

Contents

1	Basic ideas in probability	2
1.1	Experiments, Events, and Probability	2
1.1.1	The Probability of an Outcome	3
1.1.2	Events	5
1.1.3	The Probability of Events	7
1.1.4	The Gambler's Ruin	12
1.2	Conditional probability	13
1.2.1	Independence	19
1.2.2	The Monty Hall Problem, and other Perils of Conditional Probability	23
1.3	Simulation and Probability	25

CHAPTER 1

Basic ideas in probability

We need some machinery to deal with uncertainty, to account for new information, and to weigh uncertainties against one another. The appropriate machinery is probability, which allows us to reduce uncertain situations to idealized models that are often quite easy to work with.

1.1 EXPERIMENTS, EVENTS, AND PROBABILITY

If we flip a fair coin many times, we expect it to come up heads about as often as it comes up tails. If we toss a fair die many times, we expect each number to come up about the same number of times. We are performing an experiment each time we flip the coin, and each time we toss the die. We can formalize this experiment by describing the set of **outcomes** that we expect from the experiment. In the case of the coin, the set of outcomes is:

$$\{H, T\}.$$

In the case of the die, the set of outcomes is:

$$\{1, 2, 3, 4, 5, 6\}.$$

Notice that we are making a modelling choice by specifying the outcomes of the experiment, and this is typically an idealization. For example, we are assuming that the coin can only come up heads or tails (but doesn't stand on its edge; or fall between the floorboards; or land behind the bookcase; or whatever). It is often relatively straightforward to make these choices, but you should recognize them as an essential component of the model. Small changes in the details of a model can make quite big changes in the space of outcomes. We write the set of all outcomes Ω ; this is sometimes known as the **sample space**.

Worked example 1.1 *Find the lady*

We have three playing cards. One is a queen; one is a king, and one is a knave. All are shown face down, and one is chosen at random and turned up. What is the set of outcomes?

Solution: Write Q for queen, K for king, N for knave; the outcomes are $\{Q, K, N\}$

Worked example 1.2 *Find the lady, twice*

We play Find the Lady twice, replacing the card we have chosen. What is the set of outcomes?

Solution: We now have $\{QQ, QK, QN, KQ, KK, KN, NQ, NK, NN\}$

Worked example 1.3 *Children*

A couple decides to have children until either (a) they have both a boy and a girl or (b) they have three children. What is the set of outcomes?

Solution: Write B for boy, G for girl, and write them in birth order; we have $\{BG, GB, BBG, BBB, GGB, GGG\}$.

Worked example 1.4 *Monty Hall (sigh!)*

There are three boxes. There is a goat, a second goat, and a car. These are placed into the boxes at random. The goats are indistinguishable. What are the outcomes?

Solution: Write G for goat, C for car. Then we have $\{CGG, GCG, GGC\}$.

Worked example 1.5 *Monty Hall, different goats (sigh!)*

There are three boxes. There is a goat, a second goat, and a car. These are placed into the boxes at random. One goat is male, the other female, and the distinction is important. What are the outcomes?

Solution: Write M for male goat, F for female goat, C for car. Then we have $\{CFM, CMF, FCM, MCF, FMC, MFC\}$. Notice how the number of outcomes has increased, because we now care about the distinction between goats.

1.1.1 The Probability of an Outcome

We represent our model of how often a particular outcome will occur in a repeated experiment with a **probability**, a non-negative number. It is quite difficult to give a good, rigorous definition of what probability means. For the moment, we use a simple definition. Assume an outcome has probability P . Assume we repeat the experiment a very large number of times N , and each repetition is independent (more on this later; for the moment, assume that the coins/dice/whatever don't communicate with one another from experiment to experiment). Then, for about $N \times P$ of those experiments the outcome will occur (and as the number of experiments gets bigger, the fraction where the outcome occurs will get closer to P). That is, the **relative frequency** of the outcome is P .

Notice that this means that the probabilities of *outcomes* must add up to one, because each of our experiments has an outcome. We will formalize this below.

For example, if we have a coin where the probability of getting heads is $P(H) = \frac{1}{3}$, and so the probability of getting tails is $P(T) = \frac{2}{3}$, we expect this coin will come up heads in $\frac{1}{3}$ of experiments. This is not a guarantee that if you flip this coin three times, you will get one head. Instead, it means that, if you flip this coin three million times, you will very likely see very close to a million heads.

As another example, in the case of the die, we could have

$$P(1) = \frac{1}{18} \quad P(2) = \frac{2}{18} \quad P(3) = \frac{1}{18} \quad P(4) = \frac{3}{18} \quad P(5) = \frac{10}{18} \quad P(6) = \frac{1}{18}.$$

In this case, we'd expect to see five about 10,000 times in 18,000 throws.

Some problems can be handled by building a set of outcomes and reasoning about the probability of each outcome. This gets clumsy when there are large numbers of events, and is easiest when everything has the same probability, but it can be quite useful.

For example, assume we have a fair coin. We interpret this to mean that $P(H) = P(T) = \frac{1}{2}$, so that heads come up as often as tails in repeated experiments. Now we flip this coin twice - what is the probability we see two heads?

The set of outcomes is

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

and each outcome must occur equally often. So the probability is $\frac{1}{4}$.

Now consider a fair die. The space of outcomes is

$$\{1, 2, 3, 4, 5, 6\}.$$

The die is fair means that each event has the same probability. Now we toss two fair dice — with what probability do we get two threes?

The space of outcomes has 36 entries. We can write it as

$$\left(\begin{array}{cccccc} 11, & 12, & 13, & 14, & 15, & 16, \\ 21, & 22, & 23, & 24, & 25, & 26, \\ 31, & 32, & 33, & 34, & 35, & 36, \\ 41, & 42, & 43, & 44, & 45, & 46, \\ 51, & 52, & 53, & 54, & 55, & 56, \\ 61, & 62, & 63, & 64, & 65, & 66 \end{array} \right).$$

and each of these outcomes has the same probability. So the probability of two threes is $\frac{1}{36}$. The probability of getting a 2 and a 3 is $\frac{1}{18}$ — because there are two outcomes that yield this (23 and 32), and each has probability $\frac{1}{36}$.

Worked example 1.6 *Find the Lady*

Assume that the card that is chosen is chosen fairly — that is, each card is chosen with the same probability. What is the probability of turning up a Queen?

Solution: There are three outcomes, and each is chosen with the same probability, so the probability is $1/3$.

Worked example 1.7 *Find the Lady, twice*

Assume that the card that is chosen is chosen fairly — that is, each card is chosen with the same probability. What is the probability of turning up a Queen and then a Queen again?

Solution: Each outcome has the same probability, so $1/9$.

Worked example 1.8 *Children*

A couple decides to have two children. Genders are assigned to children at random, fairly, and at birth (our models have to abstract a little!). What is the probability of having a boy and then a girl?

Solution: The outcomes are $\{BB, BG, GB, GG\}$, and each has the same probability; so the probability we want is $1/4$. Notice that the order matters here; if we wanted to know the probability of having one of each gender, the answer would be different.

Worked example 1.9 *Monty Hall, indistinguishable goats, again*

Each outcome has the same probability. We choose to open the first box. With what probability will we find a goat (any goat)?

Solution: $2/3$

Worked example 1.10 *Monty Hall, yet again*

Each outcome has the same probability. We choose to open the first box. With what probability will we find the car?

Solution: $1/3$

Worked example 1.11 *Monty Hall, with distinct goats, again*

Each outcome has the same probability. We choose to open the first box. With what probability will we find a female goat?

Solution: $1/3$. The point of this example is that the sample space matters. If you care about the gender of the goat, then it's important to keep track of it; if you don't, it's probably a good idea to omit it from the sample space.

Outcomes represent all potential individual results of an experiment that we can or want to distinguish. This is quite important. For example, when we flip a coin, we could be interested if it lands on a spot that a fly landed on 10 minutes ago — this result isn't represented by our "heads" or "tails" model, and we would have to come up with an space of outcomes that does represent it. So outcomes represent the results we (a) care about and (b) can identify.

1.1.2 Events

Assume we run an experiment and get an outcome. We know what the outcome is (that's the whole point of a sample space). This means that we can tell whether the outcome we get belongs to some particular known set of outcomes. We just look in the set and see if our outcome is there. This means that *sets* of outcomes must also have a probability.

An **event** is a set of outcomes. In principle, there could be no outcome, although this is not interesting. This means that the empty set, which we write

\emptyset , is an event. The set of all outcomes, which we wrote Ω , must also be an event (although again it is not particularly interesting). **Notation:** We will write $\Omega - \mathcal{U}$ as \mathcal{U}^c ; read “the complement of \mathcal{U} ”.

There are some important logical properties of events.

- If \mathcal{U} and \mathcal{V} are events — sets of outcomes — then so is $\mathcal{U} \cap \mathcal{V}$. You should interpret this as the event that we have an outcome that is in \mathcal{U} and also in \mathcal{V} .
- If \mathcal{U} and \mathcal{V} are events, then $\mathcal{U} \cup \mathcal{V}$ is also an event. You should interpret this as the event that we have an outcome that is either in \mathcal{U} or in \mathcal{V} (or in both).
- If \mathcal{U} is an event, then $\mathcal{U}^c = \Omega - \mathcal{U}$ is also an event. You should think of this as the event we get an outcome that is not in \mathcal{U} .

This means that the set of all possible events Σ has a very important structure.

- \emptyset is in Σ .
- Ω is in Σ .
- If $\mathcal{U} \in \Sigma$ and $\mathcal{V} \in \Sigma$ then $\mathcal{U} \cup \mathcal{V} \in \Sigma$.
- If $\mathcal{U} \in \Sigma$ and $\mathcal{V} \in \Sigma$ then $\mathcal{U} \cap \mathcal{V} \in \Sigma$.
- If $\mathcal{U} \in \Sigma$ then $\mathcal{U}^c \in \Sigma$.

This means that the space of events can be quite big. For a single flip of a coin, it looks like

$$\{\emptyset, \{H\}, \{T\}, \{H, T\}\}$$

For a single throw of the die, the set of events is

$$\left\{ \begin{array}{cccccc} \emptyset, & \{1, 2, 3, 4, 5, 6\}, & & & & \\ \{1\}, & \{2\}, & \{3\}, & \{4\}, & \{5\}, & \{6\}, \\ & \{1, 2\}, & \{1, 3\}, & \{1, 4\}, & \{1, 5\}, & \{1, 6\}, \\ & & \{2, 3\}, & \{2, 4\}, & \{2, 5\}, & \{2, 6\}, \\ & & & \{3, 4\}, & \{3, 5\}, & \{3, 6\}, \\ & & & & \{4, 5\}, & \{4, 6\}, \\ & & & & & \{5, 6\}, \\ \\ & \{1, 2, 3\}, & \{1, 2, 4\}, & \{1, 2, 5\}, & \{1, 2, 6\}, & \\ & & \{1, 3, 4\}, & \{1, 3, 5\}, & \{1, 3, 6\}, & \\ & & & \{1, 4, 5\}, & \{1, 4, 6\}, & \\ & & & & \{1, 5, 6\}, & \\ & & & & \{2, 3, 4\}, & \{2, 3, 5\}, & \{2, 3, 6\}, \\ & & & & \{2, 3, 4\}, & \{2, 3, 5\}, & \{2, 3, 6\}, \\ & & & & \{2, 4, 5\}, & \{2, 4, 6\}, & \{2, 5, 6\}, \\ & & & & & \{3, 4, 5\}, & \{3, 4, 6\}, \\ & & & & & & \{3, 5, 6\}, \\ & & & & & & \{4, 5, 6\}, \\ \\ & & \{1, 2, 3, 4\}, & \{1, 2, 3, 5\}, & \{1, 2, 3, 6\}, & \\ & & & \{1, 3, 4, 5\}, & \{1, 3, 4, 6\}, & \\ & & & \{2, 3, 4, 5\}, & \{2, 3, 4, 6\}, & \\ & & & & \{3, 4, 5, 6\}, & \\ \{2, 3, 4, 5, 6\}, & \{1, 3, 4, 5, 6\}, & \{1, 2, 4, 5, 6\}, & \{1, 2, 3, 5, 6\}, & \{1, 2, 3, 4, 6\}, & \{1, 2, 3, 4, 5\}, \end{array} \right\}$$

(which gives some explanation as to why we don't usually write out the whole thing).

1.1.3 The Probability of Events

So far, we have described the probability of each outcome with a non-negative number. This number represents the relative frequency of the outcome. Straightforward reasoning allows us to extend this function to events. The probability of an event is a non-negative number; alternatively, we define a function taking events to the non-negative numbers. We require

- **The probability of every event is non-negative**, which we write $P(\mathcal{A}) \geq 0$ for all \mathcal{A} in the collection of events.
- **There are no missing outcomes**, which we write $P(\Omega) = 1$.
- **The probability of disjoint outcomes is additive**, which requires more notation. Assume that we have a collection of outcomes \mathcal{A}_i , indexed by i . We require that these have the property $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$ when $i \neq j$. This means that there is no outcome that appears in more than one \mathcal{A}_i . In turn, if we interpret probability as relative frequency, we must have that $P(\cup_i \mathcal{A}_i) = \sum_i P(\mathcal{A}_i)$.

Any function P taking events to numbers that has these properties is a probability. These very simple properties imply a series of other very important properties.

Useful facts: *The probability of events*

- $P(\mathcal{A}^c) = 1 - P(\mathcal{A})$
- $P(\emptyset) = 0$
- $P(\mathcal{A} - \mathcal{B}) = P(\mathcal{A}) - P(\mathcal{A} \cap \mathcal{B})$
- $P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \cap \mathcal{B})$
- $P(\cup_1^n \mathcal{A}_i) = \sum_i P(\mathcal{A}_i) - \sum_{i < j} P(\mathcal{A}_i \cap \mathcal{A}_j) + \sum_{i < j < k} P(\mathcal{A}_i \cap \mathcal{A}_j \cap \mathcal{A}_k) + \dots (-1)^{(n+1)} P(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \dots \cap \mathcal{A}_n)$

Proofs: *The probability of events*

$P(\mathcal{A}^c) = 1 - P(\mathcal{A})$ because \mathcal{A}^c and \mathcal{A} are disjoint, so that $P(\mathcal{A}^c \cup \mathcal{A}) = P(\mathcal{A}^c) + P(\mathcal{A}) = P(\Omega) = 1$.

$P(\emptyset) = 0$ because $P(\emptyset) = P(\Omega^c) = P(\Omega - \Omega) = 1 - P(\Omega) = 1 - 1 = 0$.

$P(\mathcal{A} - \mathcal{B}) = P(\mathcal{A}) - P(\mathcal{A} \cap \mathcal{B})$ because $\mathcal{A} - \mathcal{B}$ is disjoint from $\mathcal{A} \cap \mathcal{B}$, and $(\mathcal{A} - \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{B}) = \mathcal{A}$. This means that $P(\mathcal{A} - \mathcal{B}) + P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})$.

$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \cap \mathcal{B})$ because $P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A} \cup (\mathcal{B} \cap \mathcal{A}^c)) = P(\mathcal{A}) + P((\mathcal{B} \cap \mathcal{A}^c))$. Now $\mathcal{B} = (\mathcal{B} \cap \mathcal{A}) \cup (\mathcal{B} \cap \mathcal{A}^c)$. Furthermore, $(\mathcal{B} \cap \mathcal{A})$ is disjoint from $(\mathcal{B} \cap \mathcal{A}^c)$, so we have $P(\mathcal{B}) = P((\mathcal{B} \cap \mathcal{A})) + P((\mathcal{B} \cap \mathcal{A}^c))$. This means that $P(\mathcal{A}) + P((\mathcal{B} \cap \mathcal{A}^c)) = P(\mathcal{A}) + P(\mathcal{B}) - P((\mathcal{B} \cap \mathcal{A}))$.

$P(\cup_{i=1}^n \mathcal{A}_i) = \sum_i P(\mathcal{A}_i) - \sum_{i < j} P(\mathcal{A}_i \cap \mathcal{A}_j) + \sum_{i < j < k} P(\mathcal{A}_i \cap \mathcal{A}_j \cap \mathcal{A}_k) + \dots (-1)^{(n+1)} P(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \dots \cap \mathcal{A}_n)$ can be proven by repeated application of the previous result. As an example, we show how to work the case where there are three sets (you can get the rest by induction). $P(\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3) = P(\mathcal{A}_1 \cup (\mathcal{A}_2 \cup \mathcal{A}_3)) = P(\mathcal{A}_1) + P(\mathcal{A}_2 \cup \mathcal{A}_3) - P(\mathcal{A}_1 \cap (\mathcal{A}_2 \cup \mathcal{A}_3)) = P(\mathcal{A}_1) + (P(\mathcal{A}_2) + P(\mathcal{A}_3) - P(\mathcal{A}_2 \cap \mathcal{A}_3)) - P((\mathcal{A}_1 \cap \mathcal{A}_2) \cup (\mathcal{A}_1 \cap \mathcal{A}_3)) = P(\mathcal{A}_1) + (P(\mathcal{A}_2) + P(\mathcal{A}_3) - P(\mathcal{A}_2 \cap \mathcal{A}_3)) - P(\mathcal{A}_1 \cap \mathcal{A}_2) - P(\mathcal{A}_1 \cap \mathcal{A}_3) - (-P((\mathcal{A}_1 \cap \mathcal{A}_2) \cap (\mathcal{A}_1 \cap \mathcal{A}_3))) = P(\mathcal{A}_1) + (P(\mathcal{A}_2) + P(\mathcal{A}_3) - P(\mathcal{A}_2 \cap \mathcal{A}_3)) - P(\mathcal{A}_1 \cap \mathcal{A}_2) - P(\mathcal{A}_1 \cap \mathcal{A}_3) + P(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3)$

Looking at the useful facts should suggest a helpful analogy between the probability of an event and the “size” of the event. I find this a good way to remember equations. For example, $P(\mathcal{A} - \mathcal{B}) = P(\mathcal{A}) - P(\mathcal{A} \cap \mathcal{B})$ is easily captured — the “size” of the part of \mathcal{A} that isn’t \mathcal{B} is obtained by taking the “size” of \mathcal{A} and subtracting the size of the part that is also in \mathcal{B} . Similarly, $P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \cap \mathcal{B})$ says — you can get the “size” of $\mathcal{A} \cup \mathcal{B}$ by adding the two “sizes”, then subtracting the size of the intersection because otherwise you would count these terms twice. Some people find Venn diagrams a useful way to keep track of this argument, and Figure 1.1 is for them.

Worked example 1.12 *Odd numbers with fair dice*

We throw a fair (each number has the same probability) die twice, then add the two numbers. What is the probability of getting an odd number?

Solution: There are 36 outcomes, listed above. Each has the same probability (1/36). 18 of them give an odd number, and the other 18 give an even number. They are disjoint, so the probability is $18/36 = 1/2$

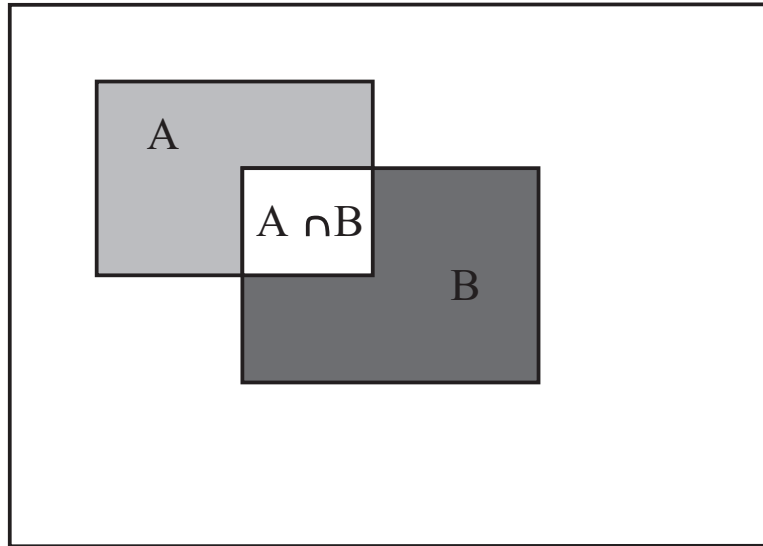


FIGURE 1.1: If you think of the probability of an event as measuring its size, many of the rules are quite straightforward to remember. Venn diagrams can sometimes help. For example, you can see that $P(\mathcal{A} - \mathcal{B}) = P(\mathcal{A}) - P(\mathcal{A} \cap \mathcal{B})$ by noticing that $P(\mathcal{A} - \mathcal{B})$ is the “size” of the part of \mathcal{A} that isn’t \mathcal{B} . This is obtained by taking the “size” of \mathcal{A} and subtracting the size of the part that is also in \mathcal{B} , i.e. the “size” of $\mathcal{A} \cap \mathcal{B}$. Similarly, you can see that $P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \cap \mathcal{B})$ by noticing that you can get the “size” of $\mathcal{A} \cup \mathcal{B}$ by adding the “sizes” of \mathcal{A} and \mathcal{B} , then subtracting the size of the intersection to avoid double counting.

Worked example 1.13 *Numbers divisible by five with fair dice*

We throw a fair (each number has the same probability) die twice, then add the two numbers. What is the probability of getting a number divisible by five?

Solution: There are 36 outcomes, listed above. Each has the same probability ($1/36$). For this event, the spots must add to either 5 or to 10. There are 4 ways to get 5. There are 3 ways to get 10. These outcomes are disjoint. So the probability is $7/36$.

Worked example 1.14 *Children*

This example is a version of of example 1.12, p44, Stirzaker, “Elementary Probability”.

A couple decides to have children. They discuss the following three strategies:

- have three children;
- have children until the first girl, or until there are three, then stop;
- have children until there is one of each gender, or until there are three, then stop.

Assume that each gender is equally likely at each birth. Let G_i be the event that there are i girls, and C be the event there are more girls than boys. Compute $P(B_1)$ and $P(C)$ in each case.

Solution: Case 1: There are eight outcomes. Each has the same probability. Three of them have a single boy, so $P(B_1) = 3/8$. $P(C) = P(C^c)$ (because C^c is the event there are more boys than than girls, AND the number of children is odd), so that $P(C) = 1/2$; you can also get this by counting outcomes.

Case 2: In this case, the outcomes are $\{G, BG, BBG\}$, but if we think about them like this, we have no simple way to compute their probability. Instead, we could use the sample space from the previous answer, but assume that some of the later births are fictitious. So the outcome G corresponds to the event $\{GBB, GBG, GGB, GGG\}$ (and so has probability $1/2$); the outcome BG corresponds to the event $\{BGB, BGG\}$ (and so has probability $1/4$); the outcome BBG corresponds to the event BBG (and so has probability $1/8$). This means that $P(B_1) = 1/4$ and $P(C) = 1/2$.

Case 3: The outcomes are $\{GB, BG, GGB, GGG, BBG, BBB\}$. Again, if we think about them like this, we have no simple way to compute their probability; so we use the sample space from the previous example with device of the fictitious births again. Then GB corresponds to the event $\{GBB, GBG\}$; BG corresponds to the event $\{BGB, BGG\}$; GGB corresponds to the event $\{GGB\}$; GGG corresponds to the event $\{GGG\}$; BBG corresponds to the event $\{BBG\}$; and BBB corresponds to the event $\{BBB\}$. Like this, we get $P(B_1) = 5/8$ and $P(C) = 1/4$.

Many probability problems are basically advanced counting exercises. One form of these problems occurs where all outcomes have the same probability. You have to determine the probability of an event that consists of some set of outcomes, and you can do that by computing

$$\frac{\text{Number of outcomes in the event}}{\text{Total number of outcomes}}$$

For example, what is the probability that three people are born on three days of the week in succession (for example, Monday-Tuesday-Wednesday; or Saturday-Sunday-Monday; and so on). We assume that the first person has no effect on the second, and that births are equally common on each day of the week. In this case, the space of outcomes consists of triples of days; the event we are interested in is a

triple of three days in succession; and each outcome has the same probability. So the event is the set of triples of three days in succession (which has seven elements, one for each starting day). The space of outcomes has 7^3 elements in it, so the probability is

$$\frac{\text{Number of outcomes in the event}}{\text{Total number of outcomes}} = \frac{7}{7^3} = \frac{1}{49}.$$

As a (very slightly) more interesting example, what is the probability that two people are born on the same day of the week? We can solve this problem by computing

$$\frac{\text{Number of outcomes in the event}}{\text{Total number of outcomes}} = \frac{7}{7 \times 7} = \frac{1}{7}.$$

An important feature of this class of problem is that your intuition can be quite misleading. This is because, although each outcome can have very small probability, the number of events can be big. For example, what is the probability that, in a room of 30 people, there is a pair of people who have the same birthday? We simplify, and assume that each year has 365 days, and that none of them are special (i.e. each day has the same probability of being chosen as a birthday).

The easy way to attack this question is to notice that our probability, $P(\{\text{shared birthday}\})$, is

$$1 - P(\{\text{all birthdays different}\}).$$

This second probability is rather easy to estimate. Each outcome in the sample space is a list of 30 days (one birthday per person). Each outcome has the same probability. So

$$P(\{\text{all birthdays different}\}) = \frac{\text{Number of outcomes in the event}}{\text{Total number of outcomes}}.$$

The total number of outcomes is easily seen to be 365^{30} , which is the total number of possible lists of 30 days. The number of outcomes in the event is the number of lists of 30 days, all different. To count these, we notice that there are 365 choices for the first day; 364 for the second; and so on. So we have

$$P(\{\text{shared birthday}\}) = 1 - \frac{365 \times 364 \times \dots \times 336}{365^{30}} = 1 - 0.2937 = 0.7063$$

which means there's really a pretty good chance that two people in a room of 30 share a birthday. There is a wide variety of problems like this; if you're so inclined, you can make a small but quite reliable profit off people's inability to estimate probabilities for this kind of problem correctly.

If we change the birthday example slightly, the problem changes drastically. If you stand up and bet that two people in the room have the same birthday, you have a probability of winning of about 0.71; but if you bet that there is someone else in the room who has the same birthday that you do, your probability of winning is $29/365$, a very much smaller number.

These combinatorial arguments can get pretty elaborate. For example, you throw 3 fair 20-sided dice. What is the probability that the sum of the faces is 14? Fairly clearly, the answer is

$$\frac{\text{The number of triples that add to 14}}{20^3}$$

but one needs to determine the number of triples that add to 14.

1.1.4 The Gambler's Ruin

Assume you bet \$1 a tossed coin will come up heads. If you win, you get \$1 and your original stake back. If you lose, you lose your stake. But this coin has the property that $P(H) = p < 1/2$. We will study what happens when you bet repeatedly.

Assume you have \$ s when you start. You will keep betting until either (a) you have \$0 (you can't borrow money) or (b) the amount of money you have accumulated is \$ j (where $j > s$ or there is nothing to do). The coin tosses are independent. We will compute p_s , the probability that you leave the table with nothing, when you start with \$ s .

Assume that you win the first bet. Then you have \$ $s + 1$, so your probability of leaving the table with nothing now becomes p_{s+1} . If you lose the first bet, then you have \$ $s - 1$, so your probability of leaving the table with nothing now becomes p_{s-1} . The coin tosses are independent, so we can write

$$p_s = pp_{s+1} + (1-p)p_{s-1}.$$

Now we also know that $p_0 = 1$ and $p_j = 0$. We need to obtain an expression for p_s . We can rearrange to get

$$p_{s+1} - p_s = \frac{(1-p)}{p} (p_s - p_{s-1})$$

(check this expression by expanding it out and comparing). Now this means that

$$p_{s+1} - p_s = \left(\frac{(1-p)}{p} \right)^2 (p_{s-1} - p_{s-2})$$

so that

$$\begin{aligned} p_{s+1} - p_s &= \left(\frac{(1-p)}{p} \right)^s (p_1 - p_0) \\ &= \left(\frac{(1-p)}{p} \right)^s (p_1 - 1). \end{aligned}$$

Now we need a simple result about series. Assume I have a series u_k , $k \geq 0$, with the property that

$$u_k - u_{k-1} = cr^{k-1}.$$

Then I can expand this expression to get

$$\begin{aligned} u_k - u_0 &= (u_k - u_{k-1}) + (u_{k-1} - u_{k-2}) + \dots + (u_1 - u_0) \\ &= c(r^{k-1} + r^{k-2} + \dots + 1) \\ &= c \left(\frac{r^k - 1}{r - 1} \right). \end{aligned}$$

If we plug our series into this result, we get

$$p_{s+1} - 1 = (p_1 - 1) \left(\frac{\left(\frac{(1-p)}{p} \right)^{s+1} - 1}{\left(\frac{(1-p)}{p} \right) - 1} \right)$$

so reindexing gives us

$$p_s - 1 = (p_1 - 1) \left(\frac{\left(\frac{1-p}{p}\right)^s - 1}{\left(\frac{1-p}{p}\right) - 1} \right)$$

Now we also know that $p_j = 0$, so we have

$$\begin{aligned} p_j &= 0 \\ &= 1 + (p_1 - 1) \left(\frac{\left(\frac{1-p}{p}\right)^j - 1}{\left(\frac{1-p}{p}\right) - 1} \right) \end{aligned}$$

meaning that

$$(p_1 - 1) = \frac{-1}{\left(\frac{\left(\frac{1-p}{p}\right)^j - 1}{\left(\frac{1-p}{p}\right) - 1}\right)}.$$

Inserting this and rearranging gives

$$p_s = \frac{\left(\frac{1-p}{p}\right)^j - \left(\frac{1-p}{p}\right)^s}{\left(\frac{1-p}{p}\right)^j - 1}.$$

This expression is quite informative. Notice that, if $p < 1/2$, then $(1-p)/p > 1$. This means that as $j \rightarrow \infty$, we have $p_s \rightarrow 1$.

1.2 CONDITIONAL PROBABILITY

If you throw a fair die twice and add the numbers, then the probability of getting a number less than six is $\frac{10}{36}$. Now imagine you know that the first die came up three. In this case, the probability that the sum will be less than six is $\frac{1}{3}$, which is slightly larger. If the first die came up four, then the probability the sum will be less than six is $\frac{1}{6}$, which is rather less than $\frac{10}{36}$. If the first die came up one, then the probability that the sum is less than six becomes $\frac{2}{3}$, which is much larger.

Each of these probabilities is an example of a **conditional probability**. We assume we have a space of outcomes and a collection of events. The conditional probability of \mathcal{B} , conditioned on \mathcal{A} , is the probability that \mathcal{B} occurs given that \mathcal{A} has definitely occurred. We write this as

$$P(\mathcal{B}|\mathcal{A})$$

One way to get an expression for $P(\mathcal{B}|\mathcal{A})$ is to notice that, because \mathcal{A} is known to have occurred, our space of outcomes or sample space is now reduced to \mathcal{A} . We know that our outcome lies in \mathcal{A} ; $P(\mathcal{B}|\mathcal{A})$ is the probability that it also lies in $\mathcal{B} \cap \mathcal{A}$.

The outcome lies in \mathcal{A} , and so it must lie in either $P(\mathcal{B} \cap \mathcal{A})$ or in $P(\mathcal{B}^c \cap \mathcal{A})$. This means that

$$P(\mathcal{B}|\mathcal{A}) + P(\mathcal{B}^c|\mathcal{A}) = 1.$$

Now recall the idea of probabilities as relative frequencies. If $P(\mathcal{C} \cap \mathcal{A}) = kP(\mathcal{B} \cap \mathcal{A})$, this means that we will see outcomes in $\mathcal{C} \cap \mathcal{A}$ about k times as often as we will see outcomes in $\mathcal{B} \cap \mathcal{A}$. But this must apply even if we know that the outcome is in \mathcal{A} . So we must have

$$P(\mathcal{B}|\mathcal{A}) \propto P(\mathcal{B} \cap \mathcal{A}).$$

Now we need to determine the constant of proportionality; write c for this constant, meaning

$$P(\mathcal{B}|\mathcal{A}) = cP(\mathcal{B} \cap \mathcal{A}).$$

We have that

$$cP(\mathcal{B} \cap \mathcal{A}) + cP(\mathcal{B}^c \cap \mathcal{A}) = cP(\mathcal{A}) = P(\mathcal{B}|\mathcal{A}) + P(\mathcal{B}^c|\mathcal{A}) = 1,$$

so that

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

Another, very useful, way to write this expression is

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

Now, since $\mathcal{B} \cap \mathcal{A} = \mathcal{A} \cap \mathcal{B}$, we must have that

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

Worked example 1.15 *Two dice*

We throw two fair dice.

- What is the probability that the sum of spots is greater than 6?
- Now we know that the first die comes up five. What is the conditional probability that the sum of spots on both dice is greater than six, conditioned on the event that the first die comes up five?

Solution:

- There are 36 outcomes, but quite a lot of ways to get a number greater than six. Recall $P(\mathcal{A}^c) = 1 - P(\mathcal{A})$. Write the event that sum is greater than six as \mathcal{S} . There are 15 ways to get a number less than or equal to six, so $P(\mathcal{S}^c) = 15/36$, which means $P(\mathcal{S}) = 21/36$.
- Write the event that the first die comes up 5 as \mathcal{F} . There are five outcomes where the first die comes up 5 and the number is greater than 6, so $P(\mathcal{F} \cap \mathcal{S}) = 5/36$. $P(\mathcal{S}|\mathcal{F}) = P(\mathcal{F} \cap \mathcal{S})/P(\mathcal{F}) = (5/36)/(1/6) = 5/6$.

Notice that $\mathcal{A} \cap \mathcal{B}$ and $\mathcal{A} \cap \mathcal{B}^c$ are disjoint sets, and that $\mathcal{A} = (\mathcal{A} \cap \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{B}^c)$. So we have

$$\begin{aligned} P(\mathcal{A}) &= P(\mathcal{A} \cap \mathcal{B}) + P(\mathcal{A} \cap \mathcal{B}^c) \\ &= P(\mathcal{A}|\mathcal{B})P(\mathcal{B}) + P(\mathcal{A}|\mathcal{B}^c)P(\mathcal{B}^c), \end{aligned}$$

a tremendously important and useful fact. Another version of this fact is also very useful. Assume we have a set of disjoint sets \mathcal{B}_i . These sets must have the property that (a) $\mathcal{B}_i \cap \mathcal{B}_j = \emptyset$ for $i \neq j$ and (b) they cover \mathcal{A} , meaning that $\mathcal{A} \cap (\cup_i \mathcal{B}_i) = \mathcal{A}$. Then we have

$$\begin{aligned} P(\mathcal{A}) &= \sum_i P(\mathcal{A} \cap \mathcal{B}_i) \\ &= \sum_i P(\mathcal{A}|\mathcal{B}_i)P(\mathcal{B}_i) \end{aligned}$$

Worked example 1.16 *Car factories*

There are two car factories, A and B . Each year, factory A produces 1000 cars, of which 10 are lemons. Factory B produces 2 cars, each of which is a lemon. All cars go to a single lot, where they are thoroughly mixed up. I buy a car.

- What is the probability it is a lemon?
- What is the probability it came from factory B ?
- The car is now revealed to be a lemon. What is the probability it came from factory B , conditioned on the fact it is a lemon?

Solution:

- Write the event the car is a lemon as \mathcal{L} . There are 1002 cars, of which 12 are lemons. The probability that I select any given car is the same, so we have $12/1002$.
- Same argument yields $2/1002$.
- Write \mathcal{B} for the event the car comes from factory B . I need $P(\mathcal{B}|\mathcal{L})$. This is $P(\mathcal{L}|\mathcal{B})P(\mathcal{B})/P(\mathcal{L}) = (1 \times 2/1002)/(12/1002) = 1/6$.

Worked example 1.17 *Royal flushes in poker - 1*

This exercise is after Stirzaker, p. 51.

You are playing a straightforward version of poker, where you are dealt five cards face down. A royal flush is a hand of AKQJ10 all in one suit. What is the probability that you are dealt a royal flush?

Solution: This is

$$\frac{\text{number of hands that are royal flushes, ignoring card order}}{\text{total number of different five card hands, ignoring card order}}.$$

There are four hands that are royal flushes (one for each suit). Now the total number of five card hands is

$$\binom{52}{5} = 2598960$$

so we have

$$\frac{4}{2598960} = \frac{1}{649740}.$$

Worked example 1.18 *Royal flushes in poker - 2*

This exercise is after Stirzaker, p. 51.

You are playing a straightforward version of poker, where you are dealt five cards face down. A royal flush is a hand of AKQJ10 all in one suit. The fifth card that you are dealt lands face up. It is the nine of spades. What now is the probability that you have been dealt a royal flush? (i.e. what is the conditional probability of getting a royal flush, conditioned on the event that one card is the nine of spades)

Solution: No hand containing a nine of spades is a royal flush, so this is easily zero.

Worked example 1.19 *Royal flushes in poker - 3*

This exercise is after Stirzaker, p. 51.

You are playing a straightforward version of poker, where you are dealt five cards face down. A royal flush is a hand of AKQJ10 all in one suit. The fifth card that you are dealt lands face up. It is the Ace of spades. What now is the probability that you have been dealt a royal flush? (i.e. what is the conditional probability of getting a royal flush, conditioned on the event that one card is the Ace of spades)

Solution: There are two ways to do this. The easiest is to notice this is the probability that the other four cards are KQJ10 of spades, which is

$$\binom{51}{4}^{-1} = \frac{1}{249900}.$$

Harder is to consider the events

\mathcal{A} = event that you receive a royal flush and last card is the ace of spades

and

\mathcal{B} = event that the last card you receive is the ace of spades,

and the expression

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

Now $P(\mathcal{A}) = \frac{1}{52}$. $P(\mathcal{A} \cap \mathcal{B})$ is given by

$$\frac{\text{number of five card royal flushes where card five is Ace of spades}}{\text{total number of different five card hands}}.$$

where we DO NOT ignore card order. This is

$$\frac{4 \times 3 \times 2 \times 1}{52 \times 51 \times 50 \times 49 \times 48}$$

yielding

$$P(\mathcal{A}|\mathcal{B}) = \frac{1}{249900}.$$

Notice the interesting part: the conditional probability is rather larger than the probability. If you see this ace, the conditional probability is $\frac{13}{5}$ times the probability that you will get a flush if you don't. Seeing this card has really made a difference.

Worked example 1.20 *False positives*

After Stirzaker, p55.

You have a blood test for a rare disease that occurs by chance in 1 person in 100,000. If you have the disease, the test will report that you do with probability 0.95 (and that you do not with probability 0.05). If you do not have the disease, the test will report a false positive with probability $1e-3$. If the test says you do have the disease, what is the probability it is correct?

Solution: Write S for the event you are sick and R for the event the test reports you are sick. We need $P(S|R)$.

$$\begin{aligned}
 P(S|R) &= \frac{P(R|S)P(S)}{P(R)} \\
 &= \frac{P(R|S)P(S)}{P(R|S)P(S) + P(R|S^c)P(S^c)} \\
 &= \frac{0.95 \times 1e-5}{0.95 \times 1e-5 + 1e-3 \times (1 - 1e-5)} \\
 &= 0.0094
 \end{aligned}$$

which should strike you as being a bit alarming. The disease is so rare that the test is almost useless.

Worked example 1.21 *False positives -2*

After Stirzaker, p55.

You want to make a blood test for a rare disease that occurs by chance in 1 person in 100,000. If you have the disease, the test will report that you do with probability p (and that you do not with probability $(1-p)$). If you do not have the disease, the test will report a false positive with probability q . You want to choose the value of p so that if the test says you have the disease, there is at least a 50% probability that you do.

Solution: Write S for the event you are sick and R for the event the test reports you are sick. We need $P(S|R)$.

$$\begin{aligned} P(S|R) &= \frac{P(R|S)P(S)}{P(R)} \\ &= \frac{P(R|S)P(S)}{P(R|S)P(S) + P(R|S^c)P(S^c)} \\ &= \frac{p \times 1e - 5}{p \times 1e - 5 + q \times (1 - 1e - 5)} \\ &\geq 0.5 \end{aligned}$$

which means that $p \geq 99999q$ which should strike you as being very alarming indeed, because $p \leq 1$ and $q \geq 0$. One plausible pair of values is $q = 1e - 5$, $p = 1 - 1e - 5$. The test has to be spectacularly accurate to be of any use.

1.2.1 Independence

As we have seen, the conditional probability of an event \mathcal{A} conditioned on another event can be very different from the probability of that event. This is because knowing that one event has occurred may significantly reduce the available outcomes of an experiment, as in example 16, and in this example.

But this does not always happen. Two events are **independent** if

$$P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B})$$

If two events \mathcal{A} and \mathcal{B} are independent, then

$$P(\mathcal{A}|\mathcal{B}) = P(\mathcal{A})$$

and

$$P(\mathcal{B}|\mathcal{A}) = P(\mathcal{B})$$

If \mathcal{A} and \mathcal{B} are independent, knowing that one of the two has occurred tells us nothing useful about whether the other will occur. For example, if we are told event \mathcal{A} with $P(\mathcal{A}) > 0$ has occurred, the sample space is reduced from Ω to \mathcal{A} . The probability that \mathcal{B} will now occur is

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

which is $P(\mathcal{B})$ if the two are independent. Again, this means that knowing that \mathcal{A} occurred tells you nothing about \mathcal{B} — the probability that \mathcal{B} will occur is the same whether you know that \mathcal{A} occurred or not.

Some events are pretty obviously independent. On other occasions, one needs to think about whether they are independent or not. Sometimes, it is reasonable to choose to model events as being independent, even though they might not be exactly independent. In several examples below, we will work with the event that a person, selected fairly and randomly from a set of people in a room, has a birthday on a particular day of the year. We assume that, for different people, the events are independent. This seems like a fair assumption, but one might want to be cautious if you know that the people in the room are drawn from a population where multiple births are common.

Example: *Drawing two cards, without replacement*

We draw two playing cards from a deck of cards. Let \mathcal{A} be the event “the first card is a queen” and let \mathcal{B} be the event that “the second card is a queen”. Then

$$P(\mathcal{A}) = \frac{4}{52}$$

and

$$P(\mathcal{B}) = \frac{4}{52}$$

but

$$P(\mathcal{A} \cap \mathcal{B}) = \frac{4 \cdot 3}{52 \cdot 51}.$$

This means that $P(\mathcal{B}|\mathcal{A}) = 3/51$; if the first card is known to be a queen, then the second card is slightly less likely to be a queen than it would otherwise. The events \mathcal{A} and \mathcal{B} are not independent.

Example: *Drawing two cards, with replacement*

We draw one playing card from a deck of cards; we write down the identity of that card, replace it in the deck, shuffle the deck, then draw another card. Let \mathcal{A} be the event “the first card is a queen” and let \mathcal{B} be the event that “the second card is a queen”. Then

$$P(\mathcal{A}) = \frac{4}{52}$$

and

$$P(\mathcal{B}) = \frac{4}{52}.$$

We also have

$$P(\mathcal{A} \cap \mathcal{B}) = \frac{4 \cdot 4}{52 \cdot 52}.$$

This means that $P(\mathcal{B}|\mathcal{A}) = 4/52$; if the first card is known to be a queen, then we know nothing about the second card. The events \mathcal{A} and \mathcal{B} are independent.

You should compare examples 1.2.1 and 1.2.1. Simply replacing a card after

it has been drawn has made the events independent. This should make sense to you: if you draw a card from a deck and look at it, you know very slightly more about what the next card should be. For example, it won't be the same as the card you have. The deck is very slightly smaller than it was, too, and there are fewer cards of the suit and rank of the card you have. However, if you replace the card you drew, then shuffle the deck, seeing the first card tells you nothing about the second card.

Worked example 1.22 *Two fair coin flips*

We flip a fair coin twice. The outcomes are $\{HH, HT, TH, TT\}$. Each has the same probability. Show that the event \mathcal{H}_1 — where the first flip comes up heads — is independent from the event \mathcal{H}_2 — where the second flip comes up heads.

Solution: $\mathcal{H}_1 = \{HT, HH\}$ and $\mathcal{H}_2 = \{TH, HH\}$. Now $P(\mathcal{H}_1) = 1/2$. Also, $P(\mathcal{H}_2) = 1/2$. Now $P(\mathcal{H}_1 \cap \mathcal{H}_2) = 1/4$, so that $P(\mathcal{H}_2|\mathcal{H}_1) = P(\mathcal{H}_2)$.

Worked example 1.23 *Independent cards*

We draw one card from a standard deck of 52 cards. The event \mathcal{A} is “the card is a red suit” and the event \mathcal{B} is “the card is a 10”. Are they independent?

Solution: These are independent because $P(\mathcal{A}) = 1/2$, $P(\mathcal{B}) = 1/13$ and $P(\mathcal{A} \cap \mathcal{B}) = 2/52 = 1/26 = P(\mathcal{A})P(\mathcal{B})$

Worked example 1.24 *Independent cards*

We take a standard deck of cards, and remove the ten of hearts. We now draw two cards from this deck. The event \mathcal{A} is “the card is a red suit” and the event \mathcal{B} is “the card is a 10”. Are they independent?

Solution: These are not independent because $P(\mathcal{A}) = 25/51$, $P(\mathcal{B}) = 3/51$ and $P(\mathcal{A} \cap \mathcal{B}) = 1/51 \neq P(\mathcal{A})P(\mathcal{B}) = 75/(51^2)$

Events $\mathcal{A}_1 \dots \mathcal{A}_n$ are **pairwise independent** if each pair is independent (i.e. \mathcal{A}_1 and \mathcal{A}_2 are independent, etc.). They are **independent** if for any collection of distinct indices $i_1 \dots i_k$ we have

$$P(\mathcal{A}_{i_1} \cap \dots \cap \mathcal{A}_{i_k}) = P(\mathcal{A}_{i_1}) \dots P(\mathcal{A}_{i_k})$$

Notice that independence is a much stronger assumption than pairwise independence.

Worked example 1.25 *Cards and pairwise independence*

We draw three cards from a properly shuffled standard deck, with replacement and reshuffling (i.e., draw a card, make a note, return to deck, shuffle, draw the next, make a note, shuffle, draw the third). Let \mathcal{A} be the event that “card 1 and card 2 have the same suit”; let \mathcal{B} be the event that “card 2 and card 3 have the same suit”; let \mathcal{C} be the event that “card 1 and card 3 have the same suit”. Show these events are pairwise independent, but not independent.

Solution: By counting, you can check that $P(\mathcal{A}) = 1/4$; $P(\mathcal{B}) = 1/4$; and $P(\mathcal{A} \cap \mathcal{B}) = 1/16$, so that these two are independent. This argument works for other pairs, too. But $P(\mathcal{C} \cap \mathcal{A} \cap \mathcal{B}) = 1/16$ which is not $1/4^3$, so the events are not independent; this is because the third event is logically implied by the first two.

We usually do not have the information required to prove that events are independent. Instead, we use intuition (for example, two flips of the same coin are likely to be independent unless there is something very funny going on) or simply choose to apply models in which some variables are independent.

Independent events can lead very quickly to very small probabilities. This can mislead intuition quite badly. For example, imagine I search a DNA database with a sample. I can show that there is a probability of a chance match of $1e-4$. There are 20,000 people in the database. Chance matches are independent. What is the probability I get at least one match, purely by chance? This is $1 - P(\text{no matches})$. But $P(\text{no matches})$ is much smaller than you think. It is $(1 - 1e-4)^{20,000}$, so the probability is about 86% that you get at least one match by chance. Notice that if the database gets bigger, the probability grows; so at 40,000 the probability of one match by chance is 98%.

People quite often reason poorly about independent events. The most common problem is known as the **gambler’s fallacy**. This occurs when you reason that the probability of an independent event has been changed by previous outcomes. For example, imagine I toss a coin that is known to be fair 20 times and get 20 heads. The probability that the next toss will result in a head has not changed at all — it is still 0.5 — but many people will believe that it has changed. This idea is also sometimes referred to as **antichance**.

It might in fact be sensible to behave as if you’re committing some version of the gambler’s fallacy in real life, because you hardly ever know for sure that your model is right. So in the coin tossing example, if the coin wasn’t known to be fair, it might be reasonable to assume that it has been weighted in some way, and so to believe that the more heads you see, the more likely you will see a head in the next toss. At time of writing, Wikipedia has some fascinating stories about the gambler’s fallacy; apparently, in 1913, a roulette wheel in Monte Carlo produced black 26 times in a row, and gamblers lost an immense amount of money betting on red. Here the gambler’s reasoning seems to have been that the universe should ensure that probabilities produce the right frequencies in the end, and so will adjust the outcome of the next spin of the wheel to balance the sums. This is an instance of the gambler’s fallacy. However, the page also contains the story of one Joseph Jagger, who hired people to keep records of the roulette wheels, and notice that one wheel favored some numbers (presumably because of some problem with balance).

He won a lot of money, until the casino started more careful maintenance on the wheels. This isn't the gambler's fallacy; instead, he noticed that the numbers implied that the wheel was not a fair randomizer. He made money because the casino's odds on the bet assumed that it was fair.

1.2.2 The Monty Hall Problem, and other Perils of Conditional Probability

Careless thinking about probability, particularly conditional probability, can cause wonderful confusion. The Monty Hall problem is a good example. The problem works like this: There are three doors. Behind one is a car. Behind each of the others is a goat. The car and goats are placed randomly and fairly, so that the probability that there is a car behind each door is the same. You will get the object that lies behind the door you choose at the end of the game. For reasons of your own, you would prefer the car to the goat.

The game goes as follows. You select a door. The host then opens a door and shows you a goat. You must now choose to either keep your door, or switch to the other door. What should you do?

You cannot tell what to do, by the following argument. Label the door you chose at the start of the game 1; label the other doors 2 and 3. Write C_i for the event that the car lies behind door i . Write G_m for the event that a goat is revealed behind door m , where m is the number of the door where the goat was revealed (which could be 1, 2, or 3). You need to know $P(C_1|G_m)$. But

$$P(C_1|G_m) = \frac{P(G_m|C_1)P(C_1)}{P(G_m|C_1)P(C_1) + P(G_m|C_2)P(C_2) + P(G_m|C_3)P(C_3)}$$

and you do not know $P(G_m|C_1)$, $P(G_m|C_2)$, $P(G_m|C_3)$, because you don't know the rule by which the host chooses which door to open to reveal a goat. Different rules lead to quite different analyses.

There are several possible rules for the host to show a goat:

- **Rule 1:** choose a door uniformly at random.
- **Rule 2:** choose from the doors with goats behind them *that are not door 1* uniformly and at random.
- **Rule 3:** if the car is at 1, then choose 2; if at 2, choose 3; if at 3, choose 1.
- **Rule 4:** choose from the doors with goats behind them uniformly and at random.

We should keep track of the rules in the conditioning, so we write $P(G_m|C_1, r_1)$ for the conditional probability that a goat was revealed behind door m when the car is behind door 1, using rule 1 (and so on).

Under rule 1, we can write

$$P(C_1|G_m, r_1) = \frac{P(G_m|C_1, r_1)P(C_1)}{P(G_m|C_1, r_1)P(C_1) + P(G_m|C_2, r_1)P(C_2) + P(G_m|C_3, r_1)P(C_3)}$$

When m is 2 or 3 we get

$$\begin{aligned} P(C_1|G_m, r_1) &= \frac{P(G_m|C_1, r_1)P(C_1)}{P(G_m|C_1, r_1)P(C_1) + P(G_m|C_2, r_1)P(C_2) + P(G_m|C_3, r_1)P(C_3)} \\ &= \frac{(1/3)(1/3)}{0(1/3) + (1/3)(1/3) + (1/3)(1/3)} \\ &= (1/2) \end{aligned}$$

but when m is 1, $P(C_1|G_m, r_1) = 0$ because there can't be both a goat and a car behind door 1. Notice that this means the host showing us a goat hasn't revealed anything about where the car is (it could be behind 1 or behind the other closed door).

Under rule 2, we can write

$$P(C_1|G_m, r_2) = \frac{P(G_m|C_1, r_2)P(C_1)}{P(G_m|C_1, r_2)P(C_1) + P(G_m|C_2, r_2)P(C_2) + P(G_m|C_3, r_2)P(C_3)}$$

When m is 2 we get

$$\begin{aligned} P(C_1|G_2, r_2) &= \frac{P(G_2|C_1, r_2)P(C_1)}{P(G_2|C_1, r_2)P(C_1) + P(G_2|C_2, r_2)P(C_2) + P(G_2|C_3, r_2)P(C_3)} \\ &= \frac{(1/2)(1/3)}{(1/2)(1/3) + 0(1/3) + 1(1/3)} \\ &= (1/3). \end{aligned}$$

We also get

$$\begin{aligned} P(C_3|G_2, r_2) &= \frac{P(G_2|C_3, r_2)P(C_3)}{P(G_2|C_1, r_2)P(C_1) + P(G_2|C_2, r_2)P(C_2) + P(G_2|C_3, r_2)P(C_3)} \\ &= \frac{1(1/3)}{(1/2)(1/3) + 0(1/3) + 1(1/3)} \\ &= (2/3). \end{aligned}$$

Notice what is happening: if the car is behind door 3, then the *only* choice of goat for the host is the goat behind 2. This means that $P(G_2|C_3, r_2) = 1$ and so the conditional probability that the car is behind door 3 is now $2/3$.

It is quite easy to make mistakes in conditional probability (the Monty Hall problem has been the subject of extensive, lively, and often quite inaccurate correspondence in various national periodicals). Several such mistakes have names, because they're so common. One is the **prosecutor's fallacy**. This often occurs in the following form: A prosecutor has evidence E against a suspect. Write I for the event that the suspect is innocent. The evidence has the property that $P(E|I)$ is extremely small; the prosecutor concludes that the suspect is guilty.

The problem here is that the conditional probability of interest is $P(I|E)$ (rather than $P(E|I)$). The fact that $P(E|I)$ is small doesn't mean that $P(I|E)$ is small, because

$$P(I|E) = \frac{P(E|I)P(I)}{P(E)} = \frac{P(E|I)P(I)}{P(E|I)P(I) + P(E|I^c)(1 - P(I))}.$$

Notice how, if $P(I)$ is large or if $P(E|I^c)$ is much smaller than $P(E|I)$, then $P(I|E)$ could be close to one. The question to look at is not how unlikely the evidence is if the subject is innocent; instead, the question is how likely the subject is to be guilty compared to some other source of the evidence. These are two very different questions.

In the previous section, we saw how the probability of getting a chance match in a large DNA database could be quite big, even though the probability of a single match is small. One version of the prosecutors fallacy is to argue that, because the probability of a single match is small, the person who matched the DNA must have committed the crime. The fallacy is to ignore the fact that the probability of a chance match to a large database is quite high.

1.3 SIMULATION AND PROBABILITY

Many problems in probability can be worked out in closed form if one knows enough combinatorial mathematics, or can come up with the right trick. Textbooks are full of these, and we've seen some. Explicit formulas for probabilities are often extremely useful. But it isn't always easy or possible to find a formula for the probability of an event in a model. An alternative strategy is to build a simulation, run it many times, and count the fraction of outcomes where that occurs. This is a simulation experiment.

This strategy rests on our view of probability as relative frequency. We expect that (say) if a coin has probability p of coming up heads, then when we flip it N times, we should see about pN heads. We can use this argument the other way round: if we flip a coin N times and see H heads, then it is reasonable to expect that the coin has probability $p = H/N$ of coming up heads. It is clear that this argument is dangerous for small N (eg try $N = 1$). But (as we shall see later) for large N it is very sound. There are some difficulties: It is important that we build independent simulations, and in some circumstances that can be difficult. Furthermore, our estimate of the probability is not exact. A simulation experiment should involve a large number of runs. Different simulation experiments will give different answers (though hopefully the difference will not be huge). But we can get an estimate of how good our estimate of the probability is — we run several simulation experiments, and look at the results as a data set. The mean is our best estimate of the probability, and the standard deviation gives some idea of how significant the change from experiment to experiment is. As we shall see later, this standard deviation gives us some idea of how good the estimate is.

I will build several examples around a highly simplified version of a real card game. This game is Magic: The Gathering, and is protected by a variety of trademarks, etc. My version — MTGDAF — isn't very interesting as a game, but is good for computing probabilities. The game is played with decks of 60 cards. There are two types of card: Lands, and Spells. Lands can be placed on the play table and stay there permanently; Spells are played and then disappear. A Land on the table can be “tapped” or “untapped”. Players take turns (though we won't deal with any problem that involves the second player, so this is largely irrelevant). Each player draws a hand of seven cards from a shuffled deck. In each turn, a player first untaps any Lands on the table, then draws a card, then plays a land onto the table (if the

player has one in hand to play), then finally can play one or more spells. Each spell has a fixed cost (of $1, \dots, 10$), and this cost is played by “tapping” a land (which is not untapped until the start of the next turn). This means that the player can cast only cheap spells in the early turns of the game, and expensive spells in the later turns.

Worked example 1.26 *MTGDAF — The number of lands*

Assume a deck of 60 cards has 24 Lands. It is properly shuffled, and you draw seven cards. You could draw $0, \dots, 7$ Lands. Estimate the probability for each, using a simulation. Furthermore, estimate the error in your estimates.

Solution: The matlab function `randperm` produces a random permutation of given length. This means you can use it to simulate a shuffle of a deck, as in listing 1.1. I then drew 10,000 random hands of seven cards, and counted how many times I got each number. Finally, to get an estimate of the error, I repeated this experiment 10 times and computed the standard deviation of each estimate of probability. This produced

0.0218 0.1215 0.2706 0.3082 0.1956 0.0686 0.0125 0.0012

for the probabilities (for 0 to 7, increasing number of lands to the right) and

0.0015 0.0037 0.0039 0.0058 0.0027 0.0032 0.0005 0.0004

for the standard deviations of these estimates.

Worked example 1.27 *MTGDAF — The number of lands*

What happens to the probability of getting different numbers of lands if you put only 15 Lands in a deck of 60? It is properly shuffled, and you draw seven cards. You could draw $0, \dots, 7$ Lands. Estimate the probability for each, using a simulation. Furthermore, estimate the error in your estimates.

Solution: You can change one line in the listing to get

0.1159 0.3215 0.3308 0.1749 0.0489 0.0075 0.0006 0.0000

for the probabilities (for 0 to 7, increasing number of lands to the right) and

0.0034 0.0050 0.0054 0.0047 0.0019 0.0006 0.0003 0.0000

for the standard deviations of these estimates.

Listing 1.1: Matlab code used to simulate the number of lands

```

simcards=[ones(24, 1); zeros(36, 1)]
% 1 if land, 0 otherwise
ninsim=10000;
nsims=10;
counts=zeros(nsims, 8);
for i=1:10
    for j=1:10000
        shuffle=randperm(60);
        hand=simcards(shuffle(1:7));
        %useful matlab trick here
        nlands=sum(hand);
        %ie number of lands
        counts(i, 1+nlands)=...
            counts(i, 1+nlands)+1;
        % number of lands could be zero
    end
end
probs=counts/ninsim;
mean(probs)
std(probs)
%%

```

Worked example 1.28 *MTGDFAF — Playing spells*

Assume you have a deck of 24 Lands, 10 Spells of cost 1, 10 Spells of cost 2, 10 Spells of cost 3, 2 Spells of cost 4, 2 Spells of cost 5, and 2 Spells of cost 6. Assume you always only play the cheapest spell in your hand (i.e. you never play two spells). What is the probability you will be able to play at least one spell on each of the first four turns?

Solution: This simulation requires just a little more care. You draw the hand, then simulate the first four turns. In each turn, you can only play a spell whose cost you can pay, and only if you have it. I used the matlab of listing 1.2 and listing 1.3; I found the probability to be 0.64 with standard deviation 0.01. Of course, my code might be wrong....

Worked example 1.29 *MTGDFAF — Playing spells*

Now we use a different distribution of cards. Assume you have a deck of 20 Lands, 9 Spells of cost 1, 5 Spells of cost 2, 5 Spells of cost 3, 5 Spells of cost 4, 5 Spells of cost 5, and 11 Spells of cost 6. Assume you always only play the cheapest spell in your hand (i.e. you never play two spells). What is the probability you will be able to play at least one spell on each of the first four turns?

Solution: This simulation requires just a little more care. You draw the hand, then simulate the first four turns. In each turn, you can only play a spell whose cost you can pay, and only if you have it. I found the probability to be 0.33 with standard deviation 0.05. Of course, my code might be wrong....

One engaging feature of the real game that is revealed by these very simple simulations is the tension between a player's goals. The player would like to have few lands — so as to have lots of spells — but doing so means that there's a bigger chance of not being able to play a spell. Similarly, a player would like to have lots of powerful (=expensive), but doing so means there's a bigger chance of not being able to play a spell. Players of the real game spend baffling amounts of time arguing with one another about the appropriate choices for a good set of cards.

Experiments reveal one nice feature about simulations which is usually true (though sometimes quite hard to prove, and occasionally not even true). You can reasonably expect that the probability you compute from a simulation behaves like normal data. Run a simulation experiment a large number of times and construct a data set whose entries are the probability from each experiment. This data should be normal. Recall that this means that, if you subtract the mean and divide by the standard deviation, you should get a histogram that looks like the standard normal curve. This is important, because it means that the right answer should be very few standard deviations away from the mean you compute. We will explore this phenomenon, and its consequences, in more detail later. Figure 28 shows some examples.

PROBLEMS

- 1.1. **Monty Hall, Rule 3:** If the host uses rule 3, then what is $P(C_1|G_2, r_3)$? Do this by computing conditional probabilities.
- 1.2. **Monty Hall, Rule 4:** If the host uses rule 4, and shows you a goat behind door 2, what is $P(C_1|G_2, r_4)$? Do this by computing conditional probabilities.

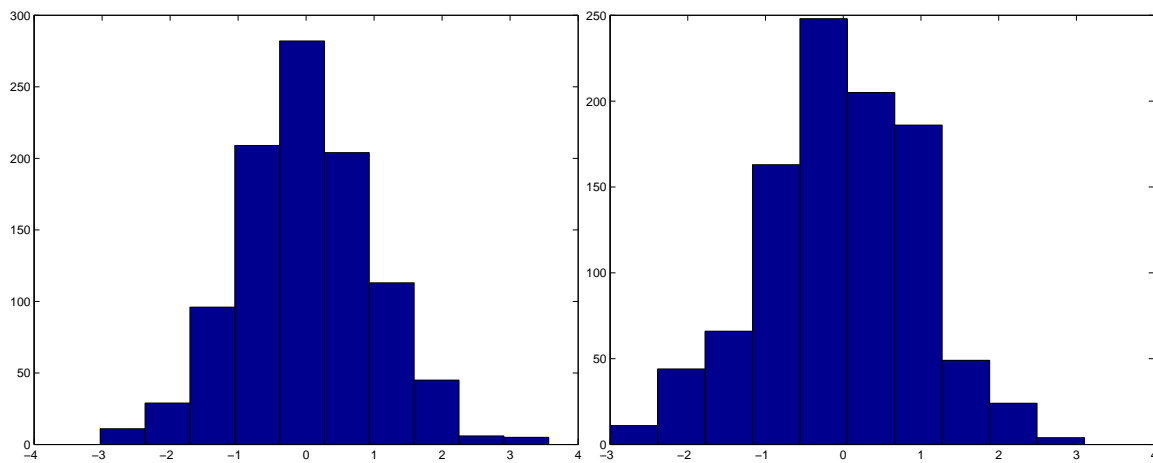


FIGURE 1.2: *Estimates of probabilities produced by simulation typically behave like normal data. On the left, I show a histogram of probabilities of having a hand of 3 Lands in the simulation of example 26; these are plotted in standard coordinates. On the right, I show a histogram of probability of playing a spell in each of the first four turns (example 29), from 1000 simulation experiments; again, these are plotted in standard coordinates. Compare these to the standard normal histograms of the previous chapter.*

Listing 1.2: Matlab code used to simulate the four turns

```

simcards=[zeros(24, 1); ones(10, 1);...
          2*ones(10, 1);3*ones(10, 1); ...
          4*ones(2, 1); 5*ones(2, 1); 6*ones(2, 1)];
nsims=10;
ninsim=1000;
counts=zeros(nsims, 1);
for i=1:nsims
    for j=1:ninsim
        % draw a hand
        shuffle=randperm(60);
        hand=simcards(shuffle(1:7));
        %reorganize the hand
        cleanhand=zeros(7, 1);
        for k=1:7
            cleanhand(hand(k)+1)=cleanhand(hand(k)+1)+1;
            % ie count of lands, spells, by cost
        end
        landsontable=0;
        [playedspell1, landsontable, cleanhand]=...
            playground(landsontable, cleanhand, shuffle, ...
                simcards, 1);
        [playedspell2, landsontable, cleanhand]=...
            playground(landsontable, cleanhand, shuffle, ...
                simcards, 2);
        [playedspell3, landsontable, cleanhand]=...
            playground(landsontable, cleanhand, shuffle, ...
                simcards, 3);
        [playedspell4, landsontable, cleanhand]=...
            playground(landsontable, cleanhand, shuffle, ...
                simcards, 4);
        counts(i)=counts(i)+...
            playedspell1*playedspell2*...
            playedspell3*playedspell4;
    end
end
end

```

Listing 1.3: Matlab code used to simulate playing a turn

```

function [playedspell , landsontable , cleanhand]=...
    playground(landsontable , cleanhand , shuffle , simcards , ...
    turn)
    % draw
ncard=simcards( shuffle(7+turn) );
cleanhand( ncard+1)=cleanhand( ncard+1)+1;
% play land
if cleanhand(1)>0
    landsontable=landsontable+1;
    cleanhand(1)=cleanhand(1)-1;
end
playedspell=0;
if landsontable>0
    i=1; done=0;
    while done==0
        if cleanhand(i)>0
            cleanhand(i)=cleanhand(i)-1;
            playedspell=1;
            done=1;
        else
            i=i+1;
            if i>landsontable
                done=1;
            end
        end
    end
end
end

```