INTERPRETATION OF THE VALUE OF EVIDENCE

Colin Graeme Girdwood AITKEN

Department of Mathematics and Statistics, University of Edinburgh, Edinburgh, United Kingdom

ABSTRACT: It has been argued for some time that the best way to evaluate evidence is to compare two probabilities, the probability of the evidence if the prosecution proposition is true and the probability of the evidence if the defence proposition is true. How these probabilities are used to evaluate evidence is explained. Various methods which have been used to interpret evidence are illustrated and their correctness or otherwise discussed. More education in the use of probabilities in a court of law is needed to ensure correct interpretations.

KEY WORDS: Bayes' theorem; Likelihood ratio; Defence fallacy; Prosecutor's fallacy.

Problems of Forensic Sciences, vol. XLVI, 2001, 14–23
Received 19 December 2000; accepted 15 September 2001

INTRODUCTION

In a trial under an adversarial system such as that used within the United Kingdom, two competing propositions are compared on the basis of evidence put forward by either the prosecution or defence. One proposition is that suggested by the prosecution, the other is that suggested by the defence. Consider an example where the evidence is that the DNA profile of an accused matches, in some sense, the DNA found at the scene of the crime. The prosecution proposition may be that the accused is guilty of the crime as charged and the defence proposition may be that the accused is innocent of the crime as charged. Often, the evidence does not enable one to test the relative merits of these, so-called, ultimate issues. It may be that one can only consider intermediate issues. These may be that the accused was or was not at the scene of the crime, the former proposition being that of the prosecution and the latter being that of the defence. Further removed still from the ultimate issue are the propositions that the accused was or was not the source of the DNA found at the scene of the crime. There is also the possibility that the defence does not need to specify any proposition other than it is not that of the prosecution [4, 5, 8, 9].

The court is interested in the truth or otherwise of the prosecution and defence propositions. It may wish to consider the probability that the prosecution proposition is true and the probability that the defence proposition is true. In contrast, the forensic scientist is interested in:

a) the probability of the evidence if the prosecution proposition is true, and

b) the probability of the evidence if the defence proposition is true.

Here an assumption is made that it makes sense to write of the "probability of the evidence". This assumption is reasonable in the context of DNA profiling where genetics enables the calculation of probabilities to be made. It may not be so reasonable in the context of eye-witness evidence for example.

EXAMPLES OF INFERENTIAL ERRORS

The distinction between the probability of the evidence and the probability of the proposition is illustrated by an example given in Darroch [6]. A rape is committed in a town. There are 10 000 men of suitable age in the town of whom 200 work underground at a mine, and who will be called miners. Evidence of mineral traces at the scene of the crime indicates that the criminal is one of the 200 miners. A suspect is identified and traces of minerals are found on his clothing. The suspect is a miner. The evidence is the mineral traces which are found at the scene of the crime and on clothing belonging to the suspect. The prosecution proposition is that the suspect is guilty. The defence proposition is that the suspect is innocent. Consider the following arguments.

One of the 200 miners is guilty; the other 199 are innocent. All 200 miners have mineral traces somewhere on some of their clothing by the nature of their job. There are 9 999 innocent men in the town, of whom 199 are miners. Of the 9 999 innocent men, 199 have mineral traces on their clothing which may be found by investigators. The mineral traces are evidence which links the person on whose clothes they have been found with the scene of the crime. The probability of finding this evidence on an innocent man is taken to be equal to the proportion of innocent men who have this evidence on their clothing. This is 199/9999 or approximately 200/10 000 or 0.02. This is taken as the probability of the evidence if the defence proposition is true and the suspect is innocent.

Consider the transpose of this argument. There are 200 miners with mineral traces on their clothing. Of these 200, 199 are innocent. The probability that a man is innocent, when the mineral traces (evidence) have been found on his clothing, is taken to be equal to the proportion of men with mineral traces on their clothing who are innocent. This is 199/200 or 0.995. The probability that a man is guilty, when the mineral traces (evidence) have been found on his clothing is taken to be equal to the proportion of men with mineral traces on their clothing who are guilty. This is 1/200 or 0.005. This is the complement of the probability of innocence. This is taken as the probability the prosecution proposition is true and the suspect is guilty if the evidence has been found on the suspect's clothing.

Table I illustrates the difference between the probabilities.

TABLE I. THE CLASSIFICATION OF 10,000 MEN INTO (MINERS, NOT MINERS) AND (GUILTY, NOT GUILTY)

	Miner	Not miner	Totals
Guilty	1	0	1
Not guilty	199	9 800	9 999
Totals	200	9 800	10 000

The total in the row labelled "not guilty" is 9 999. These are the number of innocent men. The number of those in the column labelled "miner" is 199. These are the innocent people who have mineral traces ("the evidence") somewhere on their clothing. The ratio 199/9 999 (approximately 0.02) is taken as the probability of finding the evidence on an innocent man.

The total in the column labelled "miner" is 200. The number of those that are in the row labelled "not guilty" is 199. These are the people who have mineral traces ("the evidence") somewhere on their clothing who are innocent. The ratio, 199/200 (0.995) is taken as the probability that a man is innocent when the evidence has been found on his clothing.

The probability of the evidence being found on the suspect's clothing if the suspect is innocent is 0.02. The probability the suspect is guilty if the evidence has been found on the suspect's clothing is 0.005. Thus, a small value for the probability of finding evidence on an innocent person does not mean a large value for the probability that the person is guilty. The mistake of associating a small probability of finding the evidence on an innocent person with a large probability of guilt is known as the prosecutor's fallacy or fallacy of the transposed conditional [15].

Sometimes it is not so easy to notice an occurrence of the fallacy of the transposed conditional. The following quote is taken from a report in "The (London) Times" of May 6th, 2000 of a trial for the murder of a woman. Specks of foam taken from the passenger seat in the cab of the defendant's lorry matched samples found on the dead woman's smock, and tests showed that they must have stuck to her clothing a short time before she was killed. In the report the prosecuting barrister was quoted as saying: "Spots of blood found inside the cab's sleeping bag were also shown to match that of the dead woman. The chances that they came from someone else were 3.5 million to 1".

The word “chances” is used as a synonym for “probability”. The phrase “3.5 million to 1” is a statement of odds not chance, or probability. When odds are quoted it should always be stated whether they are odds against or odds for an event. The statement is really saying, however, that the probability the spots of blood came from someone else, other than the dead woman, was 1 in 3.5 million. The origin of the blood is an intermediate issue and the statement reported in the newspaper is a probability of an intermediate issue.

The prosecution proposition is that the blood in the sleeping bag is that of the dead woman. The defence proposition is that the blood in the sleeping bag is not that of the dead woman. The evidence is that the blood in the sleeping bag matches that of the dead woman. The probability of the evidence if the defence proposition is true is 1 in 3.5 million. This is not to say that the probability the defence proposition is true is 1 in 3.5 million.

An example in which an incorrect probability was used occurred in a trial in England in November 1999. A mother was convicted of murdering her two infant children by smothering them. Testimony was given that the probability of two cot deaths occurring in the same family was 1 in 73 million. The mother was convicted. She appealed against the verdict but the appeal was rejected. A very good discussion of the original trial is given in an editorial in the “British Medical Journal” by Watkins [16]. Suspicion fell on the mother only after the second death. Thus, a relevant probability is not that of two deaths but of a second death, given that a first death has occurred. This probability is 1 in 8 500 (the square root of 1 in 73 million), a far less imposing figure. Watkins cites survey evidence [12] that, if cot deaths are random events, then two cot deaths will occur in the same family somewhere in England once every seven years. Thus, every seven years a mother could be found guilty, erroneously, of murder. However, cot deaths are not random events. Two studies imply recurrence somewhere in England about once every eighteen months [10, 11]. Watkins [16] commented towards the end of his editorial that “guidelines for using probability theory in criminal cases are urgently needed”.

PROBLEMS OF THE TRANSPOSED CONDITIONAL

Consider the following sets of statements and questions.

If I am a monkey	If I am guilty
Then I have two arms and two legs.	Then my DNA profile matches that found at the crime scene.
If I have two arms and two legs.	If my DNA profile matches that found at the crime scene.
Am I then a monkey?	Am I then guilty?

The left-hand set refers to monkeys and arms and legs. These are easily recognised concepts. The first "if-then" pair is perfectly acceptable. The answer to the question with the other pair is "don't know". In an analogous way, the first "if-then" pair for the right-hand set is perfectly acceptable. The answer to the second question depends on the number of loci used in constructing the profile. In general the answer will be "don't know". Evidence of a match may provide very strong support for the proposition that the suspect was guilty. Without other corroborating evidence it will rarely be conclusive support. The confusion of the two pairs of statements in the right-hand set is the basis of the prosecutor's fallacy and explains the alternative name of the fallacy of the transposed conditional.

An artificial example of the prosecutor's fallacy is given here with a comment on the general principles. A crime is committed and a blood stain is found at the scene. All innocent sources of the blood are eliminated and it is accepted that the stain has been left by the criminal. The size of the population from which the criminal is assumed to come is calculated to be 800 000. The blood is of a type which is present in 1% of the population. A suspect is identified by other means and found to have blood of this type. He is brought to trial.

The prosecutor comments (using "chance" as a synonym for "probability") that "there is a 1% chance that the defendant would have the crime blood type if he were innocent. Thus there is a 99% chance that he is guilty". This is an example of the prosecutor's fallacy. Remember the miners. The probability the defendant would have mineral traces on his clothing if he were innocent was 0.02. However, the probability that the defendant was guilty, given that mineral traces were found on his clothing, is 0.005. It is not 0.98 (1–0.02). In order to determine the probability of guilt information is needed of the size of the relevant population: the total number of people in the population and the number, or proportion, that would have mineral traces on their clothing.

In the blood stain example, the defence lawyer may comment that the blood type would be found in approximately 8 000 people (1% of 800 000). The evidence has thus provided a probability of 1 in 8 000 that the defendant is guilty and is of no relevance. It is these last two words which gives rise to the phrase "the defence fallacy". Before the presentation of the evidence there were 800 000 people in the pool of potential criminals, after the presentation of the evidence there were only 8 000 people in the pool. Evidence which can reduce the pool by such a proportion has to be of relevance.

Another, trivial, example may illustrate this better. An eyewitness to a crime says that the criminal was tall. There are many tall people. It is not suggested that, because of this, the eyewitness evidence is irrelevant.

In both cases, obviously, other evidence is needed before a conviction may be obtained. There needs to be evidence to eliminate the other 7 999 people of that blood type in the blood stain example. There needs to be evidence to eliminate other tall people in the eyewitness example.

The two fallacious arguments may be summarised as follows. The prosecutor argues that, because the evidence is very unlikely in an innocent person, the accused must be guilty. The defence lawyer argues that, because there is a large number of people with the evidence (even if it is only a small proportion of the population), and the accused is only one of this large number, he must be innocent.

THE LIKELIHOOD RATIO AND BAYES' THEOREM

The following example is to be found in Royall [13]. Consider a pack of 52 playing cards, face down so that the backs of the cards are visible. However, there are two possibilities of what the cards may be. It may be what is thought of as a normal pack of cards (four suits – Spades, Hearts, Diamonds, Clubs – with thirteen cards in each suit) or it may be a pack consisting of 52 aces of diamonds. It is not known which. The top card is turned over: it is the Ace of Diamonds. The probability of turning over the ace of diamonds if the pack is a normal pack is 1 in 52 (1/52). The probability of turning over the Ace of Diamonds if the pack consists solely of aces of Diamonds is 1. In other words, it is 52 times more likely that the Ace of Diamonds will be turned over if the pack consists solely of Aces of Diamonds than if it is a normal pack. Which pack is it? Does the pack contain 52 Aces of Diamonds? In a criminal context, let the prosecution proposition be that the pack consists of 52 Aces of Diamonds and the defence proposition be that the pack is a normal pack. The evidence (of an Ace of Diamonds) is 52 times more likely if the prosecution proposition is true than if the defence proposition is true. Very few, if any, readers will, however, believe that the pack is not a normal pack of cards. This is because of prior experience. The prior experience that the pack is a normal pack outweighs the evidence that supports the proposition that the pack contains 52 Aces of Diamonds.

An analogous example may be given in the context of forensic science. Blood stain evidence is found at the scene of a crime which can be assumed to come from the criminal. It is of a type which is present in one fifty-second (1/52) of the population. The probability the criminal has this blood type is 1, the probability an innocent person has this blood type is 1/52. Thus, it is 52 times more likely that the accused will have the blood type if he is guilty than if he is innocent. Is he innocent or guilty? In a criminal context, let the prosecution proposition be that the accused is guilty and the defence proposition be that the accused is innocent. The evidence is 52 times more likely if the prosecution proposition is true than if the defence hypothesis is true.

One can alter the numbers but the question remains the same. For example, suppose it is 520 or 5 200 or 52 000 or 5.2 million times more likely that the accused will have the blood type (or DNA profile) if he is guilty than if he is innocent. Is he innocent or guilty?

The phrase "more likely that" is the verbal interpretation of what is known in statistical circles as the likelihood ratio. The likelihood ratio is:

Probability of the evidence if the prosecution proposition is true / probability of the evidence if the defence proposition is true.

These probabilities are the probabilities of interest for the forensic scientist. The probabilities of interest for the courts are the probabilities that the prosecution or defence propositions are true.

The probabilities that the prosecution and defence propositions are true are complementary probabilities, that is they add up to 1; one or either but not both is true. The accused is guilty or he is innocent (not guilty). The accused was present at the crime scene or he was not. The accused is the source of the DNA at the crime scene or he is not. The ratio of two complementary probabilities is known as the odds. The odds are said to be in favour of the event in the numerator or against the event in the denominator. Thus, the ratio of the probability of the prosecution proposition of guilt to the probability of the defence proposition of innocence is known as the odds in favour of guilt or the odds against innocence. When this ratio is determined before (or prior to) the presentation of the evidence it is known as the prior odds in favour of the prosecution proposition. When it is determined after (or posterior to) the presentation of the evidence it is known as the posterior odds in favour of the prosecution proposition. The likelihood ratio is the factor which converts the prior odds into posterior odds [1]. The general result, which is a version of the theorem known as Bayes' Theorem is:

Posterior odds in favour of the prosecution proposition = the likelihood ratio × the prior odds in favour of the prosecution proposition.

A likelihood ratio is not a statement of odds. The two probabilities which contribute to the ratio are not complementary probabilities. The numerator is the probability of the evidence if the prosecution proposition is true, the denominator is the probability of the evidence if the defence proposition is true. The conditioning event is different in the two cases. In the first, it is the prosecution proposition which is taken to be true, in the second it is the defence proposition which is taken to be true. For the calculation of odds, the conditioning event has to be the same in the numerator and denominator.

For example, the posterior odds is the ratio of the probability of the prosecution proposition assuming (conditioned on) the truth of the evidence to the probability of the defence proposition assuming (conditioned on) the truth of the evidence. In both probabilities the conditioning event is the evidence. Thus with the example of the pack of cards the prior odds are so overwhelmingly in favour of the normal pack that a likelihood ratio of 52 still leaves a posterior odds strongly in favour of the normal pack.

From the general result, it can be seen that a value of the likelihood ratio greater than one supports the prosecution proposition in the sense that the posterior odds in favour of the prosecution proposition are then greater than the prior odds in favour of the prosecution proposition, by a factor corresponding to the value of the likelihood ratio. Similarly, a value less than one supports the defence proposition. A value of one (or even close to one) is neutral and the evidence may be said to be irrelevant.

THE VALUE OF THE EVIDENCE

The likelihood ratio has been called the value of the evidence [1]. A verbal convention is used by the Forensic Science Service of England and Wales [8] for the interpretation of the value in which a value greater than one is said to support the proposition proposition relative to the defence proposition, with various descriptive adjectives depending on the value. These are given in Table II.

TABLE II. VERBAL CONVENTIONS FOR THE SUPPORT, OR VALUE, V, OF THE EVIDENCE FOR THE PROSECUTION PROPOSITION. CORRESPONDING RECIPROCAL VALUES PROVIDE SUPPORT FOR THE DEFENCE PROSECUTION

Numerical range for the value V	Verbal convention
$1 < V < 10$	Limited evidence to support
$10 < V < 100$	Moderate evidence to support
$100 < V < 1000$	Moderately strong evidence to support
$1000 < V < 10{,}000$	Strong evidence to support
$10{,}000 < V$	Very strong evidence to support

Evett et al. [8] comment that it is accepted practice to use the phrase *extremely strong* for likelihood ratios of one million or more. Aitken and Taroni [2] suggest a ten point scale using the logarithm of the likelihood ratio to report the value of the evidence.

An example of the use of the likelihood ratio for interpretation in a law court is given in *R. v. Montella,* (1 NZLR High Court, 1992, 63–68). As a result of a laboratory analysis the frequency of a DNA profile in the population was 1 in 12 400. The interpretation of the evidence was as follows: "A DNA profiling examination of the samples strongly supports a contention that the semen stain on the underpants of the complainants came from the accused. It is said that the likelihood of obtaining such DNA profiling results is at least 12 400 (the reciprocal of the frequency) times greater if the semen stain originated from the accused than from another individual".

A discussion of this, and other interpretations of DNA evidence, is given in Taroni and Aitken [14] in a report of a survey of those associated with the forensic process, including students of forensic science and of forensic medicine, advocates and forensic scientists. The replies from the participants in the survey showed considerable confusion in the comprehension of the expert's statement in Montella. This is unfortunate. The likelihood ratio is the best method of evaluating and interpreting evidence [1] as it differentiates explicitly between the role of the court (judge and jury) and the scientist. The scientist comments on the value of the evidence not on the outcome of an issue.

As in the Editorial in the "British Medical Journal" [16], Taroni and Aitken [14] comment that more education is needed about the use of probabilities in courts of law. It has been shown [3] that an increase in education on the Bayesian framework leads to a reduction in the number of erroneous interpretations.

References:

1. Aitken C. G. G., Statistics and the evaluation of evidence for forensic scientists, John Wiley and Sons, Chichester 1995.
2. Aitken C. G. G., Taroni F., A verbal scale for the interpretation of evidence, *Science and Justice* 1998, vol. 38, pp. 279–281.
3. Carracedo A., Barros F., Lareu M. V. [et al.], Focusing the debate on forensic genetics (correspondence), *Science and Justice* 1996, vol. 36, pp. 204–206.
4. Cook R., Evett I. W., Jackson G. [et al.], A hierarchy of propositions: deciding which to address in forensic casework, *Science and Justice* 1998, vol. 38, pp. 231–239.
5. Cook R., Evett I. W., Jackson G. [et al.], A model for case assessment and interpretation, *Science and Justice* 1998, vol. 38, pp. 151–156.
6. Darroch J., Probability and criminal trials: some comments prompted by the Splatt trial and the Royal Commission, *The Professional Statistician* 1987, vol. 6, pp. 3–7.
7. Evett I. W., Bayesian inference and forensic science: problems and perspectives, *The Statistician* 1987, vol. 36, pp. 99–105.
8. Evett I. W., Jackson G., Lambert J. A. [et al.], Impact of principles of evidence interpretation on structure and content of statements, *Science and Justice* 2000, vol. 40, pp. 233–239.

9. Evett I. W., Jackson G., Lambert J. A., More on the hierarchy of propositions: exploring the distinction between explanations and propositions, *Science and Justice* 2000, vol. 40, pp. 3–10.
10. Guntheroth V. G., Lohmann R., Spiers P. S., Risk of sudden infant death syndrome in subsequent siblings, *Journal of Paediatrics* 1990, vol. 116, pp. 520–524.
11. Irgens I. M., Skjaerven R., Petersen D. R., Prospective assessment of recurrence risk in sudden infant death syndrome, *Journal of Paediatrics* 1984, vol. 104, pp. 349–351.
12. Reese A., In "Statistics and Justice", November 1999, http://www.stats.gla.ac.uk/allstat/
13. Royall R. M Statistical evidence: a likelihood paradigm, Chapman and Hall, London 1997.
14. Taroni F., Aitken C. G. G., Probabilistic reasoning in the law. Part 1: assessment of probabilities and explanation of the value of DNA evidence, *Science and Justice* 1998, vol. 38, pp. 165–177.
15. Thomson W. C., Schumann E. L., Interpretation of statistical evidence in criminal trials. The prosecutor's fallacy and the defence attorney's fallacy, *Law and Human Behaviour* 1987, vol. 11, pp. 167–187.
16. Watkins S. J., Conviction by mathematical error, *British Medical Journal* 2000, vol. 320, pp. 2–3.