Probably guilty: Bad mathematics means rough justice

- 28 October 2009 by **Angela Saini**
- Magazine issue 2731

SHAMBLING sleuth Columbo always gets his man. Take the society photographer in a 1974 episode of the cult US television series who has killed his wife and disguised it as a bungled kidnapping. It is the perfect crime - until the hangdog detective hits on a cunning ruse to expose it. He induces the murderer to grab from a shelf of 12 cameras the exact one used to snap the victim before she was killed. "You just incriminated yourself, sir," says a watching police officer.

If only it were that simple. Killer or not, anyone would have a 1 in 12 chance of picking the same camera at random. That kind of evidence would never stand up in court.

Or would it? In fact, such probabilistic pitfalls are not limited to crime fiction. "Statistical errors happen astonishingly often," says Ray Hill, a mathematician at the University of Salford, UK, who has given evidence in several high-profile criminal cases. "I'm always finding examples that go unnoticed in evidence statements."

The root cause is a sloppiness in analysing odds that can sully justice and even land innocent people in jail. With ever more trials resting on the "certainties" of data such as DNA matches, the problem is becoming more acute. Some mathematicians are calling for the courts to take a crash course in the true significance of the evidence put before them. Their demand: Bayesian justice for all.

That rallying call derives from the work of Thomas Bayes, an 18th-century British mathematician who showed how to calculate conditional probability - the chance of something being true if its truth depends on other things being true, too. That is precisely the kind of problem that criminal trials deal with as they sift through evidence to establish a defendant's innocence or guilt (see "Bayes on trial").

Mathematics might seem a logical fit for the courts, then. Judges and juries, though, all too often rely on gut feeling. A startling example was the rape trial in 1996 of a British man, Dennis John Adams. Adams hadn't been identified in a line-up and his girlfriend had provided an alibi. But his DNA was a 1 in 200 million match to semen from the crime scene - evidence seemingly so damning that any jury would be likely to convict him.

But what did that figure actually mean? Not, as courts and the press often assume, that there was only a 1 in 200 million chance that the semen belonged to someone other than Adams, making his innocence implausible.

It actually means there is a 1 in 200 million chance that the DNA of any random member of the public will match that found at the crime scene (see "The prosecutor's fallacy"). The difference is subtle, but significant. In a population, say, of 10,000 men who could have committed the crime, there would be a 10,000 in 200 million, or 1 in 20,000, chance that someone else is a match too. That still doesn't look good for Adams, but it's not nearly as damning.

So worried was Adams's defence team that the jury might misinterpret the odds that they called in Peter Donnelly, a statistical scientist at the University of Oxford. "We designed a questionnaire to help them combine all the evidence using Bayesian reasoning," says Donnelly (*Significance*, vol 2, p 46).

They failed, though, to convince the jury of the value of the Bayesian approach, and Adams was convicted. He appealed twice unsuccessfully, with an appeal judge eventually ruling that the jury's job was "to evaluate evidence not by means of a formula... but by the joint application of their individual common sense."

But what if common sense runs counter to justice? For David Lucy, a mathematician at Lancaster University in the UK, the Adams judgment indicates a cultural tradition that needs changing. "In some cases, statistical analysis is the only way to evaluate evidence, because intuition can lead to outcomes based upon fallacies," he says.

Norman Fenton, a computer scientist at Queen Mary, University of London, who has worked for defence teams in criminal trials, has just come up with a possible solution. With his colleague Martin Neil, he has developed a system of step-by-step pictures and decision trees to help jurors grasp Bayesian reasoning (bit.ly/1c3tgj). Once a jury has been convinced that the method works, the duo argue, experts should be allowed to apply Bayes's theorem to the facts of the case as a kind of "black box" that calculates how the probability of innocence or guilt changes as each piece of evidence is presented. "You wouldn't question the steps of an electronic calculator, so why here?" Fenton asks.

It is a controversial suggestion. Taken to its logical conclusion, it might see the outcome of a trial balance on a single calculation. Working out Bayesian probabilities with DNA and blood matches is all very well, but quantifying incriminating factors such as appearance and behaviour is more difficult. "Different jurors will interpret different bits of evidence differently. It's not the job of a mathematician to do it for them," says Donnelly.

He thinks forensics experts should be schooled in statistics so they can catch errors before they occur. Since cases such as Adams's, that has already begun to happen in the US and UK. Lawyers and jurors, however, still have far less - if any - statistical training.

As the real-life fallacies that follow show, there's no room for complacency. It is not about mathematicians trying to force their way of thinking on the world, says Donnelly: "Justice depends on getting everyone to reason properly with uncertainties."

Five fallacies to forgo

It pays to be careful when using statistics as evidence, as these examples from the legal casebook show

1. Prosecutor's fallacy

"The prosecutor's fallacy is such an easy mistake to make," says Ian Evett from the Forensic Science Service in England and Wales. It confuses two subtly different probabilities that

Bayes's formula distinguishes: P(H | E), the probability that someone is innocent if they are a match to a piece of evidence, and P(E | H), the probability that someone is a match to a piece of evidence if they are innocent (see "Bayes on trial"). The first probability is what we would like to know; the second is what forensics usually tells us.

Unfortunately, even professionals sometimes mix them up. In the 1991 rape trial of Andrew Deen in Manchester, UK, for example, an expert witness agreed on the basis of a DNA sample that "the likelihood of [the source of the semen] being any other man but Andrew Deen [is] 1 in 3 million."

That was wrong. One in 3 million was the likelihood that any innocent person in the general population had a DNA profile matching that extracted from semen at the crime scene - in other words, P(E | H). With around 60 million people in the UK, a fair few people will share that profile. Depending on how many of them might plausibly have committed the crime, the probability of Deen being innocent even though he was a match, or P(H | E), was actually far greater that 1 in 3 million.

Deen's conviction was quashed on appeal, leading to a flurry of similar appeals that have had varying success. The latest is the appeal of a Californian man jailed last year, who was discovered by police to be a DNA match to a rape and murder committed 35 years earlier.

2. Ultimate issue error

The prosecution in the Deen case stopped just short of compounding their probabilistic fallacy. In the minds of the jury, though, it probably morphed into the "ultimate issue" error: explicitly equating the (small) number P(E | H) with a suspect's likelihood of innocence.

In Los Angeles in 1968, the ultimate issue error sent Malcolm Collins and his wife Janet to jail. At first glance, the circumstances of the case left little room for doubt: an elderly lady had been robbed by a white woman with blonde hair and a black man with a moustache, who had both fled in a yellow car. The chances of finding a similar interracial couple matching that description were 1 in 12 million, an expert calculated.

The police were convinced, and without much deliberation so was the jury. They assumed that there was a 1 in 12 million chance that the couple were not the match, and that this was also the likelihood of their innocence.

They were wrong on both counts. In a city such as Los Angeles, with millions of people of all races living in it or passing through, there could well be at least one other such couple, giving the Collinses an evens or better chance of being innocent. Not to mention that the description itself may have been inaccurate - facts that helped reverse the guilty verdict on appeal.

3. Base-rate neglect

Anyone looking to DNA profiling for a quick route to a conviction should recognise that genetic evidence can be shaky. Even if the odds of finding another genetic match are 1 in a billion, in a world of 7 billion, that's another seven people with the same profile.

Fortunately, circumstantial and forensic evidence often quickly whittle down the pool of suspects. But neglecting your "base rate" - the pool of possible matches - can have you leap to false conclusions, not just in the courtroom.

Picture yourself, for example, in the doctor's surgery. You have just tested positive for a terminal disease that afflicts 1 in 10,000. The test has an accuracy of 99 per cent. What's the probability that you actually have the disease?

It is in fact less than 1 per cent. The reason is the sheer rarity of the disease, which means that even with a 99 per cent accurate test, false positives will far outweigh real ones (see diagram). That's why it is so important to carry out further tests to narrow down the odds. We lay people are not the only ones stumped by such counter-intuitive results: surveys show that 85 to 90 per cent of health professionals get it wrong too (*Behavioral and Brain Sciences*, vol 30, p 241).

4. Defendant's fallacy

It's not just prosecutors who can fiddle courtroom statistics to their advantage: defence lawyers have also been known to cherry-pick probabilities.

In 1995, for example, former American football star O. J. Simpson stood trial for the murder of his ex-wife, Nicole Brown, and her friend. Years before, Simpson had pleaded no contest to a charge of domestic violence against Brown. In an attempt to downplay that, a consultant to Simpson's defence team, Alan Dershowitz, stated that fewer than 1 in 1000 women who are abused by their husbands or boyfriends end up murdered by them.

That might well be true, but it was not the most relevant fact, as John Allen Paulos, a mathematician at Temple University in Philadelphia, Pennsylvania, later showed. As a Bayesian calculation taking in all the pertinent facts reveals, it is trumped by the 80 per cent likelihood that, if a woman is abused and later murdered, the culprit was her partner.

That may not be the whole story either, says criminologist William Thompson of the University of California, Irvine. If more than 80 per cent of all murdered women, abused or not, are killed by their partner, "the presence of abuse may have no diagnostic value at all".

5. Dependent evidence fallacy

Sometimes, mathematical logic flies out of the courtroom window long before Bayes can even be applied - because the probabilities used are wrong.

Take the dependent evidence fallacy, which was central to one of the most notorious recent miscarriages of justice in the UK. In November 1999, Sally Clark was convicted of smothering her two children as they slept. A paediatrician, Roy Meadow, testified that the odds of both dying naturally by sudden infant death syndrome (SIDS), or cot death, were 1 in 73 million. He arrived at this figure by multiplying the individual probability of SIDS in a family such as Clark's - 1 in 8500 - by itself, as if the two deaths were independent events.

But why should they be? "There may well be unknown genetic or environmental factors that predispose families to SIDS, so that a second case within the family becomes much more likely," the Royal Statistical Society explained during an appeal.

"Even three eminent judges didn't pick up on the mistake," says Ray Hill of the University of Salford, who worked for the defence team. He estimated that if one sibling dies of SIDS, the chance of another dying is as high as 1 in 60. Bayesian reasoning then produces a probability of a double cot death of around 1 in 130,000. With hundreds of thousands of children born each year in the UK, there's bound to be a double cot death every now and then.

Clark was eventually freed on appeal in 2003. Her case had a lasting effect, leading to the review of many similar cases. "I'm not aware of any cases of multiple cot deaths reaching the courts in recent years," says Hill. Clark herself never recovered from her ordeal, however. She was found dead at her home in 2007, ultimately a victim of statistical ignorance.

Angela Saini

Bayes on trial

Suppose you have a piece of evidence, E, from a crime scene - a bloodstain, or perhaps a clothing thread - that matches to a suspect. How should it affect your perception or hypothesis, H, of the suspect's innocence?

Bayes's theorem tells you how to work out the probability of H given E. It is: (the probability of H) multiplied by (the probability of E given H) divided by (the probability of E). Or in standard mathematical notation:

$$P(H \mid E) = P(H) \times P(E \mid H)/P(E)$$

Say you are a juror at an assault trial, and so far you are 60 per cent convinced the defendant is innocent: $P(H) = 0.6$. Then you're told that the blood of the defendant and blood found at the crime scene are both type B, which is found in about 10 per cent of people. How should this change your view?

What the forensics expert has given you is the probability that the evidence matches anyone in the general population, given that they are innocent: $P(E \mid H) = 0.1$. To apply Bayes's formula and find $P(H \mid E)$ - your new estimation of the defendant's innocence - you now need the quantity $P(E)$, the probability that his blood matches that at the crime scene.

This probability actually depends on the defendant's innocence or guilt. If he is innocent, it is 0.1 as it is for anyone else. If he is guilty, however, it is 1, as his blood is certain to match. This insight allows us to calculate $P(E)$ by summing the probabilities of a blood match in the case of innocence (H) or guilt (not H):

$$P(E) = [P(E \mid H) \times P(H)] + [P(E \mid \text{not } H) \times P(\text{not } H)]= (0.1 \times 0.6) + (1 \times 0.4) = 0.46$$

So according to Bayes's formula the revised probability of his innocence is:

$$P(H \mid E) = (0.6 \times 0.1) \div 0.46 = 0.13$$

As you might expect, by this measure the defendant is between four and five times guiltier than you first thought - probably.

Angela Saini is a writer based in London