square$

Thus, when we draw with replacement but without regard to order, we must be careful when we compute probabilities, since the distribution is not uniform, as it is in the other three cases. Luckily, this case is far more uncommon in applications than are the other three cases. There is one interesting application, though, that has to do with the number of integer solutions to a certain type of equation. If we look again at the way in which we arrived at the formula and let x_j denote the number of balls in slot j , we realize that we must have $x_1 + \dots + x_n = k$ and get the following observation.

Corollary 1.2. *There are $\binom{n-1+k}{k}$ nonnegative integer solutions $(x_1, \dots, x_n)$ to the equation $x_1 + \dots + x_n = k$.*

The four different ways of choosing k out of n objects are summarized in Table 1.2. Note that when we choose without replacement, k must be less than or equal to n , but when we choose with replacement, there is no such restriction.

We finish with another favorite problem from the probability literature. It combines combinatorics with previous results concerning the probability of a union.

Example 1.27 (The Matching Problem). The numbers $1, 2, \dots, n$ are listed in random order. Whenever a number remains in its original position in the permutation, we call this a “match.” For example, if $n = 5$, then there are two matches in the

TABLE 1.2 Choosing k Out of n Objects

	With Replacement	Without Replacement
With regard to order	n^k	$n(n-1) \cdots (n-k+1)$
Without regard to order	$\binom{n-1+k}{k}$	$\binom{n}{k}$

permutation 32541 and none in 23451. **(a)** What is the probability that there are no matches? **(b)** What happens to the probability in (a) as $n \rightarrow \infty$?

Before we solve this, let us try to think about part (b). Does it get easier or harder to avoid matches when n is large? It seems possible to argue for both. With so many choices, it is easy to avoid a match in each particular position. On the other hand, there are many positions to try, so it should not be too hard to get at least one match. It is not easy to have good intuition for what happens here.

To solve the problem, we first consider the complement of no matches and introduce the events

$$\begin{aligned} A &= \{\text{at least one match}\} \\ A_k &= \{\text{match in the } k\text{th draw}\}, \quad k = 1, 2, \dots, n \end{aligned}$$

so that

$$A = \bigcup_{k=1}^n A_k$$

We will apply Proposition 1.5, so we need to figure out the probabilities of the events A_k as well as all intersections of two events, three events, and so on.

First, note that there are $n!$ different permutations of the numbers $1, 2, \dots, n$. To get a match in position k , there is only one choice for that number and the rest can be ordered in $(n-1)!$ different ways. We get the probability

$$P(A_k) = \frac{\# \text{ favorable outcomes}}{\# \text{ possible outcomes}} = \frac{(n-1)!}{n!} = \frac{1}{n}$$

which means that the first sum in Proposition 1.5 equals 1. To get a match in both the i th and the j th positions, we have only one choice for each of these two positions and the remaining $n-2$ numbers can be ordered in $(n-2)!$ ways and

$$P(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$$

Since there are $\binom{n}{2}$ ways to select two events A_i and A_j , we get the following equation for the second sum in Proposition 1.5:

$$\begin{aligned} \sum_{i < j} P(A_i \cap A_j) &= \binom{n}{2} \frac{1}{n(n-1)} \\ &= \frac{n(n-1)}{2!} \times \frac{1}{n(n-1)} = \frac{1}{2!} \end{aligned}$$

Proceeding to the third sum, a similar argument gives that, for fixed $i < j < k$,

$$\sum_{i < j < k} P(A_i \cap A_j \cap A_k) = \binom{n}{3} \times \frac{1}{n(n-1)(n-2)} = \frac{1}{3!}$$

and the pattern emerges. The j th sum in Proposition 1.5 equals $1/j!$, and with the alternating signs we get

$$P(\text{at least one match}) = 1 - \sum_{j=2}^n \frac{(-1)^j}{j!} = 1 - \sum_{j=0}^n \frac{(-1)^j}{j!}$$

which finally gives

$$P(\text{no matches}) = \sum_{j=0}^n \frac{(-1)^j}{j!}$$

This is interesting. First, the probability is not monotone in n , so we cannot say that it gets easier or harder to avoid matches as n increases. Second, as $n \rightarrow \infty$, we recognize the limit as the Taylor expansion of e^{-1} and hence the probability of no matches converges to $e^{-1} \approx 0.37$ as $n \rightarrow \infty$. We can also note how rapid the convergence is; already for $n = 4$, the probability is 0.375. Thus, for all practical purposes, the probability to get no matches is 0.37 regardless of n . In Problem 36, you are asked to find the probability of exactly j matches. $\square$

1.5 CONDITIONAL PROBABILITY AND INDEPENDENCE

In this section, we introduce the important notion of *conditional probability*. The idea behind this concept is that the value of a probability can change if we get additional information. For example, the probability of contracting lung cancer is higher among smokers than nonsmokers and the probability of voting Republican is higher in Texas than in Massachusetts.

To arrive at a formal definition of conditional probabilities, we consider the example with the dartboard from Example 1.9. Suppose you throw darts repeatedly at random on a dartboard and consider only those darts that hit the number 14. In the long run, what proportion of those will also be doubles? Since the area of 14 is $142/20 = 7.1 \text{ in}^2$ and the area of the double ring inside 14 is $15/20 = 0.75 \text{ in}^2$, in the long run we expect the proportion $0.75/7.1 \approx 0.11$ of hits of 14 to also be doubles. To express this as a statement about probabilities, we can say that if we know that a dart hits 14, the probability that it is also a double is 0.11. Since the probability of 14 is $P(F) = 7.1/143$ and of both double and 14 is $P(F \cap D) = 0.75/143$, we see that the probability that a dart hits a double if we know that it hits 14 is the ratio $P(F \cap D)/P(F)$.

Now, consider a sample space in general and let A and B be two events. If we know that B occurred in an experiment, what is the probability that A also occurred? We

can draw a Venn diagram and apply the same reasoning as above. Since the fraction of area of A inside B is $P(A \cap B)/P(B)$, it seems reasonable that this is the probability we seek. This is the intuition behind the following definition.

Definition 1.4. Let B be an event such that $P(B) > 0$. For any event A , denote and define the *conditional probability of A given B* as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

We think of this as the probability of A if we know that B has occurred. Hence, to compute a conditional probability means to compute a probability given additional information.

Example 1.28. Let us revisit Mrs B and Mrs T from Example 1.7. If we introduce a third event

$$C = \{\text{he is shabby-looking}\}$$

then one way to interpret Mrs T's comment "that sounds more probable" is that

$$P(A|B \cap C) > P(A|C)$$

that is, given that more of the background is known, it seems more likely that the person is who Mrs B says he is. $\square$

Example 1.29. Roll a die and observe the number. Let

$$A = \{\text{odd outcome}\} \text{ and } B = \{\text{at least 4}\}$$

What is $P(A|B)$?

We solve this in two different ways: (1) by using the definition and (2) by intuitive reasoning. Since $P(A \cap B) = P(\{5\}) = \frac{1}{6}$ and $P(B) = \frac{1}{2}$, the definition gives

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/2} = \frac{1}{3}$$

If we think about this intuitively, to condition on the event B means that we get the additional information that the outcome is at least 4. Since one of these three outcomes is also odd and outcomes are equally likely, the conditional probability of odd is $\frac{1}{3}$. $\square$

There is no general rule for whether it is easier to use the definition or intuitive reasoning. In the previous example, the "one out of three" approach works since outcomes are equally likely, but this is not always the case.

Conditional probabilities can make it easier to compute probabilities of intersections. Say that we want to compute $P(A \cap B)$ but that it is tricky to do so

directly. However, if we can find $P(B)$ and $P(A|B)$, then the definition tells us that $P(A \cap B) = P(A|B)P(B)$ and we are done. Let us look at some examples of this.

Example 1.30. In Example 1.8, we had the events $A = \{\text{rain on Saturday}\}$ and $B = \{\text{rain on Sunday}\}$, where $P(A) = P(B) = 0.5$. Now, suppose that a rainy day is followed by another rainy day with probability 0.7. What is the probability of rain during the weekend?

We already know that the probability of a rainy weekend is

$$P(A \cup B) = 1 - P(A \cap B)$$

where we can now compute $P(A \cap B)$ as

$$P(A \cap B) = P(B|A)P(A) = 0.7 \times 0.5 = 0.35$$

and we get

$$P(A \cup B) = 0.65$$

as the probability of rain during the weekend. □

Example 1.31. From a deck of cards, draw four cards at random, without replacement. If you get j aces, draw j cards from another deck. What is the probability of getting exactly two aces from each deck?

With

$$A = \{\text{two aces from the first deck}\}$$

$$B = \{\text{two aces from the second deck}\}$$

the event of interest is $A \cap B$, and it is not that easy to figure out its probability directly. However, if we use conditional probabilities, it is simple. We get

$$P(A) = \frac{\binom{4}{2} \binom{48}{2}}{\binom{52}{4}} \quad \text{and} \quad P(B|A) = \frac{\binom{4}{2}}{\binom{52}{2}}$$

and hence

$$P(A \cap B) = P(B|A)P(A) = \frac{\binom{4}{2} \binom{48}{2}}{\binom{52}{4}} \times \frac{\binom{4}{2}}{\binom{52}{2}} \approx 0.0001$$

□

Example 1.32. The online bookseller amazon.com has a feature called the “Gold Box.” When you enter this, you are presented with 10 special offers to buy various merchandise, anything from books and DVDs, to kitchenware and the “Panasonic ER411NC nose and ear hair groomer.” The offers are presented one at a time and each time you have to decide whether to take it or to pass. If you take it, you are done and will not get to see the rest of the offers. If you pass, that offer is gone and cannot be retrieved. Suggest a strategy that gives you at least 25% chance to win the best offer.

Let us assume that the offers are presented in random order. If your strategy is to always take the first offer or if you choose at random, your chance to win is 10%. How can this be improved?

A better strategy is to let five offers pass, remember the best thus far, and take the next offer that is better. If this never happens, you are forced to take the last offer. One case in which you will certainly win is if the second best offer is among the first five and the best is among the remaining five. Thus, let

$$A = \{\text{second best offer is among the first five}\}$$

$$B = \{\text{best offer is among the last five}\}$$

so that the event of interest is $A \cap B$, which has probability

$$P(A \cap B) = P(A|B)P(B)$$

Since the offers are randomly ordered, the best offer is equally likely to be in any position and hence $P(B) = \frac{5}{10}$. Given that the best is among the last five, the second best is equally likely to be any of the remaining nine, so the probability that it is among the first five is $P(A|B) = \frac{5}{9}$ and we get

$$P(\text{get the best offer}) = \frac{5}{9} \times \frac{5}{10} \approx 0.28$$

which is larger than 0.25. Note that $A \cap B$ is not the only way in which you can get the best offer, so the true probability is in fact higher than 0.28.

Generally, if there are n offers, the same strategy gives a probability to get the best offer that is at least

$$P(A \cap B) = P(A|B)P(B) = \frac{n/2}{n-1} \times \frac{n/2}{n} = \frac{n}{4(n-1)}$$

which is greater than $\frac{1}{4}$ regardless of n [if n is odd, we can replace $n/2$ by $(n+1)/2$]. It is quite surprising that we can do so well and for example have at least 25% chance to find the best of 10 million offers. It can be shown that an even better strategy is to first discard roughly ne^{-1} offers and then take the next that is better. The probability to win is then approximately $e^{-1} \approx 0.37$ (a number that also showed up in Example 1.27). $\square$

The way in which we have defined conditional probability makes good intuitive sense. However, remember that a probability is defined as something that satisfies the three axioms in Definition 1.3. We must therefore show that whenever we condition on an event B , the definition of conditional probability does not violate any of the axioms. We state this in a proposition.

Proposition 1.9. For fixed B , $P(A|B)$ satisfies the probability axioms:

- (a) $0 \leq P(A|B) \leq 1$
- (b) $P(S|B) = 1$
- (c) If $A_1, A_2, \dots$ is a sequence of pairwise disjoint events, then

$$P\left(\bigcup_{k=1}^{\infty} A_k \mid B\right) = \sum_{k=1}^{\infty} P(A_k|B)$$

Proof. Since $A \cap B \subseteq B$, we get $0 \leq P(A \cap B) \leq P(B)$ and part (a) follows. For (b), note that $B \subseteq S$ so that $P(S \cap B) = P(B)$ and hence $P(S|B) = 1$. Finally, for (c) first note that

$$\left(\bigcup_{k=1}^{\infty} A_k\right) \cap B = \bigcup_{k=1}^{\infty} (A_k \cap B)$$

and since $A_1, A_2, \dots$ are pairwise disjoint, so are the events $A_1 \cap B, A_2 \cap B, \dots$, and we get

$$P\left(\left(\bigcup_{k=1}^{\infty} A_k\right) \cap B\right) = \sum_{k=1}^{\infty} P(A_k \cap B)$$

Divide both sides with $P(B)$ to conclude the proof. ■

It is easily realized that $P(B|B) = 1$, and with this in mind, we can think of conditioning on B as viewing B as the new sample space. The nice thing about the proposition is that we now know that conditional probabilities have all the properties of probabilities that we stated in Proposition 1.3. We restate these properties for conditional probabilities in a corollary.

Corollary 1.3. *Provided that the conditional probabilities are defined, the following properties hold:*

- (a) $P(A^c|B) = 1 - P(A|B)$
- (b) $P(B \setminus A|C) = P(B|C) - P(A \cap B|C)$

- (c) $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$
 (d) If $A \subseteq B$, then $P(A|C) \leq P(B|C)$

It is important to keep in mind that properties of probabilities hold for events to the left of the conditioning bar and that the event to the right is fixed (see Problem 37).

If we think of probability as a measure of degree of belief, we can think of conditional probability as an update of that degree, in the light of new information. Here is an example of a logical oddity that philosophers of science love to toss around to confuse the rest of us.

Example 1.33. Consider the hypothesis “all swans are white.” We can say that each observation of a white swan strengthens our belief in, or *corroborates*, the hypothesis. Also, since the two statements “all swans are white” and “all nonwhite objects are nonswans” are logically equivalent, the hypothesis is also corroborated by the observation of something that is neither white nor a swan. Thus, every sighting of a yellow dog corroborates the hypothesis that all swans are white.³

Weird, isn’t it? A zoologist trying to prove the hypothesis would certainly decide to examine swans for whiteness, rather than checking various red, green, and blue objects to make sure that they are not swans. Still, there is certainly nothing wrong with the logic, so how can the paradox be resolved? Let us try a probabilistic approach.

Suppose that we have all examinable objects in a big urn. Suppose that there are n such objects, k of which are white, and that the other $n - k$ are black (representing “nonwhite”). Suppose further that j of the objects are swans, and call the remaining objects “ravens,” another favorite bird among philosophers of science. If we do not know anything about the whiteness of swans, we may assume that the j swans are randomly spread among the n objects. Thus, when we choose a swan, the probability that it is white is $\frac{k}{n}$ (if we have very strong belief in the hypothesis to begin with, we can just introduce a lot of white “dummy objects” to make this probability anything we want). The probability that the hypothesis is true can now be thought of as the probability to get only white objects when we draw without replacement j times (assign the “swan property” to j objects). Our hypothesis is then the event

$$H = \{\text{all swans are white}\} = \{\text{get } j \text{ white objects}\}$$

Let us choose with regard to order (which does not matter to the problem, but expressions get less messy). Thus, the probability that all swans are white is

$$P(H) = \frac{k(k-1) \cdots (k-j+1)}{n(n-1) \cdots (n-j+1)}$$

³For ornithologists: This has nothing to do with *Cygnus atratus*.

We now follow two different strategies: (a) to examine swans and (b) to examine black objects. Suppose that we get a corroborating observation. How does this affect the probability of H , now pertaining to the remaining $n - 1$ objects? Let C_a and C_b be the events to get corroborating observations with the two strategies, respectively. With strategy (a), a corroborating observation means that one white swan has been removed, and the conditional probability of H becomes

$$P(H|C_a) = \frac{(k-1)(k-2)\cdots(k-j+1)}{(n-1)(n-2)\cdots(n-j+1)}$$

With strategy (b), one black raven has been removed, and we get

$$P(H|C_b) = \frac{k(k-1)\cdots(k-j+1)}{(n-1)(n-2)\cdots(n-j)}$$

Both these are larger than the original $P(H)$, so each corroborating observation indeed strengthens belief in the hypothesis. But do they do so to equal extents? Let us compare the two conditional probabilities. We get

$$\frac{P(H|C_a)}{P(H|C_b)} = \frac{n-j}{k}$$

If we now assume that the number of swans is less than the number of black objects, certainly a reasonable assumption, we have that $j < n - k$, which gives $k < n - j$, and hence

$$\frac{P(H|C_a)}{P(H|C_b)} > 1$$

so that the observation of a black raven does corroborate the hypothesis but *not as much* as the sighting of a white swan. The intuition is simple: since there are fewer swans than black objects, it is easier to check the swans. If instead $j > n - k$, strategy (b) would be preferable. If we, for example, were to corroborate the hypothesis “All Volvo drivers live outside the Vatican,” it would be better to ask a thousand Vaticanos what they drive, than to track down Volvo drivers in London and Paris to check if they happen to be vacationing Swiss Guardsmen. $\square$

1.5.1 Independent Events

In the previous section, we dealt with conditional probabilities and learned to interpret them as probabilities that are computed given additional information. It is easy to think of cases when such additional information is irrelevant and does not change the probability. For example, if we are about to flip a fair coin, the probability to get heads is $\frac{1}{2}$. Now suppose that we get the additional information that the coin was flipped once yesterday and showed heads. Since our upcoming coin flip is not affected by what happened yesterday and we know that the coin is fair, the conditional probability given this information is still $\frac{1}{2}$. With $A = \{\text{heads in next flip}\}$ and $B = \{\text{heads yesterday}\}$ we thus have $P(A) = P(A|B)$; the unconditional and conditional probabilities are the

same. Since $P(A|B) = P(A \cap B)/P(B)$, this means that $P(A \cap B) = P(A)P(B)$, and we call two events with this property *independent*.

Definition 1.5. If A and B are two events such that

$$P(A \cap B) = P(A)P(B)$$

then they are said to be *independent*.

Not surprisingly, events that are not independent are called *dependent*. In the introductory motivation for the definition, we talked about conditional and unconditional probabilities being equal. We could take this as the definition of independence, but since conditional probabilities are not always defined, we use the definition of independence above and get the following consequence.

Corollary 1.4. If $P(A|B)$ is defined, then the events A and B are independent if and only if $P(A) = P(A|B)$.

When checking for independence, it might sometimes be easier to condition on the event B^c instead of B , that is, by supposing that B did not occur. Intuitively, information regarding B and information on B^c are equivalent since saying that one occurred is the same as saying that the other one did not occur. This is stated formally as follows.

Proposition 1.10. If A and B are independent, then A and B^c are also independent.

Proof. By Proposition 1.3(b), we get

$$P(A \cap B^c) = P(A \setminus B) = P(A) - P(A \cap B)$$

and if A and B are independent, this equals

$$P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c)$$

and A and B^c are independent. ■

Example 1.34. In Example 1.30, suppose that a rainy Saturday and a rainy Sunday are independent events. What is the probability of rain during the weekend?

In this case

$$P(A \cap B) = P(A)P(B) = 0.25$$

and hence

$$P(A \cup B) = 0.75$$

which we note is higher than the 0.65 we obtained if rainy Saturdays are more likely to be followed by rainy Sundays. The reason is that under this assumption, rainy Saturdays and Sundays tend to come together more often than under the independence assumption. $\square$

Example 1.35. A card is chosen at random from a deck of cards. Consider the events

$$A = \{\text{the card is an ace}\} \quad \text{and} \quad H = \{\text{the card is a heart}\}$$

Are A and H independent?

Let us first solve this by using the definition. We have $P(A) = \frac{4}{52}$, $P(H) = \frac{1}{4}$, and $P(A \cap H) = P(\text{ace of hearts}) = \frac{1}{52}$ and hence

$$P(A \cap H) = P(A)P(H)$$

so that A and H are independent. Intuitively, the events give no information about each other. The probability of drawing an ace is $\frac{4}{52} = \frac{1}{13}$ and if we are given the information that the chosen card is a heart, the probability of an ace is still $\frac{1}{13}$. The proportion of aces is the same in the deck as within the suit of hearts. $\square$

Example 1.36. Consider the previous example but suppose that we have removed the 2 of spades from the deck. Are the events A and H still independent?

At first glance, we might think that the answer is “Yes” since the 2 of spades has nothing to do with either hearts or aces. However, the probabilities are now $P(A) = \frac{4}{51}$, $P(H) = \frac{13}{51}$, and $P(A \cap H) = P(\text{ace of hearts}) = \frac{1}{51}$ and hence

$$P(A \cap H) \neq P(A)P(H)$$

and A and H are no longer independent. Intuitively, although the 2 of spades has nothing to do with hearts or aces, its removal changes the proportion of aces in the deck from $\frac{4}{52}$ to $\frac{4}{51}$ but does not change the proportion within the suit of hearts, where it remains at $\frac{1}{13}$. Formulated as a statement about conditional probabilities, we have that

$$P(A) = \frac{4}{51} \quad \text{and} \quad P(A|H) = \frac{1}{13}$$

which are not equal. $\square$

Example 1.37. Are disjoint events independent?

It seems that disjoint events have nothing to do with each other and should thus be independent. However, this reasoning is faulty. The correct reasoning is that if we condition on one event having occurred, then the other cannot have occurred, and hence its conditional probability drops to 0. We can also see this from the definition of independence since if A and B are disjoint, then $A \cap B = \emptyset$ and hence $P(A \cap B) = P(\emptyset) = 0$, which does not equal the product $P(A)P(B)$ (assuming that neither of these probabilities equal 0). Hence, the answer in general is “absolutely not.” $\square$

In Example 1.36, the events {ace} and {hearts} are dependent. Computation yields that $P(A) = 0.078$ and $P(A|H) = 0.077$, so the difference is negligible from a practical point of view. We could say that although the events are dependent, the dependence is not strong. Compare this with the case of disjoint events where the conditional probability drops down to 0, which indicates a much stronger dependence. Dependence could also go in different directions; $P(A|B)$ could be either larger or smaller than $P(A)$. We will later return to the problem of measuring the degree of dependence in a more general context (see also Problem 42).

The following two examples illustrate how it is not always obvious which event to condition on and how it is important to find the correct such event.

Example 1.38. You know that your new neighbors have two children. Given that they have at least one daughter, what is the conditional probability that they have two daughters?

The sample space is

$$S = \{bb, bg, gb, gg\}$$

where b represents boy, g represents girl, and the order is birth order. If we assume that genders are equally likely and that genders of different children are independent, each outcome has probability $\frac{1}{4}$. Since the outcome bb is out of the question and one out of the other three outcomes has two girls, the conditional probability is $\frac{1}{3}$. Formally

$$P(gg|bg, gb, gg) = \frac{P(gg)}{P(bg, gb, gg)} = \frac{1/4}{3/4} = \frac{1}{3}$$

$\square$

Example 1.39. You know that your new neighbors have two children. One day you see the mother taking a walk with a girl. What is the probability that the other child is also a girl?

This looks like the same problem. On the basis of your observation, you rule out the outcome bb and the conditional probability of another girl is $\frac{1}{3}$. On the other hand,

since we assume that genders of different children are independent, the probability ought to be $\frac{1}{2}$.

Confusing? Let us clear it up. The first solution is incorrect, but why? While it is true that the probability of two girls, given at least one girl, is $\frac{1}{3}$, this is *not the correct event* on which to condition in this case. We are not just observing “at least one girl;” we are observing the mother walking with a *particular girl*. This distinction is important but quite subtle, and requires that we extend the sample space to be able to also describe how the mother chooses which child to walk with.⁴ Thus, we split each outcome into two, and if we denote the child that goes for the walk by an asterisk, the new sample space is

$$S = \{b^*b, bb^*, b^*g, bg^*, g^*b, gb^*, g^*g, gg^*\}$$

where, for example, b^*g means that the older child is a boy, and the younger, a girl, and that the mother takes a walk with the boy. If the mother chooses child at random, each outcome has probability $\frac{1}{8}$. It is now easy to see that four outcomes have the mother walking with a girl and that two of these have another girl, and we arrive at the solution $\frac{1}{2}$ once more (see also Problem 80). $\square$

We also want to define independence of more than two events. To arrive at a reasonable definition, let us first examine an example that highlights one of the problems that must be addressed.

Example 1.40. Flip two fair coins and consider the events

$$A = \{\text{heads in first flip}\} = \{HH, HT\}$$

$$B = \{\text{heads in second flip}\} = \{HH, TH\}$$

$$C = \{\text{different in first and second flip}\} = \{HT, TH\}.$$

Then, for example, $P(A \cap B) = P(HH) = \frac{1}{4} = P(A)P(B)$, so A and B are independent. Similarly, it is easy to show that any two of the events are independent. Hence, these events are *pairwise independent*. However, it does not seem quite right to say that the three events A , B , and C are independent since, for example, C is not independent of the event $A \cap B$. Indeed, $P(C) = \frac{1}{2}$ but $P(C|A \cap B) = 0$, since if $A \cap B$ has occurred, both flips showed heads and C is impossible. $\square$

This example indicates that in order to call three events independent, we want each event to be independent of any combination of the other two. It turns out that the following definition guarantees this (see Problem 53).

⁴Ironically, in the first edition of his excellent book *Innumeracy: Mathematical Illiteracy and Its Consequences*, John Allen Paulos described this problem a bit obscurely [4]. His terse formulation was “Consider now some randomly selected family of four. Given that Myrtle has a sibling, what is the conditional probability that her sibling is a brother?” and he went on to claim that the probability is $\frac{2}{3}$. This ambiguity was clarified in the 2001 edition.

Definition 1.6. Three events A , B , and C are called independent if the following two conditions hold:

- (a) They are pairwise independent
- (b) $P(A \cap B \cap C) = P(A)P(B)P(C)$

For more than three events, the definition is analogous and can also be extended to infinitely many events.

Definition 1.7. The events $A_1, A_2, \dots$ are called independent if

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k})$$

for all sequences of integers $i_1 < i_2 < \dots < i_k, k = 2, 3, \dots$

Sometimes events satisfying this definition are called *mutually independent*, to distinguish from *pairwise independent*, which, as we have seen, is a weaker property.

Example 1.41. Recall the experiment of rolling a die repeatedly until the first 6 appears. What is the probability that this occurs in the n th roll for $n = 1, 2, \dots$?

The event of interest is

$$B_n = \{\text{first 6 in } n\text{th roll}\}, \quad n = 1, 2, \dots$$

and let us also introduce the events

$$A_k = \{6 \text{ in } k\text{th roll}\}, \quad k = 1, 2, \dots$$

Note the difference: B_n is the event that the *first* 6 comes in the n th roll; A_k , the event that we get 6 in the k th roll but not necessarily for the first time. How do the events relate to each other? Obviously, $B_1 = A_1$. For $n = 2$, note that B_2 is the event that we do not get 6 in the first roll and that we do get 6 in the second roll. In terms of the A_k , this is $A_1^c \cap A_2$. In general

$$B_n = A_1^c \cap A_2^c \cap \dots \cap A_{n-1}^c \cap A_n$$

To compute the probability of B_n , we make two reasonable assumptions: that the die is fair and that rolls are independent. The first assumption means that $P(A_k) = \frac{1}{6}$ for

all k and the second that probabilities of intersections equal products of probabilities. Since independence carries over to complements, we get

$$\begin{aligned} P(B_n) &= P(A_1^c \cap A_2^c \cap \cdots \cap A_{n-1}^c \cap A_n) \\ &= P(A_1^c)P(A_2^c) \cdots P(A_{n-1}^c)P(A_n) \\ &= \frac{5}{6} \times \frac{5}{6} \times \cdots \times \frac{5}{6} \times \frac{1}{6} \end{aligned}$$

and we conclude that

$$P(B_n) = \frac{1}{6} \left(\frac{5}{6} \right)^{n-1}, \quad n = 1, 2, \dots$$

□

More generally, consider independent repetitions of a trial where the event A occurs with probability $p > 0$ and let E be the event that we *never* get A . With

$$B_n = \{\text{first occurrence of } A \text{ comes after the } n\text{th trial}\}$$

we have

$$E = \bigcap_{n=1}^{\infty} B_n$$

where

$$P(B_n) = P(\text{the first } n \text{ trials give } A^c) = (1 - p)^n$$

by independence. The B_n are clearly decreasing (why?), so by Proposition 1.6 we get

$$P(E) = \lim_{n \rightarrow \infty} P(B_n) = 0$$

and we summarize in the following corollary.

Corollary 1.5. *In independent repetitions of a trial, any event with positive probability occurs sooner or later.*

From Example 1.21, we can compute the probability to win the Texas Lotto jackpot (match all numbers including bonus ball) as $1/47,784,352 = 2.1 \times 10^{-8}$. This is very small, but if you keep playing, the last result tells you that you will win eventually.⁵ It may take some time, though; there are two draws a week and if you play every time for 50 years, the probability that you never win is

$$(1 - 2.1 \times 10^{-8})^{5200} \approx 0.9999$$

⁵The subtle difference between *certain* occurrence and occurrence *with probability one* is important in a more advanced study of probability theory but not for us at this point.

The probability that you win in a drawing is very low, but since there are millions of players in each draw, the probability that *somebody* wins is much higher. Suppose that 5 million number combinations are played independently and at random for a drawing. The probability that somebody wins is

$$1 - (1 - 2.1 \times 10^{-8})^{5,000,000} \approx 0.10$$

which is not that low, and it could be you.

Example 1.42 (Reliability Theory). Consider a system of two electronic components connected in series. Each component functions with probability p and the components function independent of each other. What is the probability that the system functions?

If we interpret “functions” as the natural “lets current through,” then the system functions if and only if both components function. Hence, with the events

$$\begin{aligned} A &= \{\text{system functions}\} \\ A_1 &= \{\text{first component functions}\} \\ A_2 &= \{\text{second component functions}\} \end{aligned}$$

we get $A = A_1 \cap A_2$ and by independence

$$P(A) = P(A_1)P(A_2) = p^2$$

If the components are instead connected in parallel, the system functions as long as at least one of the components function, and we have

$$A = A_1 \cup A_2$$

which gives

$$\begin{aligned} P(A) &= P(A_1 \cup A_2) = 1 - P(A_1^c \cap A_2^c) \\ &= 1 - (1 - P(A_1))(1 - P(A_2)) = 1 - (1 - p)^2 \end{aligned}$$

These are simple examples from the discipline of *reliability theory* where the probability of functioning is referred to as the *reliability* of a system. Hence, we have seen that the reliability of a series system is p^2 and that of a parallel system is $1 - (1 - p)^2$. An obvious generalization is to n components, where the reliability of a series system is p^n and that of a parallel system is $1 - (1 - p)^n$. This does not have to be about electronic components but applies to any situation where a complex system is dependent on its individual parts to function. The series system is sometimes referred to as a *weakest-link model*. $\square$

1.6 THE LAW OF TOTAL PROBABILITY AND BAYES' FORMULA

In this section, we will address one of the most important uses of conditional probabilities. The basic idea is that if a probability is hard to compute directly, it might help to break the problem up in special cases, where in each special case the conditional probability is easier to compute. For example, suppose that you buy a used car in a city where street flooding due to heavy rainfall is a common problem. You know that roughly 5% of all used cars have previously been flood-damaged and estimate that 80% of such cars will later develop serious engine problems, whereas only 10% of used cars that are not flood-damaged develop the same problems. What is the probability that your car will later run into this kind of trouble?

Here is a situation where you can compute the probability in each of two different cases, flood-damaged or not flood-damaged (and no used-car dealer worth his salt would ever let you know which).

Let us first think about this in terms of proportions. Out of every 1000 cars sold, 50 are previously flood-damaged and of those, 80%, or 40 cars, will develop serious engine problems. Among the 950 that are not flood-damaged, we expect 10%, or 95 cars, to develop the same problems. Hence, we get a total of $40 + 95 = 135$ cars out of a 1000, and the probability of future problems is 0.135.

If we introduce the events $F = \{\text{flood-damaged}\}$ and $T = \{\text{trouble}\}$, we have argued that $P(T) = 0.135$. We also know that $P(F) = 0.05$, $P(F^c) = 0.95$, $P(T|F) = 0.80$, and $P(T|F^c) = 0.10$ and the probability we computed is in fact $0.80 \times 0.05 + 0.10 \times 0.95 = 0.135$. Our probability is a weighted average of the probability in the two different cases, flood-damaged or not, and the weights are the corresponding probabilities of the cases. The example illustrates the idea behind the following important result.

Theorem 1.1 (Law of Total Probability). *Let $B_1, B_2, \dots$ be a sequence of events such that*

- (a) $P(B_k) > 0$ for $k = 1, 2, \dots$
- (b) B_i and B_j are disjoint whenever $i \neq j$
- (c) $S = \bigcup_{k=1}^{\infty} B_k$

Then, for any event A , we have

$$P(A) = \sum_{k=1}^{\infty} P(A|B_k)P(B_k)$$

Condition (a) is a technical requirement to make sure that the conditional probabilities are defined, and you may recall that a collection of sets satisfying (b) and (c) is called a *partition* of S .

Proof. First note that

$$A = A \cap S = \bigcup_{k=1}^{\infty} (A \cap B_k)$$

by the distributive law for infinite unions. Since $A \cap B_1, A \cap B_2, \dots$ are pairwise disjoint, we get

$$P(A) = \sum_{k=1}^{\infty} P(A \cap B_k) = \sum_{k=1}^{\infty} P(A|B_k)P(B_k)$$

which proves the theorem. ■

By virtue of Proposition 1.2, we realize that the law of total probability is also true for a finite union of events, $B_1, \dots, B_n$. In particular, if we choose $n = 2$ and B_1 equal to some event B , then B_2 must equal B^c , and we get the following corollary.

Corollary 1.6. *If $0 < P(B) < 1$, then*

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

The verbal description of conditions (b) and (c) in Theorem 1.1 is that we are able to find different cases that *exclude each other* and *cover all possibilities*. This way of thinking about it is often sufficient to solve problems and saves us the effort to explicitly find the sample space and the partitioning events.

Example 1.43. A sign reads HOUSTON. Two letters are removed at random and then put back together again at random in the empty spaces. What is the probability that the sign still reads HOUSTON?

There are two different cases to consider: the case where two Os are chosen, in which case the text will always be correct and the case when two different letters are chosen, in which case the text will be correct when they are put back in their original order. Clearly, these two cases exclude each other and cover all possibilities and the assumptions in the law of total probability are satisfied. Hence, without spelling out exactly what the sample space is, we can define the events

$$\begin{aligned} A &= \{\text{the sign still reads HOUSTON}\} \\ B &= \{\text{two Os are chosen}\} \end{aligned}$$

and obtain

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

If the two letters are different, they are put back in their original order with probability $\frac{1}{2}$. Hence, the conditional probabilities are

$$P(A|B) = 1 \quad \text{and} \quad P(A|B^c) = \frac{1}{2}$$

and $P(B)$ is obtained by noting that we are choosing two letters out of seven without replacement and without regard to order. The total number of ways to do this is $\binom{7}{2} = 21$, and since there is only one way to choose the two Os, we get $P(B) = \frac{1}{21}$. This gives $P(B^c) = \frac{20}{21}$ and we get

$$P(A) = 1 \times \frac{1}{21} + \frac{1}{2} \times \frac{20}{21} = \frac{11}{21}$$

which is slightly larger than $\frac{1}{2}$, as was to be expected. $\square$

Example 1.44. In the United States, the overall risk of developing lung cancer is about 0.1%. Among the 20% of the population who are smokers, the risk is about 0.4%. What is the risk that a nonsmoker will develop lung cancer?

Introduce the events $C = \{\text{cancer}\}$ and $S = \{\text{smoker}\}$. The percentages above give $P(C) = 0.001$, $P(S) = 0.20$, and $P(C|S) = 0.004$, and we wish to compute $P(C|S^c)$. The law of total probability gives

$$P(C) = P(C|S)P(S) + P(C|S^c)P(S^c)$$

which with our numbers becomes

$$0.001 = 0.004 \times 0.20 + P(C|S^c) \times 0.80$$

which we solve for $P(C|S^c)$ to get

$$P(C|S^c) = 0.00025$$

in other words, a 250 in a million risk. $\square$

Example 1.45. Here is an example of a simple game of dice that does not seem to be to your advantage but turns out to be so.

Consider three dice, A , B , and C , numbered on their six sides as follows:

Die A 1, 1, 5, 5, 5, 5

Die B 3, 3, 3, 4, 4, 4

Die C 2, 2, 2, 2, 6, 6

The game now goes as follows. You and your opponent bet a dollar each, and you offer your opponent to choose any die and roll it. Next, you choose one of the remaining dice and roll it, and whoever gets the higher number wins the money. It seems that

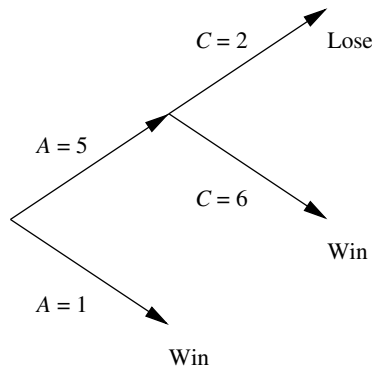


FIGURE 1.5 Tree diagram when you roll die C against die A in Example 1.45.

your opponent will have an edge since he gets to choose first. However, it turns out that once you know his choice, you can always choose so that your probability to win is more than one half! The reason for this is that, when rolled two by two against each other, these dice are such that on average A beats B , B beats C , and C beats A . The probabilities are (using A and C also to denote the numbers on dice A and C)

$$P(A \text{ beats } B) = P(A = 5) = \frac{2}{3}$$

$$P(B \text{ beats } C) = P(C = 2) = \frac{2}{3}$$

For the third case, we need to use the law of total probability and get

$$P(C \text{ beats } A) = P(C \text{ beats } A | A = 1) \times \frac{1}{3} + P(C \text{ beats } A | A = 5) \times \frac{2}{3}$$

$$= 1 \times \frac{1}{3} + P(C = 6) \times \frac{2}{3} = \frac{1}{3} + \frac{1}{3} \times \frac{2}{3} = \frac{5}{9}$$

which is also greater than $\frac{1}{2}$. Although you appear generous to let your opponent choose first, this is precisely what gives you the advantage.⁶ $\square$

Tree diagrams provide a nice way to illustrate the law of total probability. We represent each different case with a branch and look at the leaves to see which cases are of interest. We then compute the probability by first multiplying along each branch, then adding across the branches. See Figure 1.5 for an illustration of the situation in Example 1.45, where you roll die C against die A .

Sometimes, we need to condition repeatedly. For example, to compute $P(A|B)$, it may be necessary to condition further on some event C . Since a conditional probability is

⁶A completely deterministic version of this is the game “rock, paper, scissors,” in which you would always win if your opponent were to choose first. Games like these are called *nontransitive*.

a probability, this is nothing new, but the formula looks more complicated. We get

$$P(A|B) = P(A|B \cap C)P(C|B) + P(A|B \cap C^c)P(C^c|B) \quad (1.2)$$

where we note that every probability has the event B to the right of the conditioning bar. In Problem 81 you are asked to prove this.

Example 1.46 (Simpson's Paradox). In a by now famous study of gender bias at the University of California, Berkeley, it was noted that men were more likely than women to be admitted to graduate school. In 1 year, in the six largest majors, 45% of male applicants but only 30% of the female ones were admitted. To further study the bias, we divide the majors into two groups “difficult” and “easy,” referring to whether it is relatively difficult or easy to be admitted, not to the subjects themselves. It then turns out that in the “difficult” category, 26% of both men and women were admitted (actually even slightly above 26% for women and slightly below for men), so the bias obviously has to be in the other category. However, in the “easy” category, 80% of women but only 62% of men were admitted. Thus, there was no bias for difficult majors, a bias against men in easy majors, and an overall bias against women! Clearly there must be an error somewhere?

Consider a randomly chosen applicant. Let A be the event that the applicant is admitted, and let M and W be the events that the applicant is a man and a woman, respectively. We then have $P(A|M) = 0.45$ and $P(A|W) = 0.30$. Now also introduce the events D and E , for “difficult” and “easy” majors. By Table 1.3 we have, for men

$$P(A|M \cap D) = \frac{334}{1306} \approx 0.26 \quad \text{and} \quad P(A|M \cap E) = \frac{864}{1385} \approx 0.62$$

and for women

$$P(A|W \cap D) \approx 0.26 \quad \text{and} \quad P(A|W \cap E) \approx 0.80$$

and hence

$$P(A|M \cap D) = P(A|W \cap D) \quad \text{and} \quad P(A|M \cap E) < P(A|W \cap E)$$

but

$$P(A|M) > P(A|W)$$

TABLE 1.3 Numbers of Admitted, and Total Numbers (in Parentheses) of Male and Female Applicants in the Two Categories “Easy” and “Difficult” at UC Berkeley

	Male	Female
Easy major	864 (1385)	106 (133)
Difficult major	334 (1306)	451 (1702)

Thus, the conditional probabilities of being admitted are equal or higher for women in both categories but the overall probability for a woman to be admitted is lower than that of a man. Apparently, there was no error, but it still seems paradoxical. To resolve this, recall Equation 1.43, by which

$$P(A|W) = P(A|W \cap D)P(D|W) + P(A|W \cap E)P(E|W)$$

and

$$P(A|M) = P(A|M \cap D)P(D|M) + P(A|M \cap E)P(E|M)$$

and we realize that the explanation lies in the conditional probabilities $P(D|W)$, $P(E|W)$, $P(D|M)$, and $P(E|M)$, which reflect how men and women choose their majors. The probabilities that a man chooses a difficult major and an easy major, respectively, are

$$P(D|M) = \frac{1306}{2691} \approx 0.49 \quad \text{and} \quad P(E|M) \approx 0.51$$

and the corresponding probabilities for women are

$$P(D|W) = \frac{1702}{1835} \approx 0.93 \quad \text{and} \quad P(E|W) \approx 0.07$$

Thus, women almost exclusively applied for difficult majors, whereas men applied equally for difficult and easy majors, and this is the resolution of the paradox. Was it harder for women to be admitted? Yes. Was this due to gender discrimination? No. The effect on admission rates that was initially attributed to gender bias was really due to choice of major, an example of what statisticians call *confounding of factors*. The effect of gender on choice of major is a completely different issue. $\square$

The last example is a version of what is known as *Simpson's paradox*. If we formulate it as a mathematical problem, it completely loses its charm. The question then becomes if it is possible to find numbers A , a , B , b , p , and q , all between 0 and 1, such that

$$A > a \quad \text{and} \quad B > b$$

and

$$pA + (1 - p)B < qa + (1 - q)b$$

No problems here. Let $A > a > B > b$, and choose p sufficiently close to 0 and q sufficiently close to 1. Ask your mathematician friends this question, and also if there is something strange about the Berkeley admissions data, and don't be surprised if you get the answer "Yes" to both questions!

1.6.1 Bayes' Formula

We next turn to the situation when we know conditional probabilities in one direction but want to compute conditional probabilities “backward.” The following result is helpful.

Proposition 1.11 (Bayes' Formula). Under the same assumptions as in the law of total probability and if $P(A) > 0$, then for any event B_j , we have

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{k=1}^{\infty} P(A|B_k)P(B_k)}$$

Proof. Note that, by the law of total probability, the denominator is nothing but $P(A)$, and hence we must show that

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{P(A)}$$

which is to say that

$$P(B_j|A)P(A) = P(A|B_j)P(B_j)$$

which is true since both sides equal $P(A \cap B_j)$, by the definition of conditional probability. ■

Again, the obvious analog for finitely many conditioning events holds, and in particular we state the case of two such events, B and B^c , as a corollary.

Corollary 1.7. If $0 < P(B) < 1$ and $P(A) > 0$, then

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

Example 1.47. The polygraph is an instrument used to detect physiological signs of deceptive behavior. Although it is often pointed out that the polygraph is not a lie detector, this is probably the way most of us think of it. For the purpose of this example, let us retain this notion. It is debated how accurate a polygraph test is, but there are several reports of accuracies above 95% (and as a counterweight, a Web site that gladly claims “Don’t worry, the polygraph can be beaten rather easily!”). Let us assume that the polygraph test is indeed very accurate and that it decides “lie” or “truth” correctly with probability 0.95. Now consider a randomly chosen individual

who takes the test and is determined to be lying. What is the probability that this person did indeed lie?

First, the probability is *not* 0.95. Introduce the events

$L = \{\text{the person tells a lie}\}$

$L_P = \{\text{the polygraph reading says the person is lying}\}$

and let $T = L^c$ and $T_P = L_P^c$. We are given the conditional probabilities $P(L_P|L) = P(T_P|T) = 0.95$, but what we want is $P(L|L_P)$. By Bayes' formula

$$\begin{aligned} P(L|L_P) &= \frac{P(L_P|L)P(L)}{P(L_P|L)P(L) + P(L_P|T)P(T)} \\ &= \frac{0.95P(L)}{0.95P(L) + 0.05(1 - P(L))} \end{aligned}$$

and to be able to finish the computation we need to know the probability that a randomly selected person would lie on the test. Suppose that we are dealing with a largely honest population; let us say that one out of a thousand would tell a lie in the given situation. Then $P(L) = 0.001$, and we get

$$P(L|L_P) = \frac{0.95 \times 0.001}{0.95 \times 0.001 + 0.05 \times 0.999} \approx 0.02$$

and the probability that the person actually lied is only 0.02. Since lying is so rare, most detected lies actually stem from errors, not actual lies. One way to understand this is to imagine that a large number of, say, 100,000, people are tested. We then expect 100 liars and of those, 95 will be discovered. Among the remaining 99,900 truth-tellers, we expect 5%, or 4995 individuals to be misclassified as liars. Hence, out of a total of $95 + 4995 = 5090$ individuals who are classified as liars, only 95, or 2% actually are liars. A truth misclassified as a lie is called a “false-positive” and in this case, we say that the *false-positive rate* is 98%. $\square$

In the last example, there are two types of errors we can make: classifying a lie as truth, and vice versa. The probability $P(L_P|L)$ to correctly classify a lie as a lie is called the *sensitivity* of the procedure. Obviously, we want the sensitivity to be high but with increased sensitivity we may risk to misclassify more truths as lies as well. Another probability of interest is therefore the *specificity*, namely, the probability $P(T_P|T)$ that a truth is correctly classified as truth. For an extreme but illustrative example, we can achieve maximum sensitivity by classifying all statements as lies; however, the specificity is then 0. Likewise, we can achieve maximum specificity by classifying all statements as truths but then instead getting sensitivity 0. The terms are borrowed from the field of medical testing for illnesses where good procedures should be both sensitive to detect an illness but also specific for that illness. For example, using high fever to diagnose measles would have high sensitivity (not many

cases of measles will go undetected) but low specificity (many other diseases cause high fever and will be misclassified as measles).

Another probability of interest in any kind of testing situation is the false-positive rate, mentioned above. In the lie-detector example, it is $P(T|L_p)$, the probability that a detected lie is actually a truth. Also, the *false-negative rate* is $P(L|T_p)$, the probability that a detected truth is actually a lie. The sensitivity, specificity, false-positive rate, and false-negative rate are related via Bayes' formula where we also need to know the *base rate*, namely, the unconditional probability $P(T)$ of telling a lie (or having a disease). For typical examples from medical testing, see Problem 92 and subsequent problems.

Example 1.48 (The Monty Hall Problem). This problem has become a modern classic and was hotly discussed after it first appeared in the column "Ask Marilyn" in *Parade Magazine* in 1991. The problem was inspired by the game show "Let's Make a Deal" with host Monty Hall, and it goes like this. You are given the choice of three doors. Behind one door is a car; behind the others are goats. You pick a door without opening it, and the host opens another door that reveals a goat. He then gives you the choice to either open your door and keep what is behind it, or switch to the remaining door and take what is there. Is it to your advantage to switch?

At first glance, it would not seem to make a difference whether you stay or switch since the car is either behind your door or behind the remaining door. However, this is incorrect, at least if we make some reasonable assumptions. To solve the problem, we assume that the car and goats are placed at random behind the doors and that the host *always opens a door and shows a goat*. Let us further assume that in the case where you have chosen the car, he chooses which door to open at random. Now introduce the two events

$$C = \{\text{you chose the car}\}$$

$$G = \{\text{he shows a goat}\}$$

so that the probability to win after switching is $1 - P(C|G)$. But

$$P(C|G) = \frac{P(G|C)P(C)}{P(G|C)P(C) + P(G|C^c)P(C^c)} = P(C) = \frac{1}{3}$$

since $P(G|C) = P(G|C^c) = 1$. Thus, if you switch, you win the car with probability $\frac{2}{3}$, so switching is to your advantage. Note that the events G and C are in fact independent.

Intuitively, since you know that the host will always show you a goat, there is no additional information when he does. Since there are two goats and the host will always show one of them, to choose a door and then switch is equivalent to choosing the *two other doors* and telling the host to open one of them and show you a goat. Your chance of winning the car is then $\frac{2}{3}$.

One variant of the problem that has been suggested to make it easier to understand is to assume that there are not 3 but 1000 doors. One has a car, and 999 have goats.

Once you have chosen, the host opens 998 doors and shows you 998 goats. Given how unlikely it is that you found the car in the first pick, is it not obvious that you should now switch to the remaining door? You could also use one of the several computer simulations that are available online, or write your own. Still not convinced? Ask Marilyn. $\square$

Example 1.49 (The Monty Hall Problem Continued). Suppose that you are playing “Let’s Make a Deal” and have made your choice when the host suddenly realizes that he has forgotten where the car is. Since the show must go on, he keeps a straight face, takes a chance, and opens a door that reveals a goat. Is it to your advantage to switch?

Although the situation looks the same from your perspective, it is actually different since it could have happened that the host revealed the car. With C and G as above, Bayes’ formula now gives

$$\begin{aligned} P(C|G) &= \frac{P(G|C)P(C)}{P(G|C)P(C) + P(G|C^c)P(C^c)} \\ &= \frac{1 \times (1/3)}{1 \times (1/3) + (1/2) \times (2/3)} = \frac{1}{2} \end{aligned}$$

so it makes no difference whether you stay or switch. In this case, the showing of a goat behind the open door actually does give some additional information, and G and C are no longer independent. $\square$

Example 1.50 (The Island Problem). Probability theory is frequently used in courts of law, especially when DNA evidence is considered. As an example, consider the following situation. A person is murdered on an island, and the murderer must be one of the n remaining islanders. DNA evidence on the scene reveals that the murderer has a particular genotype that is known to exist in a proportion p in the general population, and we assume that the islanders’ genotypes are independent. Crime investigators start screening all islanders for their genotypes. The first one who is tested is Mr Joe Bloggs, who turns out to have the murderer’s genotype. What is the probability that he is guilty?

To solve this, we introduce the events

$G = \{\text{Mr Bloggs is guilty}\}$

$B = \{\text{Mr Bloggs’ genotype is found at the scene of the crime}\}$

so that we are asking for the probability $P(G|B)$. By Bayes’ formula

$$P(G|B) = \frac{P(B|G)P(G)}{P(B|G)P(G) + P(B|G^c)P(G^c)}$$

Here, $P(G)$ is the probability that Mr Bloggs is guilty before any genotyping has been done, and if we assume that there is no reason to suspect any particular person more than anyone else, it is reasonable to let $P(G) = \frac{1}{n}$. If Mr Bloggs is guilty, then his genotype is certain to show up at the scene of the crime, and we have $P(B|G) = 1$. If Mr Bloggs is innocent, his genotype can still show up by chance, which gives $P(B|G^c) = p$, the proportion of his genotype in the population. All put together, we get

$$P(G|B) = \frac{1 \times 1/n}{1 \times 1/n + p \times (n-1)/n} = \frac{1}{1 + (n-1)p}$$

as the probability that Mr Bloggs is guilty. $\square$

The last problem is a simple example of the general problem of how to quantify the weight of evidence in forensic identification. This “island problem” has been analyzed and discussed by lawyers and probabilists and different approaches have shown to give different results (not all correct).⁷ We will return to this in more detail in Section 2.5. For now, let us present a simple example that demonstrates how calculations can go agley.

Example 1.51. You know that your new neighbors have two children. One night you hear a stone thrown at your window and you see a child running from your yard into the neighbor’s house. It is dark, and the only thing you can see for certain is that the child is a boy. The next day you walk over to the neighbor’s house and ring the doorbell. A boy opens the door. What is the probability that he is guilty?

We will do this in two different ways. First approach: If the other child is a girl, you know that the boy is guilty and if the other child is a boy, the boy who opened the door is equally likely to be guilty or not guilty. Thus, with

$$G = \{\text{child who opened the door is guilty}\}$$

we condition on the gender of the other child and recall Example 1.39 to obtain

$$\begin{aligned} P(G) &= P(G|\text{boy})P(\text{boy}) + P(G|\text{girl})P(\text{girl}) \\ &= \frac{1}{2} \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{3}{4}. \end{aligned}$$

Second approach: Note how the situation is similar to that in the previous example, with genotype replaced by gender and Mr Bloggs replaced by the child who opened

⁷The island problem is made up (yes, really!), but there is a famous real case, *People versus Collins*, in which a couple in Los Angeles was first convicted of a crime, on the basis of circumstantial evidence, and later acquitted by the California Supreme Court. Both the initial verdict and the appeal were based on (questionable) probability arguments.

the door. In that formulation, we have $n = 2$ and $p = \frac{1}{2}$ and we get

$$P(\text{child who opened the door is guilty}) = \frac{1}{1 + 1 \times (1/2)} = \frac{2}{3}$$

There we go again; different methods give different results! As usual, we need to be very careful with which events we condition on. Let us assume that each child is equally likely to decide to go out and throw a stone at your window and that each child is equally likely to open the door. For each gender combination of two children, we thus choose at random who is guilty and who opens the door, so that each gender combination is split up into four equally likely cases. Let us use the subscript d for the child who opened the door and the superscript g for the child who is guilty. The sample space consists of the 16 equally likely outcomes

$$S = \{b_d^g b, b_d b^g, b^g b_d, b b_d^g, b_d^g g, b_d g^g, b^g g_d, b g_d^g, \\ g_d^g b, g_d b^g, g^g b_d, g b_d^g, g_d^g g, g_d g^g, g^g g_d, g g_d^g\}$$

and the event that the child who opened the door is guilty is

$$G = \{b_d^g b, b b_d^g, b_d^g g, g b_d^g, b_d^g g, g_d^g b, g_d^g g, g g_d^g\}$$

What event do we condition on? We know two things: that the guilty child is a boy and that a boy opened the door. These events are

$$A = \{b_d^g b, b_d b^g, b^g b_d, b b_d^g, b_d^g g, b^g g_d, g_d b^g, g b_d^g\} \\ B = \{b_d^g b, b_d b^g, b^g b_d, b b_d^g, b_d^g g, b_d g^g, g^g b_d, g b_d^g\}$$

and we condition on their intersection

$$A \cap B = \{b_d^g b, b_d b^g, b^g b_d, b b_d^g, b_d^g g, g b_d^g\}$$

Since four of these six outcomes are in G and the distribution on S is uniform, we get

$$P(\text{child who opened the door is guilty}) = P(G|A \cap B) = \frac{2}{3}$$

in accordance with the previous example.

The first approach gives the wrong solution but why? When we computed the probabilities $P(\text{boy})$ and $P(\text{girl})$, we implicitly conditioned on event B above but forgot to also condition on A . What we need to do is to compute $P(\text{boy})$ as

$$P(\text{other child is a boy} | A \cap B) = \frac{2}{3}$$

and not $\frac{1}{2}$. Note how the conditional probability that the other child is a boy is higher now that we also know that the guilty child is a boy. This is quite subtle and resembles the situation in Example 1.39, in the sense that we need to be careful to condition on precisely the information we have, no more and no less. We can now state the correct

version of the first solution. Everything must be computed conditioned on the event $A \cap B$, but for ease of notation let us not write this conditioning out explicitly. We get

$$\begin{aligned} P(G) &= P(G|\text{boy})P(\text{boy}) + P(G|\text{girl})P(\text{girl}) \\ &= \frac{1}{2} \times \frac{2}{3} + 1 \times \frac{1}{3} = \frac{2}{3} \end{aligned}$$

just as we should. For a variant, see Problem 99. $\square$

Example 1.52. Consider the previous example and also assume that on your way over to the neighbor's, you meet another neighbor who tells you that she saw the mother of the family take a walk with a boy a few days ago. If a boy opens the door, what is the probability that he is guilty?

By now you know how to solve this. In the previous sample space, split each outcome further into two, marking who the mother took a walk with, and proceed. The sample space now has 32 outcomes, and we will suggest a more convenient approach. We can view the various sightings of a boy as repeated sampling with replacement from a randomly chosen family. Let us convert this into a problem about black and white balls in urns.

Consider three urns, containing two balls each, such that the k th urn contains k black balls, $k = 0, 1, 2$. We first choose an urn according to the probabilities $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$ (think of the gender combinations above) and then pick balls with replacement and note their colors. If we do this j times and get only black balls, what is the probability that we have chosen the urn with only black balls? Let

$$\begin{aligned} B &= \{\text{get only black balls}\} \\ U_k &= \{\text{the } k\text{th urn chosen}\}, \quad k = 0, 1, 2 \end{aligned}$$

and compute $P(U_2|B)$. The reversed probabilities are

$$P(B|U_0) = 0, \quad P(B|U_1) = \frac{1}{2^j}, \quad P(B|U_2) = 1$$

and Bayes' formula gives

$$\begin{aligned} P(U_2|B) &= \frac{P(B|U_2)P(U_2)}{P(B|U_1)P(U_1) + P(B|U_2)P(U_2)} \\ &= \frac{1 \times (1/4)}{(1/2^j) \times (1/2) + 1 \times (1/4)} = \frac{2^{j-1}}{2^{j-1} + 1} \end{aligned}$$

In our examples with families and their children, we let urns represent families and black and white balls represent genders. Consider the probability that the other child has the same gender as the observed child. In Example 1.39, we have $j = 1$, which gives probability $\frac{1}{2}$ and in Example 1.51 we have $j = 2$, which gives probability $\frac{2}{3}$. Finally, in this example we have $j = 3$ and probability $\frac{4}{5}$. The more observations we have on boys, the stronger our belief that both children are boys. $\square$

		From mother:	
		A	a
From father:	A	AA	Aa
	a	aA	aa

FIGURE 1.6 A Punnett square illustrating possible genotypes.

1.6.2 Genetics and Probability

Genetics is a science where probability theory is extremely useful. Recall that genes occur in pairs where one copy is inherited from the mother and one from the father. Suppose that a particular gene has two different *alleles* (variants) called A and a . An individual can then have either of the three *genotypes* AA , Aa , and aa . If the parents both have genotype Aa , what is the probability that their child gets the same genotype?

We assume that each of the two gene copies from each parent is equally likely to be passed on to the child and that genes from the father and the mother are inherited independently. There are then the four equally likely outcomes illustrated in Figure 1.6, and the probability that the child also has genotype Aa is $\frac{1}{2}$ (order has no meaning here, so Aa and aA are the same). Each of the genotypes AA and aa has probability $\frac{1}{4}$. The square in the figure is an example of a *Punnett square*.

Example 1.53. An allele is said to be *recessive* if it is required to exist in two copies to be expressed and *dominant* if one copy is enough. For example, the hereditary disease *cystic fibrosis* (CF) is caused by a recessive allele of a particular gene. Let us denote this allele C and the healthy allele H so that only individuals with genotype CC get the disease. Individuals with genotype CH are *carriers*, that is, they have the disease-causing allele but are healthy. It is estimated that approximately 1 in 25 individuals are carriers (among people of central and northern European descent; it is much less common in other ethnic groups). Given this information, what is the probability that a newborn of healthy parents has CF?

Introduce the events

$$D = \{\text{newborn has CF}\}$$

$$B = \{\text{both parents are carriers}\}$$

so that

$$P(D) = P(D|B)P(B)$$

since B^c is the event that at least one parent has genotype HH , in which case the baby will also be healthy. Assuming that the mother's and father's genotypes are independent, we get

$$P(B) = \frac{1}{25} \times \frac{1}{25} = \frac{1}{625}$$

and since the child will get the disease only if it inherits the C allele from each parent, we get $P(D|B) = \frac{1}{4}$, which gives

$$P(D) = \frac{1}{625} \times \frac{1}{4} = \frac{1}{2500}$$

In other words, the *incidence* of CF among newborns is 1 in 2500, or 0.04%.

Now consider a family with one child where we know that both parents are healthy, that the mother is a carrier of the disease allele and nothing is known about the father's genotype. What is the probability that the child neither is a carrier nor has the disease?

Let E be the event we are interested in. The mother's genotype is CH , and we condition on the father's genotype to obtain

$$\begin{aligned} P(E) &= P(E|CH)P(CH) + P(E|HH)P(HH) \\ &= \frac{1}{4} \times \frac{1}{25} + \frac{1}{2} \times \frac{24}{25} = 0.49 \end{aligned}$$

where we figure out the conditional probabilities with Punnett squares. See the problem section at the end of the chapter for more on genetics. $\square$

1.6.3 Recursive Methods

Certain probability problems can be solved elegantly with recursive methods, involving the law of total probability. The general idea is to condition on a number of cases that can either be solved explicitly or lead back to the original problem. We will illustrate this in a number of examples.

Example 1.54. In the final scene of the classic 1966 Sergio Leone movie *The Good, the Bad, and the Ugly*, the three title characters, also known as “Blondie,” “Angel Eyes,” and “Tucó,” stand in a cemetery, guns in holsters, ready to draw. Let us interfere slightly with the script and assume that Blondie always hits his target, Angel Eyes hits with probability 0.9, and Tucó with probability 0.5. Let us also suppose that they take turns in shooting, that whomever is shot at shoots next (unless he is hit), and that Tucó starts. What strategy maximizes his probability of survival?

Introduce the events

$$S = \{\text{Tucó survives}\}$$

$$H = \{\text{Tucó hits his target}\}$$

Let us first suppose that Tuco tries to kill Blondie. If he fails, Blondie kills Angel Eyes, and Tuco gets one shot at Blondie. We thus have

$$P(S) = P(S|H)P(H) + P(S|H^c)P(H^c) = P(S|H)\frac{1}{2} + \frac{1}{4}$$

where we need to find $P(S|H)$, the probability that Tuco survives a shootout with Angel Eyes, who gets the first shot. If we assume an infinite supply of bullets (hey, it's a Clint Eastwood movie!), we can solve this recursively. Note how this is repeated conditioning, as in Equation (1.43), but let us ease the notation and rename the event that Tuco survives the shootout T . Now let $p = P(T)$ and condition on the three events

$$\begin{aligned} A &= \{\text{Angel Eyes hits}\} \\ B &= \{\text{Angel Eyes misses, Tuco hits}\} \\ C &= \{\text{Angel Eyes misses, Tuco misses}\} \end{aligned}$$

to obtain

$$p = P(T|A)P(A) + P(T|B)P(B) + P(T|C)P(C)$$

where $P(A) = 0.9$, $P(B) = 0.1 \times 0.5 = 0.05$, $P(C) = 0.1 \times 0.5 = 0.05$, $P(T|A) = 0$, and $P(T|B) = 1$. To find $P(T|C)$, note that if both Angel Eyes and Tuco have missed their shots, they start over from the beginning and hence $P(T|C) = p$. This gives

$$p = 0.05 + 0.05p$$

which gives $p = 0.05/0.95$, and with this strategy, Tuco has survival probability

$$P(S) = \frac{0.05}{0.95} \times 0.5 + 0.25 \approx 0.28$$

Next, suppose that Tuco tries to kill Angel Eyes. If he succeeds, he faces certain death as Blondie shoots him. If he fails, Angel Eyes will try to kill Blondie to maximize his own probability of survival. If Angel Eyes fails, Blondie kills him for the same reason and Tuco again gets one last shot at Blondie. Tuco surviving this scenario has probability $0.5 \times 0.1 \times 0.5 = 0.025$. If Angel Eyes succeeds and kills Blondie, Tuco must again survive a shootout with Angel Eyes but this time, Tuco gets to start. By an argument similar to that stated above, his probability to survive the shootout is $p = 0.5 + 0.05p$ that gives $p = 0.5/0.95$ and Tuco's survival probability is

$$P(S) = 0.025 + 0.5 \times 0.9 \times \frac{0.5}{0.95} \approx 0.26$$

not quite as good as with the first strategy.

Notice, however, that Tuco really gains from missing his shot, letting the two better shots fight it out first. The smartest thing he can do is to miss on purpose! If he aims at Blondie and misses, Blondie kills Angel Eyes and Tuco gets one last shot at Blondie.

His survival probability is 0.5. An even better strategy is to aim at Angel Eyes, miss on purpose, and give Angel Eyes a chance to kill Blondie. If Angel Eyes fails, he is a dead man and Tuco gets one last shot at Blondie. If Angel Eyes succeeds, Tuco again needs to survive the shootout which, as we just saw, has probability $p = 0.5/0.95$ and his overall survival probability is

$$P(S) = 0.1 \times 0.5 + 0.9 \times \frac{0.5}{0.95} \approx 0.52$$

When Fredric Mosteller presents a similar problem in his 1965 book *Fifty Challenging Problems in Probability* [1], he expresses some worry over the possibly unethical dueling conduct to miss on purpose. In the case of Tuco, we can safely disregard any such ethical considerations. $\square$

Example 1.55. The shootout between Tuco and Angel Eyes in the previous example is a special case of the following situation: Consider an experiment where the events A and B are disjoint and repeat the experiment until either A or B occurs. What is the probability that A occurs before B ?

First, by Corollary 1.5, we will sooner or later get either A or B . Let C be the event that A occurs before B , let $p = P(C)$, and condition on the first trial. If we get A , we have A before B for certain and if we get B , we do not. If we get neither, that is, get $(A \cup B)^c$, we start over. The law of total probability now gives

$$\begin{aligned} p &= P(C|A)P(A) + P(C|B)P(B) + P(C|(A \cup B)^c)P(A \cup B)^c \\ &= P(A) + p(1 - P(A \cup B)) \\ &= P(A) + p(1 - (P(A) + P(B))) \end{aligned}$$

and we have established an equation for p . Solving it gives

$$p = \frac{P(A)}{P(A) + P(B)}$$

$\square$

Example 1.56. Recall that a single *game* in tennis is won by the first player to win four points but that it must also be won by a margin of at least two points. If no player has won after six played points, they are at *deuce* and the first to get two points ahead wins the game. Suppose that Ann is the server and has probability p of winning a single point against Bob, and suppose that points are won independent of each other. If the players are at *deuce*, what is the probability that Ann wins the game?

We are waiting for the first player to win two consecutive points from *deuce*, so let us introduce the events

$$\begin{aligned} A &= \{\text{Ann wins two consecutive points from deuce}\} \\ B &= \{\text{Bob wins two consecutive points from deuce}\} \end{aligned}$$

with the remaining possibility that they win a point each, in which case they are back at deuce. By independence of consecutive points, $P(A) = p^2$ and $P(B) = (1 - p)^2$, and by Example 1.55 we get

$$P(\text{Ann wins}) = \frac{p^2}{p^2 + (1 - p)^2}$$

□

Example 1.57. The next sports application is to the game of badminton. The scoring system is such that you can score a point only when you win a rally as the server. If you win a rally as the receiver, the score is unchanged, but you get to serve and thus the opportunity to score. Suppose that Ann wins a rally against Bob with probability p , regardless of who serves (a reasonable assumption in badminton but would, of course, not be so in tennis, where the server has a big advantage). What is the probability that Ann scores the next point if she is the server?

If Ann is the server, the next point is scored either when she wins a rally as server or loses two consecutive rallies starting from being server. In the remaining case, the players will start over from Ann being the server with no points scored yet. Hence, we can apply the formula from Example 1.55 to the events

$$A = \{\text{Ann wins a rally as server}\}$$

$$B = \{\text{Ann loses two consecutive rallies as server}\}$$

to obtain

$$P(\text{Ann scores next point}) = \frac{P(A)}{P(A) + P(B)} = \frac{p}{p + (1 - p)^2}$$

If the players are equally good, so that $p = \frac{1}{2}$, the server thus has a $\frac{2}{3}$ probability to score the next point. □

Example 1.58 (Gambler's Ruin). Next, Ann and Bob play a game where a fair coin is flipped repeatedly. If it shows heads, Ann pays Bob one dollar, otherwise Bob pays Ann one dollar. If Ann starts with a dollars and Bob with b dollars, what is the probability that Ann ends up winning all the money and Bob is ruined?

Introduce the event

$$A = \{\text{Ann wins all the money}\}$$

and let p_a be the probability of A if Ann's initial fortune is a . Thinking a few minutes makes us realize that it is quite complicated to compute p_a directly. Instead, let us condition on the first flip and note that if it is heads, the game starts over with the

new initial fortunes $a - 1$ and $b + 1$, and if it is tails, the new fortunes are $a + 1$ and $b - 1$. Introduce the events

$$H = \{\text{heads in first flip}\} \quad \text{and} \quad T = \{\text{tails in first flip}\}$$

and apply the law of total probability to get

$$p_a = P(A|H)\frac{1}{2} + P(A|T)\frac{1}{2} = \frac{1}{2}(p_{a-1} + p_{a+1})$$

or equivalently

$$p_{a+1} = 2p_a - p_{a-1}$$

First note that $p_0 = 0$ and let $a = 1$ to obtain

$$p_2 = 2p_1$$

With $a = 2$ we get

$$p_3 = 2p_2 - p_1 = 3p_1$$

and we find the general relation

$$p_a = ap_1$$

Now, $p_{a+b} = 1$, and hence

$$p_1 = \frac{1}{a+b}$$

which finally gives the solution

$$P(\text{Ann wins all the money}) = \frac{a}{a+b}$$

By symmetry, the same argument applies to give

$$P(\text{Bob wins all the money}) = \frac{b}{a+b}$$

Note that this means that the probability that *somebody* wins is 1, which excludes the possibility that the game goes on forever, something we cannot immediately rule out.

This *gambler's ruin* problem is an example of a *random walk*. We may think of a particle that in each step decides to go up or down (or, if you prefer, left/right), and does so independent of its previous path.⁸ We can view the position after n steps as Ann's total gain, so if the walk starts in 0, Ann has won the game when it hits b , and she has lost when it hits $-a$. We refer to $-a$ and b as *absorbing barriers* (see Figure 1.7). $\square$

⁸A more romantic allegory is that of a drunken Dutchman who staggers back and forth until he either is back in his favorite *bruine cafe* or falls into the canal.

wins and p_a the probability of A if she starts with a dollars. For ease of notation, let $q = 1 - p$, the probability that Ann loses a round. We get

$$p_a = P(A|H)P(H) + P(A|T)P(T) = p_{a-1}q + p_{a+1}p$$

which gives

$$p_{a+1} = \frac{1}{p}(p_a - qp_{a-1})$$

First let $a = 1$. Since $p_0 = 0$, we get

$$p_2 = \frac{1}{p}p_1$$

which we rewrite as

$$p_2 = \left(1 + \frac{q}{p}\right)p_1$$

For $a = 2$, we get

$$\begin{aligned} p_3 &= \frac{1}{p}(p_2 - qp_1) \\ &= \frac{1}{p}\left(1 + \frac{q}{p} - q\right)p_1 = \left(1 + \frac{q}{p} + \left(\frac{q}{p}\right)^2\right)p_1 \end{aligned}$$

and the general formula emerges as

$$p_a = \left(1 + \frac{q}{p} + \left(\frac{q}{p}\right)^2 + \cdots + \left(\frac{q}{p}\right)^{a-1}\right)p_1 = \frac{1 - (q/p)^a}{1 - (q/p)}p_1$$

Finally, we use $p_{a+b} = 1$ to obtain

$$\frac{1 - (q/p)^{a+b}}{1 - (q/p)}p_1 = 1$$

which gives

$$p_1 = \frac{1 - (q/p)}{1 - (q/p)^{a+b}}$$

and the probability that Ann wins, starting from a fortune of a dollars, is thus

$$p_a = \frac{1 - (q/p)^a}{1 - (q/p)^{a+b}}$$

if $p \neq \frac{1}{2}$. The game is unfair to Ann if $p < \frac{1}{2}$ and to Bob if $p > \frac{1}{2}$. It is interesting to note that the change in winning probabilities can be dramatic for small changes of p .

For example, if the players start with 20 dollars each and $p = \frac{1}{2}$, they are equally likely to win in the end. Now change p to 0.55, so that Ann has a slight edge. Then

$$p_{20} = \frac{1 - (0.45/0.55)^{20}}{1 - (0.45/0.55)^{40}} \approx 0.98$$

so Ann is almost certain to win. See Problem 106 for an interesting application to roulette. $\square$

Example 1.61 (Penney-ante). We finish the section with a game named for Walter Penney, who in 1969 described it in an article in *Journal of Recreational Mathematics*. If a fair coin is flipped three times, there are eight outcomes

$$HHH, HHT, HTH, THH, HTT, THT, TTH, TTT$$

each of which has probability $\frac{1}{8}$. You now suggest the following game to your friend John.⁹ You bet \$1 each, he gets to choose one of the eight patterns, and you choose another. A coin is flipped repeatedly, and the sequence of heads and tails is recorded. Whoever first sees his sequence come up wins. Since all patterns are equally likely to come up in a sequence of three flips, this game seems fair. However, it turns out that after John has chosen his pattern, you can always choose so that your chance of winning is at least $\frac{2}{3}$!

The idea is to always let your sequence end with the two symbols that his begins with. Intuitively, this means that whenever his pattern is about to come up, there is a good chance that yours has come up already. For example, if he chooses *HHH*, you choose *THH*, and the only way in which he can win is if the first three flips give heads. Otherwise, the sequence *HHH* cannot appear without having a *T* before it, and thus your pattern *THH* has appeared. With these choices, your probability to win is $\frac{7}{8}$.

The general strategy is to let his first two be your last two, and never choose a palindrome. Suppose that John chooses *HTH* so that according to the strategy, you choose *HHT*. Let us calculate your probability of winning.

Let A be the event that you win, and let p be the probability of A . To find p , we condition on the first flip. If this is *T*, the game starts over, and hence $P(A|T) = p$. If it is *H*, we condition further on the second flip. If this is *H*, you win (if we start with *HH*, then *HHT* must come before *HTH*), and if it is *T*, we condition further on the third flip. If this is *H*, the full sequence is *HTH*, and you have lost. If it is *T*, the full sequence is *HTT* and the game starts over. See the tree diagram in Figure 1.8 for

⁹Named after John Haigh, who in his splendid book *Taking Chances: Winning with Probability* [3] describes this game and names the loser Doyle after Doyle Lonnegan, victim in the 1973 movie *The Sting*. I feel that Doyle has now lost enough and in this way let him get a small revenge. Hopefully, John's book has sold so well that he is able to take the loss.

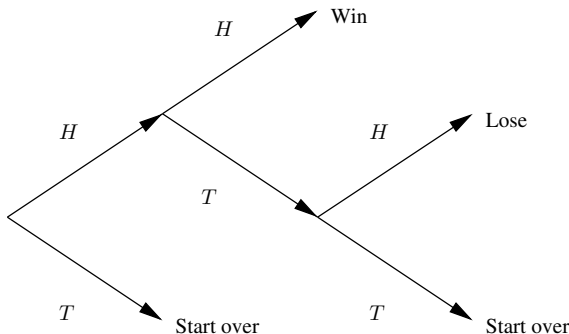


FIGURE 1.8 The four possible cases in Penney-ante when HHT competes with HTH .

an illustration of the possible cases. The law of total probability gives (ignoring the case in which you lose)

$$\begin{aligned}
 p &= P(A|T)P(T) + P(A|HH)P(HH) + P(A|HTT)P(HTT) \\
 &= p \times \frac{1}{2} + 1 \times \frac{1}{4} + p \times \frac{1}{8} = \frac{2 + 5p}{8}
 \end{aligned}$$

which we solve for p to get $p = \frac{2}{3}$. Just as in the dice game in Example 1.45, your apparent generosity to let your opponent choose first is precisely what gives you the advantage. See also Problem 108. $\square$

PROBLEMS

Section 1.2. Sample Spaces and Events

- Suggest sample spaces for the following experiments: **(a)** Three dice are rolled and their sum computed. **(b)** Two real numbers between 0 and 1 are chosen. **(c)** An American is chosen at random and is classified according to gender and age. **(d)** Two different integers are chosen between 1 and 10 and are listed in increasing order. **(e)** Two points are chosen at random on a yardstick and the distance between them is measured.
- Suggest a sample space for Example 1.8.
- Consider the experiment to toss a coin three times and count the number of heads. Which of the following sample spaces can be used to describe this experiment?
 - $S = \{H, T\}$
 - $S = \{HHH, TTT\}$
 - $S = \{0, 1, 2, 3\}$
 - $S = \{1, 2, 3\}$
 - $S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$

4. Let A , B , and C be three events. Express the following events in terms of A , B , and C : **(a)** exactly one of the events occurs. **(b)** None of the events occurs. **(c)** At least one of the events occurs. **(d)** All of the events occur.
5. The Stanley Cup final is played in best of seven games. Suppose that the good old days are brought back and that the final is played between the Boston Bruins and Montreal Canadiens. Let B_k be the event that Boston wins the k th game and describe the following events in terms of the B_k : **(a)** Boston wins game 1, **(b)** Boston loses game 1 and wins games 2 and 3, **(c)** Boston wins the series without losing any games, **(d)** Boston wins the series with one loss, and **(e)** Boston wins the first three games and loses the series.

Section 1.3. The Axioms of Probability

6. A certain thick and asymmetric coin is tossed and the probability that it lands on the edge is 0.1. If it does not land on the edge, it is twice as likely to show heads as tails. What is the probability that it shows heads?
7. Let A and B be two events such that $P(A) = 0.3$, $P(A \cup B) = 0.5$, and $P(A \cap B) = 0.2$. Find **(a)** $P(B)$, **(b)** the probability that A but not B occurs, **(c)** $P(A \cap B^c)$, **(d)** $P(A^c)$, **(e)** the probability that B does not occur, and **(f)** the probability that neither A nor B occurs.
8. Let A be the event that it rains on Saturday and B the event that it rains on Sunday. Suppose that $P(A) = P(B) = 0.5$. Furthermore, let p denote the probability that it rains on both days. Express the probabilities of the following events as functions of p : **(a)** it rains on Saturday but not Sunday. **(b)** It rains on one day but not the other. **(c)** It does not rain at all during the weekend.
9. The probability in Problem 8(b) is a decreasing function of p . Explain this intuitively.
10. People are asked to assign probabilities to the events “rain on Saturday,” “rain on Sunday,” “rain both days,” and “rain on at least one of the days.” Which of the following suggestions are consistent with the probability axioms: **(a)** 70%, 60%, 40%, and 80%, **(b)** 70%, 60%, 40%, and 90%, **(c)** 70%, 60%, 80%, and 50%, and **(d)** 70%, 60%, 50%, and 90%?
11. Two fish are caught and weighed. Consider the events $A = \{\text{the first weighs more than 10 pounds}\}$, $B = \{\text{the second weighs more than 10 pounds}\}$, and $C = \{\text{the sum of the weights is more than 20 pounds}\}$. Argue that $C \subseteq A \cup B$.
12. Let A , B , and C be three events, such that each event has probability $\frac{1}{2}$, each intersection of two has probability $\frac{1}{4}$, and $P(A \cap B \cap C) = \frac{1}{8}$. Find the probability that **(a)** exactly one of the events occurs, **(b)** none of the events occurs, **(c)** at least one of the events occurs, **(d)** all of the events occur, and **(e)** exactly two of the events occur.
13. **(a)** Let A and B be two events. Show that

$$P(A) + P(B) - 1 \leq P(A \cup B) \leq P(A) + P(B)$$

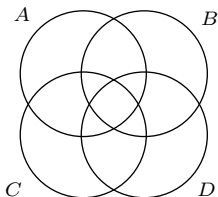
(b) Let $A_1, \dots, A_n$ be a sequence of events. Show that

$$\sum_{k=1}^n P(A_k) - (n-1) \leq P\left(\bigcup_{k=1}^n A_k\right) \leq \sum_{k=1}^n P(A_k)$$

14. A particular species of fish is known to weigh more than 10 pounds with probability 0.01. Suppose that 10 such fish are caught and weighed. Show that the probability that the total weight of the 10 fish is above 100 pounds is at most 0.1.
15. Consider the Venn diagram of four events below. If we use the “area method” to find the probability of $A \cup B \cup C \cup D$, we get

$$\begin{aligned} P(A \cup B \cup C \cup D) &= P(A) + P(B) + P(C) + P(D) \\ &\quad - P(A \cap B) - P(A \cap C) - P(B \cap D) - P(C \cap D) \\ &\quad + P(A \cap B \cap C \cap D) \end{aligned}$$

However, this does not agree with Proposition 1.5 for $n = 4$. Explain!



16. Choose a number at random from the integers $1, \dots, 100$. What is the probability that it is divisible by (a) 2, 3, or 4, (b) i, j , or k ?
17. Consider Example 1.9 where you throw a dart at random. Find the probability that you get (a) 14 or double, (b) 14, double, or triple, (c) even, double, a number higher than 10, or bull's eye.
18. Prove Proposition 1.6 by considering disjoint events $B_1, B_2, \dots$ defined by $B_1 = A_1$, $B_2 = A_2 \setminus B_1$, $\dots$, $B_k = A_k \setminus B_{k-1}, \dots$

Section 1.4. Finite Sample Spaces and Combinatorics

19. You are asked to select a password for a Web site. It must consist of five lowercase letters and two digits in any order. How many possible such passwords are there if (a) repetitions are allowed, and (b) repetitions are not allowed?
20. Consider the Swedish license plate from Example 1.16. Find the probability that a randomly selected plate has (a) no duplicate letters, (b) no duplicate digits, (c) all letters the same, (d) only odd digits, and (e) no duplicate letters and all digits equal.
21. “A thousand monkeys, typing on a thousand typewriters will eventually type the entire works of William Shakespeare” is a statement often heard in one

form or another. Suppose that one monkey presses 10 keys at random. What is the probability that he types the word HAMLET if he is **(a)** allowed to repeat letters, and **(b)** not allowed to repeat letters?

22. Four envelopes contain four different amounts of money. You are allowed to open them one by one, each time deciding whether to keep the amount or discard it and open another envelope. Once an amount is discarded, you are not allowed to go back and get it later. Compute the probability that you get the largest amount under the following different strategies: **(a)** You take the first envelope. **(b)** You open the first envelope, note that it contains the amount x , discard it and take the next amount which is larger than x (if no such amount shows up, you must take the last envelope). **(c)** You open the first two envelopes, call the amounts x and y , and discard both and take the next amount that is larger than both x and y .
23. In the early 1970s, four talented Scandinavians named Agneta, Annifrid, Benny, and Björn put a band together and decided to name it using their first name initials. **(a)** How many possible band names were there? What if a reunion is planned and the reclusive Agneta is replaced by some guy named Robert? **(b)** A generalization of (a): You are given n uppercase letters such that the numbers of $A, B, \dots, Z$ are $n_A, n_B, \dots, n_Z$, respectively (these numbers may be 0). Show that you can create

$$\frac{n!}{n_A!n_B! \cdots n_Z!}$$

different possible words. Compare with your answers in part (a).

24. Prove the following identities (rather than using the definition, try to give combinatorial arguments):

$$\text{(a)} \quad \binom{n+1}{k+1} = \binom{n}{k+1} + \binom{n}{k} \quad \text{(b)} \quad k \binom{n}{k} = n \binom{n-1}{k-1}$$

$$\text{(c)} \quad \binom{2n}{n} = \sum_{k=1}^n \binom{n}{k}^2 \quad \text{(d)} \quad \sum_{k=0}^n \binom{n}{k} = 2^n$$

25. On a chessboard (8×8 squares, alternating black and white), you place three chess pieces at random. What is the probability that they are all **(a)** in the first row, **(b)** on black squares, **(c)** in the same row, and **(d)** in the same row and on the same color?
26. In a regular coordinate system, you start at $(0, 0)$ and flip a fair coin to decide whether to go sideways to $(1, 0)$ or up to $(0, 1)$. You continue in this way, and after n flips you have reached the point (j, k) , where $j + k = n$. What is the probability that **(a)** all the j steps sideways came before the k steps up, **(b)** all the j steps sideways came either before or after the k steps up, and **(c)** all the j steps sideways came in a row?

27. An urn contains n red balls, n white balls, and n black balls. You draw k balls at random without replacement (where $k \leq n$). Find an expression for the probability that you do not get all colors.
28. You are dealt a bridge hand (13 cards). What is the probability that you do not get cards in all suits?
29. Recall Texas Lotto from Example 1.21, where five numbers are chosen among $1, \dots, 44$ and one bonus ball number from the same set. Find the probability that you match **(a)** four of the first five numbers but not the bonus ball, **(b)** three of the first five numbers and the bonus ball.
30. You are dealt a poker hand. What is the probability of getting **(a)** royal flush, **(b)** straight flush, **(c)** four of a kind, **(d)** full house, **(e)** flush, **(f)** straight, **(g)** three of a kind, **(h)** two pairs, and **(i)** one pair? (These are listed in order of descending value in poker, not in order of difficulty!)
31. From the integers $1, \dots, 10$, three numbers are chosen at random without replacement. **(a)** What is the probability that the smallest number is 4? **(b)** What is the probability that the smallest number is 4 and the largest is 8? **(c)** If you choose three numbers from $1, \dots, n$, what is the probability that the smallest number is j and the largest is k for possible values of j and k ?
32. An urn contains n white and m black balls. You draw repeatedly at random and without replacement. What is the probability that the first black ball comes in the k th draw, $k = 1, 2, \dots, n + 1$?
33. In the “Pick 3” game described in Example 1.25, suppose that you choose the “any order” option and play the numbers 111. Since there are a total of 220 cases and 1 favorable case, you think that your chance of winning is $1/220$. However, when playing this repeatedly, you notice that you win far less often than once every 220 times. Explain!
34. How many strictly positive, integer-valued solutions $(x_1, \dots, x_n)$ are there to the equation $x_1 + \dots + x_n = k$?
35. Ann and Bob shuffle a deck of cards each. Ann wins if she can find a card that has the same position in her deck as in Bob’s. What is the (approximate) probability that Ann wins?
36. Consider the matching problem in Example 1.4.17 and let n_j be the number of permutations with exactly j matches for $j = 0, 1, \dots, n$. **(a)** Find an expression for n_0 . *Hint:* How does $n_0/n!$ relate to the probability computed in the example? **(b)** Find the probability of exactly j matches, for $j = 0, 1, \dots, n$ and its limit as $n \rightarrow \infty$. *Hint:* You need to find n_j . First fix a particular set of j numbers, for example, $\{1, 2, \dots, j\}$ and note that the number of ways to match exactly those equals the number of ways to have no matches among the remaining $n - j$ numbers, which you can obtain from part (a).

Section 1.5. Conditional Probability and Independence

37. Let A and B be two events. Is it then true that $P(A|B) + P(A|B^c) = 1$? Give proof or counterexample.

38. Let A and B be disjoint events. Show that

$$P(A|A \cup B) = \frac{P(A)}{P(A) + P(B)}$$

39. Let A , B , and C be three events such that $P(B \cap C) > 0$. Show that

$$P(A \cap B \cap C) = P(A|B \cap C)P(B|C)P(C)$$

and that

$$P(A|B \cap C) = \frac{P(A \cap B|C)}{P(B|C)}$$

40. Let A and B be events, with $P(A) = \frac{1}{2}$ and $P(B) = \frac{1}{3}$. Compute both $P(A \cup B)$ and $P(A \cap B)$ if **(a)** A and B are independent, **(b)** A and B are disjoint, **(c)** A^c and B are independent, and **(d)** A^c and B are disjoint.
41. A politician considers running for election and has decided to give it two tries. He figures that the current conditions are favorable and that he has about a 60% chance of winning this election as opposed to a 50–50 chance in the next election. However, if he does win this election, he estimates that there ought to be a 75% chance of being reelected. **(a)** Find the probability that he wins both elections. **(b)** Find the probability that he wins the first election and loses the second. **(c)** If you learn that he won the second election, what is the probability that he won the first election? **(d)** If he loses the first election, what is the probability that he wins the second?
42. Consider two events A and B . We say that B gives *positive information* about A , denoted $B \nearrow A$, if $P(A|B) > P(A)$, that is if knowing B increases the probability of A . Similarly, we say that B gives *negative information* about A , denoted $B \searrow A$, if $P(A|B) < P(A)$. Are the following statements true or false? **(a)** If $B \nearrow A$, then $A \nearrow B$. **(b)** If $A \nearrow B$ and $B \nearrow C$, then $A \nearrow C$. **(c)** If $B \nearrow A$, then $B \searrow A^c$. **(d)** $A \searrow A^c$.
43. Show that both $\emptyset$ and the sample space S are independent of any event. Explain intuitively.
44. Let S be a sample space with n equally likely outcomes where n is a prime number. Show that there are no independent events (unless one of them is S or $\emptyset$).
45. A coin has probability p of showing heads. Flip it three times and consider the events $A = \{\text{at most one tails}\}$ and $B = \{\text{all flips are the same}\}$. For which values of p are A and B independent?
46. A fair coin is flipped twice. Explain the difference between the following: **(a)** the probability that both flips give heads, and **(b)** the conditional probability that the second flip gives heads given that the first flip gave heads.
47. In December 1992, a small airplane crashed in a residential area near Bromma Airport outside Stockholm, Sweden. In an attempt to calm the residents, the airport manager claimed that they should now feel safer than before since the

probability of two crashes is much smaller than the probability of one crash and hence it has now become less likely that another crash will occur in the future.¹⁰ What do you think of his argument?

48. Bob and Joe are working on a project. They each have to finish their individual tasks to complete the project and work independent of each other. When Bob is asked about the chances of him getting his part done, Joe getting his part done, and then both getting the entire project done, he estimates these to be 99%, 90%, and 95%, respectively. Is this reasonable?
49. You roll a die and consider the events A : get an even outcome, and B : get at least 2. Find $P(B|A)$ and $P(A|B)$.
50. You roll a die twice and record the largest number (if the two rolls give the same outcome, this is the largest number). **(a)** Given that the first roll gives 1, what is the conditional probability that the largest number is 3? **(b)** Given that the first roll gives 3, what is the conditional probability that the largest number is 3?
51. Roll two fair dice. Let A_k be the event that the first die gives k , and let B_n be the event that the sum is n . For which values of n and k are A_k and B_n independent?
52. The distribution of blood types in the United States according to the “ABO classification” is O:45%, A:40%, B:11%, and AB:4%. Blood is also classified according to Rh type, which can be negative or positive and is independent of the ABO type (the corresponding genes are located on different chromosomes). In the U.S. population, about 84% are Rh positive. Sample two individuals at random and find the probability that **(a)** both are A negative, **(b)** one of them is O and Rh positive, while the other is not, **(c)** at least one of them is O positive, **(d)** one is Rh positive and the other is not AB, **(e)** they have the same ABO type, and **(f)** they have the same ABO type and different Rh types.
53. Let A , B , and C be independent events. Show that A is independent of both $B \cap C$ and $B \cup C$.
54. You are offered to play the following game: A roulette wheel is spun eight times. If any of the 38 numbers (0,00,1–36) is repeated, you lose \$10, otherwise you win \$10. Should you accept to play this game? Argue by computing the relevant probability.
55. Consider the following simplified version of the birthday problem in Example 1.19. Divide the year into “winter half” and “summer half.” Suppose that the probability is p that an individual is born in the winter half. What is the probability that two people are born in the same half of the year? For which value of p is this minimized?
56. Consider the birthday problem with two people and suppose that the probability distribution of birthdays is $p_1, \dots, p_{365}$. **(a)** Express the probability that they have the same birthday as a function of the p_k . **(b)** Show that the probability in (a) is minimized for $p_k = \frac{1}{365}, k = 1, 2, \dots, 365$. If you are familiar with Lagrange

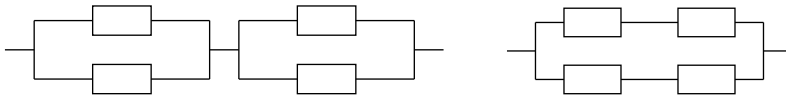
¹⁰True story!

multipliers, you can use these. Alternatively, first show that $\sum_{k=1}^{365} (p_k - \frac{1}{365} + \frac{1}{365})^2 \geq \frac{1}{365}$.

57. A certain text has one-third vowels and two-thirds consonants. Five letters are chosen at random and you are asked to guess the sequence. Find the probability that all guesses are correct if for each letter you **(a)** guess vowel or consonant with equal probabilities, **(b)** guess vowel with probability $\frac{1}{3}$ and consonant with probability $\frac{2}{3}$, and **(c)** always guess consonant.
58. Two events A and B are said to be *conditionally independent* given the event C if

$$P(A \cap B|C) = P(A|C)P(B|C)$$

- (a)** Give an example of events A , B , and C such that A and B are independent but not conditionally independent given C . **(b)** Give an example of events A , B , and C such that A and B are not independent but conditionally independent given C . **(c)** Suppose that A and B are independent events. When are they conditionally independent given their union $A \cup B$? **(d)** Since the information in C and C^c is equivalent (remember Proposition 1.10 and the preceding discussion), we might suspect that if A and B are independent given C , they are also independent given C^c . However, this is not true in general. Give an example of three events A , B , and C such that A and B are independent given C but not given C^c .
59. Roll a die twice and consider the events $A = \{\text{first roll gives at least } 4\}$, $B = \{\text{second roll gives at most } 4\}$, and $C = \{\text{the sum of the rolls is } 10\}$. **(a)** Find $P(A)$, $P(B)$, $P(C)$, and $P(A \cap B \cap C)$. **(b)** Are A , B , and C independent?
60. Roll a die n times and let A_{ij} be the event that the i th and j th rolls give the same number, where $1 \leq i < j \leq n$. Show that the events A_{ij} are pairwise independent but not independent.
61. You throw three darts independently and at random at a dartboard. Find the probability that you get **(a)** no bull's eye, **(b)** at least one bull's eye, **(c)** only even numbers, and **(d)** exactly one triple and at most one double.
62. Three fair dice are rolled. Given that there are no 6s, what is the probability that there are no 5s?
63. You have three pieces of string and tie together the ends two by two at random. **(a)** What is the probability that you get one big loop? **(b)** Generalize to n pieces.
64. Choose n points independently at random on the perimeter of a circle. What is the probability that all points are on the same half-circle?
65. Do Example 1.23 assuming that balls are instead drawn with replacement.
66. A fair coin is flipped n times. Let $A_k = \{\text{heads in } k\text{th flip}\}$, $k = 1, 2, \dots, n$, and $B = \{\text{the total number of heads is even}\}$. Show that $A_1, \dots, A_n, B$ are not independent but that if any one of them is removed, the remaining n events are independent (from Stoyanov, *Counterexamples in Probability* [9]).
67. Compute the reliability of the two systems below given each component functioning independently with probability p .



68. A system is called a “ k -out-of- n system” if it functions whenever at least k of the n components function. Suppose that components function independent of each other with probability p and find an expression for the reliability of the system.
69. You play the following game: You bet \$1, a fair die is rolled and if it shows 6 you win \$4, otherwise you lose your dollar. If you must choose the number of rounds in advance, how should you choose it to maximize your chance of being ahead (having won more than you have lost) when you quit, and what is the probability of this?
70. Suppose that there is a one-in-a-million chance that a person is struck by lightning and that there are n people in a city during a thunderstorm. (a) If n is 2 million, what is the probability that somebody is struck? (b) How large must n be for the probability that somebody is struck to be at least $\frac{1}{2}$?
71. A fair die is rolled n times. Once a number has come up, it is called *occupied* (e.g., if $n = 5$ and we get 2, 6, 5, 6, 2, the numbers 2, 5, and 6 are occupied). Let A_k be the event that k numbers are occupied. Find the probability of A_1 (easy) and A_2 (trickier).

Section 1.6. The Law of Total Probability and Bayes' Formula

72. In the United States, the overall chance that a baby survives delivery is 99.3%. For the 15% that are delivered by cesarean section, the chance of survival is 98.7%. If a baby is not delivered by cesarean section, what is its survival probability?
73. You roll a die and flip a fair coin a number of times determined by the number on the die. What is the probability that you get no heads?
74. In a blood transfusion, you can always give blood to somebody of your own ABO type (see Problem 52). Also, type O can be given to anybody and those with type AB can receive from anybody (people with these types are called *universal donors* and *universal recipients*, respectively). Suppose that two individuals are chosen at random. Find the probability that (a) neither can give blood to the other, (b) one can give to the other but not vice versa, (c) at least one can give to the other, and (d) both can give to each other.
75. You have two urns, 10 white balls, and 10 black balls. You are asked to distribute the balls in the urns, choose an urn at random, and then draw a ball at random from the chosen urn. How should you distribute the balls in order to maximize the probability to get a black ball?
76. A sign reads ARKANSAS. Three letters are removed and then put back into the three empty spaces again, at random. What is the probability that the sign still reads ARKANSAS?

77. A sign reads IDAHO. Two letters are removed and put back at random, each equally likely to be put upside down as in the correct orientation. What is the probability that the sign still reads IDAHO?
78. In the “Pick 3” game from Example 1.25, play the “any order” options and choose your three numbers at random. What is the probability that you win?
79. From a deck of cards, draw four cards at random without replacement. If you get k aces, draw k cards from another deck. What is the probability to get exactly k aces from the first deck and exactly n aces from the second deck?
80. Recall Example 1.39, where you observe a mother walking with a girl. Find the conditional probability that the other child is also a girl in the following cases: **(a)** The mother chooses the older child with probability p . **(b)** If the children are of different genders, the mother chooses the girl with probability p . **(c)** When do you get the second solution in the example that the probability equals $\frac{1}{2}$?
81. Let A , B , and C be three events. Assuming that all conditional probabilities are defined, show that

$$P(A|B) = P(A|B \cap C)P(C|B) + P(A|B \cap C^c)P(C^c|B).$$

82. Graduating students from a particular high school are classified as “weak” or “strong.” Among those who apply to college, it turns out that 56% of the weak students but only 39% of the strong students are accepted at their first choice. Does this indicate a bias against strong students?
83. In Example 1.45, if all three dice are rolled at once, which is the most likely to win?
84. Consider the introduction to Section 1.6. If your car develops engine problems, how likely is it that the dealer sold you a flood-damaged car?
85. Consider the Monty Hall problem in Example 1.48. **(a)** What is the relevance of the assumption that Monty opens a door at random in the case where you chose the car? **(b)** Suppose that there are n doors and k cars, everything else being the same. What is your probability of winning a car with the switching strategy?
86. The three prisoners Shadrach, Mesach, and Abednego learn that two of them will be set free but not who. Later, Mesach finds out that he is one of the two, and, excited, he runs to Shadrach to share his good news. When Shadrach finds out, he gets upset and complains “Why did you tell me? Now that there are only me and Abednego left, my chance to be set free is only $\frac{1}{2}$, but before it was $\frac{2}{3}$.” What do you think of his argument? What assumptions do you make?
87. A box contains two regular quarters and one fake two-headed quarter. **(a)** You pick a coin at random. What is the probability that it is the two-headed quarter? **(b)** You pick a coin at random, flip it, and get heads. What is the probability that it is the two-headed quarter?
88. Two cards are chosen at random without replacement from a deck and inserted into another deck. This deck is shuffled, and one card is drawn. If this card is an ace, what is the probability that no ace was moved from the first deck?

89. A transmitter sends 0s and 1s to a receiver. Each digit is received correctly (0 as 0, 1 as 1) with probability 0.9. Digits are received correctly independent of each other and on the average twice as many 0s as 1s are being sent. **(a)** If the sequence 10 is sent, what is the probability that 10 is received? **(b)** If the sequence 10 is received, what is the probability that 10 was sent?
90. Consider two urns, one with 10 balls numbered 1 through 10 and one with 100 balls numbered 1 through 100. You first pick an urn at random, then pick a ball at random, which has number 5. **(a)** What is the probability that it came from the first urn? **(b)** What is the probability in (a) if the ball was instead chosen randomly from all the 110 balls?
91. Smoking is reported to be responsible for about 90% of all lung cancer. Now consider the risk that a smoker develops lung cancer. Argue why this is not 90%. In order to compute the risk, what more information is needed?
92. The serious disease D occurs with a frequency of 0.1% in a certain population. The disease is diagnosed by a method that gives the correct result (i.e., positive result for those with the disease and negative for those without it) with probability 0.99. Mr Smith goes to test for the disease and the result turns out to be positive. Since the method seems very reliable, Mr Smith starts to worry, being “99% sure of actually having the disease.” Show that this is not the relevant probability and that Mr Smith may actually be quite optimistic.
93. You test for a disease that about 1 in 500 people have. If you have the disease, the test is always positive. If you do not have the disease, the test is 95% accurate. If you test positive, what is the probability that you have the disease?
94. **(a)** Ann and Bob each tells the truth with probability $1/3$ and lies otherwise, independent of each other. If Bob tells you something and Ann tells you Bob told the truth, what is the probability Bob told you the truth? **(b)** Add a third person, Carol, who is as prone to lying as Ann and Bob one. If Ann says that Bob claims that Carol told the truth, what is the probability Carol told the truth?
95. A woman witnesses a hit-and-run accident one night and reports it to the police that the escaping car was black. Since it was dark, the police test her ability to distinguish black from dark blue (other colors are ruled out) under similar circumstances and she is found to be able to pick the correct color about 90% of the time. One police officer claims that they can now be 90% certain that the escaping car was black, but his more experienced colleague says that they need more information. In order to determine the probability that the car was indeed black, what additional information is needed, and how is the probability computed?
96. Joe and Bob are about to drive home from a bar. Since Joe is sober and Bob is not, Joe takes the wheel. Bob has recently read in the paper that drunk drivers are responsible for 25% of car accidents, that about 95% of drivers are sober, and that the overall risk of having an accident is 10%. “You sober people cause 75% of the accidents,” slurs Bob, “and there are so many of you too! You should let me drive!” Joe who knows his probability theory has his answer ready. How does he respond?

97. Consider Example 1.50, where the murderer must be one of n individuals. Suppose that Joe Bloggs is initially considered the main suspect and that the detectives judge that there is a 50–50 chance that he is guilty. If his DNA matches the DNA found at the scene of the crime, what is then the probability that he is guilty?
98. Consider a parallel system of two components. The first component functions with probability p and if it functions, the second also functions with probability p . If the first has failed, the second functions with probability $r < p$, due to heavier load on the single component. **(a)** What is the probability that the second component functions? **(b)** What is the reliability of the system? **(c)** If the second component does not function, what is the probability that the first does?
99. Recall Example 1.51, where you know that the guilty child is a boy, a boy opens the door, and he has one sibling. Compute the probability that the child who opened the door is guilty if the guilty child opens the door with probability p .
100. Your new neighbors have three children. **(a)** If you are told about three independent observations of a boy, what is the probability that they have three boys? **(b)** If you get two confirmations of an observed boy and one of an observed girl, what is the probability that they have two boys and a girl? **(c)** If you get $j \geq 1$ confirmations of an observed boy and $n - j \geq 1$ of an observed girl, what is the probability that they have two boys and a girl?
101. Consider Example 1.53 about cystic fibrosis. **(a)** What is the probability that two healthy parents have a child who neither is a carrier nor has the disease? **(b)** Given that a child is healthy, what is the probability that both parents are carriers (you may disregard parents with the disease)?
102. A genetic disease or condition is said to be *sex-linked* if the responsible gene is located on either of the sex chromosomes, X and Y (recall that women have two X chromosomes and men have one each of X and Y). One example is red-green color-blindness for which the responsible gene is located on the X chromosome. The allele for color-blindness is recessive, so that one normal copy of the gene is sufficient for normal vision. **(a)** Consider a couple where the woman is color-blind and the man has normal vision. If they have a daughter, what is the probability that she is color-blind? If they have a son? **(b)** Compute the probabilities in (a) under the assumption that both parents have normal vision and the woman's father was color-blind. **(c)** It is estimated that about 7% of men are color-blind but only about 0.5% of women. Explain!
103. *Tay–Sachs Disease* is a serious genetic disease that usually leads to death in early childhood. The allele for the disease is recessive and *autosomal* (not located on any of the sex chromosomes). **(a)** In the general population, about 1 in 250 is a carrier of the disease. What incidence among newborns does this give? **(b)** Certain subpopulations are at greater risk for the disease. For example, the incidence among newborns in the Cajun population of Louisiana is 1 in 3600. What proportion of carriers does this give? **(c)** Generally, if a serious recessive disease has a carrier frequency of one in n and an incidence among newborns

- of one in N , what is the relation between n and N ? (Why is it relevant that the disease is “serious?”)
104. Consider the game of badminton in Example 1.57. **(a)** Find the probability that Ann scores the next point if she is the receiver. **(b)** Now suppose that Ann wins a rally as server with probability p_A and let the corresponding probability for Bob be p_B . If Ann serves, what is the probability that she is the next player to score?
 105. In table tennis, a set is won by the first player to reach 11 points, unless the score is 10–10, in which case serves are alternated and the player who first gets ahead by two points wins. Suppose that Ann wins a point as server with probability p_A and Bob wins a point as server with probability p_B . If the score is 10–10 and Ann serves, what is the probability that she wins the set?
 106. You are playing roulette, each time betting on “odd,” which occurs with probability $\frac{18}{38}$ and gives you even money back. You start with \$10 and decide to play until you have either doubled your fortune or gone broke. Compute the probability that you manage to double your fortune if in each round you bet \$10, \$5, \$2, and \$1 dollar, respectively. After you have found the best strategy, give an intuitive explanation of why it is the best and why it is called “bold play.”
 107. In Example 1.59, suppose that Ann wins each round with probability $p > \frac{1}{2}$. What is the probability that she eventually goes broke?
 108. The game of Penney-ante can be played with patterns of any length n . In the case $n = 1$, the game is fair (this is trivial); if $n = 2$, it can be fair or to your advantage, depending on the patterns chosen, and if $n \geq 3$, you can always choose a winning strategy. **(a)** Let $n = 2$, so that the possible patterns are HH , HT , TH , and TT . Suggest a strategy and compute your winning probability in the different cases. **(b)** Let $n = 4$ and suppose that your opponent chooses $HHHH$. Suggest how you should choose your best pattern and compute the winning probability.
 109. In the game of craps, you roll two dice and add the numbers. If you get 7 or 11 (a *natural*) you win, if you roll 2, 3, or 12 (*craps*) you lose. Any other roll establishes your *point*. You then roll the dice repeatedly until you get either 7 or your point. If you get your point first you win, otherwise you lose. Starting a new game of craps, what is the probability that you win?