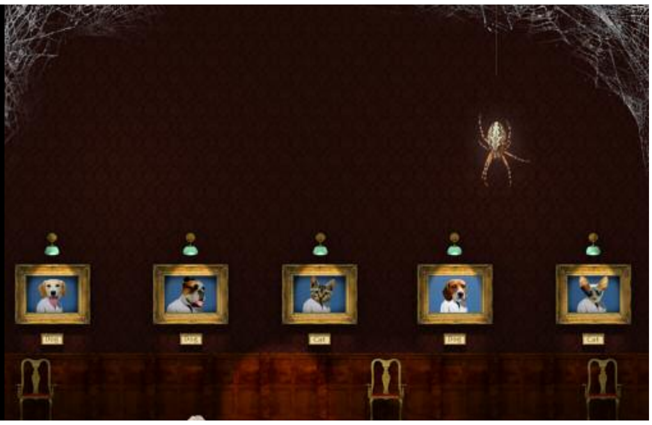
Christmas Lectures 2008

Learning from probabilities

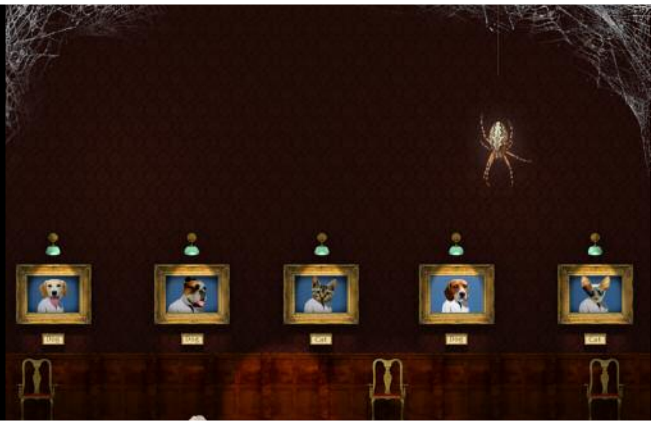


Computer scientists are now using 'Bayes' theorem' to build computers that are more intelligent and even able to learn for themselves. Here we'll show how you can use just sweets and pots and a pinch of Bayes' theorem to find out who stole a chocolate biscuit! You'll discover how likely it is that someone with chocolate smudged on their fingers is the real culprit. You can run this is an activity with your friends at home or ask your teacher to run it in class.

BAYES' THEOREM

Computer scientists work extremely hard to improve how computers work. They try to make computers more intelligent and able to solve problems better. Some scientists are developing computers that can even start learning for themselves – i.e. make intelligent decisions, like humans do. Some computer scientists are programming computers to make intelligent decisions by using 'Bayes' theorem'.

Thomas Bayes was an English reverend who lived in the eighteenth century and developed a new kind of mathematics: he worked on probability. Traditional probability is fairly easy to understand. For example, if you have three red jumpers and seven blue jumpers in a drawer what's the probability that you pull out a red one? Easy – it's 3/10 (30%). But Bayes was interested in how you can work out probabilities *backwards*. For example, in your top drawer you have two red jumpers and two blue ones and in your bottom drawer you have one red jumper and five blue ones. If you open a drawer randomly (i.e. a 50-50 chance), pull out a jumper and then see that it's blue, which drawer are you most likely to have opened? The answer's the bottom drawer because it has more blue jumpers, but what's the exact probability?



This is what Bayes' theorem is all about – working out the probabilities that different things have happened based on what outcome you see or the evidence you get. Computers often have to make decisions without having all the information, so Bayes' theorem is really useful for computers to work out what's happening around them and to learn from it.

Here's an activity that shows how Bayes' theorem works in real life.

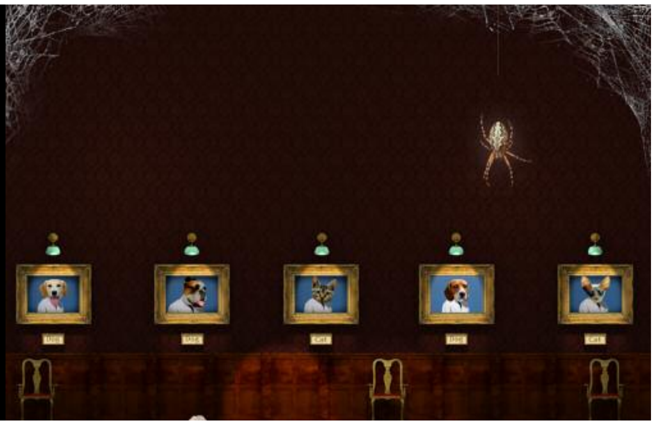
FINDING THE CULPRIT

Let's imagine a situation where Bayesian probability can be used to work out how likely something is, given a piece of evidence. To find the culprit in this case, you'll need:

- 10 small bowls or Post-it notes with a circle drawn in the middle
- a bag of uncoated raisins and a bag of chocolate-coated raisins or any two different kinds of sweets, but using a chocolate-coated type works best.

During a maths lesson Ms Muir pops out of the classroom for a moment. When she returns she discovers that someone has pinched a chocolate digestive from her special biscuit tin! She'd only been gone a few minutes and class 8B were the only people in the room the whole time, so she knows that the culprit who let the munchies get the better of them must be one of the ten pupils in the class.

Ms Muir also realises that there's a good chance that whoever pinched the biscuit has chocolate on their fingers. Shockingly, Ms Muir notices that Jon has chocolate all over his hands. So he must be the culprit, surely? (But she doesn't check everybody else's hands for evidence.) However, she also knows that lots of people like biscuits, so there's a chance that Jon may have got chocolatey fingers from his packed lunch, and not gone near her biscuit tin.



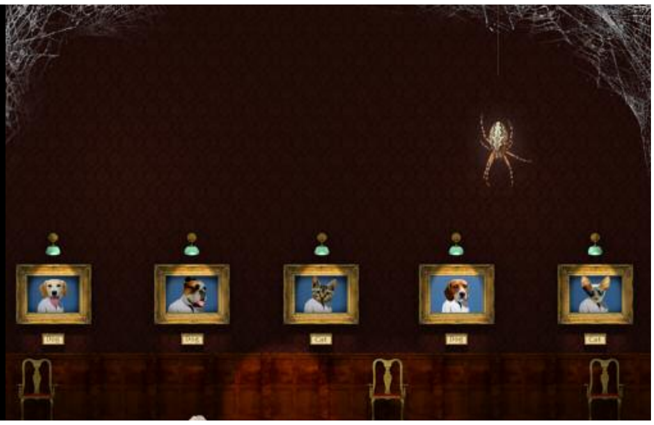
So the question is: what is the likelihood that Jon is guilty of pinching her chocolate biscuit, given that he has chocolate on his fingers? Luckily, Ms Muir knows all about Bayes' theorem and so can work out the probability of Jon being guilty or innocent from the facts of the case, which are:

- One person in a class of 10 people has stolen a biscuit, so the probability that any one of the pupils is guilty is 1/10.
- If you steal a biscuit, there is a very good chance that you get chocolatey fingers – the probability is 90%.
- However, because some pupils get chocolate in their lunch boxes (even if they are innocent of stealing any biscuits) there is a 1/5 chance they just so happen to have chocolatey fingers.

We can recreate the situation using our ten small bowls (or Post-it notes) and sweets (or raisins). The photograph on the next page shows how to set this up.

1. Place one bowl (or Post-it note) on one side and one on the other side of the desk. Label one bowl 'guilty' and the other 'innocent'.
2. There is a 90% chance that the biscuit thief got chocolate on their hands, so put 9 chocolate-coated raisins and 1 uncoated raisin in the guilty bowl. Now, if you were to close your eyes and pick a raisin out of the guilty bowl there would be a 9/10 chance it would be chocolate covered.
3. Even if someone didn't steal the biscuit, there is still a 1/5 chance that they have chocolate on their fingers. To represent this probability, put 2 chocolate-coated raisins and 8 uncoated raisins in the innocent bowl. So the chance of picking an innocent person and finding that they have chocolatey fingers is 2/10 = 1/5.
4. Since there is only 1 guilty person and 9 innocent people, we need to repeat the innocent bowl 9 times. Use the other bowls and make sure you put in the right number of chocolate-coated (2) and uncoated raisins (8) in each one.

Learning from probabilities



Now we can use this model to answer the question: what is the probability that someone is guilty of pinching the chocolate biscuit, given they have chocolate on their fingers? In our model this is the same as saying, 'if I close my eyes and pick a random raisin out of any of the bowls and then see that it's chocolate-coated, what's the probability it came from the guilty bowl?'

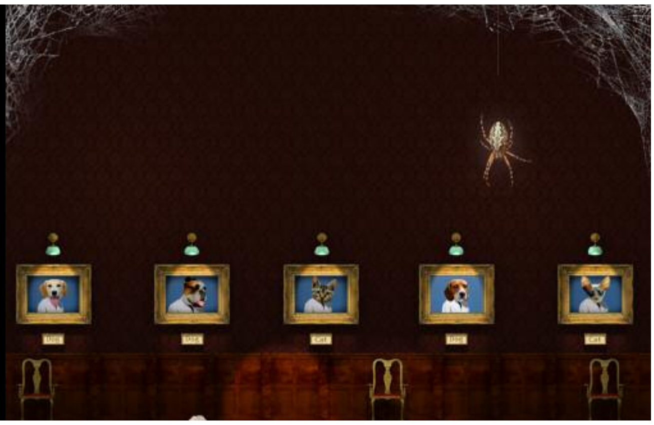
Well, there are 9 chocolate-coated raisins in the guilty bowl and 18 (9 x 2) chocolate raisins in all the innocent bowls put together. The probability that the chocolate raisin came from the guilty bowl is therefore 9/9+18 = 9/27 = 1/3.

So, even if Ms Muir found Jon with chocolate on his fingers, the probability that Jon stole the biscuit isn't certain. It is only one-third. In fact, it's much *more* likely that someone else in the class is the real culprit!

Why don't you use a piece of evidence to make up your own story about a crime and try to solve it using the Bayes' theorem? Decide what the different probabilities are, set up the bowls and sweets as in the activity above and find the overall probability that the poor accused is really innocent!

Christmas Lectures 2008

Learning from probabilities



FINDING THE CULPRIT – ADVANCED

If you enjoyed the activity on Bayesian probability, you might like to see how the answer can be worked out using some easy maths.

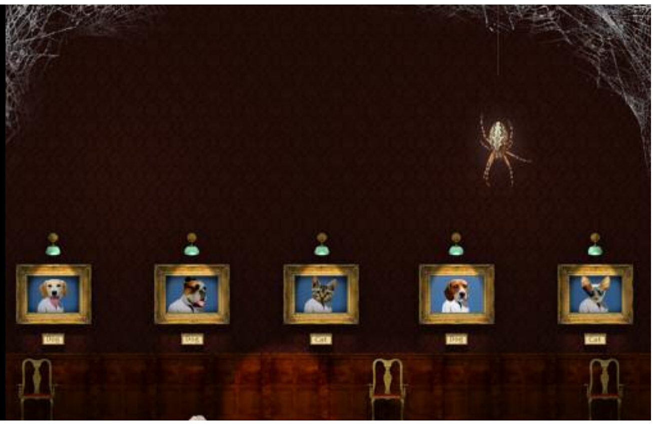
Let's just remind ourselves about the facts of this case.

1. One pupil in a class of 10 has stolen a biscuit, so the probability that any one of them is guilty is 1/10.
2. If you steal a biscuit there is a very good chance that you get chocolatey fingers. In fact, the probability is 90%.
3. However, because some pupils have chocolate in their lunch boxes, even if you are innocent of stealing any biscuits there is still a 1/5 chance you happen to have chocolatey fingers.

Let's call the act of stealing a biscuit, *S*. The complementary act, of *not* stealing a biscuit, is written *S'*. We can write the evidence of having chocolatey fingers as *C*, and so having clean fingers is *C'*. The probability of having stolen the biscuit is written as *P(S)*.

The actual question that we want to answer is 'what is the probability that someone stole the biscuit given that they have chocolate on their fingers?' This question can be written in maths as: *P(S|C)* where the | symbol means 'given that'. The probability that you have chocolatey fingers without having stolen the biscuit is: *P(C|S')*. We can now translate the three facts of the case into mathematical statements or equations.

1	P(S) = 1/10	The probability of each pupil being the guilty one is one out of ten (the number of pupils in the class where the biscuit went missing).
	P(S') = 9/10	So it must be true that nine out of the ten pupils are innocent.
2	P(C|S) = 9/10	The probability that you have chocolatey fingers after stealing the biscuit is 90%.
3	P(C|S') = 1/5	The probability that you have chocolatey fingers even if you haven't stolen the biscuit is 1/5.



If you are interested, P(S) is often called the 'prior probability' because it is what you can work out *before* the new evidence (i.e. the chocolatey fingers) came to light. P(S|C) is called the 'posterior probability' because it includes the new evidence – this is the probability we need to work out using Bayes' theorem.

Thomas Bayes realised that you could work out the probability of the event you don't know about by using the probabilities you do know. His theorem about probabilities can be written as the following equation:

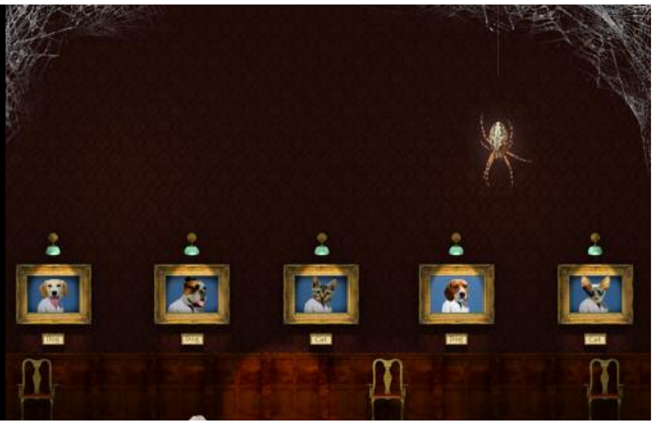
$$P(S|C) = \frac{P(C|S) \text{ x } P(S)}{P(C)}$$

P(C) is the probability of having chocolatey fingers, regardless of whether you stole the biscuit or not. So this is equal to the combination of the probability that you have chocolatey fingers after stealing *plus* the probability that you have chocolatey fingers even though you didn't steal the biscuit:

$$P(C) = P(S).P(C|S) + P(S').P(C|S')$$

So now our full Bayesian equation is:

$$P(S|C) = \frac{P(C|S) \text{ x } P(S)}{P(S).P(C|S) + P(S').P(C|S')}$$



We can put in all the probabilities that we do know (from the facts of the case) and then work out the equations.

$$P(S|C) = \frac{9/10 \times 1/10}{1/10 \times 9/10 + 9/10 \times 1/5}$$

$$P(S|C) = \frac{9/100}{9/100 + 9/50}$$

$$P(S|C) = \frac{9/100}{27/100}$$

$$P(S|C) = \frac{9}{100} \times \frac{100}{27}$$

$$P(S|C) = 1/3$$

So this bit of maths has shown us that the probability that Jon did steal the biscuit, given that he has chocolatey fingers, is one-third. This answer is the same as the one we got with the bowls and sweets!