Probability and Statistics in the Law

Philip Dawid

Department of Statistical Science

University College London

dawid@stats.ucl.ac.uk

Abstract

The field of legal reasoning is full of logical subtleties and probabilistic pitfalls. I survey a number of these, pointing out some of the problems and ambiguities, and various attempts to deal with them. Some celebrated court cases are used for illustration.

Keywords: Bayesian network, Bayes's theorem, DNA profiling, forensic identification, likelihood ratio, sudden infant deaths.

1 Introduction

Although the disciplines of Statistics and Law might seem far apart, they share some fundamental interests — in particular, the interpretation of evidence, testing of hypotheses, and decision-making under uncertainty. However, their differing backgrounds and approaches can often lead to misunderstandings, such as in the celebrated "Collins case" [17].

In recent years it has become apparent that problems arising in legal settings raise some fascinating and delicate issues of statistical logic, and that, in turn, proper application of statistical reasoning has a rôle to play in the pursuit of justice. In this paper I explore some of these logical issues, with reference to some real cases: see [6] for some further background.

2 Sudden infant deaths

There have been a number of recent cases in the UK where two or more young children in a family have died suddenly from no obvious cause, and, even though there is no specifically incriminating evidence, their mother has been convicted of murdering them. In the case of Sally Clark, a paediatrician testified at trial that the probability P that her two sons would have died of SIDS (unexplained natural causes) was 1 in 73 million. That figure was widely and properly criticised, but it can not be denied that P is extremely small. The question is: What are we to make of such "statistical evidence"?

2.1 The prosecutor's fallacy

The correct interpretation $P = \Pr(\mathcal{E} \mid \overline{G})$ (where $\mathcal{E}$ denotes the evidence — here the fact of two infant deaths — G denotes "guilt" and $\overline{G}$ "innocence") is easily distorted into: $P = \Pr(\overline{G} \mid \mathcal{E})$. After all, to say that there is 1 chance in 73 million that the children died of natural causes appears to be just the same as saying that this is the probability that the mother did not kill them — seemingly overwhelming evidence for her being guilty. This mistaken "transposition of the conditional" is so common in court, where it usually favours the prosecution, that it has been termed "the prosecutor's fallacy" (see [18] for a clear account of the prosecutor's fallacy, and suggestions as to how it might be avoided). It would have been hard for Sally Clarks's jury to ignore this seemingly powerful argument, and they did in fact convict.

2.2 Counter-argument

There is an obvious counter-argument in this case, which I presented at appeal. We are comparing two alternative hypotheses: two deaths by SIDS, and two deaths by murder. If the chance of the former is relevant, should not that of the latter be equally relevant? Using UK data, one could argue for a double murder figure of around 1 in 2 billion, to set against the SIDS figure of 1 in 73 million. One can see prosecution and defence brandishing their respective figures in adversarial combat, but the correct approach is to realise that it is their relative, not absolute, values that matter. In fact, their ratio $(1/2 \text{ billion})/(1/73 \text{ million}) = 0.0365$ can be interpreted as the odds on guilt given the evidence of the two deaths, implying a guilt probability of only 3.5%.

In the event, although the appeal court accepted that there had been some problems with the presentation of the statistical evidence at trial, it was not interested in properly identifying and understanding the logical issues involved. Sally Clark was eventually cleared on entirely unrelated grounds.

3 Identification evidence

Many criminal cases revolve about the issue of identity: is the suspect S the same person as the perpetrator C of the crime? Similar issues arise in civil cases, such as disputed paternity.

Forensic trace evidence is often brought in such cases. From the crime scene we obtain information I_C that can be assumed to apply to the criminal C — thus we may have a fingerprint, a footprint, fibres, or eye-witness evidence of sex, age, race, *etc.* With advances in DNA technology, it is now common to obtain a DNA profile of the criminal from biological material left at the scene of the crime. In addition, we have similar information I_S about the suspect S, for example his DNA profile. When this matches the crime sample, *i.e.* $I_S = I_C, = x$, say, that is clearly evidence in favour of the two samples having the same source. But how are we properly to weigh and apply this evidence?

One relevant feature of match evidence is the *match probability* P: this is the frequency with which the characteristic x occurs in the population at large. In the case of DNA profiling, the match probability can be estimated from population figures and genetic theory. Very tiny match probabilities, even as small as one in one billion, are now routine.

3.1 The prosecutor's fallacy

We henceforth implicitly condition on the suspect's characteristic: $I_S = x$. The match probability can be written as $P = \Pr(I_C = x \mid C \neq S)$. If we describe this as "the probability that the crime sample came from some one other than S", we are immediately in danger of committing the prosecutor's fallacy of § 2.1, which interprets P as $\Pr(C \neq S \mid I_C = x)$, *i.e.* the probability, in the light of the match, that S is innocent — implying that the probability of guilt G is $1 - P$. If say $P = 0.0000001$, the jury or judge might well understand that the probability is only 1 in 10 million that S is not guilty, and convict.

3.2 The defence argument

A counter-argument along the lines of § 2.2 does not succeed here, since the probability of a match under the alternative hypothesis of guilt is unity.

Instead the defence might point out that there are $N + 1$ (say) people who could have committed this crime. One of these is truly guilty, and so matches the crime trace; while we would expect to see approximately NP innocent matches out of the remaining N innocent individual. We thus expect a total of $1 + NP$ matching individuals, of whom just 1 is guilty. If all we know about S is that he matches, the probability he is guilty is $1/(1 + NP)$. Taking $N = 30$ million and again $P = 0.0000001$, we would expect 3 innocent matches, for a final guilt probability of 1 in 4 — which is certainly not evidence "beyond a reasonable doubt".

3.3 Some other arguments

The above defence argument can be varied in a number of ways [4], many of which are intuitively appealing — and have been recommended for use — but are in fact fallacious.

In all cases we assume that, prior to any evidence, any of the $N + 1$ individuals in the population is equally likely to be guilty, and that the only evidence $\mathcal{E}$ against S is that of the match: $I_S = I_C = x$. For illustration we take $N = 100$, $P = 0.004$.

Let M denote the unknown number of individuals i having $I_i = x$. We suppose that, before any samples are measured, M has the binomial distribution $\text{Bin}(N + 1; P)$. We have $\Pr(G \mid \mathcal{E}, M) = M^{-1}$, and the final guilt probability, $\Pr(G \mid \mathcal{E})$, can be obtained by taking the expectation of this quantity with respect to the conditional distribution of M, given the evidence $\mathcal{E}$.

1. The evidence tells us that $M \geq 1$, and simple conditioning on this yields

$$\Pr(G \mid \mathcal{E}) = \text{E}(M^{-1} \mid M \geq 1).$$

   For $M \sim \text{Bin}(N + 1; P)$ this is not easily expressed in closed form, but can be calculated: for our numbers it evaluates to **0.902**.

2. An alternative argument is that, given the evidence, we know that there is one guilty match, and, out of the remaining N innocent individuals, each has, independently, probability P of supplying a match. So the conditional distribution of M is $1 + \text{Bin}(N; P)$. Using this to take the expectation of M^{-1} yields

$$\Pr(G \mid \mathcal{E}) = \frac{1 - (1 - P)^{N+1}}{(N + 1) P}$$

   which, for our values, gives **0.824**.

3. Finally, the correct approach.

   We can consider the total evidence ($I_C = x, I_S = x$) as the results, both successes, of two draws, *with replacement* (since C and S could be the same individual),

from the population. The probability of this, given $M = m$, is m^{-2} (for $m \geq 1$), and, using Bayes's Theorem, the resulting conditional distribution of M is

$$\Pr(M = m \mid I_C = x, I_S = x)$$
$$= \; c\, m \begin{pmatrix} N \\ m - 1 \end{pmatrix} P^{m-1} (1 - P)^{N-m+1}$$
$$(m = 1, \dots, N + 1),$$

where the normalising constant is $c = 1/(1 + NP)$. Taking the expectation of M^{-1} with respect to this distribution then yields

$$\Pr(G \mid \mathcal{E}) = 1/(1 + NP),$$

in agreement with the original (and much simpler) defence argument. This evaluates numerically to **0.714**.

The above is just one example of the pitfalls besetting logical and probabilistic reasoning in cases at law: see [1, 10, 11, 12] for a number of other subtle issues of interpretation of forensic identification evidence.

3.4 Bayes

A serious problem with both the prosecution and the defence arguments is that they do not allow for the incorporation of any other evidence in the case. The coherent approach to combining identification and other evidence is through Bayes's Theorem: *Posterior Odds* (on G) = *Prior Odds* × *Likelihood Ratio*, where the other evidence is accounted for in the prior odds, and the likelihood ratio based on evidence $\mathcal{E}$ (where here $\mathcal{E}$ is the match evidence "$I_C = I_S = x$") is defined by:

$$LR = \frac{\Pr(\mathcal{E} \mid G)}{\Pr(\mathcal{E} \mid \overline{G})}. \tag{1}$$

Because there is typically a subjective element in assessing prior probabilities, it is often argued that experts should confine their evidence to assessment of the more "objective" likelihood ratio, leaving the court to apply Bayes's Theorem with its own prior inputs. (However, see §§ 4 and 5 below concerning

ambiguities in the definition of the likelihood ratio.)

In the case of identification evidence we can (usually) take $\Pr(\mathcal{E} \mid G) = 1$, $\Pr(\mathcal{E} \mid \overline{G}) = P$, so that the likelihood ratio is $1/P$. If the prior probability of guilt is π, the posterior probability is $\pi/(\pi + P - \pi P)$. This agrees (approximately) with the argument of the prosecutor when $\pi = 0.5$, and (exactly) with that of the defence when all $N + 1$ potential culprits are *a priori* equally likely to be the guilty party. This might be seen as support for the defence argument in the absence of any other evidence.

An interesting application of Bayes's Theorem was in the 1995 trial of Denis John Adams for sexual assault. The only prosecution evidence was a DNA match, with match probability assessed between 1 in 2 million and 1 in 200 million. The defence relied on the fact that the victim did not identify Adams at an identification parade, and also said that he did not look like the man who had raped her. In addition Adams's girlfriend testified that he had been with her at the time of the crime.

On the basis that the criminal was likely to be a local male aged between about 18–60, the prior probability of guilt, before any evidence, might be assessed at around one in 200,000. The likelihood ratio based on the DNA match is $1/P = 2$ million, say. That based on the victim's non-recognition of Adams could be assessed at, say, $0.1/0.9 = 1/9$, and that based on his girlfriend's alibi at, say, $0.25/0.5 = 1/2$. Assuming suitable independence, the posterior odds on guilt become $(1/200{,}000) \times (2{,}000{,}000) \times (1/9) \times (1/2) = 5/9$, corresponding to a posterior probability of 35% (though rising to 98% if we take $P = 1$ in 200 million).

In the actual case this argument was allowed at trial (although it does not seem to have impressed the jury, who convicted), but ruled out on appeal, on the basis that explaining how to think about probabilistic evidence "usurps the function of the jury", which "must apply its common sense". Unfortunately that leaves the door wide open to the prosecutor's fallacy and other tempting but misleading arguments.

4   Database search

In some cases where a DNA profile is found at the crime scene there may be no obvious suspect. Then a trawl may be made through a police computer DNA database in the hope that it will throw up a match. Suppose this happens: how, if at all, does the fact of the database search affect the strength of the evidence against a suspect so identified?

For definiteness, suppose that the database $\mathcal{D}$ is of size $n = 10{,}000$, that the match probability of the crime profile is $P = 1$ in 1 million, and that exactly one profile — that of S, say — in the database is found to match.

One intuition is that the database search has eliminated 9,999 individuals who would otherwise have remained alternative suspects. Given the very large initial number of alternative suspects, this has the effect of rendering the evidence in favour of S's guilt *very marginally stronger*. The relevant likelihood ratio is still close to 1 million.

An entirely different intuition proceeds by analogy with frequentist statistical approaches to testing multiple hypotheses. This would adjust the match probability to take account of the 10,000 possible ways of obtaining a match in the database, replacing it by the value, close to $10{,}000 \times (1$ in 1 million$)$ $= 1/100$, of the probability of finding a match in the database, if it does not include the criminal. And a match probability of only 1 in 100 is *vastly weaker* evidence than one of 1 in 1 million. In particular, it corresponds to a likelihood ratio in favour of guilt of 100, rather than 1 million. Stockmarr [22] has argued in favour of this likelihood ratio of 100, which relates directly to the hypothesis $H_{\mathcal{D}}$ that some one in the database is guilty, as against that of 1 million, which relates to the hypothesis H_S that S is guilty — on the grounds that the former hypothesis is data-independent, while the latter can not even be specified in advance of performing the search. However, while such data-dependence can affect frequentist inferences, its relevance to likelihood inference is

arguable.

A way of bridging the apparent chasm between these two intuitions appears when we realise that the *prior probability* of $H_\mathcal{D}$ is about 10,000 times larger than that of H_S. When we move between these hypotheses, this difference in prior odds cancels exactly with the difference in the associated likelihood ratios, so that both approaches lead to the identical posterior probability (whether for H_S or for $H_\mathcal{D}$ being unimportant, since these become logically equivalent once we have found that S is the unique profile in $\mathcal{D}$ matching the crime sample).

While this may resolve the conceptual paradox, a practical problem remains. If "objectivity" requires that we offer likelihood ratios, rather than posterior probabilities, in evidence, which should we give? — and how can we ensure that their meaning and use is properly appreciated?

For further (heated) discussion of these issues see [2, 15, 22, 16, 5].

5 Multiple perpetrators and stains

A similar problem [19, 20] arises when we know there were two criminals, two distinct DNA stains (say one on a pillow, one on a sheet) have been found at the scene of the crime, and there is a single suspect, S, who matches one of them — say the pillow stain — with its associated match probability P. How is the strength of the evidence against S affected by the multiplicity of stains?

Once again there is a choice of hypotheses to compare, these being logically equivalent in the light of the findings, but not in advance. A first approach compares "S left one of the two stains" with "S did not leave either stain'; a second compares "S left the pillow stain" with "S did not leave either stain"; and yet a third compares "S left the pillow stain" with "S did not leave the pillow stain". Under some assumptions, the associated likelihood ratios are, respectively, $\frac{1}{2}P$, P, and $\frac{1}{2}P \times (2 - \delta)/(1 - \delta)$, where δ is the prior probability that S is guilty. And once again,

the differences between these disappear after they are combined with their varying relevant prior odds. In [8] I argue that it is the first of these likelihood ratios that relates most directly to the relevant issue: that of the guilt of S. But one must also take into account the knowledge that there were two culprits, which effectively doubles the prior probability of S's guilt as compared with a single-suspect case.

6 Mixed stains

In many cases, *e.g.* involving a rape or scuffle, a crime trace may clearly[1] be a mixture of biological material from more than one individual. We may or may not know how many contributors are involved, or the identity of some of them. It is sometimes possible to separate out the components of different contributors, *e.g.* by taking into account the differing amounts of DNA at different bands, but this is unreliable.

Suppose we have a suspect S who "matches" the crime trace, in that all his bands are contained in it. What is the strength of the DNA evidence against him? This can involve complex and subtle calculations and be sensitive to assumptions made.

6.1 O. J. Simpson

In the celebrated trial of O. J. Simpson for double murder, one of the crime samples could be explained as a mixture of blood from Simpson and one of the victims, Ron Goldman. At a certain locus, Simpson had genotype AB, Goldman AC, and the crime sample had ABC. In pre-trial depositions[2], the prosecution argued that the relevant match probability P should be taken as the frequency of Simpson's genotype AB — about 5%. (Such a P would be multiplied by similar figures calculated for other loci to obtain an overall match probability). The defence argued that P should be the total probability of any of the genotypes, AA, AB, AC, BB, BC, CC, that would have "matched" the crime sample:

---

[1]For example, because it has more than two bands at some locus.

[2]http://tinyurl.com/2fhsx

about 39%.

However, on the assumption that the mixture consists of Goldman and the culprit, the culprit must have type AB, BB or BC. These have combined probability 21%, and it is the reciprocal of this figure for P that yields the correct likelihood ratio. If we did not know Goldman's genotype, or thought that the other contributor was some one else, we need to conduct a more complex calculation to obtain the relevant likelihood ratio. Interpreting this as P^{-1}, we again obtain $P \approx 21\%$ (though this is an accidental concurrence of two potentially different figures).

7 Missing suspect

When a suspect, or other relevant party, is not available for DNA profiling, useful information can sometimes be obtained by profiling relatives — although the analysis then required can be both conceptually and computationally challenging.

7.1 Hanratty

In 1962 James Hanratty was executed for rape and murder. In 1998 a DNA profile, assumed to be from the culprit, was extracted from some items that had been stored since the crime. Its associated match probability was around 1 in 2.5 million. Ever ready to fall for the prosecutor's fallacy, the Press duly reported this as "There is a 1 in 2.5 million chance that Hanratty was not the A6 killer" — even though, since Hanratty's DNA was unavailable, there was no more evidence against him than against any one else.

Hanratty's mother and brother now offered their own DNA for profiling — and this failed to exclude him. Again reports of the above match probability circulated as evidence of his guilt. In fact, the actual likelihood ratio, based on the indirect evidence of his relatives' DNA, was around 440.

Finally his body was exhumed, and a direct match obtained. Although the defence attempted to attribute this to contamination, it is generally agreed that the case is now closed.

7.2 Disputed paternity

Problems of disputed paternity necessarily resolve around indirect "matching" of the DNA of the putative father with the true father. When profiles from mother, child and putative father are available, the likelihood ratio in favour of paternity can be calculated by standard formulae. When the putative father's profile is unavailable, profiles may be obtained from his relatives: for example, two full brothers, and an undisputed child and its (different) mother. Although the logical steps in calculating the likelihood ratio are clear in principle (though not always so to the forensic and other experts directly involved in such work), the computational difficulties of implementing them can be severe.

8 Bayesian networks

The technology of Bayesian networks [3] has proved valuable in addressing complex problems of DNA interpretation: Figure 1 shows a graphical representation of the paternity case described in § 7.2: see [14] for further details.
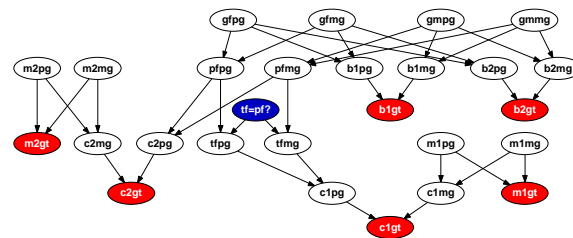

Figure 1: Bayes net representation of a complex paternity case

This technology is also being applied to problems such as mixed stains [21], mutation [13, 7], contamination and laboratory errors; and, more generally, as a decision aid for representing and manipulating mixed masses of evidence [9].

9 Conclusions

Seemingly straightforward problems of legal reasoning can quickly lead to complexity, controversy and confusion: the above examples

are just a few amongst many. The whole field forms a particularly challenging testbed for the whole range of ideas and dialogues about reasoning under uncertainty.

Acknowledgements

References

[1] D. J. Balding and P. J. Donnelly. Inference in forensic identification (with Discussion). *Journal of the Royal Statistical Society, Series A*, 158:21–53, 1995.

[2] D. J. Balding and P. J. Donnelly. DNA profile evidence when the suspect is identified through a database search. *Journal of Forensic Sciences*, 41:603–7, 1996.

[3] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Springer, New York, 1999.

[4] A. P. Dawid. The island problem: Coherent use of identification evidence. In P. R. Freeman and A. F. M. Smith, editors, *Aspects of Uncertainty: A Tribute to D. V. Lindley*, chapter 11, pages 159–170. John Wiley and Sons, Chichester, 1994.

[5] A. P. Dawid. Comment on Stockmarr's "Likelihood ratios for evaluating DNA evidence, when the suspect is found through a database search" (with response by Stockmarr). *Biometrics*, 57:976–980, 2001.

[6] A. P. Dawid. Bayes's theorem and weighing evidence by juries. *Proceedings of the British Academy*, 113:71–90, 2002. (Published in book form as *Bayes's Theorem*, edited by Richard Swinburne).

[7] A. P. Dawid. An object-oriented Bayesian network for estimating mutation rates. In *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, Jan 3-6 2003, Key West, Florida*, (ed. C. M. Bishop and B. J. Frey). `http://tinyurl.com/39bmh`, 2003.

[8] A. P. Dawid. Which likelihood ratio? In discussion of [20]. *Law, Probability and Risk*, 2004. To appear.

[9] A. P. Dawid and I. W. Evett. Using a graphical method to assist the evaluation of complicated patterns of evidence. *Journal of Forensic Sciences*, 42:226–231, 1997.

[10] A. P. Dawid and J. Mortera. In discussion of [1]. *Journal of the Royal Statistical Society, Series A*, 158:46, 1995.

[11] A. P. Dawid and J. Mortera. Coherent analysis of forensic identification evidence. *Journal of the Royal Statistical Society, Series B*, 58:425–443, 1996.

[12] A. P. Dawid and J. Mortera. Forensic identification with imperfect evidence. *Biometrika*, 85:835–849, 1998.

[13] A. P. Dawid, J. Mortera, and V. L. Pascali. Non-fatherhood or mutation? A probabilistic approach to parental exclusion in paternity testing. *Forensic Science International*, 124:55–61, 2001.

[14] A. P. Dawid, J. Mortera, V. L. Pascali, and D. W. van Boxel. Probabilistic expert systems for forensic inference from genetic markers. *Scandinavian Journal of Statistics*, 29:577–595, 2002.

[15] P. J. Donnelly and R. D. Friedman. DNA database searches and the legal consumption of scientific evidence. *Michigan Law Review*, 97:931–984, 1999.

[16] I. W. Evett, L. A. Foreman, and B. S. Weir. Letter to the Editor (with responses by A. Stockmarr and B. Devlin). *Biometrics*, 56:1274–1275, 2000.

[17] W. B. Fairley and F. Mosteller. A conversation about Collins. In W. B. Fairley and F. Mosteller, editors, *Statistics in Public Policy*, pages 369–379. Addison-Wesley, Reading, Massachusetts, 1977.

[18] G. Gigerenzer. *Reckoning with Risk: Learning to Live with Uncertainty*. Allen Lane: The Penguin Press, 2002.

[19] R. W. J. Meester and M. Sjerps. The evidential value in the DNA database search controversy and the two stain problem. *Biometrics*, 59:727–732, 2003.

[20] R. W. J. Meester and M. Sjerps. Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence (with discusion by A. P. Dawid, D. J. Balding, J. S. Buckleton and C. M. Triggs). *Law, Probability and Risk*, 2004. To appear.

[21] J. Mortera, A. P. Dawid, and S. L. Lauritzen. Probabilistic expert systems for DNA mixture profiling. *Theoretical Population Biology*, 63:191–205, 2003.

[22] A. Stockmarr. Likelihood ratios for evaluating DNA evidence when the suspect is found through a database search. *Biometrics*, 55:671–677, 1999.