Week 2

20 May 2021

7 Conditional probability

A lot of times we want to know the probability of something occurring while also having some information about what has already happened. As a quick example, imagine your friend has just rolled two dice, but they kept the results from you. They ask you what are the chances that they got snake eyes (two 1s). They then let you see one of the dice and it's a 1! We now have more information and the question becomes whether this new information will change the probability that they rolled two 1s. This is the idea of conditional probability.

Example 7.1 Let's actually go into this example a little deeper and see what happens. First we start by figuring out the initial probability that your friend rolled two 1s on their die. Since each dice has 6 sides there are 36 different possibilities:

$$\begin{array}{cccccc}
(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\
(2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\
(3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\
(4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\
(5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\
(6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6)
\end{array}$$

The set of these thirty-six outcomes gives us the sample space Ω. Our friend is asking what the chances are that they rolled two 1s. Looking at our sample space, we notice that $A = \{(1,1)\}$ which means our probability is __________________.

At this point your friend shows you that one of the two die is in fact a 1! What is the probability that the other dice is a 1 as well? We might think that since the other dice has 6 options, that the chance should be

Common student mistake

$\frac{1}{6}$, but that's not exactly correct. Your friend has a *choice* on which die they showed you. It could have been the first one or the second one and so it's not as simple as $\frac{1}{6}$. Instead, what we have to do is look at all of our original 36 possibilities and ask which ones have a 1 in them to see our new sample space. We're left with:

So our sample space has 11 elements and only 1 of them is $(1, 1)$. Our probability is $\frac{1}{11}$ which is much less than $\frac{1}{6}$.

How do we represent this mathematically? Notice that our event A never changed. The event stayed as "two 1s" and so $A = \{(1, 1)\}$ throughout the whole process. Instead, we added more information. Our friend told us that "one of the dice is in fact a 1". Therefore we have a new event B which is "one of the dice is 1" which gives us the set of 11 elements from above. Our sample space is therefore B and our new set is whatever is in both A and B.

In other words, the *conditional probability of A given B* is

Wikipedia: Conditional probability

Although this is nice, we want to try and work with probabilities as much as we can. So what we do is we convert the previous formula into probabilities:

Another way to write this is the following:

We call this formula the *multiplication rule*. Intuitively what this is saying is if the event B happens around 1/2 the time and if 1/4 the time that B happens the event A happens , then A and B happen about 1/8 the time.

7.1 Tree diagrams

We can actually represent all of this with what are known as tree diagrams. A *tree diagram* is basically a scheme which allows us to graphically represent sample spaces for conditional probability. The best way to see a tree diagram is through an example.

Example 7.2 Suppose we have three bags of chocolate, each one containing three different types of chocolate: White (W), Dark (D) and Milk (M). But we know that the the quantities in each bag are different! If the first bag has 2 whites and 1 milk, the second bag has 2 dark and 2 milk and the third bag has 1 white, 1 dark and 2 milk, what is the probability that we will get a dark chocolate if we randomly draw a chocolate from a random bag?

Common Student Mistake: Before we use conditional probability, let's talk about a common student mistake at this point. A common mistake is to just add up all of the quantities ($2+1$ white, $1+2+2$ milk and $2+1$ dark) and say that since there are 3 dark chocolates and 11 chocolates in total, you have a $3/11$ chance of getting a dark chocolate! This is wrong because you are completely forgetting about the bags! If you chose the first bag, you have absolutely no chance of getting dark chocolate. It's important to keep track of *all* the information given in a problem. Every word counts.

With the warning out of the way this is a perfect time to use conditional probability and tree diagrams. For this we set up our diagram in the following way:

Example 7.3 Let's look at another example that might deal more with real life. Say you're midway through a class and want to know what your chances are of passing the class. You need to pass both exams to pass the class. You've studied hard for the exam coming up and you're sure that you have a 95% chance of passing the first exam. The second exam is harder to predict though and you think that if you pass the first exam, then you have a 90% chance of passing the second exam, but if you fail the first exam your chances will go down to 75%. What is the chance you will pass the class?

We start off by making a tree diagram as before.

What are your chances of passing at least one exam? What about passing exactly one exam?

This second example is a good time to talk about average conditional probabilities. We had set our event A to be "pass exactly one exam" and we can condition that on the event B which is "pass the first exam".

What does this look like mathematically? We actually already saw this above! We had

But notice that this is just a partition! We can extend this argument to any partition to get the *law of total probability*.

Wikipedia: Law of total probability

Note that the book calls this the *Rule of average conditional probabilities*.

Theorem 7.4 (Law of total probability) *If $A_1, A_2, \ldots A_n$ is a partition of Ω then*

8 Independence

In the examples we saw in the previous section, our second event was always *dependent* on our first event. But this is not always the case. Think of rolling two die. If I roll the first dice, it doesn't magically change the probability for the second dice. Since the second die is not dependent on the first we say the two events are independent.

Mathematically, we think of this as the probability of an event A not changing no matter whether B occurs or not:

In this case say A and B are *independent*. A much more simple way to view this definition is the following:

Wikipedia: Independence

Working this around we have:

This gives us the *multiplication rule for independent events*

The multiplication rule is often taken to be the definition of independence (for example, Wikipedia uses this as the definition of independence). Any of the three definitions I have given are all valid responses for the definition of independence.

Example 8.1 Let's look at flipping a coin twice as an example. We first will draw the tree diagram:

We easily see that our events are independent just by looking at the diagram.

9 Sequence of events

We started looking at doing one event followed by another, but what happens if we have a chain of multiple events? Say I flip 10 different coins, or I draw 7 cards, or I pull out 34 bars of chocolate from a bag. How do we find the probability of a sequence of events?

Let's first start off slow. If we have two events we saw:

$$P(A \cap B) = P(B)P(A \mid B) = P(A)P(B \mid A)$$

How about if we have three events A, B, and C? Since we're going in order C is dependent on A and B, *i.e.*, $A \cap B$. So we have:

$$P((A \cap B) \cap C) = P(A \cap B)P(C \mid (A \cap B)) = P(A)P(B \mid A)P(C \mid (A \cap B))$$

We should start seeing a pattern at this point. It turns out, we can keep doing this forever and we get a *multiplication rule for n events*

Example 9.1 Let's look at a quick example of how this might work using our tree diagrams. Suppose we want to flip a coin two times with the condition that if we flip a tails, then we draw a number between 1 and 3 out of a bag. Once we've drawn a number out, then we stop. If the second flip is a heads, then we consider that as drawing a 0. The tree diagram would look like:

What is the probability that we flipped exactly 1 tails?

What is the probability that we pulled a 1 out of the bag?

We can also extend the notion of independence to an arbitrary sequence of events. If we have n events $A_1, A_2, \ldots, A_n$, then we say they are *independent* if they are pairwise independent. What this means is for any two events A_i and A_j then A_i and A_j are independent. In this case, the multiplication rule gives us:

An easy example of this is to just think of a coin toss. If you flip multiple coins the coin flip of one coin doesn't influence another coin's result. So the order that you flip the coins doesn't matter.

Example 9.2 We finish with an extremely standard example of how things in probability have unexpected results, can be complicated and benefit from having structure. This problem is called the *birthday problem*. It's a simple question: Say there are n people in a room. What is the probability that at least two people have the same birthday? (We ignore leap days...sorry)

10 Bayes' Rule

Let's go back to the chocolate in a bag example from before. Recall that we had three different bags which contained different chocolates. The first bag has 2 whites and 1 milk, the second bag has 2 dark and 2 milk and the third bag has 1 white, 1 dark and 2 milk. We were told we first randomly chose a bag and then pick out a chocolate. The question we asked at the time was, what are the chances the chocolate is dark.

Now, we're going to ask a slightly reverse question. Say that I randomly pick a bag and pull out a chocolate and I show you that I pulled out a dark chocolate. I then ask you, which bag do you think I pulled the chocolate out of. Basically, we want to try and go backwards with probability. If you look at the bags, the second bag has the most dark chocolates so we'd predict that the second bag was the bag I most likely pulled the chocolate from.

Let's analyze this mathematically.

Let's try and generalize this. Let A be our event "pulled a dark chocolate". Then we have three "prior" events which are the three bag choices: B_1, B_2, B_3. What we wanted to calculate was ________________. To do this we did:

This is called *Bayes' Rule*.

Wikipedia: Bayes' rule

Bayes' rule is sometimes called Bayes' theorem, Bayes' law or Bayes-Price theorem.

Note that we're starting to notice that $P(A \mid B)$ and $P(B \mid A)$ are both things we can measure! So it makes sense to start differentiating which event comes before another event if we have a sequence of events. If event A happens before event B then $P(A \mid B)$ is called the *posterior*

Wikipedia: Posterior probability
Wikipedia: Likelihood
Wikipedia: Prior probability

probability of A given B and $P(B \mid A)$ is known as the *likelihood of B given a fixed A*. The probability of event A (*i.e.*, $P(A)$) is referred to as the *prior probability*.

Recall that earlier I asked the following question: "Say that I randomly pick a bag and pull out a chocolate and I show you that I pulled out a dark chocolate. What are your chances of guessing which bag I pulled the chocolate out of?" Let's next say that I'm being manipulative and I pose the following question instead of the one from before. "Say that I pick a bag and pull out a chocolate and I show you that I pulled out a dark chocolate. What are your chances of guessing which bag I pulled the chocolate out of?"

Common student error: It's the same as before!

It's actually not the same as before! What is different about the second question? ______________________________. Usually when we use the word "randomly" we are making the *implicit assumption* that all outcomes are equally likely to occur. So when I "randomly chose" a bag, each bag has $\frac{1}{3}$ chance of being chosen. But what if it's no longer random? This actually makes the problem *much* harder because you no longer know the values of $P(B_i)$ from above. At this point, it's better to keep them as variables since you don't have enough information to solve the problem. This gives the following: