Chapter 2

Probability

From *Probability, For the Enthusiastic Beginner* (Draft version, March 2016)
David Morin, morin@physics.harvard.edu

Having learned in Chapter 1 how to count things, we can now talk about probability. We will find that in many situations it is a trivial matter to generate probabilities from our counting results. So we will be justly rewarded for the time and effort we spent in Chapter 1.

The outline of this chapter is as follows. In Section 2.1 we give the definition of probability. Although this definition is fairly easy to apply in most cases, there are a number of subtleties that come up. These are discussed in Appendix A. In Section 2.2 we present the various rules of probability. We show how these can be applied in a few simple examples, and then we work through a number of more substantial examples in Section 2.3. In Section 2.4 we present four classic probability problems that many people find counterintuitive. Section 2.5 is devoted to *Bayes' theorem*, which is a relation between certain conditional probabilities. Finally, in Section 2.6 we discuss *Stirling's formula*, which gives an approximation to the ubiquitous factorial, $n!$.

2.1 Definition of probability

Probability gives a measure of how likely it is for something to happen. It can be defined as follows:

> DEFINITION OF PROBABILITY: Consider a very large number of identical trials of a certain process; for example, flipping a coin, rolling a die, picking a ball from a box (with replacement), etc. If the probability of a particular event occurring (for example, getting a Heads, rolling a 5, or picking a blue ball) is p, then the event will occur in a fraction p of the trials, on average.

Some examples are:

- The probability of getting a Heads on a coin flip is 1/2 (or equivalently 50%). This is true because the probabilities of getting a Heads or a Tails are equal, which means that these two outcomes must each occur half of the time, on average.

- The probability of rolling a 5 on a standard 6-sided die is 1/6. This is true because the probabilities of rolling a 1, 2, 3, 4, 5, or 6 are all equal, which means that these six outcomes must each happen one sixth of the time, on average.

- If there are three red balls and seven blue balls in a box, then the probabilities of picking a red ball or a blue ball are, respectively, 3/10 and 7/10. This follows from the fact that the probabilities of picking each of the ten balls are all equal (or at least let's assume they are), which means that each ball will be picked one tenth of the time, on average. Since there are three red balls, a red ball will therefore be picked 3/10 of the time, on average. And since there are seven blue balls, a blue ball will be picked 7/10 of the time, on average.

Note the inclusion of the words "on average" in the above definition and examples. We'll discuss this in detail in the subsection below.

Many probabilistic situations have the property that they involve a number of different possible outcomes, *all of which are equally likely*. For example, Heads and Tails on a coin are equally likely to be tossed, the numbers 1 through 6 on a die are equally likely to be rolled, and the ten balls in the above box are all equally likely to be picked. In such a situation, the probability of a certain scenario happening is given by

$$p = \frac{\text{number of desired outcomes}}{\text{total number of possible outcomes}} \qquad \text{(for equally likely outcomes)} \quad (2.1)$$

Calculating a probability then simply reduces to a matter of counting the number of desired outcomes, along with the total number of outcomes. For example, the probability of rolling an even number on a die is 1/2, because there are three desired outcomes (2, 4, and 6) and six total possible outcomes (the six numbers). And the probability of picking a red ball in the above example is 3/10, as we already noted, because there are three desired outcomes (picking any of the three red balls) and ten total possible outcomes (the ten balls). These two examples involved trivial counting, but we'll encounter many examples where it is more involved. This is why we did all of that counting in Chapter 1!

It should be stressed that Eq. (2.1) holds only under the assumption that all of the possible outcomes are equally likely. But this usually isn't much of a restriction, because this assumption will generally be valid in the setups we'll be dealing with in this book. In particular, it holds in setups dealing with permutations and subgroups, both of which we studied in detail in Chapter 1. Our ability to count these sorts of things will allow us to easily calculate probabilities via Eq. (2.1). Many examples are given in Section 2.3 below.

There are three words that people often use interchangeably: "probability," "chance," and "odds." The first two of these mean the same thing. That is, the statement, "There is a 40% chance that the bus will be late," is equivalent to the statement, "There is a 40% probability that the bus will be late." However, the word "odds" has a different meaning; see Problem 2.1 for a discussion of this.

The importance of the words "on average"

The above definition of probability includes the words "on average." These words are critical, because the definition wouldn't make any sense if we omitted them and instead went with something like: "If the probability of a particular event occurring is *p*, then the event will occur in *exactly* a fraction *p* of the trials." This can't be a valid definition of probability, for the following reason. Consider the roll of one die, for which the probability of each number occurring is 1/6. This definition would imply that on one roll of a die, we will get 1/6 of a 1, and 1/6 of a 2, and so on. But this is nonsense; you can't roll 1/6 of a 1. The number of times a 1 appears on one roll must of course be either zero or one. And in general for many rolls, the number must be an integer, 0, 1, 2, 3, ...

There is a second problem with this definition, in addition to the problem of non integers. What if we roll a die six times? This definition would imply that we will get exactly $(1/6) \cdot 6 = 1$ of each number. This prediction is a little better, in that at least the proposed numbers are integers. But it still can't be correct, because if you actually do the experiment and roll a die six times, you will find that you are certainly *not* guaranteed to get each of the six numbers exactly once. This scenario *might* happen (we'll calculate the probability in Section 2.3.4 below), but it is more likely that some numbers will appear more than once, while other numbers won't appear at all.

Basically, for a small number of trials (such as six), the fractions of the time that the various events occur will most likely not look much like the various probabilities. This is where the words "very large number" in our original definition come in. The point is that if you roll a die a huge number of times, then the fractions of the time that each of the six numbers appears will be *approximately* equal to 1/6. And the larger the number of rolls, the closer the fractions will generally be to 1/6.

In Chapter 5 we'll explain why the fractions are expected to get closer and closer to the actual probabilities, as the number of trials gets larger and larger. For now, just take it on faith that if you flip a coin 100 times, the probability of obtaining either 49, 50, or 51 Heads isn't so large. It happens to be about 24%, which tells you that there is a decent chance that the fraction of Heads will deviate moderately from 1/2. However, if you flip a coin 100,000 times, the probability of obtaining Heads between 49% and 51% of the time is 99.999999975%, which tells you that there is virtually no chance that the fraction of Heads will deviate much from 1/2. If you increase the number of flips to 10^9 (a billion), this result is even more pronounced; the probability of obtaining Heads in the narrow range between 49.99% and 50.01% of the time is 99.999999975% (the same percentage as above). We'll discuss such matters in detail in Section 5.2. For more commentary on the words "on average," see the last section in Appendix A.

2.2 The rules of probability

So far we've talked only about the probabilities of single events, for example, rolling an even number on a die, getting a Heads on a coin toss, or picking a blue ball from a box. We'll now consider two (or more) events. Reasonable questions we

can ask are: What is the probability that both of the events occur? What is the probability that either of the events occurs? The rules presented below will answer these questions. We'll provide a few simple examples for each rule, and then we'll work through some longer examples in Section 2.3.

2.2.1   AND: The "intersection" probability, $P(A \text{ and } B)$

Let A and B be two events. For example, if we roll two dice, we can let $A = \{$rolling a 2 on the left die$\}$ and $B = \{$rolling a 5 on the right die$\}$. Or we might have $A = \{$picking a red ball from a box$\}$ and $B = \{$picking a blue ball without replacement after the first pick$\}$. What is the probability that A and B both occur? In answering this question, we must consider two cases: (1) A and B are independent events, or (2) A and B are dependent events. Let's look at each of these in turn. In each case, the probability that A and B both occur is known as the *joint probability*.

Independent events

Two events are said to be *independent* if they don't affect each other, or more precisely, if the occurrence of one doesn't affect the probability that the other occurs. An example is the first setup mentioned above – rolling two dice, with $A = \{$rolling a 2 on the left die$\}$ and $B = \{$rolling a 5 on the right die$\}$. The probability of obtaining a 5 on the right die is 1/6, independent of what happens with the left die. And similarly the probability of obtaining a 2 on the left die is 1/6, independent of what happens with the right die. Independence requires that neither event affects the other. The events in the second setup mentioned above with the balls in the box are *not* independent; we'll talk about this below.

   Another example of independent events is picking *one* card from a deck, with $A = \{$the card is a king$\}$ and $B = \{$the (same) card is a heart$\}$. The probability of the card being a heart is 1/4, independent of whether or not it is a king. And the probability of the card being a king is 1/13, independent of whether or not it is a heart. Note that it is possible to have two different events even if we have only one card. This card has two qualities (its suit and its value), and we can associate an event with each of these qualities.

REMARK: A note on terminology: The words "event" and "outcome" sometimes mean the same thing in practice, but there is technically a difference. An *outcome* is the result of an experiment. If we draw a card from a deck, then there are 52 possible outcomes; for example, the 4 of clubs, the jack of diamonds, etc. An *event* is a set of outcomes. For example, an event might be "drawing a heart." This event contains 13 outcomes, namely the 13 cards that are hearts. A given card may belong to many events. For example, in addition to belonging to the A and B events in the preceding paragraph, the king of hearts belongs to the events $C = \{$the card is red$\}$, $D = \{$the card's value is higher than 8$\}$, $E = \{$the card is the king of hearts$\}$, and so on. As indicated by the event E, an event may consist of a single outcome. An event may also be the empty set (which occurs with probability 0), or the entire set of all possible outcomes (which occurs with probability 1), which is known as the *sample space*. ♣

The "And" rule for independent events is:

- *If events A and B are independent, then the probability that they both occur equals the product of their individual probabilities:*

$$\boxed{P(A \text{ and } B) = P(A) \cdot P(B)} \tag{2.2}$$

We can quickly apply this rule to the two examples mentioned above. The probability of rolling a 2 on the left die and a 5 on the right die is

$$P(2 \text{ and } 5) = P(2) \cdot P(5) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} . \tag{2.3}$$

This agrees with the fact that one out of the 36 pairs of (ordered) numbers in Table 1.5 is "2, 5." Similarly, the probability that a card is both a king and a heart is

$$P(\text{king and heart}) = P(\text{king}) \cdot P(\text{heart}) = \frac{1}{13} \cdot \frac{1}{4} = \frac{1}{52} . \tag{2.4}$$

This makes sense, because one of the 52 cards in a deck is the king of hearts.

The logic behind Eq. (2.2) is the following. Consider N trials of a given process, where N is very large. In the case of the two dice, a trial consists of rolling both dice. The outcome of such a trial takes the form of an ordered pair of numbers. The first number is the result of the left roll, and the second number is the result of the right roll. On average, the fraction of the outcomes that have a 2 as the first number is $(1/6) \cdot N$.

Let's now consider only this "2-first" group of outcomes and ignore the rest. Then on average, a fraction $1/6$ *of these outcomes* have a 5 as the second number. This is where we are invoking the independence of the events. As far as the second roll is concerned, the set of $(1/6) \cdot N$ trials that have a 2 as the first roll is no different from any other set of $(1/6) \cdot N$ trials, so the probability of obtaining a 5 on the second roll is simply $1/6$. Putting it all together, the average number of trials that have both a 2 as the first number *and* a 5 as the second number is $1/6$ of $(1/6) \cdot N$, which equals $(1/6) \cdot (1/6) \cdot N$.

In the case of general probabilities $P(A)$ and $P(B)$, it is easy to see that the two $(1/6)$'s in the above result get replaced by $P(A)$ and $P(B)$. So the average number of outcomes where A and B both occur is $P(A) \cdot P(B) \cdot N$. And since we performed N trials, the fraction of outcomes where A and B both occur is $P(A) \cdot P(B)$, on average. From the definition of probability in Section 2.1, this fraction is the probability that A and B both occur, in agreement with Eq. (2.2).

If you want to think about the rule in Eq. (2.2) in terms of a picture, then consider Fig. 2.1. Without worrying about specifics, let's assume that different points within the overall square represent different outcomes. And let's assume that they're all equally likely, which means that the area of a region gives the probability that an outcome located in that region occurs (assuming that the area of the whole region is 1). The figure corresponds to $P(A) = 0.2$ and $P(B) = 0.4$. Outcomes to the left of the vertical line are ones where A occurs, and outcomes to the right of the vertical line are ones where A doesn't occur. Likewise for B and outcomes above and below the horizontal line.
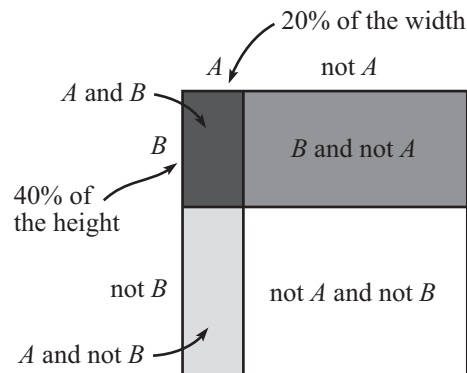
**Figure 2.1:** A probability square for independent events.

From the figure, we see that not only is 40% of the entire square above the vertical line, but also that 40% of the left vertical strip (where A occurs) is above the vertical line, and likewise for the right vertical strip (where A doesn't occur). In other words, B occurs 40% of the time, independent of whether or not A occurs. Basically, B couldn't care less what happens with A. Similar statements hold with A and B interchanged. So this type of figure, with a square divided by horizontal and vertical lines, does indeed represent independent events.

The darkly shaded "A and B" region is the intersection of the region to the left of the vertical line (where A occurs) and the region above the horizontal line (where B occurs). Hence the word "intersection" in the title of this section. The area of the darkly shaded region is 20% of 40% (or 40% of 20%) of the total area, that is, $(0.2)(0.4) = 0.08$ of the total area. The total area corresponds to a probability of 1, so the darkly shaded region corresponds to a probability of 0.08. Since we obtained this probability by multiplying $P(A)$ by $P(B)$, we have therefore given a pictorial proof of Eq. (2.2).

Dependent events

Two events are said to be *dependent* if they *do* affect each other, or more precisely, if the occurrence of one *does* affect the probability that the other occurs. An example is picking two balls in succession from a box containing two red balls and three blue balls (see Fig. 2.2), with $A = \{$choosing a red ball on the first pick$\}$ and $B = \{$choosing a blue ball on the second pick, *without replacement* after the first pick$\}$. If you pick a red ball first, then the probability of picking a blue ball second is 3/4, because there are three blue balls and one red ball left. On the other hand, if you *don't* pick a red ball first (that is, if you pick a blue ball first), then the probability of picking a blue ball second is 2/4, because there are two red balls and two blue balls left. So the occurrence of A certainly affects the probability of B.

Another example might be something like: $A = \{$it rains at 6:00$\}$ and $B = \{$you walk to the store at 6:00$\}$. People are generally less likely to go for a walk when it's raining outside, so (at least for most people) the occurrence of A affects the probability of B.

2.2. The rules of probability
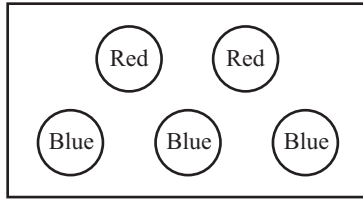
Figure 2.2: A box with two red balls and three blue balls.

The "And" rule for dependent events is:

- *If events A and B are dependent, then the probability that they both occur equals*

$$P(A \text{ and } B) = P(A) \cdot P(B|A) \tag{2.5}$$

  where $P(B|A)$ stands for the probability that B occurs, given that A occurs. It is called a "conditional probability," because we are assuming a given condition, namely that A occurs. It is read as "the probability of B, given A."

There is actually no need for the "dependent" qualifier in the first line of this rule, as we'll see in the second remark near the end of this section.

The logic behind Eq. (2.5) is the following. Consider N trials of a given process, where N is very large. In the above setup with the balls in a box, a "trial" consists of picking two balls in succession, without replacement. On average, the fraction of the outcomes in which a red ball is drawn on the first pick is $P(A) \cdot N$. Let's now consider only these outcomes and ignore the rest. Then a fraction $P(B|A)$ *of these outcomes* have a blue ball drawn second, by the definition of $P(B|A)$. So the number of outcomes where A and B both occur is $P(B|A) \cdot P(A) \cdot N$. And since we performed N trials, the fraction of outcomes where A and B both occur is $P(A) \cdot P(B|A)$, on average. This fraction is the probability that A and B both occur, in agreement with the rule in Eq. (2.5).

The reasoning in the previous paragraph is equivalent to the mathematical identity,

$$\frac{n_{A \text{ and } B}}{N} = \frac{n_A}{N} \cdot \frac{n_{A \text{ and } B}}{n_A}, \tag{2.6}$$

where n_A is the number of trials where A occurs, etc. By definition, the lefthand side of this equation equals $P(A \text{ and } B)$, the first term on the righthand side equals $P(A)$, and the second term on the righthand side equals $P(B|A)$. So Eq. (2.6) is equivalent to the relation,

$$P(A \text{ and } B) = P(A) \cdot P(B|A), \tag{2.7}$$

which is Eq. (2.5). In terms of the Venn-diagram type of picture in Fig. 2.3, Eq. (2.6) is the statement that the darkly shaded area (which represents $P(A \text{ and } B)$) equals the area of the A region (which represents $P(A)$) multiplied by the fraction of the A region that is taken up by the darkly shaded region. This fraction is $P(B|A)$, by definition.
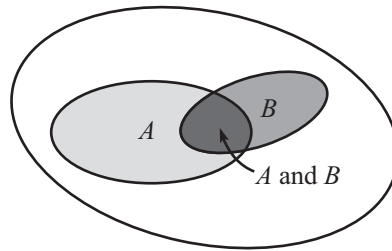
Figure 2.3: Venn diagram for probabilities of dependent events.

As in Fig. 2.1, we're assuming in Fig. 2.3 that different points within the overall boundary represent different outcomes, and that they're all equally likely. This means that the area of a region gives the probability that an outcome located in that region occurs (assuming that the area of the whole region is 1). We're using Fig. 2.3 for its qualitative features only, so we're drawing the various regions as general blobs, as opposed to the specific rectangles in Fig. 2.1, which we used for a quantitative calculation.

Because the "*A* and *B*" region in Fig. 2.3 is the intersection of the *A* and *B* regions, and because the intersection of two sets is usually denoted by $A \cap B$, you will often see the $P(A \text{ and } B)$ probability written as $P(A \cap B)$. That is,

$$P(A \cap B) \equiv P(A \text{ and } B). \tag{2.8}$$

But we'll stick with the $P(A \text{ and } B)$ notation in this book.

There is nothing special about the order of *A* and *B* in Eq. (2.5). We could just as well interchange the letters and write $P(B \text{ and } A) = P(B) \cdot P(A|B)$. However, we know that $P(B \text{ and } A) = P(A \text{ and } B)$, because it doesn't matter which event you say first when you say that two events both occur. So we can also write $P(A \text{ and } B) = P(B) \cdot P(A|B)$. Combining this with Eq. (2.5), we see that we can write $P(A \text{ and } B)$ in two different ways:

$$\begin{aligned} P(A \text{ and } B) &= P(A) \cdot P(B|A) \\ &= P(B) \cdot P(A|B). \end{aligned} \tag{2.9}$$

The fact that $P(A \text{ and } B)$ can be written in these two ways will be critical when we discuss Bayes' theorem in Section 2.5.

Example (Balls in a box): Let's apply Eq. (2.5) to the setup with the balls in the box in Fig. 2.2 above. Let A = {choosing a red ball on the first pick} and B = {choosing a blue ball on the second pick, *without replacement* after the first pick}. For shorthand, we'll denote these events by Red_1 and Blue_2, where the subscript refers to the first or second pick. We noted above that $P(\text{Blue}_2|\text{Red}_1) = 3/4$. And we also know that $P(\text{Red}_1)$ is simply 2/5, because there are initially two red balls and three blue balls. So Eq. (2.5) gives the probability of picking a red ball first and a blue ball second (without replacement after the first pick) as

$$P(\text{Red}_1 \text{ and } \text{Blue}_2) = P(\text{Red}_1) \cdot P(\text{Blue}_2|\text{Red}_1) = \frac{2}{5} \cdot \frac{3}{4} = \frac{3}{10}. \tag{2.10}$$

We can verify that this is correct by listing out all of the possible pairs of balls that can be picked. If we label the balls as 1, 2, 3, 4, 5, and if we let 1, 2 be the red balls, and 3, 4, 5 be the blue balls, then the possible outcomes are shown in Table 2.1. The first number stands for the first ball picked, and the second number stands for the second ball picked.

	Red first		Blue first		
Red second	—	2 1	3 1	4 1	5 1
	1 2	—	3 2	4 2	5 2
Blue second	1 3	2 3	—	4 3	5 3
	1 4	2 4	3 4	—	5 4
	1 5	2 5	3 5	4 5	—

Table 2.1: Ways to pick two balls from the box in Fig. 2.2, without replacement.

The "—" entries stand for the outcomes that aren't allowed; we can't pick two of the same ball, because we're not replacing the ball after the first pick. The dividing lines are drawn for clarity. The internal vertical line separates the outcomes where a red or blue ball is drawn on the first pick, and the internal horizontal line separates the outcomes where a red or blue ball is drawn on the second pick. The six pairs in the lower left corner are the outcomes where a red ball (numbered 1 and 2) is drawn first and a blue ball (numbered 3, 4, and 5) is drawn second. Since there are 20 possible outcomes in all, the desired probability is $6/20 = 3/10$, in agreement with Eq. (2.10).

Table 2.1 also gives a verification of the $P(\text{Red}_1)$ and $P(\text{Blue}_2|\text{Red}_1)$ probabilities we wrote down in Eq. (2.10). $P(\text{Red}_1)$ equals $2/5$ because eight of the 20 entries are to the left of the vertical line. And $P(\text{Blue}_2|\text{Red}_1)$ equals $3/4$ because six of these eight entries are below the horizontal line.

The task of Problem 2.4 is to verify that the second expression in Eq. (2.9) also gives the correct result for $P(\text{Red}_1 \text{ and } \text{Blue}_2)$ in this setup.

We can think about the rule in Eq. (2.5) in terms of a picture analogous to Fig. 2.1. If we consider the above example with the red and blue balls, then the first thing we need to do is recast Table 2.1 in a form where equal areas yield equal probabilities. If we get rid of the "—" entries in Table 2.1, then all entries have equal probabilities, and we end up with Table 2.2.

1 2	2 1	3 1	4 1	5 1
1 3	2 3	3 2	4 2	5 2
1 4	2 4	3 4	4 3	5 3
1 5	2 5	3 5	4 5	5 4

Table 2.2: Rewriting Table 2.1.

In the spirit of Fig. 2.1, this table becomes the square shown in Fig. 2.4. The upper left region corresponds to red balls on both picks. The lower left region

corresponds to a red ball and then a blue ball. The upper right region corresponds to a blue ball and then a red ball. And the lower right region corresponds to blue balls on both picks. This figure makes it clear why we formed the product $(2/5) \cdot (3/4)$ in Eq. (2.10). The 2/5 gives the fraction of the outcomes that lie to the left of the vertical line (these are the ones that have a red ball first), and the 3/4 gives the fraction *of these outcomes* that lie below the horizontal line (these are the ones that have a blue ball second). The product of these fractions gives the overall fraction (namely 3/10) of the outcomes that lie in the lower left region.
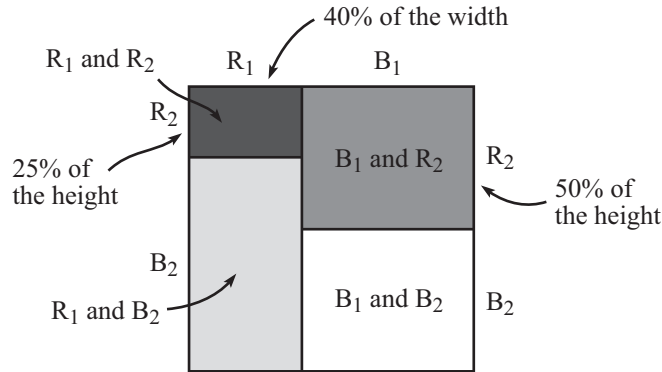
Figure 2.4: Pictorial representation of Table 2.2.

The main difference between Fig. 2.4 and Fig. 2.1 is that the one horizontal line in Fig. 2.1 is now two different horizontal lines in Fig. 2.4. The heights of the horizontal lines in Fig. 2.4 depend on which vertical strip we're dealing with. This is the visual manifestation of the fact that the red/blue probabilities on the second pick depend on what happens on the first pick.

REMARKS:

1. The method of explicitly counting the possible outcomes in Table 2.1 shows that you don't *have* to use the rule in Eq. (2.5), or similarly the rule in Eq. (2.2), to calculate probabilities. You can often instead just count up the various outcomes and solve the problem from scratch. However, the rules in Eqs. (2.2) and (2.5) allow you to take a shortcut that avoids listing out all the outcomes, which might be rather difficult if you're dealing with large numbers.

2. The rule in Eq. (2.2) for independent events is a special case of the rule in Eq. (2.5) for dependent events. This is true because if A and B are independent, then $P(B|A)$ is simply equal to $P(B)$, because the probability of B occurring is just $P(B)$, independent of whether or not A occurs. Eq. (2.5) then reduces to Eq. (2.2) when $P(B|A) = P(B)$. Therefore, there was technically no need to introduce Eq. (2.2) first. We could have started with Eq. (2.5), which covers all possible scenarios, and then showed that it reduces to Eq. (2.2) when the events are independent. But pedagogically, it is often better to start with a special case and then work up to the more general case.

3. In the above "balls in a box" example, we encountered the conditional probability $P(\text{Blue}_2|\text{Red}_1)$. We can also talk about the "reversed" conditional probability, $P(\text{Red}_1|\text{Blue}_2)$. However, since the second pick happens *after* the first pick, you might wonder how much sense it makes to talk about the probability of the Red_1

event, *given* the Blue$_2$ event. Does the second pick somehow influence the first pick, even though the second pick hasn't happened yet? When you make the first pick, are you being affected by a mysterious influence that travels backward in time?

No, and no. When we talk about $P(\text{Red}_1|\text{Blue}_2)$, or about any other conditional probability in the example, everything we might want to know can be read off from Table 2.1. Once the table has been created, we can forget about the temporal order of the events. By looking at the Blue$_2$ pairs (below the horizontal line), we see that $P(\text{Red}_1|\text{Blue}_2) = 6/12 = 1/2$. This should be contrasted with $P(\text{Red}_1|\text{Red}_2)$, which is obtained by looking at the Red$_2$ pairs (above the horizontal line); we find that $P(\text{Red}_1|\text{Red}_2) = 2/8 = 1/4$. Therefore, the probability that your first pick is red *does* depend on whether your second pick is blue or red. But this doesn't mean that there is a backward influence in time. All it says is that if you perform a large number of trials of the given process (drawing two balls, without replacement), and if you look at all of the cases where your second pick is blue (or conversely, red), then you will find that your first pick is red in $1/2$ (or conversely, $1/4$) of these cases, on average. In short, the second pick has no causal influence on the first pick, but the after-the-fact *knowledge* of the second pick affects the *probability* of what the first pick *was*.

4. A trivial yet extreme example of dependent events is the two events: *A*, and "not *A*." The occurrence of *A* highly affects the probability of "not *A*" occurring. If *A* occurs, then "not *A*" occurs with probability zero. And if *A* doesn't occur, then "not *A*" occurs with probability 1. ♣

In the second remark above, we noted that if *A* and *B* are independent (that is, if the occurrence of one doesn't affect the probability that the other occurs), then $P(B|A) = P(B)$. Similarly, we also have $P(A|B) = P(A)$. Let's prove that one of these relations implies the other. Assume that $P(B|A) = P(B)$. Then if we equate the two righthand sides of Eq. (2.9) and use $P(B|A) = P(B)$ to replace $P(B|A)$ with $P(B)$, we obtain

$$P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$$
$$\implies P(A) \cdot P(B) = P(B) \cdot P(A|B)$$
$$\implies P(A) = P(A|B). \tag{2.11}$$

So $P(B|A) = P(B)$ implies $P(A|B) = P(A)$, as desired. In other words, if *B* is independent of *A*, then *A* is also independent of *B*. We can therefore talk about two events being independent, without worrying about the direction of the independence. The condition for independence is therefore either of the relations,

$$\boxed{P(B|A) = P(B) \quad \text{or} \quad P(A|B) = P(A)} \qquad \text{(independence)} \tag{2.12}$$

Alternatively, the condition for independence may be expressed by Eq. (2.2),

$$\boxed{P(A \text{ and } B) = P(A) \cdot P(B)} \qquad \text{(independence)} \tag{2.13}$$

because this equation implies (by comparing it with Eq. (2.5), which is valid in any case) that $P(B|A) = P(B)$.

2.2.2   OR: The "union" probability, $P(A \text{ or } B)$

Let A and B be two events. For example, let $A = \{\text{rolling a 2 on a die}\}$ and $B = \{\text{rolling a 5 on the *same* die}\}$. Or we might have $A = \{\text{rolling an even number (that is, 2, 4, or 6) on a die}\}$ and $B = \{\text{rolling a multiple of 3 (that is, 3 or 6) on the *same* die}\}$. A third example is $A = \{\text{rolling a 1 on one die}\}$ and $B = \{\text{rolling a 6 on *another* die}\}$. What is the probability that either A or B (or both) occurs? In answering this question, we must consider two cases: (1) A and B are exclusive events, or (2) A and B are nonexclusive events. Let's look at each of these in turn.

Exclusive events

Two events are said to be *exclusive* if one precludes the other. That is, they can't both happen. An example is rolling one die, with $A = \{\text{rolling a 2 on the die}\}$ and $B = \{\text{rolling a 5 on the *same* die}\}$. These events are exclusive because it is impossible for one number to be both a 2 and a 5. (The events in the second and third scenarios mentioned above are *not* exclusive; we'll talk about this below.) Another example is picking one card from a deck, with $A = \{\text{the card is a diamond}\}$ and $B = \{\text{the card is a heart}\}$. These events are exclusive because it is impossible for one card to be both a diamond and a heart.

The "Or" rule for exclusive events is:

- *If events A and B are exclusive, then the probability that either of them occurs equals the sum of their individual probabilities:*

$$\boxed{P(A \text{ or } B) = P(A) + P(B)} \tag{2.14}$$

The logic behind this rule boils down to Fig. 2.5. The key feature of this figure is that there is *no overlap* between the two regions, because we are assuming that A and B are exclusive. If there were a region that was contained in both A and B, then the outcomes in that region would be ones for which A and B both occur, which would violate the assumption that A and B are exclusive. The rule in Eq. (2.14) is simply the statement that the area of the union (hence the word "union" in the title of this section) of regions A and B equals the sum of their areas. There is nothing fancy going on here. This statement is no deeper than the statement that if you have two separate bowls, the total number of apples in the two bowls equals the number of apples in one bowl plus the number of apples in the other bowl.

We can quickly apply this rule to the two examples mentioned above. In the example with the die, the probability of rolling a 2 or a 5 on one die is

$$P(2 \text{ or } 5) = P(2) + P(5) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}. \tag{2.15}$$

This makes sense, because two of the six numbers on a die are the 2 and the 5. In the card example, the probability of a card being either a diamond or a heart is

$$P(\text{diamond or heart}) = P(\text{diamond}) + P(\text{heart}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}. \tag{2.16}$$
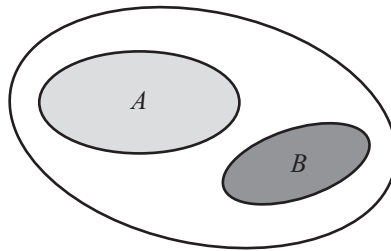
Figure 2.5: Venn diagram for the probabilities of exclusive events.

This makes sense, because half of the 52 cards in a deck are diamonds or hearts.

A special case of Eq. (2.14) is the "Not" rule, which follows from letting $B =$ "not A."

$$P(A \text{ or } (\text{not } A)) = P(A) + P(\text{not } A)$$
$$\implies 1 = P(A) + P(\text{not } A)$$
$$\implies P(\text{not } A) = 1 - P(A). \tag{2.17}$$

The first equality here follows from Eq. (2.14), because A and "not A" are certainly exclusive events; you can't both have something and not have it. To obtain the second line in Eq. (2.17), we have used $P(A \text{ or } (\text{not } A)) = 1$, which holds because every possible outcome belongs to either A or "not A."

Nonexclusive events

Two events are said to be *nonexclusive* if it is possible for both to happen. An example is rolling one die, with A = {rolling an even number (that is, 2, 4, or 6)} and B = {rolling a multiple of 3 (that is, 3 or 6) on the *same* die}. If you roll a 6, then A and B both occur. Another example is picking one card from a deck, with A = {the card is a king} and B = {the card is a heart}. If you pick the king of hearts, then A and B both occur.

The "Or" rule for nonexclusive events is:

- *If events A and B are nonexclusive, then the probability that either (or both) of them occurs equals*

$$\boxed{P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)} \tag{2.18}$$

The "or" here is the so-called "inclusive or," in the sense that we say "A or B occurs" if either *or both* of the events occur. As with the "dependent" qualifier in the "And" rule in Eq. (2.5), there is actually no need for the "nonexclusive" qualifier in the "Or" rule here, as we'll see in the third remark below.

The logic behind Eq. (2.18) boils down to Fig. 2.6. The rule in Eq. (2.18) is the statement that the area of the union of regions A and B equals the sum of their areas *minus the area of the overlap*. This subtraction is necessary so that we don't double

count the region that belongs to both A and B. This region isn't "doubly good" just because it belongs to both A and B. As far as the "A or B" condition goes, the overlap region is just the same as any other part of the union of A and B.
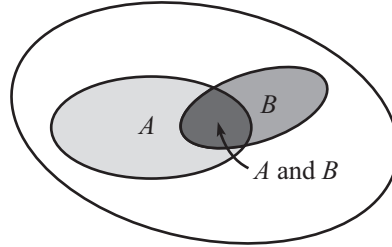
Figure 2.6: Venn diagram for the probabilities of nonexclusive events.

In terms of a physical example, the rule in Eq. (2.18) is equivalent to the statement that if you have two bird cages that have a region of overlap, then the total number of birds in the cages equals the number of birds in one cage, plus the number in the other cage, minus the number in the overlap region. In the situation shown in Fig. 2.7, we have $7 + 5 - 2 = 10$ birds (which oddly all happen to be flying at the given moment).
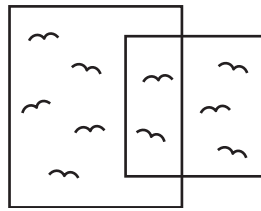
Figure 2.7: Birds in overlapping cages.

Things get more complicated if you have three or more events and you want to calculate probabilities like $P(A \text{ or } B \text{ or } C)$. But in the end, the main task is to keep track of the overlaps of the various regions; see Problem 2.2.

Because the "A or B" region in Fig. 2.6 is the union of the A and B regions, and because the union of two sets is usually denoted by $A \cup B$, you will often see the $P(A \text{ or } B)$ probability written as $P(A \cup B)$. That is,

$$P(A \cup B) \equiv P(A \text{ or } B). \tag{2.19}$$

But we'll stick with the $P(A \text{ or } B)$ notation in this book.

We can quickly apply Eq. (2.18) to the two examples mentioned above. In the example with the die, the only way to roll an even number *and* a multiple of 3 on a single die is to roll a 6, which happens with probability 1/6. So Eq. (2.18) gives the probability of rolling an even number or a multiple of 3 as

$$P(\text{even or mult of 3}) = P(\text{even}) + P(\text{mult of 3}) - P(\text{even and mult of 3})$$

$$= \frac{1}{2} + \frac{1}{3} - \frac{1}{6} = \frac{4}{6} = \frac{2}{3}. \tag{2.20}$$

This makes sense, because four of the six numbers on a die are even numbers or multiples of 3, namely 2, 3, 4, and 6. (Remember that whenever we use "or," it means the "inclusive or.") We subtracted off the 1/6 in Eq. (2.20) so that we didn't double count the roll of a 6.

In the card example, the only way to pick a king *and* a heart with a single card is to pick the king of hearts, which happens with probability 1/52. So Eq. (2.18) gives the probability that a card is a king or a heart as

$$P(\text{king or heart}) = P(\text{king}) + P(\text{heart}) - P(\text{king and heart})$$
$$= \frac{1}{13} + \frac{1}{4} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}. \tag{2.21}$$

This makes sense, because 16 of the 52 cards in a deck are kings or hearts, namely the 13 hearts, plus the kings of diamonds, spades, and clubs; we already counted the king of hearts. As in the previous example with the die, we subtracted off the 1/52 here so that we didn't double count the king of hearts.

REMARKS:

1. If you want, you can think of the area of the union of *A* and *B* in Fig. 2.6 as the area of *only A*, plus the area of *only B*, plus the area of "*A* and *B*." (Equivalently, the number of birds in the cages in Fig. 2.7 is 5 + 3 + 2 = 10.) This is easily visualizable, because these three areas are the ones you see in the figure. However, the probabilities of *only A* and of *only B* are often a pain to deal with, so it's generally easier to think of the area of the union of *A* and *B* as the area of *A*, plus the area of *B*, minus the area of the overlap. This way of thinking corresponds to Eq. (2.18).

2. As we mentioned in the first remark on page 66, you don't *have* to use the above rules of probability to calculate things. You can often instead just count up the various outcomes and solve the problem from scratch. In many cases you're doing basically the same thing with the two methods, as we saw in the above examples with the die and the cards.

3. As with Eqs. (2.2) and (2.5), the rule in Eq. (2.14) for exclusive events is a special case of the rule in Eq. (2.18) for nonexclusive events. This is true because if *A* and *B* are exclusive, then $P(A \text{ and } B) = 0$, by definition. Eq. (2.18) then reduces to Eq. (2.14) when $P(A \text{ and } B) = 0$. Likewise, Fig. 2.5 is a special case of Fig. 2.6 when the regions have zero overlap. There was therefore technically no need to introduce Eq. (2.14) first. We could have started with Eq. (2.18), which covers all possible scenarios, and then showed that it reduces to Eq. (2.14) when the events are exclusive. But as in Section 2.2.1, it is often better to start with a special case and then work up to the more general case. ♣

2.2.3 (In)dependence and (non)exclusiveness

Two events are either independent or dependent, and they are also either exclusive or nonexclusive. There are therefore $2 \cdot 2 = 4$ combinations of these characteristics. Let's see which combinations are possible. You'll need to read this section *very* slowly if you want to keep everything straight. This discussion is given for curiosity's sake only, in case you were wondering how the dependent/independent characteristic relates to the exclusive/nonexclusive characteristic. There is no need

to memorize the results below. Instead, you should think about each situation individually and determine its properties from scratch.

- EXCLUSIVE AND INDEPENDENT: This combination isn't possible. If two events are independent, then their probabilities are independent of each other, which means that there is a nonzero probability (namely, the product of the individual probabilities) that both events happens. Therefore, they cannot be exclusive.

  Said in another way, if two events A and B are exclusive, then the probability of B given A is zero. But if they are also independent, then the probability of B is independent of what happens with A. So the probability of B must be zero, period. Such a B is a very uninteresting event, because it never happens.

- EXCLUSIVE AND DEPENDENT: This combination is possible. An example consists of the events

$$A = \{\text{rolling a 2 on a die}\},$$
$$B = \{\text{rolling a 5 on the } same \text{ die}\}. \qquad (2.22)$$

  Another example consists of A as one event and $B = \{\text{not } A\}$ as the other. In both of these examples the events are exclusive, because they can't both happen. Furthermore, the occurrence of one event certainly affects the probability of the other occurring, in that the probability $P(B|A)$ takes the extreme value of zero, due to the exclusive nature of the events. The events are therefore quite dependent (in a negative sort of way). In short, *if two events are exclusive, then they are necessarily also dependent.*

- NONEXCLUSIVE AND INDEPENDENT: This combination is possible. An example consists of the events

$$A = \{\text{rolling a 2 on a die}\},$$
$$B = \{\text{rolling a 5 on } another \text{ die}\}. \qquad (2.23)$$

  Another example consists of the events $A = \{\text{getting a Heads on a coin flip}\}$ and $B = \{\text{getting a Heads on another coin flip}\}$. In both of these examples the events are clearly independent, because they involve different dice or coins. And the events *can* both happen (a fact that is guaranteed by their independence, as mentioned in the "Exclusive and Independent" case above), so they are nonexclusive. In short, *if two events are independent, then they are necessarily also nonexclusive.* This statement is the logical "contrapositive" of the corresponding statement in the "Exclusive and Dependent" case above.

- NONEXCLUSIVE AND DEPENDENT: This combination is possible. An example consists of the events

$$A = \{\text{rolling a 2 on a die}\},$$
$$B = \{\text{rolling an even number on the } same \text{ die}\}. \qquad (2.24)$$

Another example consists of picking balls *without replacement* from a box with two red balls and three blue balls, with the events being A = {picking a red ball on the first pick} and B = {picking a blue ball on the second pick}. In both of these examples the events are dependent, because the occurrence of A affects the probability of B. (In the die example, $P(B|A)$ takes on the extreme value of 1, which isn't equal to $P(B) = 1/2$. Also, $P(A|B) = 1/3$, which isn't equal to $P(A) = 1/6$. Likewise for the box example.) And the events can both happen, so they are nonexclusive.

To sum up, we see that all exclusive events must be dependent, but nonexclusive events can be either independent or dependent. Similarly, all independent events must be nonexclusive, but dependent events can be either exclusive or nonexclusive. These facts are summarized in Table 2.3, which indicates which combinations are possible.
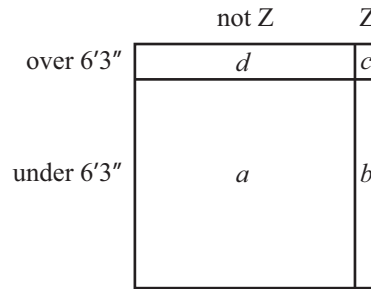
	Independent	Dependent
Exclusive	NO	YES
Nonexclusive	YES	YES

Table 2.3: Relations between (in)dependence and (non)exclusiveness.

2.2.4 Conditional probability

In Eq. (2.5) we introduced the concept of conditional probability, with $P(B|A)$ denoting the probability that B occurs, given that A occurs. In this section we'll talk more about conditional probabilities. In particular, we'll show that two probabilities that you might naively think are equal are in fact not equal. Consider the following example.

Fig. 2.8 gives a pictorial representation of the probability that a random person's height is greater than 6′3″ (6 feet, 3 inches) or less than 6′3″, along with the probability that a random person's last name begins with Z or not Z. We haven't tried to mimic the exact numbers, but we have indicated that the vast majority of people are under 6′3″ (this case takes up most of the vertical span of the square), and also that the vast majority of people have a last name that doesn't begin with Z (this case takes up most of the horizontal span of the square). We'll assume that the probabilities involving heights and last-name letters are independent. This independence manifests itself in the fact that the horizontal and vertical dividers of the square are straight lines (as opposed to, for example, the shifted lines in Fig. 2.4). This independence makes things a little easier to visualize, but it isn't critical in the following discussion.

Figure 2.8: Probability square for independent events (height, and first letter of last name).

Let's now look at some conditional probabilities. Let the areas of the four rectangles in Fig. 2.8 be a, b, c, d, as indicated. The area of a region represents the probability that a given person is in that region. Let Z stand for "having a last name that begins with Z," and let U stand for "being under $6'3''$ in height."

Consider the conditional probabilities $P(Z|U)$ and $P(U|Z)$. $P(Z|U)$ deals with the subset of cases where we know that U occurs. These cases are associated with the area below the horizontal dividing line in the figure. So $P(Z|U)$ equals the fraction of the area *below the horizontal line* (which is $a + b$) that is also *to the right of the vertical line* (which is b). This fraction $b/(b + a)$ is very small.

In contrast, $P(U|Z)$ deals with the subset of cases where we know that Z occurs. These cases are associated with the area to the right of the vertical dividing line in the figure. So $P(U|Z)$ equals the fraction of the area *to the right of the vertical line* (which is $b + c$) that is also *below the horizontal line* (which is b). This fraction $b/(b + c)$ is very close to 1. To sum up, we have

$$P(Z|U) = \frac{b}{b + a} \approx 0,$$

$$P(U|Z) = \frac{b}{b + c} \approx 1. \tag{2.25}$$

We see that $P(Z|U)$ is *not* equal to $P(U|Z)$. If we were dealing with a situation where $a = c$, then these conditional probabilities *would* be equal. But that is an exception. In general, the two probabilities are not equal.

If you're too hasty in your thinking, you might say something like, "Since U and Z are independent, one doesn't affect the other, so the conditional probabilities should be the same." This conclusion is incorrect. The correct statement is, "Since U and Z are independent, one doesn't affect the other, so the conditional probabilities are equal to the corresponding unconditional probabilities." That is, $P(Z|U) = P(Z)$ and $P(U|Z) = P(U)$. But $P(Z)$ and $P(U)$ are vastly different, with the former being approximately zero, and the latter being approximately 1.

In order to make it obvious that the two conditional probabilities $P(A|B)$ and $P(B|A)$ aren't equal in general, we picked an example where the various probabilities were all either close to zero or close to 1. We did this solely for pedagogical purposes; the non-equality of the conditional probabilities holds in general (except in the $a = c$ case). Another extreme example that makes it clear that the two conditional probabilities are different is: The probability that a living thing is human,

given that it has a brain, is very small; but the probability that a living thing has a brain, given that it is human, is 1.

The takeaway lesson here is that when thinking about the conditional probability $P(A|B)$, the order of A and B is critical. Great confusion can arise if one forgets this fact. The classic example of this confusion is the "Prosecutor's fallacy," discussed below in Section 2.4.3. That example should convince you that a lack of basic knowledge of probability can have significant and possibly tragic consequences in real life.

2.3 Examples

Let's now do some examples. Introductory probability problems generally fall into a few main categories, so we've divided the examples into the various subsections below. There is no better way to learn how to solve probability problems (or any kind of problem, for that matter) than to just sit down and do a bunch of them, so we've presented quite a few.

If the statement of a given problem lists out the specific probabilities of the possible outcomes, then the rules in Section 2.2 are often called for. However, in many problems you encounter, you'll be calculating probabilities from scratch (by counting things), so the rules in Section 2.2 generally don't come into play. You simply have to do lots of counting. This will become clear in the examples below. For all of these, *be sure to try the problem for a few minutes on your own before looking at the solution*.

In virtually all of these examples, we'll be dealing with situations in which the various possible outcomes are equally likely. For example, we'll be tossing coins, picking cards, forming committees, forming permutations, etc. We will therefore be making copious use of Eq. (2.1),

$$p = \frac{\text{number of desired outcomes}}{\text{total number of possible outcomes}} \quad \text{(for equally likely outcomes)} \quad (2.26)$$

We won't, however, bother to specifically state each time that the different outcomes are all equally likely. Just remember that they are, and that this fact is necessary for Eq. (2.1) to be valid.

Before getting into the examples, let's start off with a problem-solving strategy that comes in very handy in certain situations.

2.3.1 The art of "not"

There are many setups in which the easiest way to calculate the probability of a given event A is not to calculate it directly, but rather to calculate the probability of "not A" and then subtract the result from 1. This yields $P(A)$ because we know from Eq. (2.17) that $P(A) = 1 - P(\text{not } A)$. The event "not A" is called the *complement* of the event A.

The most common situation of this type involves a question along the lines of, "What is the probability of obtaining at least one of such-and-such?" The "at least" part appears to make things difficult, because it could mean one, or two, or three, etc.

It would be at best rather messy, and at worst completely intractable, to calculate the individual probabilities of all the different numbers and then add them up to obtain the answer. The "at least one" question is very different from the "exactly one" question.

The key point that simplifies things is that the only way to *not* get at least one of something is to get *exactly zero* of it. This means that we can just calculate the probability of getting zero, and then subtract the result from 1. We therefore need to calculate only *one* probability, instead of a potentially large number of probabilities.

Example (At least one 6): Three dice are rolled. What is the probability of obtaining at least one 6?

Solution: We'll find the probability of obtaining zero 6's and then subtract the result from 1. In order to obtain zero 6's, we must obtain something other than a 6 on the first die (which happens with 5/6 probability), and likewise on the second die (5/6 probability again), and likewise on the third die (5/6 probability again). These are independent events, so the probability of obtaining zero 6's equals $(5/6)^3 = 125/216$. The probability of obtaining at least one 6 is therefore $1 - (5/6)^3 = 91/216$, which is about 42%.

If you want to solve this problem the long way, you can add up the probabilities of obtaining exactly one, two, or three 6's. This is the task of Problem 2.11.

REMARK: Beware of the following incorrect reasoning for this problem: There is a 1/6 chance of obtaining a 6 on each of the three rolls. The total probability of obtaining at least one 6 therefore seems like it should be $3 \cdot (1/6) = 1/2$. This is incorrect because we're trying to find the probability of "a 6 on the first roll" *or* "a 6 on the second roll" *or* "a 6 on the third roll." (This "or" combination is equivalent to obtaining at least one 6. Remember that when we write "or," we mean the "inclusive or.") But from Eq. (2.14) (or its simple extension to three events) it is appropriate to add up the individual probabilities *only if the events are exclusive.* For nonexclusive events, we must subtract off the "overlap" probabilities, as we did in Eq. (2.18); see Problem 2.2(d) for the case of three events. The above three events (rolling 6's) are clearly nonexclusive, because it is possible to obtain a 6 on, say, both the first roll *and* the second roll. We have therefore double (or triple) counted many of the outcomes, and this is why the incorrect answer of 1/2 is larger than the correct answer of 91/216. The task of Problem 2.12 is to solve this problem by using the result in Problem 2.2(d) to keep track of all the double (and triple) counting.

Another way of seeing why the "$3 \cdot (1/6) = 1/2$" reasoning can't be correct is that it would imply that if we had, say, 12 dice, then the probability of obtaining at least one 6 would be $12 \cdot (1/6) = 2$. But probabilities larger than 1 are nonsensical. ♣

2.3.2 Picking seats

Situations often come up where we need to assign various things to various spots. We'll generally talk about assigning people to seats. There are two common ways to solve problems of this sort: (1) You can count up the number of desired outcomes,

along with the total number of outcomes, and then take their ratio via Eq. (2.1), or (2) you can imagine assigning the seats one at a time, finding the probability of success at each stage, and using the rules in Section 2.2, or their extensions to more than two events. It's personal preference which method you use. But it never hurts to solve a problem both ways, of course, because that allows you to double check your answer.

Example 1 (Middle in the middle): Three chairs are arranged in a line, and three people randomly take seats. What is the probability that the person with the middle height ends up in the middle seat?

First solution: Let the people be labeled from tallest to shortest as 1, 2, and 3. Then the 3! = 6 possible orderings are

$$1\,2\,3 \quad 1\,3\,2 \quad 2\,1\,3 \quad 2\,3\,1 \quad 3\,1\,2 \quad 3\,2\,1 \qquad (2.27)$$

We see that two of these (1 2 3 and 3 2 1) have the middle-height person in the middle seat. So the probability is 2/6 = 1/3.

Second solution: Imagine assigning the people randomly to the seats, and let's assign the middle-height person first, which we are free to do. There is a 1/3 chance that this person ends up in the middle seat (or any other seat, for that matter). So 1/3 is the desired answer. Nothing fancy going on here.

Third solution: If you want to assign the tallest person first, then there is a 1/3 chance that she ends up in the middle seat, in which case there is zero chance that the middle-height person ends up there. There is a 2/3 chance that the tallest person *doesn't* end up in the middle seat, in which case there is a 1/2 chance that the middle-height person ends up there (because there are two seats remaining, and one yields success). So the total probability that the middle-height person ends up in the middle seat is

$$\frac{1}{3}\cdot 0 + \frac{2}{3}\cdot\frac{1}{2} = \frac{1}{3}. \qquad (2.28)$$

REMARK: The preceding equation technically comes from one application of Eq. (2.14) and two applications of Eq. (2.5). If we let T stand for tallest and M stand for middle-height, and if we use the notation T_{mid} to mean that the tallest person is in the middle seat, etc., then we can write

$$P(M_{\mathrm{mid}}) = P(T_{\mathrm{mid}} \text{ and } M_{\mathrm{mid}}) + P(T_{\mathrm{not\ mid}} \text{ and } M_{\mathrm{mid}})$$
$$= P(T_{\mathrm{mid}}) \cdot P(M_{\mathrm{mid}}|T_{\mathrm{mid}}) + P(T_{\mathrm{not\ mid}}) \cdot P(M_{\mathrm{mid}}|T_{\mathrm{not\ mid}})$$
$$= \frac{1}{3}\cdot 0 + \frac{2}{3}\cdot\frac{1}{2} = \frac{1}{3}. \qquad (2.29)$$

Eq. (2.14) is relevant in the first line because the two events "T_{mid} and M_{mid}" and "$T_{\mathrm{not\ mid}}$ and M_{mid}" are exclusive events, since T can't be both in the middle seat and not in the middle seat.

However, when solving problems of this kind, although it is sometimes helpful to explicitly write down the application of Eqs. (2.14) and (2.5) as we just did, this often isn't necessary. It is usually quicker to imagine a large number of trials and then calculate the number of these trials that yield success. For example, if we do 600 trials

of the present setup, then $(1/3) \cdot 600 = 200$ of them (on average) have T in the middle seat, in which case failure is guaranteed. Of the other $(2/3) \cdot 600 = 400$ trials where T isn't in the middle seat, half of them (which is $(1/2) \cdot 400 = 200$) have M in the middle seat. So the desired probability is $200/600 = 1/3$. In addition to being more intuitive, this method is safer than just plugging things into formulas (although it's really the same reasoning in the end). ♣

Example 2 (Order of height in a line): Five chairs are arranged in a line, and five people randomly take seats. What is the probability that they end up in order of decreasing height, from left to right?

First solution: There are $5! = 120$ possible arrangements of the five people in the seats. But there is only one arrangement where they end up in order of decreasing height. So the probability is $1/120$.

Second solution: If we randomly assign the tallest person to a seat, there is a $1/5$ chance that she ends up in the leftmost seat. Assuming that she ends up there, there is a $1/4$ chance that the second tallest person ends up in the second leftmost seat (because there are only four seats left). Likewise, the chances that the other people end up where we want them are $1/3$, then $1/2$, and then $1/1$. (If the first four people end up in the desired seats, then the shortest person is guaranteed to end up in the rightmost seat.) So the probability is $1/5 \cdot 1/4 \cdot 1/3 \cdot 1/2 \cdot 1/1 = 1/120$.

The product of these five probabilities comes from the extension of Eq. (2.5) to five events (see Problem 2.2(b) for the three-event case), which takes the form,

$$P(A \text{ and } B \text{ and } C \text{ and } D \text{ and } E) = P(A) \cdot P(B|A) \cdot P(C|A \text{ and } B)$$
$$\cdot P(D|A \text{ and } B \text{ and } C) \qquad (2.30)$$
$$\cdot P(E|A \text{ and } B \text{ and } C \text{ and } D).$$

We will use similar extensions repeatedly in the examples below.

Alternatively, instead of assigning people to seats, we can assign seats to people. That is, we can assign the first seat to one of the five people, and then the second seat to one of the remaining four people, and so on. Multiplying the probabilities of success at each stage gives the same product as above, $1/5 \cdot 1/4 \cdot 1/3 \cdot 1/2 \cdot 1/1 = 1/120$.

Example 3 (Order of height in a circle): Five chairs are arranged in a circle, and five people randomly take seats. What is the probability that they end up in order of decreasing height, going clockwise? The decreasing sequence of people can start anywhere in the circle. That is, it doesn't matter which seat has the tallest person.

First solution: As in the previous example, there are $5! = 120$ possible arrangements of the five people in the seats. But now there are *five* arrangements where they end up in order of decreasing height. This is true because the tallest person can take five possible seats, and once her seat is picked, the positions of the other people are uniquely determined if they are to end up in order of decreasing height. The probability is therefore $5/120 = 1/24$.

Second solution: If we randomly assign the tallest person to a seat, it doesn't matter where she ends up, because all five seats in the circle are equivalent. But given that she ends up in a certain seat, the second tallest person needs to end up in the seat next to her in the clockwise direction. This happens with probability 1/4. Likewise, the third tallest person has a 1/3 chance of ending up in the next seat in the clockwise direction. And then 1/2 for the fourth tallest person, and 1/1 for the shortest person. The probability is therefore $1/4 \cdot 1/3 \cdot 1/2 \cdot 1/1 = 1/24$.

If you want, you can preface this product with a "5/5" for the tallest person, because there are five possible seats she can take (this is the denominator), and there are also five successful seats she can take (this is the numerator) because it doesn't matter where she ends up.

Example 4 (Three girls and three boys): Six chairs are arranged in a line, and three girls and three boys randomly pick seats. What is the probability that the three girls end up in the three leftmost seats?

First solution: The total number of possible seat arrangements is 6! = 720. There are 3! = 6 different ways that the three girls can be arranged in the three leftmost seats, and 3! = 6 different ways that the three boys can be arranged in the other three (the rightmost) seats. So the total number of successful arrangements is $3! \cdot 3! = 36$. The desired probability is therefore $3!3!/6! = 36/720 = 1/20$.

Second solution: Let's assume that the girls pick their seats first, one at a time. The first girl has a 3/6 chance of picking one of the three leftmost seats. Then, given that she is successful, the second girl has a 2/5 chance of success, because only two of the remaining five seats are among the left three. And finally, given that she too is successful, the third girl has a 1/4 chance of success, because only one of the remaining four seats is among the left three. If all three girls are successful, then all three boys are guaranteed to end up in the three rightmost seats. The desired probability is therefore $3/6 \cdot 2/5 \cdot 1/4 = 1/20$.

Third solution: The $3!3!/6!$ result in the first solution looks suspiciously like the inverse of the binomial coefficient $\binom{6}{3} = 6!/3!3!$. This suggests that there is another way to solve the problem. And indeed, imagine randomly choosing three of the six seats for the girls. There are $\binom{6}{3}$ ways to do this, all equally likely. Only one of these is the successful choice of the three leftmost seats, so the desired probability is $1/\binom{6}{3} = 3!3!/6! = 1/20$.

2.3.3 Socks in a drawer

Picking colored socks from a drawer is a classic probabilistic setup. As usual, if you want to deal with such setups by counting things, then subgroups and binomial coefficients will come into play. If, however, you want to imagine picking the socks in succession, then you'll end up multiplying various probabilities and using the rules in Section 2.2.

Example 1 (Two blue and two red):   A drawer contains two blue socks and two red socks. If you randomly pick two socks, what is the probability that you obtain a matching pair?
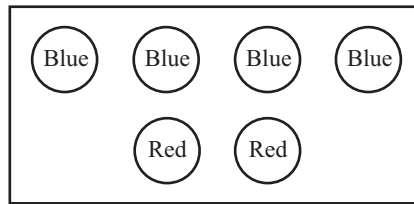
First solution: There are $\binom{4}{2} = 6$ possible pairs you can pick. Of these, two are matching pairs (one blue pair, one red pair). So the probability is $2/6 = 1/3$. If you want to list out all the pairs, they are (with 1 and 2 being the blue socks, and 3 and 4 being the red socks):

$$\textbf{1, 2} \qquad 1, 3 \qquad 1, 4 \qquad 2, 3 \qquad 2, 4 \qquad \textbf{3, 4} \tag{2.31}$$

The pairs in bold are the matching pairs.

Second solution: After you pick the first sock, there is one sock of that color (whatever it may be) left in the drawer, and two of the other color. So of the three socks left, one gives you a matching pair, and two don't. The desired probability is therefore $1/3$. See Problem 2.9 for a generalization of this example.

Example 2 (Four blue and two red):   A drawer contains four blue socks and two red socks, as shown in Fig. 2.9. If you randomly pick two socks, what is the probability that you obtain a matching pair?

Figure 2.9: A box with four blue socks and two red socks.

First solution: There are $\binom{6}{2} = 15$ possible pairs you can pick. Of these, there are $\binom{4}{2} = 6$ blue pairs and $\binom{2}{2} = 1$ red pair. The desired probability is therefore

$$\frac{\binom{4}{2} + \binom{2}{2}}{\binom{6}{2}} = \frac{7}{15} . \tag{2.32}$$

Second solution: There is a $4/6$ chance that the first sock you pick is blue. If this happens, there is a $3/5$ chance that the second sock you pick is also blue (because there are three blue and two red socks left in the drawer). Similarly, there is a $2/6$ chance that the first sock you pick is red. If this happens, there is a $1/5$ chance that the second sock you pick is also red (because there are one red and four blue socks left in the drawer). The probability that the socks match is therefore

$$\frac{4}{6} \cdot \frac{3}{5} + \frac{2}{6} \cdot \frac{1}{5} = \frac{14}{30} = \frac{7}{15} . \tag{2.33}$$

If you want to explicitly justify the sum on the lefthand side here, it comes from the sum on the righthand side of the following relation (with B_1 standing for a blue sock on the first pick, etc.):

$$P(B_1 \text{ and } B_2) + P(R_1 \text{ and } R_2) = P(B_1) \cdot P(B_2|B_1) + P(R_1) \cdot P(R_2|R_1). \quad (2.34)$$

However, equations like this can be a bit intimidating, so it's often better to think in terms of a large set of trials, as mentioned in the remark in the first example in Section 2.3.2.

2.3.4 Coins and dice

There is never a shortage of probability examples involving dice rolls or coin flips.

Example 1 (One of each number): Six dice are rolled. What is the probability of obtaining exactly one of each of the numbers 1 through 6?

First solution: The total number of possible (ordered) outcomes for what all six dice show is 6^6, because there are six possibilities for each die. How many outcomes are there that have each number appearing once? This is simply the question of how many permutations there are of six numbers, because we need all six numbers to appear, but it doesn't matter in what order. There are 6! permutations, so the desired probability is

$$\frac{6!}{6^6} = \frac{5}{324} \approx 1.5\%. \quad (2.35)$$

Second solution: Let's imagine rolling six dice in succession, with the goal of having each number appear once. On the first roll, we get what we get, and there's no way to fail. So the probability of success on the first roll is 1. However, on the second roll, we don't want to get a repeat of the number that appeared on the first roll (whatever that number happened to be). Since there are five "good" options left, the probability of success on the second roll is 5/6. On the third roll, we don't want to get a repeat of either of the numbers that appeared on the first and second rolls, so the probability of success on the third roll (given success on the first two rolls) is 4/6. Likewise, the fourth roll has a 3/6 chance of success, the fifth has 2/6, and the sixth has 1/6. The probability of complete success all the way through is therefore

$$1 \cdot \frac{5}{6} \cdot \frac{4}{6} \cdot \frac{3}{6} \cdot \frac{2}{6} \cdot \frac{1}{6} = \frac{5}{324}, \quad (2.36)$$

in agreement with the first solution. Note that if we write the initial 1 here as 6/6, then this expression becomes $6!/6^6$, which is the fraction that appears in Eq. (2.35).

Example 2 (Three pairs): Six dice are rolled. What is the probability of getting three pairs, that is, three different numbers that each appear twice?

Solution: We'll count the total number of (ordered) ways to get three pairs, and then we'll divide that by the total number of possible (ordered) outcomes for the six rolls, which is 6^6.

There are two steps in the counting. First, how many different ways can we pick the three different numbers that show up? We need to pick three numbers from six, so the number of ways is $\binom{6}{3} = 20$.

Second, given the three numbers that show up, how many different (ordered) ways can two of each appear on the dice? Let's says the numbers are 1, 2, and 3. We can imagine plopping two of each of these numbers down on six blank spots (which represent the six dice) on a piece of paper. There are $\binom{6}{2} = 15$ ways to pick where the two 1's go. And then there are $\binom{4}{2} = 6$ ways to pick where the two 2's go in the four remaining spots. And then finally there is $\binom{2}{2} = 1$ way to pick where the two 3's go in the two remaining spots.

The total number of ways to get three pairs is therefore $\binom{6}{3} \cdot \binom{6}{2} \cdot \binom{4}{2} \cdot \binom{2}{2}$. So the probability of getting three pairs is

$$p = \frac{\binom{6}{3} \cdot \binom{6}{2} \cdot \binom{4}{2} \cdot \binom{2}{2}}{6^6} = \frac{20 \cdot 15 \cdot 6 \cdot 1}{6^6} = \frac{25}{648} \approx 3.9\%. \tag{2.37}$$

If you try to solve this problem in a manner analogous to the second solution in the previous example (that is, by multiplying probabilities for the successive rolls), then things get a bit messy because there are many different scenarios that lead to three pairs.

Example 3 (Five coin flips): A coin is flipped five times. Calculate the probabilities of getting the various possible numbers of Heads (0 through 5).

Solution: We'll count the number of (ordered) ways to get the different numbers of Heads, and then we'll divide that by the total number of possible (ordered) outcomes for the five flips, which is 2^5.

There is only $\binom{5}{0} = 1$ way to get zero Heads, namely TTTTT. There are $\binom{5}{1} = 5$ ways to get one Heads (such as HTTTT), because there are $\binom{5}{1}$ ways to choose the one coin that shows Heads. There are $\binom{5}{2} = 10$ ways to get two Heads, because there are $\binom{5}{2}$ ways to choose the two coins that show Heads. And so on. The various probabilities are therefore

$$P(0) = \frac{\binom{5}{0}}{2^5}, \quad P(1) = \frac{\binom{5}{1}}{2^5}, \quad P(2) = \frac{\binom{5}{2}}{2^5},$$

$$P(3) = \frac{\binom{5}{3}}{2^5}, \quad P(4) = \frac{\binom{5}{4}}{2^5}, \quad P(5) = \frac{\binom{5}{5}}{2^5}. \tag{2.38}$$

Plugging in the values of the binomial coefficients gives

$$P(0) = \frac{1}{32}, \quad P(1) = \frac{5}{32}, \quad P(2) = \frac{10}{32},$$

$$P(3) = \frac{10}{32}, \quad P(4) = \frac{5}{32}, \quad P(5) = \frac{1}{32}. \tag{2.39}$$

The sum of all these probabilities correctly equals 1. The physical reason for this is that the number of Heads must be *something*, which means that the sum of all the

probabilities must be 1. (This holds for any number of flips, of course, not just 5.) The mathematical reason is that the sum of the binomial coefficients (the numerators in the above fractions) equals 2^5 (which is the denominator). See Section 1.8.3 for the explanation of this.

2.3.5 Cards

We already did a lot of card counting in Chapter 1 (particularly in Problem 1.10), and some of those results will be applicable here. As we have mentioned a number of times, exercises in probability are often just exercises in counting. There is effectively an endless number of probability questions we can ask about cards. In the following examples, we will always assume a standard 52-card deck.

Example 1 (Royal flush from seven cards): A few variations of poker involve being dealt seven cards (in one way or another) and forming the best five-card hand that can be made from these seven cards. What is the probability of being able to form a Royal flush in this setup? A Royal flush consists of $10, J, Q, K, A$, all from the same suit.

Solution: The total number of possible seven-card hands is $\binom{52}{7} = 133,784,560$. The number of seven-card hands that contain a Royal flush is $4 \cdot \binom{47}{2} = 4,324$, because there are four ways to choose the five Royal flush cards (the four suits), and then $\binom{47}{2}$ ways to choose the other two cards from the remaining $52 - 5 = 47$ cards in the deck. The probability is therefore

$$\frac{4 \cdot \binom{47}{2}}{\binom{52}{7}} = \frac{4,324}{133,784,560} \approx 0.0032\%. \tag{2.40}$$

This is larger than the result for five-card hands. In that case, only four of the $\binom{52}{5} = 2,598,960$ hands are Royal flushes, so the probability is $4/2,598,960 \approx 0.00015\%$, which is about 20 times smaller than 0.0032%. As an exercise, you can show that the ratio happens to be exactly 21.

Example 2 (Suit full house): In a five-card poker hand, what is the probability of getting a "full house" of suits, that is, three cards of one suit and two of another? (This isn't an actual poker hand worth anything, but that won't stop us from calculating the probability!) How does your answer compare with the probability of getting an actual full house, that is, three cards of one value and two of another? Feel free to use the result from part (a) of Problem 1.10.

Solution: There are four ways to choose the suit that appears three times, and $\binom{13}{3} = 286$ ways to choose the specific three cards from the 13 of this suit. And then there are three ways to choose the suit that appears twice from the remaining three suits, and $\binom{13}{2} = 78$ ways to choose the specific two cards from the 13 of this suit. The total

number of suit-full-house hands is therefore $4 \cdot \binom{13}{3} \cdot 3 \cdot \binom{13}{2} = 267{,}696$. Since there is a total of $\binom{52}{5}$ possible hands, the desired probability is

$$\frac{4 \cdot \binom{13}{3} \cdot 3 \cdot \binom{13}{2}}{\binom{52}{5}} = \frac{267{,}696}{2{,}598{,}960} \approx 10.3\%. \tag{2.41}$$

From part (a) of Problem 1.10, the total number of actual full-house hands is 3,744, which yields a probability of $3{,}744/2{,}598{,}960 \approx 0.14\%$. It is therefore much more likely (by a factor of about 70) to get a full house of suits than an actual full house of values. (You can show that the exact ratio is 71.5.) This makes intuitive sense; there are more values than suits (13 compared with four), so it is harder to have all five cards involve only two values as opposed to only two suits.

Example 3 (Only two suits): In a five-card poker hand, what is the probability of having all of the cards be members of at most two suits? (A single suit falls into this category.) The suit full house in the previous example is a special case of "at most two suits." This problem is a little tricky, at least if you solve it a certain way; be careful about double counting some of the hands!

First solution: If two suits appear, then there are $\binom{4}{2} = 6$ ways to pick them. For a given choice of two suits, there are $\binom{26}{5}$ ways to pick the five cards from the $2 \cdot 13 = 26$ cards of these two suits. It therefore seems like there should be $\binom{4}{2} \cdot \binom{26}{5} = 394{,}680$ different hands that consist of cards from at most two suits.

However, this isn't correct, because we double (or actually triple) counted the hands that involve only one suit (the flushes). For example, if all five cards are hearts, then we counted such a hand in the heart/diamond set of $\binom{26}{5}$ hands, and also in the heart/spade set, and also in the heart/club set. We counted it three times when we should have counted it only once. Since there are $\binom{13}{5}$ hands that are heart flushes, we have included an extra $2 \cdot \binom{13}{5}$ hands, so we need to subtract these from our total. Likewise for the diamond, spade, and club flushes. The total number of hands that involve at most two suits is therefore

$$\binom{4}{2}\binom{26}{5} - 4 \cdot 2 \cdot \binom{13}{5} = 394{,}680 - 10{,}296 = 384{,}384. \tag{2.42}$$

The desired probability is then

$$\frac{\binom{4}{2}\binom{26}{5} - 8 \cdot \binom{13}{5}}{\binom{52}{5}} = \frac{384{,}384}{2{,}598{,}960} \approx 14.8\%. \tag{2.43}$$

This is larger than the result in Eq. (2.41), as it should be, because suit full houses are a subset of the hands that involve at most two suits.

Second solution: There are three general ways that we can have at most two suits: (1) all five cards can be of the same suit (a flush), (2) four cards can be of one suit, and one card of another, or (3) three cards can be of one suit, and two cards of another; this is the suit full house from the previous example. We will denote these types of hands by $(5,0)$, $(4,1)$, and $(3,2)$, respectively. How many hands of each type are there?

There are $4 \cdot \binom{13}{5} = 5,148$ hands of the $(5,0)$ type, because there are $\binom{13}{5}$ ways to pick five cards from the 13 cards of a given suit, and there are four suits. From the previous example, there are $4 \cdot \binom{13}{3} \cdot 3 \cdot \binom{13}{2} = 267,696$ hands of the $(3,2)$ type. To figure out the number of hands of the $(4,1)$ type, we can use exactly the same kind of reasoning as in the previous example. This gives $4 \cdot \binom{13}{4} \cdot 3 \cdot \binom{13}{1} = 111,540$ hands. Adding up these three results gives the total number of "at most two suits" hands as

$$4 \cdot \binom{13}{5} + 4 \cdot \binom{13}{4} \cdot 3 \cdot \binom{13}{1} + 4 \cdot \binom{13}{3} \cdot 3 \cdot \binom{13}{2}$$

$$= 5,148 + 111,540 + 267,696$$

$$= 384,384, \tag{2.44}$$

in agreement with the first solution. (The repetition of the "384" here is due in part to the factors of 13 and 11 in all of the terms in the first line of Eq. (2.44). These numbers are factors of 1001.) The hands of the $(3,2)$ type account for about 2/3 of the total, consistent with the fact that the 10.3% result in Eq. (2.41) is about 2/3 of the 14.8% result in Eq. (2.43).

2.4 Four classic problems

Let's now look at four classic probability problems. No book on probability would be complete without a discussion of the "Birthday Problem" and the "Game-Shown Problem." Additionally, the "Prosecutor's Fallacy" and the "Boy/Girl Problem" are two other classics that are instructive to study in detail. All four of these problems have answers that might seem counterintuitive at first, but they eventually make sense if you think about them long enough!

After reading the statement of each problem, be sure to try solving it on your own before looking at the solution. If you can't solve it on your first try, set it aside and come back to it later. There's no hurry; the problem will still be there. There are only so many classic problems like these, so don't waste them. If you look at a solution too soon, the opportunity to solve it is gone, and it's never coming back. If you do eventually need to look at the solution, cover it up with a piece of paper and read one line at a time, to get a hint. That way, you can still (mostly) solve it on your own.

2.4.1 The Birthday Problem

We'll present the Birthday Problem first. Aside from being a very interesting problem, its unexpected result allows you to take advantage of unsuspecting people and win money on bets at parties (as long as they're large enough parties, as we'll see!).

Problem: How many people need to be in a room in order for there to be a greater than 1/2 probability that at least two of them have the same birthday? By "same birthday" we mean the same day of the year; the year may differ. Ignore leap years.

(At this point, as with all of the problems in this section, don't read any further until you've either solved the problem or thought hard about it for a long time.)

Solution: If there was ever a problem that called for the "art of not" strategy in Section 2.3.1, this is it. There are many different ways for there to be *at least* one common birthday (one pair, two pairs, one triple, etc.), and it is completely intractable to add up all of these individual probabilities. It is *much* easier (and even with the italics, this is a vast understatement) to calculate the probability that there *isn't* a common birthday, and then subtract this from 1 to obtain the probability that there *is* at least one common birthday.

The calculation of the probability that there *isn't* a common birthday proceeds as follows. Let there be n people in the room. We can imagine taking them one at a time and randomly plopping their names down on a calendar, with the (present) goal being that there are no common birthdays. The first name can go anywhere. But when we plop down the second name, there are only 364 "good" days left, because we don't want the day to coincide with the first name's day. The probability of success for the second name is therefore 364/365. Then, when we plop down the third name, there are only 363 "good" days left (assuming that the first two people have different birthdays), because we don't want the day to coincide with either of the other two days. The probability of success for the third name is therefore 363/365. Similarly, when we plop down the fourth name, there are only 362 "good" days left (assuming that the first three people have different birthdays). The probability of success for the fourth name is therefore 362/365. And so on.

If there are n people in the room, the probability that all n birthdays are distinct (that is, there *isn't* a common birthday among any of the people; hence the superscript "no" below) therefore equals

$$P_n^{\text{no}} = 1 \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \frac{362}{365} \cdot \frac{361}{365} \cdot \dots \cdot \frac{365 - (n-1)}{365}. \qquad (2.45)$$

If you want, you can write the initial 1 here as 365/365, to make things look nicer. Note that the last term involves $(n-1)$ and not n, because $(n-1)$ is the number of names that have already been plopped down. As a double check that this $(n-1)$ is correct, it works for small numbers like $n = 2$ and 3. You should always perform a simple check like this whenever you write down *any* expression involving a parameter such as n.

We now just have to multiply out the product in Eq. (2.45) to the point where it becomes smaller than 1/2, so that the probability that there *is* a common birthday is larger than 1/2. With a calculator, this is tedious, but not horribly painful. We find that $P_{22}^{\text{no}} = 0.524$ and $P_{23}^{\text{no}} = 0.493$. If P_n^{yes} is the probability that there *is* a common birthday among n people, then $P_n^{\text{yes}} = 1 - P_n^{\text{no}}$, so $P_{22}^{\text{yes}} = 0.476$ and $P_{23}^{\text{yes}} = 0.507$. Since our original goal was to have $P_n^{\text{yes}} > 1/2$ (or equivalently $P_n^{\text{no}} < 1/2$), we see that there must be at least 23 people in a room in order for there to be a greater than 50% chance that at least two of them have the same birthday. The probability in the $n = 23$ case is 50.7%.

The task of Problem 2.14 is to calculate the probability that among 23 people, *exactly* two of them have a common birthday. That is, there aren't two different pairs with common birthdays, or a triple with the same birthday, etc.

REMARK: The $n = 23$ answer to our problem is much smaller than most people would expect. As mentioned above, it therefore provides a nice betting opportunity. For $n = 30$, the probability of a common birthday increases to 70.6%, and most people would still find it hard to believe that among 30 people, there are probably two who have the same birthday. Table 2.4 lists various values of n and the probabilities, $P_n^{\text{yes}} = 1 - P_n^{\text{no}}$, that at least two people have a common birthday.

n	10	20	23	30	50	60	70	100
P_n^{yes}	11.7%	41.1%	50.7%	70.6%	97.0%	99.4%	99.92%	99.99997%

Table 2.4: Probability of a common birthday among n people.

Even for $n = 50$, most people would probably be happy to bet, at even odds, that no two people have the same birthday. But you'll win the bet 97% of the time.

One reason why many people can't believe the $n = 23$ result is that they're asking themselves a different question, namely, "How many people (in addition to me) need to be present in order for there to be at least a 1/2 chance that someone else has *my* birthday?" The answer to this question is indeed much larger than 23. The probability that *no one* out of n people has a birthday on a *given day* is simply $(364/365)^n$, because each person has a 364/365 chance of not having that particular birthday. For $n = 252$, this is just over 1/2. And for $n = 253$, it is just under 1/2; it equals 0.4995. Therefore, you need to come across 253 other people in order for the probability to be greater than 1/2 that at least one of them *does* have *your* birthday (or any other particular birthday). See Problem 2.16 for further discussion of this. ♣

2.4.2 The Game-Show Problem

We'll now discuss the Game-Show Problem. In addition to having a variety of common incorrect solutions, this problem also also a long history of people arguing vehemently in favor of those incorrect solutions.

Problem: A game-show host offers you the choice of three doors. Behind one of these doors is the grand prize, and behind the other two are goats. The host (who knows what is behind each of the doors) announces that after you select a door (without opening it), he will open one of the other two doors and purposefully reveal a goat. You select a door. The host then opens one of the other doors and reveals the promised goat. He then offers you the chance to switch your choice to the remaining door. To maximize the probability of winning the grand prize, should you switch or not? Or does it not matter?

Solution: We'll present three solutions, one right and two wrong. You should decide which one you think is correct before reading beyond the third solution. Cover up the page after the third solution with a piece of paper, so that you don't inadvertently see which one is correct.

- REASONING 1: Once the host reveals a goat, the prize must be behind one of the two remaining doors. Since the prize was randomly located to begin with, there must be equal chances that the prize is behind each of the two remaining doors. The probabilities are therefore both 1/2, so it doesn't matter if you switch.

If you want, you can imagine a friend (who is aware of the whole procedure of the host announcing that he will open a door and reveal a goat) entering the room *after* the host opens the door. This person sees two identical unopened doors (he doesn't know which one you initially picked) and a goat. So for him there must be a 1/2 chance that the prize is behind each unopened door. The probabilities for you and your friend can't be any different, so you also say that each unopened door has a 1/2 chance of containing the prize. It therefore doesn't matter if you switch.

- REASONING 2: There is initially a 1/3 chance that the prize is behind any of the three doors. So if you don't switch, your probability of winning is 1/3. No actions taken by the host can change the fact that if you play a large number n of these games, then (roughly) $n/3$ of them will have the prize behind the door you initially pick.

  Likewise, if you switch to the other unopened door, there is a 1/3 chance that the prize is behind that door. (There is obviously a goat behind at least one of the other two doors, so the fact that the host reveals a goat doesn't tell you anything new.) Therefore, since the probability is 1/3 whether or not you switch, it doesn't matter if you switch.

- REASONING 3: As in the first paragraph of Reasoning 2, if you don't switch, your probability of winning is 1/3.

  However, if you switch, your probability of winning is greater than 1/3. It increases to 2/3. This can be seen as follows. Without loss of generality, assume that you pick the first door. (You can repeat the following reasoning for the other doors if you wish. It gives the same result.) There are three equally likely possibilities for what is behind the three doors: PGG, GPG, and GGP, where P denotes the prize and G denotes a goat. If you don't switch, then in only the first of these three cases do you win, so your odds of winning are 1/3 (consistent with the first paragraph of Reasoning 2). But if you do switch from the first door to the second or third, then in the first case PGG you lose, but in the other two cases you win, because the door not opened by the host has the prize. (The host has no choice but to reveal the G and leave the P unopened.) Therefore, since two out of the three equally likely cases yield success if you switch, your probability of winning if you switch is 2/3. So you do in fact want to switch.

Which of these three solutions is correct? Don't read any further until you've firmly decided which one you think is right.

The third solution is correct. The error in the first solution is the statement, "there must be equal chances that the prize is behind each of the two remaining doors." This is simply not true. The act of revealing a goat breaks the symmetry between the two remaining doors, as explained in the third solution. One door is the one you initially picked, while the other door is one of the two that you didn't pick. The fact that there are two possibilities doesn't mean that their probabilities have to be equal, of course!

The error in the supporting reasoning with your friend (who enters the room after the host opens the door) is the following. While it *is* true that both probabilities are 1/2 for your friend, they aren't both 1/2 for *you*. The statement, "the probabilities for you and your friend can't be any different," is false. You have information that your friend doesn't have; you know which of the two unopened doors is the one you initially picked and which is the door that the host chose to leave unopened. (And as seen in the third solution, this information yields probabilities of 1/3 and 2/3.) Your friend doesn't have this critical information. Both doors look the same to him. Probabilities can certainly be different for different people. If I flip a coin and peek and see a Heads, but I don't show you, then the probability of a Heads is 1/2 for you, but 1 for me.

The error in the second solution is that the act of revealing a goat *does* give you new information, as we just noted. This information tells you that the prize isn't behind that door, and it also distinguishes between the two remaining unopened doors. One is the door you initially picked, while the other is one of the two doors that you didn't initially pick. As seen in the third solution, this information has the effect of increasing the probability that the goat is behind the other door. Note that another reason why the second solution can't be correct is that the two probabilities of 1/3 don't add up to 1.

To sum up, it should be no surprise that the probabilities are different for the switching and non-switching strategies *after* the host opens a door (the probabilities are obviously the same, equal to 1/3, whether or not a switch is made *before* the host opens a door), because the host gave you some of the information he had about the locations of things.

REMARKS:

1. If you still doubt the validity of the third solution, imagine a situation with 1000 doors containing one prize and 999 goats. After you pick a door, the host opens 998 other doors and reveals 998 goats (and he said beforehand that he was going to do this). In this setup, if you don't switch, your chances of winning are 1/1000. But if you do switch, your chances of winning are 999/1000, which can be seen by listing out (or imagining listing out) the 1000 cases, as we did with the three PGG, GPG, and GGP cases in the third solution. It is clear that the switch should be made, because the *only* case where you lose after you switch is the case where you had initially picked the prize, and this happens only 1/1000 of the time.

   In short, a huge amount of information is gained by the revealing of 998 goats. There is initially a 999/1000 chance that the prize is somewhere behind the other 999 doors, and the host is kindly giving you the information of exactly which door it is (in the highly likely event that it is in fact one of the other 999).

2. The clause in the statement of the problem, "The host announces that after you select a door (without opening it), he will open one of the other two doors and purposefully reveal a goat," is crucial. If it is omitted, and it is simply stated that, "The host then opens one of the other doors and reveals a goat," then it is impossible to state a pre-ferred strategy. If the host doesn't announce his actions beforehand, then for all you know, he *always* reveals a goat (in which case you should switch, as we saw above). Or he *randomly* opens a door and just happened to pick a goat (in which case it doesn't matter if you switch, as you can show in Problem 2.18). Or he opens a door and reveals a goat if and only if your initial door has the prize (in which case you definitely should

not switch). Or he could have one procedure on Tuesdays and another on Fridays, each of which depends on the color of the socks he's wearing. And so on.

3. As mentioned above, this problem is infamous for the intense arguments it lends itself to. There's nothing terrible about getting the wrong answer, nor is there anything terrible about not believing the correct answer for a while. But concerning arguments that drag on and on, it doesn't make any sense to argue about this problem for more than, say, 20 minutes, because at that point everyone should stop and just *play the game*! You can play a number of times with the switching strategy, and then a number of times with the non-switching strategy. Three coins with a dot on the bottom of one of them are all you need.[1] Not only will the actual game yield the correct answer (if you play enough times so that things average out), but the patterns that form will undoubtedly convince you of the correct reasoning (or reinforce it, if you're already comfortable with it). Arguing endlessly about an experiment, when you can actually *do* the experiment, is as silly as arguing endlessly about what's behind a door, when you can simply open the door.

4. For completeness, there is one subtlety we should mention here. In the second solution, we stated, "No actions taken by the host can change the fact that if you play a large number n of these games, then (roughly) $n/3$ of them will have the prize behind the door you initially pick." This part of the reasoning was correct; it was the "switching" part of the second solution that was incorrect. After doing Problem 2.18 (where the host randomly opens a door), you might disagree with the above statement, because it will turn out in that problem that the actions taken by the host *do* affect this $n/3$ result. However, the above statement is still correct for "*these* games" (the ones governed by the original statement of this problem). See the second remark in the solution to Problem 2.18 for further discussion. ♣

2.4.3   The Prosecutor's Fallacy

We now present one of the most classic problems/paradoxes in the subject of probability. This classic nature is due in no small part to the problem's critical relevance to the real world. After reading the statement of the problem below, you should think carefully and settle on an answer before looking at the solution. The discussion of conditional probability in Section 2.2.4 gives a hint at the answer.

Problem: Consider the following scenario. Detectives in a city, say, Boston (whose population we will assume to be one million), are working on a crime and have put together a description of the perpetrator, based on things such as height, a tattoo, a limp, an earing, etc. Let's assume that only one person in $10,000$ fits the description. On a routine patrol the next day, police officers see a person fitting the description. This person is arrested and brought to trial based solely on the fact that he fits the description.

During the trial, the prosecutor tells the jury that since only one person in $10,000$ fits the description (a true statement), it is highly unlikely (far beyond a reasonable doubt) that an innocent person fits the description (again a true statement); it is

---

[1]You actually don't need three objects. It's hard to find three exactly identical coins anyway. The "host" can simply roll a die, without showing the "contestant" the result. Rolling a 1 or 2 can mean that the prize is located behind the first door, a 3 or 4 the second, and a 5 or 6 the third. The game then basically involves calling out door numbers.

therefore highly unlikely that the defendant is innocent. If you were a member of the jury, would you cast a "guilty" vote? If yes, what is your level of confidence? If no, what is wrong with the prosecutor's reasoning?

Solution: We'll assume that we are concerned only with people living in Boston. There are one million such people, so if one person in 10,000 fits the description, this means that there are 100 people in Boston who fit it (one of whom is the perpetrator). When the police officers pick up someone fitting the description, this person could be any one of these 100 people. So the probability that the defendant in the courtroom is the actual perpetrator is only 1/100. In other words, there is a 99% chance that the person is innocent. A guilty verdict (based on the given evidence) would therefore be a horrible and tragic vote.

The above (correct) reasoning is fairly cut and dry, but it contradicts the prosecutor's reasoning. The prosecutor's reasoning must therefore be incorrect. But what exactly is wrong with it? It seems quite plausible at every stage. To isolate the flaw in the logic, let's list out the three separate statements the prosecutor made in his argument:

1. Only one person in 10,000 fits the description.

2. It is highly unlikely (far beyond a reasonable doubt) that an innocent person fits the description.

3. It is therefore highly unlikely that the defendant is innocent.

As we noted above when we posed the problem, the first two of these statements are true. Statement 1 is true by assumption, and Statement 2 is true basically because 1/10,000 is a small number. Let's be precise about this and work out the exact probability that an innocent person fits the description. Of the one million people in Boston, the number who fit the description is $(1/10,000)(10^6) = 100$. Of these 100 people, only one is guilty, so 99 are innocent. And the total number of innocent people is $10^6 - 1 = 999,999$. The probability that an innocent person fits the description is therefore

$$\frac{\text{innocent and fitting description}}{\text{innocent}} = \frac{99}{999,999} \approx 9.9 \cdot 10^{-5} \approx \frac{1}{10,000} \,. \qquad (2.46)$$

As expected, the probability is essentially equal to 1/10,000.

Now let's look at the third statement above. This is where the error is. This statement is false, because Statement 2 simply does not imply Statement 3. We know this because we have already calculated the probability that the defendant is innocent, namely 99%. This correct probability of 99% is vastly different from the incorrect probability of 1/10,000 that the prosecutor is trying to mislead you with. However, even though the correct result of 99% tells us that Statement 3 must be false, where exactly is the error? After all, at first glance Statement 3 *seems* to follow from Statement 2. The error is the confusion of conditional probabilities. In detail:

- Statement 2 deals with the probability of fitting the description, *given* innocence. The (true) statement is equivalent to, "*If* a person is innocent, *then* there is a very small probability that he fits the description." This probability is the conditional probability $P(\text{D}|\text{I})$, with D for description and I for innocence.

- Statement 3 deals with the probability of innocence, *given* that the description is fit. The (false) statement is equivalent to, "*If* a person (such as the defendant) fits the description, *then* there is a very small probability that he is innocent." This probability is the conditional probability $P(\text{I}|\text{D})$.

These two conditional probabilities are *not* the same. The error is the assumption (or implication, on the prosecutor's part) that they are. As we saw above, $P(\text{D}|\text{I}) = 99/999,999 \approx 0.0001$, whereas $P(\text{I}|\text{D}) = 0.99$. These two probabilities are markedly different.

Intuitively, $P(\text{D}|\text{I})$ is very small because a very small fraction of the population (in particular, a very small fraction of the innocent people) fit the description. And $P(\text{I}|\text{D})$ is very close to 1 because nearly everyone (in particular, nearly everyone who fits the description) is innocent. This state of affairs is indicated in Fig. 2.10. (This a just a rough figure; the areas aren't actually in the proper proportions.) The large oval represents the 999,999 innocent people, and the small oval represents the 100 people who fit the description.
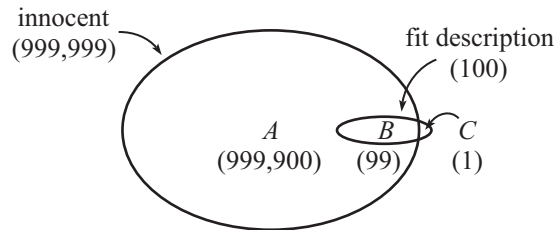
Figure 2.10: The different types of people in the prosecutor's fallacy.

There are three basic types of people in the figure: There are $A = 999,900$ innocent people who don't fit the description, $B = 99$ innocent people who do fit the description, and $C = 1$ guilty person who fits the description. (The fourth possibility – a guilty person who doesn't fit the description – doesn't exist.) The two conditional probabilities that are relevant in the above discussion are then

$$P(\text{D}|\text{I}) = \frac{B}{\text{innocent}} = \frac{B}{B+A} = \frac{99}{999,999},$$

$$P(\text{I}|\text{D}) = \frac{B}{\text{fit description}} = \frac{B}{B+C} = \frac{99}{100}. \qquad (2.47)$$

Both of these probabilities have B in numerator, because B represents the people who are innocent *and* fit the description. But the A in the first denominator is much larger than the C in second denominator. Or said in another way, B is a very small fraction of the innocent people (the large oval in Fig. 2.10), whereas it is a very large fraction of the people who fit the description (the small oval in Fig. 2.10).

The prosecutor's faulty reasoning has been used countless times in actual court cases, with tragic consequences. Innocent people have been convicted, and guilty people have walked free (the argument can work in that direction too). These consequences can't be blamed on the jury, of course. It is inevitable that many jurors will fail to spot the error in the reasoning. It would be silly to think that the entire population should be familiar with this issue in probability. Nor can the blame be put on the attorney making the argument. This person is either (1) overzealous and/or incompetent, or (2) entirely within his/her right to knowingly make an invalid argument (as distasteful as this may seem). In the end, the blame falls on either (1) the opposing attorney for failing to rebut the known logical fallacy, or (2) a legal system that in some cases doesn't allow a final rebuttal.

2.4.4 The Boy/Girl Problem

The well-known Boy/Girl Problem can be stated in many different ways, with answers that may or may not be the same. Three different formulations are presented below, and a fourth is given in Problem 2.19. Assume in all of them that any process involved in the scenario is completely random. That is, assume that any child is equally likely to be a boy or a girl (even though this isn't quite true in real life), and assume that there is nothing special about the person you're talking with, and assume that there are no correlations between children (as there are with identical twins), and so on.

Problem:

(a) You bump into a random person on the street who says, "I have two children. At least one of them is a boy." What is the probability that the other child is also a boy?

(b) You bump into a random person on the street who says, "I have two children. The older one is a boy." What is the probability that the other child is also a boy?

(c) You bump into a random person on the street who says, "I have two children, one of whom is this boy standing next to me." What is the probability that the other child is also a boy?

Solution:

(a) The key to all three of these formulations is to list out the various equally likely possibilities for the family's children, while taking into account only the "I have two children" information, and *not yet* the information about the boy. With B for boy and G for girl, the family in the present scenario in part (a) can be of four types (at least *before* the parent gives you information about the boy), each with probability 1/4:

$$\boxed{BB} \quad \boxed{BG} \quad \boxed{GB} \quad GG$$

Ignore the boxes for a moment. In each pair of letters, the first letter stands for the older child, and the second letter stands for the younger child.

Note that there are indeed four equally likely possibilities (BB, BG, GB, GG), as opposed to just three equally likely possibilities (BB, BG, GG), because the older child has a 50-50 chance of being a boy or a girl, as does the younger child. The BG and GB cases each get counted once, just as the HT and TH cases each get counted once when flipping two coins, where the four equally likely possibilities are HH, HT, TH, TT.

Under the assumption of general randomness stated in the problem, we are assuming that you are equally likely (at least *before* the parent gives you information about the boy) to bump into a parent of any one of the above four types of two-child families.

Let us *now* invoke the information that at least one child is a boy. This information tells us that you can't be talking with a GG parent. The parent must be a BB, BG, or GB parent, all equally likely. (They are equally likely, because they are all equivalent with regard to the "at least one of them is a boy" statement.) These are the boxed families in the above list. Of these three cases, only the BB case has the other child being a boy. The desired probability that the other child is a boy is therefore 1/3.

If don't trust the reasoning in the preceding paragraph, just imagine performing many trials of the setup. This is always a good strategy when solving probability problems. Imagine that you encounter 1000 random parents of two children. You will encounter about 250 of each of the four types of parent. The 250 GG parents have nothing to do with the given setup, so we must discard them. Only the other 750 parents (BB, BG, GB) are able to provide the given information that at least one child is a boy. Of these 750 parents, 250 are of the BB type and thereby have a boy as the other child. The desired probability is therefore 250/750 = 1/3.

(b) As in part (a), *before* the information about the boy is taken into account, there are four equally likely possibilities for the children (again ignore the boxes for a moment):

$$\boxed{\text{BB}} \quad \boxed{\text{BG}} \quad \text{GB} \quad \text{GG}$$

But once the parent tells you that the older child is a boy, the GB and GG cases are ruled out; remember that the first letter in each pair corresponds to the older child. So you must be talking with a BB or BG parent, both equally likely. Of these two cases, only the BB case has the other child being a boy. The desired probability that the other child is a boy is therefore 1/2.

(c) This version of the problem is a little trickier, because there are now *eight* equally likely possibilities (before the information about the boy is taken into account), instead of just four. This is true because for each of the four types of families in the above lists, the parent may choose to take either of the children for a walk (with equal probabilities, as we are assuming for everything). The

eight equally likely possibilities are therefore shown in Fig. 2.5 (again ignore the boxes for a moment). The bold letter indicates the child you encounter.

| **BB** | **BG** | GB | GG |

| **BB** | B**G** | **GB** | GG |

Table 2.5: The eight types of families, accounting for the child present.

Once the parent tells you that one of the children is the boy standing there, four of the eight possibilities are ruled out. Only the four boxed pairs in Fig. 2.5 (the ones with a bold **B**) satisfy the condition that the child standing there is a boy. Of these four (equally likely) possibilities, two of them have the other child being a boy. The desired probability that the other child is a boy is therefore 1/2.

REMARKS:

1. We used the given assumption of general randomness many times in the above solutions. One way to make things *non*random is to assume that the parent who is out for a walk is chosen randomly with equal 1/3 probabilities of being from BB families, or GG families, or one-boy-and-one-girl families. This is an artificial construction, because it means that a given BG or GB family (which together make up half of all two-child families) is less likely to be chosen than a given BB or GG family. This violates our assumption of general randomness. In this scenario, you can show that the answers to parts (a), (b), and (c) are 1/2, 2/3, and 2/3.

   Another way to make things nonrandom is to assume that in part (c) a girl is always chosen to go on the walk if the family has at least one girl. The answer to part (c) is then 1, because the only way a boy will be standing there is if both children are boys. On the other hand, if we assume that a boy is always chosen to go on the walk if the family has at least one boy, then the answer to part (c) is 1/3. This is true because for BB, the other child is a boy; and for both BG and GB (for which the boy is always chosen to go on the walk), the other child is a girl. Basically, the middle four pairs in Table 2.5 will all have a bold **B**, so they will all be boxed. There are countless ways to make things nonrandom, so unless we make an assumption of general randomness, there is no way to solve the problem.

2. Let's compare the scenarios in parts (a) and (b), to see exactly why the probabilities differ. In part (a), the parent's statement rules out the GG case. The BB, BG, and GB cases survive, with the BB families representing 1/3 of all of the possibilities. If the parent then changes the statement, "at least one of them is a boy" to "the older one is a boy," we are now in the realm of part (b). The GB case is now also ruled out (in addition to the GG case). So only the BB and BG cases survive, with the BB families representing 1/2 of all of the possibilities. This is why the probability jumps from 1/3 to 1/2 in going from part (a) to part (b). An additional group of families (GB) is ruled out.

   Let's now compare the scenarios in parts (a) and (c), to see exactly why the probabilities differ. As in the preceding paragraph, the parent's statement in part (a) rules out the GG case. If the parent then makes the additional statement "…and there he is over there next to that tree," we are now in the realm of part (c). Which additional families are ruled out? Well, in part (a), you could be talking with a parent in any of

the families in Table 2.5 except the two GG entries. So there are six valid possibilities. But as soon as the parent adds the "and there he is" comment, the unboxed **G**B and B**G** entries are ruled out. So a larger fraction of the valid possibilities (now two out of four, instead of two out of six) have the other child being a boy.

3. Having gone through all of the above reasonings and the comparisons of the different cases, we should note that there is actually a much quicker way of obtaining the probabilities of 1/2 in parts (b) and (c). If the parent says that the older child is a boy, or that one of the children is the boy standing next to her, then the parent is making a statement *solely about a particular child* (the older one, or the present one). The parent is saying nothing about the other child (the younger one, or the absent one). We therefore know nothing about that child. So by our assumption of general randomness, the other child is equally likely to be a boy or a girl. This should be contrasted with part (a). In that scenario, when the parent says that at least one child is a boy, the parent is *not* making a claim about a *specific child*, but rather about the *collective set* of the two children together. We are therefore not able to uniquely define the "other child" and simply say that the answer is 1/2. The answer depends on both children together, and it turns out to be different from 1/2 (namely 1/3).

4. There is a subtlety in this problem that we should address: How does the parent decide what information to give you? A reasonable rule could be that in part (a) the parent says, "At least one child is a boy," if she is able to; otherwise she says, "At least one child is a girl." This is consistent with all of our above reasoning. But consider what happens if we tweak the rule so that now the parent says, "At least one child is a girl," if she is able to; otherwise she says, "At least one child is a boy." In this case, the answer to part (a) is 1, because the only parents making the "boy" statement are the BB parents. This minor tweak completely changes the problem.

   If you want to avoid this issue, you can rephrase part (a) as: You bump into a random person on the street and ask, "Do you have (exactly) two children? If so, is at least one of them a boy?" In the cases where the answers to both of these questions are "yes," what is the probability that the other child is also a boy? Alternatively, you can just remove the parent and pose the problem as: Consider all two-child families that have at least one boy. What is the probability that both children are boys? This phrasing isn't as catchy as the original, but it gets rid of the above issue.

5. In the various lists of types of families in the above solutions, only the boxed types were applicable. The unboxed ones didn't satisfy the conditions given in the statement of the problem, so we discarded them. This act of discarding the unboxed types is equivalent to using the conditional-probability statement in Eq. (2.5), which can be rearranged to say

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)} \ . \tag{2.48}$$

For example, in part (a) if we let $A = \{$at least 1 boy$\}$ and $B = \{$2 boys$\}$, then we obtain

$$P\big((2 \text{ boys})|(\text{at least 1 boy})\big) = \frac{P\big((\text{at least 1 boy}) \text{ and } (2 \text{ boys})\big)}{P(\text{at least 1 boy})} \ . \tag{2.49}$$

The lefthand side of this equation is the probability we're trying to find. On the righthand side, we can rewrite $P((\text{at least 1 boy}) \text{ and } (2 \text{ boys}))$ as just $P(2 \text{ boys})$, because $\{2 \text{ boys}\}$ is a subset of $\{$at least 1 boy$\}$. So we have

$$P\big((2 \text{ boys})|(\text{at least 1 boy})\big) = \frac{P(2 \text{ boys})}{P(\text{at least 1 boy})} = \frac{1/4}{3/4} = \frac{1}{3} \ . \tag{2.50}$$

The preceding equations might look a bit intimidating, which is why we took a more intuitive route in the above solution to part (a), where we imagined doing 1000 trials and then discarding the 250 GG families. Discarding these families accomplishes the same thing as having the P(at least 1 boy) term in the denominator in Eq. (2.50); namely, they both signify that we are concerned only with families that have at least one boy. This remark leads us into the following section on Bayes' theorem.

6. If you thought that some of the answers to this problem were counterintuitive, then, well, you haven't seen anything yet! Tackle Problem 2.19 and you'll see why. ♣

2.5 Bayes' theorem

We now introduce Bayes' theorem, which gives a relation between certain conditional probabilities. The theorem is relevant to much of what we have been discussing in this chapter, particularly Section 2.4. We have technically already derived everything we need for the theorem (and we have actually already been using the theorem without realizing it), so the proof will be very quick. There are three common forms of the theorem. After we prove these, we'll do an example and then present a helpful way of thinking about the theorem in terms of pictures.

Theorem 2.1 (Bayes' theorem) *The "simple form" of Bayes' theorem is*

$$P(A|Z) = \frac{P(Z|A) \cdot P(A)}{P(Z)} \tag{2.51}$$

The "explicit form" is (with "~A" shorthand for "not A")

$$P(A|Z) = \frac{P(Z|A) \cdot P(A)}{P(Z|A) \cdot P(A) + P(Z|\sim A) \cdot P(\sim A)} \tag{2.52}$$

And the "general form" is

$$P(A_k|Z) = \frac{P(Z|A_k) \cdot P(A_k)}{\sum_i P(Z|A_i) \cdot P(A_i)} \tag{2.53}$$

where the A_i are a complete and mutually exclusive set of events. That is, every possible outcome belongs to one (hence the "complete") and only one (hence the "mutually exclusive") of the A_i.

Proof: The simple form of Bayes' theorem in Eq. (2.51) follows from what we noted back in Eq. (2.9). Since the order of A and Z doesn't matter in $P(A \text{ and } Z)$, we can write down two different expressions for this probability:

$$\begin{aligned} P(A \text{ and } Z) &= P(A|Z) \cdot P(Z) \\ &= P(Z|A) \cdot P(A). \end{aligned} \tag{2.54}$$

If we equate the two righthand sides of these equations and divide through by $P(Z)$, we obtain Eq. (2.51).

The explicit form in Eq. (2.52) follows from the fact that the $P(Z)$ in the denominator of Eq. (2.51) can be written as

$$P(Z) = P(Z \text{ and } A) + P(Z \text{ and } \sim A)$$
$$= P(Z|A) \cdot P(A) + P(Z|\sim A) \cdot P(\sim A). \qquad (2.55)$$

The first line here comes from the fact that every outcome is a member of either A or $\sim A$, and the second line comes from two applications of Eq. (2.54).

The general form in Eq. (2.53) is obtained by replacing the A in Eq. (2.51) with A_k and noting that

$$P(Z) = \sum_i P(Z \text{ and } A_i)$$
$$= \sum_i P(Z|A_i) \cdot P(A_i). \qquad (2.56)$$

The first line here comes from the fact that every outcome is a member of exactly one of the A_i, and the second line comes from n applications (where n is the number of A_i) of Eq. (2.54). Note that Eq. (2.52) is a special case of Eq. (2.53), with $A_1 = A$ and $A_2 = \sim A$, and with $k = 1$ (so $A_k = A$). Note also that all of the numerators on the righthand sides of the three formulations of the theorem are equal to $P(A \text{ and } Z)$ or $P(A_k \text{ and } Z)$, from Eq. (2.54).   ∎

As promised, these proofs were very quick. All we needed was Eq. (2.54) and the fact that $P(Z) = \sum_i P(Z \text{ and } A_i)$, which holds because the A_i are mutually exclusive and complete. However, even though the proofs were quick, and even though the theorem isn't anything we didn't already know (since we already knew the two ingredients in the preceding sentence), the theorem can still be a bit intimidating, especially the general form in Eq. (2.53). So we'll do an example to get some practice. But first some remarks.

REMARKS:

1. In Eq. (2.53) the $P(A_i)$ are known as the *prior* probabilities, the $P(Z|A_i)$ are known as the *conditional* probabilities, and $P(A_k|Z)$ is known as the *posterior* probability. The prior and conditional probabilities are the ones you are given (at least in this book; see the following remark), and the posterior probability is the one you are trying to find.

2. Since Bayes' theorem is simply a restatement of what we already know, you might be wondering what good it is and why it comes up so often when people talk about probability. Does it actually give us anything new? Well, yes and no. The theorem itself doesn't give us anything new, but the way in which it is *used* does.

   It would take many pages to do justice to this topic, but in a nutshell, there are two main types of probability reasoning. *Frequentist* reasoning (which is what we are using in this book) defines probability by imagining a large number of trials. In contrast, *Bayesian* reasoning doesn't require a large number of trials. The difference between these two reasonings shows up when one gets into statistical inference, that is, when one tries to estimate probabilities by gathering data (which we won't do in this book). In the end, the difference comes down to how one treats the prior probabilities $P(A_i)$

in Eq. (2.53). A frequentist considers them to be definite quantities (based on the frequencies obtained in large numbers of trials), whereas a Bayesian considers them to be unknowns whose values are given by specified distributions (determined in some manner). However, this difference is moot in this book, because we will always deal with situations where the prior probabilities take on definite values that are given. In this case, the frequentist and Bayesian reasonings are identical. They both boil down to Eq. (2.54). ♣

Let's now do an example. A common setup where Bayes' theorem is relevant involves false positives on a diagnostic test, so that's the setup we'll use here. After working through the example, we'll see how we can alternatively make use of a particularly helpful type of picture. There are many different probabilities that appear in Eq. (2.53), and it can be hard to remember what the theorem says or to get an intuitive feel for what's going on. In contrast, a quick glance at a figure such as Fig. 2.14 below makes it easy to remember the theorem and understand it intuitively.

Example (False positives): A hospital administers a test to see if a patient has a certain disease. Assume that we know the following three things:

- 2% of the overall population has the disease.

- If a person *does* have the disease, then the test has a 95% chance of correctly indicating that the person has it. (So 5% of the time, the test incorrectly indicates that the person doesn't have the disease.)

- If a person *does not* have the disease, then the test has a 10% chance of incorrectly indicating that the person has it; this is a "false positive" result. (So 90% of the time, the test correctly indicates that the person doesn't have the disease.)

The question we want to answer is: If a patient tests positive, what is the probability that they[2] actually have the disease?

We'll answer this question first by pretending that we haven't seen Bayes' theorem, and then by using the theorem. The reasoning will be exactly the same in both solutions, because in the first solution we'll actually be using Bayes' theorem without realizing it.

First solution: Imagine taking a large number of people (say, 1000) from the general population and testing them for the disease. A given person either has the disease or doesn't (two possibilities), and their test is either positive or negative (two possibilities). So there are $2 \cdot 2 = 4$ different types of people, with regard to the disease and the test. Let's make a probability tree to determine how many people of each type there are; see Fig. 2.11. The three given facts correspond to the three forks in the tree:

- The first fact tells us that of the given 1000 people, 2% (which is 20 people) have the disease (on average), while 98% (which is 980 people) don't have the disease.

- The second fact tells us that of the 20 people with the disease, 95% (which is 19 people) test positive, while 5% (which is 1 person) tests negative.

---

[2] I am using "they" as a gender-neutral singular pronoun, in protest of the present failing of the English language.

• The third fact tells us that of the 980 people without the disease, 10% (which is 98 people) test positive, while 90% (which is 882 people) test negative.
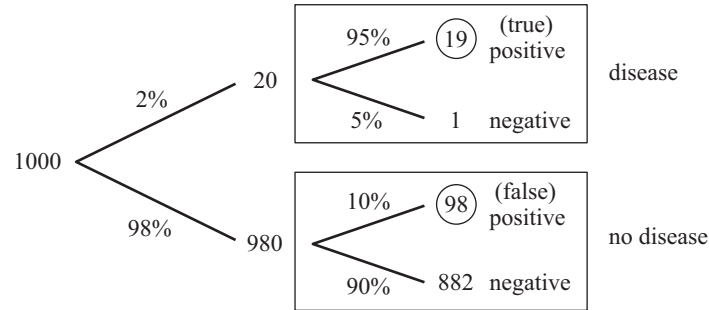
Figure 2.11: The probability tree for yes/no disease and positive/negative test.

The answer to the above question (namely, "If a patient tests positive, what is the probability that they actually have the disease?") can now simply be read off from the tree. The total number of people who test positive is the sum of the two circled numbers, which is $19 + 98 = 117$. And of these 117 people, only 19 have the disease. So our answer is

$$p = \frac{19}{19 + 98} = \frac{19}{117} = 16\%. \tag{2.57}$$

If we want to write this directly in terms of the given probabilities, then if we recall how we arrived at the numbers 19 and 98, we obtain

$$p = \frac{(0.95)(0.02)}{(0.95)(0.02) + (0.10)(0.98)} = 0.16. \tag{2.58}$$

Second solution: We'll use the "explicit form" of Bayes' theorem in Eq. (2.52), which is a special case of the "general form" in Eq. (2.53). In the notation of Eq. (2.52) we have

$$A = \text{have disease},$$
$$\sim A = \text{don't have disease},$$
$$Z = \text{test positive}. \tag{2.59}$$

Our goal is to calculate $P(A|Z)$, that is, the probability of having the disease, given a positive test. From the given facts in the three bullet points, we know that

$$P(A) = 0.02,$$
$$P(Z|A) = 0.95,$$
$$P(Z|\sim A) = 0.10. \tag{2.60}$$

Plugging these probabilities into Eq. (2.52) gives

$$P(A|Z) = \frac{P(Z|A) \cdot P(A)}{P(Z|A) \cdot P(A) + P(Z|\sim A) \cdot P(\sim A)}$$
$$= \frac{(0.95)(0.02)}{(0.95)(0.02) + (0.10)(0.98)} = 0.16, \tag{2.61}$$

in agreement with the first solution. This is the same expression as in Eq. (2.58), which is consistent with the fact that (as we mentioned above) our reasoning in the first solution was equivalent to using Bayes' theorem.

REMARK: We see that if a person tests positive, they have only a 16% chance of actually having the disease. This answer might seem surprisingly low. After all, the test seems fairly reliable; it gives the correct result 95% of the time if a person has the disease, and 90% of the time if a person doesn't have the disease. So how did we end up with an answer that is much smaller than either of these two percentages?

The explanation is that because the percentage of people with the disease is so tiny (2%), the small percentage (10%) of false positives among the non-disease people yields a *number* of false positives that is significantly larger than the *number* of true positives. Basically, 10% of 98% of 1000 (which is 98) is significantly larger than 95% of 2% of 1000 (which is 19). The 98 false positives dominate the 19 true positives. Although the 10% false-positive rate is small, it isn't small enough to prevent the smallness of the 2% disease rate from controlling the outcome. A takeaway from this discussion is that one must be very careful when testing for rare diseases. If the disease is very rare, then the test must be extremely accurate, otherwise a positive test isn't meaningful.

If we decrease the 10% percentage (that is, reduce the percentage of false positives) and/or increase the 2% percentage (that is, increase the percentage of people with the disease), then the answer to our original question will increase. That is, a larger fraction of the people who test positive will actually have the disease. For example, if we assume that 40% of the population have the disease (so 60% don't have it), and if we keep all the other percentages in the problem the same, then Eq. (2.58) becomes

$$p = \frac{(0.95)(0.40)}{(0.95)(0.40) + (0.10)(0.60)} = 0.86. \tag{2.62}$$

This probability is closer to 1 than in the original scenario, because if we have 1000 people, then the 60 (instead of the earlier 98) false positives are dominated by the 380 (instead of the earlier 19) true positives. You can verify these numbers.

In the limit where the 10% false-positive percentage in the original scenario goes to zero, or the 2% disease percentage goes to 100%, the number of false positives goes to zero. This is true because if $10\% \to 0\%$ then the test never incorrectly says that a person has the disease when they don't; and if $2\% \to 100\%$ then the entire population has the disease, so every positive test is a true one. In either of these limits, the answer to our question goes to 1 (or 100%); a positive test always correctly indicates the disease. ♣

In the first solution above, we calculated the various numbers and probabilities by using a probability tree. We can alternatively use a figure along the lines of Fig. 2.4. In the following discussion we'll pretend that we haven't seen Bayes' theorem, and then we'll circle back to the theorem and show in Fig. 2.14 how the different ingredients in the theorem correspond to the different parts of the figure.

Fig. 2.12 shows a pictorial representation of the probability tree in Fig. 2.11. The overall square represents the given 1000 people.[3] A vertical line divides the

---

[3]When drawing a figure like this, the area of a region can represent either the probability of being in that region, or the actual number of outcomes/people/etc. in that region. The usage should be clear from the context. We're using actual numbers here.

square into two rectangles – a very thin one on the left representing the 20 people with the disease, and a wide one on the right representing the 980 people without the disease. These two rectangles are further divided into the people who test positive (the shaded lower regions, with 19 and 98 people) or test negative (the unshaded upper regions, with 1 and 882 people). The desired probability of a person having the disease if they test positive equals the 19 true positives (the darkly shaded thin rectangle) divided by the total $19 + 98 = 117$ number of positives (both shaded regions).
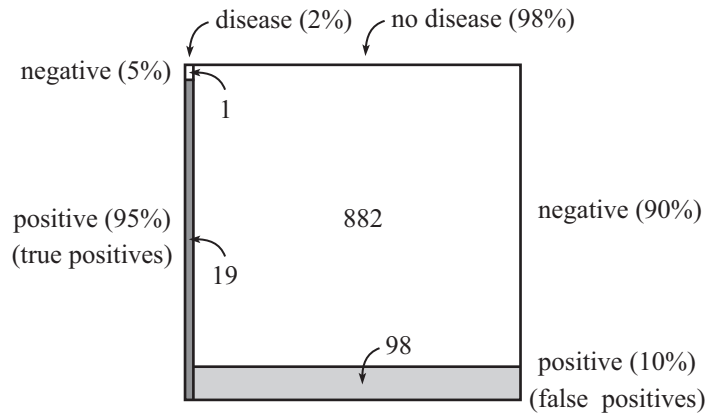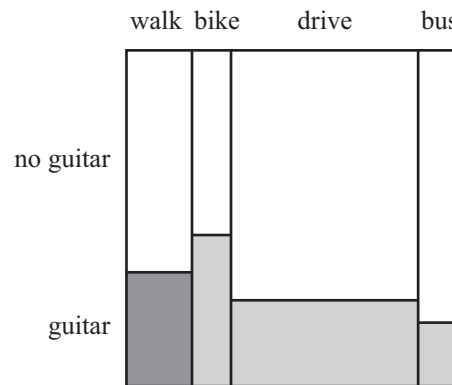

Figure 2.12: The probability square for yes/no disease and positive/negative test.

In Fig. 2.12 there are only two types of people in the population – those with the disease and those without it. As an example of a more general setup, let's consider how people commute to work. We'll assume that we are given the percentages of people who walk, bike, drive, take the bus, etc. And then for each of these types, we'll assume that we are also given the percentage who have a particular attribute – for example, the ability to play the guitar. We can then ask questions such as, "If we pick a random person (among those who commute to work) from the set of people who can play the guitar, what is the probability that this person walks to work?" If we compare this question to our earlier one involving the disease testing, we see that guitar playing is analogous to testing positive, and walking to work is analogous to having the disease. It's just that now we have many types of commuters instead of only two types of disease carriers (carriers or non carriers).

To answer the above question, we can draw a figure analogous to Fig. 2.12; see Fig. 2.13 with some made-up percentages for the various types of commuters. These percentages are undoubtedly completely unrealistic, but they're good enough for the sake of an example.

For simplicity, we'll assume that there are only four possible ways to commute to work. If the guitar players are represented by the shaded regions, then the answer to our question is obtained by dividing the area of the darkly shaded region (which represents the guitar players who are walkers) by the total area of all the shaded regions (which represents all of the guitar players). Mathematically, the preceding sentence is equivalent to dividing the first equality in Eq. (2.54) through by $P(Z)$

walk bike　　drive　　bus

Figure 2.13: The probability square for a hypothetical commuting example.

and then letting A = "walk" and Z = "guitar":

$$P(\text{walk}|\text{guitar}) = \frac{P(\text{walk and guitar})}{P(\text{guitar})}$$

$$= \frac{\text{dark shaded area}}{\text{total shaded area}} \, . \qquad (2.63)$$

Assuming that there are only four possible ways to commute to work, we need to be given eight pieces of information:

- We need to be given the four percentages of people who walk, bike, drive, or take the bus. (Actually, since these percentages must add up to 100%, there are only three independent bits of information here.) These percentages determine the relative widths of the vertical rectangles in Fig. 2.13. The analogous information in the "False positives" example was contained in the first bullet point on page 99 (the percentage of people who have the disease).

- For each of the four types of commuters, we need to be given the percentage who play the guitar. These four percentages determine the heights of the shaded areas within the vertical rectangles in Fig. 2.13. The analogous information in the "False positives" example was contained in the second and third bullet points on page 99.

Of course, if we are simply given the area of the darkly shaded region (which represents the number of guitar players who are walkers), and also the total area of all the shaded regions (which represents the total number of guitar players), then we can just divide the first of these two pieces of information by the second, and we're done. But in most situations, we're given the above eight (or whatever the relevant number is) pieces of information instead of these two, and the main task is to determine these two.

If you want to instead think in terms of a probability tree, as in Fig. 2.11, then in the present commuting example, the initial fork has four branches (for the walk/bike/drive/bus options), and then each of these four options splits into two possibilities (guitar or no guitar). We therefore end up with four circled numbers (the

guitar players) instead of the two in Fig. 2.11, and we need to divide one of these (the one in the walking branch) by the sum of all four.

The interpretation of Bayes' theorem in terms of a figure like Fig. 2.13 is summarized in Fig. 2.14. In this figure, we are considering areas to represent probabilities instead of actual numbers (although either way is fine), because heights and widths then represent the relevant probabilities. It is invariably much more intuitive to think of the theorem in terms of a figure instead of algebraic manipulations, so when you think of Bayes' theorem, you'll probably want to think of Fig. 2.14.
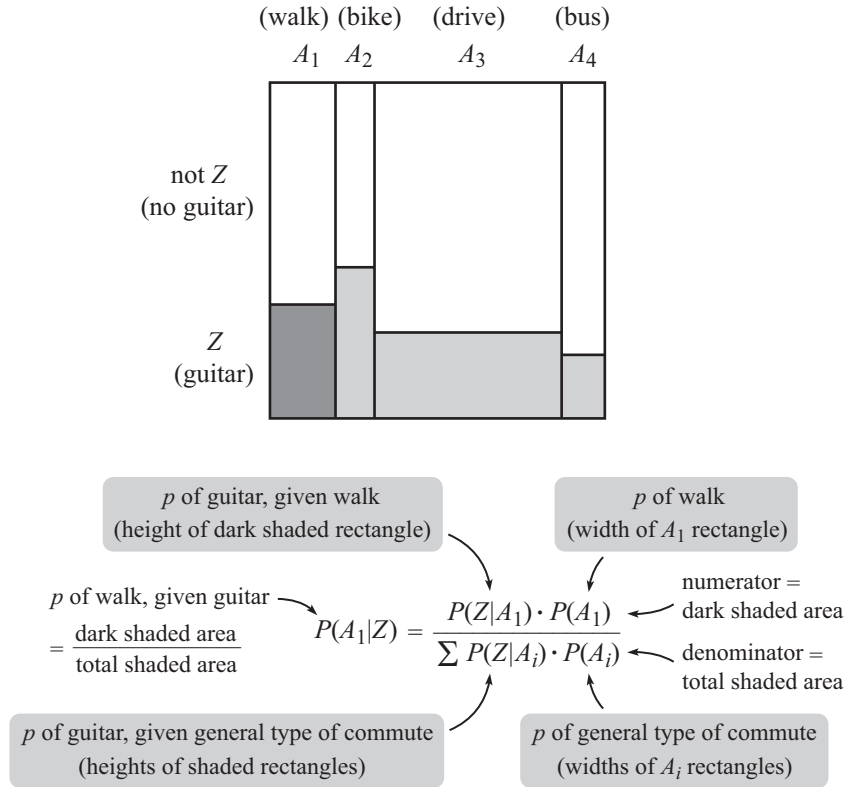
Figure 2.14: Pictorial representation of Bayes' theorem.

SMALLCAPS REMARKS:

1. It is often the case that you aren't given $P(Z)$ in the simple form of Bayes' theorem in Eq. (2.51), but instead need to calculate it via $\sum P(A_i) \cdot P(Z|A_i)$ or $P(Z|A) \cdot P(A) + P(Z|\sim A) \cdot P(\sim A)$, as we did in the "False positives" example. So the general form of Bayes' theorem in Eq. (2.53) or the explicit form in Eq. (2.52) is often the relevant one.

2. When using Bayes' theorem to calculate $P(A_1|Z)$, remember that in the notation of Fig. 2.14, the *first* letter A_1 in $P(A_1|Z)$ is one of the many A_i that divide up the *horizontal* span of the square, while the *second* letter Z is associated with the *vertical* span of the shaded areas.

3. In setups involving Bayes' theorem, there can be an arbitrary number n of the A_i columns in Fig. 2.14. (We've drawn the case with $n = 4$.) But each column is divided into only *two* regions, namely the Z region and the not-Z region. Of course, the not-Z region might very well be broken down into other regions, but that isn't relevant here.

   If you wish, you can think of there being only *two* columns, namely the A_1 column and the "not-A_1" column, which consists of all the other A_i. However, if you are given information for each of the A_i, then you will need to consider them separately. But after calculating all the relevant numbers, it is certainly fine to lump all the other A_i together into a single "not-A_1" column. Fig. 2.14 then becomes Fig. 2.15. The lightly shaded area here is the same as the total lightly shaded area in Fig. 2.14. Fig. 2.15 corresponds to the explicit form of Bayes' theorem in Eq. (2.52), while Fig. 2.14 corresponds to the general form in Eq. (2.53).
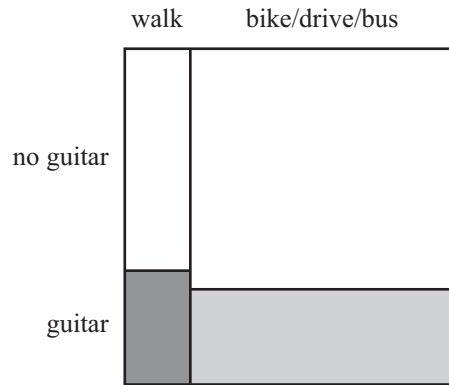
Figure 2.15: Grouping all of the nonwalkers together.

4. The essence of Bayes' theorem comes down to the fact that $P(A \text{ and } Z)$ can be written in the two different ways given in Eq. (2.54). In terms of Fig. 2.14, you can think of $P(A_1 \text{ and } Z)$, which is the area of the darkly shaded rectangle, in two different ways. It is a certain fraction (namely $P(A_1|Z)$) of the overall shaded area (namely $P(Z)$); this leads to the first equality in Eq. (2.54). And $P(A_1 \text{ and } Z)$ is also a certain fraction (namely $P(Z|A_1)$) of the leftmost (walking) rectangle area (namely $P(A_1)$); this leads to the second equality in Eq. (2.54).

   Said in another way, the *number* of guitar players who are walkers equals the *number* of walkers who are guitar players. This common number equals the area of the darkly shaded rectangle (which is the probability $P(A_1 \text{ and } Z)$) multiplied by the total number of people. Note that the first sentence above is *not* true (in general) if the word "number" is replaced by "fraction." That is, it is *not* true that the *fraction* of guitar players who are walkers equals the *fraction* of walkers who are guitar players. Equivalently, it is *not* true that $P(A_1|Z) = P(Z|A_1)$. Instead, these two conditional probabilities are related according to Eq. (2.51).

5. In Section 2.4 we solved the game-show problem, the prosecutor's fallacy, and the boy/girl problem without using Bayes' theorem. However, if we *had* used the theorem, the reasoning would have been basically the same, just as the reasoning that led to Eq. (2.58) in the "False positives" example was basically the same as the reasoning that led to Eq. (2.61). We chose to discuss the problems in Section 2.4 before discussing Bayes' theorem, so that it would be clear that the problems are still perfectly solvable

even if you've never heard of the theorem. If you want to solve the prosecutor's fallacy by explicitly using Bayes' theorem, see Problem 2.21. ♣

2.6   Stirling's formula

Stirling's formula gives an approximation to $n!$ that is valid for large n, in the sense that the larger n is, the better the approximation is. By "better," we mean that as n gets large, the approximation gets closer and closer to $n!$ in a *multiplicative* sense (as opposed to an additive sense). That is, the *ratio* of the approximation and $n!$ approaches 1. (The additive *difference* between the approximation and $n!$ gets larger and larger as n grows, but we don't care about that.) Stirling's formula is given by:

$$\boxed{n! \approx n^n e^{-n} \sqrt{2\pi n}} \qquad \text{(Stirling's formula)} \qquad (2.64)$$

Here e is the base of the natural logarithm, equal to $e \approx 2.71828$. See Appendix B for a discussion of e, often referred to as *Euler's number*. There are various proofs of Stirling's formula, but they generally involve calculus, so we'll just accept the formula here. It does indeed give an accurate approximation to $n!$ (an extremely accurate one, if n is large), as you can see from Table 2.6, where $S(n)$ stands for the $n^n e^{-n} \sqrt{2\pi n}$ Stirling approximation. Even if n is just 10, the approximation is off by only about 0.8%. And although there is never any need to use the formula for small numbers like 1 or 5, it works surprisingly well in those cases too.

n	$n!$	$S(n)$	$S(n)/n!$
1	1	0.922	0.922
5	120	118.0	0.983
10	$3.629 \cdot 10^6$	$3.599 \cdot 10^6$	0.992
100	$9.3326 \cdot 10^{157}$	$9.3249 \cdot 10^{157}$	0.9992
1000	$4.02387 \cdot 10^{2567}$	$4.02354 \cdot 10^{2567}$	0.99992

Table 2.6: Showing the accuracy of Stirling's formula.

You will note that for the powers of 10 in the table, the ratios of $S(n)$ to $n!$ all take the same form, namely decimals with an increasing number of 9's and then a 2. It's actually not a 2, because we rounded off, but it's essentially the same rounding off for all the numbers. This isn't a coincidence. It follows from a more accurate version of Stirling's formula, but we won't get into that here.

Stirling's formula will be critical in Chapter 5 when we talk about approximations to certain probability distributions. But for now, it is relevant when dealing with binomial coefficients of large numbers, because these binomial coefficients involve the factorials of large numbers. There are two main benefits to using Stirling's formula:

- Depending on the type of calculator you have, you might get an error message when you plug in the factorial of a number that is too big. Stirling's

formula allows you to avoid this problem if you first simplify the expression that results from Stirling's formula (using the letter n to stand for the specific number you're dealing with), and *then* plug the simplified result into your calculator.

- If you use Stirling's formula and arrive at a simplified answer in terms of n (we'll call this a *symbolic* answer since it's written in terms of the symbol n instead of specific numbers), you can then plug in your specific value of n. Or you can plug in any other value, for that matter. The benefit of having a symbolic answer in terms of n is that you don't need to solve the problem from scratch every time you're given a new value of n. You simply need to plug the new value of n into your symbolic answer.

These two benefits are illustrated in the following example.

Example (50 out of 100): A coin is flipped 100 times. Calculate the probability of obtaining exactly 50 Heads.

Solution: In 100 flips, there are 2^{100} possible outcomes (all equally likely), of which $\binom{100}{50}$ have exactly 50 Heads. The probability of obtaining exactly 50 Heads is therefore

$$P(50) = \frac{1}{2^{100}} \binom{100}{50} = \frac{1}{2^{100}} \cdot \frac{100!}{50! \, 50!} \, . \tag{2.65}$$

Now, although this is the correct answer, your calculator might not be able to handle the large factorials. But even if it can, let's use Stirling's formula so that we can produce a symbolic answer. To this end, we'll replace the number 50 with the letter n (and hence 100 with $2n$). In terms of n, we can write down the probability of obtaining exactly n Heads in $2n$ flips, and then we can use Stirling's formula (applied to both n and $2n$) to simplify the result. The first steps of this simplification will actually go in the wrong direction and create a big mess, but nearly everything will cancel out in the end. We obtain:

$$P(n) = \frac{1}{2^{2n}} \binom{2n}{n} = \frac{1}{2^{2n}} \cdot \frac{(2n)!}{n! \, n!} \approx \frac{1}{2^{2n}} \cdot \frac{(2n)^{2n} e^{-2n} \sqrt{2\pi(2n)}}{(n^n e^{-n} \sqrt{2\pi n})^2}$$

$$= \frac{1}{2^{2n}} \cdot \frac{2^{2n} n^{2n} e^{-2n} \cdot 2\sqrt{\pi n}}{n^{2n} e^{-2n} \cdot 2\pi n}$$

$$= \frac{1}{\sqrt{\pi n}} \, . \tag{2.66}$$

A simple answer indeed! And the "π" is a nice touch, too. In our specific case with $n = 50$, we have

$$P(50) \approx \frac{1}{\sqrt{\pi \cdot 50}} \approx 0.07979 \approx 8\%. \tag{2.67}$$

This is small, but not negligible. If we instead have $n = 500$, we obtain $P(500) \approx 2.5\%$. This is the probability of obtaining exactly 500 Heads in 1000 coin flips. As noted above, we can just plug in whatever number we want, and not have to redo the entire calculation!

The $1/\sqrt{\pi n}$ result in Eq. (2.66) is extremely clean. It is much simpler than the expression in Eq. (2.65), and *much* simpler than the expressions in the first two lines of Eq. (2.66). True, it's only an approximate result, but it's a good one. The exact result in Eq. (2.65) happens to be about 0.07959, so for $n = 50$ the ratio of the approximate result in Eq. (2.67) to the exact result is 1.0025. In other words, the approximation is off by only 0.25%. That's plenty good for most purposes.

When you derive a symbolic approximation like Eq. (2.66), you gain something and you lose something. You lose some truth, of course, because your answer technically isn't correct (although invariably its accuracy is quite sufficient). But you gain a great deal of information about how the answer depends on your input number, n. And along the same lines, you gain some aesthetics. The resulting symbolic answer is invariably nice and concise, so it allows you to easily see how the answer depends on n. For example, in our coin-flipping example, the expression in Eq. (2.66) is proportional to $1/\sqrt{n}$. This means that if we increase n by a factor of, say, 100, then $P(n)$ decreases by a factor of $\sqrt{100} = 10$. So without doing any work, we can quickly use the $P(50) \approx 8\%$ result to deduce that $P(5000) \approx 0.8\%$. In short, there is *far* more information contained in the symbolic result in Eq. (2.66) than in the numerical 8% result obtained directly from Eq. (2.65).

2.7   Summary

In this chapter we learned about probability. In particular, we learned:

- The probability of an event is defined to be the fraction of the time the event occurs in a very large number of identical trials. In many situations the possible outcomes are all equally likely, in which case the probability of a certain class of outcomes occurring is

$$p = \frac{\text{number of desired outcomes}}{\text{total number of possible outcomes}} \quad \text{(for equally likely outcomes)}$$

$$(2.68)$$

- The various "and" and "or" rules of probability are:

  1. For any two (possibly dependent) events,

     $$P(A \text{ and } B) = P(A) \cdot P(B|A). \qquad (2.69)$$

  2. In the special case of independent events, we have $P(B|A) = P(B)$, so Eq. (2.69) reduces to

     $$P(A \text{ and } B) = P(A) \cdot P(B). \qquad (2.70)$$

  3. For any two (possibly nonexclusive) events,

     $$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B). \qquad (2.71)$$

  4. In the special case of exclusive events, we have $P(A \text{ and } B) = 0$, so Eq. (2.71) reduces to

     $$P(A \text{ or } B) = P(A) + P(B). \qquad (2.72)$$

- *A* and *B* are independent events if any one of the following relations is true:

$$P(B|A) = P(B),$$
$$P(A|B) = P(A),$$
$$P(A \text{ and } B) = P(A) \cdot P(B). \tag{2.73}$$

- The conditional probabilities $P(A|B)$ and $P(B|A)$ are not equal, in general.

- Two common ways to calculate probabilities are: (1) count up the number of desired outcomes, along with the total number of possible outcomes, and use Eq. (2.68) (assuming that the outcomes are equally likely), and (2) imagine things happening in succession (for example, picking seats or rolling dice), and then multiply the relevant probabilities. The results for some problems, in particular the Birthday Problem and the Game-Show Problem, might seem surprising at first, but you can avoid confusion by methodically using one (or both) of these strategies.

- Bayes' theorem takes a variety of forms; see Eqs. (2.51)–(2.53). The last of these is the "general form" of the theorem:

$$P(A_k|Z) = \frac{P(Z|A_k) \cdot P(A_k)}{\sum_i P(Z|A_i) \cdot P(A_i)}. \tag{2.74}$$

The theorem tells us how the conditional probability $P(A_k|Z)$ is obtained from the set of conditional probabilities $P(Z|A_i)$.

- Stirling's formula, which gives an approximation to $n!$, takes the form,

$$n! \approx n^n e^{-n} \sqrt{2\pi n} \qquad \text{(Stirling's formula)} \tag{2.75}$$

This approximation is very helpful for simplifying binomial coefficients. We will use it a great deal in Chapter 5.

2.8 Exercises

See **www.people.fas.harvard.edu/~djmorin/book.html** for a supply of problems without included solutions.

2.9 Problems

Section 2.1: Definition of probability

2.1. **Odds** ∗

If an event occurs with probability p, then the *odds* in favor of the event occurring are defined to be "p to $(1 - p)$." (And similarly, the odds *against* the event occurring are defined to be "$(1 - p)$ to p.") In other words, the odds are simply the ratio of the probabilities of the event occurring (namely p) and

not occurring (namely $1-p$). It is customary to write "$p:(1-p)$" as shorthand for "p to $(1-p)$." (The odds are sometimes also written as the ratio $p/(1-p)$. But this fraction can look like a probability, which may cause confusion, so we'll avoid this notation.) In practice, the probabilities p and $1-p$ are usually multiplied through by the smallest number that turns them into integers. For example, odds of $1/3:2/3$ are generally written as $1:2$. Find the odds of the following events:

(a) Getting a Heads on a coin toss.

(b) Rolling a 5 on a die.

(c) Rolling a multiple of 2 or 3 on a die.

(d) Randomly picking a day of the week with more than six letters.

Section 2.2: The rules of probability

2.2. **Rules for three events** $**$

(a) Consider three events, A, B, and C. If they are all independent of each other, show that

$$P(A \text{ and } B \text{ and } C) = P(A) \cdot P(B) \cdot P(C). \qquad (2.76)$$

(b) If they are (possibly) dependent, show that

$$P(A \text{ and } B \text{ and } C) = P(A) \cdot P(B|A) \cdot P(C|A \text{ and } B). \qquad (2.77)$$

(c) If they are all mutually exclusive, show that

$$P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C). \qquad (2.78)$$

(d) If they are (possibly) nonexclusive, show that

$$
\begin{aligned}
P(A \text{ or } B \text{ or } C) = {} & P(A) + P(B) + P(C) \\
& - P(A \text{ and } B) - P(A \text{ and } C) - P(B \text{ and } C) \\
& + P(A \text{ and } B \text{ and } C). \qquad (2.79)
\end{aligned}
$$

2.3. **"Or" rule for four events** $***$

Parts (a), (b), and (c) of Problem 2.2 generalize quickly to more than three events, but part (d) is trickier. Derive the "or" rule for four (possibly) nonexclusive events. That is, derive the rule analogous to Eq. (2.79).

2.4. **Red and blue balls** $*$

Show that the second expression in Eq. (2.9), with $A = \text{Red}_1$ and $B = \text{Blue}_2$, gives the correct result of $3/10$ for $P(\text{Red}_1 \text{ and } \text{Blue}_2)$ in the "balls in a box" example on page 64.

2.5. **Dependent events** ∗

Calculate the overall probability of *B* occurring in the scenario described by Fig. 2.16.
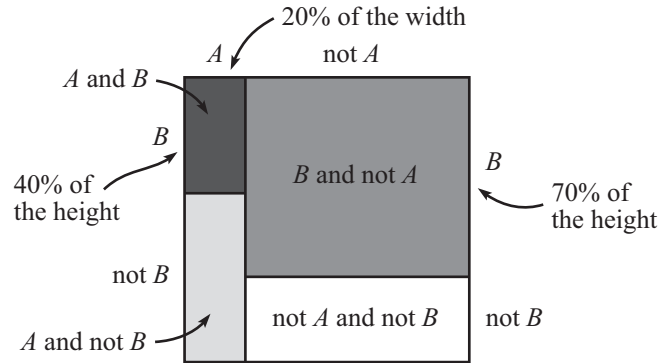
Figure 2.16: A hypothetical probability square.

2.6. **A single horizontal line** ∗

There is an asymmetry in Fig. 2.16. Because there is a single vertical line but two horizontal lines, it is easy to read off the *P*(*A*) and *P*(not *A*) probabilities, but not easy to read off the *P*(*B*) and *P*(not *B*) probabilities. Hence the calculation in Problem 2.5. Redraw Fig. 2.16 with a single horizontal line and two vertical lines (while keeping the areas (probabilities) of the four sub-rectangles the same, of course).

2.7. **Proofreading** ∗∗

Two people each proofread the same book. One person finds 100 errors, and the other finds 60. There are 20 errors common to both people. Assume that all errors are equally likely to be found (which is undoubtedly not true in practice), and also that the discovery of an error by one person is independent of the discovery of that error by the other person. Given these assumptions, roughly how many errors does the book have? *Hint:* Draw a picture similar to Fig. 2.1, and then find the probability of each person finding a given error.

Section 2.3: Examples

2.8. **Red balls, blue balls** ∗∗

Three boxes sit on a table. One box contains two red balls, another contains two blue balls, and the third contains one red ball and one blue ball. You choose one of the boxes at random, and then you draw a ball from that box. If it turns out to be a red ball, what is the probability that the other ball in the box is also red?

2.9. **Sock pairs** ∗∗

    (a) Four red socks and four blue socks are in a drawer. You reach in and pull out two socks at random. What is the probability that you obtain a matching pair?

    (b) Answer the same question, but now in the general case with n red socks and n blue socks.

    (c) Presumably you answered the above questions by counting the relevant pairs of socks. Can you think of a quick probability argument, requiring no counting, that gives the answer to part (b) (and part (a))?

2.10. **Sock pairs, again** ∗∗

    (a) As in Problem 2.9, four red socks and four blue socks are in a drawer. You reach in and pull out two socks at random. You then reach in and pull out two more socks (without looking at the socks in the first pair). What is the probability that the second pair you pull out is a matching pair? Answer this by calculating the probabilities, given that the first pair is (or is not) a matching pair.

    (b) You should find that the answer to part (a) is the same as the answer to part (a) of Problem 2.9. Can you think of a quick probability argument, requiring no counting, that explains why this is the case? The reasoning will work in the general case with n red socks and n blue socks. And it will also work if you draw a third pair, or a fourth pair, etc. (without looking at any of the other pairs).

2.11. **At least one 6** ∗∗

Three dice are rolled. What is the probability of obtaining at least one 6? We solved this in Section 2.3.1, but your task here is to solve it the long way, by adding up the probabilities of obtaining exactly one, two, or three 6's.

2.12. **At least one 6, by the rules** ∗∗

Three dice are rolled. What is the probability of obtaining at least one 6? We solved this in Section 2.3.1, and again in Problem 2.11. But your task here is to solve it by using Eq. (2.79) from Problem 2.2, with each of the three letters in that formula standing for a 6 on each of the three dice.

2.13. **Rolling sixes** ∗∗

This problem was posed by Samuel Pepys to Isaac Newton in 1693 and is therefore known as the Newton-Pepys problem.

    (a) 6 dice are rolled. What is the probability of obtaining at least one 6?

    (b) 12 dice are rolled. What is the probability of obtaining at least two 6's?

    (c) 18 dice are rolled. What is the probability of obtaining at least three 6's? Which of the above three probabilities is the largest?

Section 2.4: Four classic problems

2.14. **Exactly one pair** ∗∗

If there are 23 people in a room, what is the probability that *exactly* two of them have a common birthday? That is, we don't want two different pairs with common birthdays, or three people with a common birthday, etc.

2.15. **My birthday** ∗∗

(a) You are in a room with 100 other people. Let p be the probability that at least one of these 100 people has your birthday. Without doing any calculations, state whether p is larger, smaller, or equal to, 100/365.

(b) Now calculate the exact value of p.

2.16. **My birthday, again** ∗∗

We saw at the end of Section 2.4.1 that 253 is the answer to the question, "How many people (in addition to me) need to be present in order for there to be at least a 1/2 chance that someone else has *my* birthday?" We solved this by finding the smallest n for which $(364/365)^n$ is less than 1/2. Answer this question again, by making use of the approximation in Eq. (7.14) in Appendix C. What is the answer in the general case where there are N days in a year instead of 365? Assume that N is large.

2.17. **My birthday, yet again** ∗∗

With 253 other people in a room, what is the probability that *exactly* one of these people has your birthday? Exactly two? Exactly three?

2.18. **A random game-show host** ∗∗

Consider the following variation of the Game-Show Problem we discussed in Section 2.4.2. A game-show host offers you the choice of three doors. Behind one of these doors is the grand prize, and behind the other two are goats. The host announces that after you select a door (without opening it), he will *randomly* open one of the other two doors. You select a door. The host then randomly opens one of the other doors, and the result happens to be a goat. He then offers you the chance to switch your choice to the remaining door. Should you switch or not? Or does it not matter?

2.19. **Boy/girl problem with general information** ∗∗∗

This problem is an extension of the Boy/Girl Problem from Section 2.4.4. You should study that problem thoroughly before tackling this one. As in the original versions of the problem, assume that all processes are completely random. The new variation is the following:

You bump into a random person on the street who says, "I have two children. At least one of them is a boy whose birthday is in the summer." What is the probability that the other child is also a boy?

What if the clause is changed to, "whose birthday is on August 11th"? Or "who was born during a particular minute on August 11th"? Or more generally, "who has a particular characteristic that occurs with probability *p*"? *Hint*: Make a table of all of the various possibilities, analogous to the tables in Section 2.4.4.

Section 2.5: Bayes' theorem

2.20. **A second test** ∗∗

Consider the setup in the "False positives" example in Section 2.5. If we instead perform *two* successive tests on each person, what is the probability that a person who tests positive both times actually has the disease?

2.21. **Bayes' theorem for the prosecutor's fallacy** ∗∗

In Section 2.4.3 we discussed the prosecutor's fallacy. Explain the fallacy again here, but now by using Bayes' theorem. In particular, determine $P(I|D)$ (the probability of being innocent, given that the description is satisfied) by drawing a figure analogous to Fig. 2.14

2.22. **Black balls and white balls** ∗∗

One box contains two black balls, and another box contains one black ball and one white ball. You pick one of the boxes at random and draw a ball *n* times, with replacement after each draw. If a black ball is drawn all *n* times, what is the probability that you picked the box with two black balls?

2.10 Solutions

2.1. **Odds**

(a) The probability of getting a Heads is 1/2, as is the probability of not getting a Heads. So the desired odds are $1/2 : 1/2$, or equivalently $1 : 1$. These are known as "even odds."

(b) The probability of rolling a 5 is 1/6, and the probability of not rolling a 5 is 5/6. So the desired odds are $1/6 : 5/6$, or equivalently $1 : 5$.

(c) There are four desired outcomes $(2, 3, 4, 6)$, so the "for" and "against" probabilities are 4/6 and 2/6, respectively. The desired odds are therefore $4/6 : 2/6$, or equivalently $2 : 1$.

(d) Tuesday, Wednesday, Thursday, and Saturday all have more than six letters, so the "for" and "against" probabilities are 4/7 and 3/7, respectively. The desired odds are therefore $4/7 : 3/7$, or equivalently $4 : 3$.

Note that to convert from odds to probability, the odds of $a : b$ in favor of an event occurring are equivalent to a probability of $a/(a + b)$ that the event occurs.

2.2. **Rules for three events**

(a) We can use the same type of reasoning that we used in Section 2.2.1. If we perform a large number of trials, then *A* occurs in a fraction $P(A)$ of them. (It is understood here that the words "on average" follow all statements of this form.)

And then *B* occurs in a fraction $P(B)$ *of these* trials, because the events are independent, which means that the occurrence of *A* doesn't affect the probability of *B*. So the fraction of the total number of trials where *A* and *B* both occur is $P(A) \cdot P(B)$. And then *C* occurs in a fraction $P(C)$ *of these* trials, because *C* is independent of *A* and *B*. So the fraction of the total number of trials where all three of *A*, *B*, and *C* occur is $P(A) \cdot P(B) \cdot P(C)$. The desired probability is therefore $P(A) \cdot P(B) \cdot P(C)$. If you want to visualize this geometrically, you'll need to use a cube instead of the square in Fig. 2.1.

This reasoning can easily be extended to an arbitrary number of independent events. The probability of all of the events occurring is simply the product of all of the individual probabilities.

(b) The reasoning in part (a) works again, with only slight modifications. If we perform a large number of trials, then *A* occurs in a fraction $P(A)$ of them. And then *B* occurs in a fraction $P(B|A)$ *of these* trials, by definition. So the fraction of the total number of trials where *A* and *B* both occur is $P(A) \cdot P(B|A)$. And then *C* occurs in a fraction $P(C|A \text{ and } B)$ *of these* trials, by definition. So the fraction of the total number of trials where all three of *A*, *B*, and *C* occur is $P(A) \cdot P(B|A) \cdot P(C|A \text{ and } B)$. The desired probability is therefore $P(A) \cdot P(B|A) \cdot P(C|A \text{ and } B)$.

Again, this reasoning can easily be extended to an arbitrary number of (possibly) dependent events. For four events, we just need to tack on the factor $P(D|A \text{ and } B \text{ and } C)$, and so on.

(c) Since the events are all mutually exclusive, we don't have to worry about any double counting. The total number of trials where *A* or *B* or *C* occurs is simply the sum of the number of trials where *A* occurs, plus the number where *B* occurs, plus the number where *C* occurs. The same statement must be true if we substitute the word "fraction" for "number," because the fractions are related to the numbers via division by the total number of trials. And since the fractions are the probabilities, we end up with the desired result, $P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C)$. If there are more events, we simply have more terms in the sum.

(d) This rule is more involved than the preceding three. Let's think of the probabilities in terms of areas, as we did in Section 2.2.2. The generic situation for three events is shown in Fig. 2.17. For simplicity, we've chosen the three regions to be circles with the same size, but this of course isn't necessary. The various overlap regions are shown, with the juxtaposition of two letters standing for their intersection. So *AB* means "*A* and *B*." The labels might appear to suggest otherwise, but remember that *A* includes the whole circle, and not just the white part. Similarly, *AB* includes the dark *ABC* region too, and not just the lighter region where the *AB* label is.

Our goal is to determine the total area contained in the three circles, because this represents the probability of "*A* or *B* or *C*." We can add up the areas of the *A*, *B*, and *C* circles, but then we need to subtract off the areas that we double counted. These areas are the pairwise overlaps of the circles, that is, *AB*, *AC*, and *BC* (remember that each of these regions includes the dark *ABC* region in the middle). At this point, we've correctly counted all of the white and light gray regions exactly once. But what about the *ABC* region in the middle? We counted it three times in the *A*, *B*, and *C* regions, but then we subtracted it off three times in the *AB*, *AC*, and *BC* regions. So at the moment, we haven't counted it at all. We therefore need to add it on once. Then every part of the
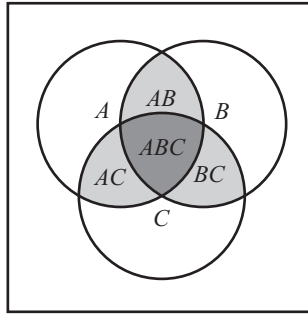
Figure 2.17: Venn diagram for three nonexclusive events.

union of the circles will be counted exactly once. The total area is therefore

$$\text{Total area} = A + B + C - AB - AC - BC + ABC, \qquad (2.80)$$

where we are using the regions' labels to stand for their areas. Translating this from a statement about areas to a statement about probabilities yields the desired result,

$$\begin{aligned} P(A \text{ or } B \text{ or } C) = {} & P(A) + P(B) + P(C) \\ & - P(A \text{ and } B) - P(A \text{ and } C) - P(B \text{ and } C) \\ & + P(A \text{ and } B \text{ and } C). \end{aligned} \qquad (2.81)$$

2.3. **"Or" rule for four events**

As in Problem 2.2(d), we'll discuss things in terms of areas. If we add up the areas of four regions, A, B, C, and D, then we have double counted the pairwise overlaps, so we need to subtract these off. There are six of these regions: AB, AC, AD, BC, BD, and CD. But then what about the triple overlaps, such as ABC? We counted ABC three times in the A, B, and C regions, but then we subtracted it off three times in the AB, AC, and BC regions. So at the moment, we haven't counted it at all. We therefore need to add it on once. (This is the same reasoning as in Problem 2.2(d).) Likewise for ABD, ACD, and BCD. Finally, what about the quadruple overlap region, $ABCD$? We counted this four times in the single regions (like A), then we subtracted it off six times in the double regions (like AB), and then we added it on four times in the triple regions (like ABC). So at the moment, we have counted it $4 - 6 + 4 = 2$ times. Since we want to count it only one time, we need to subtract it off once. The total area is therefore

$$\begin{aligned} \text{Total area} = {} & A + B + C + D \\ & - AB - AC - AD - BC - BD - CD \\ & + ABC + ABD + ACD + BCD \\ & - ABCD. \end{aligned} \qquad (2.82)$$

Writing this in terms of probabilities gives the result,

$$
\begin{aligned}
P(A \text{ or } B \text{ or } C \text{ or } D) = {} & P(A) + P(B) + P(C) + P(D) \\
& - P(A \text{ and } B) - P(A \text{ and } C) - P(A \text{ and } D) \\
& - P(B \text{ and } C) - P(B \text{ and } D) - P(C \text{ and } D) \\
& + P(A \text{ and } B \text{ and } C) + P(A \text{ and } B \text{ and } D) \\
& + P(A \text{ and } C \text{ and } D) + P(B \text{ and } C \text{ and } D) \\
& - P(A \text{ and } B \text{ and } C \text{ and } D).
\end{aligned}
\tag{2.83}
$$

REMARK: You might think that it's a bit of a coincidence that at every stage, we either overcounted or undercounted each region *once*. Equivalently, the coefficient of every term in Eqs. (2.82) and (2.83) is ± 1. The same thing is true in the case of three events in Eqs. (2.80) and (2.81). Likewise in the case of two events in Eq. (2.18), and trivially in the case of one event. Is it also true for larger numbers of events? Indeed it is, and the binomial expansion is the key to understanding why.

We won't go through every step, but if you want to think about it, the main points to realize are: First, the numbers 4, 6, and 4 in the above counting in the four-event case are actually the binomial coefficients $\binom{4}{1}$, $\binom{4}{2}$, $\binom{4}{3}$. This makes sense because, for example, the number of regions of double overlap (like AB) that contain the region $ABCD$ is simply the number of ways to pick two letters from four letters, which is $\binom{4}{2}$. Second, the "alternating sum" $\binom{4}{1} - \binom{4}{2} + \binom{4}{3}$ equals 2 (which means that we have overcounted the $ABCD$ region by one time), because this is what you obtain when you expand the righthand side of $0 = (1-1)^4$ with the binomial expansion. (This is a nice little trick.) And third, you can show how this generalizes to a larger number n of events. For even n, the alternating sum of the relevant binomial coefficients is 2, as we just saw for $n = 4$. For odd n, the alternating sum is zero, which means that we have undercounted by one time. (The relevant binomial coefficients are all but the first and last in the expansion of $(1-1)^n$, and these two coefficients are either 1 and 1 for even n, or 1 and -1 for odd n.) For example, $\binom{5}{1} - \binom{5}{2} + \binom{5}{3} - \binom{5}{4} = 0$. This "alternating sum" rule for counting is known as the *inclusion–exclusion principle*. ♣

2.4. **Red and blue balls**

By counting the various kinds of pairs in Table 2.1, we find $P(\text{Blue}_2) = 12/20 = 3/5$ (by looking at all 20 pairs), and $P(\text{Red}_1|\text{Blue}_2) = 6/12 = 1/2$ (by looking at only the 12 pairs below the horizontal line). So we have

$$
\begin{aligned}
P(\text{Red}_1 \text{ and } \text{Blue}_2) &= P(\text{Blue}_2) \cdot P(\text{Red}_1|\text{Blue}_2) \\
&= \frac{3}{5} \cdot \frac{1}{2} = \frac{3}{10},
\end{aligned}
\tag{2.84}
$$

in agreement with Eq. (2.10). As mentioned in the third remark on page 66, it still makes sense to talk about $P(\text{Red}_1|\text{Blue}_2)$, even though the second pick happens after the first pick.

2.5. **Dependent events**

FIRST SOLUTION: This problem is equivalent to finding the fraction of the total area that lies above the horizontal line segments in Fig. 2.16. The upper left region is $40\% = 2/5$ of the area that lies to the left of the vertical line, which itself is $20\% = 1/5$ of the total area. And the upper right region is $70\% = 7/10$ of the area that lies to the right of the vertical line, which itself is $80\% = 4/5$ of the total area. The fraction of the total area that lies above the horizontal line segments is therefore

$$
\frac{1}{5} \cdot \frac{2}{5} + \frac{4}{5} \cdot \frac{7}{10} = \frac{2}{25} + \frac{14}{25} = \frac{16}{25} = 64\%.
\tag{2.85}
$$

Second solution: We'll use the rule in Eq. (2.5) twice. First, note that

$$P(B) = P(A \text{ and } B) + P((\text{not } A) \text{ and } B). \qquad (2.86)$$

This is true because either A happens or it doesn't. We can apply Eq. (2.5) to each of the two terms in Eq. (2.86) to obtain

$$P(B) = P(A) \cdot P(B|A) + P(\text{not } A) \cdot P(B| \text{not } A)$$

$$= \frac{1}{5} \cdot \frac{2}{5} + \frac{4}{5} \cdot \frac{7}{10} = \frac{2}{25} + \frac{14}{25} = \frac{16}{25} = 64\%, \qquad (2.87)$$

which is exactly the same equation as in the first solution. This is no surprise, of course, because the two solutions are actually same. They are simply presented in a different language. Comparing the solutions makes it clear how conditional probabilities like $P(B|A)$ are related to fractional areas.

2.6. A single horizontal line

As usual, let the total area of the square in Fig. 2.16 be 1. Then from the given lengths along the sides of the square, we find that the upper two areas (probabilities) are 0.08 and 0.56, for a total of 0.64; this is $P(B)$. And the lower two areas are 0.12 and 0.24, for a total of 0.36; this is $P(\text{not } B)$. The single horizontal line in Fig. 2.18 must therefore be 64% of the way down from the top of the square. And the two vertical lines must be $0.08/0.64 = 12.5\%$ and $0.12/0.36 = 33.3\%$ of the way from the left side. The four areas are the same (by construction) as in Fig. 2.16. It's just that in Fig. 2.18, the $P(B) = 0.64$ probability is clear by simply looking at the figure. If we wanted to calculate $P(A)$ from Fig. 2.18, we would have to do a calculation analogous to the one we did in Problem 2.5.
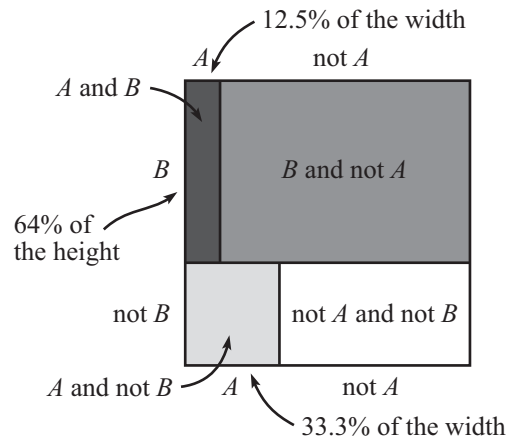

Figure 2.18: Redrawing Fig. 2.16 with a single horizontal line.

2.7. Proofreading

The breakdown of the errors is shown in Fig. 2.19. If the two people are labeled A and B, then 20 errors are found by both A and B, 80 are found by A but not B, and 40 are found by B but not A.

If we consider only the 100 errors found by A, we see that 20 of them are found by B, which is a $1/5$ fraction. Since we are assuming that B finding a given error is

Figure 2.19: Breakdown of errors found by *A* and *B*.

independent of *A* finding it, we see that if *B* finds 1/5 of the errors found by *A*, then he must find 1/5 of the complete set of errors (on average). So 1/5 is the probability that *B* finds any given error. Therefore, since we know that *B* found a total of 60 errors, the total number *N* of errors in the book must be given by $60/N = 1/5 \implies N = 300$. The unshaded region in Fig. 2.19 therefore represents $300 - 80 - 20 - 40 = 160$ errors. This is the number that both people missed.

We can also do things the other way around. If we consider only the 60 errors found by *B*, we see that 20 of them are found by *A*, which is a 1/3 fraction. By the same reasoning as above, this 1/3 is the probability that *A* finds any given error. And since we know that *A* found a total of 100 errors, the total number *N* must be given by $100/N = 1/3 \implies N = 300$, as above.

Another method (although in the end it's the same as the above methods) is the following. Let the area of the unshaded region in Fig. 2.19 be *x*. Then if we look at how the areas of the two vertical rectangles are divided by the horizontal line, we see that the ratio of *x* to 40 must equal the ratio of 80 to 20. So $x = 160$, as we found above. Alternatively, if we look at how the areas of the two horizontal rectangles are divided by the vertical line, we see that the ratio of *x* to 80 must equal the ratio of 40 to 20. So again, $x = 160$.

It is quite fascinating that you can get a sense of the total number of errors just by comparing the results of two readers' independent proofreadings. There is no need to actually find all the errors and count them up, if you only want to make a rough estimate. The larger the numbers involved, the better the estimate, in a multiplicative sense.

2.8. **Red balls, blue balls**

Let's ignore for a moment the fact that you happen to draw a red ball. Without this condition, there are six equally likely results of the process; you are equally likely to draw any of the six balls in the boxes. This fact can be argued by symmetry (there is nothing special about any of the balls). Or you can break down the probabilities: you have a 1/3 chance of drawing a given box, and then a 1/2 chance of drawing a given ball in that box. So all of the probabilities are equal to $(1/3)(1/2) = 1/6$.

Let's use the numbers 1 through 6 to label the balls: 1 and 2 are the two red balls in the first box, 3 and 4 are the two blue balls in the second box, and 5 and 6 are, respectively, the red and blue balls in the third box. If you play *n* games, where *n* is large, you will obtain approximately *n*/6 of each of the numbers 1 through 6.

Let's now invoke the fact that you draw a red ball. This means that the *n*/2 games

where you draw a blue ball (3, 4, and 6) aren't relevant. Only the $n/2$ games where you draw a red ball (1, 2, and 5) are relevant. And *of these games*, 2/3 have the ball coming from the first box (in which case the other ball is red), and 1/3 have the ball coming from the third box (in which case the other ball is blue). The desired probability that the other ball is red is therefore 2/3.

REMARKS:

1. The statement of the problem asks for the probability that the other ball in the box is also red, given that you draw a red ball. Since the word "probability" is used, it is understood that we must consider a large number of trials and look at what happens, on average, in these trials. Although the setup in the problem mentions only one trial, we must consider many. The given question, namely "If it turns out to be a red ball, what is the probability that the other ball in the box is also red?," is really just shorthand for the question, "If you run a large number of trials and look only at the ones where the drawn ball is red, in what fraction of these trials is the other ball in the box also red?"

2. In the statement of the problem, the clause, "You choose one of the boxes at random," is critical. Consider the alternative question: "Someone gives you a box containing either two red balls, two blue balls, or one of each. You draw a ball from this box. If it turns out to be a red ball, what is the probability that the other ball in the box is also red?" This question is unanswerable, because for all you know, the person always gives you a box with two red balls. Or perhaps she always gives you a box with one ball of each color, and you just happened to pick the red ball. Maybe it's 90% the former and 10% the latter, or maybe it depends on the day of the week. There is no way to tell what happens in a large number of trials. Even if you *do* perform a large number of trials and throw away the ones where you pick a blue ball, there is still no way to determine the probability associated with a future trial, because at any point the person might change her rules for the type of box she gives you.

3. What if, instead of three equally likely boxes sitting on the table, we have a single box and we color each of the two balls red or blue, based on coin tosses? There are then four equally likely possibilities for the contents of the box: RR, RB, BR, and BB. We therefore effectively have four equally likely boxes instead of three. You can show, with a quick modification of our original reasoning, that the answer is now 1/2 instead of 2/3.

   This result of 1/2 makes intuitive sense, due to the following alternative reasoning. Imagine picking a ball, without looking at it. The other ball has a 1/2 chance of being red, because its color is determined by a coin flip. Now look at the ball you picked. The other ball still has a 1/2 chance of being red, because your act of looking at the ball you picked can't change the color of the other ball. Therefore, if the ball you picked is red, then the other ball has a 1/2 chance of being red. Of course, the same thing is true if the ball you picked is blue, but those trials don't have anything to do with the given setup where you pick a red ball. ♣

2.9. **Sock pairs**

(a) The total number of possible pairs that you can draw from the eight socks in the drawer is $\binom{8}{2} = 28$. The number of ways that you can draw a red pair from the four red socks is $\binom{4}{2} = 6$. Likewise for the four blue socks. So there are

12 ways in all that you can draw a matching pair. The desired probability is therefore $12/28 = 3/7$.

(b) If there are now n red and n blue socks in the drawer, the total number of possible pairs that you can draw is $\binom{2n}{2} = 2n(2n-1)/2$. The number of ways that you can draw a red pair from the n red socks is $\binom{n}{2} = n(n-1)/2$. Likewise for the n blue socks. So there are $n(n-1)$ ways in all that you can draw a matching pair. The desired probability is therefore

$$\frac{n(n-1)}{2n(2n-1)/2} = \frac{n-1}{2n-1} .\tag{2.88}$$

If $n = 4$, this yields the probability of $3/7$ that we obtained in part (a).

(c) For the quick probability argument, imagine drawing the two socks in succession. The first sock is either red or blue. Whichever color it is, there are now $n-1$ socks remaining of that color. And there are $2n-1$ socks remaining in all. So the probability that the second sock has the same color as the first is $(n-1)/(2n-1)$.

For large n, this result approaches $1/2$. This makes sense because if n is large, the removal of the first sock from the drawer only negligibly changes the distribution of socks from 50-50. So you're basically flipping a coin with the second sock.

2.10. **Sock pairs, again**

(a) We know from Problem 2.9 that there is a $3/7$ probability of obtaining a matching first pair, and hence a $4/7$ probability of obtaining a non-matching first pair. So there is a $3/7$ probability that we are left with two socks of one color and four of the other, and there is a $4/7$ probability that we are left with three socks of each color.

In the first of these two cases, there are $\binom{6}{2} = 15$ possible pairs we can draw for our second pair, of which $\binom{2}{2} + \binom{4}{2} = 1 + 6 = 7$ are matching pairs. The probability that the second pair is matching, given that the first pair is matching (which happens with probability $3/7$), is therefore $7/15$.

Similarly, in the second of the two cases, there are again $\binom{6}{2} = 15$ possible pairs we can draw for our second pair, of which $\binom{3}{2} + \binom{3}{2} = 3 + 3 = 6$ are matching pairs. The probability that the second pair is matching, given that the first pair *isn't* matching (which happens with probability $4/7$), is therefore $6/15$.

The desired probability (that the second pair is matching) is therefore

$$\frac{3}{7} \cdot \frac{7}{15} + \frac{4}{7} \cdot \frac{6}{15} = \frac{21 + 24}{105} = \frac{3}{7} .\tag{2.89}$$

You can apply the same reasoning to the general case with n red and n blue socks, but it gets a bit messy. In any event, there is no need to work through the algebra, because there is a much quicker line of reasoning in part (b) below.

(b) We'll be general from the start here. That is, we'll assume that we have n socks of each color, and that we successively draw n pairs until there are no socks left in the drawer. We claim that all n pairs have the same $(n-1)/(2n-1)$ probability of matching, assuming that we haven't looked at any of the other pairs yet. This assumption is important; we must not have any knowledge of the other pairs. If we *do* have knowledge, then this affects the probabilities for future pairs. For

example, in part (a) above, we saw that if the first pair is matching, the second pair has a 7/15 chance of matching. But if the first pair isn't matching, the second pair has a 6/15 chance of matching.

Imagine drawing the $2n$ socks in succession and lining them up on a table. We can label them as s_1, s_2, s_3, ..., s_{2n}. We can then divide them into n pairs, (s_1, s_2), (s_3, s_4), ..., (s_{2n-1}, s_{2n}). If we ask for the probability that, say, the third pair (socks s_5 and s_6) is matching (assuming we haven't looked at any of the other pairs), we can *now* imagine looking at this particular pair. And if we look at s_5 first and then at s_6, we can use the reasoning in part (c) of Problem 2.9 to say that the probability of a matching pair is $(n-1)/(2n-1)$. This reasoning works for any of the n pairs; there is nothing special about a specific pair (assuming we haven't looked at any of the other pairs). All pairs therefore have equal $(n-1)/(2n-1)$ probabilities of being matching pairs.

The point here is that if you don't look at the pairs you've already picked, then for all practical purposes the present pair you're picking is the first pair. The order in which you draw the pairs therefore doesn't matter, so the desired probabilities are all equal.

2.11. **At least one 6**

The probability of obtaining exactly one 6 equals $\binom{3}{1} \cdot (1/6)(5/6)^2$, because there are $\binom{3}{1} = 3$ ways to pick which die is the 6. And then given this choice, there is a 1/6 chance that the die is in fact a 6, and a $(5/6)^2$ chance that both of the other dice are not 6's.

The probability of obtaining exactly two 6's equals $\binom{3}{2} \cdot (1/6)^2(5/6)$, because there are $\binom{3}{2} = 3$ ways to pick which two dice are the 6's. And then given this choice, there is a $(1/6)^2$ chance that they are in fact both 6's, and a 5/6 chance that the other die is not a 6.

The probability of obtaining exactly three 6's equals $\binom{3}{3} \cdot (1/6)^3$, because there is just $\binom{3}{3} = 1$ way for all three dice to be 6's. And then there is a $(1/6)^3$ chance that they are in fact all 6's.

The total probability of obtaining at least one six is therefore

$$\binom{3}{1} \cdot \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2 + \binom{3}{2} \cdot \left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right) + \binom{3}{3} \cdot \left(\frac{1}{6}\right)^3 = \frac{75}{216} + \frac{15}{216} + \frac{1}{216}$$

$$= \frac{91}{216}, \tag{2.90}$$

in agreement with the result in Section 2.3.1.

REMARK: If we add this result to the probability of obtaining zero 6's, which is $(5/6)^3$, the sum is 1, because we have now taken into account every possible outcome. This fact was what we used to solve the problem the quick way in Section 2.3.1, after all. But let's pretend that we don't know the sum is 1, and let's verify this explicitly. If we write $(5/6)^3$ suggestively as $\binom{3}{0} \cdot (5/6)^3$, then our goal is to show that

$$\binom{3}{0} \cdot \left(\frac{5}{6}\right)^3 + \binom{3}{1} \cdot \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2 + \binom{3}{2} \cdot \left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right) + \binom{3}{3} \cdot \left(\frac{1}{6}\right)^3 = 1. \tag{2.91}$$

This is indeed a true statement, because the lefthand side is simply the binomial expansion of $(5/6 + 1/6)^3 = 1$. This makes it clear why the sum of the probabilities of

the various outcomes will still be 1, even if we have, say, an eight-sided die (again, forgetting that we know intuitively that the sum must be 1). The only difference is that we now have the expression $(7/8 + 1/8)^3 = 1$, which is still true. And any other exponent (that is, any other number of rolls) will also yield a sum of 1, as we know it must. ♣

2.12. **At least one 6, by the rules**

We'll copy Eq. (2.79) here:

$$
\begin{aligned}
P(A \text{ or } B \text{ or } C) = {}& P(A) + P(B) + P(C) \\
& - P(A \text{ and } B) - P(A \text{ and } C) - P(B \text{ and } C) \\
& + P(A \text{ and } B \text{ and } C).
\end{aligned}
\tag{2.92}
$$

The lefthand side of this equation is the probability of obtaining at least one 6. (Remember that the "or" is the "inclusive or.") So our task is to evaluate the righthand side, which involves three different types of terms.

The probability of obtaining a 6 on any given die (without caring what happens with the other two dice) is 1/6, so

$$
P(A) = P(B) = P(C) = \frac{1}{6}.
\tag{2.93}
$$

The probability of obtaining 6's on two given dice (without caring what happens with the third die) is $(1/6)^2$, so

$$
P(A \text{ and } B) = P(A \text{ and } C) = P(B \text{ and } C) = \frac{1}{36}.
\tag{2.94}
$$

The probability of obtaining 6's on all three dice is $(1/6)^3$, so

$$
P(A \text{ and } B \text{ and } C) = \frac{1}{216}.
\tag{2.95}
$$

Eq. (2.92) therefore gives the probability of obtaining at least one 6 as

$$
3 \cdot \frac{1}{6} - 3 \cdot \frac{1}{36} + \frac{1}{216} = \frac{108 - 18 + 1}{216} = \frac{91}{216},
\tag{2.96}
$$

in agreement with the result in Section 2.3.1 and Problem 2.11.

2.13. **Rolling sixes**

(a) In all three parts of this problem, there are far fewer ways to fail to obtain the specified number of 6's than to succeed. So we'll calculate the probability of failure and then subtract that from 1 to obtain the probability of success.

If 6 dice are rolled, the probability of obtaining zero 6's is $(5/6)^6$. The probability of obtaining at least one 6 is therefore

$$
1 - \left(\frac{5}{6}\right)^6 = 0.665.
\tag{2.97}
$$

(b) If 12 dice are rolled, the probability of obtaining zero 6's is $(5/6)^{12}$, and the probability of obtaining exactly one 6 is $\binom{12}{1}(1/6)^1(5/6)^{11}$, because there are $\binom{12}{1}$ possibilities for the one die that shows a 6. The probability of obtaining at least two 6's is therefore

$$
1 - \left(\frac{5}{6}\right)^{12} - \binom{12}{1}\left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^{11} = 0.619.
\tag{2.98}
$$

(c) Similarly, if 18 dice are rolled, the probability of obtaining zero 6's is $(5/6)^{18}$, the probability of obtaining exactly one 6 is $\binom{18}{1}(1/6)^1(5/6)^{17}$, and the probability of obtaining exactly two 6's is $\binom{18}{2}(1/6)^2(5/6)^{16}$. The probability of obtaining at least three 6's is therefore

$$1 - \left(\frac{5}{6}\right)^{18} - \binom{18}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^{17} - \binom{18}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{16} = 0.597. \qquad (2.99)$$

We see that the probability in part (a) is the largest.

REMARK: We can also pose the problem with larger numbers of rolls. For example, if 600 dice are rolled, what is the probability of obtaining at least 100 6's? Or more generally, if $6n$ dice are rolled, what is the probability of obtaining at least n 6's? From the same type of reasoning as above, the answer in the general case is

$$1 - \sum_{k=0}^{n-1}\binom{6n}{k}\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{6n-k}. \qquad (2.100)$$

For large n, it is intractable to evaluate this sum by hand. But it's easy to use a computer to evaluate it for any n. For $n = 10$, 100, and 1000 we obtain probabilities of, respectively, 0.554, 0.517, and 0.505. These probabilities decrease with n, and they appear to approach the nice simple answer of $1/2$ in the $n \to \infty$ limit. See Problem 5.2 for an explanation of where this $1/2$ comes from. ♣

2.14. **Exactly one pair**

There are $\binom{23}{2}$ possible pairs that can have the common birthday. Let's look at one particular pair and calculate the probability that these two people have a common birthday, while everyone else has a *unique* birthday. We'll then multiply this result by $\binom{23}{2}$ to account for all the possible pairs.

The probability that a given pair has a common birthday is $1/365$, because the first person's birthday can be chosen to be any day, and then the second person has a $1/365$ chance of matching that day. We then need the 21 other people to have 21 different birthdays, none of which is the same as the pair's birthday. The first of these people can end up in any of the remaining 364 days; this happens with probability $364/365$. The second of these people can end up in any of the remaining 363 days; this happens with probability $363/365$. And so on, until the 21st of these people can end up in any of the remaining 344 days; this happens with probability $344/365$.

The total probability that exactly one pair has a common birthday is therefore

$$\binom{23}{2} \cdot \frac{1}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \frac{362}{365} \cdot \cdots \cdot \frac{344}{365}. \qquad (2.101)$$

Multiplying this out gives $0.363 = 36.3\%$. This is smaller than the "at least one common birthday" result of 50.7% that we found in Section 2.4.1 for 23 people, as it must be. The remaining $50.7\% - 36.3\% = 14.4\%$ probability corresponds to occurrences of two different pairs with common birthdays, or three people with a common birthday, etc.

2.15. **My birthday**

(a) p is smaller than $100/365$. If the events "Person A having your birthday" and "Person B having your birthday," etc., were all mutually exclusive, then p would

be equal to 100/365. But these events are not mutually exclusive, because it is certainly possible for two (or more) of the people to have your birthday. These multiple-event probabilities are counted twice (or more) in the naive 100/365 result. So they must be subtracted off in order to obtain the correct probability. The correct probability is therefore smaller than 100/365.

Note that if we replace the number 100 here by 365 (or anything larger), then the "smaller" answer is obvious, because the probability p is certainly smaller than 365/365 = 1. This suggests (although it doesn't prove) that the answer for the number 100 (or any other number) is "smaller." The one exception is where 100 is replaced by 1, that is, where there is only one other person in the room. In this case we don't have to worry about double counting any probabilities, so the answer is exactly 1/365.

(b) The probability that *no one* out of the 100 people has your birthday equals $(364/365)^{100}$. The probability that at least one of them *does* have your birthday is therefore

$$p = 1 - \left(\frac{364}{365}\right)^{100} = 0.24. \tag{2.102}$$

This is indeed smaller than 100/365 = 0.27. It is only slightly smaller, though, because the multiple-event probabilities are small.

2.16. **My birthday, again**

We may as well be general right from the start and assume that there are N days in a year. We can eventually set $N = 365$. If there are N days in a year, then the probability that *no one* out of n people has your birthday equals $(1 - 1/N)^n$. This is an exact expression, but we can simplify it by making use of the approximation in Eq. (7.14), namely $(1 + a)^n \approx e^{na}$. With $a \equiv -1/N$ here, $(1 - 1/N)^n$ becomes

$$\left(1 - \frac{1}{N}\right)^n \approx e^{-n/N}. \tag{2.103}$$

Our goal is to have this probability be smaller than 1/2, so that the probability that someone *does* have your birthday is larger than 1/2. Taking the log of both sides of $e^{-n/N} < 1/2$ gives

$$-\frac{n}{N} < \ln\left(\frac{1}{2}\right) \quad \Longrightarrow \quad -\frac{n}{N} < -\ln 2 \quad \Longrightarrow \quad \frac{n}{N} > \ln 2 \tag{2.104}$$

$$\Longrightarrow \quad n > N \ln 2 \approx (0.693)N.$$

Therefore, if $n > N \ln 2$, it is more likely than not that at least one of the n people has your birthday. For $N = 365$, we find that $N \ln 2$ is slightly less than 253, so this agrees with the (exact) result we obtained by simply taking the nth power of 364/365. Since $\ln 2$ is very close to 0.7, a quick approximation to the answer to this problem is $(0.7)N$.

2.17. **My birthday, yet again**

ONE PERSON: The probability that a specific person has your birthday is 1/365. Since we want *exactly* one person to have your birthday, we want none of the other 252 people to have it; this occurs with probability $(364/365)^{252}$. There are 253 ways to pick the specific person who has your birthday, so the total probability that exactly one of the 253 people has your birthday is

$$253 \cdot \frac{1}{365} \cdot \left(\frac{364}{365}\right)^{252} = 0.347. \tag{2.105}$$

Two PEOPLE: The probability that two specific people have your birthday is $(1/365)^2$. The probability that none of the other 251 people have your birthday is $(364/365)^{251}$. There are $\binom{253}{2}$ ways to pick the two specific people who have your birthday, so the total probability that exactly two of the 253 people have your birthday is

$$\binom{253}{2}\left(\frac{1}{365}\right)^2\left(\frac{364}{365}\right)^{251} = 0.120. \tag{2.106}$$

THREE PEOPLE: By similar reasoning, the probability that exactly three of the 253 people have your birthday is

$$\binom{253}{3}\left(\frac{1}{365}\right)^3\left(\frac{364}{365}\right)^{250} = 0.0276. \tag{2.107}$$

The pattern is clear. The probability that exactly k people have your birthday is

$$P(k) = \binom{253}{k}\left(\frac{1}{365}\right)^k\left(\frac{364}{365}\right)^{253-k}. \tag{2.108}$$

For $k = 0$, this gives the $(364/365)^{253} \approx 1/2$ probability (obtained at the end of Section 2.4.1 and in Problem 2.16) that no one has your birthday. Note that the $P(k)$ probabilities are simply the terms in the binomial expansion:

$$\left(\frac{1}{365} + \frac{364}{365}\right)^{253} = \sum_{k=0}^{253}\binom{253}{k}\left(\frac{1}{365}\right)^k\left(\frac{364}{365}\right)^{253-k}. \tag{2.109}$$

Since the lefthand side of this equation equals 1, we see that the sum of the $P(k)$ also equals 1. This must be the case, of course, because the number of other people who have your birthday has to be *something*.

2.18. **A random game-show host**

We'll solve this problem by listing out the various possibilities. Without loss of generality, assume that you pick the first door. (You can repeat the following reasoning for the other doors if you wish. It gives the same result.) There are three equally likely possibilities for what is behind the three doors: PGG, GPG, and GGP, where P denotes the prize and G denotes a goat. For each of these three possibilities, since you picked the first door, the host opens either the second or third door (with equal probabilities). So there are six equally likely results of his actions. These are shown in Fig. 2.7, with the bold letters signifying the object revealed.

	PGG	GPG	GGP
open 2nd door	P**G**G	G**P**G̶	G**G**P
open 3rd door	P**G**G	GP**G**	G̶G̶**P**̶

Table 2.7: There are six equally likely scenarios with a randomly opened door, assuming that you pick the first door.

We now note that the two results where the prize is revealed (the crossed-out G**P**G and G**G**P results) are not relevant to this problem, because we are told that the host happens to reveal a goat. Only the four other results are relevant:

$$\text{P}\mathbf{G}\text{G} \qquad \text{PG}\mathbf{G} \qquad \text{GP}\mathbf{G} \qquad \text{GG}\mathbf{P}$$

They are all still equally likely, so their probabilities must each be 1/4. We see that if you *don't* switch from the first door, you win on the first two of these results and lose on the second two. And if you *do* switch, you lose on the first two and win on the second two. So either way, your probability of winning is 1/2. It therefore doesn't matter if you switch.

REMARKS:

1. In the original version of the problem in Section 2.4.2, the probability of winning was 2/3 if you switched. How can it possibly decrease to 1/2 in the present random version, when in both versions the exact same thing happened, namely the host revealed a goat?

   The difference is due to the two cases where the host reveals the prize in the random version (the **GPG** and **GGP** cases). You don't benefit from these cases in the random version, because we are told in the statement of the problem that they don't exist. But in the original version, they represent guaranteed success if you switch, because the host is forced to open the other door, which is a goat.

   But still you may say, "If there are two setups, and if I pick, say, the first door in each, and if the host reveals a goat in each (by prediction in one case, and by random pick in the other), then *exactly the same thing happens in both setups.* How can the resulting probabilities (for winning on a switch) be different?" The answer is that although the two outcomes are the same, probabilities have nothing to do with *two* setups. Probabilities are defined only for a *large number* of setups. And if you play a large number of these pairs of games (prediction in one, random pick in the other), then in 1/3 of the pairs the host will reveal different things (a goat in the prediction version and the prize in the random version). These cases yield success in the original prediction version, but they are irrelevant in the random version. They are effectively thrown away there.

2. We will now address the issue mentioned in the fourth remark in Section 2.4.2. We correctly stated in Section 2.4.2 that in the original version of the problem, "No actions taken by the host can change the fact that if you play a large number n of these games, then (roughly) $n/3$ of them will have the prize behind the door you initially pick." However, in the present random version of the problem, something *does* affect the probability that the prize is behind the door you initially pick. It is now 1/2 instead of 1/3. So can something affect this probability or not?

   Well, yes and no. If *all* of the n games are considered (as in the original version), then $n/3$ of them have the prize behind the initial door, and that's that. However, the random version of the problem involves throwing away 1/3 of the games (the ones where the host reveals the prize), because it is assumed in the statement of the problem that the host happens to reveal a goat. So for the *remaining games* (which are 2/3 of the initial total, hence $2n/3$), 1/2 of them now have the prize behind your initial door.

   If you play a large number n of games of each version (including the $n/3$ games that are thrown away in the random version), then the actual *number* of games that have the prize behind your initial door is the same, namely $n/3$. It's just that in the original version this number can be thought of as 1/3 of n, whereas in the random version it can be thought of as 1/2 of $2n/3$. So in the end, the thing that influences the probability (that the initial door you pick has the prize) and changes it from 1/3 to 1/2 isn't the opening of a door, but rather the throwing away of 1/3 of the games. Since no games are thrown away in the original

version, the above statement in quotes is correct (with the key phrase being "*these* games").

3. As with the original version of the problem, if you find yourself arguing about the answer for an excessive amount of time, you should just *play the game* a bunch of times (at least a few dozen, to get good enough statistics). The randomness can be determined by a coin toss. As mentioned above, you will end up throwing away 1/3 of the games (the ones where the host reveals the prize). ♣

2.19. Boy/girl problem with general information

Let's be general right from the start and consider the case where the boy has a particular characteristic that occurs with probability p. (So $p = 1/4$ if the characteristic is a summer birthday.) As in all of the versions of this problem in Section 2.4.4, we'll list out the various possibilities in a table, *before* the parent's additional information (beyond "I have two children") is taken into account. It is still the case that the BB, BG, GB, and GG types of two-child families are all equally likely, with a 1/4 probability for each. We are again ordering the children in a given pair by age; the first letter is associated with the older child. But we could just as well order them by, say, height or shoe size.

In the present version of the problem, there are now various different subtypes within each type of family, depending on whether or not the children have the given characteristic (which occurs with probability p). For example, if we look at the BB types, there are four possibilities for the occurrence(s) of the characteristic. With "y" standing for "yes, the child has the characteristic," and "n" standing for "no, the child doesn't have the characteristic," the four possibilities are B_yB_y, B_yB_n, B_nB_y, and B_nB_n. (In the second possibility here, for example, the older boy has the characteristic, and the younger boy doesn't.) Since y occurs with probability p, we know that n occurs with probability $1 - p$. The probabilities associated with each of the four possibilities are therefore equal to the 1/4 probability that BB occurs, multiplied by, respectively, p^2, $p(1 - p)$, $(1 - p)p$, and $(1 - p)^2$.

The same reasoning holds with the BG, GB, and GG types, so we obtain a total of $4 \cdot 4 = 16$ distinct possibilities. These are listed in Table 2.8 (ignore the boxes for a moment). The four subtypes in any given row all have the same occurrence(s) of the characteristic, so they all have the same probability; this probability is listed on the right. The subtypes in the middle two rows all have equal probabilities. As mentioned above, in the case where the given characteristic is "having a birthday in the summer," p equals 1/4. So the probabilities associated with the four rows in that case are equal to 1/4 multiplied by, respectively, 1/16, 3/16, 3/16, and 9/16.

Before the parent gives you the additional information, all 16 of the subtypes in the table are possible. But after the statement is made that there is at least one boy with the given characteristic (that is, there is at least one B_y in the pair of children), only seven subtypes remain. These are indicted with boxes. The other nine subtypes are ruled out.

We now simply observe that the three boxes in the left-most column in the table have the other child being a boy, while the four other boxes in the second and third columns have the other child being a girl. The desired probability that the other child is a boy is therefore equal to the sum of the probabilities of the left three boxes, divided by the sum of the probabilities of all seven boxes. This gives (ignoring the common factor of 1/4 in all of the probabilities)

$$P_{\text{BB}} = \frac{p^2 + 2 \cdot p(1 - p)}{3 \cdot p^2 + 4 \cdot p(1 - p)} = \frac{2p - p^2}{4p - p^2} = \frac{2 - p}{4 - p} \ . \tag{2.110}$$

	BB	BG	GB	GG	Probability
yy	B_yB_y	B_yG_y	G_yB_y	G_yG_y	$(1/4) \cdot p^2$
yn	B_yB_n	B_yG_n	G_yB_n	G_yG_n	$(1/4) \cdot p(1-p)$
ny	B_nB_y	B_nG_y	G_nB_y	G_nG_y	$(1/4) \cdot p(1-p)$
nn	B_nB_n	B_nG_n	G_nB_n	G_nG_n	$(1/4) \cdot (1-p)^2$

Table 2.8: The 16 types of families.

In the case where the given characteristic is "having a birthday in the summer," p equals 1/4. Plugging this into Eq. (2.110) gives the probability that the other child is also a boy as $P_{BB} = 7/15 = 0.467$.

If the given characteristic is "having a birthday on August 11th," then $p = 1/365$, which yields $P_{BB} = 729/1459 = 0.4997 \approx 1/2$.

If the given characteristic is "being born during a particular minute on August 11th," then p is essentially equal to zero, so Eq. (2.110) tells us that P_{BB} is essentially equal to 1/2. This makes sense, because if $p = 0$, then the $p(1-p)$ probability for the middle two rows in Table 2.8 is much larger than the p^2 probability for the top row. Of course, *all* of these probabilities are very small in the small-p limit, but p^2 is much smaller than $p(1-p) \approx p$ when p is small. So we can ignore the top row. We are then left with four boxes, two of which are BB and two of which are BG/GB. The desired probability therefore equals 1/2.

Another somewhat special case is $p = 1/2$. (You can imagine that every child flips a coin, and we're concerned with the children who get Heads.) In this case we have $p = 1-p$, so all of the probabilities in the righthand column in Table 2.8 are equal. All 16 entries in the table therefore have equal probabilities (namely 1/16). Determining probabilities is then just a matter of counting boxes, so the answer to the problem is 3/7, because three of the seven boxes are of the BB type.

REMARKS:

1. The above $P_{BB} \approx 1/2$ result in the $p \approx 0$ case leads to the following puzzle. Let's say that you bump into a random person on the street who says, "I have two children. At least one of them is a boy." At this stage, you know that the probability that the other child is also a boy is 1/3, from part (a) of the original problem in Section 2.4.4. But if the parent then adds, "...who was born during a particular minute on August 11th," then we just found that the probability that the other child is also a boy jumps to (essentially) 1/2. Why exactly did this jump take place?

   In the original scenario in Section 2.4.4, there were three equally likely possibilities after the parent gave the additional information, namely BB, BG, and GB. Only 1/3 of these cases (namely BB) had the other child being a boy. In the new scenario (with $p \approx 0$), there are four equally likely possibilities after the parent gives the additional information, namely B_yB_n, B_nB_y, B_yG_n, and G_nB_y. (As mentioned above, we're ignoring the top row in Table 2.8 since $p \approx 0$.)

So in the new scenario, 1/2 of these cases (the two BB cases) have the other child being a boy. The critical point here is that BB now counts *twice*, whereas it counted only once in the original scenario. This is due to the fact that a BB parent is *twice as likely* (compared with a BG or GB parent) to be able to say that a boy was born during a particular minute on August 11th, because with two boys there are two chances to achieve this highly improbable characteristic. In contrast, a BB parent is *no more likely* (compared with a BG or GB parent) to be able to say simply that at least one child is a boy.

2. In the other extreme where the given characteristic is "being born on *any* day," we have $p = 1$. (This clearly isn't much of a characteristic, since it is satisfied by everyone.) So Eq. (2.110) gives $P_{BB} = 1/3$. In this $p = 1$ case, only the entries in the top row in Table 2.8 have nonzero probabilities. We are therefore in the realm of the first scenario in Section 2.4.4, where we started off with the four types of families (BB, BG, GB, GG) and then ruled out the GG type, yielding a probability of 1/3. It makes sense that the 1/3 answer in the $p = 1$ case is the same as the 1/3 answer in the first scenario in Section 2.4.4, because the "being born on *any* day" statement provides no additional information. So the setup is equivalent to the first scenario in Section 2.4.4, where the parent provided no additional information (beyond the fact that one child was a boy). ♣

2.20. **A second test**

The relevant probability tree is obtained by simply tacking on one more iteration of branches to Fig. 2.11. The result is shown in Fig. 2.20. (We've again arbitrarily started with 1000 people.) We are concerned only with the two numbers 18.05 and 9.8, because these are the only numbers associated with positive results for both tests (labeled as "++"). The desired probability is therefore

$$p = \frac{18.05}{18.05 + 9.8} = 64.8\%. \tag{2.111}$$

This is significantly larger than the result of 16% in the original example in Section 2.5.

Note that since we are concerned only with two of the final eight numbers, there was actually no need to draw the entire probability tree. The two relevant numbers are obtained from the products,

$$(1000)(0.02)(0.95)(0.95) = 18.05,$$
$$(1000)(0.98)(0.1)(0.1) = 9.8. \tag{2.112}$$

These products make it clear how to proceed in the general case of n tests. If we perform n successive tests on each person, then the probability that a person who tests positive all n times actually has the disease is

$$p = \frac{(0.02)(0.95)^n}{(0.02)(0.95)^n + (0.98)(0.1)^n}. \tag{2.113}$$

If $n = 1$ then $p = 0.16$, as we found in the original example. If, say, $n = 4$, then $p = 99.4$. Here the smallness of the $(0.1)^n$ factor in Eq. (2.113) wins out over the smallness of the 0.02 factor. In this case, although not many people have the disease, the number of people who falsely test positive all four times is even smaller. If n is large, then p is essentially equal to 1.
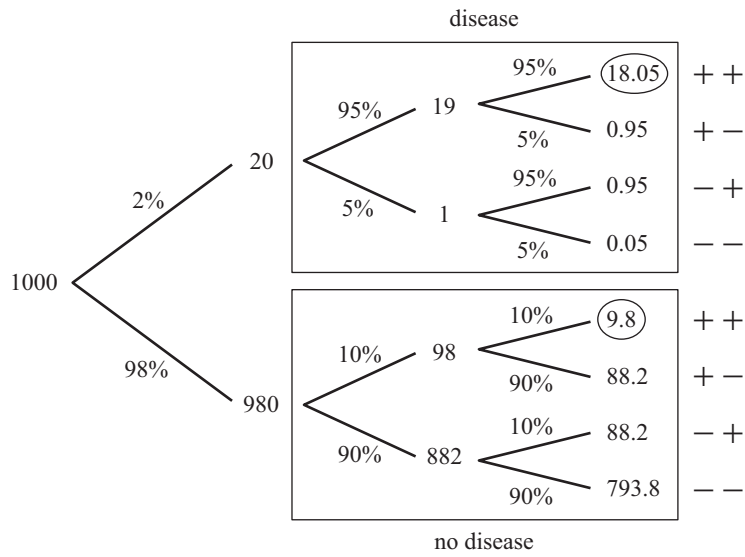
disease

no disease

Figure 2.20: The probability tree for two tests.

2.21. **Bayes' theorem for the prosecutor's fallacy**

A given person is either innocent or guilty, and either fits the description or doesn't. Our goal is to find $P(I|D)$. From the second remark at the end of Section 2.5, we want the horizontal span of our square to be associated with the innocent and guilty possibilities, and the vertical span to be associated with the description or not-description possibilities. The result is shown in Fig. 2.21. This figure contains the same information as Fig. 2.10, but in rectangular instead of oval form.
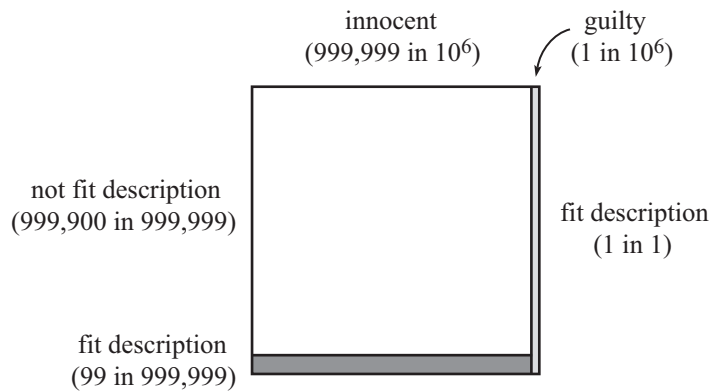
innocent
(999,999 in 10^6)

guilty
(1 in 10^6)

not fit description
(999,900 in 999,999)

fit description
(1 in 1)

fit description
(99 in 999,999)

Figure 2.21: The probability square for the prosecutor's fallacy.

We haven't draw things to scale, because if we did, both of the shaded rectangles would be too thin to see. The thin vertical rectangle represents the single guilty person, and the rest of the square represents the 999,999 innocent people. The guilty person fits the description, so the entire thin vertical rectangle is shaded. (As in Fig. 2.10,

the fourth possible group of people – guilty and not fitting the description – has zero people in it.) Only about 0.01% of the innocent people fit the description, so the darkly shaded rectangle is very squat. The desired probability $P(I|D)$ equals the number of people in the darkly shaded region (namely 99) divided by the total number of people in both shaded regions (namely 99 + 1). So the desired probability $P(I|D)$ of being innocent, given that the description is satisfied, equals $99/100 = 0.99$.

As we mentioned in Section 2.4.3, $P(I|D)$ (which is close to 1) is *not* equal to $P(D|I)$ (which is close to 0). The former is the ratio of the darkly shaded area to the total shaded area, while the latter is the ratio of the darkly shaded area to the area of the entire left vertical rectangle (the whole square minus the one guilty person).

If you want to use the simple form of Bayes' theorem in Eq. (2.51), instead of using the probability square in Fig. 2.21, you can write

$$P(I|D) = \frac{P(D|I) \cdot P(I)}{P(D)} = \frac{\frac{99}{999,999} \cdot \frac{999,999}{10^6}}{\frac{1}{10^4}} = \frac{99}{100}, \qquad (2.114)$$

as desired. You can verify that the various probabilities we used here are correct.

2.22. **Black balls and white balls**

Our goal is to calculate the probability that the box you pick is the one with two black balls, given that all n draws are black. We'll denote this probability by $P(B_2|nB)$. The two possibilities for the box you pick are the B_2 box with two black balls, and the B_1 box with one black ball. So with this notation, the general form (which is the same as the explicit form in this case) of Bayes' theorem in Eq. (2.53) gives

$$P(B_2|nB) = \frac{P(nB|B_2) \cdot P(B_2)}{P(nB|B_2) \cdot P(B_2) + P(nB|B_1) \cdot P(B_1)} . \qquad (2.115)$$

We are given that $P(B_1) = P(B_2) = 1/2$, and we also know that $P(nB|B_2) = 1$ and $P(nB|B_1) = (1/2)^n$. So Bayes' theorem gives

$$P(B_2|nB) = \frac{1 \cdot (1/2)}{1 \cdot (1/2) + (1/2)^n \cdot (1/2)}$$
$$= \frac{1}{1 + 1/2^n} = \frac{2^n}{2^n + 1} . \qquad (2.116)$$

If $n = 1$, then $P(B_2|nB) = 2/3$. And if $n = 10$, then $P(B_2|nB) = 1024/1025 \approx 99.9\%$.

If you want to solve the problem without explicitly using Bayes' theorem, the math turns out to be essentially the same. Imagine doing a large number N of trials of the given process. On average, you will pick each of the two boxes $N/2$ times. All n draws will be black in all of the $N/2$ cases where you pick B_2. But all n draws will be black in only $1/2^n$ of the $N/2$ cases where you pick B_1. The other $1 - 1/2^n$ fraction of the cases (where you draw at least one white ball) aren't relevant here. You are therefore dealing with the B_2 box in $N/2$ of the $N/2 + (N/2)/2^n$ times that you draw n black balls. The desired probability is then

$$P(B_2|nB) = \frac{N/2}{N/2 + (N/2)/2^n}$$
$$= \frac{1}{1 + 1/2^n}, \qquad (2.117)$$

as above.