

Chapter 7 Conditional Probability

7.1 Conditional Probability

So far we've talked simply about the probability of various propositions. But sometimes we're not interested in the absolute probability of a proposition, we're interested in its **conditional** probability. That is, we're interested in the probability of the proposition *assuming* or *conditional on* some other proposition obtaining.

For example, imagine we're trying to decide whether to go to a party. At first glance, we might think that one of the factors that is relevant to our decision is the probability that it will be a successful party. But on second thought that isn't particularly relevant at all. If the party is going to be unpleasant if we are there (because we'll annoy the host) but quite successful if we aren't there, then it might be quite probable that it will be a successful party, but that will be no reason at all for us to go. What matters is the probability of it being a good, happy party *conditional* on our being there.

It isn't too hard to visualise how conditional probability works if we think of measures over lines on the truth table. If we assume that something, call it B is true, then we should 'zero out', i.e. assign probability 0, to all the possibilities where B doesn't obtain. We're now left with a measure over only the B -possibilities. The problem is that it isn't a normalised measure. The values will only sum to $Pr(B)$, not to 1. We need to renormalise. So we divide by $Pr(B)$ and we get a probability back. In a formula, we're left with

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

We can work through an example of this using a table that we've seen once or twice in the past.

p	q	r	
T	T	T	0.0008
T	T	F	0.008
T	F	T	0.08
T	F	F	0.8
F	T	T	0.0002
F	T	F	0.001
F	F	T	0.01
F	F	F	0.1

Assume now that we're trying to find the conditional probability of p given q . We could do this in two different ways.

First, we could set the probability of any line where q is false to 0. So we will get the following table.

p	q	r	
T	T	T	0.0008
T	T	F	0.008
T	F	T	0
T	F	F	0
F	T	T	0.0002
F	T	F	0.001
F	F	T	0
F	F	F	0

The numbers don't sum to 1 any more. They sum to 0.01. So we need to divide everything by 0.01. It's sometimes easier to conceptualise this as multiplying by $\frac{1}{Pr(q)}$, i.e. by multiplying by 100. Then we'll end up with.

p	q	r	
T	T	T	0.08
T	T	F	0.8
T	F	T	0
T	F	F	0
F	T	T	0.02
F	T	F	0.1
F	F	T	0
F	F	F	0

And since p is true on the top two lines, the 'new' probability of p is 0.88. That is, the conditional probability of p given q is 0.88. As we were writing things above, $Pr(p|q) = 0.88$.

Alternatively we could just use the formula given above. Just adding up rows gives us the following numbers.

$$\begin{aligned}
 Pr(p \wedge q) &= 0.0008 + 0.008 = 0.0088 \\
 Pr(q) &= 0.0008 + 0.008 + 0.0002 + 0.001 = 0.01
 \end{aligned}$$

Then we can apply the formula.

$$\begin{aligned}
 Pr(p|q) &= \frac{Pr(p \wedge q)}{Pr(q)} \\
 &= \frac{0.0088}{0.01} \\
 &= 0.88
 \end{aligned}$$

7.2 Bayes Theorem

It is often easier to calculate conditional probabilities in the 'inverse' direction to what we are interested in. That is, if we want to know $Pr(A|B)$, it might be much easier to discover $Pr(B|A)$. In these cases, we use Bayes Theorem to get

the right result. I'll state Bayes Theorem in two distinct ways, then show that the two ways are ultimately equivalent.

$$\begin{aligned} Pr(A|B) &= \frac{Pr(B|A)Pr(A)}{Pr(B)} \\ &= \frac{Pr(B|A)Pr(A)}{Pr(B|A)Pr(A) + Pr(B|\neg A)Pr(\neg A)} \end{aligned}$$

These are equivalent because $Pr(B) = Pr(B|A)Pr(A) + Pr(B|\neg A)Pr(\neg A)$. Since this is an independently interesting result, it's worth going through the proof of it. First note that

$$\begin{aligned} Pr(B|A)Pr(A) &= \frac{Pr(A \wedge B)}{Pr(A)}Pr(A) \\ &= Pr(A \wedge B) \\ Pr(B|\neg A)Pr(\neg A) &= \frac{Pr(\neg A \wedge B)}{Pr(\neg A)}Pr(\neg A) \\ &= Pr(\neg A \wedge B) \end{aligned}$$

Adding those two together we get

$$\begin{aligned} Pr(B|A)Pr(A) + Pr(B|\neg A)Pr(\neg A) &= Pr(A \wedge B) + Pr(\neg A \wedge B) \\ &= Pr((A \wedge B) \vee (\neg A \wedge B)) \\ &= Pr(B) \end{aligned}$$

The second line uses the fact that $A \wedge B$ and $\neg A \wedge B$ are inconsistent, which can be verified using the truth tables. And the third line uses the fact that $(A \wedge B) \vee (\neg A \wedge B)$ is equivalent to B , which can also be verified using truth tables. So we get a nice result, one that we'll have occasion to use a bit in what follows.

$$Pr(B) = Pr(B|A)Pr(A) + Pr(B|\neg A)Pr(\neg A)$$

So the two forms of Bayes Theorem are the same. We'll often find ourselves in a position to use the second form.

One kind of case where we have occasion to use Bayes Theorem is when we want to know how significant a test finding is. So imagine we're trying to decide whether the patient has disease D , and we're interested in how probable it is that the patient has the disease conditional on them returning a test that's positive for the disease. We also know the following background facts.

- In the relevant demographic group, 5% of patients have the disease.
- When a patient has the disease, the test returns a positive result 80% of the time
- When a patient does not have the disease, the test returns a negative result 90% of the time

So in some sense, the test is fairly reliable. It usually returns a positive result when applied to disease carriers. And it usually returns a negative result when applied to non-carriers. But as we'll see when we apply Bayes Theorem, it is very unreliable in another sense. So let A be that the patient has the disease, and B be that the patient returns a positive test. We can use the above data to generate some 'prior' probabilities, i.e. probabilities that we use prior to getting information about the test.

- $Pr(A) = 0.05$, and hence $Pr(\neg A) = 0.95$

- $Pr(B|A) = 0.8$
- $Pr(B|\neg A) = 0.1$

Now we can apply Bayes theorem in its second form.

$$\begin{aligned}
 Pr(A|B) &= \frac{Pr(B|A)Pr(A)}{Pr(B|A)Pr(A) + Pr(B|\neg A)Pr(\neg A)} \\
 &= \frac{0.8 \times 0.05}{0.08 \times 0.05 + 0.1 \times 0.95} \\
 &= \frac{0.04}{0.04 + 0.095} \\
 &= \frac{0.04}{0.135} \\
 &\approx 0.296
 \end{aligned}$$

So in fact the probability of having the disease, conditional on having a positive test, is less than 0.3. So in that sense the test is quite unreliable.

This is actually a quite important point. The fact that the probability of B given A is quite high does not mean that the probability of A given B is equally high. By tweaking the percentages in the example I gave you, you can come up with cases where the probability of B given A is arbitrarily high, even 1, while the probability of A given B is arbitrarily low.

Confusing these two conditional probabilities is sometimes referred to as the *prosecutors' fallacy*, though it's not clear how many actual prosecutors are guilty of it! The thought is that some prosecutors start with the premise that the probability of the defendant's blood (or DNA or whatever) matching the blood at the crime scene, conditional on the defendant being innocent, is 1 in a billion (or whatever it exactly is). They conclude that the probability of the defendant being innocent, conditional on their blood matching the crime scene, is about 1 in a billion. Because of derivations like the one we just saw, that is a clearly invalid move.

7.3 Conditionalisation

The following two concepts seem fairly closely related.

- The probability of some hypothesis H given evidence E
- The new probability of hypothesis H when evidence E comes in

In fact these are distinct concepts, though there are interesting philosophical questions about how intimately they are connected.

The first one is a *static* concept. It says, at one particular time, what the probability of H is given E . It doesn't say anything about whether or not E actually obtains. It doesn't say anything about changing your views, or your probabilities. It just tells us something about our current probabilities, i.e. our current measure on possibility space. And what it tells us is what proportion of the space where E obtains is occupied by possibilities where H obtains. (The talk of 'proportion' here is potentially misleading, since there's no physical space to measure. What we care about is the measure of the $E \wedge H$ space as a proportion of the measure of the E space.)

The second one is a *dynamic* concept. It says what we do when evidence E actually comes in. Once this happens, old probabilities go out the window, because we have to adjust to the new evidence that we have to hand. If E indicates H , then the probability of H should presumably go up, for instance.

Because these are two distinct concepts, we'll have two different symbols for them. We'll use $Pr(H|E)$ for the static concept, and $Pr_E(H)$ for the dynamic concept. So $Pr(H|E)$ is what the current probability of H given E is, and $Pr_E(H)$ is what the probability of H will be when we get evidence E .

Many philosophers think that these two should go together. More precisely, they think that a rational agent always updates by *conditionalisation*. That's just to say that for any rational agent, $Pr(H|E) = Pr_E(H)$. When we get evidence E , we always replace the probability of H with the probability of H given E .

The conditionalisation thesis occupies a quirky place in contemporary philosophy. On the one hand it is almost universally accepted, and an extremely interesting set of theoretical results have been built up using the assumption it is true. (Pretty much everything in Bayesian philosophy of science relies in one way or another on the assumption that conditionalisation is correct. And since Bayesian philosophy of science is a thriving research program, this is a non-trivial fact.) On the other hand, there are remarkably few direct, and plausible, arguments in favor of conditionalisation. In the absence of a direct argument we can say two things.

First, the fact that a lot of philosophers (and statisticians and economists etc) accept conditionalisation, and have derived many important results using it, is a reason to take it seriously. The research programs that are based around conditionalisation do not seem to be degenerating, or failing to produce new insights. Second, in a lot of everyday applications, conditionalisation seems to yield sensible results. The simplest cases here are cases involving card games or roulette wheels where we can specify the probabilities of various outcomes in advance.

Let's work through a very simple example to see this. A deck of cards has 52 cards, of which 13 are hearts. Imagine we're about to draw 2 cards, without replacement, from that deck, which has been well-shuffled. The probability that the first is a heart is $\frac{13}{52}$, or, more simply, $\frac{1}{4}$. If we assume that a heart has been taken out, e.g. if we draw a heart with the first card, the probability that we'll draw another heart is $\frac{12}{51}$. That is, conditional on the first card we draw being a heart, the probability that the second is a heart is $\frac{12}{51}$.

Now imagine that we do actually draw the first card, and it's a heart. What should the probability be that the next card will be a heart? It seems like it should be $\frac{12}{51}$. Indeed, it is hard to see what else it could be. If A is *The first card drawn is a heart* and B is *The second card drawn is a heart*, then it seems both $Pr(A|B)$ and $Pr_B(A)$ should be $\frac{12}{51}$. And examples like this could be multiplied endlessly.

The support here for conditionalisation is not just that we ended up with the same result. It's that we seem to be making the same calculations both times. In cases like this, when we're trying to figure out $Pr(A|B)$, we pretend we're trying to work out $Pr_B(A)$, and then stop pretending when we've worked out the calculation. If that's always the right way to work out $Pr(A|B)$, then $Pr(A|B)$ should always turn out to be equal to $Pr_B(A)$. Now this argument goes by fairly quickly obviously, and we might want to look over more details later on, but it's easy to see we might take conditionalisation to be a plausible model for updating probabilities.

