FORENSIC IDENTIFICATION THEN AND NOW

Julia Mortera

Università Roma Tre, Rome, Italy

A. Philip Dawid

University of Cambridge, UK

Abstract *Here we first give a summary of some historical cases where a statistical analysis was used in court cases: a case in a Chinese handbook dated 1247; a disputed inheritance dated 1865; the identification of inscriptions dated 30 AD; forensic identification in the Talmud; the miscarriage of justice in the Dreyfus affair. We then show the important impact that Bayesian networks have had on the solution of complex forensic identification problems.*

Keywords: *Bayesian networks, database search, disputed paternity, DNA mixtures, likelihood ratio, statistics and the law.*

1. FORENSIC STATISTICS: THEN

In this first part of the paper we will review some historical cases where statistics was first introduced in fact finding and identification issues concerning the law.

1.1 THE BEGINNINGS OF FORENSIC SCIENCE

The first recorded use of forensics in the solution of a crime comes from a Chinese handbook for coroners called *The Washing Away of Wrongs*, produced in 1247. One of the many case studies it contains follows the investigation of a roadside stabbing. The coroner examined the slashes on the victim's body, then tested an assortment of blades on a cow carcass. He concluded that the murder weapon was a sickle. But knowing what caused the wounds was a long way from identifying whose hand had wielded the blade, so he turned to possible motives. The victim's possessions were intact, which ruled out robbery. According to his widow, he had no enemies. The best lead was the revelation that the victim hadn't been able to satisfy a man who had recently demanded the repayment of a debt.

The coroner accused the moneylender, who denied the charge. But, tenacious as any TV detective, he ordered all 70 adults in the neighbourhood to stand in a line,

their sickles at their feet. There were no visible traces of blood on any of the sickles. But within seconds a fly landed enthusiastically on the moneylender's blade, attracted by minute traces of blood. A second fly followed, then another. When confronted again by the coroner, the moneylender gave a full confession. He'd tried to clean his blade, but his attempt at concealment had been foiled by the insect informers humming quietly at his feet.

1.2 HANDWRITING IDENTIFICATION

Possibly one of the earliest uses of probabilistic reasoning in the analysis of forensic identification evidence dates back to 1865 with the Howland will case. Sylvia Ann Howland died in 1865, leaving roughly half her estate of some 2 million dollars to various legatees, with the residue to be held in trust for the benefit of Hetty Robinson, Sylvia Howland's niece. Hetty claimed her right to inherit the entire estate on the basis of an earlier will. The Executor contended that two of the three signatures on the will Hetty provided were forged. In the ensuing case of Robinson v. Mandell, Benjamin Peirce analysed the contested signature of Sylvia Ann Howland on her will. He compared 30 downstrokes in the signature with corresponding downstrokes from a different genuine signature. He showed that under a binomial model this amount of agreement was highly improbable. He showed that the number of overlapping downstrokes between two signatures closely followed the binomial distribution, considering that each downstroke was an independent event. When the admittedly genuine signature on the first page of the contested will was compared with that on the second, all 30 downstrokes coincided, suggesting that the second signature was a tracing of the first.

Benjamin Peirce took the stand and asserted that, given the independence of each downstroke, the probability that all 30 downstrokes should coincide in two genuine signatures was $\frac{1}{2.666\times10^{21}}$. He assumed a probability of coincidence roughly equal to 0.0227. He went on to observe "So vast improbability is practically an impossibility. Such evanescent shadows of probability cannot belong to actual life. They are unimaginably less than those least things which the law cares not for. [*omissis*] The coincidence which has occurred here must have had its origin in an intention to produce it. It is utterly repugnant to sound reason to attribute this coincidence to any cause but design."

1.3 TALPIYOT TOMBS

In 1980, a burial tomb was unearthed in East Talpiyot neighbourhood of Jerusalem containing ossuaries (limestone coffins) bearing inscriptions thought to be "Yeshua son of Yehosef", "Marya" and "Yoseh"; names which match those of New Testament (NT) figures. However, these names were in common use at those times. Feuerverger (2008) analyses the plausibility that the names inscribed on the ossuaries match those of the New Testament (NT) figures. The evidence on which the analysis is based is the distribution of names (onomasticon) in the era when the tomb was dated, around 30 AD. Mortera and Vicard (2008) show an example of how an object-oriented Bayesian network could have been used for evaluating the weight of two pieces of identification inference relevant to the ossuary findings: that from onomasticon together with that from DNA profiling of the bones found in the named tombs.

1.4 FORENSIC IDENTIFICATION IN THE TALMUD

In the context of identification the Talmud law gives guidelines on objects whose identity is unknown and reference is made to the likelihood that they derive from a specific source in order to determine their legal status, *i.e.* whether they be permitted or forbidden, ritually clean or unclean, *etc.* (Rabinovitch, 1969). For example, only meat which has been slaughtered in the prescribed manner is *kasher, i.e.* permitted for food. In *Hullin 11a* the rule is "Follow the majority", so when finding, for example, a piece of unidentified meat and there are nine shops selling *kasher* meat and only one not, then the meat can be eaten.

In the law of inheritance the Talmud evaluates the probability that a pregnant woman whose husband dies will bear a live male child. This probability must be less than one-half; as given in *Yevamoth 119a*: *A minority [of pregnant women] miscarry and of all the live births half are male and half female. Add the minority of those who miscarry to the half who bear females and the males are in a minority*.

A ruling of the 2nd century that states: *(Yevamoth 64b)*: *[A mother] had one child circumcised and he died; a second one and he died; one must not circumcise the third*, shows that the independence assumption of certain events cannot be assumed, unlike how the independence assumption was erroneously made in the year 2000 in the Sally Clark case (a young woman wrongfully accused of murdering her two babies, based on flawed probabilistic reasoning presented in court by a pediatrician Professor Sir Roy Meadow). For further details on the Sally Clark case see `(http://www.sallyclark.org.uk), (http://www.statslab.cam.ac.uk/~apd/SallyClark_report.doc).`

1.5 L'AFFAIRE DREYFUS

Alfred Dreyfus was born in 1859 in Mulhouse, then located in Alsace, into a prosperous Jewish family. He left his native town for Paris in 1871 in response to the annexation of the province by Germany following the Franco-Prussian War.

In 1894, while an artillery captain for the General Staff of France, Alfred Dreyfus was suspected of providing secret military information to the German government and was condemned to life imprisonment for espionage on Devil's Island.

Alphonse Bertillon was a witness for the prosecution in the Dreyfus affair in 1894 and again in 1899. (Bertillon founded the first laboratory for criminal identification in France. He was head of the identification branch of the Préfecture of Police. He applied anthropometry or Bertillonage to law enforcement and created an identification system based on physical measurements as well as many other forensic techniques). He testified as a handwriting expert and claimed that the incriminating document (known as the "bordereau") contained strong evidence pointing to Dreyfus's handrwriting, and where it differed the discrepencies were deliberate. However, he was not a handwriting expert, and his convoluted and flawed evidence was a significant contributing factor to one of the most infamous miscarriages of justice. Using a complex system of measurements (based on geometric transformations, probability calculus and military cryptography), he attempted to prove that Dreyfus had disguised his handwriting and forged an imitation of his own handwriting. Both courts martial accepted Bertillon's analysis, and Dreyfus was convicted. The verdict of the second court martial caused a huge scandal, and it was eventually overturned.

"J'accuse…!" was an open letter published on in January 1898 in the newspaper L'Aurore by the influential writer Émile Zola. In the letter, Zola addressed the President of France, Félix Faure, and accused the government of anti-semitism and the unlawful jailing of Alfred Dreyfus. Zola pointed out judicial errors and lack of serious evidence. The letter was printed on the front page of the newspaper and caused a stir in France and abroad. Zola was prosecuted for and found guilty of libel on 23 February 1898. To avoid imprisonment, he fled to England, returning home in June 1899.

Bertillon used the notion from military cryptography that coded messages are often written using a "key" (here using the word *intérêt*) repeated many times. He overlaid the bordereau on the "key" and saw that many letters matched, but many did not. So he made a second almost identical key except for vertical lines set to be the same distance apart as those of the bordereau. He now placed the bordereau first over one and then over the other key, shifted over by one letter, and counted the

letters *e, n, r, t* (the most frequent in French writing) that matched the same letters in the word *intérêt*. He computed the expected values of the most frequent letters in French writing, according to the frequency of their occurrence in the key. Now, the bordereau contained about 800 letters and, for example, 60 were *r*'s, so one would expect to find 1/7 (there being 1 *r* in the 7 letters of intérêt) of the *r*'s in the bordereau lying over an *r* in the key, *i.e.* between 8 or 9 *r*'s. Whereas, Bertillon found 17.

In 1906, Dreyfus appealed his case again, to obtain the annulment of his guilty verdict. The fallacy in Bertillon's reasoning was revealed only then in the Court of Appeal by three mathematicians, Poincaré, Appell and Darboux. They stated "If one takes certain coincidences as evidence, and one shows that there had been a priori few chances for those to happen, have we the right to conclude that they cannot be the effect of chance?" Furthermore, they showed that by using two keys, Bertillon was basically doubling the probability of coincidence of certain letters being overlaid. The verdict was overturned and Dreyfus was also awarded the Cross of the *Légion d'Honneur*, which stated, "a soldier who has endured an unparallelled martyrdom."

Charles M. F. W. Esterhazy was a commissioned officer in the French armed forces and was a spy for the German Empire and the actual perpetrator of the act of treason of which Captain Alfred Dreyfus was wrongfully accused and convicted. When later comparing the writing of Esterhazy with the bordereau even Bertillon stated that there was a perfect match.

Furthermore, Bertillon stated that there were four coincidences out of the 26 initial and final letters of the 13 repeated polysyllabic words in the document. He evaluated the probability of an isolated coincidence as 0.2 and calculated a probability of $0.2^4 = 0.0016$ that four such coincidences would occur in normal writing. But 0.2^4 is the probability of four coincidences out of four; that of four or more out of 13 is some 400 times greater, approximately 0.7. For further details on *L'Affaire Dreyfus* see Schneps and Colmez (2013).

2. FORENSIC STATISTICS: NOW

Legal applications of probabilistic and statistical reasoning have a long history, some examples have been given in Section § 1. For an excellent overview of the history of probabilistic reasoning in courts, see Zabell, (1988).

Forensic statistics is experiencing a period of rapid change because of the tremendous evolution in DNA profiling, DNA database searching, cybersecurity *etc.* DNA evidence has transformed the proof of identity in criminal litigation, but it has also introduced daunting problems of statistical analysis into the process.

Since the pioneering work of Jeffreys et al. (1985), genetic fingerprinting or DNA profiling has become an indispensable tool for identification of individuals in the investigative and judicial process associated with criminal cases, in paternity and immigration cases, and in other contexts. Problems of forensic identification from DNA evidence can become extremely challenging, both logically and computationally, in the presence of complicating features such as missing data on individuals, mixed DNA trace evidence, heterogeneous populations, mutation *etc*. The features of probabilistic expert systems (PES) and Bayesian networks (BNs) have been exploited to handle a wide variety of complex problems of forensic identification such as: kinship analysis (Corradi and Ricciardi, 2013; Corradi et al., 2003; Dawid et al., 2002a; Dawid et al., 2007; Vicard et al., 2008; Green and Mortera, 2009; Aitken and Taroni, 2004; Taroni et al., 2010); integrating forensic information from various sources (Taroni et al., 2014; Biedermann et al., 2008; Garbolino and Taroni, 2002); identification of individuals in DNA mixtures with models based on discrete allelic information (Mortera, 2003; Mortera et al., 2003; Lauritzen and Mortera, 2002) as well as continuous gamma distributed peak height information (Cowell et al., 2006; Cowell et al., 2007a; Cowell et al., 2011; Cowell et al., 2007b).

Besides the main applications in DNA evidence evaluation, *i.e.* inference of source and relatedness testing as discussed in the previous paragraphs, Bayesian networks have also been developed for the study of a variety of further topics that gravitate around the evaluation of forensic analyses such as cross-transfer evidence in criminal cases and error rates.

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; nor that explicit probability arguments should become routine in courts of law. In the previous sections we gave some historical cases where incorrect probabilistic arguments were used in courts, leading to the condemnation of innocent suspects.

Nowadays, there are however increasing numbers of cases – such as DNA identification, or the Sally Clark case – where evidence about probabilities is clearly relevant, and the court would stand to benefit from advice about how to handle them.

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.

3. PROBABILITY LOGIC

In a case at law, let ε 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 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: $\Pr(H_0 \mid \varepsilon)$ and $\Pr(H_1 \mid \varepsilon)$. 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 $\Pr(\varepsilon \mid H_0)$ and $\Pr(\varepsilon \mid H_1)$: the probability that the evidence would have arisen, under each of the competing scenarios.

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

$$\frac{\Pr\left(H_1 \middle| \varepsilon\right)}{\Pr\left(H_0 \middle| \varepsilon\right)} = \frac{\Pr\left(H_1\right)}{\Pr\left(H_0\right)} \times \frac{\Pr\left(\varepsilon \middle| H_1\right)}{\Pr\left(\varepsilon \middle| H_0\right)}. \tag{1}$$

The left-hand side of (1) is the *posterior odds* for comparing H_1 and H_0, given the evidence ε: this is a simple transformation of $\Pr(H_1 \mid \varepsilon)$, the desired *posterior probability* of H_1.

The second term on the right-hand side of (1) is constructed out of the directly assessed terms $\Pr(\varepsilon \mid H_0)$ and $\Pr(\varepsilon \mid H_1)$: it is the *likelihood ratio* (for H_1, as against H_0) engendered by the evidence ε. It is noteworthy that only the ratio of these terms enters, their absolute values being otherwise irrelevant.

To complete (1) we need the term $\Pr(H_1)/\Pr(H_0)$, the *prior odds* for comparing H_1 and H_0 (*i.e.*, before the evidence ε 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 (1).

We can express (1) in words as:

POSTERIOR ODDS = PRIOR ODDS × LIKELIHOOD RATIO.

When ε denotes all the evidence in the case, all the probabilities in (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 formalising the legal doctrine of "presumption of innocence" (which of course is not the same as an *assumption* of innocence). When ε denotes a piece of evidence presented in mid-process, all the probabilities in (1) must be conditioned on the evidence previously presented: in particular, the "prior" probabilities could themselves have been calculated using (1), as posterior probabilities based on earlier evidence.

Notwithstanding the unarguable correctness of (1), it is often replaced by other, more "intuitive", probabilistic arguments, that can be very misleading.

3.1 THE PROSECUTOR'S FALLACY

In a criminal trial, an item of evidence ε may be offered in proof of the guilt, G, of a defendant S, on the basis that the probability of ε would be very low if S were not guilty ($\overline{G}$). For example, in the trial of Sally Clark for double infanticide (Dawid, 2005; Dawid, 2008), 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 this as 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 $\Pr(\varepsilon \mid \overline{G})$ as $\Pr(\overline{G} \mid \varepsilon)$. 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" (Thompson and Schumann, 1987).

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.

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

3.2 FORENSIC IDENTIFICATION

A particularly fertile field where the prosecutor's fallacy flourishes is that of *identification evidence*. Here, unlike the case for 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 ε is presented which bears on this. This may be, for example, eye-witness evidence (as in the celebrated "Collins case" (Fairley and 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, fibres, footprints, fingerprints, bitemarks, 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.

3.3 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.

Computerised search typically allows us to identify every individual in the database whose DNA profile matches the crime trace. Let P denote the initial probability that an individual matches, independently for different individuals. 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 $1/NP$ 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 towards 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.

4. 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 ε 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 ε available, from all sources, is once again crystallised in the *likelihood ratio*, $\mathrm{LR} = \Pr(\varepsilon \mid H_1)/\Pr(\varepsilon \mid H_0)$. If we wish to compare more that two hypotheses, we require the full *likelihood function*, a function of the various hypotheses H being entertained (and of course the evidence ε):

$$\mathrm{lik}(H) \propto \Pr(\varepsilon \mid H). \tag{2}$$

The proportionality sign in (2) indicates that we have omitted a factor that does not depend on H , although it can depend on ε . 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:

$$\Pr(H \mid \mathcal{E}) \propto \Pr(H) \times \text{lik}(H). \tag{3}$$

Again the omitted proportionality factor in (3) 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, $\sum_H \Pr(H \mid \mathcal{E})=1$, is satisfied.

4.1 GENETIC BACKGROUND

To proceed further we need some basic genetic facts about DNA profiles, which we summarise very briefly below: for more details see *e.g.* Butler (2005).

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. CAGT*) in the 4-letter alphabet of the genetic code is repeated exactly 5 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]

4.2 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 *non-paternity*: 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 1.

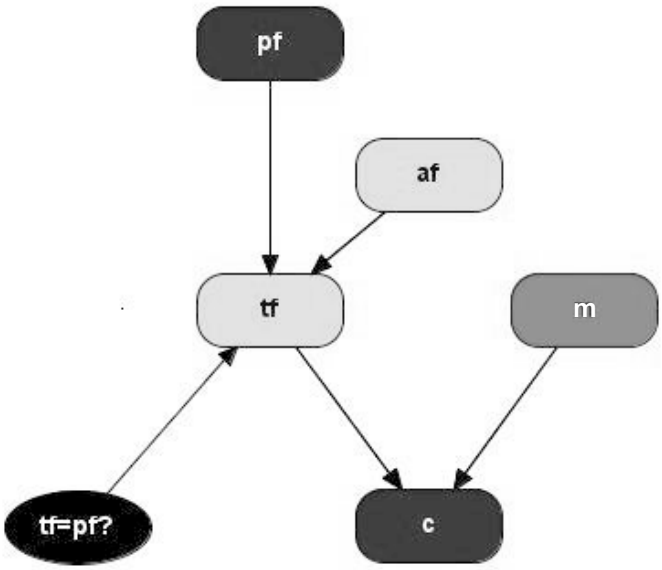

Figure 1: Pedigree for simple disputed paternity.

[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 and Nichols, 1994; Dawid and Mortera, 1996; Green and Mortera, 2009).

Because of our independence assumptions, we can analyse 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_1, we just apply Mendel's laws of segregation; under non-paternity, H_0 we require (estimates of) the frequencies of relevant marker alleles among the population. Using (1) this can then be combined with the prior odds of paternity, based on external background evidence *B*, in order 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 `pfgt`= {10, 12}. The estimated population frequencies of alleles 10 and 12 are, respectively, 0.284 and 0.260. In this case, conditioning on the genotypes of mother and putative father (which makes no difference to the answer), 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 favour of paternity, based on marker *D*7 alone, is 1.93.

4.3 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.

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 ε for the case. Suppose we wish to compute the likelihood ratio in favour of the hypothesis that the victim and suspect contributed to the mixture: H_1: v & s, as against the hypothesis that the victim and an unknown individual u contributed to

the mixture: H_0: *v* & *u*. It is not difficult to show that in this case the LR is

$$\mathrm{LR} = \frac{1}{p_A^2 + 2p_A p_B + 2p_A p_C}, \tag{4}$$

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

5. 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 formulae.

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 § 5.5 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 (4) become relatively complex when the number of potential contributors to the mixture becomes large; or if we want to use quantitative data (peak areas), which contain important additional information about the composition of the mixture; or to allow for uncertainty in allele frequencies and/or population substructure.

To handle such cases sophisticated probabilistic modelling 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" (OOBN) form, as implemented in commercial software such as Hugin 6 [3]. Bayesian networks for evaluating DNA evidence were introduced by Dawid et al. (2002). Further description and developments can be found in Mortera (2003); Mortera et al. (2003); Vicard et al. (2004); Cowell et al. (2004); Dawid et al. (2006); Dawid et al. (2007); 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.

5.1 SIMPLE DISPUTED PATERNITY

We use the example in § 4.2 of simple disputed paternity to introduce some basic ingredients of forensic OOBNs.

---

[3] Obtainable from `www.hugin.com`

In fact Figure 1 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 1 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); 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 alleles and their associated frequencies in the relevant population.

The internal structure of **child** is displayed in Figure 2.

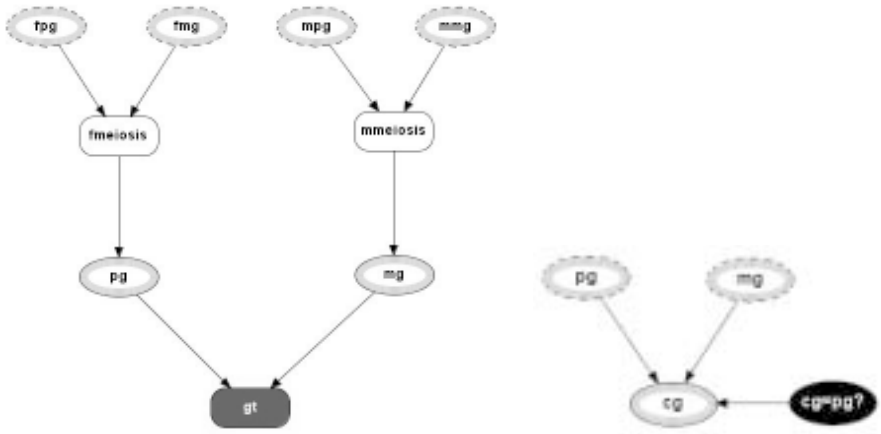

Figure 2: Networks child and meiosis.

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_1 (`tf=pf`) when it takes the value *true* and H_0 (`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.

5.2 COMPLEX DISPUTED PATERNITY

Figure 3 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.

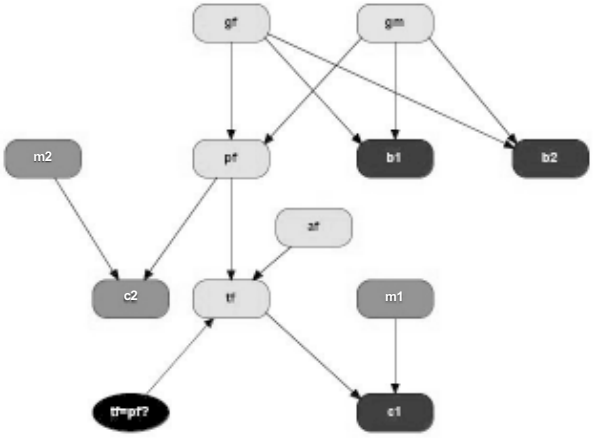

Figure 3: Pedigree for incomplete paternity case.

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 ε 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 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 1300, meaning that the observed DNA evidence is 1300 times more probable on the hypothesis of paternity than it would be were we to assume non-paternity. According to Evett and Weir (1998), 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.

5.3 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 et al., 2001; Dawid et al., 2003; Dawid, 2003; Vicard and Dawid, 2004; Vicard et al., 2004). 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 2, as shown in Figure 4, 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.

Once an appropriate network **mut** has been built, and **meiosis** modified as described above, pedigree networks constructed as in Sections 5.1 or 5.2 will now automatically incorporate the additional possibility of mutation.

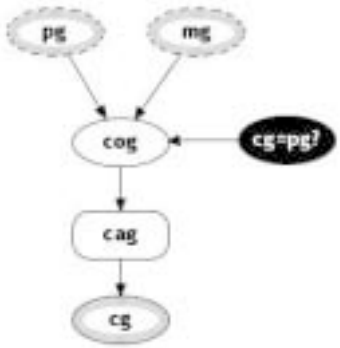

Figure 4: Revised network meiosis, incorporating mutation.

5.4 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; Dawid et al., 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 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 impact 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 labelled 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 favour 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 `bgt`={12, 12}, the overall paternity ratio LR = $L_D \times L_B$ achieves a maximum value of 1027.3, at `pr(silent)` = 0.0000642, even though it vanishes for pr(silent) = 0.

Table 1: Disputed paternity with brother too. `mgt`= {12, 15}, `pfgt`= {14, 14}, `cgt`= {12, 12}. Likelihood ratio in favour of paternity allowing for silent alleles: L_D, without brother's genotype. L_B, further (multiplicative) effect of brother's genotype.

		LB with `bgt` =						
pr(`silent`)	L_D	{16, 20}	{12, 17}	{12, 14}	{14, 17}	{14, 14}	{16, 16}	{12, 12}
0	0	1	1	0.546	0.546	1.0000	6.12	1595
0.000015	0.472	1	1	0.546	0.546	0.9999	6.07	403.7
0.0001	2.473	1	1	0.546	0.546	1	6.13	3334
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

5.5 BAYESIAN NETWORKS FOR ANALYSING MIXED DNA PROFILES

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

Figure 5 shows a top-level network which can be used for analysing a mixture with two contributors, *p*1 and *p*2. 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. Boolean node `p1=s?` represents the hypothesis that contributor *p*1 is the suspect *s*. Node `p1gt`, the genotype of *p*1, is an instance of a network **query** which 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 of 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.

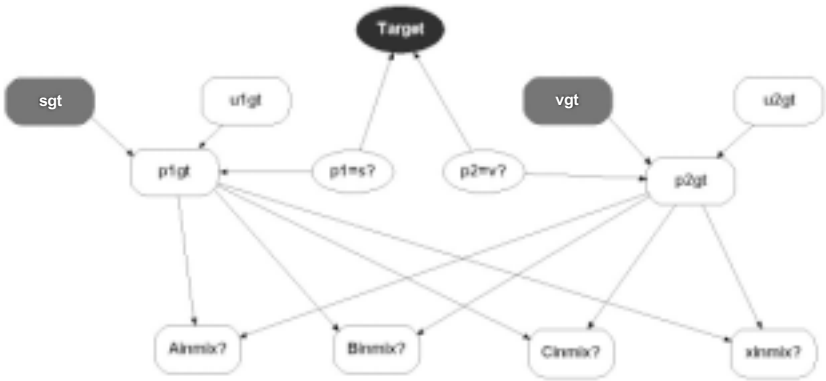

Figure 5: Bayesian network for DNA mixture from two contributors.

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 can not 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 $\Pr_x(\varepsilon \mid H)$ for the probability of the evidence under this hypothesis. Thus the likelihood ratio LR needed to evaluate the DNA evidence ε – comprising the DNA profiles of the victim, the suspect and the mixed trace – in favour of a hypothesis H_1 against an alternative hypothesis H_0 is

$$\mathrm{LR} = \frac{\Pr_{x_1}\left(\varepsilon \middle| H_1\right)}{\Pr_{x_0}\left(\varepsilon \middle| H_0\right)},$$

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 favourable 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 favourable 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 $\Pr_x(\varepsilon \mid H)$ can be no larger than the probability that all the alleles of the x unknown contributors are in the mixed trace. This implies (Lauritzen and Mortera, 2002):

$$\Pr_x\left(\varepsilon \middle| H\right) \leq \prod_{m=1}^{M} k_m^{2x}$$

where, for each marker m, k_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_0 is any alternative hypothesis yielding likelihood L_0, we need not consider an alternative hypothesis H with more than $b(L_0)$ unknown contributors, where

$$b\left(y\right) = \frac{\ln y}{2\sum_{m=1}^{M} \ln k_m}$$

since that would yield a likelihood smaller than L_0.

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 2 gives the likelihoods (arbitrarily normalised to sum to 1) based on a network which 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 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).

		with silent allele	
Hypothesis	**without silent**	**exact**	**2*p* rule**
s & *v* & 2*u*	0.0017	0.0039	0.0836
s & 2*u*	0.0015	0.0032	0.0598
v & 2*u*	0.0015	0.0031	0.0719
2*u*	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 favour of H_1: *s* & *v* against H_0: *v* & *u*, when correctly accounting for a silent allele, is 35.7, as compared to 5.3 based on the 2*p* 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 27.5.

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, by means of a Bayesian network (Cowell et al., 2007b). 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 6 shows the top level network for two contributors, involving six markers, each an instance of a lower level network **marker** as shown in Figure 7. This network is an extended version of the one shown in Figure 5, incorporating additional structure to model the quantitative peak area information. In particular, the nodes `Aweight` *etc.* in **marker** are instances of a class network that models the quantitative information on the peak weight.

Cowell et al. (2006, 2007b) analyse the data shown in Table 3, taken from Evett et al. (1998), involving a 6-marker mixed profile with between 2 and 4 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 repeat numbers 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.

Figure 6: 6-marker OOBN for mixture using peak areas, 2 contributors (reproduced from Cowell et al. (2004)).

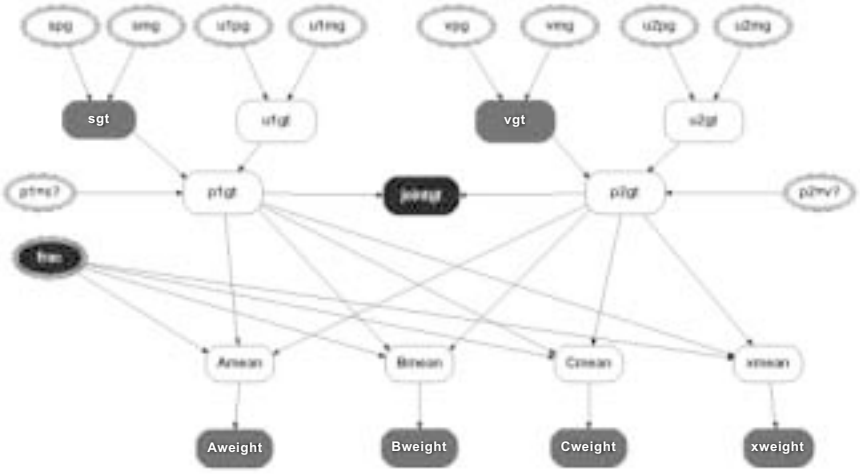

Figure 7: Network marker with three observed allele peaks.

Table 3: Data for mixed trace with two contributors. The starred values are the suspect's alleles.

Marker	**D8**			**D18**			**D21**			
Alleles	10*	11	14*	13*	16	17	59	65	67*	70*
Peak area	6416	383	5659	38985	1914	1991	1226	1434	8816	8894

Marker	**FGA**			**THO1**		**VWA**			
Alleles	21*	22*	23	8*	9.3*	16*	17	18*	19
Peak area	16099	10538	1014	17441	22368	4669	931	4724	188

5.6 RECENT DEVELOPMENTS

Cowell et al. (2015) develop a statistical model for the quantitative peak information obtained from an electropherogram of a forensic DNA sample which works directly with the peak height information, and allows the introduction of a threshold such that the dropout of an allele is interpreted as failure for its associated peak to be observed above the threshold. Another common artefact is stutter, whereby an allele that is present in the sample is mis-copied at some stage in the amplification process. Another artefact is known as dropin, referring to the occurrence of small unexpected peaks in the DNA amplification. This can for example be due to sporadic contamination of a sample either at source or in the forensic laboratory. Current technology allows for the amplification of very small amounts of DNA, even as little as contained within one cell. In these cases many of these artefacts can occur. These artefacts are simply represented in a coherent way in this model.

The parameters of the model, and their standard errors, are estimated by maximum likelihood in the presence of multiple unknown contributors, exploiting a Bayesian network representation for efficient computation. The model can efficiently both find likelihood ratios for evidential calculations, and deconvolve the mixtures for the purpose of finding likely profiles of one or more unknown contributors to a DNA mixture. It is readily extended to simultaneous analysis of more than one mixture where one can see that the combination of evidence from several traces may give an evidential strength close to that of a single source trace and thus this modeling of peak height information provides for a very efficient mixture analysis. A gamma model is used for the peak heights which is based on Cowell et al. (2007a, 2011).

Recently Mortera et al. (2016) applied this model to analyse a complex disputed paternity case, where the DNA of the putative father was extracted from his corpse that had been inhumed for over 20 years. This DNA was contaminated and appeared to be a mixture of at least two individuals. Furthermore, the mother's DNA was not available. The DNA mixture was analysed so as to predict the most

probable genotypes of each contributor. The major contributor's profile was then used to compute the likelihood ratio for paternity. We also showed how to take into account a dropout allele and the possibility of mutation in paternity testing.

6. CONCLUSIONS

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 non-genetic 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 object-oriented Bayesian networks we have constructed a flexible computational toolkit, and used it to analyse complex cases of DNA profile evidence, accounting appropriately for such features as missing individuals, mutation, silent alleles and mixed DNA traces.

ACKNOWLEDGMENTS

Philip Dawid was supported by an Emeritus Fellowship from the Leverhulme Trust. We thank Joe Gastwirth for useful discussions.

REFERENCES

Aitken, C.G. and Taroni, F. (2004). *Statistics and The Evaluation of Evidence for Forensic Scientists*. 2nd edition. John Wiley & Sons.

Anderson, T.J., Schum, D.A. and Twining, W.L. (2005). *Analysis of Evidence*, (second edn). Cambridge University Press. With on-line appendix on *Probability and Proof* by Philip Dawid at http://tinyurl.com/7g3bd.

Balding, D.J. and 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.

Biedermann, A., Bozza, S. and Taroni, F. (2008). Decision theoretic properties of forensic identification: underlying logic and argumentative implications. *Forensic Science International*, **177**(2): 120-132.

Butler, J.M. (2005). *Forensic DNA Typing: Biology, Technology and Genetics of STR Markers.* Second Edition. Elsevier Academic Press.

Corradi, F. and Ricciardi, F. (2013). Evaluation of kinship identification systems based on short tandem repeat DNA profiles. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **62**(5): 649-668.

Cowell, R.G., Graversen, T., Lauritzen, S.L. and Mortera, J. (2015). Analysis of DNA mixtures with artefacts. *Journal of the Royal Statistical Society: Series C (with discussion)*, **64**: 1-48.

Cowell, R.G., Lauritzen, S.L. and Mortera, J. (2004). Identification and separation of DNA mixtures using peak area information using a probabilistic expert system. Statistical Research Paper 25, Cass Business School, City University.

Cowell, R.G., Lauritzen, S.L. and Mortera, J. (2006). MAIES: A tool for DNA mixture analysis. In *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*, (ed. R. Dechter and T. Richardson editors), pp. 90–7. Morgan Kaufmann Publishers, San Francisco.

Cowell, R.G., Lauritzen, S.L. and Mortera, J. (2007a). A gamma model for DNA mixture analyses. *Bayesian Analysis*, **2**(2): 333-348.

Cowell, R.G., Lauritzen, S.L. and Mortera, J. (2007b). Identification and separation of DNA mixtures using peak area information. *Forensic Science International*, **166**(1): 28-34.

Cowell, R.G., Lauritzen, S.L. and Mortera, J. (2011). Probabilistic expert systems for handling artifacts in complex DNA mixtures. *Forensic Science International: Genetics*, **5**: 202-209.

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

Dawid, A.P. (2005). Probability and Proof. On-line Appendix to Anderson et al. (2005). `http://tinyurl.com/7g3bd.`

Dawid, A.P. (2008). Statistics and the law. In *Evidence*, (ed. K. Tybjerg, J. Swenson Wright, and A. Bell). Cambridge University Press. 119-148

Dawid, A.P. and 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. and Pascali, V.L. (2003). Mutations and the probabilistic approach to incompatible paternity tests. In *Progress in Forensic Genetics 9*, International Congress Series, Vol. 1239, (ed. B. Brinkmann and A. Carracedo), pp. 637–8. Elsevier Science, Amsterdam. International Society for Forensic Genetics. `doi:10.1016/S0531-5131(02)00845-2.`

Dawid, A.P., Mortera, J. and 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. and van Boxel, D.W. (2002). Probabilistic expert systems for forensic inference from genetic markers. *Scandinavian Journal of Statistics*, **29**: 577-795.

Dawid, A.P., Mortera, J. and Vicard, P. (2006). Representing and solving complex DNA identification cases using Bayesian networks. In *Progress in Forensic Genetics 11*, International Congress Series, Vol. 1288, (ed. A. Amorim, F. Corte-Real and N. Morling), pp. 484-91. Elsevier, Amsterdam.

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

Evett, I.W., Gill, P.D. and Lambert, J.A. (1998). Taking account of peak areas when interpreting mixed DNA profiles. *Journal of Forensic Sciences*, **43**: 62-69.

Evett, I.W. and Weir, B.S. (1998). *Interpreting DNA Evidence*. Sinauer, Sunderland, MA.

Fairley, W.B. and Mosteller, F. (1977). A conversation about Collins. In *Statistics in Public Policy*, (ed. W.B. Fairley and F. Mosteller), pp. 369–79. Addison-Wesley, Reading, Massachusetts.

Feuerverger, A. (2008). Statistical analysis of an archeological find. *Annals of Applied Statistics*, **2**(1): 3-54.

Garbolino, P. and Taroni, F. (2002). Evaluation of scientific evidence using Bayesian networks. *Forensic Science International*, **125**(2): 149-155.

Green, P.J. and Mortera, J. (2009). Sensitivity of inferences in forensic genetics to assumptions about founder genes. *Annals of Applied Statistics*, **3**(2): 731-763.

Jeffreys, A. J., Wilson, A. J. and Thein, S. L. (1985). Individual specific 'fingerprints' of human DNA. *Nature*, **316**: 76-79.

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

Mortera, J. (2003). Analysis of DNA mixtures using Bayesian networks. In *Highly Structured Stochastic Systems*, (ed. P.J. Green, N.L. Hjort and S. Richardson), chapter 1B, pp. 39–44. Oxford University Press.

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

Mortera, J., Vecchiotti, C., Zoppis, S. and Merigioli, S. (2016). Paternity testing that involves a DNA mixture. *Forensic Science International: Genetics*, **23**: 50–54 doi: 10.1016/ j.fsigen.2016.02.014.

Mortera, J. and Vicard, P. (2008). Discussion of: Statistical analysis of an archeological find. *Annals of Applied Statistics*, **2**(1): 91-96.

National Research Council (1996). *The Evaluation of Forensic DNA Evidence*. Washington, D.C.: National Academy Press.

Rabinovitch, N.L. (1969). Studies on the history of probability and statistics. XXII: Probability in the Talmud. *Biometrika*, **56**(1): 437 – 441.

Schneps, L. and Colmez, C. (2013). *Math on Trial: How Numbers Get Used and Abused in the Courtroom*. Basic Books, New York, NY.

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.

Taroni, F., Aitken, C., Garbolino, P. and Biedermann, A. (2006). *Bayesian Networks and Probabilistic Inference in Forensic Science*, Statistics in Practice. John Wiley and Sons, Chichester.

Taroni, F., Biedermann, A., Bozza, S., Garbolino, P. and Aitken, C. (2014). *Bayesian Networks for Probabilistic Inference and Decision Analysis in Forensic Science*. John Wiley & Sons.

Taroni, F., Bozza, S., Biedermann, A., Garbolino, P. and Aitken, C. (2010). *Data Analysis in Forensic Science: a Bayesian Decision Perspective*. John Wiley & Sons.

Thompson, W.C. and Schumann, E.L. (1987). Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy. *Law and Human Behavior*, **11**: 167-187.

Vicard, P. and 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. and Lauritzen, S.L. (2004). Estimation of mutation rates from paternity cases using a Bayesian network. Research Report 249, Department of Statistical

Science, University College London.

Vicard, P., Dawid, A.P., Mortera, J. and Lauritzen, S.L. (2008). Estimating mutation rates from paternity casework. *Forensic Science International: Genetics*, **2**(1): 9-18.

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