Basic Ideas in Probability

3

We will perform experiments—which could be pretty much anything, from flipping a coin, to eating too much saturated fat, to smoking, to crossing the road without looking—and reason about the outcomes (mostly bad for the examples I gave). But these outcomes are uncertain, and we need to weigh those uncertainties against one another. If I flip a coin, I could get heads or tails, and there's no reason to expect to see one more often than the other. If I eat too much saturated fat or smoke, I will very likely have problems, though I might not. If I cross the road without looking, I may be squashed by a truck or I may not. Our methods need also to account for information. If I look before I cross the road, I am much less likely to be squashed. Probability is the machinery we use to describe and account for the fact that some outcomes are more frequent than others.

3.1   Experiments, Outcomes and Probability

Imagine you repeat the same experiment numerous times. You do not necessarily expect to see the same result each time. Some results might occur more frequently than others. We account for this tendency using probability. To do so, we need to be clear about what results an experiment can have. For example, you flip a coin. We might agree that the only possible results are a head or a tail, thus ignoring the possibilities that (say) a bird swoops down and steals the coin; the coin lands and stays on edge; the coin falls between the cracks in the floor and disappears; and so on. By doing so, we have idealized the experiment.

3.1.1   Outcomes and Probability

We will formalize experiments by specifying the set of **outcomes** that we expect from the experiment. Every run of the experiment produces exactly one of the set of possible outcomes. We never see two or more outcomes from a single experiment, and we never see no outcome. The advantage of doing this is that we can count how often each outcome appears.

> **Definition 3.1 (Sample Space)**   The sample space is the set of all outcomes, which we usually write Ω.

> **Worked example 3.1 (Find the Lady)**   We have three playing cards. One is a queen; one is a king, and one is a jack. 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, J for jack; the outcomes are $\{Q, K, J\}$

Worked example 3.2 (Find the Lady, Twice) We play find the lady twice, replacing the card we have chosen. What is the sample space?

Solution We now have $\{QQ, QK, QN, KQ, \; KK, KJ, JQ, JK, JJ\}$

Worked example 3.3 (A Poor Choice of Strategy for Planning a Family)   A couple decides to have children. As they know no mathematics, they decide to have children until a girl then a boy are born. What is the sample space? Does this strategy bound the number of children they could be planning to have?

Solution Write B for boy, G for girl. The sample space looks like any string of B's and G's that (a) ends in GB and (b) does not contain any other GB. In regular expression notation, you can write such strings as B^*G^+B. There is a lower bound on the length of the string (two), but no upper bound. As a family planning strategy, this is unrealistic, but it serves to illustrate the point that sample spaces don't have to be finite to be tractable.

Remember this: *Sample spaces are required, and need not be finite*

We represent our model of how often a particular outcome will occur in a repeated experiment with a **probability**, a non-negative number. This number gives the relative frequency of the outcome of interest, when an experiment is repeated a very large number of times.

Assume that we repeat an experiment N times. Assume also that the coins, dice, whatever involved in each repetition of the experiment don't communicate with one another from experiment to experiment (or, equivalently, that experiments don't "know" about one another). We say that an outcome A has probability P if (a) outcome A occurs in about $N \times P$ of those experiments and (b) as N gets larger, the fraction of experiments where outcome A occurs will get closer to P. We write $\#(A)$ for the number of times outcome A occurs. We interpret P as

$$\lim_{N \to \infty} \frac{\#(A)}{N}.$$

We can draw two important conclusions immediately.

- For any outcome A, $0 \le P(A) \le 1$.
- $\sum_{A_i \in \Omega} P(A_i) = 1$.

Remember that every run of the experiment produces exactly one outcome. The probabilities add up to one because each experiment must have one of the outcomes in the sample space. Some problems can be handled by building a set of outcomes and reasoning about the probability of each outcome. This is particularly useful when the outcomes *must* have the same probability, which happens rather a lot.

Worked example 3.4 (A Biased Coin) Assume 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 flip this coin three million times. How many times do we see heads?

Solution $P(H) = \frac{1}{3}$, so we expect this coin will come up heads in $\frac{1}{3}$ of experiments. This means that we will very likely see very close to a million heads. Later on, we will be able to be more precise.

Remember this: *The probability of an outcome is the frequency of that outcome in a very large number of repeated experiments. The sum of probabilities over all outcomes must be one.*

3.2   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 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 we should be able to predict the probability of a *set* of outcomes from any reasonable model of an experiment. For example, we might roll a die and ask what the probability of getting an even number is. We would like our probability models to be able to predict the probability of sets of outcomes.

> **Definition 3.2 (Event)**   An event is a set of outcomes. I will usually write events as sets (so, for example, $\mathcal{E}$).

Assume we are given a discrete sample space Ω. A natural choice of an event space is the collection of all subsets of Ω. It turns out that this is not the only possible choice, but we will ignore this point. So far, we have described the probability of each outcome with a non-negative number. We can extend this idea of probability to deal with events in a straightforward way.

The set of all outcomes, which we wrote Ω, must be an event. We must have $P(\Omega) = 1$ (because we said that every run of an experiment produces one outcome, and that outcome must be in Ω). In principle, there could be no outcome, although this never happens. This means that the empty set, which we write $\varnothing$, is an event, and we have $P(\varnothing) = 0$.

Any given outcome must be an event, because an event is a set of outcomes. Now assume A and B are two distinct outcomes, and write $\mathcal{E} = \{A, B\}$ for the event that contains both. We must have that $P(\mathcal{E}) = P(A) + P(B)$, because the number of times repeated experiments produce an outcome in $\mathcal{E}$ is given by the number of times we see A plus the number of times we see B. Now assume that C_i are N distinct outcomes, and $\mathcal{F}$ is the event that contains all of them, and no other outcomes. Then we must have $P(\mathcal{F}) = \sum_i P(C_i)$ (because we observe an outcome in $\mathcal{F}$ whenever we see any of the outcomes C_i). In turn, this means that if $\mathcal{E}$ and $\mathcal{F}$ are disjoint events, $P(\mathcal{E} \cup \mathcal{F}) = P(\mathcal{E}) + P(\mathcal{F})$. All this yields a straightforward set of properties, collected in a box below.

> **Useful Facts 3.1 (Basic Properties of the Probability Events)**
> We have
>
> - The probability of every event is between zero and one; in equations
>
> $$0 \le P(\mathcal{A}) \le 1$$
>
>   for any event $\mathcal{A}$.
> - Every experiment has an outcome; in equations,
>
> $$P(\Omega) = 1.$$
>
> - The probability of disjoint events is additive; writing this in equations requires some notation. Assume that we have a collection of events $\mathcal{A}_i$, indexed by i. We require that these have the property $\mathcal{A}_i \cap \mathcal{A}_j = \varnothing$ when $i \ne j$. 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
>
> $$P(\cup_i \mathcal{A}_i) = \sum_i P(\mathcal{A}_i)$$

3.2.1   Computing Event Probabilities by Counting Outcomes

If you can compute the probability of each outcome in an event $\mathcal{F}$, computing the probability of the event is straightforward. The outcomes are each disjoint events, so you just add the probabilities. A common, and particularly useful, case occurs when you know each outcome in the sample space has the same probability. In this case, computing the probability of an event is an exercise in counting. You can show

$$P(\mathcal{F}) = \frac{\text{Number of outcomes in } \mathcal{F}}{\text{Total number of outcomes in } \Omega}$$

(look at the exercises).

Worked example 3.5 (Odd Numbers with Fair Dice)   We throw a fair (each number has the same probability) six-sided die twice, then add the two numbers. What is the probability of getting an odd number?

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

Worked example 3.6 (Numbers Divisible by Five with Fair Dice)   We throw a fair (each number has the same probability) six-sided die twice, then add the two numbers. What is the probability of getting a number divisible by five?

Solution   There are 36 outcomes. 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, so the probability is 7/36.

Sometimes a bit of fiddling with the space of outcomes makes it easy to compute what we want. Examples 3.8 and 3.47 show cases where you can use fictitious outcomes as an accounting device to simplify a computation.

Worked example 3.7 (Children—1)   A couple decides to have children. They decide simply to have three children. Assume that three births occur, each birth results in one child, and boys and girls are equally likely at each birth. Let $\mathcal{B}_i$ be the event that there are i boys, and $\mathcal{C}$ be the event there are more girls than boys. Compute $P(\mathcal{B}_1)$ and $P(\mathcal{C})$.

Solution   There are eight outcomes. Each has the same probability. Three of them have a single boy, so $P(\mathcal{B}_1) = 3/8$. Four of these outcomes have more girls than boys, so $P(\mathcal{C}) = 1/2$.

Worked example 3.8 (Children—2)   A couple decides to have children. They decide to have children until the first girl is born, or until there are three, and then stop. Assume that each birth results in one child, and boys and girls are equally likely at each birth. Let $\mathcal{B}_i$ be the event that there are i boys, and $\mathcal{C}$ be the event there are more girls than boys. Compute $P(\mathcal{B}_1)$ and $P(\mathcal{C})$.

Solution   In this case, we could write the outcomes as $\{G, BG, BBG\}$, 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. This gives us natural collection of events for which it is easy to compute probabilities. Having one girl corresponds to the event $\{Gbb, Gbg, Ggb, Ggg\}$, where I have used lowercase letters to write the fictitious later births; the probability is 1/2. Having a boy then a girl corresponds to the event $\{BGb, BGg\}$ (and so has probability 1/4). Having two boys then a girl corresponds to the event $\{BBG\}$ (and so has probability 1/8). Finally, having three boys corresponds to the event $\{BBB\}$ (and so has probability 1/8). This means that $P(\mathcal{B}_1) = 1/4$ and $P(\mathcal{C}) = 1/2$.

Counting outcomes in an event can require pretty elaborate combinatorial arguments. One form of argument that is particularly important is to reason about permutations and combinations. You should recall that the number of distinct permutations of N items is $N!$.

Worked example 3.9 (Card Hands) You draw a hand of seven cards from a properly shuffled standard deck of cards. With what probability do receive 2–8 of hearts, *in that order*?

Solution There are numerous ways to do this, but I'll use permutations. There are 52! different orderings of a properly shuffled deck of cards. This is the total number of outcomes. The number of outcomes in the event comes by noticing that any outcome in the event is an ordering of the cards where the first seven cards are 2–8 of hearts, in that order. So there are 45! outcomes in the event, because you can reorder the remaining 45 cards arbitrarily. This means the probability is

$$\frac{45!}{52!}.$$

The number of combinations of k items, chosen from N, *where the order does not matter*, is given by

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

Worked example 3.10 (Card Hands—2) You draw a hand of seven cards from a properly shuffled standard deck of cards. With what probability do receive 2–8 of hearts, *in any order*?

Solution There are 52! different orderings of a properly shuffled deck of cards, so 52! outcomes Of these, 45! have the first seven cards 2–8 of hearts. There are 7! orderings of these cards. So the number of outcomes in the event is 45! 7! and the probability is

$$\frac{7!45!}{52!}$$

Alternatively, there are $\binom{N}{k}$ hands of seven distinct cards, ignoring the order in which they are obtained. Only one such hand contains 2–8 of hearts, so the probability is

$$\frac{1}{\binom{52}{7}}$$

(and you should check this reasoning got us to the same answer as the previous argument).

Worked example 3.11 (Card Hands—3) You draw a hand of seven cards from a properly shuffled standard deck of cards. With what probability does your hand contain 2–8 of *any suit*? The cards don't have to have the same suit, and they can arrive *in any order*.

Solution From the previous example, there are 52! orderings of a properly shuffled deck and so 52! outcomes in total. There are 45! orderings that fix the first seven cards to some specified values, as in Worked example 3.9. The number of hands of seven cards that works is obtained by (a) choosing a suit for each card then (b) counting the number of different orders. This yields $4^7 7!45!$ outcomes in the event, so the probability is
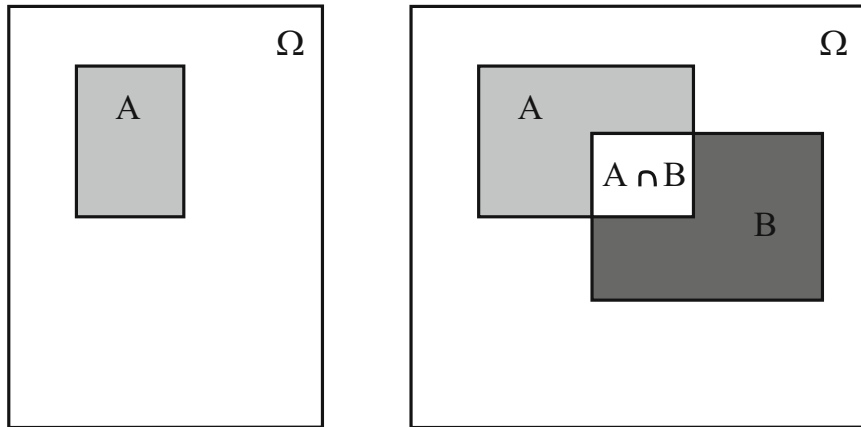
$$\frac{4^7 7!45!}{52!}.$$

Fig. 3.1 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. On the *left*, a Venn diagram to help remember that $P(\mathcal{A}) + P(\mathcal{A}^c) = 1$. The "size" of Ω is 1, outcomes lie either in $\mathcal{A}$ or $\mathcal{A}^c$, and the two don't intersect. On the *right*, 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

> **Remember this:** *In some problems, you can compute the probabilities of events by counting outcomes.*

3.2.2   The Probability of Events

There is an analogy between probability and "size" which is helpful in deriving and remembering expressions for the probability of events. Think about the probability of an event as the "size" of that event. This "size" is relative to Ω, which has "size" 1. I find this a good way to remember equations. Some people find Venn diagrams a useful way to keep track of this argument, and Fig. 3.1 is for them.

Notice that $\mathcal{A}$ and $\mathcal{A}^c$ don't overlap, and together make up all of Ω. So the "size" of $\mathcal{A}$ and the "size" of $\mathcal{A}^c$ should add to the "size" of Ω and so

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

Notice the "size" of the part of $\mathcal{A}$ that isn't in $\mathcal{B}$ is obtained by taking the "size" of $\mathcal{A}$ and subtracting the "size" of $\mathcal{A} \cap \mathcal{B}$—that is, the part of $\mathcal{A}$ that is also in $\mathcal{B}$. This means that

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

Notice the "size" of $\mathcal{A} \cup \mathcal{B}$ is obtained by adding the two "sizes", then subtracting the "size" of the intersection because otherwise you would double-count the part where the two sets overlap. This means that

$$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \cap \mathcal{B}).$$

I have collected these expressions, which you should remember, in box 3.2. The "size" analogy can be made precise by thinking about "size" in the right way; I won't bother, because doing so takes effort without really enhancing the underlying intuition. I prove the expressions are right without using the "size" analogy below.

Useful Facts 3.2 (Properties of the Probability of Events)

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

Proposition $P(\mathcal{A}^c) = 1 - P(\mathcal{A})$

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

Proposition $P(\emptyset) = 0$

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

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

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

Proposition $P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \cap \mathcal{B})$

Proof $P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A} \cup (\mathcal{B} \cap \mathcal{A}^c)) = P(\mathcal{A}) + P((\mathcal{B} \cap \mathcal{A}^c))$. 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 $P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P((\mathcal{B} \cap \mathcal{A}^c)) = P(\mathcal{A}) + P(\mathcal{B}) - P((\mathcal{B} \cap \mathcal{A}))$.

Proposition $P(\cup_1^n \mathcal{A}_i) = \sum_i P(\mathcal{A}_i) - \sum_{i<j} P(\mathcal{A}_i \cap \mathcal{A}_j) + \sum_{i<j<k} P(\mathcal{A}_i \cap \mathcal{A}_j \cap \mathcal{A}_k) + \ldots (-1)^{(n+1)} P(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \ldots \cap \mathcal{A}_n)$

Proof This 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).

$$P(\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3)$$

$$= P(\mathcal{A}_1 \cup (\mathcal{A}_2 \cup \mathcal{A}_3))$$

$$= P(\mathcal{A}_1) + P(\mathcal{A}_2 \cup \mathcal{A}_3)$$

$$- P(\mathcal{A}_1 \cap (\mathcal{A}_2 \cup \mathcal{A}_3))$$

$$= P(\mathcal{A}_1) + (P(\mathcal{A}_2) + P(\mathcal{A}_3) - P(\mathcal{A}_2 \cap \mathcal{A}_3))$$

$$- P((\mathcal{A}_1 \cap \mathcal{A}_2) \cup (\mathcal{A}_1 \cap \mathcal{A}_3))$$

$$= P(\mathcal{A}_1) + (P(\mathcal{A}_2) + P(\mathcal{A}_3) - P(\mathcal{A}_2 \cap \mathcal{A}_3))$$

$$-P(\mathcal{A}_1 \cap \mathcal{A}_2) - P(\mathcal{A}_1 \cap \mathcal{A}_3)$$

$$-(-P((\mathcal{A}_1 \cap \mathcal{A}_2) \cap (\mathcal{A}_1 \cap \mathcal{A}_3)))$$

$$= P(\mathcal{A}_1) + P(\mathcal{A}_2) + P(\mathcal{A}_3)$$

$$-P(\mathcal{A}_2 \cap \mathcal{A}_3) - P(\mathcal{A}_1 \cap \mathcal{A}_2) - P(\mathcal{A}_1 \cap \mathcal{A}_3)$$

$$+P(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3)$$

3.2.3   Computing Probabilities by Reasoning About Sets

The rule $P(\mathcal{A}^c) = 1 - P(\mathcal{A})$ is occasionally useful for computing probabilities on its own. More commonly, you need other reasoning as well. The next problem illustrates an important feature of questions in probability: your intuition can be quite misleading. One problem is that the number of outcomes can be bigger or smaller than you expect.

Worked example 3.12 (Shared Birthdays) What is the probability that, in a room of 30 people, there is a pair of people who have the same birthday?

Solution 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). This model isn't perfect (there tend to be slightly more births roughly 9 months after: the start of spring; blackouts; major disasters; and so on) but it's workable. 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 compute. 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

$$P(\{\text{shared birthday}\})$$

$$= 1 - \frac{365 \times 364 \times \ldots 336}{365^{30}}$$

$$\approx 1 - 0.29 = 0.71$$

which means there's really a pretty good chance that two people in a room of 30 share a birthday.

If we change the birthday example slightly, the problem changes drastically. If you stand up in a room of 30 people and bet that two people in the room have the same birthday, you have a probability of winning of about 0.71. If you bet that there is someone else in the room who has the same birthday that you do, your probability of winning is very different.

Worked example 3.13 (Shared Birthdays) You bet there is someone else in a room of 30 people who has the same birthday that you do. Assuming you know nothing about the other 29 people, what is the probability of winning?

Solution The easy way to do this is
$$P(\{\text{winning}\}) = 1 - P(\{\text{losing}\}).$$
Now you will lose if everyone has a birthday different from you. You can think of the birthdays of the others in the room as a list of 29 days of the year. If your birthday is on the list, you win; if it's not, you lose. The number of losing lists is the number of lists of 29 days of the year such that your birthday is not in the list. This number is easy to get. We have 364 days of the year to choose from for each of 29 locations in the list. The total number of lists is the number of lists of 29 days of the year. Each list has the same probability. So

$$P(\{\text{losing}\}) = \frac{364^{29}}{365^{29}}$$

and

$$P(\{\text{winning}\}) \approx 0.0765.$$

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 (Examples 3.12 and 3.13 are reliably profitable; you could probably do quite well out of Examples 3.45 and 3.46).

The rule $P(\mathcal{A}-\mathcal{B}) = P(\mathcal{A}) - P(\mathcal{A}\cap\mathcal{B})$ is also occasionally useful for computing probabilities on its own; more commonly, you need other reasoning as well.

Worked example 3.14 (Dice) You flip two fair six-sided dice, and add the number of spots. What is the probability of getting a number divisible by 2, but not by 5?

Solution There is an interesting way to work the problem. Write $\mathcal{D}_n$ for the event the number is divisible by n. Now $P(\mathcal{D}_2) = 1/2$ (count the cases; or, more elegantly, notice that each die has the same number of odd and even faces, and work from there). Now $P(\mathcal{D}_2 - \mathcal{D}_5) = P(\mathcal{D}_2) - P(\mathcal{D}_2 \cap \mathcal{D}_5)$. But $\mathcal{D}_2 \cap \mathcal{D}_5$ contains only three outcomes ($6, 4, 5, 5$ and $4, 6$), so $P(\mathcal{D}_2 - \mathcal{D}_5) = 18/36 - 3/36 = 5/12$

Sometimes it is easier to reason about unions than to count outcomes directly.

Worked example 3.15 (Two Fair Dice) I roll two fair six-sided dice. What is the probability that the result is divisible by either 2 or 5, or both?

Solution Write $\mathcal{D}_n$ for the event the number is divisible by n. We want $P(\mathcal{D}_2 \cup \mathcal{D}_5) = P(\mathcal{D}_2) + P(\mathcal{D}_5) - P(\mathcal{D}_2 \cap \mathcal{D}_5)$. From Example 3.14, we know $P(\mathcal{D}_2) = 1/2$ and $P(\mathcal{D}_2 \cap \mathcal{D}_5) = 3/36$. By counting outcomes, $P(\mathcal{D}_5) = 7/36$. So $P(\mathcal{D}_2 \cup \mathcal{D}_5) = (18 + 7 - 3)/36 = 22/36$.

3.3   Independence

Some experimental results do not affect others. For example, if I flip a coin twice, whether I get heads on the first flip has no effect on whether I get heads on the second flip. As another example, I flip a coin; the outcome does not affect whether I get hit on the head by a falling apple later in the day. We refer to events with this property as **independent**.

Here is a pair of events that is not independent. Imagine I throw a six-sided die. Write $\mathcal{A}$ for the event that the die comes up with an odd number of spots, and write $\mathcal{B}$ for the event that the number of spots is either 3 or 5. Now these events are
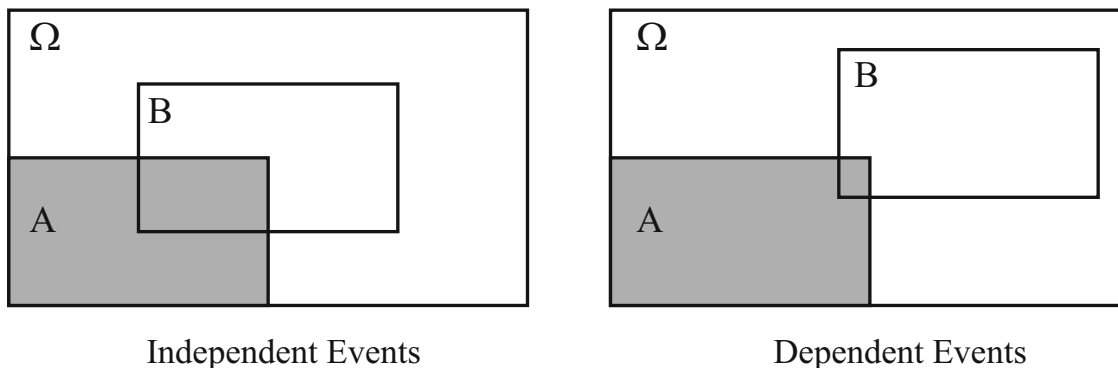
Independent Events   Dependent Events

Fig. 3.2 On the *left*, $\mathcal{A}$ and $\mathcal{B}$ are independent. $\mathcal{A}$ spans 1/4 of Ω, and $\mathcal{A} \cap \mathcal{B}$ spans 1/4 of $\mathcal{B}$. This means that knowing whether an outcome is in $\mathcal{A}$ or not doesn't affect the probability that it is in $\mathcal{B}$. 1/4 of the outcomes of Ω lie in $\mathcal{A}$, and 1/4 of theoutcomes in $\mathcal{B}$ lie in $\mathcal{A} \cap \mathcal{B}$. On the *right*, they are not. Very few of the outcomes in $\mathcal{B}$ lie in $\mathcal{B} \cap \mathcal{A}$, so that observing $\mathcal{B}$ means that $\mathcal{A}$ becomes less likely, because very few of the outcomes in $\mathcal{B}$ also lie in $\mathcal{A} \cap \mathcal{B}$

interrelated in an important way. If I *know* that $\mathcal{B}$ has occurred, I also know that $\mathcal{A}$ has occurred—I don't need to check separately, because $\mathcal{B}$ implies $\mathcal{A}$.

Here is an example of a weaker interaction that results in events not being independent. Write $\mathcal{C}$ for the event that the die comes up with an odd number of spots, and write $\mathcal{D}$ for the event that the number of spots is larger than 3. These events are interrelated. The probability of each event separately is 1/2. If I *know* that $\mathcal{C}$ has occurred, then I know that the die shows either 1, 3, or 5 spots. One of these outcomes belongs to $\mathcal{D}$, and two do not. This means that knowing that $\mathcal{C}$ has occurred tells you something about whether $\mathcal{D}$ has occurred. Independent events do *not* have this property. This means that the probability that they occur together has an important property, given in the box below.

Definition 3.3 (Independent Events) Two events $\mathcal{A}$ and $\mathcal{B}$ are **independent** if and only if

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

The "size" analogy helps motivate this expression. We think of $P(A)$ as the "size" of $\mathcal{A}$ relative to Ω, and so on. Now $P(\mathcal{A} \cap \mathcal{B})$ measures the "size" of $\mathcal{A} \cap \mathcal{B}$—that is, the part of $\mathcal{A}$ that lies inside $\mathcal{B}$. But if $\mathcal{A}$ and $\mathcal{B}$ are independent, then the "size" of $\mathcal{A} \cap \mathcal{B}$ relative to $\mathcal{B}$ should be the same as the "size" of $\mathcal{A}$ relative to Ω (Fig. 3.2). Otherwise, $\mathcal{B}$ affects $\mathcal{A}$, because $\mathcal{A}$ is more (or less) likely when $\mathcal{B}$ has occurred.

So for $\mathcal{A}$ and $\mathcal{B}$ to be independent, we must have

$$\text{"Size" of } \mathcal{A} = \frac{\text{"Size" of piece of } \mathcal{A} \text{ in } \mathcal{B}}{\text{"Size" of } \mathcal{B}},$$

or, equivalently,

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

which yields our expression.

Worked example 3.16 (Fair Dice) The space of outcomes for a fair six-sided die is

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

The die is fair, so each outcome has the same probability. Now we toss two fair six-sided dice. The outcome for each die is independent of that for the other. With what probability do we get two threes?

(continued)

Solution

$$P \text{ (first die yields 3} \cap \text{second die yields 3)}$$
$$= P(\text{first die yields 3}) \times$$
$$P(\text{second die yields 3})$$
$$= (1/6)(1/6)$$
$$= 1/36$$

Worked example 3.17 (Find the Lady, Twice) Recall the setup of Worked example 3.1. Assume that the card that is chosen is chosen fairly—that is, each card is chosen with the same probability. The game is played twice, and the cards are reshuffled between games. What is the probability of turning up a Queen and then a Queen again?

Solution The events are independent, so 1/9.

You can use Definition 3.3 (i.e. $\mathcal{A}$ and $\mathcal{B}$ are independent if and only if $P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B})$) to tell whether events are independent or not. Quite small changes to a problem affect whether events are independent, as in the worked example below.

Worked example 3.18 (Cards and Independence) We shuffle a standard deck of 52 cards and draw one card. The event $\mathcal{A}$ is "the card is a red suit" and the event $\mathcal{B}$ is "the card is a 10". (1): Are $\mathcal{A}$ and $\mathcal{B}$ independent?

Now we take a standard deck of cards, and remove the ten of hearts. We shuffle this deck, and draw one card. The event $\mathcal{C}$ is "the card drawn from the modified deck is a red suit" and the event $\mathcal{D}$ is "the card drawn from the modified deck is a 10". (2): Are $\mathcal{C}$ and $\mathcal{D}$ independent?

Solution (1): $P(\mathcal{A}) = 1/2$, $P(\mathcal{B}) = 1/13$ and in Example 3.44 we determined $P(\mathcal{A} \cap \mathcal{B}) = 2/52$. But $2/52 = 1/26 = P(\mathcal{A})P(\mathcal{B})$, so they are independent.

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

The probability of a sequence of independent events can become very small very quickly, and this often misleads people.

Worked example 3.19 (Accidental DNA Matches) I search a DNA database with a sample. Each time I attempt to match this sample to an entry in the database, there is a probability of an accidental chance match of $1e - 4$. Chance matches are independent. There are 20,000 people in the database. What is the probability I get at least one match, purely by chance?

Solution This is $1 - P(\text{no chance matches})$. But $P(\text{no chance matches})$ is much smaller than you think. We have

$$P(\text{no chance matches})$$

$$= P \begin{pmatrix} \text{no chance match to record 1} \cap \\ \text{no chance match to record 2} \cap \\ \dots \cap \\ \text{no chance match to record 20,000} \end{pmatrix}$$

$$= P(\text{no chance match to a record})^{20,000}$$

$$= (1 - 1e - 4)^{20,000}$$

$$\approx 0.14$$

so the probability is about 0.86 that you get at least one match by chance. If you're surprised, look at the exponent. Notice that if the database gets bigger, the probability grows; so at 40,000 the probability of one match by chance is 0.98.

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. At time of writing, Wikipedia has some fascinating stories about the gambler's fallacy which suggest that it's quite a common mistake. People may interpret, say, a run of 20 heads as evidence that either the coin isn't fair, or the tosses aren't independent.

Remember this: *Independence can mislead your intuition. There are two common problems. The first happens because the probability of a set of independent events can become very small very quickly, so that modelling events that aren't independent as independent can lead to trouble (as in Worked example 3.19). The second happens because most people want to believe that the universe keeps track of independent events to ensure that probability calculations work (the gambler's fallacy).*

3.3.1 Example: Airline Overbooking

We can now quite easily study airline overbooking. Airlines generally sell more tickets for a flight than there are seats on the aircraft, because some passengers don't turn up on time, usually for random reasons. If the airline only sold one ticket per seat, their planes would likely have empty seats—which are lost profit—on each flight. If too many passengers turn up for a flight, the airline hopes that someone will accept a reasonable sum of money to take the next flight. Overbooking is sensible, efficient behavior and good for passengers *if* sensibly administered by the airline. This is because ticket prices should be at their lowest when each plane is just full, and there is quite likely some passenger who will take money to fly at some other time.

To choose the number of extra tickets sold, the airline needs to think about the probability of having to pay out (which we compute below) and the amount of money they will need to pay. We don't have the tools to discuss how much the airline may need to pay, which depends quite a lot on passenger behavior, details of the schedule for the next flight, and so on. On occasion, the strategy can get expensive for the airline. While I was revising this text for publication, an airline managed to hit headlines by having airport security drag a passenger off a flight. Details of the resulting settlement were not publicised, but it can't have been cheap for the airline.

Worked example 3.20 (Overbooking—1) An airline has a regular flight with six seats. It always sells seven tickets. Passengers turn up for the flight with probability p, and do so independent of other passengers. What is the probability that the flight is overbooked?

Solution This is like a coin-flip problem; think of each passenger as a biased coin. With probability p, the biased coin comes up T (for *t*urn up) and with probability $(1 - p)$, it turns up H (for no-s*h*ow). This coin is flipped seven times, and we are interested in the probability that there are seven T's. This is p^7, because the flips are independent.

Worked example 3.21 (Overbooking—2) An airline has a regular flight with six seats. It always sells eight tickets. Passengers turn up for the flight with probability p, and do so independent of other passengers. What is the probability that the flight is overbooked?

Solution Now we flip the coin eight times, and are interested in the probability of getting more than six T's. This is the union of two disjoint events (seven T's and eight T's). For the case of seven T's, one flip must be H; there are eight choices for this flip. For the case of eight T's, all eight flips must be T, and there is only one way to achieve this. So the probability the flight is overbooked is

$$P(\text{overbooked}) = P(7\ T\text{'s} \cup 8\ T\text{'s})$$

$$= P(7\ T\text{'s}) + P(8\ T\text{'s})$$

$$= 8p^7(1-p) + p^8$$

Worked example 3.22 (Overbooking—3) An airline has a regular flight with six seats. It always sells eight tickets. Passengers turn up for the flight with probability p, and do so independent of other passengers. What is the probability that six passengers arrive? (i.e. the flight is not overbooked or underbooked).

Solution Now we flip the coin eight times, and are interested in the probability of getting exactly six T's. The probability that a particular set of six passengers arrives is given by the probability of getting any given string of six T's and two H's. This must have probability $p^6(1-p)^2$. But there are a total of $\frac{8!}{2!6!}$ distinct such strings. So the probability that six passengers arrive is

$$\frac{8!}{2!6!}p^6(1-p)^2 = 28p^6(1-p)^2.$$

Worked example 3.23 (Overbooking—4) An airline has a regular flight with s seats. It always sells t tickets. Passengers turn up for the flight with probability p, and do so independent of other passengers. What is the probability that u passengers turn up?

Solution Now we flip the coin t times, and are interested in the probability of getting u T's. There are

$$\frac{t!}{u!(t-u)!}$$

disjoint outcomes with u T's and $t-u$ H's. Each such outcome is independent, and has probability $p^u(1-p)^{t-u}$. So

$$P(u \text{ passengers turn up}) = \frac{t!}{u!(t-u)!}p^u(1-p)^{t-u}$$

Worked example 3.24 (Overbooking—5) An airline has a regular flight with s seats. It always sells t tickets. Passengers turn up for the flight with probability p, and do so independent of other passengers. What is the probability that the flight is oversold?

Solution We need $P(\{s + 1 \text{ turn up}\} \cup \{s + 2 \text{ turn up}\} \cup \ldots \cup \{t \text{ turn up}\})$. But the events $\{i \text{ turn up}\}$ and $\{j \text{ turn up}\}$ are disjoint if $i \neq j$. So we can exploit Example 3.23, and write

$$P(\text{overbooked}) = P(\{s + 1 \text{ turn up}\})$$

$$+ P(\{s + 2 \text{ turn up}\}) +$$

$$\ldots P(\{t \text{ turn up}\})$$

$$= \sum_{i=s+1}^{t} P(\{i \text{ turn up}\})$$

$$= \sum_{i=s+1}^{t} \frac{t!}{i!(t-i)!} p^i (1-p)^{t-i}$$

3.4   Conditional Probability

Imagine we have two events $\mathcal{A}$ and $\mathcal{B}$. If they are independent, then the probability that they occur together is straightforward to compute. But if $\mathcal{A}$ and $\mathcal{B}$ are not independent, then knowing that one event has occurred can have a significant effect on the probability the other will occur. Here are two extreme examples. If $\mathcal{A}$ and $\mathcal{B}$ are the same, then knowing that $\mathcal{A}$ occurred means you know that $\mathcal{B}$ occurred, too. If $\mathcal{A} = \mathcal{B}^c$, then knowing that $\mathcal{A}$ occurred means you know that $\mathcal{B}$ did not occur. A less extreme example appears below.

Worked example 3.25 (The Probability of Events That Are Not Independent) You throw a fair six-sided die twice and add the numbers. First, compute the probability of getting a number less than six. Second, imagine you *know* that the first die came up three. Compute the probability the sum will be less than six. Third, imagine you *know* that the first die came up four. Compute the probability the sum will be less than six. Finally, imagine you *know* that the first die came up one. Compute the probability the sum will be less than six.

Solution The probability of getting a number less than six is $\frac{10}{36}$. If the first die comes up three, then the question is what is the probability of getting a number less than three on the second die, which is $\frac{1}{3}$. If the first die comes up four, then the question is what is the probability of getting a number less than two on the second die, which is $\frac{1}{6}$. Finally, if the first die comes up one, then the question is what is the probability of getting a number less than five on the second die, which is $\frac{2}{3}$.

Notice how, in Worked example 3.25, knowing what happened to the first die can have a significant effect on the probability of the event.

Definition 3.4 (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

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

From the examples, it should be clear to you that for some cases $P(\mathcal{B}|\mathcal{A})$ is the same as $P(\mathcal{B})$, and for other cases it is not.

3.4.1 Evaluating Conditional Probabilities

To get an expression for $P(\mathcal{B}|\mathcal{A})$, 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 $\mathcal{B} \cap \mathcal{A}$ or in $\mathcal{B}^c \cap \mathcal{A}$, and it cannot lie in both. 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 outcomes in $\mathcal{C} \cap \mathcal{A}$ will appear k times as often as outcomes in $\mathcal{B} \cap \mathcal{A}$. But this must apply even if we know in advance that the outcome is in $\mathcal{A}$. This means that, if $P(\mathcal{C} \cap \mathcal{A}) = kP(\mathcal{B} \cap \mathcal{A})$, then $P(\mathcal{C}|\mathcal{A}) = kP(\mathcal{B}|\mathcal{A})$. In turn, 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

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

so that

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

I find the "size" metaphor helpful here. We have that $P(\mathcal{B}|\mathcal{A})$ measures the probability that an outcome is in $\mathcal{B}$, given we *know* it is in $\mathcal{A}$. From the "size" perspective, $P(\mathcal{B}|\mathcal{A})$ measures the "size" of $(\mathcal{A} \cap \mathcal{B})$ *relative to* $\mathcal{A}$. So our expression makes sense, because the fraction of the event $\mathcal{A}$ that is also part of the event $\mathcal{B}$ is given by the "size" of the intersection divided by the "size" of $\mathcal{A}$.

Another, very useful, way to write the expression $P(\mathcal{B}|\mathcal{A}) = P(\mathcal{B} \cap \mathcal{A})/P(\mathcal{A})$ 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 3.26 (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?

(continued)

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 $P(\mathcal{L}) = 12/1002$.
- Same argument yields $P(\mathcal{B}) = 2/1002$.
- Write $\mathcal{B}$ for the event the car comes from factory B. I need $P(\mathcal{B}|\mathcal{L}) = P(\mathcal{L} \cap \mathcal{B})/P(\mathcal{L}) = P(\mathcal{L}|\mathcal{B})P(\mathcal{B})/P(\mathcal{L})$. I have $P(\mathcal{L}|\mathcal{B})P(\mathcal{B})/P(\mathcal{L}) = (1 \times 2/1002)/(12/1002) = 1/6$.

Worked example 3.27 (Royal Flushes in Poker—1) 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

$$\binom{52}{5} = 2{,}598{,}960$$

so we have

$$\frac{4}{2{,}598{,}960} = \frac{1}{649{,}740}.$$

Worked example 3.28 (Royal Flushes in Poker—2) 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. What is the conditional probability of getting a royal flush, conditioned on the event that this 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 3.29 (Royal Flushes in Poker—3) 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 your 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 Now consider the events

$$\mathcal{A} = \text{you get a royal flush } and \text{ the last card}$$
$$\text{is the ace of spades}$$

and

$$\mathcal{B} = \text{the last card you get is the ace of spades,}$$

and the expression

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

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

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

This is

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

yielding

$$P(\mathcal{A}|\mathcal{B}) = \frac{1}{249,900}.$$

Notice the interesting part: seeing this card has really made a difference.

Worked example 3.30 (Two Dice) We throw two fair six-sided dice. 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 Write the event that the first die comes up 5 as $\mathcal{F}$, and the event the sum is greater than six as $\mathcal{S}$. There are five outcomes where the first die comes up 5 and the number is greater than 6, so $P(\mathcal{F} \cap \mathcal{S}) = 5/36$. Now

$$P(\mathcal{S}|\mathcal{F}) = P(\mathcal{F} \cap \mathcal{S})/P(\mathcal{F}) = (5/36)/(1/6)$$
$$= 5/6.$$

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, because $P(\mathcal{A}) = P(\mathcal{A} \cap \mathcal{B}) + P(\mathcal{A} \cap \mathcal{B}^c)$, we have

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

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

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

It is wise to be suspicious of your intuitions when thinking about problems in conditional probability. There is a really big difference between $P(\mathcal{A}|\mathcal{B})$ and $P(\mathcal{B}|\mathcal{A})P(\mathcal{A})$. Not respecting this difference can lead to serious problems (Sect. 3.4.4), and seems to be easy to do. The division sign in the expression

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

can have alarming effects; as a result, most people have quite poor intuitions about conditional probability.

Remember this: *Here is one helpful example. If you buy a lottery ticket ($\mathcal{L}$), the probability of winning ($\mathcal{W}$) is small. So $P(\mathcal{W}|\mathcal{L})$ may be very small. But $P(\mathcal{L}|\mathcal{W})$ is 1—the winner is always someone who bought a ticket.*

Useful Facts 3.3 (Conditional Probability Formulas)
You should remember the following formulas:

- $P(\mathcal{B}|\mathcal{A}) = \dfrac{P(\mathcal{A}|\mathcal{B})P(\mathcal{B})}{P(\mathcal{A})}$
- $P(\mathcal{A}) = P(\mathcal{A}|\mathcal{B})P(\mathcal{B}) + P(\mathcal{A}|\mathcal{B}^c)P(\mathcal{B}^c)$
- Assume (a) $\mathcal{B}_i \cap \mathcal{B}_j = \varnothing$ for $i \neq j$ and (b) $\mathcal{A} \cap (\cup_i \mathcal{B}_i) = \mathcal{A}$; then $P(\mathcal{A}) = \sum_i P(\mathcal{A}|\mathcal{B}_i)P(\mathcal{B}_i)$

3.4.2   Detecting Rare Events Is Hard

It is hard to detect rare events. This nuisance is exposed by conditional probability reasoning. I have set these examples in a medical framework, but the problem occurs in pretty much any application domain. The issue comes up again and again in discussions of screening tests for diseases. Two recent important controversies have been around whether screening mammograms are a good idea, and whether screening for prostate cancer is a good idea. There is an important issue here. There are real harms that occur when a test falsely labels a patient as ill. First, the patient is distressed and frightened. Second, necessary medical interventions might be quite unpleasant and dangerous. This means it takes thought to tell whether screening does more good (by finding and helping sick people) than harm (by frightening and hurting well people).

Worked example 3.31 (False Positives)   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 that you actually have the disease?

Solution   Write $\mathcal{S}$ for the event you are sick and $\mathcal{R}$ for the event the test reports you are sick. We need $P(\mathcal{S}|\mathcal{R})$. We have

$$P(\mathcal{S}|\mathcal{R}) = \frac{P(\mathcal{R}|\mathcal{S})P(\mathcal{S})}{P(\mathcal{R})}$$

$$= \frac{P(\mathcal{R}|\mathcal{S})P(\mathcal{S})}{P(\mathcal{R}|\mathcal{S})P(\mathcal{S}) + P(\mathcal{R}|\mathcal{S}^c)P(\mathcal{S}^c)}$$

$$= \frac{0.95 \times 1e - 5}{0.95 \times 1e - 5 + 1e - 3 \times (1 - 1e - 5)}$$

$$= 0.0094$$

which should strike you as being a bit alarming. Notice what is happening here. There are two ways that the test could come back positive: either you have the disease, or the test is producing a false positive. But the disease is so rare that it's much more likely you have a false positive result than you have the disease.

If you want to be strongly confident you have detected a very rare event, you need an extremely accurate detector. The next example shows how to compute how accurate the detector needs to be. The degree of accuracy required is often well beyond anything current technologies can reach. You should remember this example the next time someone tells you their test is, say, 90% accurate—such a test could also be completely useless.

Worked example 3.32 (False Positives −2) You want to *design* 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 $\mathcal{S}$ for the event you are sick and $\mathcal{R}$ for the event the test reports you are sick. We need $P(\mathcal{S}|\mathcal{R})$. We have

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

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.

3.4.3 Conditional Probability and Various Forms of Independence

Two events are independent if
$$P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B}).$$

In turn, if two events $\mathcal{A}$ and $\mathcal{B}$ are independent, then

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

and

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

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.

Useful Facts 3.4 (Conditional Probability for Independent Events)
If two events $\mathcal{A}$ and $\mathcal{B}$ are independent, then
$$P(\mathcal{A}|\mathcal{B}) = P(\mathcal{A})$$

and

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

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. There are weaker kinds of independence that are sometimes useful.

Definition 3.5 (Pairwise Independence) Events $\mathcal{A}_1 \ldots \mathcal{A}_n$ are **pairwise independent** if each pair is independent (i.e. $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent, etc.).

Worked example 3.33 (Pairwise Independence is a Weaker Property than Independence) This means that you can have events that are pairwise independent, but not independent. 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.

Definition 3.6 (Conditional Independence) Events $\mathcal{A}_1 \ldots \mathcal{A}_n$ are **conditionally independent** conditioned on event $\mathcal{B}$ if

$$P(\mathcal{A}_1 \cap \ldots \cap \mathcal{A}_n | \mathcal{B}) = P(\mathcal{A}_1 | \mathcal{B}) \ldots P(\mathcal{A}_n | \mathcal{B})$$

Worked example 3.34 (Cards and Conditional Independence) We remove a red 10 and a red 6 from a standard deck of playing cards. We shuffle the remaining cards, and draw one card. Write $\mathcal{A}$ for the event that the card drawn is a 10, $\mathcal{B}$ for the event the card drawn is red, and $\mathcal{C}$ for the event that the card drawn is either a 10 or a 6. Show that $\mathcal{A}$ and $\mathcal{B}$ are not independent, but are conditionally independent conditioned on $\mathcal{C}$.

Solution We have $P(\mathcal{A}) = 3/50$, $P(\mathcal{B}) = 24/50$, $P(\mathcal{A} \cap \mathcal{B}) = 1/50$, so

$$P(\mathcal{A}|\mathcal{B}) = \frac{1/50}{24/50} = \frac{1}{24} \neq P(\mathcal{A})$$

so $\mathcal{A}$ and $\mathcal{B}$ are not independent. We have also that $P(\mathcal{A}|\mathcal{C}) = 1/2$ and $P(\mathcal{B}|\mathcal{C}) = 2/6 = 1/3$. Now

$$P(\mathcal{A} \cap \mathcal{B}|\mathcal{C}) = 1/6 = P(\mathcal{A}|\mathcal{C})P(\mathcal{B}|\mathcal{C})$$

so $\mathcal{A}$ and $\mathcal{B}$ are conditionally independent conditioned on $\mathcal{C}$.

3.4.4 Warning Example: The Prosecutor's Fallacy

Treat conditional probability with great care, because the topic confuses a lot of people, even people you might expect not to be confused. One important mistake is the **prosecutor's fallacy**, which has a name because it's such a common error. A prosecutor has evidence $\mathcal{E}$ against a suspect. Write $\mathcal{I}$ for the event that the suspect is innocent. Things get interesting when $P(\mathcal{E}|\mathcal{I})$ is small. The prosecutor argues, incorrectly, that the suspect must be guilty, because $P(\mathcal{E}|\mathcal{I})$ is so small. The argument is incorrect because $P(\mathcal{E}|\mathcal{I})$ is irrelevant to the issue. What matters is $P(\mathcal{I}|\mathcal{E})$, which is the probability you are innocent, given the evidence.

The distinction is very important, because $P(\mathcal{I}|\mathcal{E})$ could be big even if $P(\mathcal{E}|\mathcal{I})$ is small. In the expression

$$P(\mathcal{I}|\mathcal{E}) = \frac{P(\mathcal{E}|\mathcal{I})P(\mathcal{I})}{P(\mathcal{E})}$$

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

notice that if $P(\mathcal{I})$ is large or if $P(\mathcal{E}|\mathcal{I}^c)$ is much smaller than $P(\mathcal{E}|\mathcal{I})$, then $P(\mathcal{I}|\mathcal{E})$ could be close to one even if $P(\mathcal{E}|\mathcal{I})$ is small.

This fallacy can be made even more mischievous. Assume the prosecutor incorrectly adopts a model that items of evidence are independent (or even just conditionally independent, conditioned on $\mathcal{I}$) when they're not. Then this model could result in an estimate of $P(\mathcal{E}|\mathcal{I})$ that is much smaller than it should be.

The prosecutor's fallacy has contributed to a variety of miscarriages of justice, with real, and shocking, consequences. One famous incident occurred in the UK, involving a mother, Sally Clark, who was convicted of murdering two of her children. Expert evidence by paediatrician Roy Meadow argued that the probability of both deaths resulting from Sudden Infant Death Syndrome was extremely small. Her first appeal cited, among other grounds, statistical error in the evidence. The appeals court rejected this appeal, calling the statistical point "a sideshow". This prompted a great deal of controversy, both in the public press and various professional journals, including a letter from the then president of the Royal Statistical Society to the Lord Chancellor, pointing out that "*statistical evidence . . . (should be) . . . presented only by appropriately qualified statistical experts*". A second appeal (on other grounds) followed, and was successful. The appellate judges specifically criticized the statistical evidence, although it was not a point of appeal. Clark never recovered from this horrific set of events and died in tragic circumstances shortly after the second appeal. Roy Meadow was then struck off the rolls for serious professional misconduct as an expert witness, a ruling he appealed successfully. You can find a more detailed account of this case, with pointers to important documents including the letter to the Lord Chancellor (which is well worth reading), at http://en.wikipedia.org/wiki/Roy_Meadow; there is further material on the prosecutors fallacy at http://en.wikipedia.org/wiki/Prosecutor%27s_fallacy.

This story is not just about problems with the criminal law. There is a very significant difference between the meaning of $P(\mathcal{E}|\mathcal{I})$ and the meaning of $P(\mathcal{I}|\mathcal{E})$. When you use conditional probabilities, you need to be sure which one is important to you.

> **Remember this:** *You need to be careful reasoning about conditional probability and about independent events. These topics mislead intuition so regularly that some errors have names. Be very careful.*

3.4.5 Warning Example: The Monty Hall Problem

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. The goats are interchangeable, and, for reasons of your own, you would prefer the car to a goat. 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?

This problem is known as the Monty Hall problem, and is a relatively simple exercise in conditional probability. But careless thinking about probability, particularly conditional probability, can cause wonderful confusion. The Monty Hall problem has been the subject of extensive, lively, and often quite inaccurate correspondence in various national periodicals—it seems to catch the attention, which is why I describe it in some detail.

Notice that you cannot tell what to do *using the information provided*, by the following argument. Label the door you chose at the start of the game 1; 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.

Here are some 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.

It should be straightforward for you to come up with other possible rules. 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). This means we are interested in

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

Notice that each of these rules is consistent with your observations—what you saw could have occurred under any of these rules. You have to know which rule the host uses to proceed. You should be aware that in many of the discussions of this problem, people assume without comment that the host uses rule 2, then proceed with this assumption.

Worked example 3.35 (Monty Hall, Rule One) Assume the host uses rule one, and shows you a goat behind door two. What is $P(C_1|G_2, r_1)$?

Solution To work this out, we need to know $P(G_2|C_1, r_1)$, $P(G_2|C_2, r_1)$ and $P(G_2|C_3, r_1)$. Now $P(G_2|C_2, r_1)$ must be zero, because the host could not reveal a goat behind door two if there was a car behind that door. Write O_2 for the event the host *chooses* to open door two, and B_2 for the event there happens to be a goat behind door two. These two events are independent—the host chose the door uniformly at random. We can compute

$$P(G_2|C_1, r_1) = P(O_2 \cap B_2|C_1, r_1)$$

$$= P(O_2|C_1, r_1)P(B_2|C_1, r_1)$$

$$= (1/3)(1)$$

$$= 1/3$$

where $P(B_2|C_1, r_1) = 1$ because we conditioned on the fact there was a car behind door one, so there is a goat behind each other door. This argument establishes $P(G_2|C_3, r_1) = 1/3$, too. So $P(C_1|G_2, r_1) = 1/2$—the host showing you the goat does not motivate you to do anything, because if $P(C_1|G_2, r_1) = 1/2$, then $P(C_3|G_2, r_1) = 1/2$, too—there's nothing to choose between the two closed doors.

Worked example 3.36 (Monty Hall, Rule Two) Assume the host uses rule two, and shows you a goat behind door two. What is $P(C_1|G_2, r_2)$?

Solution To work this out, we need to know $P(G_2|C_1, r_2)$, $P(G_2|C_2, r_2)$ and $P(G_2|C_3, r_2)$. Now $P(G_2|C_2, r_2) = 0$, because the host chooses from doors with goats behind them. $P(G_2|C_1, r_2) = 1/2$, because the host chooses uniformly and at random from doors with goats behind them that are not door one; if the car is behind door one, there are two such doors. $P(G_2|C_3, r_2) = 1$, because there is only one door that (a) has a goat behind it and (b) isn't door one. Plug these numbers into the formula, to get $P(C_1|G_2, r_2) = 1/3$. This is the source of all the fuss. It says

that, if you know the host is using rule two, you should switch doors if the host shows you a goat behind door two (because $P(C_3|G_2, r_2) = 2/3$).

Notice what is happening: if the car is behind door three, then the *only* choice of goat for the host is the goat behind two. So by choosing a door under rule two, the host is signalling some information to you, which you can use. By using rule three, the host can tell you precisely where the car is (exercises).

Many people find the result of Example 3.36 counterintuitive. Each time I've taught this material, I've had lively discussions with students and with teaching assistants. Some people object to the extent of newspaper columns, letters to the editor, arguments on the internet, etc. One example that some people find helpful is an extreme case. Imagine that, instead of three doors, there are 1002. The host is using rule two, modified in the following way: open all but one of the doors that are not door one, choosing only doors that have goats behind them to open. You choose door one; the host opens 1000 doors—say, all but doors one and 1002. What would you do?

3.5   Extra Worked Examples

3.5.1   Outcomes and Probability

Worked example 3.37 (Children) A co-uple 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 3.38 (Monty Hall (Sigh!) with Indistinguishable Goats) 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 for our purposes; equivalently, we do not care about the difference between goats. What is the sample space?

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

Worked example 3.39 (Monty Hall with Distinguishable Goats) 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 is the sample space?

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.

Worked example 3.40 (Find the Lady, with Even Probabilities) Recall the problem of Worked example 3.1. 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 3.41 (Monty Hall, Indistinguishable Goats, Even Probabilities)   Recall the problem of Worked example 3.39. Each outcome has the same probability. We choose to open the first box. With what probability will we find a goat (any goat)?

Solution   There are three outcomes, each has the same probability, and two give a goat, so $2/3$

Worked example 3.42 (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   There are three places the car could be, each has the same probability, so $1/3$

Worked example 3.43 (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   Using the reasoning of the previous example, but substituting "female goat" for "car", $1/3$. 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 a good idea to omit it from the sample space.

3.5.2   Events

Worked example 3.44 (Drawing a Red Ten) I shuffle a standard pack of cards, and draw one card. What is the probability that it is a red ten?

Solution   There are 52 cards, and each is an outcome. Two of these outcomes are red tens; so we have $2/52 = 1/26$.

Worked example 3.45 (Birthdays in Succession) We stop three people at random, and ask the day of the week on which they are born. What is the probability that they are born on 3 days of the week in succession (for example, the first on Monday; the second on Tuesday; the third on Wednesday; or Saturday-Sunday-Monday; and so on).

Solution   We assume that births are equally common on each day of the week. The space of outcomes consists of triples of days, and each outcome has the same probability. The event is the set of triples of 3 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

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

Worked example 3.46 (Shared Birth-days)   We stop two people at random. What is the probability that they were born on the same day of the week?

(continued)

Solution The day the first person was born doesn't matter; the probability the second person was born on that day is $1/7$. Or you could count outcomes explicitly to get

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

Worked example 3.47 (Children—3) *This example is a version of example 1.12, p44, in Stirzaker, "Elementary Probability".* A couple decides to have children. They decide to have children until there is one of each gender, or until there are three, and then stop. Assume that each birth results in one child, and each gender is equally likely at each birth. Let $\mathcal{B}_i$ be the event that there are i boys, and $\mathcal{C}$ be the event there are more girls than boys. Compute $P(\mathcal{B}_1)$ and $P(\mathcal{C})$.

Solution We could write the outcomes as $\{GB, BG, GGB, GGG, BBG, BBB\}$. 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 the device of the fictitious births again. The important events are $\{GBb, GBg\}$; $\{BGb, BGg\}$; $\{GGB\}$; $\{GGG\}$; $\{BBG\}$; and $\{BBB\}$. Like this, we get $P(\mathcal{B}_1) = 5/8$ and $P(\mathcal{C}) = 1/4$.

3.5.3 Independence

Worked example 3.48 (Children) A couple decides to have two children. Genders are assigned to children at random, fairly, at birth and independently at each birth (our models have to abstract a little!). What is the probability of having a boy and then a girl?

Solution

$$P(\text{first is boy} \cap \text{second is girl})$$
$$= (1/2)(1/2)$$
$$= 1/4$$

Worked example 3.49 (Programs) We sample the processes on a computer at random intervals. Write $\mathcal{A}$ for the event that program A is observed to be running in a sample, $\mathcal{B}$ for the event that program B is observed to be running in a sample, and $\mathcal{N}$ for the (Nasty) event that program C is observed to be behaving badly in a sample. We find $P(\mathcal{A} \cap \mathcal{N}) = 0.07$; $P(\mathcal{B} \cap \mathcal{N}) = 0.05$; $P(\mathcal{A} \cap \mathcal{B} \cap \mathcal{N}) = 0.04$; and $P(\mathcal{N}) = 0.1$. Are $\mathcal{A}$ and $\mathcal{B}$ conditionally independent conditioned on $\mathcal{N}$?

Solution This is a straightforward calculation. You should get $P(\mathcal{A}|\mathcal{N}) = 0.7$; $P(\mathcal{B}|\mathcal{N}) = 0.5$; $P(\mathcal{A} \cap \mathcal{B}|\mathcal{N}) = 0.4$; and so $P(\mathcal{A} \cap \mathcal{B}|\mathcal{N}) \neq P(\mathcal{A}|\mathcal{N}) \times P(\mathcal{B}|\mathcal{N})$, and they are not conditionally independent—there is some form of interaction here.

Worked example 3.50 (Independent Test Results)   You have a blood test for a rare disease. We study the effect of repeated tests. Write $\mathcal{S}$ for the event that the patient is sick; $\mathcal{D}_i^+$ for the event that the i'th repetition of the test reports positive; and $\mathcal{D}_i^-$ for the event that the i'th repetition of the test reports negative. The test has $P(\mathcal{D}^+|\mathcal{S}) = 0.8$ and $P(\mathcal{D}^-|\overline{\mathcal{S}}) = 0.8$, and $P(\mathcal{S}) = 1e - 5$. This blood test has the property that, if you repeat the test, results are conditionally independent conditioned on the true result, meaning that $P(\mathcal{D}_1^+ \cap \mathcal{D}_2^+|\overline{\mathcal{S}}) = P(\mathcal{D}_1^+|\overline{\mathcal{S}})P(\mathcal{D}_2^+|\overline{\mathcal{S}})$. Assume you test positive once; twice; and ten times. In each case, what is the posterior probability that you are sick?

Solution I will work the case for two positive tests. We need $P(\mathcal{S}|\mathcal{D}_1^+ \cap \mathcal{D}_2^+)$. We have

$$P(\mathcal{S}|\mathcal{D}_1^+ \cap \mathcal{D}_2^+) = \frac{P(\mathcal{D}_1^+ \cap \mathcal{D}_2^+|\mathcal{S})P(\mathcal{S})}{P(\mathcal{D}_1^+ \cap \mathcal{D}_2^+)}$$

$$= \frac{P(\mathcal{D}_1^+ \cap \mathcal{D}_2^+|\mathcal{S})P(\mathcal{S})}{P(\mathcal{D}_1^+ \cap \mathcal{D}_2^+|\mathcal{S})P(\mathcal{S}) + P(\mathcal{D}_1^+ \cap \mathcal{D}_2^+|\overline{\mathcal{S}})P(\overline{\mathcal{S}})}$$

$$= \frac{0.8 \times 0.8 \times 1e{-}5}{0.8 \times 0.8 \times 1e{-}5 + 0.2 \times 0.2 \times (1{-}1e{-}5)}$$

$$\approx 1.6e - 4.$$

You should check that once yields a posterior of approximately 4e-5, and ten times yields a posterior of approximately 0.91. This isn't an argument for repeating tests; rather, you should regard it as an indication of how implausible the assumption of conditional independence of test results is.

3.5.4   Conditional Probability

Worked example 3.51 (Card Games)   You have two decks of 52 standard playing cards. One has been shuffled properly. The other is organized as 26 black cards, then 26 red cards. You are shown one card from one deck, which turns out to be black; what is the posterior probability that you have a card from the shuffled deck?

Solution Write $\mathcal{S}$ for the event the card comes from the shuffled deck, and $\mathcal{B}$ the event you are given a black card. We want

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

$$= \frac{P(\mathcal{B}|\mathcal{S})P(\mathcal{S})}{P(\mathcal{B}|\mathcal{S})P(\mathcal{S}) + P(\mathcal{B}|\overline{\mathcal{S}})P(\overline{\mathcal{S}})}$$

$$= \frac{(1/2) \times (1/2)}{(1/2) \times (1/2) + 1 \times (1/2)}$$

$$= 1/3$$

Worked example 3.52 (Finding a Common Disease) A disease occurs with probability 0.4 (i.e. it is present in 40% of the population). You have a test that detects the disease with probability 0.6, and produces a false positive with probability 0.1. What is the posterior probability you have the disease if the test comes back positive?

Solution Write $\mathcal{S}$ for the event you are sick, and $\mathcal{P}$ for the event the test comes back positive. We want

$$
\begin{aligned}
P(\mathcal{S}|\mathcal{P}) &= \frac{P(\mathcal{P}|\mathcal{S})P(\mathcal{S})}{P(\mathcal{P})} \\
&= \frac{P(\mathcal{P}|\mathcal{S})P(\mathcal{S})}{P(\mathcal{P}|\mathcal{S})P(\mathcal{S}) + P(\mathcal{P}|\overline{\mathcal{S}})P(\overline{\mathcal{S}})} \\
&= \frac{0.6 \times 0.4}{0.6 \times 0.4 + 0.1 \times 0.6} \\
&= 0.8
\end{aligned}
$$

Notice that if the disease is quite common, even a rather weak test is helpful.

Worked example 3.53 (Which Disease Do You Have?) Disease A occurs with probability 0.1 (i.e. it is present in 20% of the population), and disease B occurs with probability 0.2. It is not possible to have both diseases. You have a single test. This test reports positive with probability 0.8 for a patient with disease A, with probability 0.5 for a patient with disease B, and with probability 0.01 for a patient with no disease. What is the posterior probability you have either disease, or neither, if the test comes back positive?

Solution We are interested in $\mathcal{A}$ (the event you have disease A), $\mathcal{B}$ (the event you have disease B), and $\mathcal{W}$ (the event you are well). Write $\mathcal{P}$ for the event the test comes back positive. We want $P(\mathcal{A}|\mathcal{P})$, $P(\mathcal{B}|\mathcal{P})$ and $P(\mathcal{W}|\mathcal{P}) = 1 - P(\mathcal{A}|\mathcal{P}) - P(\mathcal{B}|\mathcal{P})$. We have

$$
\begin{aligned}
P(\mathcal{A}|\mathcal{P}) &= \frac{P(\mathcal{P}|\mathcal{A})P(\mathcal{A})}{P(\mathcal{P})} \\
&= \frac{P(\mathcal{P}|\mathcal{A})P(\mathcal{A})}{P(\mathcal{P}|\mathcal{A})P(\mathcal{A}) + P(\mathcal{P}|\mathcal{B})P(\mathcal{B}) + P(\mathcal{P}|\mathcal{W})P(\mathcal{W})} \\
&= \frac{0.8 \times 0.1}{0.8 \times 0.1 + 0.5 \times 0.2 + 0.01 \times 0.7} \\
&\approx 0.43
\end{aligned}
$$

A similar calculation yields $P(\mathcal{B}|\mathcal{P})) \approx 0.53$ and $P(\mathcal{W}|\mathcal{P}) \approx 0.04$. The low probability of a false positive means that a positive result very likely comes from some disease. Even though the test isn't particularly sensitive to disease B, the fact B is twice as common as A means a positive result is somewhat more likely to have come from B than from A.

Worked example 3.54 (Fraud or Psychic Powers?) You want to investigate the powers of a putative psychic. You blindfold this person, then flip a fair coin 10 times. Each time, the subject correctly tells you whether it came up heads or tails. There are three possible explanations: chance, fraud, or psychic powers. What is the posterior probability of each, conditioned on the evidence.

Solution We have to do some modelling here. We must choose reasonable numbers for the prior of chance ($\mathcal{C}$), fraud ($\mathcal{F}$), and psychic powers ($\mathcal{P}$). There's little reliable evidence for psychic powers to date, so we can choose $P(\mathcal{P}) = 2\epsilon$ (where ϵ is a very small number), and allocate the remaining probability evenly between $\mathcal{C}$ and $\mathcal{F}$. Write $\mathcal{E}$ for the event the subject correctly calls 10 flips of a fair coin. We have $P(\mathcal{E}|\mathcal{C}) = (1/2)^{10}$. Assume that fraud and psychic powers are efficient, so that $P(\mathcal{E}|\mathcal{F}) = P(\mathcal{E}|\mathcal{P}) = 1$. Then we have

$$
\begin{aligned}
P(\mathcal{P}|\mathcal{E}) &= \frac{P(\mathcal{E}|\mathcal{P})P(\mathcal{P})}{P(\mathcal{E}|\mathcal{P})P(\mathcal{P}) + P(\mathcal{E}|\mathcal{C})P(\mathcal{C}) + P(\mathcal{E}|\mathcal{F})P(\mathcal{F})} \\
&= \frac{2\epsilon}{2\epsilon + (1/2)^{10} \times (0.5 - \epsilon) + (0.5 - \epsilon)} \\
&\approx 4\epsilon
\end{aligned}
$$

and $P(\mathcal{F}|\mathcal{E})$ is rather close to 1. I'd check how well the blindfold works; it's a traditional failure point in experiments like this.

3.6   You Should

3.6.1   Remember These Definitions

3.6.2   Remember These Terms

3.6.3   Remember and Use These Facts

3.6.4   Remember These Points

3.6.5 Be Able to

- Write out a set of outcomes for an experiment.
- Construct an event space.
- Compute the probabilities of outcomes and events.
- Determine when events are independent.
- Compute the probabilities of outcomes by counting events, when the count is straightforward.
- Compute a conditional probability.

Problems

Outcomes

3.1 You roll a four sided die. What is the space of outcomes?

3.2 King Lear decides to allocate three provinces (1, 2, and 3) to his daughters (Goneril, Regan and Cordelia—read the book) at random. Each gets one province. What is the space of outcomes?

3.3 You randomly wave a flyswatter at a fly. What is the space of outcomes?

3.4 You read the book, so you know that King Lear had family problems. As a result, he decides to allocate two provinces to one daughter, one province to another daughter, and no provinces to the third. Because he's a bad problem solver, he does so at random. What is the space of outcomes?

The Probability of an Outcome

3.5 You roll a fair four sided die. What is the probability of getting a 3?

3.6 You shuffle a standard deck of playing cards and draw a card. What is the probability that this is the king of hearts?

3.7 A roulette wheel has 36 slots numbered 1–36. Of these slots, the odd numbers are red and the even numbers are black. There are two slots numbered zero, which are green. The croupier spins the wheel, and throws a ball onto the surface; the ball bounces around and ends up in a slot (which is chosen fairly and at random). What is the probability the ball ends up in slot 2?

Events

3.8 At a particular University, 1/2 of the students drink alcohol and 1/3 of the students smoke cigarettes.

(a) What is the largest possible fraction of students who do neither?
(b) It turns out that, in fact, 1/3 of the students do neither. What fraction of the students does both?

Computing Probabilities by Counting Outcomes

3.9 Assume each outcome in Ω has the same probability. In this case, show

$$P(\mathcal{E}) = \frac{\text{Number of outcomes in } \mathcal{E}}{\text{Total number of outcomes in } \Omega}$$

3.10 You roll a fair four sided die, and then a fair six sided die. You add the numbers on the two dice. What is the probability the result is even?

3.11 You roll a fair 20 sided die. What is the probability of getting an even number?

3.12 You roll a fair five sided die. What is the probability of getting an even number?

3.13 *I am indebted to Amin Sadeghi for this exercise.* You must sort four balls into two buckets. There are two white, one red and one green ball.

(a) For each ball, you choose a bucket independently and at random, with probability $\frac{1}{2}$. Show that the probability each bucket has a colored ball in it is $\frac{1}{2}$.
(b) You now choose to sort these balls in such a way that each bucket has two balls in it. You can do so by generating a permutation of the balls uniformly and at random, then placing the first two balls in the first bucket and the second two balls in the second bucket. Show that there are 16 permutations where there is one colored ball in each bucket.
(c) Use the results of the previous step to show that, using the sorting procedure of that step, the probability of having a colored ball in each bucket is $\frac{2}{3}$.
(d) Why do the two sorting procedures give such different outcomes?

The Probability of Events

3.14 You flip a fair coin three times. What is the probability of seeing HTH? (i.e. Heads, then Tails, then Heads)

3.15 You shuffle a standard deck of playing cards and draw a card.

(a) What is the probability that this is a king?
(b) What is the probability that this is a heart?
(c) What is the probability that this is a red card (i.e. a heart or a diamond)?

3.16 A roulette wheel has 36 slots numbered 1–36. Of these slots, the odd numbers are red and the even numbers are black. There are two slots numbered zero, which are green. The croupier spins the wheel, and throws a ball onto the surface; the ball bounces around and ends up in a slot (which is chosen fairly and at random).

(a) What is the probability the ball ends up in a green slot?
(b) What is the probability the ball ends up in a red slot with an even number?
(c) What is the probability the ball ends up in a red slot with a number divisible by 7?

3.17 You flip a fair coin three times. What is the probability of seeing two heads and one tail?

3.18 You remove the king of hearts from a standard deck of cards, then shuffle it and draw a card.

(a) What is the probability this card is a king?
(b) What is the probability this card is a heart?

3.19 You shuffle a standard deck of cards, then draw four cards.

(a) What is the probability all four are the same suit?
(b) What is the probability all four are red?
(c) What is the probability each has a different suit?

3.20 You roll three fair six-sided dice and add the numbers. What is the probability the result is even?

3.21 You roll three fair six-sided dice and add the numbers. What is the probability the result is even *and* not divisible by 20?

3.22 You shuffle a standard deck of cards, then draw seven cards. What is the probability that you see no aces?

3.23 Show that $P(\mathcal{A} - (\mathcal{B} \cup \mathcal{C})) = P(\mathcal{A}) - P(\mathcal{A} \cap \mathcal{B}) - P(\mathcal{A} \cap \mathcal{C}) + P(\mathcal{A} \cap \mathcal{B} \cap \mathcal{C})$.

3.24 You draw a single card from a standard 52 card deck. What is the probability that it is red?

3.25 You remove all heart cards from a standard 52 card deck, then draw a single card from the result.

(a) What is the probability that the card you draw is a red king?
(b) What is the probability that the card you draw is a spade?

Permutations and Combinations

3.26 You shuffle a standard deck of playing cards, and deal a hand of 10 cards. With what probability does this hand have five red cards?

3.27 Magic the Gathering is a popular card game. Cards can be land cards, or other cards. We consider a game with two players. Each player has a deck of 40 cards. Each player shuffles their deck, then deals seven cards, called their *hand*.

(a) Assume that player one has 10 land cards in their deck and player two has 20. With what probability will *each* player have four lands in their hand?
(b) Assume that player one has 10 land cards in their deck and player two has 20. With what probability will player one have two lands and player two have three lands in hand?
(c) Assume that player one has 10 land cards in their deck and player two has 20. With what probability will player two have more lands in hand than player one?

3.28 The previous exercise divided Magic the Gathering cards into lands vs. other. We now recognize four kinds of cards: land, spell, creature and artifact. We consider a game with two players. Each player has a deck of 40 cards. Each player shuffles their deck, then deals seven cards, called their *hand*.

(a) Assume that player one has 10 land cards, 10 spell cards, 10 creature cards and 10 artifact cards in their deck. With what probability will player one have at least one of each kind of card in hand?
(b) Assume that player two has 20 land cards, 5 spell cards, 7 creature cards and 8 artifact cards in their deck. With what probability will player two have at least one of each kind of card in hand?
(c) Assume that player one has 10 land cards, 10 spell cards, 10 creature cards and 10 artifact cards in their deck;. and player two has 20 land cards, 5 spell cards, 7 creature cards and 8 artifact cards in their deck. With what probability will at least one of the players have at least one of each kind card in hand?

3.29 You take a standard deck of 52 playing cards and shuffle it. Compute the probability that, in the shuffled deck, there is at least one pair of cards following one another in increasing order (i.e. a 2 followed by a 3, or a 3 followed by a 4, etc.). This isn't particularly easy, but the probability is higher than most people realize; you can surprise your friends and make money with this information.

Independence

3.30 Event $\mathcal{A}$ has $P(\mathcal{A}) = 0.5$. Event $\mathcal{B}$ has $P(\mathcal{B}) = 0.2$. We also know that $P(\mathcal{A} \cup \mathcal{B}) = 0.65$. Are A and B independent? Why?

3.31 Event $\mathcal{A}$ has $P(\mathcal{A}) = 0.5$. Event $\mathcal{B}$ has $P(\mathcal{B}) = 0.5$. These events are independent. What is $P(\mathcal{A} \cup \mathcal{B})$?

3.32 You take a standard deck of cards, shuffle it, and remove both red kings. You then draw a card.

(a) Is the event {card is red} independent of the event {card is a queen}?
(b) Is the event {card is black} independent of the event {card is a king}?

3.33 You flip a fair coin seven times. What is the probability that you see three H's and two T's?

3.34 An airline sells T tickets for a flight with S seats, where $T > S$. Passengers turn up for the flight independently, and the probability that a passenger with a ticket will turn up for a flight is p_t. The pilot is eccentric, and will fly only if precisely E passengers turn up, where $E < S$. Write an expression for the probability the pilot will fly.

Conditional Probability

3.35 You roll two fair six-sided dice. What is the conditional probability the sum of numbers is greater than three, conditioned on the first die coming up even.

3.36 I claim event $\mathcal{B}$ has probability ϵ, that $P(\mathcal{A}|\mathcal{B}) = 1$, and that $P(\mathcal{B}|\mathcal{A}) = \epsilon/2$. Can such a probability distribution exist?

3.37 You take a standard deck of cards, shuffle it, and remove one card. You then draw a card.

(a) What is the conditional probability that the card you draw is a red king, conditioned on the removed card being a king?
(b) What is the conditional probability that the card you draw is a red king, conditioned on the removed card being a red king?
(c) What is the conditional probability that the card you draw is a red king, conditioned on the removed card being a black ace?

3.38 A royal flush is a hand of five cards, consisting of Ace, King, Queen, Jack and 10 of a single suit. Poker players like this hand, but don't see it all that often.

(a) You draw three cards from a standard deck of playing cards. These are Ace, King, Queen of hearts. What is the probability that the next two cards you draw will result in a getting a royal flush? (this is the conditional probability of getting a royal flush, conditioned on the first three cards being AKQ of hearts).

3.39 You roll a fair five-sided die, and a fair six-sided die.

(a) What is the probability that the sum of numbers is even?
(b) What is the conditional probability that the sum of numbers is even, conditioned on the six-sided die producing an odd number?

3.40 You take a standard deck of playing cards, shuffle it, and remove 13 cards without looking at them. You then shuffle the resulting deck of 39 cards, and draw three cards. Each of these three cards is red. What is the conditional probability that every card you removed is black?

3.41 Magic the Gathering is a popular card game. Cards can be land cards, or other cards. We will consider a deck of 40 cards, containing 10 land cards and 30 other cards. A player shuffles that deck, and draws seven cards but does not look at them. The player then chooses one of these cards at random; it is a land.

(a) What is the conditional probability that the original hand of seven cards is all lands?
(b) What is the conditional probability that the original hand of seven cards contains only one land?

3.42 Magic the Gathering is a popular card game. Cards can be land cards, or other cards. We will consider a deck of 40 cards, containing 10 land cards and 30 other cards. A player shuffles that deck, and draws seven cards but does not look at them. The player then chooses three of these cards at random; each of these three is a land.

(a) What is the conditional probability that the original hand of seven cards is all lands?
(b) What is the conditional probability that the original hand of seven cards contains only three lands?

3.43 You take a standard deck of playing cards, and remove one card at random. You then draw a single card. Write $\mathcal{S}$ for the event that the card you remove is a six. Write $\mathcal{N}$ for the event that the card you remove is not a six. Write $\mathcal{R}$ for the event that the card you remove is red. Write $\mathcal{B}$ for the event the card you remove is black.

(a) Write $\mathcal{A}$ for the event you draw a 6. What is $P(\mathcal{A}|\mathcal{S})$?
(b) Write $\mathcal{A}$ for the event you draw a 6. What is $P(\mathcal{A}|\mathcal{N})$?
(c) Write $\mathcal{A}$ for the event you draw a 6. What is $P(\mathcal{A})$?
(d) Write $\mathcal{D}$ for the event you draw a red six. Are $\mathcal{D}$ and $\mathcal{A}$ independent? why?
(e) Write $\mathcal{D}$ for the event you draw a red six. What is $P(\mathcal{D})$?

3.44 A student takes a multiple choice test. Each question has N answers. If the student knows the answer to a question, the student gives the right answer, and otherwise guesses uniformly and at random. The student knows the answer to 70% of the questions. Write $\mathcal{K}$ for the event a student knows the answer to a question and $\mathcal{R}$ for the event the student answers the question correctly.

(a) What is $P(\mathcal{K})$?
(b) What is $P(\mathcal{R}|\mathcal{K})$?
(c) What is $P(\mathcal{K}|\mathcal{R})$, as a function of N?
(d) What values of N will ensure that $P(\mathcal{K}|\mathcal{R}) > 99\%$?

3.45 Write the event a patient has an illness as $\mathcal{I}$. Write the event that a test reports the patient has the illness as $\mathcal{R}$. Assume that $P(\mathcal{R}|\mathcal{I}^c) = 0.1$. We have that $P(\mathcal{I}|\mathcal{R}) = 0.5$.

(a) Compute $P(\mathcal{I})$ as a function of $P(\mathcal{R}|\mathcal{I})$, and plot it.
(b) What is the smallest possible value of $P(\mathcal{I})$? For what value of $P(\mathcal{R}|\mathcal{I})$ does this occur?
(c) Now plot the smallest possible value of $P(\mathcal{I})$ for different values of $P(\mathcal{R}|\mathcal{I}^c)$, assuming that $P(\mathcal{R}|\mathcal{I}) = 0.99$.

The Monty Hall Problem

3.46 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.

3.47 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.