Bayes' Theorem and the Prosecutor's Fallacy

The spotty dice problem

Suppose I have two six-sided dice. One has spots on all six faces, the other has one spot on one face only. I put them in a bag and draw one out at random and throw it. It shows a spot. What is the probability that I drew the 6-spotted die?

This is a classic problem in what is known as conditional probability. Obviously the *a priori* probability of drawing the 6-spotted die out of the bag is 1/2 (because there are two dice in the bag) but that is not what is required. What I want to know is that *given that the die drawn shows a spot*, what is the probability of it being the 6-spotted die? Intuitively we feel that it is much more likely to be the 6-spotted die than the single-spotted die but is this correct?

Now there are 12 possibilities all of equal probability, 6 for each die. Out of these 12 possibilities, 7 show a spot; and out of these 7 possibilities, 6 are due to the six-spotted die and 1 is due to the single-spotted die. Given that I see a spot, the chances that I drew the six-spotted die are therefore 6/7. Our intuition is not wrong.

The calculation in this case is relatively easy but in order to work out the probabilities in more complicated cases, we need some help. This comes in the form of

Bayes' Theorem

Bayes' theorem states that the probability of an outcome **X** occurring *given* a possibly related condition **C** is equal to the *a priori* probability of both **X and C** occurring divided by the *a priori* probability of **C** on its own. In symbols

$$p(\mathbf{X} \mid \mathbf{C}) = \frac{p(\mathbf{X} \,\&\, \mathbf{C})}{p(\mathbf{C})} \tag{1}$$

(By *a priori* I mean the probability of the event occurring without any other conditions.)

Let's see if Bayes' theorem gets the same result. First we must define **X** and **C** very carefully.

X : the die I drew was the 6-spotted die
C : the die cast shows a spot

First, what is the *a priori* probability of **X & C** occurring? That is to say: what is the probability that a) I drew the 6-spotted die **and** b) the die shows a spot?

Now obviously if I drew the 6-spotted die it was bound to show a spot so the probability I want is the same as the probability that I drew the 6-spotted die so

$p(\mathbf{X} \,\&\, \mathbf{C}) = 1/2$

Second, what is the *a priori* probability the the die drawn shows a spot. We have seen that there are 12 possibilities, 7 of which show a spot so

$p(\mathbf{C}) = 7/12$

$$p(\mathbf{X} \mid \mathbf{C}) = \frac{p(\mathbf{X} \,\&\, \mathbf{C})}{p(\mathbf{C})} = \frac{1/2}{7/12} = 6/7 \tag{2}$$

Now suppose you have 6 dice one of which has six spots. You pull one die out of the bag and throw it and it shows a spot. What does your intuition tell you about the probability that the die you drew was the 6-spotted die? Let's see if you are right.

In this case there are 36 possibilities of which 6 + 5 = 11 show a spot and of these 11, 6 are due to the 6-spotted die. The probability that I drew the latter is therefore 6/11.

The Bayes' calculation is as follows:

$p(\mathbf{X}\ \&\ \mathbf{C}) = p(\mathbf{X}) = 1/6$

$p(\mathbf{C}) = 11/36$

$p(\mathbf{X} \mid \mathbf{C}) = 1/6$ over $11/36 = 6/11$.

I hope you are now convinced that Bayes' theorem really works. We shall now use it to find the answer in the general case where we have a large number (n) of 6-sided dice in the bag. The calculation goes as follows. **X** and **C** are the same as before:

X : the die I drew was the 6-spotted die
C : the die cast shows a spot

$p(\mathbf{X}\ \&\ \mathbf{C}) = p(\mathbf{X}) = 1/n$

This time there are $6n$ possibilities (n dice each with 6 faces). Of these possibilities the total number which show a spot are 6 + 1 + 1 + 1 + … for a total of n dice. This can be written as $6 + (n - 1)$ which means that

$p(\mathbf{C}) = (6 + (n - 1))/6n$

so
$$p(\mathbf{X} \mid \mathbf{C}) = \frac{1/n}{(6 + (n-1))/6n} = \frac{6}{5+n} \tag{3}$$

(Putting $n = 2$ and $n = 6$ gives us the correct answers of 6/7 and 6/11 respectively)

As a lead in to my discussion of the prosecutor's fallacy, we need to generalise the spotted dice problem to the case where I have n dice each of which has m faces.

$p(\mathbf{X}\ \&\ \mathbf{C}) = 1/n$ (as before)

Since each die has m faces there are mn possibilities. Of these possibilities the total number which show a spot are $m + 1 + 1 + 1 + \ldots$ or $m + (n - 1)$ which means that

$p(\mathbf{C}) = (m + (n - 1))/mn$

so
$$p(\mathbf{X} \mid \mathbf{C}) = \frac{1/n}{(m + (n-1))/mn} = \frac{m}{m+n-1} \tag{4}$$

For example, if we have 100 dice ($n = 100$) each of which has 10 faces ($m = 10$), all of them having one spot except one die which has 10 spots, what is the probability that, given that the die drawn shows a spot, the die drawn is the 10-spotted die?

$$p(\mathbf{X} \mid \mathbf{C}) = \frac{10}{10+100-1} = \frac{10}{109} = 0.0917 \text{ or approximately } 9\% \tag{5}$$

I hope that these simple examples have convinced you that conditional probabilities are not difficult either to calculate or to understand and that usually your intuition is not far wrong either.

Now let's take a trip to the Old Bailey

The prosecutor's fallacy

A girl has been found brutally raped and murdered on the overnight train from Euston to Glasgow. A certain Mr Scroggins has been accused of the crime. The prosecuting barrister stands to make his case:

“M'lud. Ladies and gentlemen of the jury. You have heard that the police have been most diligent in tracking down the perpetrator of this heinous crime. After trawling through their extensive database of DNA samples, they have discovered only one person in the database that had DNA matching the DNA found on the poor helpless victim at the scene of the crime and that person was the defendant who now stands before you in the dock. You have also heard the testimony of the expert witness Professor Noitall who has categorically stated that the chance of finding a match between two samples of DNA from different people is one in a million[1]. I repeat – 1 in 1,000,000. I submit that the probability that the defendant is innocent is also one in a million. I rest my case."

What jury could resist such compelling evidence?

And yet the conclusion is wrong – horribly, horribly wrong.

Much depends on the size of the 'extensive' database. Let us suppose that the incident takes place at a time when the database includes the whole of the male population of the UK – say, 30 million men[2]. If Professor Noitall is right then the police would expect to find approximately 30 matches of which one belonged to the perpetrator and 29 to innocent people. We are told, however, that only one match was found. This is unlikely but by no means impossible and will happen occasionally.

Now we need to know that, given that there was a match between the two samples, what is the probability that the defendant is guilty. The situation is exactly the same as the spotty dice game but this time there are 30 million dice in the bag and each dice has 1 million faces. Having pulled out an individual out of the database and found it to come up with a match, we want to know what is the probability that it is the man whose DNA is *bound* to match the sample from the crime scene and not one of the men whose DNA matches it by accident. We can use our formula to calculate the answer:

$n = 30{,}000{,}000$
$m = 1{,}000{,}000$

X: Mr Scroggins is guilty
C: A match was found between Mr Scroggins' DNA and that found at the scene

$$p(\mathbf{X} \mid \mathbf{C}) = \frac{1{,}000{,}000}{1{,}000{,}000 + 30{,}000{,}000 - 1} \approx \frac{1}{31}$$

In other words, the defendant is almost certainly innocent.

How can the jury have been persuaded so easily to make such a catastrophic mistake? It can't be because the jury are, understandably, unfamiliar with Bayes' Theorem. There must be another reason why our intuitive sense of probability has been so drastically led astray. I believe the reason is something like this.

Professor Noitall testified that the chance of a match between two different people was 1 in a million. More specifically what he is saying is the *if Mr Scroggins is innocent* there would be a 1 in a million chance of finding a match. The prosecuting barrister has taken this to mean that *if there is a match*, there is a 1 in a million chance of Mr Scroggins being innocent. He has swapped the outcome and the condition. Mathematically he has assumed that $p(\mathbf{X} \mid \mathbf{C})$ is the same as $p(\mathbf{C} \mid \mathbf{X})$. This is emphatically not the case. Let us see why. From equation (1) we have:

$$p(\mathbf{X} \mid \mathbf{C}) = \frac{p(\mathbf{X} \,\&\, \mathbf{C})}{p(\mathbf{C})} \quad \text{and} \quad p(\mathbf{C} \mid \mathbf{X}) = \frac{p(\mathbf{C} \,\&\, \mathbf{X})}{p(\mathbf{X})} \tag{6}$$

Now $p(\mathbf{X} \,\&\, \mathbf{C})$ is the same as $p(\mathbf{C} \,\&\, \mathbf{X})$ so we can eliminate this expression to get

$$p(\mathbf{X} \mid \mathbf{C}) = p(\mathbf{C} \mid \mathbf{X}) \frac{p(\mathbf{X})}{p(\mathbf{C})} \tag{7}$$

$p(\mathbf{X} \mid \mathbf{C})$ is *not* the same as $p(\mathbf{C} \mid \mathbf{X})$.

1 In fact the accepted figure these days is more like 1 in a *billion*
2 Currently in the UK the police hold a database containing the DNA records of about 5 million individuals

But this doesn't really explain why the barrister's argument sounds so persuasive. I believe that the reason has more to do with psychology than mathematics. The prosecutor has taken the very small probability figure supplied by his expert and planted it into the minds of the jury who then reason as follows: this figure of 1 in a million is obviously important so it must have something to do with Mr Scoggin's guilt or innocence. Clearly it can't be the case that there is a 1 in a million chance that he is guilty otherwise the prosecutor would never have brought the case in the first place; it follows that there must be a 1 in a million chance that he is innocent!

Now we should not reject this argument as being stupid. It is indeed the case that the figure of 1 in a million is important and relevant to the case. It is also true that the fact that this probability is very small is closely related to the conclusion that Mt Scroggins is very probably guilty. But we need to show mathematically why this argument is faulty. Equation (3) tells us that the probability of the defendant's *guilt* is:

$$p(\mathbf{X} \mid \mathbf{C}) = \frac{m}{m + n - 1} \tag{8}$$

Now the probability of his *innocence* is (1 – the probability of his *guilt*) so:

$$\text{probability of innocence} = 1 - \frac{m}{m + n - 1} = \frac{n - 1}{m + n - 1} \tag{9}$$

In practice, n (the number of potential suspects) is quite large. We can therefore ignore the minus ones on both the top and bottom of the fraction to get a simpler formula which will do just as well:

$$\text{probability of innocence} = \frac{n}{m + n} \tag{10}$$

We can now see clearly exactly how the figure m (= 1,000,000) supplied by the professor is related to the probability of the defendant's guilt or innocence. Because m appears on the bottom of the equation the larger the value of m the smaller the probability of the defendant's innocence. But the crucial point to make is that the calculation of the probability involves knowing n – the number of potential suspects. If n is large, the probability of the defendant being *innocent* approaches 1.

In all practical cases, n is much smaller than m so we can simplify the formula even further to:

$$\text{probability of innocence} = \frac{n}{m} = n \times (1/m) \tag{11}$$

DNA evidence

So does this mean that DNA evidence is worthless? Far from it. DNA evidence *on its own* is worthless but combined with other evidence which restricts the number of possible suspects it is invaluable. Suppose that the steward on the train recognises Mr Scroggins and testifies that he was on the train that night, a fact confirmed by the discovery of the relevant train ticket on his person. In addition there were about 200 other men on the train.

Now it is not necessary for the police to check the DNA of all the other men on the train. The fact remains that there is a match between Mr Scroggin's DNA and the DNA found on the victim. How does this alter the probabilities? The situation is now the same as pulling one 1,000,000-sided die out of a bag of 200 and throwing a spot. The chances of Mr Scroggins being *innocent* under these circumstances are:

$$\text{probability of innocence} = \frac{200}{1000000 + 200} \approx 0.0002 \quad \text{or } 0.02\% \tag{12}$$

In other words, Mr Scroggins is almost certainly guilty.

DNA evidence and the law

In 'A Guide to DNA' published by the Forensic Science Service[3] lawyers are given the following advice:

> Presenting probabilistic evidence in court has many pitfalls but there is one, in particular, that has led to successful appeals (e.g. in Doheny and Adams). Consider, for example, where there is a full SGM Plus match between the defendant and semen recovered from a rape victim... Then the scientist might say:
>
> "The probability of a match if the semen came from another person is one in a billion."
>
> In an effort to simplify things for the benefit of the jury, counsel might be tempted to invite the scientist to agree to the following paraphrase:
>
> "The probability that the semen came from another person is one in a billion."
>
> The second sentence is undoubtedly simpler but it does not follow from the first sentence. This is what is known as the "prosecutor's fallacy", though statisticians know it as the "fallacy of the transposed conditional". This is, undoubtedly, what the learned judges had in mind in the judgement in Doheny and Adams where they say:
>
> "The scientist should not be asked his opinion on the likelihood that it was the defendant who left the crime stain, nor when giving evidence should he use terminology which may lead the jury to believe that he is expressing such an opinion."
>
> As a general rule, the role of the scientist is to advise the court of the probability of the evidence given the proposition, in this example, of the probability of a match if the semen came from some unknown person. It is properly the role of the jury to consider the probability of the proposition given the evidence, in this example, the probability that the semen did come from some unknown person.

I do not think this is good advice. Telling the jury that the probability of a match given the defendant's innocence is of no help at all in determining the probability of his guilt and only serves to confuse the issue. As we have seen, to deduce the probability of the proposition given the evidence requires a thorough understanding of Bayes' Theorem. This cannot be expected of a jury. It is essential therefore that the jury should have access to advice from someone who does have such an understanding. In the majority of cases the scientist who is expert enough to explain to the jury the intricacies of DNA matching should be competent enough to answer a question of the form:

"Given that the defendant's DNA matches that in the semen found at the scene and that other evidence restricts the number of possible suspects to n (where n is the population of Manchester or the number of people on the train or some other relevant number), what is, in your opinion, the probability that the defendant is guilty?"

I imagine the exchange going something like this:

Prosecutor: "I understand that the probability of finding a match by, accident as it were, between two unrelated people is of the order of 1 in a billion. Is that correct?"

Prof Noitall: "*It is*"

"So if you were to compare the DNA found at the scene of the crime with the DNA of every man on this planet, how many matches would you expect to find?"

"*Well there are about 5 billion men on Earth so you would expect to get around 5 accidental matches.*"

"All of them innocent?"

---

3 https://www.cps.gov.uk/legal/assets/uploads/files/lawyers'%20DNA%20guide%20KSWilliams %20190208%20(i).pdf

“*No – one would be guilty*” (a titter from the court)

“But you wouldn't know which?”

“*No*”

“But you surely wouldn't suspect Mr Tong who lives in China or Mr Abogo who lives in Africa?”

“*No – not unless they were on the train that night.*” (Laughter from the court)

“Ah yes. Indeed. Now there were about 200 men on the train that night and since there was no way anyone could have got on or off the train, we have established that it must have been one of these men who committed the crime. Moreover we know that the defendant was on the train too. Does this make a difference to the calculation of the probability of the defendants guilt or innocence?”

“*Yes, it makes a big difference.*”

“So in order to calculate the probability that the DNA from the sample found on the victim originated from the defendant, you need two figures. a) the probability of an accidental match and b) the number of possible suspects. Is this correct?”

“*it is.*”

“Could you show us haw to perform this calculation?”

“*Certainly. The probability that the DNA found at the scene originated from the defendant is simply the probability of an accidental match multiplied by the number of potential suspects* (see equation 11). *In the case before us the probability is 1 in a billion times 200 which is 1 in 5 million.*”

“Let us be clear what you are saying. The probability that the match which exists between the DNA found at the scene and the DNA of the defendant was accidental is 1 in 5 million. Is that correct?”

“*It is.*”

“How confident therefore are you that, on the basis of the evidence presented to you, the DNA found at the scene originated form the defendant?”

“*I am absolutely confident that that is the case, beyond reasonable doubt.*”

“Thank you. I have no further questions, M'lud.”

© J Oliver Linton, Carr Bank, December 2017