24

PROBABILITY AND EVIDENCE

JULIA MORTERA AND PHILIP DAWID

INTRODUCTION

Legal applications of probabilistic and statistical reasoning have a long history, having exercised pioneers such as Nicolas Bernoulli, Condorcet, Laplace, Poisson, and Cournot (Zabell, 1988). After a period of neglect, interest has resurfaced in recent years, and the topic has given rise to many challenging problems.

Evidence presented in a case at law can be regarded as data and the issue to be decided by the court as a hypothesis under test. The relationship between these may be immediate, or else indirect, involving a long chain or tangled web of intermediate propositions. In any case, there will almost always be some uncertainty about the ultimate issue, the evidence, and the way in which these are related. Such uncertainty can, in principle at least, be described probabilistically. In a legal setting, where it is understood that different "reasonable men" (and women) can reasonably hold a range of opinions, it is natural to take a subjective interpretation of probability, regarding it as a measure of a specific individual's uncertainty about a specific (not necessarily repeatable) event in the light of specified information. This interpretation should be borne in mind in the sequel. In particular, under it there is no obstacle to assigning a noncategorical probability to the guilt of the suspect in the light of evidence presented while at the same time believing that it has a definite, though currently unknown, truth value that might even be revealed if only enough evidence were available.

We do not suggest that judges and juries are likely to have (or should be expected to acquire) a sophisticated understanding of probability or facility in manipulating probabilities or that explicit probability arguments should become routine in courts of law. There are, however, increasing numbers of cases—such as DNA identification or the Sally Clark case (see the section The Prosecutor's Fallacy)—where evidence about probabilities is clearly relevant, and the court would stand to benefit from advice about how to handle them. This does not, of course, imply that it will appreciate this need or be ready to accept such advice: For example, at the first appeal of Sally Clark, the expert statistical evidence of Philip Dawid and Ian Evett was essentially dismissed by the court on the grounds that statistics is "hardly rocket science."

Sometimes—but all too rarely—there will be extensive relevant frequency data in the light of which all reasonable subjective probabilities for some event should essentially agree with its observed relative frequency. In other cases, all parties may be willing to accept an expert witness's assessments of some probabilities. Yet other probabilities, relevant for the juror or other judicial

decision maker, will be subject to subjective vagueness, although we will usually be able to distinguish between "reasonable" and "unreasonable" probability assessments. But even where probability values can be agreed on, their correct handling is far from obvious or intuitive, and fallacious intuitions, arguments, and inferences abound.

PROBABILITY LOGIC

In a case at law, let $\mathcal{E}$ denote one or more items of evidence (perhaps its totality). We need to consider how this evidence affects the comparison of the hypotheses, H_0 and H_1 say, offered by either side. Thus, in a criminal case with a single charge against a single defendant, the evidence might be that the defendant's DNA profile matches the one found at the crime scene. Hypothesis H_0, offered by the defence, is that the defendant is innocent ($\overline{G}$); the prosecution hypothesis, H_1, is that of guilt (G).

The adjudicator needs to assess his or her conditional probability for either hypothesis, *given* the evidence: $\Pr(H_0|\mathcal{E})$ and $\Pr(H_1|\mathcal{E})$. However, it will not usually be possible to assess these directly, and they will have to be constructed out of other, more basic, ingredients. In particular, it will often be reasonable to assess directly $\Pr(\mathcal{E}|H_0)$ and $\Pr(\mathcal{E}|H_1)$: the probability that the evidence would have arisen in each of the competing scenarios.

Bayes's theorem—a trivial consequence of the definition of conditional probability—tells us that

$$\frac{\Pr(H_1|\mathcal{E})}{\Pr(H_0|\mathcal{E})} = \frac{\Pr(H_1)}{\Pr(H_0)} \times \frac{\Pr(\mathcal{E}|H_1)}{\Pr(\mathcal{E}|H_1)}. \qquad (24.1)$$

The left-hand side of (24.1) is the *posterior odds* for comparing H_1 and H_0 given the evidence $\mathcal{E}$: This is a simple transformation of $\Pr(H_1|\mathcal{E})$, the desired *posterior probability* of H_1.

The second term on the right-hand side of (24.1) is constructed out of the directly assessed terms $\Pr(\mathcal{E}|H_0)$ and $\Pr(\mathcal{E}|H_1)$: It is the *likelihood ratio* (for H_1, as against H_0) engendered by the evidence $\mathcal{E}$. It is noteworthy that only the ratio of these terms enters, their absolute values being otherwise irrelevant.

To complete (24.1), we need the term $\Pr(H_1)/\Pr(H_0)$, the *prior odds* for comparing H_1 and H_0 (i.e., before evidence $\mathcal{E}$ is incorporated). This might reasonably vary from one individual juror to another, so that it would not be appropriate to treat it as a subject for direct evidence. For this reason, forensic experts are often instructed to give their evidence in the form of a likelihood ratio, it being left to the adjudicator to combine this appropriately with the prior assessment, using (24.1).

We can express (24.1) in words as

POSTERIOR ODDS
= PRIOR ODDS × LIKELIHOOD RATIO.

When $\mathcal{E}$ denotes all the evidence in the case, all the probabilities in (24.1) are unconditional; in particular, the prior odds should be assessed on the basis that there is no evidence to distinguish the suspect from any other potential suspect—this can be regarded as one way of formalizing the legal doctrine of "presumption of innocence" (which, of course, is not the same as an *assumption* of innocence). When $\mathcal{E}$ denotes a piece of evidence presented in midprocess, all the probabilities in (24.1) must be conditioned on the evidence previously presented: In particular, the "prior" probabilities could themselves have been calculated using (24.1), as posterior probabilities based on earlier evidence.

Notwithstanding the unarguable correctness of (24.1), it is often replaced by other, more "intuitive" probabilistic arguments that can be very misleading.

The Prosecutor's Fallacy

In a criminal trial, an item of evidence $\mathcal{E}$ may be offered in proof of the guilt, G, of a defendant S, on the basis that the probability of $\mathcal{E}$ would be very low if S were not guilty ($\overline{G}$). For example, in the trial of Sally Clark for double infanticide (Dawid, 2005, 2007), an expert medical witness testified that the probability that both her babies would have died from natural causes was one in 73 million.[1] If, as appears very natural, we describe this figure as "the probability that the babies died by innocent means," it is all too easy to misinterpret

[1]This figure has itself been widely and properly criticized, but that is not the issue here.

this as the probability (on the basis of the evidence of the deaths) that Sally is innocent—such a tiny figure seeming to provide incontrovertible proof of her guilt. Mathematically, this is equivalent to misinterpreting $\Pr(\mathcal{E}|\overline{G})$ as $\Pr(\overline{G}|\mathcal{E})$. For obvious reasons, this error is known as "transposing the conditional" or, because it typically produces seemingly convincing evidence of guilt, "the prosecutor's fallacy."

The prosecutor's fallacy is a seductive and widespread mode of reasoning, affecting the general public, the media, lawyers, jurors, and judges alike. Although we do not have access to the deliberations of Sally Clark's jury, it has generally been considered that their "guilty" verdict was strongly influenced by such mistaken reasoning.

Forensic Identification

A particularly fertile field where the prosecutor's fallacy flourishes is that of *identification evidence*. Here, unlike the case of Sally Clark, it is undisputed that a crime has been committed: The issue before the court is whether or not the suspect, S, is indeed the culprit C. Thus, the hypothesis G of guilt is equivalent to that of identity, $C = S$. Evidence $\mathcal{E}$ is presented that bears on this. This may be, for example, eyewitness evidence (as in the celebrated "Collins case" [Fairley & Mosteller, 1977], which kick-started modern interest in the interpretation of probabilities in the law) or forensic evidence of a *match* between some characteristic of the crime scene (the "crime trace") and a similar characteristic measured on the suspect. Examples include handwriting, rifling marks on bullets, glass fragments, fibers, footprints, fingerprints, bite marks, and, of especial importance and power, DNA profiles. It is common in such a case for the jury to be told something like "The probability of this DNA match arising from an innocent man is only one in one billion" and for all parties to misinterpret this number, in line with the prosecutor's fallacy, as the probability of S's innocence (Balding & Donnelly, 1995).

The Island Problem

The "island problem" (Eggleston, 1983, Appendix 3) is a toy example that well illustrates the uses and misuses of statistical logic in forensic identification.

A murder has been committed on an island, cut off from the outside world, on which $N+1$ inhabitants remain. Forensic evidence at the scene consists of a measurement, $I_C = x$, on a "crime trace" characteristic I_C, which can be assumed to come from the criminal C. The initial probability of any given islander having the characteristic x is assessed as P, independently for different islanders. Moreover, before observing any evidence, all inhabitants of the island are considered to have the same probability of being the culprit. The mainland police arrive and arrest a random islander, S. It is found that S matches the crime trace: $I_S = x$. There is no further relevant evidence. How should this match evidence be used to assess the claim that S is the murderer?

We shall consider a number of arguments that have been used to address this question. Those in the sections Defence Counterargument, Bayesian Argument, and Supreme Court Variation 3 below yield the correct answer, the remainder being fallacious: We leave it to the reader to identify the reasons. For illustration, following Eggleston, we take $N = 100$, $P = 0.004$.

Prosecutor's Fallacy

Prosecuting counsel, arguing according to his favorite fallacy, asserts that the probability that S is guilty is $1 - P$, or 0.996, and that this proves guilt "beyond a reasonable doubt."

Defence Counterargument

Counsel for the defence points out that while the guilty party must have characteristic x, the expected further number having this characteristic among the remaining N innocent islanders is NP. Hence, the set of islanders having this characteristic can be taken to have size $1 + NP$. The match evidence places S in this set but does not otherwise distinguish him from any of the other members of it. Since just one of these is guilty, the probability that this is S is thus $1/(1 + NP)$, or 0.714—indicative, perhaps, but *not* "beyond a reasonable doubt."

Bayesian Argument

Conditioning all the time on the evidence $I_C = x$ from the crime scene (which, we assume, of itself has no bearing on the issue of guilt) and

taking $\mathcal{E}$ to be the additional "match evidence" $I_S = x$, the probability of this evidence would be $\Pr(\mathcal{E}|G) = 1$ if S were guilty ($S = C$) and $\Pr(\mathcal{E}|\overline{G}) = P$ if S were innocent. Hence, the *likelihood ratio* in favor of guilt, on the basis of the match evidence, is

$$\text{LR} := \frac{\Pr(\mathcal{E}|G)}{\Pr(\mathcal{E}|\overline{G})} = \frac{1}{P},$$

or LR = 250.

While this seems strong evidence in favor of guilt, a complete probabilistic argument must also incorporate the prior odds on guilt before taking account of the match evidence. We can argue that in the absence of any other evidence, S is no more or less likely to be the culprit than any other islander, so that the prior probability of guilt is $1/(N+1)$, corresponding to prior odds on guilt of $1/N$.

We can now apply Bayes's theorem (24.1) to obtain the posterior odds on guilt

$$(1/N) \times (1/P) = 1/NP. \tag{24.2}$$

The corresponding posterior probability of guilt is

$$\Pr(G|\mathcal{E}) = \frac{1}{1+NP}, \tag{24.3}$$

or 0.714.

Note that this Bayesian argument could be readily modified to incorporate additional evidence if available—it is merely necessary to adjust the prior odds appropriately (either informally or formally by means of yet another application of Bayes's theorem) to take that into account.

We see that in the absence of additional evidence, this result accords with that of the defence argument above.

Supreme Court Argument

In its appeal judgment on the "Collins case," the Supreme Court of California argued on the following lines. Denote by M the unknown number of islanders possessing characteristic x. Before obtaining any evidence, we can take M to have the binomial distribution $\text{Bin}(N+1;P)$. Now, we have observed that S has characteristic x and so have learned that $M \geqslant 1$. If $M = 1$, there is no other matching individual and S must be guilty; however, if there is a nonnegligible probability that $M > 1$, so that S is not the only matching individual, this would be a source of doubt as to S's guilt. Hence, the Supreme Court calculated

$$\begin{aligned}\Pr(M > 1 | M \geqslant 1) \\ = \quad & \frac{1-(1-P)^{N+1}-(N+1)P(1-P)^N}{1-(1-P)^{N+1}},\end{aligned}$$

which, for our illustrative figures, yields 0.19. An approximately 20% chance of there being another islander having the matching characteristic could be considered enough to raise reasonable doubt as to S's guilt.

Supreme Court: Variation 1

The above line of argument can be developed further, as follows. With no other evidence, we can take $\Pr(G|M = m) = m^{-1}$. As above, we condition the initial $\text{Bin}(N+1;P)$ for M on the known fact that $M \geqslant 1$, to obtain

$$\Pr(G|\mathcal{E}) = \text{E}\left(M^{-1}|M \geqslant 1\right).$$

This is not simply expressible algebraically but can be calculated numerically: For our illustrative figures, it yields $\Pr(G|\mathcal{E}) = 0.902$.

Supreme Court: Variation 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, a 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

$$\Pr(G|\mathcal{E}) = \frac{1-(1-P)^{N+1}}{(N+1)P}, \tag{24.4}$$

which, for our values, gives 0.824.

Supreme Court: Variation 3

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 $\{m/(N+1)\}^2$, and using Bayes's theorem, the resulting conditional distribution of M is

$$\Pr(M = m | I_C = x, I_S = x) = c\,m \binom{N}{m-1} P^{m-1} (1-P)^{N-m+1} \quad (m = 1, \ldots, N+1),$$

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

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

or 0.714—in agreement with the Bayesian and defence arguments.

The Effect of Search

We have so far supposed that the suspect S was selected at random from the island population and, quite fortuitously, was found to match the crime trace. More realistically, the police might trawl through the population until they discover an individual who provides a match. Because this will yield further information beyond the mere fact of a match, we can expect the resulting inference to differ from that appropriate to the "lucky match" case previously considered.

If the search delivers a (first) match for the $(q+1)$th individual examined, then q necessarily innocent parties have been eliminated, thereby reducing the size of the remaining suspect population from N to $N-q$. Intuitively, it would seem that formulas (24.2) and (24.3) given above must therefore be adjusted by making this substitution, so yielding

$$\Pr(G|q) = \frac{1}{1+(N-q)P}. \qquad (24.5)$$

This is correct, although the full analysis is more subtle since it must account for the probabilistic nature of the outcome q of the search (Dawid & Mortera, 1995).

Formula (24.5) can only be applied when we know q, the number of nonmatching individuals examined before the matching suspect S is found. But whatever the value of q, (24.5) will yield a value at least as large as (24.3). It follows that if we know that a search has been conducted to identify a suspect but are not told q, the answer given by Formula (24.3) must be too small. In fact, in this case of a known search of unknown size, the appropriate answer is now given by Formula (24.4).

Database Search

Search scenarios are common in cases where a DNA trace is found at the crime scene and, in the absence of any obvious suspect, a search for a match is made through a police database of DNA profiles. Such databases can be very large—by December 2005, the U.K. database comprised around 3 million profiles, with about 3,000 "matches" being made per month.

Computerized search typically allows us to identify every individual in the database whose DNA profile matches the crime trace. Suppose that there is exactly one such individual, S. If the initial suspect population is of size $N+1$ and the database is of size $n+1$, then the search has eliminated n individuals from the suspect population, and so, if there is no other evidence to distinguish among those remaining, the odds on S being guilty are increased from $1/NP$, as in (24.2), to $1/(N-n)P$. (If there is other evidence for or against S, this could be expressed as a likelihood ratio and combined with the above odds using Bayes's theorem. It is also possible to account for evidence pointing the finger toward or away from other individuals.)

When n is small in relation to N, the effect of the database search is only a small increase in the probability that S is guilty. This is fortunate, since evidence that a search was conducted to identify the suspect is usually inadmissible in court. Ignoring it will typically make little difference, and to the extent that it does, it will be to the advantage of the defendant.

However, at the other extreme, where the whole population is searched ($n = N$) and S is the only individual found to match, we obtain infinite odds, corresponding to certainty, that S is guilty—as is obviously appropriate in this case.

Alternative Arguments

Other arguments, with very different implications, have also been brought to bear on this problem.

One *frequentist* view, recommended by the U.S. National Research Council (1996), treats the

problem as analogous to that of multiple statistical hypothesis testing, where the strength of the evidence has to be adjusted to account for the very fact that a search has been conducted. It is argued that since *any* match found in the database would have resulted in a prosecution, the relevant "match probability" is no longer the probability, P, that S would match the crime trace (if innocent) but the probability, approximately $(n+1)P$, that *some* match would be found in the database (if all its members were innocent). The impact of the evidence, as measured by the match probability, is thus attenuated by a factor of $n+1$, the size of the database. Even if this is only a very small fraction of the total population, it can be very large in absolute size, which would appear to render the match evidence essentially worthless.

A closely related *likelihood* viewpoint is taken by Stockmarr (1999). He claims that it is not appropriate to assess a likelihood for the hypothesis H_S that S is guilty, since that hypothesis could not even have been formulated before the search was conducted. Hence, he claims, we should instead focus on the hypothesis H_D—which *can* be formulated before the search—that the database D contains the culprit. When the search then turns up a single match, the corresponding likelihood ratio in favor of H_D (as against its negation) is about $1/(n+1)P$ (as compared with $1/P$ in favor of the "data-dependent" hypothesis H_S). Moreover, whoever the (unique) matching individual turns out to be, the hypothesis H_D becomes logically equivalent to the hypothesis that this matcher is the culprit, which is the proposition that will be put before the court. Consequently, the strength of the evidence is more appropriately measured by a likelihood ratio of $1/(n+1)P$ than one of $1/P$.

We can reconcile this view with the analysis given in the section Database Search if we remember that a likelihood ratio is only one of the ingredients in Bayes's theorem (Dawid, 2001). If we replace H_S by H_D, not only will the likelihood ratio change, but so too will the prior odds: Because the database contains $(n+1)$ individuals, a priori, the odds on the culprit being one of these will be about $(n+1)$ times greater than the odds on his or her being the specific individual S. It turns out that on performing this replacement of the hypothesis, the change to the prior odds exactly cancels with that to the likelihood ratio. There is thus no net effect on the posterior odds: Both approaches deliver the same ultimate verdict.

Which Likelihood Ratio?

The above analysis does, however, lead to problems for the forensic scientist, who is, quite properly, trained to testify as to "the likelihood ratio" generated by the evidence and not directly as to the posterior probability. When, as above, we have a choice as to how to frame the hypotheses, there is no unique likelihood ratio (although the posterior probability will be unaffected by this indeterminacy). In that case, it would seem more helpful to the court to present the likelihood ratio for the hypotheses of direct interest: that S is, or is not, the culprit.

A related issue arises when it can be assumed that the crime was committed by two persons, each of whom has left a DNA trace at the scene (say one on a pillow and one on a sheet). S is arrested and it is found that his DNA matches the trace from the pillow, which has population frequency P. Under reasonable assumptions, it can be shown (Dawid, 2004) that the likelihood ratio in favor of the hypothesis that S was one of the culprits, as against his innocence, is $1/(2P)$. But (given the evidence) S is guilty if and only if he left the stain on the pillow and taking this as the hypothesis at issue leads to a likelihood ratio (as against S's innocence) of $1/P$. Other ways of framing the hypotheses yield yet other results (Meester & Sjerps, 2004).

Once again these different answers can be reconciled by taking proper account of the differing prior probabilities. But if one value is to be given to the court as "the likelihood ratio," what should it be? The first value quoted above, $1/(2P)$, does directly address the question at issue: Is S guilty or not? On the other hand, the very existence of two culprits makes it a priori about twice as probable that S is guilty as would hold for the case of a single-culprit crime. If the court is used to thinking about this latter case, and is not attuned to the need to double the prior probability, one might argue, as a pragmatic solution, that the "correct" likelihood ratio, $1/(2P)$, should be doubled, so as

to build this correction in automatically—which would bring us back to the value $1/P$.

Complex Patterns of Evidence

The difficulties of assessing a single item of evidence are compounded when we want to account for the complex interrelationships between the many items of evidence in a case. To organize the evidence it is then helpful to construct a diagrammatic representation of all the evidence and hypotheses in the problem and the relationships between them. This idea was first suggested by Wigmore (1937) (see Anderson, Schum, & Twining, 2005, for an introduction to the "Wigmore chart" method). More recently, the methods of graphical modeling and Bayesian networks—also known as probabilistic expert systems (Cowell, Dawid, Lauritzen, & Spiegelhalter, 1999)—have been applied. Such a network contains a node for each variable in the problem, with arrows between nodes to denote probabilistic dependence of a "child" node on all its "parents." To complete the description, we need the numerical or algebraic specification of the associated conditional probabilities.

Example

Dawid and Evett (1997) consider a fictitious burglary case, described as follows:

> An unknown number of offenders entered a commercial premises late at night through a hole, which they cut in a metal grille. Inside, they were confronted by a security guard who was able to set off an alarm before one of the intruders punched him in the face, causing his nose to bleed.
>
> The intruders left from the front of the building just as a police patrol car was arriving and they dispersed on foot, their getaway car having made off at the first sound of the alarm. The security guard said that there were four men, but the light was too poor for him to describe them, and he was confused because of the blow he had received. The police in the patrol car saw the offenders only from a considerable distance away. They searched the surrounding area and, about 10 min later, one of them found the suspect trying to "hot wire" a car in an alley about a quarter of a mile from the incident.
>
> At the scene, a tuft of red fibers was found on the jagged end of one of the cut edges of the grille. Blood samples were taken from the guard and the suspect. The suspect denied having anything to do with the offence. He was wearing a jumper and jeans, which were taken for examination.
>
> A spray pattern of blood was found on the front and right sleeve of the suspect's jumper. The blood type was different from that of the suspect but the same as that from the security guard. The tuft from the scene was found to be red acrylic. The suspect's jumper was red acrylic. The tuft was indistinguishable from the fibers of the jumper by eye, microspectrofluorimetry and thin layer chromatography (TLC). The jumper was well worn and had several holes, though none could clearly be said to be a possible origin for the tuft.

In this example, there are three general kinds of evidence: eyewitness, blood, and fiber; and for each kind a variety of individual evidential items. We can summarize the salient features of the evidence against the suspect as follows:

- *Eyewitness*
  - G: The evidence of the security guard
  - W: The evidence of the police officer who arrested the suspect
- *Blood*
  - R: The bloodstain in the form of a spray on the suspect's jumper
  - X_1: Suspect's blood type
  - X_2: Guard's blood type
  - Y_2: Blood type of blood spray on jumper
- *Fibers*
  - X_3: Properties of the suspect's jumper
  - Y_1: Properties of fiber tuft

The uncertain hypotheses and variables that enter are

- *Hypotheses*
  - C: Whether the suspect was or was not one of the offenders
  - A: The identity of the person who left the fibers on the grille

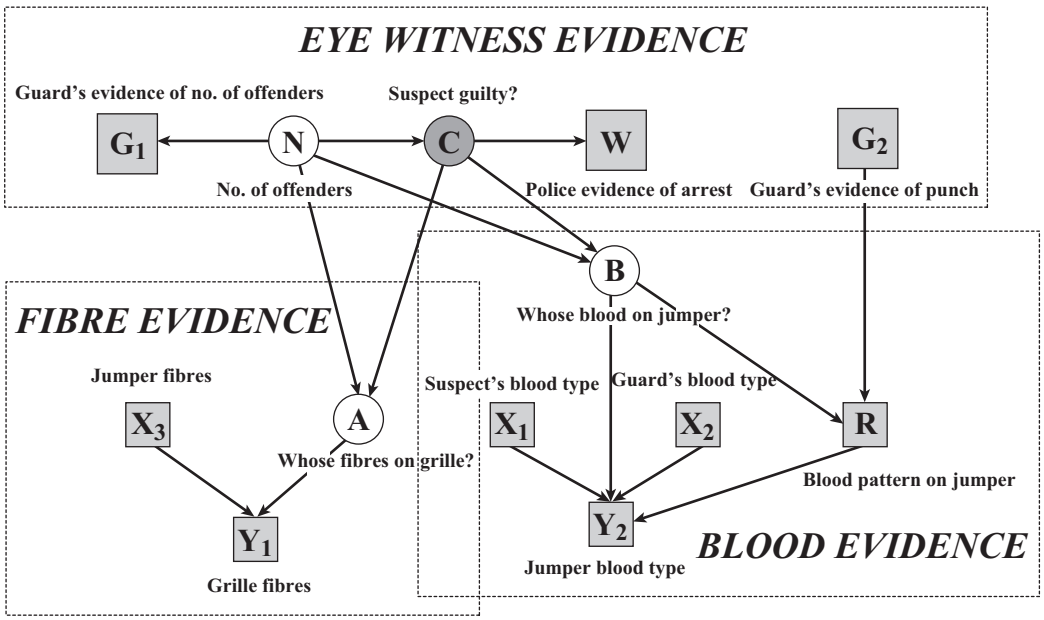

Figure 24.1 Bayesian network for burglary example.

B: The identity of the person who punched the guard

N: The number of offenders

Of these, the specific charge before the court is $C = \text{TRUE}$; the others are included to provide a complete account of the problem.

Figure 24.1 shows a graphical representation of the problem as a Bayesian network. The evidence items are shown as squares and the hypotheses as circles. Variable Y_2, the measurement of the blood type of the spray on the jumper, is dependent on X_1, the suspect's blood type (because it might be a self-stain) and the guard's blood type X_2. But information is also provided by R, the variable that describes the shape of the stain, because that sheds light on whether or not it might be a self-stain. In turn, the shape of the stain is influenced by the way in which the guard was punched, G_2, and B, the identity of the person who did it; while B is, in turn, influenced by whether or not the suspect was one of the offenders, variable C, and also the number of offenders, N.

Dawid and Evett (1997) describe how the graph can be used to read off implicit properties of independence: For example, to show that conditionally on knowing A and N, the pair of variables (B,R) is independent of the pair (G_1,Y_1). These properties can then be used to simplify the algebraic and numerical identification of the overall likelihood ratio for comparing the hypotheses $C = \text{TRUE}$ or $C = \text{FALSE}$, based on the evidence.

Taroni, Aitken, Garbolino, and Biedermann (2006) give a detailed account of theory and applications of Bayesian networks in problems of forensic inference. See Baio and Corradi (2006) and Cavallini and Corradi (2005) for further interesting examples.

Forensic Genetics

Most of the logic so far presented applies in principle to any kind of identification evidence. But forensic DNA evidence has some additional special features, principally owing to its pattern of inheritance from parent to child. These make it possible to use it to address queries such as the following:

Disputed paternity:
Is individual A the father of individual B?

Disputed inheritance:
Is A the daughter of deceased B?

Immigration:
Is A the mother of B? How is A related to B?

Criminal case—mixed trace:
Did A and B both contribute to a stain found at the scene of the crime? Who contributed to the stain?

Disasters:
Was A among the individuals involved in a disaster? Who were those involved?

In a simple disputed paternity case, the evidence $\mathcal{E}$ will comprise DNA profiles from mother, child, and putative father. Hypothesis H_1 is that the putative father is the true father, while hypothesis H_0 might be that the true father is some other individual, whose DNA profile can be regarded as randomly drawn from the population. We can also entertain other hypotheses, such as that one of one or more other identified individuals is the father or that the true father is the putative father's brother.

In a complex criminal case, we might find a stain at the scene of the crime having the form of a *mixed trace*, containing DNA from more than one individual. DNA profiles are also taken from the victim and a suspect. We can entertain various hypotheses as to just who—victim, suspect, person or persons unknown—contributed to the mixed stain.

When we are only comparing two hypotheses H_0 and H_1, the impact of the totality of the DNA evidence $\mathcal{E}$ available, from all sources, is once again crystallized in the *likelihood ratio*, $\mathrm{LR} = P(\mathcal{E}|H_1)/P(\mathcal{E}|H_0)$. If we wish to compare more than two hypotheses, we require the full *likelihood function*, a function of the various hypotheses H being entertained (and, of course, the evidence $\mathcal{E}$):

$$\mathrm{LR}(H) \propto \Pr(\mathcal{E}|H). \tag{24.6}$$

The proportionality sign in (24.6) indicates that we have omitted a factor that does not depend on H, although it can depend on $\mathcal{E}$. Such a factor is of no consequence and need not be specified, since it disappears on forming ratios of likelihoods for different hypotheses on the same evidence. Only such relative likelihoods are required, not absolute values.

We also now need to specify the prior probabilities, $\Pr(H)$, for the full range of hypotheses H. Then, posterior probabilities in the light of the evidence are again obtained from Bayes's theorem, which can now be expressed as

$$\Pr(H|\mathcal{E}) \propto \Pr(H) \times \mathrm{LR}(H). \tag{24.7}$$

Again, the omitted proportionality factor in (24.7) does not depend on H, although it might depend on $\mathcal{E}$. It can be recovered, if desired, as the unique such factor for which the law of total probability, $\sum_H \Pr(H|\mathcal{E}) = 1$, is satisfied.

Genetic Background

To proceed further, we need some basic genetic facts about DNA profiles, which we summarize very briefly below: See, for example, Buckleton, Triggs, and Walsh (2005) for more details.

A gene is a particular sequence of the four *bases*, represented by the letters A, C, G, and T, that carry the genetic information in DNA. A specific position on a chromosome is called a *locus*; since chromosomes come in pairs, there are two genes at any locus. A *DNA profile* consists of measurements on a number of *forensic markers*, which are specially selected loci, on different chromosomes. Current technology uses around 12–20 *short tandem repeat* (STR) markers. Each such marker has a finite number (up to around 20) of possible values, or *alleles*, generally positive integers. For example, an allele value of 5 indicates that a certain word (e.g., CAGGTG) in the four-letter alphabet of the genetic code is repeated exactly five times in the DNA sequence at that locus on a chromosome.

An individual's *DNA profile* comprises a collection of *genotypes*, one for each marker. Each genotype consists of an unordered pair of alleles, one inherited from the father and one from the mother (though one cannot distinguish which is which). When both alleles are identical, the individual is *homozygous* at that marker, and only a

single allele value is observed; else the individual is *heterozygous*. In most cases, a DNA profile can be measured without error, even from a single cell.

Assuming *Mendelian segregation*, at each marker a parent passes a copy of just one of his two alleles, randomly chosen, to his or her child, independently of the other parent and independently for each child. Distinct forensic markers are located on different chromosomes, so segregate independently. It is often reasonable to assume *random mating* within an appropriate population, which then implies independence of alleles both within markers (*Hardy-Weinberg equilibrium*) and across markers (*linkage equilibrium*). Databases have been gathered from which allele frequency distributions, for various populations, can be estimated for each forensic marker. On the basis of these values and the independence assumptions, a *profile probability* can be assigned to any DNA profile, measuring its rarity in the population.[2]

Simple Disputed Paternity

A man is alleged to be the father of a child, but disputes this. DNA profiles are obtained from the mother `m`, the child `c`, and the putative father `pf`. On the basis of these data, we wish to assess the likelihood ratio for the hypothesis of *paternity*: H_1: `tf=pf`, the true father is the putative father; as against that of *nonpaternity*: H_0: `tf = af`—where `af` denotes an unspecified alternative father, treated as unrelated to `pf` and randomly drawn from the population.

The disputed pedigree can be represented as in Figure 24.2.

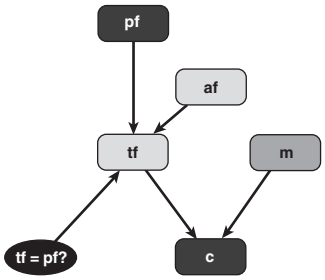

Figure 24.2 Pedigree for simple disputed paternity.

Because of our independence assumptions, we can analyze the markers one at a time, finally multiplying their associated likelihood ratio values together to obtain the overall likelihood ratio based on the full collection of markers.

Consider now the measured genotypes, from all three parties, for some fixed marker. Under paternity, H_0, we just apply Mendel's laws of segregation; under nonpaternity, H_1, we require (estimates of) the frequencies of relevant marker alleles among the population. Using (24.1), this can then be combined with the prior odds of paternity, based on external background evidence B, to obtain the posterior odds for paternity. As an illustrative example, suppose that the data, for marker D7, are child's genotype `cgt`$=\{12,12\}$, mother's genotype `mgt`$=\{10,12\}$, putative father's genotype and `pfgt`$=\{10,12\}$. The estimated population frequencies of alleles 10 and 12 are, respectively, 0.284 and 0.260. In this case, by conditioning on the genotypes of mother and putative father, we see that the child's genotype will be as observed if and only if both the mother and the true father contributed allele 12 to the child. This event has probability 0.5×0.5 if the true father is the putative father, and probability 0.5×0.260 if the true father is, instead, some unrelated individual from the population. Thus, the likelihood ratio in favor of paternity, based on marker $D7$ alone, is 1.93.

DNA Mixtures

A *mixed DNA profile* is typically obtained from an unidentified biological stain or other trace thought to be associated with a crime. This commonly occurs in rape cases, in robberies where an object might have been handled by more than one individual, and also in a scuffle or brawl. For a mixed DNA trace, there is no constraint on the number of distinct alleles observed for each marker, since the trace might have been formed as a mixture of biological material from more than one person.

[2] Although we do not develop this here, one should really allow for the fact that allele frequency estimates based on finite databases remain uncertain. This raises some subtle new issues (Balding & Nichols, 1994; Dawid & Mortera, 1996).

In simple cases of DNA mixtures when using only the qualitative allele information, algebraic formulae for calculating the likelihoods of all hypotheses involving a specified set of known and unknown contributors to the mixture can be computed (assuming Hardy-Weinberg equilibrium and known allele frequencies).

To illustrate, suppose that for a single DNA marker, we have a three-allele crime trace $\{A,B,C\}$, and individual profiles from a victim, $v = \{B,C\}$ and a suspect, $s = \{A\}$. These together with the allele frequencies constitute the evidence $\mathcal{E}$ for the case. Suppose we wish to compute the likelihood ratio in favor of the hypothesis that the victim and suspect contributed to the mixture, H_0: $v \& s$, as against the hypothesis that the victim and an unknown individual u contributed to the mixture, H_1: $v \& u$. It is not difficult to show that in this case

$$\mathrm{LR} = \frac{1}{p_A^2 + 2p_A p_B + 2p_A p_C}, \qquad (24.8)$$

where p_i is the frequency of allele i in the population.

Bayesian Networks for Forensic DNA Identification

In more complex scenarios than those described above, it can become difficult or impossible to obtain the required probabilistic formulas.

In cases of disputed paternity, it commonly occurs that the DNA profiles of one or more of the "principal actors" in the pedigree are not available; but there is indirect evidence, in the form of DNA profiles of various known relatives. In the section Complex Disputed Paternity below, we consider such a case, where the putative father is unavailable for testing, but we have DNA from two of his brothers and an undisputed child of his by another woman. The analysis of all the data is clearly now much more complex. Likewise the appropriate extensions of (24.8) become relatively complex when the number of potential contributors to the mixture becomes large; if we want to use quantitative data (peak areas), which contain important additional information about the composition of the mixture, and to allow for uncertainty in allele frequencies and/or population substructure.

To handle such cases, sophisticated probabilistic modeling tools are required. Again, Bayesian networks, together with their associated computational methodology and technology, have been found valuable for this, particularly in their "object-oriented" Bayesian networks (OOBN) form, as implemented in commercial software such as HUGIN 6.[3] Bayesian networks for evaluating DNA evidence were introduced by Dawid, Mortera, Pascali, and van Boxel (2002). Further description and developments can be found in Mortera (2003, chap. 1B); Mortera, Dawid, and Lauritzen (2003); Vicard, Dawid, Mortera, and Lauritzen (2008); Cowell, Lauritzen, and Mortera (2007); Dawid, Mortera, and Vicard (2006); Dawid, Mortera, and Vicard (2007); and Taroni et al. (2006).

For some illustrative cases, we describe below how we can construct a suitable OOBN representation of a complex DNA identification problem incorporating all the individuals involved and the relationships between them.

Simple Disputed Paternity

We use the example of simple disputed paternity given in the section Forensic Genetics to introduce some basic ingredients of forensic OOBNs.

In fact, Figure 24.2 is just the relevant "top-level" network, constructed using the graphical interface to HUGIN 6. Each node (except the hypothesis node `tf=pf?`) in Figure 24.2 is itself an "instance" of another generic ("class") network, with further internal structure. In what follows, **bold face** will indicate a network class, and `teletype face` will indicate a node or instance. We describe only selected features here. A fuller description of OOBN networks for paternity casework can be found in Dawid et al. (2007) and Dawid et al. (2006).

Each of `m`, `pf`, and `af` is an instance of a class **founder**, while `c` is an instance of class **child** and `tf` is an instance of class **query**.

Within **founder** (not shown) we have two instances (maternal and paternal genes) of a class **gene**, which embodies the relevant repertory of

[3] Obtainable from http://www.hugin.com.

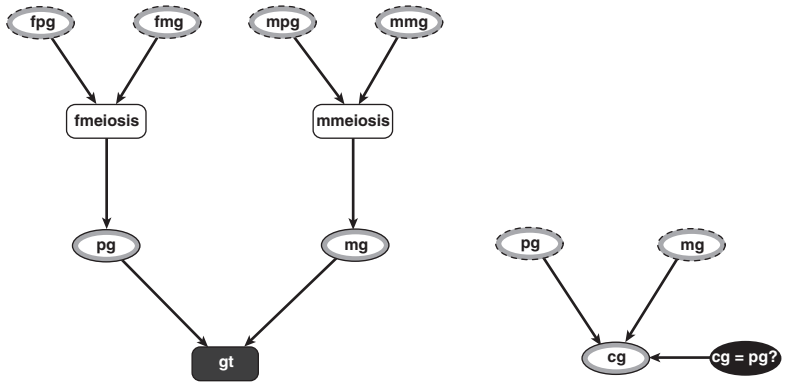

Figure 24.3 Networks **child** and **meiosis**.

alleles and their associated frequencies in the relevant population.

The internal structure of **child** is displayed in Figure 24.3.

On the paternal (left hand) side of **child**, the input nodes `fpg` and `fmg` represent the child's father's paternal and maternal genes. These are then copied into nodes `pg` and `mg` of an instance `fmeiosis` of a class network **meiosis**, whose output node `cg` is obtained by flipping a fair coin (node `cg=pg?`) to choose between `pg` and `mg`; this is then copied to `pg` (child's paternal gene) in network **child**. A similar structure holds for the maternal (right hand) side of **child**. Finally, `pg` and `mg` are copied into an instance `gt` of a network class **genotype**, which forgets the information on parental origin (this is also a feature of **founder**). Any DNA evidence on the individual is entered here.

The hypothesis node `tf=pf?` embodies H_0 (`tf` = `pf`) when it takes the value *true* and H_1 (`tf` = `af`) when *false*; it feeds into the instance `tf` of class **query** to implement this selection. We initially, and purely nominally, set both hypotheses as equally probable, so that, after propagation of evidence, the ratio of their posterior probabilities yields the paternity ratio based on this marker. By entering the data for each marker into the appropriate Bayesian network, we can thus easily calculate the associated likelihood ratio for paternity.

We build a separate such network for each STR marker, incorporating the appropriate repertoire of alleles and their frequencies. On entering the available DNA data, we can compute the associated likelihood ratio. Finally, we multiply these together across all markers to obtain the overall likelihood ratio.

Once supplied with the basic building blocks **founder**, **child**, and **query**, we can connect them together in different ways, much like a child's construction set, to represent a wide range of similar problems. An illustration is given in the next section.

Complex Disputed Paternity

Figure 24.4 is a OOBN representation of a disputed paternity case, where we have DNA profiles from a disputed child `c1` and from its mother `m1` but not from the putative father `pf`. We do, however, have DNA from `c2`, an undisputed child of `pf` by a different, observed, mother `m2`, as well as from two undisputed full brothers `b1` and `b2` of `pf`. The sibling relationship is made explicit by the incorporation of the unobserved grandfather `gf` and grandmother `gm`, parents of `pf`, `b1` and `b2`. The "hypothesis node" `tf=pf?` again indicates whether the true father `tf` is `pf` or is an alternative father `af`, treated as randomly drawn from the population.

Nodes `gf`, `gm`, `m1`, `m2`, and `af` are all instances of class **founder**; `pf`, `b1`, `b2`, `c1`, and `c2` are instances of class **child**; `tf` is an instance of class **query**.

The DNA evidence $\mathcal{E}$ consisted of the 6 DNA profiles, each comprising 10 STR markers, from `m1`, `m2`, `c1`, `c2`, `b1`, and `b2`. By entering the data for each marker into the Bayesian network (incorporating the appropriate alleles for that marker and their frequencies), we can thus easily

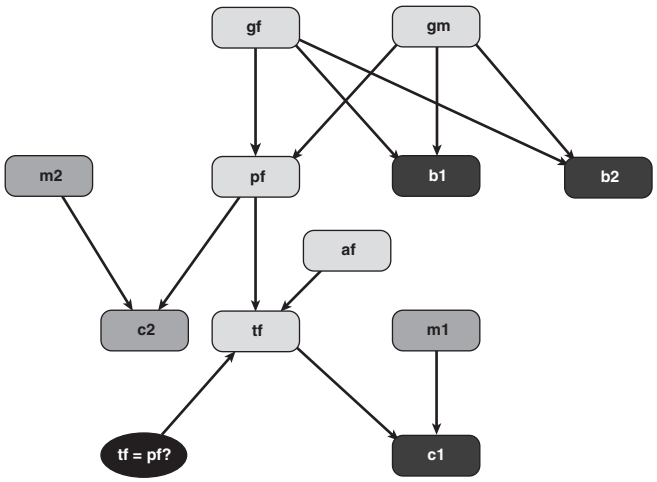

Figure 24.4 Pedigree for incomplete paternity case.

calculate the associated likelihood ratio for paternity. The overall paternity ratio is then given by their product.

For this particular case, this overall paternity ratio evaluates to around 1,300, meaning that the observed DNA evidence is 1,300 times more probable on the hypothesis of paternity than it would be were we to assume nonpaternity. According to Evett and Weir (1998, chap. 9), such a value might be considered as offering "very strong support" to the hypothesis of paternity (although paternity applications such as this will never produce the kind of likelihood ratio value, sometimes in the billions, that can occur when DNA profiling evidence is used to match a suspect to a crime). However, it is important to remember, in all cases, that the likelihood ratio derived from the DNA evidence is only one element of the whole story, which also involves prior probabilities, and perhaps further likelihood ratios based on other evidence in the case. All these ingredients need to be combined appropriately, using Bayes's theorem, to produce the final probability of paternity.

Mutation

It is easy to modify these networks to incorporate a variety of additional complications. One such is the possibility of *mutation* of genes in transmission from parent to child, which could lead to a true father appearing to be excluded (Dawid, Mortera, & Pascali, 2001; Dawid, Mortera, Dobosz, & Pascali, 2003; Dawid, 2003; Vicard & Dawid, 2004; Vicard et al., 2008). We must now distinguish between a child's *original gene* cog, identical with one of the parent's own genes, and the *actual gene* cag available to the child, which may differ from cog because of mutation. We elaborate the class network **meiosis** of Figure 24.3, as shown in Figure 24.5 by passing its original output cog ("child's original gene") through an instance cag ("child's actual gene") of a new network **mut**, constructed to implement whatever model is used to describe how the value of cog is stochastically altered by mutation. The output of cag is then copied to cg. Thus, **meiosis** now represents the result of mutation acting on top of Mendelian segregation.

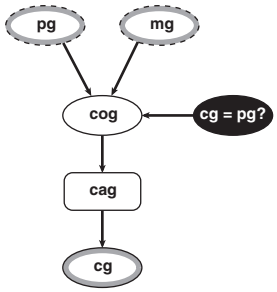

Figure 24.5 Revised network **meiosis**, incorporating mutation.

Table 24.1 Disputed paternity with brother too. `mgt` = {12, 15}, `pfgt` = {14, 14}, `cgt` = {12, 12}.

`pr(silent)`	L_D	L_B *with* `bgt` =						
		{16,20}	*{12,17}*	*{12,14}*	*{14,17}*	*{14,14}*	*{16,16}*	*{12,12}*
0	0	1	1	0.546	0.546	1	6.13	3,334
0.000015	0.472	1	1	0.546	0.546	1.0000	6.12	1,595
0.0001	2.473	1	1	0.546	0.546	0.9999	6.07	403.7
0.001	7.485	1	1	0.551	0.551	0.9992	5.54	46.07
0.01	8.100	1	1	0.590	0.590	0.9932	3.19	5.45

NOTE: Likelihood ratio in favor of paternity allowing for silent alleles: L_D, without brother's genotype. L_B, further (multiplicative) effect of brother's genotype.

Once an appropriate network **mut** has been built, and **meiosis** modified as described above, pedigree networks constructed as in the section Silent Allelles will now automatically incorporate the additional possibility of mutation.

Silent Alleles

Yet another complication that is easily handled by simple modifications to lower-level networks is the possibility that some alleles may not be recorded by the equipment, so that a truly heterozygous genotype appears homozygous (Dawid et al., 2007, 2006). This may be due to sporadic equipment failure, in which case it is not inherited and we talk of a *missed* allele; or to an inherited biological feature, in which case we refer to the allele as *silent*.

In some cases, making proper allowance for these possibilities can have a dramatic effect. Table 24.1 shows results for a particular case where, in addition to the genotypes `mgt`, `pfgt`, and `cgt` of mother, putative father, and child, we also have the genotype `bgt` of the putative father's brother. These refer to the single STR marker vWA.

If we had complete data on the genotypes `mgt`, `pfgt`, and `cgt`, the additional data `bgt` would have no effect whatsoever on the paternity ratio, since the child's genotype is conditionally independent of information on the putative father's brother given the mother and putative father's genotypes. In the case shown, in the absence of silence we would have an exclusion. Allowing for silence at various rates, but using only the data on the basic family triplet, gives the paternity ratios in the column labeled L_D, from which we already see that a small probability of silence can, in fact, lead to a paternity ratio greater than 1—now constituting evidence in favor of paternity. The remaining columns show the *additional* (multiplicative) effect of using the information on the brother's genotype `bgt` for various cases. The first row shows that even as the probability of silence tends to 0, its disturbing effect can be very substantial. In fact, when $\texttt{bgt} = \{12, 12\}$, the overall paternity ratio $\mathrm{LR} = L_D \times L_B$ achieves a maximum value of $1,027.3$, at $\texttt{pr(silent)} = 0.0000642$, even though it vanishes for $\texttt{pr(silent)} = 0$.

Bayesian Networks for Analyzing Mixed DNA Profiles

Bayesian networks have also been constructed to address the challenging problems that arise in the interpretation of mixed trace evidence, as described in the section DNA Mixtures. Typically, one would be interested in testing whether the victim and suspect contributed to the mixture, H_0: $v \,\&\, s$, against the hypothesis that the victim and an unknown individual contributed to the mixture, H_1: $v \,\&\, u$. One might alternatively consider an additional unknown individual u_2 instead of the victim, with hypotheses H_0: $u_2 \,\&\, s$ versus H_1: $u_2 \,\&\, u_1$.

Figure 24.6 shows a top-level network that can be used for analyzing a mixture with two contributors, $p1$ and $p2$. Nodes `sgt`, `vgt`, `u1gt`, and `u2gt` are all instances of a network class **genotype** and represent the suspect's, the victim's, and two unknown individuals' genotypes.

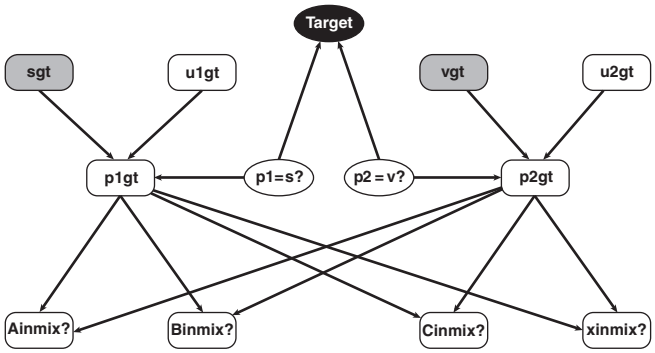

Figure 24.6 Bayesian network for DNA mixture from two contributors.

Boolean node `p1=s?` represents the hypothesis that contributor $p1$ is the suspect s. Node `p1gt`, the genotype of $p1$, is an instance of a network **query** that selects between the two genotypes `sgt` or `u1gt` according to the true/false state of the Boolean node `p1=s?`. A similar relationship holds between nodes `p2gt`, `vgt`, `u1gt`, and `p1=v?`. Possible genotype information on the suspect and/or the victim is entered and propagated from nodes `sgt` and `vgt`. The `target` node is the logical combination of the two Boolean nodes `p1=s?` and `p2=v?` and represents the four different hypotheses described above. `Ainmix?` determines whether allele A is in the mixture: This will be so if at least one A allele is present in either `p1gt` or `p2gt`. Similarly, for `Binmix?`, `Cinmix?`, `Dinmix?`, and `xinmix?` (where `x` refers to all the alleles that are not observed). Information on the alleles seen in the mixture is entered and propagated from these nodes.

The modular structure of Bayesian networks supports easy extension to mixtures with more contributors, as in cases where a rape victim declares that she has had one consensual partner in addition to the unidentified rapist or that she has been victim of multiple rape. Simple modification of the network handles such scenarios, so long as the total number of contributors can be assumed known.

In general, however, although the evidence of the trace itself will determine a lower bound to this total, there is in principle no upper bound. Thus, if in a trace we see that the maximum number of alleles in any marker is three, we know that the minimum number of contributors that could have produced this trace is two, but we cannot be sure that there were only two. However, it is often possible to set a relatively low-upper limit to the number it is reasonable to consider. We allow, as contributors to the mixture, persons with known DNA profiles, such as the victim and suspect and possibly also unknown individuals. Each of the various hypotheses H we might consider will involve a specification, x, for the number of unknown contributors. Although not strictly necessary, for extra clarity we write $\Pr_x(\mathcal{E}|H)$ for the probability of the evidence under this hypothesis. Thus, the likelihood ratio (LR) needed to evaluate the DNA evidence $\mathcal{E}$—comprising the DNA profiles of the victim, the suspect, and the mixed trace—in favor of a hypothesis H_0 against an alternative hypothesis H_1 is

$$\mathrm{LR} = \frac{\Pr_{x_0}(\mathcal{E}|H_0)}{\Pr_{x_1}(\mathcal{E}|H_1)},$$

where x_i denotes the number of unknown individuals involved in the hypothesis H_i.

When computing the weight of evidence, one should give the defendant the benefit of any doubt or uncertainty and so present the most favorable reasonable scenario for the defence. This implies that we should seek and use a lower bound for the value of the LR as we vary our assumptions within reasonable limits. And this, in turn, requires that we use an upper limit for the number of unknown contributors it is reasonable to consider. If the evidence is incriminating even in this most favorable case, it will be even more so for a larger number of unknown contributors.

To aid in setting such an upper limit, we can use the fact that $\Pr_x(\mathcal{E}|H)$ can be no larger than the probability that all the alleles of the x unknown contributors are in the mixed trace. This implies (Lauritzen & Mortera, 2002)

$$\Pr_x(\mathcal{E}|H) \leq \prod_{m=1}^{M} \kappa_m^{2x},$$

where, for each marker m, κ_m is the total probability that a randomly chosen allele will be one of those seen in the mixed trace. From this, it follows that if H_1 is any alternative hypothesis, yielding likelihood L_1, we need not consider an alternative hypothesis H with more than $b(L_1)$ unknown contributors, where

$$b(y) = \frac{\ln y}{2\sum_{m=1}^{M} \ln \kappa_m},$$

since that would yield a likelihood smaller than L_1.

Once it has been agreed to limit attention to some maximum total number of potential contributors, cases where the number of unknown contributors is itself uncertain can again be addressed using a Bayesian network, now including nodes for the number of unknown contributors and the total number of contributors (Mortera et al., 2003). This can be used for computing the posterior distribution of the total number of contributors to the mixture, as well as likelihood ratios for comparing all plausible hypotheses.

The modular structure of the Bayesian networks can be used to handle still further complex mixture problems. For example, we can consider together missing individuals, silent alleles, and a mixed crime trace simply by piecing together the appropriate modules.

The issue of silent alleles in a mixed trace arose in the celebrated case of *People v. O. J. Simpson* (Los Angeles County Case BA097211). At VNTR marker D2S44, the crime trace showed a three-band profile *ABC*, the victim had profile *AC* and the suspect had profile *AB*. The population allele frequencies are taken as $p_A = 0.0316$, $p_B = 0.0842$, and $p_C = 0.0926$ and the frequency of a silent allele as $p_n = 0.05$. For this marker, Table 24.2 gives the likelihoods (arbitrarily normalized to sum to 1) based on a network that handles silent alleles and allows for up to two unknown contributors. Results are shown both ignoring and allowing for silent alleles, and also for a "simplified" rough rule for accounting for silence, recommended in the report of the National Research Council (1996), which replaces the frequency p^2 by the much larger quantity $2p$.

Table 24.2 O. J. Simpson case: Likelihoods for hypotheses as to constitution of mixed trace, for suspect s, victim v, and varying number of contributors u (allowing for silent alleles).

Hypothesis	*Without Silent*	*With Silent Allele*	
		Exact	$2p$ *Rule*
s & v & $2u$	0.0017	0.0039	0.0836
s & $2u$	0.0015	0.0032	0.0598
v & $2u$	0.0015	0.0031	0.0719
$2u$	0.0006	0.0008	0.0027
s & v & u	0.0392	0.0578	0.1886
s & u	0.0271	0.0340	0.0878
v & u	0.0253	0.0315	0.0805
s & v	0.9031	0.8657	0.4251

Note that the likelihood ratio in favor of H_0: $s\&v$ against H_1: $v\&u$, when correctly accounting for a silent allele, is 27.5, as compared with 5.3 based on the $2p$ rule. This clearly shows that in this case the rule recommended by the National Research Council is over conservative. Without accounting for the possibility of a silent allele, the likelihood ratio is 35.7.

So far we have only used qualitative information, namely which allele values are present in the mixture and the other profiles. A more sensitive analysis additionally uses measured "peak areas," which give quantitative information on the amounts of DNA involved. This requires much more detailed modeling, but again this can be effected by means of a Bayesian network (Cowell et al., 2007). Because the mixture proportion `frac` of DNA contributed by one of the parties is a common quantity across markers, we must now handle them all simultaneously within one "super network." Figure 24.7 shows the top-level network for two contributors, involving six markers, each an instance of a lower-level network **marker** as shown in Figure 24.8. This network is an extended version of the one shown in Figure 24.6, incorporating additional structure to model the quantitative peak area information. In particular, the nodes `Aweight` etc. in **marker** are instances

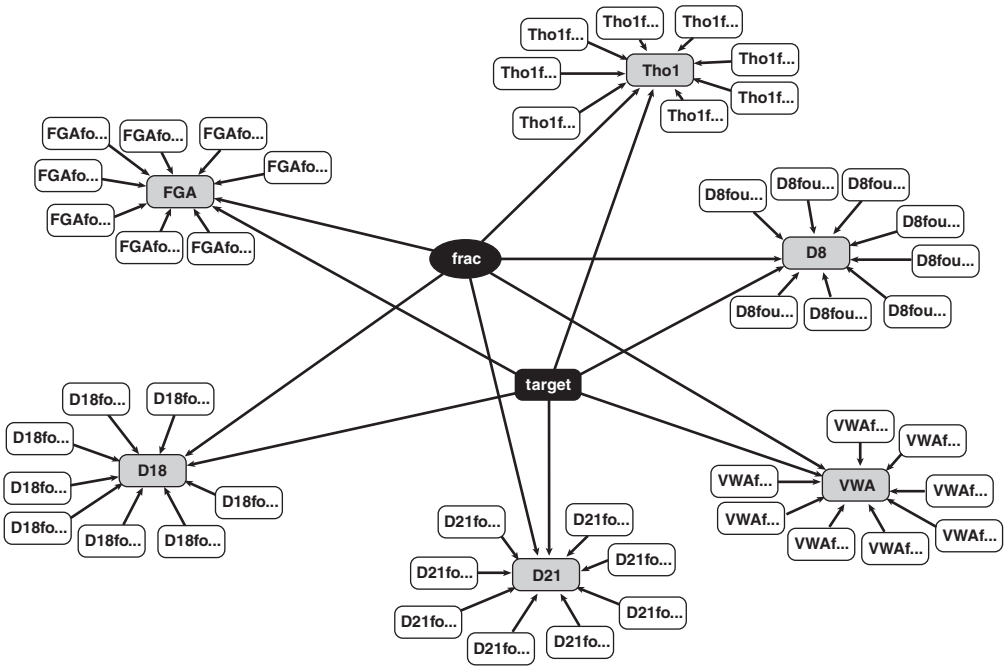

Figure 24.7 Six-marker OOBN for mixture using peak areas, two contributors.

SOURCE: Cowell, R. G., Lauritzen, S. L., & Mortera, J. (2007). Identification and separation of DNA mixtures using peak area information. *Forensic Science International, 166*, 28–34.

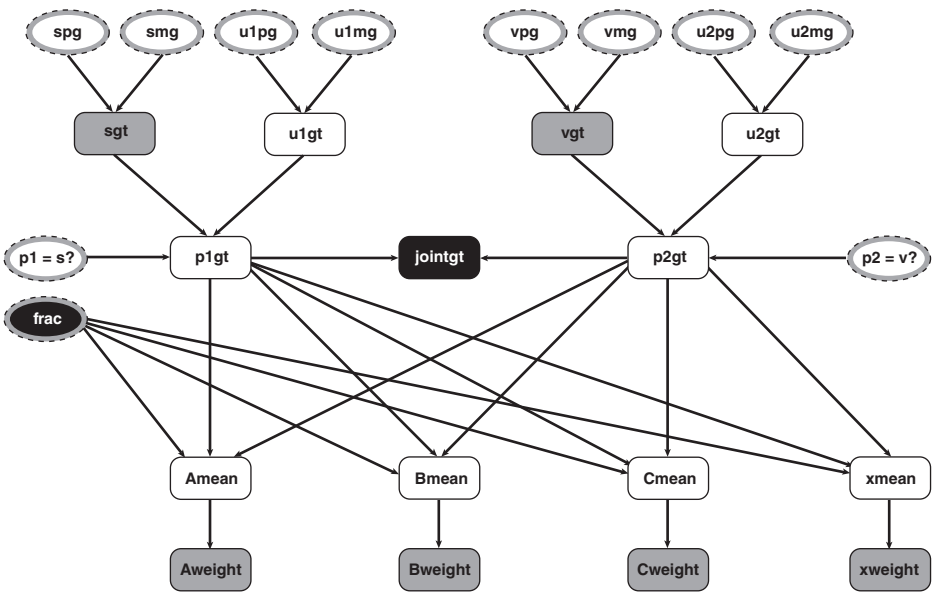

Figure 24.8 Network **marker** with three observed allele peaks.

Table 24.3 Data for mixed trace with two contributors.

Marker	*D8*			*D18*			*D21*			
Alleles	10*	11	14*	13*	16	17	59	65	67*	70*
Peak area	6,416	383	5,659	38,985	1,914	1,991	1,226	1,434	8,816	8,894

Marker	*FGA*			*THO1*		*vWA*			
Alleles	21*	22*	23	8*	9.3*	16*	17	18*	19
Peak area	16,099	10,538	1,014	17,441	22,368	4,669	931	4,724	188

NOTE: The starred values are the suspect's alleles.

of a class network that models the quantitative information on the peak weight.

Cowell et al. (2007) analyze the data shown in Table 24.3, taken from Evett, Gill, and Lambert (1998), involving a six-marker mixed profile with between two and four distinct observed bands per marker and a suspect whose profile is contained in these. It is assumed that this profile is a mixture either of the suspect and one other unobserved contributor or of two unknowns. Using only the allele values as data, the likelihood ratio for the suspect being a contributor to the mixture is calculated to be around 25,000. On taking account of the peak areas also, this rises to about 170,000,000.

CONCLUSION

We hope we have stimulated the reader's interest in the application of probability and statistical reasoning to forensic science. There are many challenging logical subtleties, ambiguities, and probabilistic pitfalls in legal reasoning, some of which we have illustrated. Some of the issues arising in this context have valuable lessons for other applications of statistics, such as confidentiality of census data (Skinner, 2007).

We have also aimed to show the usefulness of Bayesian networks for representing and solving a wide variety of complex forensic problems. Both genetic and nongenetic information can be represented in the same network. A particularly valuable feature is the modular structure of Bayesian networks, which allows a complex problem to be broken down into simpler structures that can then be pieced back together in many ways, so allowing us to address a wide range of forensic queries. In particular, using OOBN we have constructed a flexible computational toolkit and used it to analyze complex cases of DNA profile evidence, accounting appropriately for such features as missing individuals, mutation, silent alleles, and mixed DNA traces.

REFERENCES

Anderson, T. J., Schum, D. A., & Twining, W. L. (2005). *Analysis of evidence* (2nd ed.). Cambridge, UK: Cambridge University Press.

Baio, G., & Corradi, F. (2006). Handling manipulated evidence. *Forensic Science International*, *169*, 181–187.

Balding, D. J., & Donnelly, P. J. (1995). Inference in forensic identification (with discussion). *Journal of the Royal Statistical Society Series A*, *158*, 21–53.

Balding, D. J., & Nichols, R. A. (1994). DNA profile match probability calculation: How to allow for population stratification, relatedness, database selection and single bands. *Forensic Science International*, *64*, 125–140.

Buckleton, J. S., Triggs, C. M., & Walsh, S. J. (Eds.). (2005). *Forensic DNA evidence interpretation.* Boca Raton, FL: CRC Press.

Cavallini, D., & Corradi, F. (2005). OOBN for forensic identification through searching a DNA profiles database. In R. G. Cowell & Z. Ghahramani (Eds.), *Proceedings of the workshop on artificial intelligence and statistics 2005* (pp. 41–48). Key West, FL: Society for Artificial Intelligence and Statistics. (Retrieved from http://www.gatsby.ucl.uk/aistats)

Cowell, R. G., Dawid, A. P., Lauritzen, S. L., & Spiegelhalter, D. J. (1999). *Probabilistic networks and expert systems*. New York: Springer.

Cowell, R. G., Lauritzen, S. L., & Mortera, J. (2007). Identification and separation of DNA mixtures using peak area information. *Forensic Science International*, *166*, 28–34.

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*, *57*, 976–980.

Dawid, A. P. (2003). An object-oriented Bayesian network for estimating mutation rates. In C. M. Bishop & B. J. Frey (Eds.), *Proceedings of the ninth international workshop on artificial intelligence and statistics* (pp. 41–48). Key West, FL: Society for Artificial Intelligence and Statistics. (Retrieved from http://www.tinyurl.com/39bmh)

Dawid, A. P. (2004). Which likelihood ratio? In discussion of Meester and Sjerps (2004). *Law, Probability and Risk*, *3*, 65–71.

Dawid, A. P. (2005). *Probability and proof.* (Online appendix to Anderson et al. (2005). Retrieved from http://www.tinyurl.com/7g3bd)

Dawid, A. P. (2007). Statistics and the law. In K. Tybjerg, J. Swenson-Wright, & A. Bell (Eds.), *Evidence.* Cambridge, UK: Cambridge University Press.

Dawid, A. P., & Evett, I. W. (1997). Using a graphical method to assist the evaluation of complicated patterns of evidence. *Journal of Forensic Sciences*, *42*, 226–231.

Dawid, A. P., & Mortera, J. (1995). In discussion of Balding and Donnelly (1995). *Journal of the Royal Statistical Society Series A*, *158*, 46.

Dawid, A. P., & Mortera, J. (1996). Coherent analysis of forensic identification evidence. *Journal of the Royal Statistical Society Series B*, *58*, 425–443.

Dawid, A. P., Mortera, J., Dobosz, M., & Pascali, V. L. (2003). Mutations and the probabilistic approach to incompatible paternity tests. In B. Brinkmann & A. Carracedo (Eds.), *International Congress series: Vol. 1239. Progress in forensic genetics 9.* Amsterdam: Elsevier Science.

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

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

Dawid, A. P., Mortera, J., & Vicard, P. (2006). Representing and solving complex DNA identification cases using Bayesian networks. In A. Amorim, F. Corte-Real, & N. Morling (Eds.), *International Congress Series: Vol. 1288. Progress in forensic genetics 11* (pp. 484–491). Amsterdam: Elsevier.

Dawid, A. P., Mortera, J., & Vicard, P. (2007). Object-oriented Bayesian networks for complex forensic DNA profiling problems. *Forensic Science International*, *169*, 195–205.

Eggleston, R. (1983). *Evidence, proof and probability* (2nd ed.). London: Weidenfeld & Nicolson.

Evett, I. W., Gill, P. D., & Lambert, J. A. (1998). Taking account of peak areas when interpreting mixed DNA profiles. *Journal of Forensic Sciences.*

Evett, I. W., & Weir, B. S. (1998). *Interpreting DNA evidence*. Sunderland, MA: Sinauer.

Fairley, W. B., & Mosteller, F. (1977). A conversation about Collins. In W. B. Fairley & F. Mosteller (Eds.), *Statistics in public policy* (pp. 369–379). Reading, MA: Addison-Wesley.

Lauritzen, S. L., & Mortera, J. (2002). Bounding the number of contributors to mixed DNA stains. *Forensic Science International*, *130*, 125–126.

Meester, R. W. J., & Sjerps, M. (2004). Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence (with discussion by A. P. Dawid, D. J. Balding, J. S. Buckleton, and C. M. Triggs). *Law, Probability and Risk*, *3*, 51–86.

Mortera, J. (2003). Analysis of DNA mixtures using Bayesian networks. In P. J. Green, N. L. Hjort, & S. Richardson (Eds.), *Highly structured stochastic systems* (pp. 39–44). Oxford, UK: Oxford University Press.

Mortera, J., Dawid, A. P., & Lauritzen, S. L. (2003). Probabilistic expert systems for DNA mixture profiling. *Theoretical Population Biology*, *63*, 191–205.

National Research Council. (1996). *The evaluation of forensic DNA evidence*. Washington, DC: National Academy Press.

Skinner, C. J. (2007). The probability of identification: Applying ideas from forensic statistics to disclosure risk assessment. *Journal of the Royal Statistical Society Series A*, *170*, 195–212.

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

Taroni, F., Aitken, C., Garbolino, P., & Biedermann, A. (2006). *Bayesian networks and probabilistic inference in forensic science*. Chichester, UK: Wiley.

Vicard, P., & Dawid, A. P. (2004). A statistical treatment of biases affecting the estimation of mutation rates. *Mutation Research*, *547*, 19–33.

Vicard, P., Dawid, A. P., Mortera, J., & Lauritzen, S. L. (2008). Estimating mutation rates from paternity casework. *Forensic Science International: Genetics*, *2*, 9–18.

Wigmore, J. H. (1937). *The science of judicial proof* (3rd ed.). Boston: Little, Brown.

Zabell, S. L. (1988). The probabilistic analysis of testimony. *Journal of Statistical Planning and Inference*, *20*, 327–354.