Statistics and the law

A minefield for juries

By John Croucher

The subject of statistics strikes terror into the heart of many professionals, whether they be medical practitioners, members of the legal fraternity or any other professional who should know something about how it all works (but probably do not).

Pity, then, the beleaguered juror, who must navigate his or her way through conflicting testimony from those who are supposed to know. And a little knowledge is dangerous, as highlighted by the tragic Sally Clark case in the UK.

By a majority verdict of 10:2,[1] Clark had been convicted in 1999 of murdering her two sons, both of whom had died of apparent sudden infant death syndrome (SIDS). She was sentenced to life imprisonment. As a lawyer, a convicted child-killer and the daughter of a police officer, she did it tough in prison. She served more than three years of her sentence before her conviction was quashed in 2003.[2] She never was able to put it behind her and she was found dead in her home on 16 March 2007. It was described as 'one of the great miscarriages of justice in modern British legal history';[3] and the misuse of statistical evidence played a significant role.

Statistical errors

An eminent paediatrician gave evidence to the jury that the probability of two babies dying of SIDS in Clark's circumstances (affluent, middle-class, non-smoking) was one in 73 million. Putting this into words, rather than numbers, he said that 'one sudden infant death in a family is a tragedy, two is suspicious and three is murder unless proven otherwise'. This figure was outrageously incorrect. Indeed, statisticians commenting on the Clark case had been very disturbed from the outset that such a serious statistical error had been made and one that had no doubt influenced the jury, especially given the comments of the trial judge in his summing up that such evidence was 'compelling'. It even prompted a letter from the President of the Royal Statistical Society, Professor Peter Green, to the Lord Chancellor that outlined the statistical flaws made at the trial and implored him 'to ensure that statistical evidence is presented only by qualified statistical experts, as would be the case for any other form of expert evidence'. Indeed the correct probability of a family that has already had a cot death having a second cot death is more like one in 100—a far cry from one in 73 million.

There has been much written on the statistical errors made in the Clark trial and they serve as an excellent example of just how things can go horribly wrong if they are accepted as fact. They can also have a compelling influence on jurors in their weighing up of evidence. The theory of 'probability' is widely misunderstood by most members of the general public, including jurors, along with those (unfortunate) students forced to study it by compulsion, even on a small scale, as part of their chosen major. Indeed, many sadly try to put it to one side as soon as their degree is over. But invariably statistics manages to intrude into some professional lives such that it can't just be ignored. And so it is with legal practitioners, many of whom find statistics a real challenge. This has led to erroneous conclusions based on the evidence that have become known as a variety of 'fallacies', including those of both the prosecutor and defence. Whether or not the underlying mathematics is fully understood, it is essential that jurors can correctly interpret

Professor John S Croucher is a Professor of Statistics at the Macquarie Graduate School of Management, Macquarie University in Sydney.

the information provided by witnesses, expert or otherwise. Without an ability to interpret evidence provided in the form of statistical information, seemingly damning figures ('one in 73 million') are left hanging in the air to confound counsel, judge and jury alike. Statistical evidence has to be introduced and evaluated properly if it is to have an effective role in evidence.

Ordinary and conditional probability

Without delving into the deeper intricacies of probability theory, there are a few examples that can always be relied upon to give pause for thought—and to reveal the perils of statistics and surprising results to the uninitiated. There is a vital difference between 'ordinary' (or 'unconditional') probability and 'conditional' probability. To illustrate the former, it is instructive to consider the well known 'birthday problem'. This goes along the lines of: 'how many people do you need in a random sample before the probability of having matching birthdays (same day and month) is at least 50%?' A typical response might be about 183 since there are 365 days in a year. In fact the right answer is only 23, a figure that seems incredible but is nevertheless correct. A variation on this problem is: 'how many people do you need in a random sample before the probability of two people having a match on the last two digits on their home telephone number is at least 50%?' Although there are 100 possible two-digit numbers, only 13 people are required in the sample to achieve this. It is 'unconditional' because there is no other information given. These two gems alone are often enough to convince people that probability theory may well be beyond them.

△ Without an ability to interpret evidence provided in the form of statistical information, seemingly damning figures ... are left hanging in the air to confound counsel, judge and jury alike. △

'Conditional probability' is a probability calculated with the knowledge that some other event has occurred. The information you know alters the probability. For example, in the absence of other facts, the probability of throwing a six on a fair six-sided die is one in six. However, if you were now told that the outcome was an even number, this probability reduces to one in three. That is, this knowledge has changed the answer markedly since the number of possible outcomes has been reduced from six to only three. This is usually expressed in words as 'the conditional probability of throwing a six given that the outcome is even'.

A similar situation often arises in the legal context. If a person is selected at random then, in the absence of other information, the probability that they are male is about one in two (or ½ or 50%). However, if you are now told that the person has a beard. the conditional probability they are male changes to (essentially) one, or 100%. In a trial situation this is to all intents and purposes what a judge and/or jury is trying to do. That is, to find the conditional probability that an accused is guilty or innocent given the evidence. That is why statistical evidence has been introduced in criminal trials—as a pointer to the probability in an evidentiary sense of the particular fact in issue.

As a particular instance, it is often conjectured that juries confuse the direction of conditional probabilities with dire consequences. For example, with DNA testing becoming more widespread at crime scenes, which of the following two conditional probabilities should the court be considering:

A. the probability that the DNA found at the crime scene matches that of the accused if the accused is innocent; or

B. the probability that the accused is innocent if the DNA found at the crime scene matches theirs?

An inexperienced person may well say that these probabilities are really the same thing, but this is far from the case and they can in fact differ to a very large degree. We have already noted that the conditional probability of throwing a six on a fair die given the outcome is even is one in three. However, in the reverse, the conditional probability that the outcome is even given that a six has been thrown is one (a certain event). So which of *A* or *B* in the DNA example is the one that should be of interest? The correct answer is at the end of this article!

Independent events

The issue of *independence* is also one that is often misused and misunderstood in the legal context. Statistically, two 'events' (facts) are said to be 'independent' if the occurrence of one of them is totally unaffected by the occurrence of the other. Although it may be arguable on some occasions whether two events might be really independent, in most cases it seems clear-cut. For example, the outcomes of two tosses of a coin are readily seen to be independent events since a coin has no memory and cannot remember what it landed on the first toss. On the other hand, the event of a person being pregnant is certainly not independent of the fact that they are female.

Independent events are more straightforward to deal with since various probabilities can be calculated reasonably easily. In particular, the probability of two (or more) independent events occurring is simply the product of the probabilities of each individual event occurring. In the case of the coin, for example, the probability of obtaining a head on a single toss is ½. The probability of obtaining two heads in two tosses is ½ x ½ = ¼, since the events are independent. This can extend to any number of tosses so that the probability of tossing, say, five heads in a row = ½ x ½ x ½ x ½ x ½ = 1/32 or about 3% of the time.

Dangerous mistakes

The danger comes when the events are *not* independent but their probabilities are multiplied by the naïve user anyway. This was one of the major criticisms made of the one in 73 million figure given in the Sally Clark case. The expert provided an estimate that the probability of a randomly chosen baby in the socio-economic circumstances of that of Clark's dying of SIDS was about one in 8500. He, therefore, concluded that the probability of *two* such deaths could be obtained by squaring this value. This yields 1/8500 x 1/8500 or about one in 73 million. This appears to be powerful evidence against the accused. But is the event of a second child dying of SIDS really independent of the event of the first child also dying of SIDS? If they are *not* independent then it is nonsense to multiply the probabilities since the answer can be spectacularly incorrect, as is the case here. There are many elements in calculating what would be the probability of *dependent* events occurring (like two children in the same family dying of the same cause), but it cannot be calculated simply by using the multiply rule that is used for *independent* events.

Let's return to the example about the ½ probability that a random person in the population is male—and add to it. Suppose we estimate that the percentage of people walking down the main street of a city during business hours at any given time have the characteristics listed are as shown in brackets: male (50%); suit coat (10%); suit trousers (10%); black shoes (20%); case (10%); tie (15%); glasses (25%); moustache (10%); beard (15%); dark hair (30%).

There is nothing particularly startling about these figures. But suppose that an eyewitness to a crime stated that the perpetrator had all of these characteristics. In a population of about, say, 60 million people, how many people would we expect would match that description?
If we assume that the characteristics are independent, the probability that an individual has *all* of them can be found by multiplying the individual probabilities. This yields 0.50 x 0.10 x 0.10 x 0.20 x 0.10 x 0.15 x 0.25 x 0.10 x 0.15 x 0.30 = 0.000000017. Therefore the 'expected' number of people who have all of these characteristics in a population of 60 million = 0.000000017 x 60,000,000 = 1. That is, just *one* person.

> △ The danger comes when the events are not independent but their probabilities are multiplied by the naïve user anyway. △

You might therefore conclude that if you could find a person with all of those characteristics somewhere in Australia or even the UK then you would have got your offender! A careful look at the characteristics, however, shows that we are describing a male who is wearing a suit, has dark hair, beard and moustache, wears glasses and is carrying a case. It is obvious that there are probably many thousands of people who match that description, not just one. The problem is of course that the characteristics are far from independent (it is easy to see why) and it is quite ridiculous to multiply them together. Although it may seem obvious, similar erroneous calculations to these have appeared in court proceedings around the world to show that the chance of finding another person with characteristics similar to an accused is extremely low. This naturally has the effect of making the accused look more guilty.

More sophisticated calculations

There are other types of cases involving probability that require more sophisticated calculations. For example, suppose that your child is about to be vaccinated and you ask the medical practitioner about the risk that it will kill him or her. You are told that the risk is one in 200,000 and, it being so low, you agree to go ahead but the child subsequently dies from the vaccination. You are naturally devastated and subsequently discover that of the 800,000 children who received the vaccination there were in fact six who died as a direct result. Your calculations tell you that if the risk of dying had been really 1 in 200,000 then there should have been only four deaths in 800,000, not six. This clearly means that the medical practitioner lied about the true risk. Or did they?

To help answer this question, suppose that we take a fair coin and toss it six times. We might anticipate that half of the outcomes would be a head and so we would 'expect' three heads. But suppose we actually obtained four heads in those six tosses. Would that necessarily mean that the coin was not a fair one? Almost certainly not since there is some statistical variation that must be allowed for and we will not always get exactly what we expect, even if the original premise of the coin being fair is true.

This is also the case for the vaccination question where there is some margin of error within which it may be quite likely that the medical practitioner was still correct. Only precise calculations involving the probability of obtaining the given number of deaths or greater, based on the assumption that the information provided was accurate, can answer a problem such as this.

There are many other legal matters in which statistics can play an important role in arriving at the correct conclusion and it is very important for jurors to be aware of some of the more common pitfalls and range of situations to which it applies. In most cases this will still mean enlisting a statistical expert who has done the actual number crunching and analysis but at least they should have some confidence that it has been done in a correct manner.

Finally, as promised, the answer to the question posed earlier about which probability is the one that should be considered in the DNA problem. The correct answer is *B*. As the guilt or innocence of the accused is the relevant question, the probability to consider is the person's innocence, given that their DNA matches the DNA found at the crime scene.

Endnotes

1. See the article by Nicholas Cowdery on the issue of majority verdicts, earlier in this issue of *Reform*.
2. *R v Clark* [2000] EWCA Crim 54; [2003] EWCA Crim 1020.
3. G Wansell, 'Whatever the coroner may say, Sally Clark died of a broken heart', *The Independent* (London), 18 March 2007.