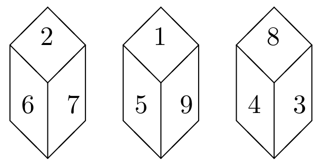
Probability

Introduction: We've already seen some basic probability in class and in the tutorials, but let's reviews the strange dice problem to remind ourselves.

We play a game called strange dice, which involves the three strange dice below (each hidden face has the same number as the face opposite it).



Die A Die B Die C

The rules of the game are simple: you choose a die, and then I choose a die. Then we both roll the dice, and whoever rolls a higher number wins. What's the winning strategey?

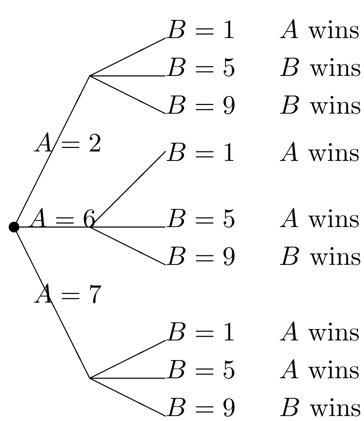
1. Show that Die A beats Die B with a probability of 5/9.

   We have to enumerate all the possible rolls die A and die B can have, and then figure out how many times A beats B. Die A can roll 2, 6, or7.

   We start by considering the case that A rolls 2. In this case, B can roll 1, 5, or 9, so he beats A 2/3 of the time. In other words, if A rolls a 1 he only beats B 1/3 of the time. Now what happens if A rolls 6? In this case B can still roll 1, 5, or 9, so (provided A rolls 6 we see that A beats B 2/3 of the time. Last, we see what happens when A rolls 7. In this case, B again rolls 1, 5, or 9, so again A beats B 2/3 of the time. Finally, A rolls each of his numbers with equal probability of 1/3, so to get the probability that A beats B we add up 1/3 of all the probabilities listed above to get

$$\text{Prob. that A beats B } = \frac{1}{3}\cdot\frac{1}{3}+\frac{1}{3}\cdot\frac{2}{3}+\frac{1}{3}\cdot\frac{2}{3}=\frac{5}{9}.$$

   We can represent these probabilities with a tree diagram as follows.



2. Show that Die B beats Die C with a probability of 5/9.

   Apply the same reasoning as in part (a).

3. Show that Die C beats Die A with a probability of 5/9.

   Apply the same reasoning as in part (a).

4. Is there a winning strategey for the person who chooses first? How about the person who chooses second?

   Here's we winning strategy for the player choosing second: If the first player chooses A then the second player should choose C. If the first player chooses B then the second player should choose A. If the first player chooses C then the second player chould choose B.

   The player choosing first has no winning strategy. They lose with probabily 5/9 regardless.

As you can see, the basic skills we need from these probability problems are an efficient way to organize information (which is why we talked about graphs) and some basic counting.

For the rest of these notes we'll cover some other basic ideas in probability (such as conditional probabilities, independent vs. dependent events, and random variables) using the same tools we've developed.

Before we begin, it's useful to introduce a little bit of notation. We write the set of all possible outcomes as X, and we call X the **sample space**. For any $x \in X$ we denote the probability of an outcome x by $\Pr(x)$, and read this as "the probability that x occurs." More generally, we can think of a subset $A \subset X$ as an event, and we can measure $\Pr(A) = \Pr(x \in A)$, which is the probability that a randomly chosen $x \in X$ is an element of A. For instance, in the strange dice problem above the sample space is the set of all possible die rolls for each of A and B, and an outcome can be $A = 6, B = 1$. The event we're trying to measure is the probability that $A > B$, and the computation we just did shows $\Pr(A > B) = \frac{5}{9}$, which is short-hand for saying "the probability that die A beats die B is 5/9."

Conditional probability: Sometimes you want to measure the probability of some event x given that y has already happened. For instance, the probability that it will rain on a given day is fairly low, but if it is cloudy in the morning, the probability it will rain that particular day is higher.

Here's another example. Suppose we roll two fair,6-sided dice. The probability that but the number we roll are odd is $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. However, suppose we know that the sum of these numbers is 10. Now what is the probability that both the numbers we rolled are odd?

We will write conditional probabilities as $\Pr(A|B)$, which we read as "the probability that A happens, given that B has does occur." We can compute this by the following rule (which is the product rule for probabilities of two events).

Rule 1. *If A and B are two events we can measure, and $\Pr(B) \neq 0$ then*

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)},$$

where $\Pr(A \cap B)$ the the probability that both A and B occurs. We can rewrite this as

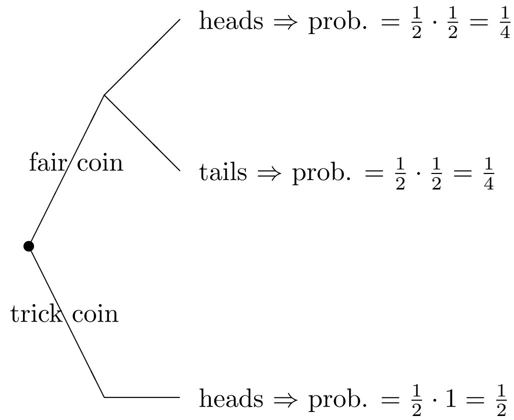
$$\Pr(A \cap B) = \Pr(B)\Pr(A|B).$$

Notice that this last expression is valid even if $\Pr(B) = 0$, as in this case both sides of the equation are 0.

In fact, the rule for conditional probabilities explains why we can compute probabilities using these tree diagrams. As we traverse along a path in the tree, we multiply the probabilities, which ends up computing the probabilities that a sequence of events occurs, which is precisely what we're trying to compute.

Here's an example. Suppose you flip one of two coins. One coin is a fair coin: it lands on the heads side half the time and on the tails side the other half. The other coin is a trick coin, and always lands on the heads side. Choose a coin at random (which means you choose the fair coin half the time and the trick coin the other half of the time), and flip it. If the coin comes up heads, what is the probability you flipped the fair coin?

We have two events we need to keep track of: we chose a coin and then we flipped it. We can again represent all the possible outcomes as a tree diagram.



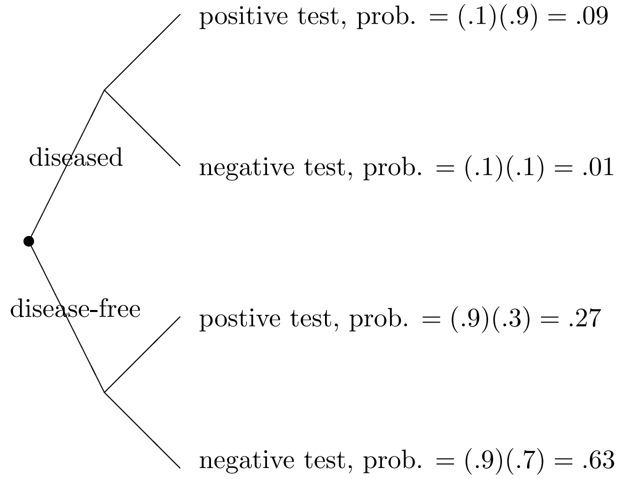
Now let's let A be the event that we choose the fair coin, and B be the event that we flipped heads. We have that $\Pr(A) = \frac{1}{2}$, and the tree diagram above shows us that $\Pr(B) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$. We want to know the probability that we chose the fair coin given that the did flip heads; by the product rule this is

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

We can consider another example, that of testing for an infectious disease. Suppose that in the city of Cape Town, about 10% of the population carries a fatal, infectious disease. We have a test to see if a given person has the disease, but the test is not perfect. More precisely:

- If you have the disease, there's a 10% chance the test say you don't (*i.e* a false negative).
- If you don't have the disease, there's a 30% chance the test will say you do (*i.e.* a false positive).

Well, suppose you test positive for this fatal disease. What is the probability that you actually have it? Just as we did with the previous example, we can write all the outcomes with a tree diagram.



Now we have two events to measure. Let A be the event you have this disease, and let B be the event you test positive for the disease. We have $\Pr(A) = .1$ and $\Pr(B) = .09 + .27 = .36$. We also have, from the tree diagram, $\Pr(A \cap B) = .09$. Now we can compute the conditional probability. Given that you do test positive for the disease, the probability that you actually have it is

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{.09}{.36} = .25.$$

This answer might surprise you: you've tested positive for this disease, but there's only a 1 in 4 chance you have it! However, it does make sense if you think about it. There are two ways you could test positive: either you do have the disease and the test is accurate, or you don't have the disease and the test is inaccurate. There are many many more people you don't have the disease than those that do, which is exactly why most of the positive tests are false positives.

Exercise: Suppose you test for this disease twice and both times, it comes back positive. Now what is the probability you have it? How about if you test twice, and the result comes back positive the first time, but negative the second?

There is something we need to be careful about with conditional probability, which we illustrate with the following example. At some carnivals, you can gamble playing a game called carnival dice. First you pick a number from 1 to 6, and then you roll three dice. You win if one of the three dice shows the number you picked, and otherwise you lose. It sounds plausible that you have a 50% chance of winning, because the number you pick can show up on any of the three dice you roll with a probability of $\frac{1}{6}$, and

$$\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

However, this is not true, as a more careful analysis reveals. Let A_i be the event that the ith die shows the number you pick. Indeed, our rule for counting the union of sets tells us

Rule 2. *Let A_1, A_2, and A_3 be sets of events in our sample space. Then*

$$\begin{aligned} \Pr(A_1 \cup A_2 \cup A_3) &= \Pr(A_1) + \Pr(A_2) + \Pr(A_3) - \Pr(A_1 \cap A_2) - \Pr(A_2 \cap A_3) - \Pr(A_3 \cap A_1) \\ &\quad + \Pr(A_1 \cap A_2 \cap A_3). \end{aligned}$$

Let's apply this rule to carnival dice, using the fact that all the dice are independent. We guess a number. The probability any one die matches this number is

$$\Pr(A_i) = \frac{1}{6};$$

the probability any two dice matches this number is

$$\Pr(A_i \cap A_j) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36};$$

the probability all three dice match the chosen number is

$$\Pr(A_1 \cap A_2 \cap A_3) = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{216}.$$

Now, by our rule for computing probabilities of unions of sets, we have

$$\Pr(\text{win}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} - \frac{1}{36} - \frac{1}{36} + \frac{1}{216} = \frac{91}{216} \simeq 41\%.$$

These are not very good odds of winning.

There are several other rules for computing conditional probabilities, and they all follow from the rules we have for counting unions of sets. The fundamental rule you need to remember is

$$\Pr(A \cap B) = \Pr(A|B) \cdot \Pr(B),$$

and everything else will follow.

Independent events: Some times we would like two events to be independent. For instance, if we flip two fair coins on opposite sides of the room, one coin shouldn't influence the other.

Definition 1. *Two events A and B in a probability sample space are independent if*

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B).$$

Some times we will assume two events are independent, usually because it is a physically reasonable thing to do. Other times we will check our data to see whether or not two events are independent.

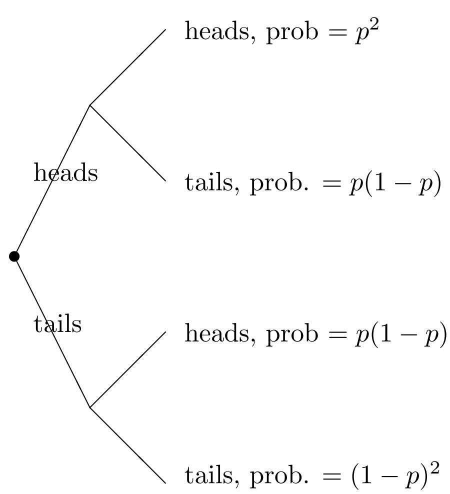
Let's look at an example in flipping two fair coins. Let A be the even that the first coin comes up heads, and B be the event that second coin comes up heads.

Exercise: Check that $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$ by drawing out the tree diagram.

Ok, this first example is more or less what we expected, so we examine another example. Let A be the even that the first coin comes up heads, and B the event that the two coins come up the same.

Exercise: Show that these two events are also independent, also by drawing out the tree diagram. Can you offer an intuitive explanation for why this is?

Let's now suppose the coins we're flipping are not exactly fair. The coins are weighted so they both come up heads with probability p, where p is some number between 0 and 1.



Again, let A be the event that the first coin comes up heads, and let B be the event that both coins come up the same. We have

$$\begin{aligned} \Pr(A) &= p \\ \Pr(B) &= p^2 + (1-p)^2 = 1 - 2p + 2p^2 \\ \Pr(A \cap B) &= p^2 \end{aligned}$$
$$\Pr(A) \cdot \Pr(B) = p(1 - 2p + 2p^2).$$

Finally we can compare these last two lines to answer the question: for which p are the events A and B independent? If A and B are independent we must have

$$p^2 = p(1 - 2p + 2p^2) \Rightarrow p = 0, \frac{1}{2}, 1.$$

Exercise: Suppose the first coin comes up heads with probability p and the second coin comes up heads with probability $1 - p$. Again, let A be the event that the first coin comes up heads, and let B be the event that the two coins come up the same. For which p are these two events independent?

We might have a very large sample space and several (*i.e.* more than two) events, and we can still ask for them to be independent. In what follows we talk about three events being independent, but the formulas generalize very naturally to any number of events.

Definition 2. *Let A_1, A_2, A_3 be events we want to measure in some probability sample space. We say that A_1, A_2, A_3 are mutually independent if $\Pr(A_i \cap A_j) = \Pr(A_i) \cdot \Pr(A_j)$ whenever $i \neq j$, and if $\Pr(A_1 \cap A_2 \cap A_3) = \Pr(A_1) \cdot \Pr(A_2) \cdot \Pr(A_3)$. On the other hand, we say that A_1, A_2, A_3 are pairwise independent if any two of them are independent.*

If we have k events, we can say they're mutually independent if any $(k-1)$-tuple of them are mutally independent and

$$\Pr(A_1 \cap \cdots \cap A_k) = \Pr(A_1) \cdots \Pr(A_k).$$

The definition of pairwise independent stays the same. By the definition, any set of mutually independent events are also pairwise independent; in fact, pairwise independence is the first thing you have to check. However, once you have four or more events, or a large sample space, it's hard to check if several events are mutually independent.

Let's consider an example to show that pairwise independence is not the same as mutual independence. Flip three fair coins, and define the events

- event A_1 is the first coin matching the second coin
- event A_2 is the second coin matching the third coin
- even A_3 is the third coin matching the first coin.

We'll show that A_1, A_2, A_3 are pairwise independent but not mutually independent.

First we write out the sample space. We write out each outcome in the sample space as a series of H or T, for each of the coin flips, so that we have

$$\{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}.$$

Because we've flipped independent, fair coins, each of these outcomes occurs with equal probability. There are eight elements of the sample space, so each outcome has probability $\frac{1}{8}$.

Next we compute the probability of A_1. This event can occur if we have HHH, HHT, TTH, TTT, so that

$$\Pr(A_1) = \Pr(HHH) + \Pr(HHT) + \Pr(TTH) + \Pr(TTT) = \frac{1}{2}.$$

A similar computation shows that $\Pr(A_2) = \frac{1}{2} = \Pr(A_3)$. What about $\Pr(A_1 \cap A_2)$? This event can correspond to either HHH or TTT, so we have

$$\Pr(A_1 \cap A_2) = \Pr(HHH) + \Pr(TTT) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \Pr(A_1) \cdot \Pr(A_2).$$

In fact, the only way we can have $x \in A_i \cap A_j$ for $i \neq j$ is to have $x = HHH$ or $x = TTT$. In other words, the only way for any two distinct pair of coins to agree is if either they're all heads or they're all tails. Thus we have, for $i \neq j$,

$$\Pr(A_i \cap A_j) = \Pr(HHH) + \Pr(TTT) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \Pr(A_i) \cdot \Pr(A_j),$$

and so we've just shown that the three events A_1, A_2, A_3 are pairwise independent.

The last thing we need to check is the triple intersection. We have

$$\Pr(A_1 \cap A_2 \cap A_3) = \Pr(HHH) + \Pr(TTT) = \frac{1}{4} \neq \Pr(A_1) \cdot \Pr(A_2) \cdot \Pr(A_3) = \frac{1}{8},$$

and so the events A_1, A_2, A_3 are not mutually independent.

Exercise: Do the same analysis for four coins.

As a last example in this question, we'll consider the following question: if all 118 students in 1044 show up for a lecture, what is the probability that two of them share a birthday? We'll have to make two assumptions to model this problem.

- Birthdays are randomly distributed throughout the year, so that any one person is equally likely to have his or her birthday on Jan. 15 as on Feb. 10 or Nov. 26. (In fact, this is not quite true.)
- The birthdays of all the students are mutually independent.

Now we'll analyze a more general problem, which includes this birthday question. Suppose there are N days in the year, we have m people with birthdays $\{b_1, b_2, \ldots, b_m\}$ where each birthday $b_1, \ldots, b_m$ is a number from 1 to N. What is the probability that all the numbers $b_1, \ldots, b_m$ are distinct?

Exercise: Notice that we need to have $N \geq m$ in order that it is possible to have all the birthdays on distinct days. Explain why this is.

The tree diagram for this example is very large, so we won't draw it. However, we can still do the computation. The sample space is the set of sequences of integers

$$X = \{(b_1, \ldots, b_m) : 1 \leq b_i \leq N\}, \qquad |X| = N^m.$$

The event we're trying to measure the probability of is

$$A = \{x = (b_1, \ldots, b_m) \in X : b_i \neq b_j \text{ whenever } i \neq j\}.$$

Then we have the probability that there are two people in the room with the same birthday as $1 - \Pr(A)$.

Observe that, because we have assumed the birthdays are independent, all of these sequences have the same probability. This means that for any sequence of birthdays $x = (b_1, \ldots, b_m)$ we have $\Pr(x) = N^{-m}$. Now we can compute $\Pr(A)$ by summing all the probabilities of the elements of A. However, all these outcomes in A have the same probability, so we only need to figure out how many elements A has. Let's count the number of elements of A: we have N choices for the first term in the sequence (*i.e.* the birthday of the first student), $N - 1$ choices for the second, and so on down to the mth student. We have

$$\Pr(A) = N^{-m}(\#A) = \frac{N(N-1)(N-2)\cdots(N-m+1)}{N^m} = \frac{N!}{(N-m)!N^m}.$$

We can substitute in $N = 365$ (ignoring leap years) and $m = 118$ to get

$$\Pr(A) = \frac{365!}{247! \cdot 365^{118}} \simeq 5.34 \times 10^{-10},$$

which is a very tiny number. This means we're almost certain to have two people in our class with the same birthday.

The answer we just computed is sort of satisfying, but not quite. What is would really like to know is: how many people do you need to get in a room before the probability that two of them have the same birthday is at least 1/2? So, let's turn our formula around. We'd like to find the least m such that

$$\Pr(A) = \frac{365!}{(365-m)!365^m} < \frac{1}{2}.$$

Now, we'd like to solve for m, but this is too difficult. Instead, we will find a different formula for $\Pr(A)$.

We do this as follows. Number the students from 1 to m, and let E_i be the event that the ith person has a birthday different from the first $i-1$ people. For instance, E_2 is the even that the second person doesn't have the same person as the first, and E_3 is the event that the third person has a different birthday from the first and second (though in the case of E_3 the first two people can have the same birthday). Now, by our formula for conditional probabilities, we have

$$\begin{aligned} \Pr(A) &= \Pr(E_1 \cap E_2 \cap \cdots \cap E_m) \\ &= \Pr(E_1) \cdot \Pr(E_2 | E_1) \cdots \Pr(E_m | E_1 \cap E_2 \cap \cdots \cap E_{m-1}). \end{aligned}$$

This formula with all the conditional probabilities looks scary, but it's actually not. For instance, $\Pr(E_1) = 1$, because there are no conditions to satisfy. Then E_2 occurs when the first and second person have different birthdays, which means

$$\Pr(E_2 | E_1) = 1 - \frac{1}{365}.$$

Remember that all the birthdays are spread randomly throughout the year, and that the days of all the students birthdays are mutually independent, so that

$$\Pr(E_3 | E_1 \cap E_2) = 1 - \frac{2}{365}.$$

By a similar reasoning, we have

$$\Pr(E_k | E_1 \cap \cdots E_{k-1}) = 1 - \frac{k}{365},$$

which (plugging into our formula above) means

$$\Pr(A) = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{m-1}{365}\right),$$

which is the formula we're looking for.

Hmm, this is still not quite satisfying. We have another formula, but it still seems very hard to set $\Pr(A) = 1/2$ and solve for m. So, instead of solving exactly for m, we make an approximation. We exponentiate and take a logarithm to get

$$\Pr(A) = e^{\ln(1-1/365) + \ln(1-2/365) + \cdots + (1-(m-1)/365)}.$$

Now we can approximate (hint: Taylor's theorem)

$$\frac{-x}{1-x} \le \ln(1-x) \le -x,$$

so that

$$e^{-\frac{m(m-1)}{2(365-m)}} \leq \Pr(A) \leq e^{-\frac{m(m-1)}{2\cdot 365}}.$$

Now, we've already figured out that the number of students where the probability is 1/2 is going to be much less than 365, so we may as well say

$$\Pr(A) \simeq e^{-\frac{m(m-1)}{2\cdot 365}}.$$

Now it's easy to solve for m, and we see conclude what is known as the birthday principle:

If there are more than $\sqrt{2\ln(2)\cdot 365}$ people in a room, then there is a greater than 1/2 probability that two of them have the same birthday. In fact, we can apply the same reasoning to years with N days in them (in case you move to Mars or Venus). If there are more than $\sqrt{2\ln(2)N}$ people in a room, then there is a greater than 1/2 probability that two of them have the same birthday.