



Sally Clark was cleared of murder in 2003 after statistical evidence was found to be flawed.

Improve statistics in court

Experts must agree a set of acceptable ways to assess and present forensic evidence, says **Norman Fenton**.

Misuse of statistics is a blight on the law. Verdicts have been influenced by incorrect statistical reasoning in dozens of widely documented cases^{1,2}. Sally Clark, for instance, was wrongly convicted of the murder of her two infant sons in a British court in 1999. She was cleared in 2003 after further investigation revealed that the probability of cot death had been calculated wrongly as being far too low. Many more cases go unnoticed.

Most common fallacies of statistical reasoning can be avoided by applying Bayes' theorem, a rule that allows the evidence to be weighted. Yet the Bayesian approach is widely misunderstood and mistrusted in court. A year ago a UK Court of Appeal ruling (known as *Regina v. T.*)³, dealt a further blow. Quashing a murder conviction in which the prosecution had relied heavily on Bayesian methods to present footwear-matching evidence, the judge said that Bayesian methods were an inadmissible way to present expert evidence — except for DNA and “possibly other areas where there is a firm statistical base”. Such restrictions are a backward step for justice. The consequences will be that expert evidence will be

misinterpreted and widely suppressed.

Forensic, statistical and legal experts around the world have reacted to the *Regina v. T.* ruling with concern and criticism⁴. To confront such challenges, I am setting up an international consortium of statisticians, forensic scientists and academic and practising lawyers (80 people signed up in the first 2 months) to develop guidelines for when and how Bayesian reasoning should be used to present evidence. We will agree initial objectives in a workshop in London in December (see go.nature.com/agn3or).

UPDATED ODDS

Bayes' theorem is the accepted rule for updating the probability of a hypothesis given new evidence. Importantly, the formula can be used to weight the impact of pieces of evidence individually or in combination. Suppose, for example, that blood found at the scene of a crime is of a type that is prevalent in one in every thousand people, and the defendant has the same blood type. Clearly, the match increases the probability that it was the defendant's blood at the scene. But by how much?

To answer this we need to compare the prosecution likelihood (the probability of seeing the evidence if the prosecution hypothesis is true) with the defence likelihood (the probability of seeing the evidence if the defence hypothesis is true). In this example, the former would be equal to one, and the latter would be one in a thousand. So we are 1,000 times more likely to observe the evidence if the prosecution hypothesis is true than if the defence hypothesis is true.

A simple measure of the impact of evidence is the likelihood ratio, the prosecution likelihood divided by the defence likelihood (1,000 in this case). Values above one favour the prosecution (the higher the better); those below one favour the defence (the lower the better). A value of exactly one means that the evidence is worthless (the prosecution and defence are affected equally). The likelihood ratio is extremely valuable, but to draw definitive conclusions we need Bayes' theorem to tell us how the odds change when new evidence is added: our updated (posterior) odds equal the prior odds for the prosecution hypothesis multiplied by the likelihood ratio.

So if, for example, there were 10,000 other adults who could have been at the scene of the crime, our prior odds would have been 10,000 to 1 against the prosecution hypothesis. Once we see the evidence, the revised odds still favour the defence, but they have dropped to 10 to 1 against the prosecution (equivalently the probability that the defendant was not at the scene has gone from 99.99% to about 91%).

A common error — known as the ‘prosecutor fallacy’ — is to assume that the (revised) probability of the defence hypothesis is the same as the defence likelihood. A prosecutor might state, for example, that ‘the probability that the defendant was not at the scene given this evidence is one in a thousand’, when actually it is 91%. This is one of the most common statistical legal mistakes.

SUMMARY

- Statistical methods for weighing evidence are being blocked in recent court rulings
- Lawyers may be unwilling to quantify subjective evidence, preventing legal conclusions from being drawn
- The difficulty of presenting complex probability calculations hinders their widespread acceptance
- An international consortium of statisticians, lawyers and forensic scientists is drawing up guidelines for the use of statistics in court

Such an error might be spotted through intuitive reasoning that tallies with the Bayes result: of the 10,000 other adults, about 10 should have the same blood type as the defendant, so the blood match tells us that the defendant is one of 11 who could have been at the scene (see 'Bayesian reasoning'). However, the explanation is rarely so simple.

LEGAL RESISTANCE

Despite its potential utility, Bayes' theorem is not trusted by much of the legal profession. There are two main reasons: subjectivity is often misunderstood, and Bayesian arguments are difficult to present in a way that everyone will comprehend.

Subjective judgement about uncertain evidence is at the heart of the jury trial process. And most lawyers are fiercely resistant to the idea that a numerical figure can be attached to this uncertainty. Suppose, for example, that a defendant was known to be part of a mob, one of whom committed an attack. An eyewitness for the prosecution estimates there were 50 other people in the mob. So the odds that the defendant committed the attack are 50 to 1 against. But this is subjective — another witness for the defence might claim there were 100 other people in the mob.

Bayesians would seek to assess such subjective information by considering a range of odds, from 50 to 1 to 100 to 1, which is acceptable to both sides. Few lawyers would use such subjective numbers. But that often means that they cannot make important conclusions. For example, if further evidence arose, say with a likelihood ratio (as in the blood-match example) of 1,000, then the revised odds would range from 20 to 1 to 10 to 1, and strongly favour the prosecution. Similarly, in a medical negligence case in 2010, I showed that the claimant's argument was favoured irrespective of the evidence supplied by both sides⁵.

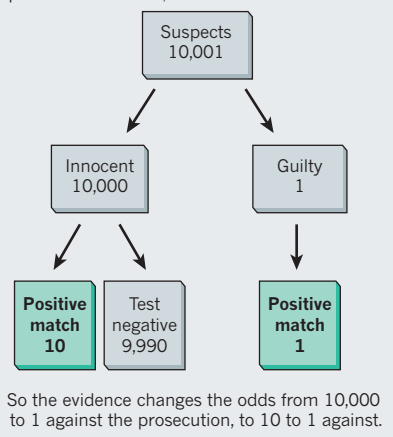
Misunderstandings of subjectivity have also restricted the types of evidence to which lawyers assume that Bayes and likelihood ratios can be applied. The requirement of a 'firm statistical base' as ruled by the judge in *Regina v. T.*, for example, has been interpreted in many cases as meaning that no subjective data can be used. But all probabilities — including DNA matches from the most comprehensive database — involve some subjective judgement⁶. It is better to acknowledge that subjectivity is inevitable and evaluate it using Bayes' theorem than to forego the evidence.

The fact that only the simplest Bayesian arguments can be explained from first principles in a way that laypeople will comprehend also limits their acceptance in court. A

NATURE.COM
See our News special on science in court:
go.nature.com/ez6pwk

BAYESIAN REASONING

Although only 1 of 10,001 suspects for a hypothetical crime is guilty, 11 would have a positive match for a blood type with a prevalence of 1 in 1,000.



real case may need to make complex assumptions or include multiple pieces of dependent evidence, and these cannot be relayed in straightforward decision trees. Most Bayesian calculations are so complicated that software tools are needed to complete them⁷. So, in the case of *Regina v. Adams*⁷, even when the defence expert presented the Bayesian calculations (balancing subjective evidence for the defence with the prosecution's DNA-match probability) to the jury from first principles, the exercise wasn't successful. It also backfired when the appeal judge ruled against such use of Bayes' theorem in future trials².

CONSENSUS NEEDED

The *Regina v. T.* ruling has drawn fierce criticism from many experts (including some lawyers) who appreciate the benefits of Bayes' theorem. They regard the ruling as a constraint on accepted scientific practice⁵; although it has also been praised as an attempt to rule out probabilistic forensic evidence that is based on 'unscientific' data. Lacking a definition of what is 'scientific', lawyers are erring on the side of caution and rejecting the use of likelihood ratios, even in areas in which their use used to be standard (such as in fibre and glass matching). Experts are then left to make vague assertions about how well the evidence supports a hypothesis.

There have been isolated attempts to improve the understanding of probability within the law. Last year, for instance, the UK Royal Statistical Society's statistics and law working group issued guidance aimed at judges, lawyers, forensic scientists and expert witnesses⁸. But legal practice will change only when a critical mass of international experts, supported by key members of the judiciary, reaches a consensus on two points: when Bayesian reasoning about evidence can and cannot be applied;

and how to get it accepted in court without having to present the calculations from first principles.

Wider acceptance of Bayesian analysis also requires lawyers, expert witnesses and others to understand that there is a crucial difference between the genuinely disputable (subjective) prior assumptions, and the (objective) Bayesian calculations required to compute the conclusions from the different disputed assumptions. Lay observers must accept that they can question only the assumptions that go into the Bayesian calculations and not the calculations themselves. By considering ranges of subjective assumptions we can address the most persistent objections to the use of the theorem. Acceptance of emerging Bayesian software tools will remove the need to go through the calculations in court from first principles.

Proper use of probabilistic reasoning has the potential to improve the efficiency, transparency and fairness of the criminal justice system. Bayesian reasoning can help experts to formulate accurate and informative opinions; courts to determine the admissibility of evidence and identify which cases should and should not be pursued; and lawyers to explain, and jurors to evaluate, the weight of evidence during a trial. It can also help to identify any errors and unjustified assumptions in expert opinions.

There is still widespread disagreement about the type of evidence to which Bayesian reasoning should be applied and how it should be presented. There are ways to overcome these technical barriers, but cultural barriers still remain between the fields of science and law, and these will be broken down only by achieving a critical mass of relevant experts and stakeholders, united in their objectives. The international consortium is building towards such a consensus. ■

Norman Fenton is professor of risk information management at Queen Mary, University of London and chief executive of Agena, Cambridge, UK.
e-mail: norman@eeecs.qmul.ac.uk

1. Aitken, C. G. G. & Taroni, F. *Statistics and the Evaluation of Evidence* 2nd edn (John Wiley & Sons, 2004).
2. Fenton, N. E. & Neil, M. *Aust. J. Legal Phil.* (in the press); available at http://www.eecs.qmul.ac.uk/~norman/papers/fenton_neil_prob_fallacies_June2011web.pdf
3. *Regina v. T.* [2010] available at <http://www.bailii.org/ew/cases/EWCA/Crim/2010/2439.pdf>
4. Berger, C. E. H., Buckleton, J., Champod, C., Evett, I. & Jackson, G. *Sci. Justice* **51**, 43–49 (2011).
5. Fenton, N. & Neil, M. *J. Biomed. Inform.* **43**, 485–495 (2010).
6. Thompson, W. C., Taroni, F. & Aitken, C. G. G. *J. Foren. Sci.* **48**, 47–54 (2003).
7. *Regina v. Adams* [1996] 2 Cr App R 467, [1996] Crim LR 898, *CA and Regina v. Adams* [1998] 1 Cr App R 377
8. Aitken, C., Roberts, P. & Jackson, G. *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings* (Royal Statistical Society, 2010).