

Bayes's Theorem And Weighing Evidence by Juries

A. P. Dawid
University College London

August 3, 2001

1 Statistics and the Law

1.1 Evidence

At first sight, there may appear to be little connection between Statistics and Law. On closer inspection it can be seen that the problems they tackle are in many ways identical — although they go about them in very different ways. In a broad sense, each subject can be regarded as concerned with the *Interpretation of Evidence*. I owe my own introduction to the common ground between the two activities to my colleague William Twining, Professor of Jurisprudence at University College London, who has long been interested in probability in the law. In our discussions we quickly came to realise that, for both of us, the principal objective in teaching our students was the same: to train them to be able to interpret a mixed mass of evidence. That contact led to my contributing some lectures on uses and abuses of probability and statistics in the law to the University of London Intercollegiate LIM course on Evidence and Proof (and an Appendix on “Probability and Proof” to Anderson and Twining (1991)), as well as drawing me into related research (Dawid 1987; Dawid 1994; Dawid and Mortera 1996; Dawid and Mortera 1998). To my initial surprise, I found here a rich and stimulating source of problems, simultaneously practical and philosophical, to challenge my logical and analytical problem-solving skills. For general background on some of the issues involved, see Eggleston (1983); Robertson and Vignaux (1995); Aitken (1995); Evett and Weir (1998); Gastwirth (2000).

The current state of legal analysis of evidence seems to me similar to that of science before Galileo, in thrall to the authority of Aristotle and loth to concede the need to break away from old habits of thought. Galileo had the revolutionary idea that scientists should actually look at how the world behaves. It may be equally revolutionary to suggest that lawyers might look at how others have approached the problem of interpretation of evidence, and that they might even have something to learn from them. It is my strong belief (though I do not

expect it to be shared by many lawyers) that statisticians, with their training in logical analysis of the impact of evidence on uncertainty, have much to contribute towards identifying and clarifying many delicate issues in the interpretation of legal evidence. And I believe that by far the greatest clarity and progress will come from the exploitation of the Bayesian philosophy and methodology of statistics. Although other statistical approaches are possible, I consider that these are at best unnecessary and at worst dangerously misleading.

My message is not a new one. Some of the earliest work on Probability (by such great minds as James and Nicholas Bernoulli, Condorcet, Laplace, Poisson, Cournot) was concerned with, or motivated by, problems of quantification and combination of legal evidence and judgement. However, both statisticians and lawyers seem to have lost this thread. I hope I can persuade both communities that it is well worth picking up and following.

2 Testing Between Two Hypotheses — in Court

In a criminal case, each charge on the indictment sheet can be regarded as a hypothesis under test. For simplicity we restrict attention to the case of a single defendant and a single charge against that defendant. The proposition that the defendant is *guilty* of the charge (a proposition which typically combines issues both of fact and of law) is the *Prosecution Hypothesis*; we shall henceforth denote it by G . It is specific and clearly defined — what statisticians would call a ‘simple hypothesis’.

The task of the Defence is to cast doubt on G , or, what is equivalent, to argue for the reasonableness of the contrary proposition \bar{G} , which asserts the falsity of G . We call \bar{G} the *Defence Hypothesis*. It can be very non-specific — a ‘composite hypothesis’, in statistical parlance. The Defence is not obliged to identify or argue in favour of any more specific alternative hypothesis to G , although this will often be appropriate. In that case, \bar{G} might also reduce to a simple hypothesis.

Let us denote the conjunction of one or more items of evidence (perhaps the totality of the evidence) in the case by E . Our task is to use the evidence E to cast light on the two hypotheses, G and \bar{G} , before the court.

2.1 Bayesian Approach

Since we start out with uncertainty about the hypotheses and the evidence, we should try and quantify that uncertainty appropriately. A vast body of thoughtful logical and philosophical analysis of uncertainty has concluded that the only appropriate quantification of uncertainty, of any kind, is in terms of the *Calculus of Probability*. This is the basis of the modern Bayesian approach to statistical inference. An important aspect of this philosophy is that complete objectivity is an illusion, and thus that there is no such thing as ‘the’ probability of any uncertain event — rather, each individual is entitled to his or her own subjective probability. This is not, however, to say that anything goes: in the light

of whatever relevant evidence may be available, certain opinions will be more reasonable than others. In simple statistical problems, it can be shown that differing initial subjective distributions will be brought into ever closer and closer agreement when updated through the incorporation of sufficiently extensive observational evidence. In legal applications the conditions for this convergence may not apply, but even so there will be certain probabilistic ingredients and conclusions that can be regarded as reasonable by all reasonable parties.

From the standpoint of any individual juror, the end-point of his (or her) analysis of the evidence E heard should be his *posterior probability* of guilt given the evidence:

$$P(G | E) \tag{1}$$

(where we use the notation $P(A | B)$ to denote the conditional probability of an uncertain event A , in the light of known or hypothesised information B). This is a direct expression of the juror's remaining uncertainty as to the validity of the Prosecution Hypothesis, after taking into account the evidence E . However, it will not typically be appropriate to assess this directly; rather, it should be constructed out of other, more basic and defensible, primitive ingredients. This can be done using the 'odds' form of *Bayes's theorem*:

$$\frac{P(G | E)}{P(\bar{G} | E)} = \frac{P(G)}{P(\bar{G})} \times \frac{P(E | G)}{P(E | \bar{G})}, \tag{2}$$

or, in words:

$$\textit{Posterior Odds} = \textit{Prior Odds} \times \textit{Likelihood Ratio}.$$

The first term on the right-hand side of (2), the *prior odds*, measures the relative degrees of belief as between the Prosecution and the Defence hypotheses, *before* the evidence E has been incorporated. There may be a range of reasonable values that jurors could hold for this in the light of previous evidence. The second term, the *likelihood ratio*, LR, involves the probabilities accorded to the new evidence E by each of the two competing hypotheses. In some cases (essentially when both G and \bar{G} can be framed as simple hypotheses) these ingredients will be moderately 'objective' and agreed upon, and thus so also will be the likelihood ratio. It might then seem appropriate to present the value of LR, or its constituent probabilities, as a summary of the impact of the evidence, leaving individual jurors to combine it with their own prior beliefs, using equation (2). Once the posterior odds, say Ω , has been calculated, the desired posterior probability of guilt, $P(G | E)$, is just $\Omega / (1 + \Omega)$. (When one or both of the hypotheses is composite, more complex calculations may be required, incorporating prior assessments even into the calculation of the likelihood ratio).

Simple though it is, both the logic and the calculation involved in equation (2) will be beyond most judges (and even some jurors). It could then be helpful to present the impact of the evidence by means of a table, incorporating the relevant value of LR, and showing how any hypothetical prior probability of guilt, $P(G)$, would be updated by the evidence to the posterior probability

$P(G|E)$. If the resulting posterior probability is sufficiently high (and it would be up to the Judge to advise on what this means), then the verdict “Guilty” would be appropriate. Table 1 shows the impact of a likelihood ratio value of 100 (*i.e.* the probability of the evidence E is 100 times greater under the Prosecution hypothesis than it is under the Defence hypothesis). If it is appropriate to regard satisfaction as to guilt “beyond reasonable doubt” as attained when the posterior probability $P(G|E)$ exceeds 99%, then we see that a juror should be willing to convict whenever his prior probability, before incorporating E , exceeded 50%. (Note, however, that when E is the only evidence in the case, before E is admitted the suspect should be treated no differently from any other member of the population, so that a prior probability of even 0.001 could be regarded as unreasonably high.)

Prior Prob.	0.001	0.01	0.1	0.3	0.5	0.7	0.9
Post. Prob.	0.09	0.50	0.92	0.98	0.99	0.996	0.999

Table 1: The impact of a likelihood ratio $LR = 100$

The application of equation (2) can be done either ‘wholesale’, with E representing the totality of the evidence in the case; or ‘piecemeal’, with sequential incorporation of several items of evidence. In the latter case, the probabilities on the right-hand side should be regarded as evaluated conditional on all previously incorporated evidence. This evaluation might have been done formally, using previous applications of (2) to account for earlier items of evidence, or more informally, if the previous evidence seems too ‘soft’ to justify that. One has to be particularly aware of the possibility that this incorporation of previous evidence might alter the probabilities constituting the likelihood ratio. This will happen if, conditional on the relevant hypothesis, the new item E is not independent of previous evidence. But quite frequently this independence assumption is justified, and then the LR value is unaffected by the previously incorporated evidence.

2.2 Neyman-Pearson and Bayes

The Bayesian approach to statistical inference has been the subject of controversy for at least 150 years, at times almost being consigned to oblivion, and at others (increasingly so in modern times) being strongly in the ascendant. Other schools of statistical inference have grown up which purport to banish subjectivity by using probability in more indirect ways. The most successful of these is the Neyman-Pearson approach, which assesses and selects decision rules by examining their probabilistic performance under various hypotheses. We introduce and examine this in the context of a real court case with which the author was involved.

2.2.1 Regina v Sally Clark

Sally Clark's first child Christopher died unexpectedly at the age of about three months, when Sally was the only person in the house. The death was initially treated as a case of Sudden Infant Death Syndrome (SIDS, or 'Cot Death'). Sally's second child Harry was born the following year. Harry died in similar circumstances at the age of two months. Soon after, Sally was arrested and tried for murdering both children. There was some medical evidence suggesting smothering, although this could also be explained as arising from attempts at resuscitation. For the moment we disregard the medical evidence, but return to this at § 2.4 below.

At trial, a professor of paediatrics testified that the probability that, in a family like Sally's, two babies would both die of SIDS was around 1 in 73 million. This was based on a professionally executed study which estimated the probability of a single SIDS death in such a family at one in 8500 — which figure was then squared by the witness, to account for two deaths. Although this calculation is extremely dubious, being based on unrealistic assumptions of independence, I will accept it here, simply for the sake of argument.

Sally Clark was convicted of murder. Although we cannot know how the jury regarded the statistical evidence, it is reasonable to speculate that it was strongly influenced by the extremely small probability value of 1 in 73 million that both deaths could have arisen from SIDS, regarding this as ruling out the possibility of death by natural causes.

2.2.2 Neyman-Pearson approach

Let us see how a Neyman-Pearson statistician might go about making inference from the evidence. For simplicity, suppose that the only possible causes of infant death are SIDS and murder, and discount the possibility that these might both occur in the same family. Consider the decision rule: "If two babies in a family both die of unexplained causes, decide that their mother murdered them". Under the hypothesis that a mother does indeed murder two of her babies, the probability of making an error using this rule is 0. Under the contrary hypothesis that she does not, an error will be made if and only if they both die of SIDS, which will occur with probability 1 in 73 million. The rule appears amazingly accurate, under either hypothesis, and thus seems a good one to follow. And, if we apply it to the case of Sally Clark, we must decide that she murdered her babies.

2.2.3 Criticism

Even without looking at the above argument through Bayesian spectacles, it does not need much thought to realise that there is something missing. We are considering two hypotheses: death by SIDS, and death by murder. The probability of the former plays a major role in the above Neyman-Pearson argument. But should not the probability of the latter be just as relevant? Office of National Statistics data for 1997 show that, out of a total of 642093 live births,

seven babies were murdered in the first year of life. This yields an estimate of the probability of being murdered, for one child, of 1.1×10^{-5} . If we were to treat this in the same (admittedly dubious) way as the corresponding probability for SIDS, and square it to account for the two deaths, we could argue that the probability that Sally Clark's two babies were both murdered is about 1 in 8.4 billion (we here use the American billion, *i.e.* one thousand million) — and if this were the argument presented to the jury, how could it but conclude that the possibility of murder is so improbable as to be inconceivable?

Once this is pointed out, it should be clear that the figure quoted of 1 in 73 million for two deaths due to SIDS cannot be significant in itself, but only in relation to the 'counterbalancing probability' of two deaths due to murder.

2.3 Two Bayesian arguments

Only the Bayesian approach can be relied upon to take automatic correct account of all the evidence. However, there can be more than one way of formulating the problem for Bayesian analysis. Here we consider two such formulations, and verify that they yield the same answer.

2.3.1 Restricted hypotheses

Taking into account that the two babies died of something, and that we only consider SIDS and murder as possible causes of death, we can formulate two hypotheses, exactly one of which must be the case:

M: Sally Clark murdered her babies.

S: Sally Clark's babies died of SIDS.

The evidence *E* before the court is just the fact that the babies died.

By Bayes's Theorem, the posterior odds for comparing the two hypotheses is given by:

$$\frac{P(M|E)}{P(S|E)} = \frac{P(M)}{P(S)} \times \frac{P(E|M)}{P(E|S)}. \quad (3)$$

The first term on the right-hand side is just the ratio of the overall probabilities of two deaths, under the two hypotheses. Using the hypothetical figures above, this is (1/8.4 billion)/(1/73 million), or about 0.009. As for the second, likelihood ratio, term, *E* is certain to occur under either hypothesis, so that this is just unity, and the posterior odds is again $0.009 = 9/1000$ — or over 100:1 against. We conclude that it is over 100 times more probable that the two babies died of SIDS than that they were murdered.

2.3.2 Unrestricted hypotheses

It could be objected that the evidence *E* has been used twice in the above calculation: once to delimit the hypotheses, and again to form a likelihood ratio. Even though the latter use was completely ineffective, since the likelihood ratio

was unity, there might still be a concern because the hypotheses being compared were only formulated in the light of the evidence: in particular, hypotheses which do not lead to the babies' deaths should perhaps not have been excluded, prior to taking the evidence of those deaths into account. To verify that this use of the evidence to refine the hypotheses is not in fact of any importance, we conduct an alternative analysis, in which no such refinement takes place. That is, we now specify the hypotheses as G , that Sally Clark murdered her babies, and \bar{G} , that she did not. Then G is identical with M , but \bar{G} is not the same as S , since it does not assume that the babies will die at all — it is just the negation of G .

Once again, Bayes's Theorem, as given by (2), can be applied. We have $P(E|G) = 1$, while $P(E|\bar{G})$ is 1/73 million. The likelihood ratio in favour of guilt is thus 73 million — a seemingly overwhelming weight of evidence in favour of murder. However, we must not forget the first term on the right of (2), the prior odds.

In fact we have $P(G) = P(M) = 1/8.4$ billion; so the prior odds, $P(G)/P(\bar{G})$, is also essentially 1/8.4 billion. Combining the two factors we obtain once again $(1/8.4 \text{ billion})/(1/73 \text{ million})$, or 0.009, as the posterior odds on guilt.

It is thus entirely immaterial which of the above two Bayesian arguments is followed. My personal preference in this particular instance is for the former, because it deals symmetrically with the two competing hypotheses about the deaths, thus directly addressing the criticism voiced in §2.2.3. However, this preference is of no importance so long as the correct Bayesian calculations are performed; and, as we shall see, the latter line of argument is a more natural approach in many problems.

2.4 Additional evidence

In the case of Sally Clark, there was additional medical evidence, including haemorrhages to the children's brains and eyes. Once again, so long as the laws of probability are followed correctly, it does not matter whether this evidence is incorporated before or after the analysis based purely on the fact of death. In this case it is more straightforward to bring it in at the end, as a new likelihood ratio term. Thus suppose (purely for illustration) that the probability of observing the specific medical signs in the case, on the hypothesis that the babies were murdered, is assessed at 1 in 20; while the probability of the same signs being observed on the hypothesis that they died of SIDS can be taken as 1 in 100. The effect of the medical evidence is then to multiply the posterior odds on guilt resulting from the previous analysis by 5. Combined with the hypothetical figures used before, which gave posterior odds of 0.009 based just on the fact of death, this would yield updated posterior odds of 0.043, corresponding to a posterior probability of 4.2%. While the choice of specific numbers here is clearly subject to some vagueness, the Bayesian approach does make it very clear exactly what features of the evidence must be taken into account. In particular, it could be quite misleading to suggest simply that the medical signs observed are 'consistent with' the babies having been smothered. What is

required is a quantitative estimate of how many times more likely that evidence would be under the murder hypothesis than it would be under the hypothesis of accidental death followed by attempts at resuscitation.

3 Identification evidence

In many criminal and civil cases, identification of a suspect as the actual perpetrator of an offence is based on an alleged ‘match’ between trace evidence taken from the scene and a sample taken from the suspect. A fingerprint on the murder weapon is compared with that of the suspect; a footprint found at the scene is compared with shoes belonging to the suspect; the refractive index of fragments of glass found on the suspect’s clothing is compared with that of the broken window at the scene of a burglary; fibres left at the scene are compared, by eye, microspectro-fluorimetry, and thin layer chromatography, with others taken from the suspect’s jumper. Increasingly, especially in cases of murder, rape and alleged paternity, the match is based on DNA profiles, and may well be the only evidence presented.

In the following we assume that the fact that a crime has been committed is not in dispute — the only issue is the identity of the culprit.

A forensic scientist will typically summarise the strength of potentially incriminating identification evidence in terms of an associated *match probability*: an assessment of the probability that the observed match would arise if the suspect were in fact innocent. Match probability values of the order of one in a billion are now routine for DNA profiling. Intuitively, the smaller the match probability, the stronger the evidence against the suspect; and very tiny probabilities appear to be incontrovertibly convincing evidence. However, there are many subtleties and pitfalls in the interpretation of such forensic identification evidence, which are all too often not appreciated.

A typical problem can be formulated as follows. Let M denote the evidence as to the match, and B any other (background) evidence in the case. The full evidence is thus $E = M \& B$. Typically, though by no means universally, it is appropriate assume that, if the suspect were indeed guilty, then the two samples would be bound to match, and we shall assume this. That is, we take

$$P(M | G) = 1. \tag{4}$$

The match probability measures how probable the match would be if the suspect were innocent: that is, $P(M | \bar{G})$. Typically this is calculated on the hypothesis that, in this case, the true perpetrator would be entirely unrelated to the suspect. However, this could be misleading if there were any possibility that the suspect and the perpetrator might be related, either closely or distantly. Taking such a possibility into account can greatly affect the value of a DNA profile match probability.

For the sake of illustration, we shall proceed on the assumption that the match probability is one in 10 million:

$$P(M | \bar{G}) = 10^{-7}. \tag{5}$$

3.1 Speeches in Court

3.1.1 The Prosecution argument

Counsel for the Prosecution argues as follows:

“Ladies and gentlemen of the jury, the probability of the observed match between the sample at the scene of the crime and that of the suspect having arisen by innocent means is one in 10 million. This is an entirely negligible probability, and we must therefore conclude that, with a probability overwhelmingly close to 1, the suspect is guilty. You have no alternative but to convict.”

3.1.2 The Defence argument

Counsel for the Defence argues as follows:

“In the general population, there are about 30 million individuals who might possibly have committed this crime. One of these is the true culprit. Of the remaining 30 million innocent individuals, each has a probability of one in 10 million of providing a match to the crimes sample. We should therefore expect there to be 3 innocent individuals who match. We know that the suspect matches, but he could still be any one of the four matching individuals, of which 1 is guilty and three are innocent. So the probability that he is guilty is only one quarter. This falls far short of the criminal criterion of ‘beyond reasonable doubt’. You must acquit.”

3.2 The Bayesian argument

Whatever one thinks of the above two arguments, neither can be generally appropriate, since neither makes any allowance for the incorporation of the other evidence (B) in the case. However, this can readily be done by means of the correct Bayesian argument. We now present this, and then return to reconsider the above Prosecution and Defence arguments.

We start from the position that the other evidence B has already been presented, so that all the probabilities we consider are already conditional on B . We then apply Bayes’s theorem to incorporate, in addition, the match evidence M , so obtaining posterior probabilities given the totality of the evidence, $E = M \& B$. With this understanding, (2) becomes:

$$\begin{aligned} \frac{P(G|E)}{P(\bar{G}|E)} &= \frac{P(G|M \& B)}{P(\bar{G}|M \& B)} \\ &= \frac{P(G|B)}{P(\bar{G}|B)} \times \frac{P(M|G \& B)}{P(M|\bar{G} \& B)}. \end{aligned} \tag{6}$$

We previously assumed (4) that, if the suspect is guilty, then he will certainly provide a match — and this property can not be affected by learning any further background information. That is, $P(M|G \& B) = 1$.

It will likewise normally be appropriate (particularly for DNA evidence) to disregard the background information when assessing the probability that the suspect would match were he in fact innocent. This will hold when, conditional on innocence, the background evidence and the match identification evidence can be regarded as arising independently. Assuming this in the case at hand, from (5) we have $P(M | \bar{G} \& B) = P(M | \bar{G}) = 10^{-7}$.

The likelihood ratio for the identification evidence, $P(M | G \& B) / P(M | \bar{G} \& B)$, is thus 10^7 , unaffected by the other evidence B .

The other term needed for (6) is the prior odds, $P(G | B) / P(\bar{G} | B)$. But, while this is indeed prior to the introduction of the identification evidence, it is posterior to the background evidence B . The initial incorporation of that evidence might have been done formally, by means of an earlier application of Bayes's theorem, or informally. But in any event it needs to be taken into account in making a realistic assessment of the relevant prior odds. Whatever such assessment is made, the final odds on guilt is obtained by multiplying this prior odds by 10^7 .

3.3 Reconsideration of Counsels' arguments

We now look more closely at Defence Counsel's argument.

Let us accept that the appropriate size of the 'catchment area' of potential perpetrators is indeed 30 million. We further suppose that, apart from the DNA evidence, there is no additional background evidence relating to the suspect, other than that he belongs to this catchment population. Then it would be appropriate to take, as the prior probability of guilt, $P(G | B) = 1/30$ million, and the prior odds would be essentially the same — certainly such a judgement would seem to be in the spirit of the 'presumption of innocence'. Applying Bayes's theorem (6), the posterior odds on guilt are thus $(1/30 \text{ million}) \times (10 \text{ million}) = 1/3$, corresponding to a posterior probability of one quarter. That is to say, the Defence argument agrees with a Bayesian analysis in which the only background evidence used is the size of the population of potential perpetrators. It is implicit that, before the identification evidence is taken into account, anyone in this population is as likely as any other to be the true culprit. In cases where the identification evidence is the sole evidence presented, the Defence argument is thus essentially valid; however, it is not generally appropriate to use it as it stands when there is additional evidence in the case.

In a case of 'naked identification evidence' the story-line of the Defence argument may well be more appealing than a dry application of Bayes's theorem, and would deliver the correct inference. However, one problem with reframing the Bayesian argument in this way is that, when the match probability is sufficiently tiny, the 'expected number of innocent matches' can be much less than 1. If, for example, the match probability in the case above had been one in one billion (as is now fairly routine in DNA profiling), rather than one in 10 million, we would expect only 0.03 innocent matches in a population of size 30 million. Applied formally, the Defence argument still applies: of those matching, there is one who is guilty and 0.03 who are innocent, and so the posterior probability

of guilt is $1/1.03 = 0.971$, exactly as would emerge from the Bayesian calculation. But when it requires them to juggle with small fractions of a person, the intuitive appeal of the Defence argument to the jury may well disappear.

We now turn to reconsider the Prosecution argument. This is often termed the ‘Prosecutor’s Fallacy’, and indeed as presented it does turn on a serious misrepresentation. This is not usually deliberate, but rather a consequence of the difficulty of expressing statements of conditional probability clearly and unambiguously in English, which sometimes seems to have been deliberately constructed to facilitate probabilistic misunderstanding (and I am sure that English is no different from any other natural language in this respect).

The match probability of 1 in 10 million is a measure of the probability, $P(M | \bar{G})$, of obtaining the match, on the hypothesis that the suspect is innocent. (Note that this is just the kind of conditional probability, for some observed event conditional on one or more entertained hypotheses, which underlies the Neyman-Pearson logic criticised in § 2.2 above.) However, the Prosecutor misinterprets this as $P(\bar{G} | M)$, the probability that the suspect is innocent, on the evidence of the match. This common and seductive error is also known as ‘transposing the conditional’. In general, there is absolutely no reason for the two conditional probabilities, $P(M | \bar{G})$ and $P(\bar{G} | M)$, to be similar. The actual connection between them is governed by Bayes’s theorem, which also requires other input, namely the prior probability of guilt, $P(G)$.

It is not hard to show that the transposition involved in the Prosecution argument *will* lead to an (approximately) valid result only in the special case that $P(G | B) = \frac{1}{2}$ — *i.e.* when, on the basis of other evidence in the case, it can be regarded as equally probable that the suspect is or is not guilty. There may occasionally be cases where this is a reasonable assumption, but it is far from being so automatically. Contrast it with the implicit assumption underlying the Defence argument — that the prior probability that the suspect is guilty is one in 30 million (or whatever the appropriate value for the population of possible perpetrators may be).

In summary, seen through Bayesian spectacles, both the Prosecution and the Defence arguments are generally inappropriate. Each becomes reasonable under certain specific assumptions about the prior probability of guilt, but these implicit assumptions are radically different in the two cases — which in turn explains why their answers are so completely different. Only the complete Bayesian argument is flexible enough to allow incorporation of whatever might be a reasonable prior judgment as to guilt, based on the other background evidence in the case.

3.4 Regina v Denis John Adams

The case of Denis John Adams illustrates both the Bayesian approach to reasoning about identification evidence in the presence of other evidence, and some of the pitfalls besetting presentation of this approach in court.

Adams was arrested for rape. The only evidence linking him to the crime, other than the fact that he lived in the local area, was a match between his DNA and that of semen obtained from the victim. The relevant match probability was said to be one in 200 million, although the Defence challenged this, suggesting that a figure of one in 20 million or even one in 2 million could be more appropriate.

All other evidence in the case was in favour of the defendant. The victim did not pick Adams out at an identification parade, and had said that he did not resemble her attacker. Adams was 37 and looked older; the victim claimed the rapist was in his early twenties. Furthermore, Adams's girlfriend testified that he had spent the night of the attack with her, and this alibi remained unchallenged.

At trial, with the consent of both sides and the Court, the jury was given instruction in the correct (Bayesian) way to combine all the evidence, introducing in turn (i) the prior probability, ahead of all specific evidence; (ii) likelihood ratios engendered by the defence evidence; and finally (iii) the identification evidence. Attention was drawn to the relevant probability questions to address, and the jurors were asked to assess their own probability values for these. The jury was guided through a practice calculation; however, care was taken at all times not to propose any specific probability values (other than match probabilities) to the jury. Below I shall insert specific numbers, but these are entirely hypothetical, and used merely to illustrate the general shape of the argument as it might be conducted by a juror following the instructions given.

It was indicated that there were approximately 150,000 males aged between 18 and 60 in the local area who, absent any other evidence, might have committed the crime. In order to allow for some possibility (assessed at something like a 25% chance) that the attacker came from outside the area, a prior probability of guilt of the order of one in 200,000 might be considered reasonable. (The prior odds would then be essentially the same.)

With regard to the evidence that the victim did not recognise Adams as her attacker, one might assess the conditional probability of this happening, if Adams were truly guilty, at around 10%; and its conditional probability, were he innocent, at around 90%. On forming the required ratio of these figures, a likelihood ratio of $1/9$ is obtained. As for the alibi evidence, one could assess this might be proffered with probability 25% if he were guilty, as against 50% if innocent, leading to a likelihood ratio of $1/2$. If we assume that the two items of defence evidence would arise independently, given either guilt or innocence, then we can multiply them together to obtain an overall 'defence likelihood ratio' of $1/18$.

Applying Bayes's Theorem to combine the prior odds of one in 200,000 and the above likelihood ratio of $1/18$, the odds in favour of guilt, after taking into account the defence evidence but before incorporating the DNA evidence, would be assessed at one in 3.6 million.

Now the DNA match evidence by itself provides a likelihood ratio of between 200 million and 2 million in favour of guilt. Overwhelming though this may seem, it has to be taken in conjunction with the counterbalancing defence

evidence, by applying it (using Bayes's theorem) to the previously calculated odds of one in 3.6 million. When this is done, we find that the posterior probability of guilt, given the totality of the evidence, varies from 0.98 (using a match probability of one in 200 million) to 0.36 (using one in 2 million). The defence argued that, in the light of all the evidence, Adams's guilt had not been established beyond a reasonable doubt.

We cannot know what went on in the jury room, but the jury returned a verdict of guilty. The case then went to appeal. The Appeal Court roundly rejected the attempt to school the jury in the rational analysis of probabilistic evidence, saying that "it trespasses on an area peculiarly and exclusively within the province of the jury", and that "to introduce Bayes's theorem, or any similar method, into a criminal trial plunges the jury into inappropriate and unnecessary realms of theory and complexity". The task of the jury was said to be to "evaluate evidence and reach a conclusion not by means of a formula, mathematical or otherwise, but by the joint application of their individual common sense and knowledge of the world to the evidence before them". While one may well applaud this restatement of the traditional rôle of the jury, it fails to address the problem that common sense usually fares extremely badly when it comes to manipulating probabilities, and in particular common experience simply does not encompass such tiny probabilities as arise with DNA match evidence.

The appeal was granted on the basis that the trial judge had not adequately dealt with the question of what the jury should do if they did not want to use Bayes's theorem. A retrial was ordered. Once again attempts were made to describe the Bayesian approach to the integration of all the evidence, once again the jury convicted, once again the case went to appeal and once again the Bayesian approach was rejected as inappropriate to the courtroom — although this time the appeal was dismissed.

At the time of writing, it seems that Bayesian arguments, while not formally banned from court, need to be presented with a good deal of circumspection. One possible approach (recommended by the Court of Appeal in the case of *R. v. Doheny*) uses a variation on the Defence argument presented in §3.1.2. In that argument, the defendant was identified as one of four matching individuals, just one of whom is guilty. Absent other evidence, we can regard all four as equally likely to be the guilty party, leading to a probability of 1 in 4 that it is the defendant. However, we could go on to take into account other specific evidence, for or against the defendant, so generating differing probabilities of guilt for these four. This could be done formally, using Bayes's theorem to combine a suitable likelihood ratio based on the other evidence with a 'prior' odds of 1/3, or informally by the application of 'common sense'. Thus in the Adams case (taking a base population of 200,000 and a match probability of one in 2 million), the number of innocent individuals matching the DNA would be expected to be 1/10. Taking into account the likelihood ratio of 1/18 from the other evidence means we must count Adams himself as 1/18 of an individual. The probability that it is Adams, rather than anyone else, who is the guilty party is thus calculated as $(1/18)/\{(1/18) + (1/10)\} = 0.36$, as before. But it

must be admitted that such juggling with parts of people might be distasteful to the jury!

There is an additional wrinkle to the Adams story. At appeal, the Defence pointed out that Denis John Adams had a full brother, whose DNA had not been investigated. The probability that his brother had the same DNA profile as he did was calculated as one in 220, and it was submitted that this weakened the impact of DNA evidence against Denis John Adams. The Appeal Court dismissed this point on the grounds that there was no evidence that the brother might have committed the offence — ignoring the fact that, in the absence of the DNA match, neither was there any such evidence against Denis John Adams. The Bayesian argument, in its original or variant forms, does allow one to account for different match probabilities due to genetic relatedness. In effect, the brother adds an additional 1/120 of a person to the pool (previously 1/10) of matching individuals. This makes little numerical difference to the calculation as conducted above, but would have a much larger effect if the match probability for unrelated individuals had been taken as one in 200 million, so yielding 1/1000 as the number of unrelated matching individuals. Also, taking into account the existence of a number of individuals (unknown as well as known) related to Denis John Adams to various degrees could further add to the pool, perhaps boosting the figure of 1/120 substantially.

4 Databases and search

The police now have computer databases of tens of thousands of DNA profiles, and these look set to get much larger still. Increasingly, in crimes such as rape or murder where there is no obvious suspect, such a database is ‘trawled’ to see if it contains a profile matching that found at the scene. If it does, the matching individual may be arrested and charged, even in the absence of any other evidence linking him to the crime.

Because a jury might be unfairly prejudiced by the information that a defendant’s profile had previously been entered into a police database, the fact that he was identified in this way would typically not be admissible as evidence in court. Nevertheless, for the sake of rational analysis, it is reasonable to ask how — if at all — the fact of the database search should affect the impact of the DNA match. Once again we can identify two mutually contradictory lines of argument.

4.1 The Prosecution line

Prosecuting Counsel argues:

“A database of 10,000 profiles has been searched, and only the defendant’s profile has been found to match that found at the scene. We have thus eliminated 9,999 potential alternative suspects. This must make it more likely that the defendant is guilty.”

According to this argument, the evidence for the defendant's guilt is stronger (although, given the initial extremely large number of potential suspects, perhaps only marginally) than it would have been had he been identified without searching the database. It also gives some comfort that the usual practice of hiding the database trawl from the jury is of little consequence, and is if anything conservative, erring on the side of the defendant.

4.2 The Defence line

Counsel for the Defence argues:

“The DNA profile found at the scene of the crime occurs in the population with a frequency of one in one million. The police database contains 10,000 profiles. The probability that, on searching the database, a match will be found is thus $10,000 \times 1/1,000,000 = 1/100$. This figure, rather than one in one million, is the relevant match probability. Clearly it is not nearly small enough to be regarded as convincing evidence against my client.”

From this point of view, the effect of the database search is to weaken, very dramatically, the strength of the evidence against the defendant.

4.3 Disagreeing to disagree

The comparative merits of the above two arguments have been fiercely debated in the statistical community, with those of a Neyman-Pearson frame of mind favouring the Defence line (National Research Council 1996; Stockmarr 1999), and Bayesians arguing in favour of the Prosecution line, which indeed can be rephrased as a Bayesian probability calculation (Dawid and Mortera 1996; Balding and Donnelly 1996; Donnelly and Friedman 1999; Dawid 2001). Rarely can there have been such an important application of statistics in which the differing intuitions and approaches of the two schools lead to answers so vividly and violently opposed. Defence Counsel, and Neyman-Pearsonites, claim that the upward adjustment of the match probability is essential, to allow for the fact that the hypothesis under test (that the specific matching individual identified by the search did indeed commit the crime) was not known in advance, but only formulated in the light of the data examined. Prosecuting Counsel, and Bayesians, counter-claim that this is irrelevant, and that it is the singular guilt of the actual defendant that is at issue, not the whole database that is on trial. One argument that seems particularly telling to me is to consider what happens when we take matters to extremes. Thus suppose that (as may indeed soon become the case) the police database is fully comprehensive, containing the DNA profiles of all members of the population. And suppose that exactly one of these is found to match the DNA profile from the scene of the crime. On purely logical (and commonsense) grounds, we then know for sure that we must have identified the true culprit — and this is exactly the import of the Prosecution line. But similarly to take the Defence line to extremes would be to regard

the evidence against the defendant as weakened to the point of non-existence, which is clearly absurd.

Statisticians on both sides of this divide (and here I speak as one firmly planted on the Prosecution/Bayesian side) can only despair of the inability of those on the other to appreciate the strengths of the ‘correct’ arguments, and the weaknesses of the ‘incorrect’ arguments. An important moral is that, while all agree that arguments phrased in terms of probabilities are essential to such rational argument, there are various different ways of formulating such arguments — and it does matter how this is done. I personally find that looking at the world through Bayesian spectacles gives a crystal clear image, and clearly brings into focus many issues that would otherwise remain fuzzy and confused.

5 Concluding Remarks

The last message is the one that I would like the reader to take home. Bayesian Statistics is just the logic of rational inference in the presence of uncertainty. It is a valuable intellectual resource, bringing clarity to the formulation and analysis of many perplexing problems. I believe that it could be of far greater significance in the Law than has thus far been allowed or appreciated.

I conclude by briefly mentioning some other problems of legal reasoning that have been greatly clarified by examining them from a Bayesian perspective.

5.1 Genetic heterogeneity

There has been considerable technical discussion about how to allow, for purposes of DNA identification, for the fact that the genetic structure of the population is heterogeneous, so that a given DNA profile may generate different ‘match probabilities’, depending on the subpopulation that is used to determine the frequencies of its constituent alleles (National Research Council 1992, Chapter 3). And again there has been disagreement and confusion about appropriate ways to handle this problem (Foreman *et al.* 1997; Roeder *et al.* 1998). A fully Bayesian approach (Dawid 1998; Dawid and Pueschel 1999) reveals that at least some of the seeming disagreements are merely semantic, while identifying other inadequacies in the arguments presented.

5.2 Combining evidence

The Bayesian machinery is ideally suited to the modelling and analysis of complex inter-relations between many variables. Recent years have seen the development of powerful computational systems based on natural graphical representations (Cowell *et al.* 1999) — an enterprise that can be seen as developing on the pioneering work of Wigmore (1931) (see in particular Kadane and Schum (1996)). Dawid and Evett (1997) show how these ideas can be used to organise and implement the combination of evidence from a variety of sources, such as forensic evidence obtained from fibre analysis and from bloodstains.

5.3 Missing data

Particularly challenging problems arise when DNA evidence is not available on the principal suspect, but indirect evidence can be obtained by typing close relatives. This arises quite commonly in paternity suits, and occasionally in criminal cases, but hitherto there has been no clear understanding of how to analyse such problems. By building a suitable graphical computer model, it is possible readily to calculate the appropriate likelihood ratio, taking correct account of the information that is actually available (Dawid *et al.* 2001).

Acknowledgements

I am grateful to Peter Donnelly and Julia Mortera for comments on an earlier draft of this article.

References

- Aitken, C. G. G. (1995). *Statistics and the Evaluation of Evidence for Forensic Scientists*. John Wiley and Sons, Chichester.
- Anderson, T. and Twining, W. (1991). *Analysis of Evidence*. Weidenfeld and Nicolson, London.
- Balding, D. J. and Donnelly, P. J. (1996). DNA profile evidence when the suspect is identified through a database search. *J. Forensic Sci.*, **41**, 603–7.
- Cowell, R. G., Dawid, A. P., Lauritzen, S., and Spiegelhalter, D. J. (1999). *Probabilistic Networks and Expert Systems*. Springer, New York.
- Dawid, A. P. (1987). The difficulty about conjunction. *The Statistician*, **36**, 91–7.
- Dawid, A. P. (1994). The island problem: Coherent use of identification evidence. In *Aspects of Uncertainty: A Tribute to D. V. Lindley*, (ed. P. R. Freeman and A. F. M. Smith), chapter 11, pp. 159–70. John Wiley and Sons, Chichester.
- Dawid, A. P. (1998). Modelling issues in forensic inference. In *1997 ASA Proceedings, Section on Bayesian Statistics*, pp. 182–6.
- Dawid, A. P. (2001). Comment on Stockmarr’s “Likelihood ratios for evaluating DNA evidence, when the suspect is found through a database search” (with response by Stockmarr). *Biometrics*. In Press.
- Dawid, A. P. and Evett, I. W. (1997). Using a graphical method to assist the evaluation of complicated patterns of evidence. *J. Forensic Sciences*, **42**, 226–31.
- Dawid, A. P. and Mortera, J. (1996). Coherent analysis of forensic identification evidence. *Journal of the Royal Statistical Society, Series B*, **58**, 425–43.
- Dawid, A. P. and Mortera, J. (1998). Forensic identification with imperfect evidence. *Biometrika*, **85**, 835–49.

- Dawid, A. P., Mortera, J., Pascali, V. L., and van Boxel, D. W. (2001). Probabilistic expert systems for forensic inference from genetic markers. *Scandinavian Journal of Statistics*. In Press.
- Dawid, A. P. and Pueschel, J. (1999). Hierarchical models for DNA profiling using heterogeneous databases. In *Bayesian Statistics 6*, (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith). Oxford University Press, Oxford. In Press.
- Donnelly, P. and Friedman, R. D. (1999). DNA database searches and the legal consumption of scientific evidence. *Michigan Law Review*, **97**, 931–84.
- Eggleston, R. (1983). *Evidence, Proof and Probability*, (second edn). Weidenfeld and Nicolson, London.
- Evett, I. W. and Weir, B. S. (1998). *Interpreting DNA Evidence*. Sinauer, Sunderland, MA.
- Foreman, L. A., Smith, A. F. M., and Evett, I. W. (1997). Bayesian analysis of deoxyribonucleic acid profiling data in forensic identification applications (with Discussion). *Journal of the Royal Statistical Society, Series A*, **160**, 429–69.
- Gastwirth, J. L. (ed.) (2000). *Statistical Science in the Courtroom*. Springer-Verlag, New York.
- Kadane, J. B. and Schum, D. A. (1996). *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*. John Wiley and Sons, New York.
- National Research Council (1992). *DNA Technology in Forensic Science*. Washington, D. C.: National Academy Press.
- National Research Council (1996). *The Evaluation of Forensic DNA Evidence*. Washington, D. C.: National Academy Press.
- Robertson, B. and Vignaux, G. A. (1995). *Interpreting Evidence*. John Wiley and Sons, Chichester.
- Roeder, K., Escobar, M., Kadane, J. B., and Balazs, I. (1998). Measuring heterogeneity in forensic databases using hierarchical Bayes models. *Biometrika*, **85**, 269–87.
- Stockmarr, A. (1999). Likelihood ratios for evaluating DNA evidence when the suspect is found through a database search. *Biometrics*, **55**, 671–7.
- Wigmore, J. H. (1931). *The Science of Judicial Proof*, (second edn). Little, Brown, Boston.