MAT220: Lecture 7 Handout
Conditional Probability

Contents

Last lecture we introduced probability distributions and random variables. We saw that probabilities do not always come from simply counting equally-likely outcomes. Instead, in a general probability distribution, each outcome may have its own probability attached to it.

In this lecture, we develop *conditional probability.* This means probability after we have learned extra information. Conditional probability can often feel very weird and counterintuitive. However, if we follow the rules of probability carefully, and if we keep track of the sample space, then the calculations will make sense. We will also study decision trees, which give us a practical way to organise multi-step probability problems. Finally, we discuss three famous examples: the Monty Hall problem, the Birthday Paradox, and medical testing.

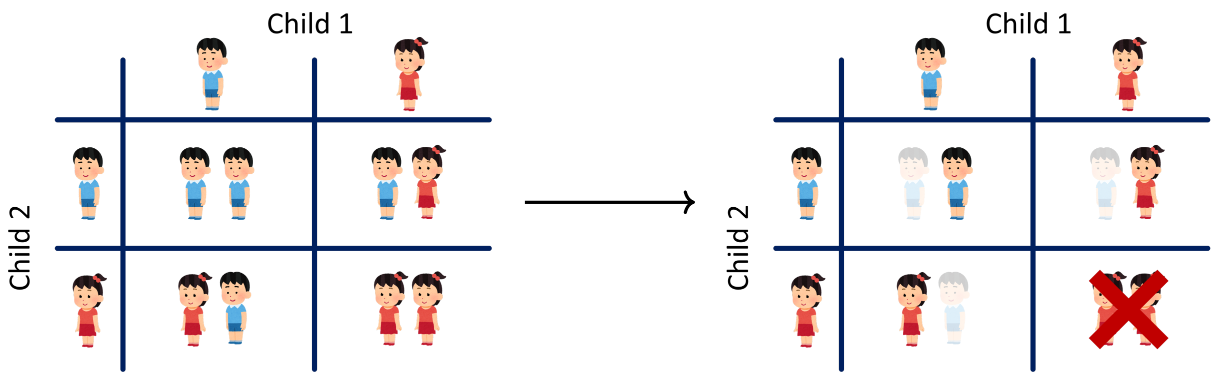
1 Conditional Probability in General

Suppose that you know your neighbour has two children. You go over to their house and knock on the door, and a boy answers the door. What is the probability that the other child is a boy?

Amazingly, the answer is actually $\frac{1}{3}$. This is quite surprising when heard for the first time, however, it is indeed true. Let's break down the problem briefly. Note first that there are four possible ordered outcomes:

$$S = \{BB, BG, GB, GG\}.$$

Naively, we would expect there to be a $\frac{1}{4}$ chance of any of these outcomes. However, by seeing that at least one child is a boy, we receive an information update: we now *know* that the "GG" option is no longer available. This reduces the space of possibilities to:



In other words, the sample space has effectively been reduced to:

$$\{(\text{Boy, Boy}), (\text{Girl, Boy}), (\text{Boy, Girl})\}.$$

Only one of these three outcomes has two boys. Therefore:

$$P(\text{two boys, given at least one boy}) = \frac{1}{3}.$$

We have just demonstrated a calculation of *conditional probability*, which calculates uncertainty based on information updates or extra assumptions. As you can see, the results of conditional probabilities are often surprising and counterintuitive. Today we will introduce a few methods to navigate this potentially confusing mess. We will start with a brief review of some theory that we have seen in the past few lectures, and then we will apply this to some basic computations. After that, we will work through *decision trees*, which allow us to organise probabilities into a very understandable system of outcomes. Finally, we will end today with some famous examples of counterintuitive results involving conditional probability: The Monty Hall problem, the Birthday paradox, and the validity of medical testing.

1.1 A brief review

Recall that in classical probability, all individual outcomes are equally likely. As we discussed in Lecture 6, this is a special assumption that offers us a unique way to determine probabilities:

$$P(A) = \frac{|A|}{|S|}.$$

The formula above says that since every outcome is assumed to be equally likely, we can determine $P(A)$ by simply counting up the size of any event A and calculating the proportion of the sample space S that A occupies.

In contrast, when working with general probability, we relax this assumption that every outcome has an equal chance of happening. Writing ω for an individual outcome, in general probability we do not assume that all probabilities $p(\omega)$ are the same. As a consequence, we cannot simply count up the size of an event A in order to determine its probability. Instead, we must add up the probabilities of the outcomes contained within A. The result of this sum will then be the probability $P(A)$:

$$P(A) = \sum_{\omega \text{ in } A} p(\omega).$$

As we mentioned in the previous lecture, the formulas for compound events are still valid in the general setting. Their justification is different, however, they can still be used as expected.

1. To determine the probability of an event A *not* happening, we may use the formula:

$$P(\text{not } A) = 1 - P(A).$$

2. To determine the probability of two events A and B happening at the same time, we may use either of the formulas:

$$P(A \text{ and } B) = \begin{cases} P(A) \cdot P(B), & \text{if } A \text{ and } B \text{ are independent,} \\ P(A \mid B) \cdot P(B), & \text{are dependent, provided } P(B) \neq 0. \end{cases}$$

3. To determine the probability of either of two events A or B happening (or both), we may use either of the formulas:

$$P(A \text{ or } B) = \begin{cases} P(A) + P(B), & \text{if } A \text{ and } B \text{ are mutually exclusive,} \\ P(A) + P(B) - P(A \text{ and } B), & \text{if } A \text{ and } B \text{ are not mutually exclusive.} \end{cases}$$

We mentioned in Lecture 5 that the symbol $P(A \mid B)$ represents the probability of A occurring given that we *know* B will definitely occur. In that discussion, we mentioned that for classical probabilities, we can use the formula for $P(A \text{ and } B)$ to discover the formula for $P(A \mid B)$. Recall that if we intentionally insert a copy of $1 = \frac{|B|}{|B|}$ into the classical version of our Golden Formula, we get:

$$P(A \text{ and } B) = \frac{|A \cap B|}{|S|} = \frac{|A \cap B|}{|S|} \cdot \frac{|B|}{|B|} = \frac{|A \cap B|}{|B|} \cdot \frac{|B|}{|S|}.$$

The second factor is just $P(B)$, and the first factor gives us a new formula:

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

In probabilistic terms, this formula determines the probability of A given that we start with B as our sample space instead of S. Put differently, it is the probability of A given that B is guaranteed.

In the case of general probability theory, there is a similar formula:

$$P(A \mid B) = \frac{\sum_{\omega \in A \cap B} p(\omega)}{\sum_{\omega \in B} p(\omega)} = \sum_{\omega \in A \cap B} \frac{p(\omega)}{P(B)}, \qquad \text{provided } P(B) \neq 0.$$

This formula appears to be quite unattractive. For now, it is good enough for you to recognise that it exists , and in a moment we will demystify this with a clear example.

In either case (classical or otherwise), we may always rearrange the formula for $P(A \text{ and } B)$ to obtain another expression of conditional probability:

$$P(A \text{ and } B) = P(A \mid B) \cdot P(B) \quad \longrightarrow \quad P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}, \qquad P(B) \neq 0.$$

The same process applies if we decide to switch the roles of A and B:

$$P(B \text{ and } A) = P(B \mid A) \cdot P(A) \quad \longrightarrow \quad P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}, \qquad P(A) \neq 0.$$

These formulas give us a useful method for finding conditional probabilities: if we happened to know both $P(A \text{ and } B)$ and the probability of the event that we are assuming to know, then we can use either of the two formulas above to calculate conditional probabilities. We will now demonstrate this via two examples.

1.2 Example 1: Conditional Probability in a classical setting

In Lecture 5 we saw conditional probability briefly. At that time, we touched on the following example. Suppose that you roll a fair dice, and consider the two events:

$$A = \{\text{the dice lands on an even number}\},$$

and

$$B = \{\text{the dice lands on a number less than 4}\}.$$

Since we are working with an ordinary 6-sided dice, the sample space and events A and B can be written in set notation as:

$$S = \{1, 2, 3, 4, 5, 6\}, \qquad A = \{2, 4, 6\}, \qquad B = \{1, 2, 3\}.$$

Now, imagine that as you throw the dice and it is spinning up in the air, time slows down for a moment and God descends from the heavens. He whispers in your ear "the dice will land on a number less than 4". Now, in that moment, you would absolutely know that the event B will happen. In turn, this might adjust your uncertainty for event A happening.

To figure out the new probability of A happening given that B is guaranteed, we can effectively reduce the entire sample space to the smaller subset of numbers less than 4:

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

In this restricted system, we notice that there are now three possible outcomes, and only one of them happens to be an even number. Therefore, the probability of landing an even number will now *change* from the initial $\frac{1}{2}$ to $\frac{1}{3}$. What we have just done is calculate the conditional probability of event A *given* B. In symbols, the conditional probability $P(A \mid B)$ is calculated by rearranging the formula for the probability of A and B:

$$P(A \text{ and } B) = P(A \mid B) \cdot P(B) \quad \longrightarrow \quad P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}.$$

By inspection, we can see that the intersection of the sets A and B is the set $A \cap B = \{2\}$. Thus, in symbols, the argument we made above can be written:

$$P(A \mid B) = \frac{|A \cap B|}{|B|} = \frac{|\{2\}|}{|\{1,2,3\}|} = \frac{1}{3}.$$

Perhaps it is simpler to directly calculate the probabilities $P(A \text{ and } B)$ and $P(B)$ first, and then just substitute them into the formula:

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{1}{6} \div \frac{3}{6} = \frac{1}{3}.$$

1.3 Example 2: Conditional Probability in a general case

Now suppose that we replace our dice with an unfair one, with new probabilities:

$$p(1) = 0.1, \quad p(2) = 0.2, \quad p(3) = 0.1, \quad p(4) = 0.1, \quad p(5) = 0.1, \quad p(6) = 0.4.$$

Using the same events and sample space as before:

$$S = \{1,2,3,4,5,6\}, \qquad A = \{2,4,6\}, \qquad B = \{1,2,3\}.$$

Observe now that we are *not* working with classical probability, meaning that we cannot simply count up the sizes of subsets in order to determine probabilities. Instead, we have to add up the probabilities of the individual outcomes contained within each event. So, the probabilities of A and B are now:

$$\begin{aligned} B = \{1,2,3\} &\Longrightarrow P(B) = p(1) + p(2) + p(3) = 0.1 + 0.2 + 0.1 = 0.4 \\ A = \{2,4,6\} &\Longrightarrow P(A) = p(2) + p(4) + p(6) = 0.2 + 0.1 + 0.4 = 0.7 \end{aligned}$$

So, according to this unfair system, there is a 70% chance of landing on an even number whenever we throw the dice. Suppose now that, again, when we throw the dice up into the air, God comes down and tells us that the outcome will definitely be a number less than 4. In this situation, our uncertainty regarding event A will again change, because we have removed the possibilities of landing a 4 , a 5, or a 6. Again, based on our given information, the new system has a reduced sample space:

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

Roughly speaking, the probabilities of the outcomes $1, 2$ and 3 now need to be adjusted so that they all add up to 1 in this new system. Mathematically, the way to do that is to divide each of them by the probability $P(B) = 0.4$. Observe, when we do that, the sum of all the possible outcomes indeed adds to 1:

$$\frac{p(1)}{P(B)} + \frac{p(2)}{P(B)} + \frac{p(3)}{P(B)} = \frac{0.1}{0.4} + \frac{0.2}{0.4} + \frac{0.1}{0.4} = \frac{0.4}{0.4} = 1.$$

So, the conditional probabilities of individual outcomes inside B can be found by dividing their original probabilities by the total probability of the reduced sample space. In particular, if we look at the new probability for landing on a 2, we have:

$$p(2 \mid B) = \frac{p(2)}{P(B)} = \frac{0.2}{0.4} = \frac{1}{2}.$$

Here, $p(2 \mid B)$ is shorthand for $P(\{2\} \mid B)$. On some level, this makes sense: notice that in the original system, the three outcomes $1, 2$ and 3 have probabilities $0.1, 0.2$ and 0.1, respectively. The probabilities of landing a 1 or a 3 are equal at 0.1, and the probability of landing a 2 is twice as much at 0.2. So, if we were to suddenly restrict our attention to only these three outcomes, we would expect that this ratio is still maintained: $P(2 \mid B)$ should be twice as big as $P(1 \mid B)$ and $P(3 \mid B)$, and these latter two should be equal. The only way to achieve this is with the probabilities

$$P(1 \mid B) = \frac{1}{4}, \quad P(2 \mid B) = \frac{1}{2}, \quad \text{and} \quad P(3 \mid B) = \frac{1}{4}.$$

In order to calculate the conditional probability $P(A \mid B)$ in this system, we can first calculate $P(A \text{ and } B)$ and then use the same formula as before. To emphasize, we are *not* working in classical probability, so to determine the probability of the compound event A and B, we need to look inside the intersection $A \cap B$ and add up each of the probabilities of the outcomes in that set. Since $A \cap B = \{2\}$, there is only one outcome (namely, landing on a 2), and therefore there is only one term in the sum. Simply, we conclude that $P(A \text{ and } B) = 0.2$.

Now, we can use the formula for conditional probability to confirm our previous argument:

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{0.2}{0.4} = 0.5,$$

which is the same answer as before.

Exercise

Calculate the conditional probability $P(B \mid A)$ for the unfair dice.

Solution

We want to calculate $P(B \mid A)$, meaning the probability that the dice lands on a number less than 4, given that we already know it landed on an even number. The intersection is

$$A \cap B = \{2\},$$

which gives us $P(A \text{ and } B) = 0.2$, as before. However, now we need to calculate $P(A)$ before applying the formula. We have:

$$P(A) = p(2) + p(4) + p(6) = 0.2 + 0.1 + 0.4 = 0.7,$$

and therefore

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)} = \frac{0.2}{0.7} = \frac{2}{7}.$$

1.4 Bayes' Theorem

As the previous exercise demonstrates, the two conditional probabilities

$$P(A \mid B) \qquad \text{and} \qquad P(B \mid A)$$

are not the same in general. However, they are still related to each other. According to our discussion in Section 1.1, we have two different ways to determine the probability $P(A \text{ and } B)$:

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

We may set these two equal to each other and rearrange to discover that:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}.$$

This relationship is known as *Bayes' Theorem*.

Bayes' Theorem

For a pair of events A and B, if $P(A) \neq 0$ and $P(B) \neq 0$ then

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}.$$

Bayes' Theorem is particularly useful when we want to calculate one conditional probability when we know the other. For example, based on our example of the unfair dice in the previous section, we have:

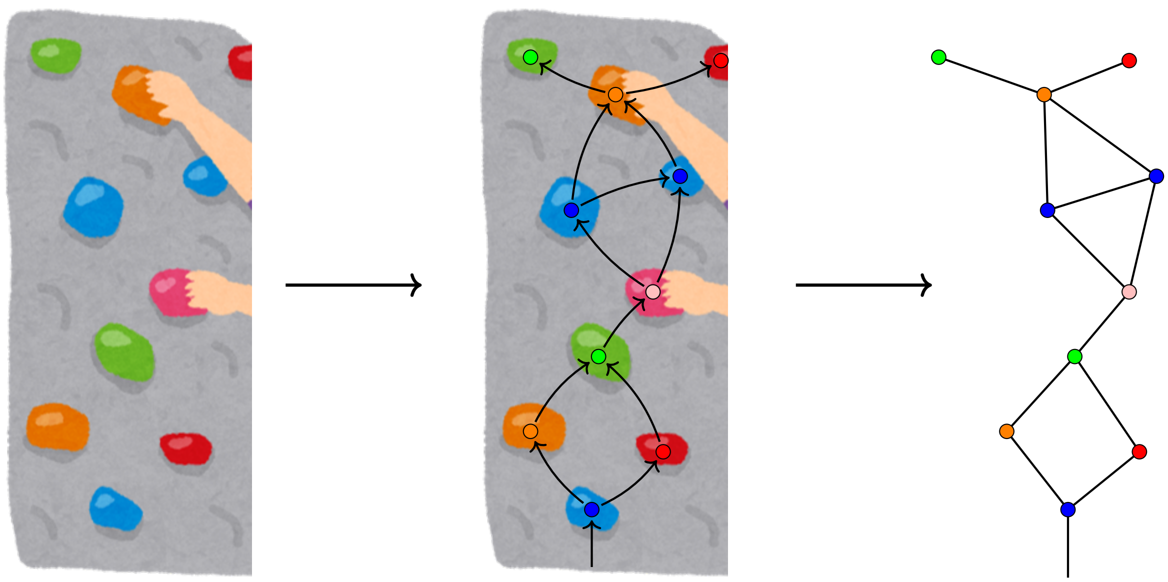
$$P(B \mid A) = \frac{P(A \mid B) \cdot P(B)}{P(A)} = \frac{0.5 \times 0.4}{0.7} = \frac{2}{7}.$$

2 Decision Trees

Conditional probabilities naturally arise when studying situations in which successive choices might affect each other. In small systems, it can be extremely helpful to simply try to map out the space of all possible things that can happen. One method for doing this is to use a *decision tree*, which is a diagram that branches out depending on each step of a multi-stage process. To motivate this, we will start with a rather complicated example.

2.1 A motivating example: the rock climber

Suppose that we follow the hand of somebody climbing up a rock face. We can idealise this situation by creating a sort of "graph" which displays how the network of coloured rocks is connected to each other:



In this picture, as the climber moves up the rock, their hand will trace a path from the bottom blue rock upwards to either the green or red rock at the top. Notice that at some stages of this process, the rock climber is presented with a choice of how to proceed. For example, at the very start of the rockface, the climber can either choose to go from the blue rock to the orange, or perhaps to the red. At other points, the climber cannot choose what to do, it is simply determined. For example, from the first orange rock they must follow the path to the next green rock.

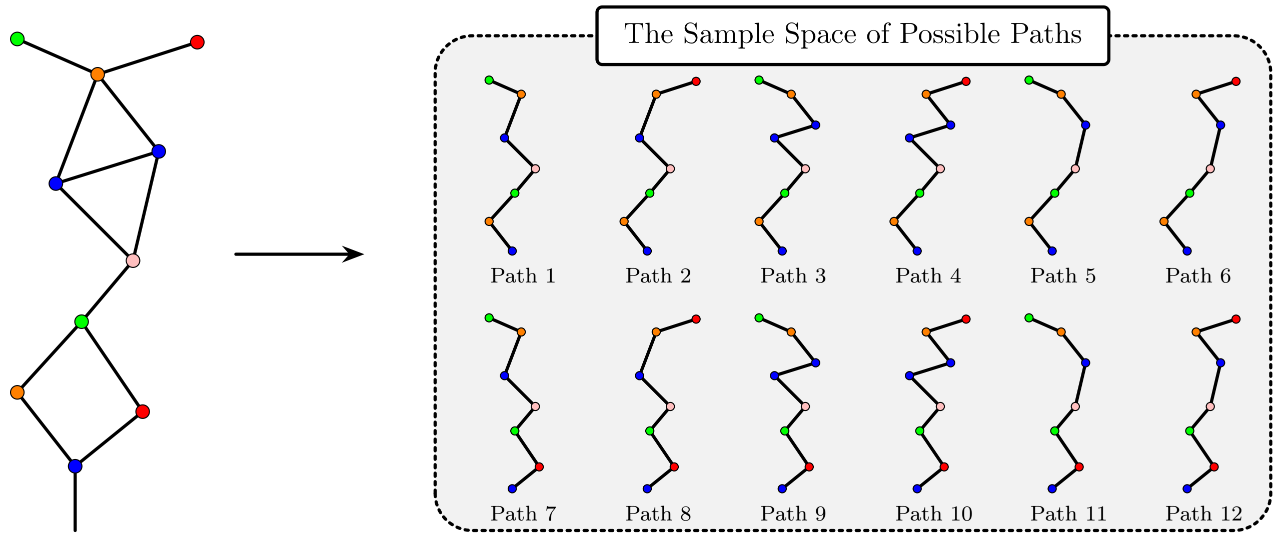
Suppose that every time the climber makes a choice, there is a 50% chance that they pick the left rock, and a 50% chance that they pick the right rock. Our central question is:

> What is the probability that the climber takes any given path?

Assuming that there are 12 paths, a first natural guess might be to apply the Golden Formula of classical probability:

$$P(\text{a given path}) = \frac{1}{\text{total number of paths}}.$$

Firstly, we should describe the sample space, which in this case is the space of all paths from the bottom to the top. With a bit of time, you can convince yourself that there are actually 12 unique paths through the rock face. They are pictured below:



Thus the sample space in this setting is the set

$$S = \{\text{Path 1}, \text{Path 2}, \text{Path 3}, \ldots, \text{Path 12}\}.$$

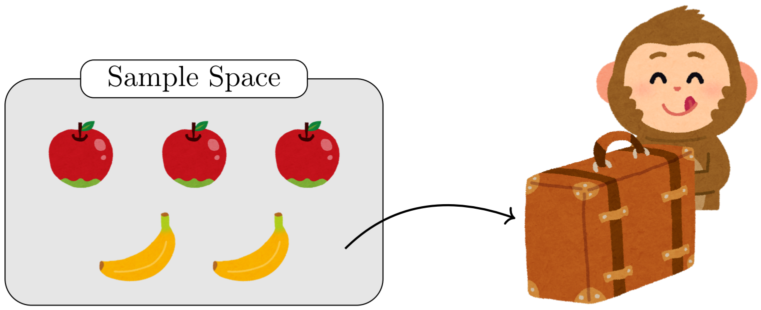
It seems reasonable, therefore, to try and use the above formula to calculate the probability of a given path. For example, for Path 1 we can guess:

$$P(\text{Path 1}) \stackrel{?}{=} \frac{1}{\text{total number of paths}} = \frac{1}{12}.$$

As a matter of fact, this is not correct. Later on we will calculate this properly, but for now we have to simply observe: in the space of paths, notice that some paths are more complicated than others. These different paths **will not** be equally likely outcomes, so the above formula is not the correct approach. We will return to this in a moment. But first, we will illustrate the concept again in a simpler setting.

2.2 Example: the jungle and the thief

Suppose that you are in the jungle, and you leave your bag unattended. A local monkey climbs down from his tree and rummages through your bag randomly. As it so happens, there are 5 pieces of fruit in your bag that you packed for lunch: 3 apples and 2 bananas. The monkey starts stealing fruit randomly and eats whatever he finds before looking in the bag again. In probability theory, this process is called *drawing without replacement*.



Exercise 1

(a) What is the probability that the monkey steals an apple first?
(b) What is the probability that the monkey steals a banana first?

Solution

There are 5 fruits total, so we may perform a simple count:

(a) $P(\text{apple first}) = \frac{3}{5}$.
(b) $P(\text{banana first}) = \frac{2}{5}$.

Now, we will focus on a successive question:

> What is the probability he steals a banana *and then* an apple?

It is clear that we should be using a conditional probability here, since the second outcome is clearly affected by the first.

To be precise about the ordering, let B_1 mean "banana on the first draw" and let A_2 mean "apple on the second draw". Then "banana then apple" is the event "B_1 and A_2".

We can calculate this by using the conditional probability formula. We may calculate the probability using the formula:

$$P(B_1 \text{ and } A_2) = P(A_2 \mid B_1) \cdot P(B_1).$$

Let's evaluate this formula one step at a time:

Step 1: We saw previously that $P(B_1) = \frac{2}{5}$.

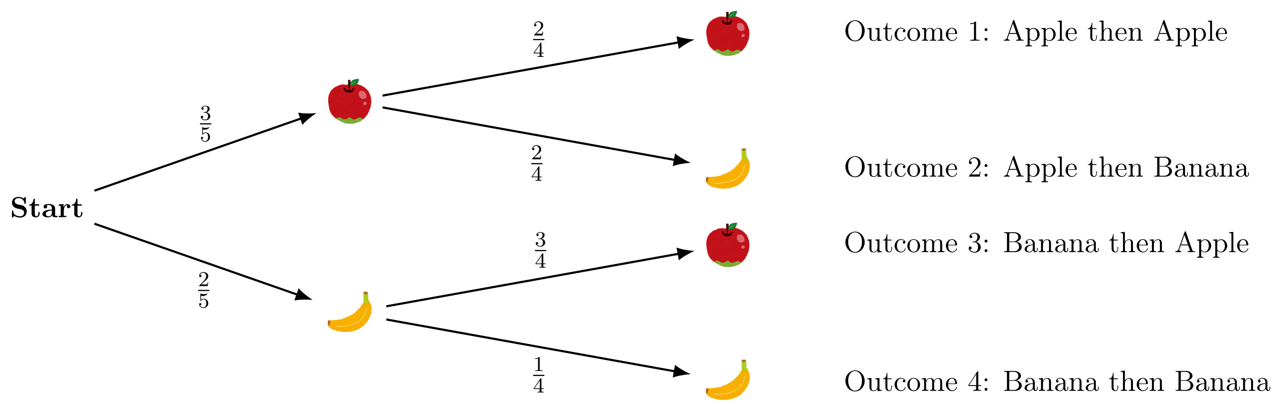
Step 2: We now compute $P(A_2 \mid B_1)$. Observe that if a banana is removed (and eaten by the monkey), then on the second attempt at taking fruit from the bag, the monkey now has a *different* sample space. In the first attempt at taking fruit, there were 5 pieces of fruit. However, since a banana has already been removed, now there are only 4 pieces of fruit left: 3 apples and 1 banana. This new sample space has size 4, and there are 3 apples that the monkey could select. So, the conditional probability will be:

$$P(A_2 \mid B_1) = \frac{3}{4}.$$

Step 3: We now take these two numbers and multiply them to get the probability of taking a banana and then an apple:

$$P(B_1 \text{ and } A_2) = P(B_1)\,P(A_2 \mid B_1) = \frac{2}{5} \cdot \frac{3}{4} = \frac{3}{10}.$$

Similar reasoning can be applied to any of the other possibilities for our cheeky monkey. We can arrange this entire process into a tree, drawn below:



As you can see, the calculation we did earlier, namely:

$$P(\text{Banana then Apple}) = P(B_1) \cdot P(A_2 \mid B_1) = \frac{2}{5} \cdot \frac{3}{4} = \frac{3}{10},$$

can also be visually confirmed by checking the branch that corresponds to Outcome 3 in the tree above , because here we are simply multiplying the numbers written along the branch corresponding to Outcome 3. In fact, we have just seen our first instance of a general trick: if we can arrange the outcomes of a multi-step experiment as a tree, then the probability of any outcome can be obtained by multiplying every probability that occurs along that branch.

We can use this observation to quickly calculate the other three outcomes in the tree above:

- Outcome 1: $P(A \text{ then } A) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10}$,
- Outcome 2: $P(A \text{ then } B) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10}$,
- Outcome 4: $P(B \text{ then } B) = \frac{2}{5} \cdot \frac{1}{4} = \frac{1}{10}$.

Observe that the sum of these four probabilities is

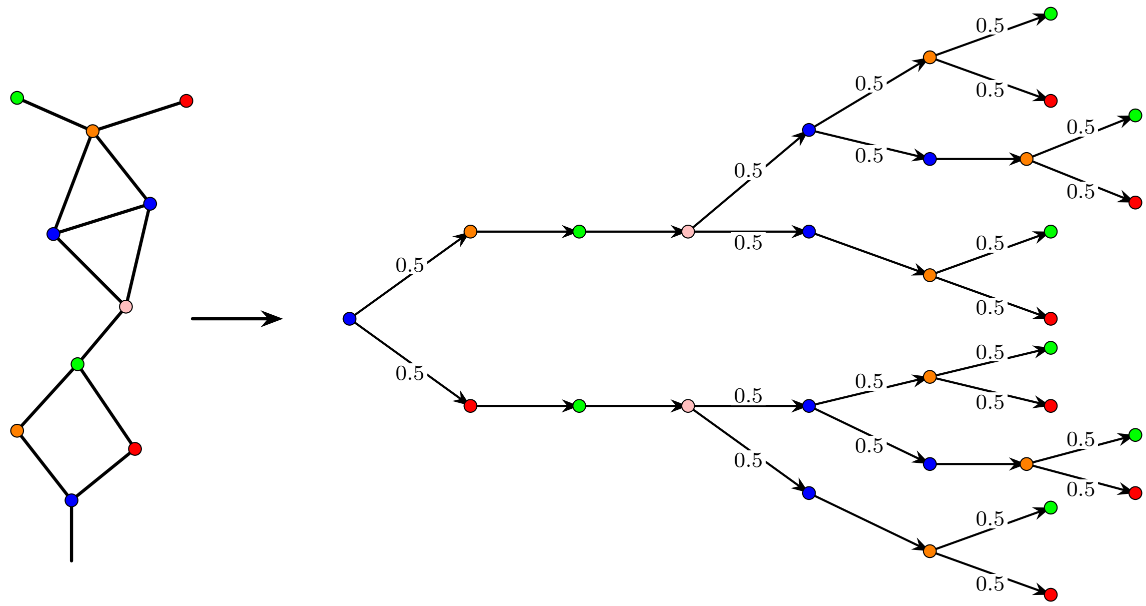
$$\frac{3}{10} + \frac{3}{10} + \frac{3}{10} + \frac{1}{10} = 1.$$

This is the same as saying that it is guaranteed that one of the outcomes must occur.

2.3 Returning to the Rock Climber

We've just seen that a decision tree can be used to organize all possible outcomes of a multi-step experiment. To find the probability of a particular path, we simply need to multiply the probabilities that are written along that particular path. We can use this technique to figure out the probabilities associated with each of our rock climber's routes through the wall.

To begin with, we write the tree for this system and then label each "choice point" with the assumed 50-50 probabilities. The full tree is drawn below.



The above tree is organised so that each path is in order: the first path corresponds to Path 1, the second path corresponds to Path 2, and so on. You should observe here that some paths are more complicated than others , for instance, Path 12 looks simpler than Path 9 or 10. This is because there are some routes up the wall that only require making three choices, and others require making four choices. As a matter of fact, any path that involves making the choice to move to the left from the pink rock will be more complicated than those paths that involve the choice to move to the right instead. In any case, we can use our theory of decision trees to calculate the probability of any given path up the rock face. As with the example of the klepto-monkey, in this case we simply multiply the values written along the branch to obtain the overall probability. These are summarised below.

Probabilities of Each Path

We make a distinction between the number of choices that each path requires:

- Paths $1, 2, 3, 4, 7, 8, 9$ and 10 all require making four choices. On the decision tree, there are thus 4 copies of 0.5 written on the branches corresponding to these paths, so their overall probability will be:
$$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{16}.$$
- Paths $5, 6, 11$ and 12 require making only *three* total choices. Therefore, the probability of taking each one of these paths is greater than the others:
$$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}.$$

We may always double-check that our calculations are correct by confirming that all of our answers add up to 1. We have:

$$8\left(\frac{1}{16}\right) + 4\left(\frac{1}{8}\right) = \frac{1}{2} + \frac{1}{2} = 1,$$

which is correct. Notice here that we have also discovered something else: there is a 50% probability that the climber will take one of the simpler-looking paths, and there is a 50% probability that they will not.

2.4 Exercise: Drawing Without Replacement

A box has 6 balls inside: 3 red, 2 blue and 1 green. You take 2 balls at random, without replacing the first ball back into the box.

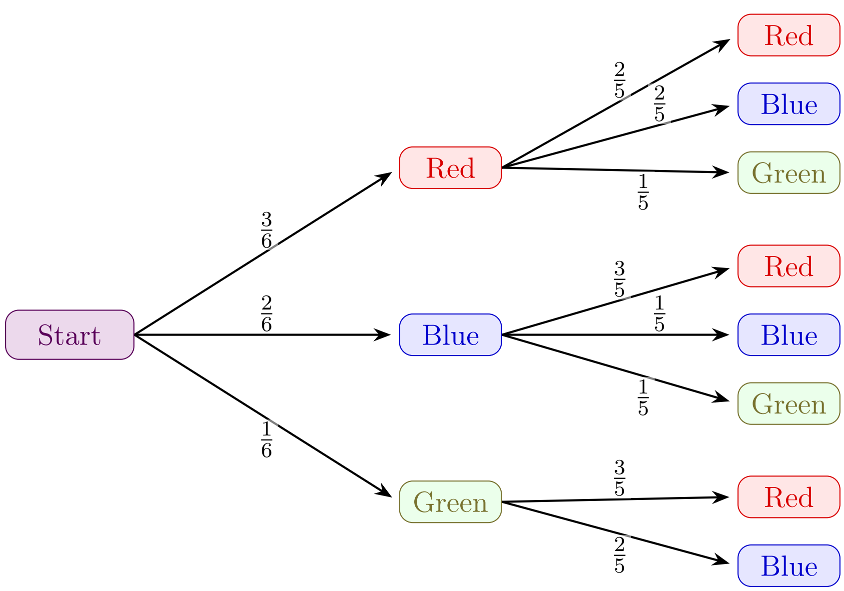
Exercise 2

(a) What is the probability of getting 2 reds in a row?

(b) What is the probability of getting a green then a blue?

(c) What is the probability of getting a blue and a red ball, in either order?

Before getting to the full solution, let's observe what the tree looks like. For completeness, we have labelled all of the probabilities along each arrow:



As you can see, on the bottom branches there are no longer three possible outcomes . Why? Because there is only 1 green ball in the box. If we select it first, then the box would only have red and blue balls left over. This tree is fully populated with all of the probabilities and conditional probabilities. Instead of calculating these directly using formulas, we can simply argue heuristically, in a similar manner to that of our monkey thief of Section 2.2. For example: to calculate the probability of picking another red ball after just having drawn a red, we are being asked to look at the situation in which one red has already been removed, meaning that the sample space has now been reduced to only 2 reds, plus the 2 blues and 1 green ball that remain in the box. Since there are 2 possible red balls to select out of 5, the conditional probability $P(\text{red} \mid \text{red})$ will be $\frac{2}{5}$. With this in mind, we will now answer the questions above.

Solution

(a) Red then red:

$$P(\text{red then red}) = \frac{3}{6} \cdot \frac{2}{5} = \frac{6}{30} = \frac{1}{5}.$$

(b) Green then blue:

$$P(\text{green then blue}) = \frac{1}{6} \cdot \frac{2}{5} = \frac{2}{30} = \frac{1}{15}.$$

(c) Blue and red means either red then blue or blue then red:

$$P(\text{red then blue}) = \frac{3}{6} \cdot \frac{2}{5} = \frac{1}{5},$$

and

$$P(\text{blue then red}) = \frac{2}{6} \cdot \frac{3}{5} = \frac{1}{5}.$$

These are mutually exclusive outcomes, so:

$$P(\text{blue and red}) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}.$$

3 Famous Examples

3.1 The Monty Hall Problem

Suppose that you are on a game show. There are three doors. Behind one door is a car, and behind the other two doors are goats. The doors are randomized, so you do not know what is behind which door. So, you pick door 1, hoping that there is a car behind. The host, who knows what is behind the doors, then opens a different door and reveals a goat. Assume that the host always opens a different door containing a goat and always offers you the chance to switch. You now have a choice:

stay with your original door   or   switch to the remaining unopened door.

Should you switch?

Amazingly, under these standard assumptions, the answer is yes: it makes more sense to switch. The probability of winning the car by staying is $\frac{1}{3}$, while the probability of winning the car by switching doors is $\frac{2}{3}$. This is rather confusing, so we will argue it in three different ways.

3.1.1 Argument 1: transfer the probability

Let

$$A = \{\text{your first choice is the door with the car behind it}\}.$$

Then:

$$P(A) = \frac{1}{3},$$

and therefore:

$$P(\text{not } A) = \frac{2}{3}.$$

If your first choice was correct, then switching loses. If your first choice was wrong, then switching wins, because the host removes the other goat. Therefore:

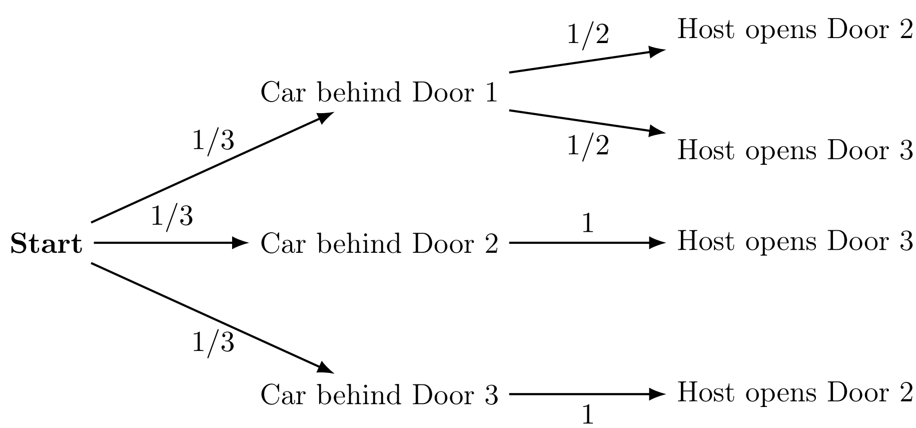
$$P(\text{switching wins}) = P(\text{your first choice was wrong}) = \frac{2}{3}.$$

3.1.2 Argument 2: make the game bigger

Suppose instead that there were 1000 doors, where one door has a car behind it and the other 999 doors have goats. Suppose that you picked door number 1 as before. The host, knowing where the car actually is, then opens 998 of the doors and reveals goats behind them. When you are left with only two doors left: your door and the other, you may start to doubt your original choice. If there were 1000 doors, your original choice would have a probability of $\frac{1}{1000}$ of being correct. That probability does not suddenly become larger just because the host opened doors that he already knew were goats. Instead, your door carries the same $\frac{1}{1000}$ probability of containing the car, and the other remaining door is overwhelmingly more likely to have the car. Clearly, switching doors would be a better idea.

3.1.3 Argument 3: a decision tree

Assuming that you initially picked Door 1, the space of possible outcomes is below. It should be emphasized here that the host *knows* which door the car is actually behind. If he wants to intentionally pick a door that reveals a goat, then sometimes he has a choice and sometimes he does not. When the host has two goat doors available, assume he chooses between them with equal probability.



If the car happened to be behind Door 1, then switching loses. According to the decision tree above, the probability of this happening will be the sum of the probabilities for the top two branches, giving a result of $\frac{1}{3}$. If the car is *not* behind Door 1, then it is either behind Door 2 or Door 3. The probability of this happening is $\frac{2}{3}$. So, we see that switching will cause you to win the car with a probability of $\frac{2}{3}$, and switching will cause you to *lose* the car with a probability of $\frac{1}{3}$.

3.2 The Birthday Paradox

Let A be the event "at least two people in the room share a birthday". In this case, the probability $P(A)$ will change radically depending on n, the number of people in the room. For example, if $n = 2$, then we would expect that the chance that these two people share a birthday is extremely low. However, for $n > 365$, it is guaranteed that two people in the room must share a birthday, since there are only 365 possible birthdays if we ignore leap years. As we can see, when n is very low, so is $P(A)$, and when n is high, $P(A)$ eventually becomes equal to 1. Here is a question:

> At what value of n does $P(A)$ become more than 50%?

Put differently: how many people do there need to be in the room so that it is *more likely than not* that two people share a birthday? In fact, the answer is quite surprising: we only require 23 people!

This surprising fact is known as the "Birthday paradox". It is an example of a "veridical paradox", which is simply a mathematical result that is so surprising that it *feels wrong*. However, as we will now demonstrate, the Birthday Paradox follows from some basic applications of probability.

3.2.1 Resolving the Birthday Paradox

In order to resolve the birthday paradox, we are going to flip it on its head and rephrase the problem. Instead of computing $P(A)$, we are going to compute $P(\text{not } A)$, the probability of the complement of A. Recall that the complement event is the probability-theoretic analogue of the word "not". So here, the complement event A^c means "everybody in the room has a distinct birthday".

Recall that the probability of the complement is simply $1 - P(A)$. So, if $P(\text{not } A) < 50\%$, then $P(A) > 50\%$, because

$$P(A) + P(\text{not } A) = 1.$$

Let's assume that the choice of each birthday is independent and equally likely, so that any birthday has probability $1/365$. We will now calculate $P(\text{not } A)$ for low n to see how the process works.

For $n = 2$, we are trying to calculate the probability that two people have distinct birthdays. Clearly we are allowed to pick Person 1's birthday freely. But, in order for the two birthdays to be distinct, we must pick a birthday for Person 2 that avoids our first choice. There are 364 days that are distinct from our first choice, so (if treated as a random event), the probability that Person 2 has a distinct birthday will be

$$P(\text{not } A) = \frac{\text{number of days distinct from Person 1's birthday}}{\text{total number of days in a year}} = \frac{364}{365} \approx 0.997 = 99.7\%.$$

Suppose now that we add a third person to the room, i.e., $n = 3$. We have just established that the probability of two distinct birthdays is $\frac{364}{365}$. But, we also need the third person's choice of birthday to be distinct from Person 1's *and* Person 2's. The probability for the next person's birthday is conditional on the previous birthdays. To calculate the probability of an "and" statement like this, we need to multiply the associated probabilities:

$$
\begin{aligned}
P(\text{not } A) &= \frac{364}{365} \times \frac{\text{number of days distinct from Person 1 and Person 2's birthday}}{\text{total number of days in a year}} \\
&= \frac{364}{365} \cdot \frac{363}{365} \approx 0.992 = 99.2\%.
\end{aligned}
$$

We can keep repeating this pattern for larger and larger n. Every time that we increase the value of n by 1, we see that we keep "using up" available birthdays for the next person to be able to pick from. This means that we keep needing to multiply by smaller and smaller numbers after every step. For example:

$$
\begin{aligned}
\text{For } n = 4 : P(\text{not } A) &= \left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\left(\frac{362}{365}\right) \approx 0.983 = 98.3\% \\
\text{For } n = 5 : P(\text{not } A) &= \left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\left(\frac{362}{365}\right)\left(\frac{361}{365}\right) \approx 0.973 = 97.3\% \\
\text{For } n = 6 : P(\text{not } A) &= \left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\left(\frac{362}{365}\right)\left(\frac{361}{365}\right)\left(\frac{360}{365}\right) \approx 0.960 = 96.0\%
\end{aligned}
$$

For general n, we obtain the formula:

$$
P(\text{not } A) = \left(\frac{365}{365}\right)\left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\cdots\left(\frac{365-(n-1)}{365}\right).
$$

The paradox will be resolved by finding the smallest value of n so that the probability $P(\text{not } A)$ drops below 50%, because that would mean that $P(A)$ goes *above* 50%. If we simply use the formula for bigger and bigger values of n, we will see that for $n = 23$ we have:

$$
P(\text{not } A) = \left(\frac{365}{365}\right)\left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\cdots\left(\frac{343}{365}\right) \approx 0.493 < 0.5.
$$

For comparison, when $n = 22$, we get $P(\text{not } A) \approx 0.524 > 0.5$. So $P(A) > 50\%$ when $n = 23$.

3.3 Medical Testing

Suppose that 1% of a population has a disease. There happens to be a medical test for this particular disease, in which:

- if a person has the disease, then the test detects this 99% of the time, but
- if a person does *not* have the disease, then the test will give a false positive 5% of the time.

Suppose that you take the test and you test positive. What is the probability that you have the disease? Amazingly, the answer is about 16.7%. We will now demonstrate how.

To begin with, let's rewrite the above conditions using conditional probabilities.

Let:

$$
D = \{\text{the person has the disease}\},
$$

and

$$
T = \{\text{the person tests positive}\}.
$$

The question is asking for the conditional probability:

$$P(D \mid T),$$

that is, the probability of actually having the disease after testing positive. Note that this is not the same as the other way around: $P(T \mid D)$ calculates the probability of a person testing positive, given that they actually have the disease. The setup of the question tells us that

$$P(T \mid D) = 0.99, \qquad P(D) = 0.01, \qquad P(D^c) = 0.99, \qquad \text{and} \qquad P(T \mid D^c) = 0.05.$$

We can use this information in combination with Bayes' Theorem to calculate the missing conditional probability:

$$P(D \mid T) = \frac{P(T \mid D)P(D)}{P(T)}.$$

In order to do this, we need to first calculate $P(T)$, the overall probability of testing positive.

Observe that there are two possible ways to test positive:

- either the person has the disease and correctly tests positive, or
- the person does not have the disease but falsely tests positive.

These are mutually exclusive, so we can reinterpret the event T as an "or" statement and calculate its probability as:

$$P(T) = P(T \mid D)P(D) + P(T \mid D^c)P(D^c) = (0.99)(0.01) + (0.05)(0.99) = 0.0594.$$

Using this in Bayes' theorem, we recover the infamous 16.7%:

$$P(D \mid T) = \frac{0.0099}{0.0594} = \frac{1}{6} \approx 0.167.$$

Why this feels surprising

The test is very good at detecting the disease when somebody actually has it. However, the disease is rare and there are a relatively large number of false positives. These false positives drag down the reliability of a single positive result, causing such a surprising result.