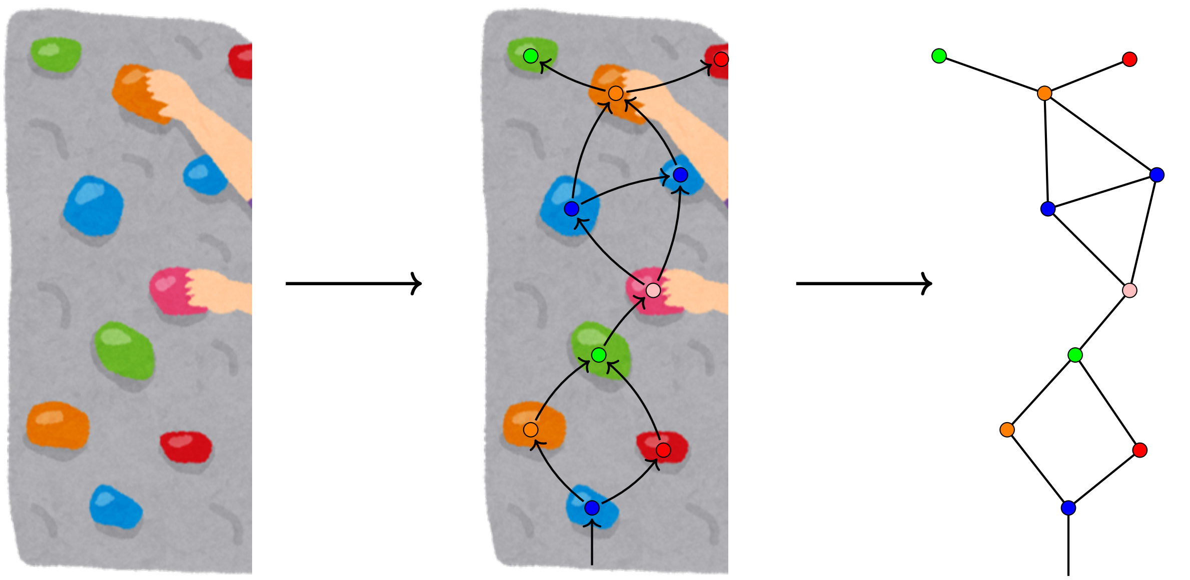
MAT120: Lecture 11 Handout

Successive Probability

Last lecture we introduced probability as a way to quantify uncertainty, and we learned how to compute classical probabilities by counting outcomes. In this lecture we develop *conditional probability*, meaning probability after we learn extra information. We also learn a practical tool, *decision trees*, for organizing multi-step experiments. Finally, we start counting outcomes for coin flips, which leads into binomial experiments next lecture.

1 Introduction: 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. As a matter of fact, there are 12 unique paths from the bottom to the top. What is the probability that the climber takes any given path?

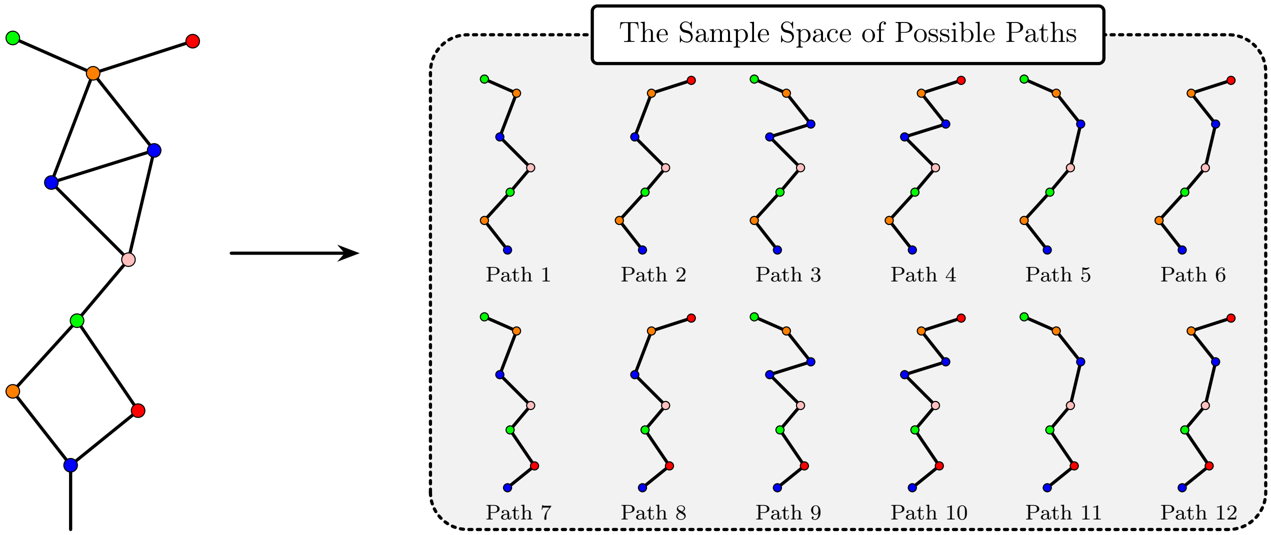
1.1 Trying to solve with Naive Counting

Last lecture, we saw a fundamental formula that works in our setting of classical probability:

$$P(\text{event}) = \frac{\text{number of outcomes favourable to event}}{\text{total number of outcomes}}.$$

In short, we can often calculate the probability of a general event by counting up the number of outcomes that correspond to that event, and then dividing by the total number of possibilities.

In the case of our rock climber, we can try to do the same thing. 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 12 unique paths through the rock face. They are pictured below:



Thus the sample space in this setting is the set

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

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:

$$P(\text{Path } 1) \stackrel{?}{=} \frac{\text{number of outcomes favourable to event}}{\text{total number of outcomes}} = \frac{1}{12}.$$

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.

2 Decision trees

2.1 What is a Decision Tree?

Last lecture we briefly introduced the notion of conditional probability, which made its way into the formula we used to calculate the probability of the compound event "A and B". Recall that for dependent events A and B, the probability of the event "A and B" is given by the formula:

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

or equivalently,

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

Here the term $P(B \mid A)$ is the *conditional probability*, which roughly means "the probability of B given that we assume A". Generally speaking, the conditional probability is about the change in probability that comes about by updating information based on assumptions.

Last lecture, we saw that the conditional probability can be defined by the formula

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)} \qquad \text{(provided } P(A) > 0\text{)}.$$

When working with classical probability, each outcome is equally-likely, which means that the above formula can be calculated by counting:

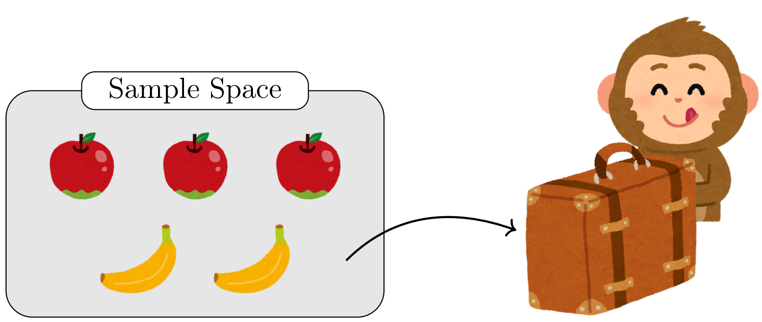
$$P(B \mid A) = \frac{\text{size of } A \cap B}{\text{size of } A}.$$

This is a way to count the size of the set-theoretic *intersection* $A \cap B$: we restrict attention to the outcomes in the event A, and then count how many of those outcomes also lie in B.

Conditional probabilities can become particularly useful when we consider *successive probabilities* – the act of taking a multi-step experiment with different potential outcomes at different steps. A decision tree is a diagram that exhaustively lists all the possible outcomes of one of these multi-step experiments. The collection of outcomes "split off" from each other, and are represented as branches of the tree.[1] Simply, decision trees help us to organise the space of possible outcomes, so that the probabilities of each individual outcome are easy to calculate.

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: there are 3 apples and 2 bananas. The monkey starts stealing fruit randomly and eats whatever he finds before looking in the bag again. In probabilistic terminology, this 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?

[1] To be slightly more precise: a branch is a line from the start of the tree to the very end.

Solution

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

(a) $P(\text{apple first}) = \dfrac{3}{5}$.

(b) $P(\text{banana first}) = \dfrac{2}{5}$.

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. Here, we let event A correspond to "The Monkey takes an Apple" and event B correspond to "The Monkey takes a Banana". We may then interpret the compound event "B and A" as "The Monkey takes a Banana *and then* an Apple". We may therefore calculate the probability using the formula:

$$P(B_1 \cap A_2) = P(B_1) \cdot P(A_2 \mid B_1).$$

We take this formula one step at a time:

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

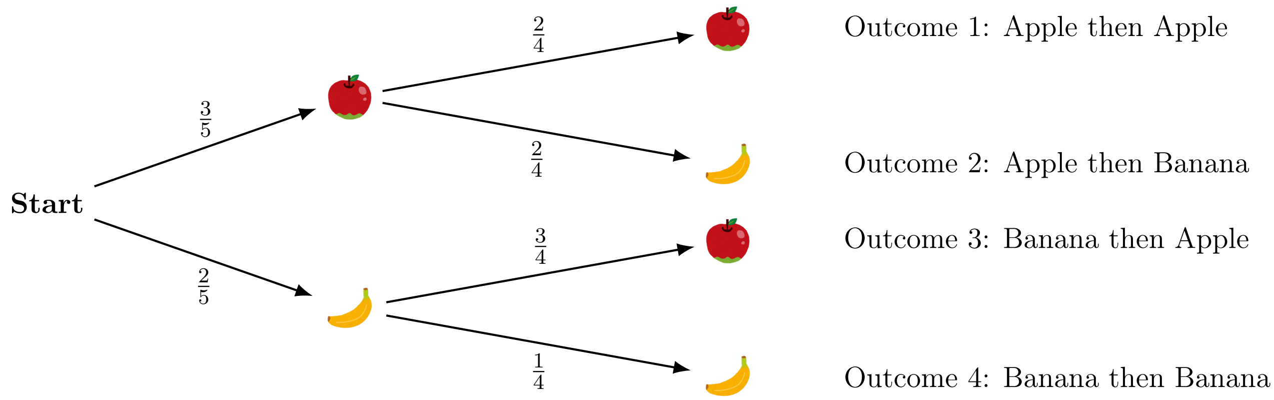
Step 2: We now compute $P(A_2 \mid B_1)$. Observe that if a banana is removed (and eaten by the monkey), then on the second attempt at taking fruit from the bag, the Monkey now has a *different* sample space. In the first attempt at taking food, there were 5 pieces of fruit. However, since a banana has 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:

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

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 – here we are simply multiplying the numbers written on the third branch. 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 on 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

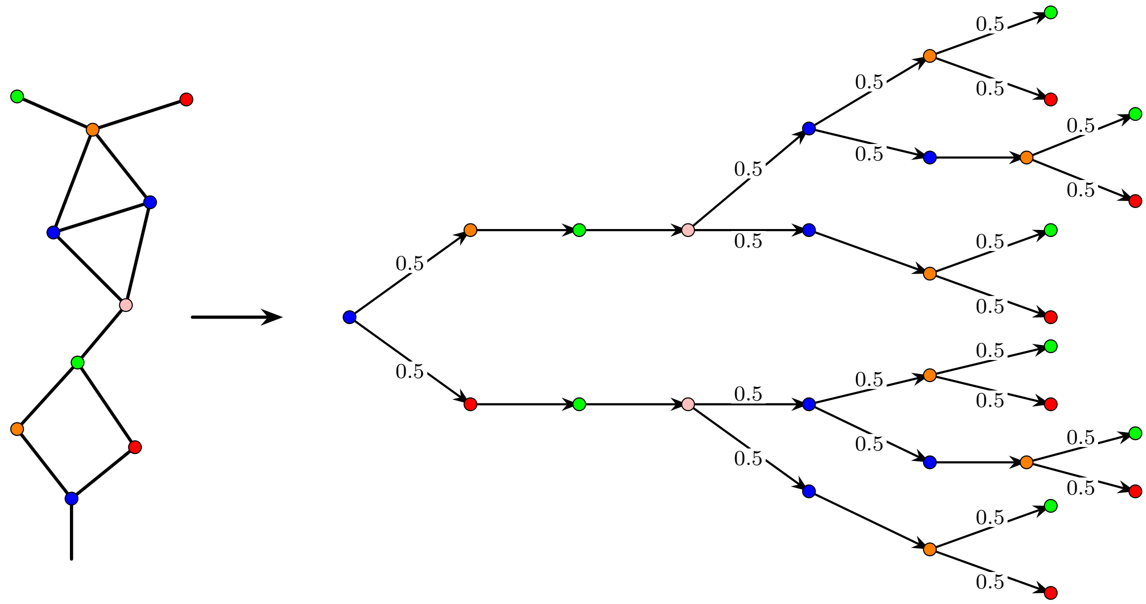
$$\frac{3}{10} + \frac{3}{10} + \frac{3}{10} + \frac{1}{10} = 1.$$

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 branch corresponds to Path 1, the second branch 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 turn left at the pink node will be more complicated than those paths that involve the choice to turn right.

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 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 accidentally discovered another finding: 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.

3 Counting outcomes

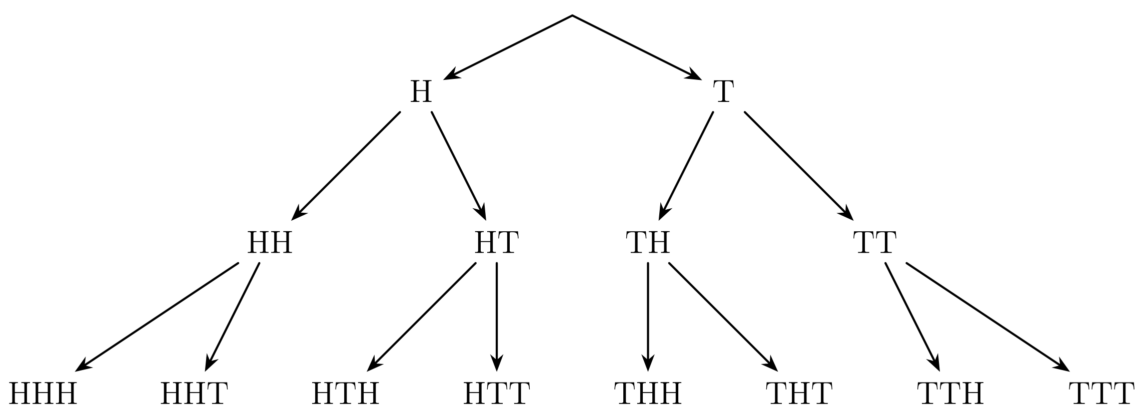
3.1 Flipping a Coin 3 times

Suppose we flip a coin 3 times.

Exercise 2

(a) What is the probability of exactly 2 tails?
(b) What is the probability of exactly 1 tail?

This sample space has equally-likely outcomes, so our best approach is simply to try and determine the probability by counting the number of favourable outcomes in each case. Again, if we wanted to, we could arrange these as a tree. For later convenience, we will draw this tree with the start of the experiment at the top.



Solution to Exercise 2

Based on the tree above, we see that the sample space for 3 coin flips is a set of size 8. We are working with equally-likely outcomes, so we can calculate the probabilities with a direct count.

(a) Exactly 2 tails: there are three outcomes favourable to this event, i.e. $\{HTT, THT, TTH\}$, so $P(\text{Exactly 2 tails}) = \frac{3}{8}$.
(b) Exactly 1 tail: there are three outcomes favourable to this event, i.e. $\{THH, HTH, HHT\}$, so $P(\text{Exactly 1 tail}) = \frac{3}{8}$.

3.2 Counting Tails for n-many Coin Flips

Based on the previous example, we may think that it's quite easy to calculate probabilities involving coin flips. However, the situation becomes quite complicated over time. Notice first how the size of each row in our probability tree grows with time: in the tree above, our first row has two outcomes, our second row has four outcomes, and our third row has eight outcomes. If we were to flip the coin again, then there are two more outcomes for every string of H's and T's that already exists. So, we will double the number of possible outcomes again, meaning that the fourth row will have 16 outcomes, that is, there are 16 possible ways that four coin flips can be. In general, when we flip a coin n times, there will be 2^n total outcomes.

Suppose that we wanted to count some specific number of Tails after n flips, let's say exactly r-many tails. In principle, we could use the same techniques from before by counting the number of outcomes that have precisely r tails, and then divide it by 2^n. However, this process quickly becomes difficult. For example: how many outcomes of 6 coin flips have exactly 2 tails? This would potentially involve sifting through the $2^6 = 64$ different outcomes of H's and T's to find the ones that have 2 T's and 4 H's.[2]

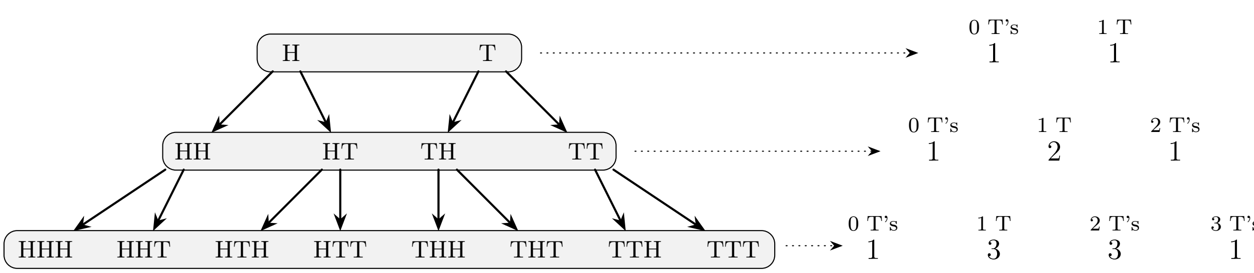
In practice, it would be nice if there were a clever way for us to be able to count up the number of outcomes with exactly r tails without needing to write out an overwhelming number of H's and T's and then read through them. As a matter of fact, there *is* a clever way to do this, and we have already seen how.

3.2.1 Counting exactly r-many tails in n flips

To illustrate our technique, let's again observe the tree above, and count up all of the possible outcomes that we find.

- 1 flip: if we flip a coin once, then the set of outcomes is $\{H, T\}$. There is one outcome with 0 T's present, and one outcome with 1 T present.
- 2 flips: flipping twice gives outcomes $\{HH, HT, TH, TT\}$. There is one outcome with 0 T's (namely the outcome HH), there are 2 outcomes that have 1 T (namely HT and TH) and there is one outcome with 2 T's (namely TT)
- 3 flips: here our set of outcomes is size 8, and is the final row of the tree above. We have already performed this count: we know that there is one outcome with 0 T's (namely HHH), there are 3 outcomes with 1 T (namely HHT, THH, HTH), there are 3 outcomes with 2 T's (namely HTT, THT, TTH) and there is one outcome with 3 T's (namely TTT).

To make this more suggestive, we can write our "T-counter" next to the tree:



As you can see, the triangle on the right-hand side of our diagram looks suspiciously like the first few rows of Pascal's triangle. As a matter of fact, this *is* Pascal's triangle. Why? Because counting up

[2] If this feels like a comfortable process for you, then consider $n = 10$ and $r = 4$ instead.

the possible outcomes of coin flips with exactly r tails is like saying "we have n positions to choose from, and I want to see all of the ways that I can select r of them to be tails".

If we redraw Pascal's triangle with annotations for the number of possible tails, we can read off the numbers without needing to make any complicated counts:

0 T's: 1							$2^0 = 1$ outcome
0 T's: 1	1 T: 1						$2^1 = 2$ outcomes
0 T's: 1	1 T: 2	2 T's: 1					$2^2 = 4$ outcomes
0 T's: 1	1 T: 3	2 T's: 3	3 T's: 1				$2^3 = 8$ outcomes
0 T's: 1	1 T: 4	2 T's: 6	3 T's: 4	4 T's: 1			$2^4 = 16$ outcomes
0 T's: 1	1 T: 5	2 T's: 10	3 T's: 10	4 T's: 5	5 T's: 1		$2^5 = 32$ outcomes
0 T's: 1	1 T: 6	2 T's: 15	3 T's: 20	4 T's: 15	5 T's: 6	6 T's: 1	$2^6 = 64$ outcomes

Using Pascal's triangle is significantly easier than writing out all possible combinations of H's and T's. For example, to determine the probability of getting exactly 2 tails in 6 total coin flips, we can inspect the triangle, go to the row $n = 6$ and the entry $r = 2$ to see that there are 15 outcomes that have precisely 2 tails, and there are $2^6 = 64$ total outcomes. We can immediately substitute these values into our counting formula to conclude that

$$P(2 \text{ Tails in } 6 \text{ flips}) = \frac{\text{number of ways we can get 2 tails in 6 flips}}{\text{total number of outcomes of 6 flips}} = \frac{15}{64} \approx 23.4\%.$$

Pascal's triangle has entries $C_{n,r}$. Generally speaking, we can use our "counting combinations" technique in order to create a general formula for the probability of finding exactly r tails amongst n coin flips:

$$P(r \text{ Tails in } n \text{ flips}) = \frac{\text{number of ways we can get } r \text{ tails in } n \text{ flips}}{\text{total number of outcomes of } n \text{ flips}} = \frac{C_{n,r}}{2^n}.$$

Of course, the probability of getting a head is the same as the probability of getting a tail, so the above formula is symmetric in that it can be used to calculate the probability of exactly r heads in n flips.

4 Bonus: The Birthday Paradox

Let A be the event "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. 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 not a paradox in the sense of the Barbershop Paradox that we saw earlier in the course – in that case, we ended up with infinite loops of true/false for which logic's function broke down. In contrast, the Birthday paradox 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.

4.1 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 $P(A^c)$, the probability of the complement of A. Recall that the complement event is the probability theory's 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 $P(A^c) < 50\%$, then $P(A) > 50\%$, because

$$P(A) + P(A^c) = 1.$$

Let's assume that the choice of each birthday is equally likely, so that any birthday has the probability 1/365. We will now calculate $P(A^c)$ 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

$$P(A^c) = \frac{\text{number of days distinct from Person 1's birthday}}{\text{total number of days in a year}} = \frac{364}{365} \approx 0.997 = 99.7\%.$$

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 third person's choice of birthday to be distinct from Person 1's *and* Person 2's. Randomly selecting the next person's birthday is a *conditional* event, because it depends on the outcome of the previous selection. To calculate the probability of an "and" statement like this, we need to multiply the associated probabilities:

$$\begin{aligned} P(A^c) &= \frac{364}{365} \times \frac{\text{number of days distinct from Person 1 and Person 2's birthday}}{\text{total number of days in a year}} \\ &= \frac{364}{365} \cdot \frac{363}{365} \approx 0.992 = 99.2\%. \end{aligned}$$

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 $(n+1)^{st}$ 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:

$$\text{For } n = 4 : P(A^c) = \left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\left(\frac{362}{365}\right) \approx 0.983 = 98.3\%$$
$$\text{For } n = 5 : P(A^c) = \left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\left(\frac{362}{365}\right)\left(\frac{361}{365}\right) \approx 0.973 = 97.3\%$$
$$\text{For } n = 6 : P(A^c) = \left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\left(\frac{362}{365}\right)\left(\frac{361}{365}\right)\left(\frac{360}{365}\right) \approx 0.960 = 96.0\%$$

For general n, we obtain the formula:

$$P(A^c) = \left(\frac{365}{365}\right)\left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\cdots\left(\frac{365-(n-1)}{365}\right).$$

The paradox will be resolved by finding the smallest value of n so that the probability $P(A^c)$ 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:

$$P(A^c) = \left(\frac{365}{365}\right)\left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\cdots\left(\frac{342}{365}\right) \approx 0.490 < 0.5.$$

So $P(A) > 50\%$ when $n = 23$.