Book title here
Bayesian Networks in Forensic Science

1.1 Introduction

Solving forensic identification problems frequently requires complex probabilistic argument and computation. The technology of Bayesian Networks (BNs), also called Probabilistic Expert Systems (Cowell et al. 1999), supplies a valuable tool that can streamline and help solve such problems. Aitken and Gammerman (1989) gave one of the first applications of Bayesian networks in forensic science.

Building a Bayesian Network has two stages. We start by constructing a graphical representation of the problem, involving a node for each relevant variable or hypothesis, with connexions between these that encode the way in which they depend, probabilistically, on each other. We then add quantitative information about those probabilistic dependencies. Once the BN is set up in a suitable software environment, one enters the available evidence on the observed variables, and the system will compute the resulting remaining uncertainty about hypotheses of interest.

The aim of this Chapter is to give an overview of the applications of BNs in forensic statistics with a particular emphasis on forensic genetics. This is largely done by example. Some of these are too complex to describe here in all their details, but the interested reader can find those details in the further references.

In §1.2 we recall the basic logical structure of probabilistic argument in the legal context. Section 1.3 introduces BNs by way of two illustrative examples: first, a specific problem of identification from eyewitness evidence, and secondly a more generic structure for forensic identification. In §1.4 we describe a valuable extension of BNs as OOBNs (object-oriented Bayesian networks), whose key features are illustrated with a further elaboration of the problem of identification from eyewitness testimony. In §1.5 we turn to consider the special features of forensic genetics identification problems, and how these can be incorporated into a BN. Illustrative examples include simple criminal identification, disputed paternity, and complex criminal cases involving family relationships. We also describe how a forensic BN can be easily modified to allow for additional complications, such as the possibility of mutation. In §1.6 we show how to construct BNs to analyse mixed DNA profiles, using discrete or continuous information, while in §1.7 we describe networks for representing
Bayesian Networks in Forensic Science

uncertain allele frequencies and heterogeneous reference populations. Some concluding remarks are given in §1.8. The Appendix briefly outlines some of the essential underlying theory of Bayesian Networks.

1.2 Probability Logic

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

The odds form of Bayes’s theorem—a trivial consequence of the definition of conditional probability—tells us that

\[
\frac{\Pr(H_1|E)}{\Pr(H_0|E)} = \frac{\Pr(H_1)}{\Pr(H_0)} \times \frac{\Pr(E|H_1)}{\Pr(E|H_0)}.
\]

The left-hand side of (1.1) is the posterior odds for comparing $H_1$ and $H_0$, given the evidence $E$. When $H_0$ and $H_1$ are considered to exhaust all possibilities, this is a simple transformation of $\Pr(H_1|E)$, the posterior probability of $H_1$.

The second term on the right-hand side of (1.1) is constructed out of the directly assessed terms $\Pr(E|H_0)$ and $\Pr(E|H_1)$: it is the likelihood ratio $LR = \Pr(E|H_1)/\Pr(E|H_0)$ (for $H_1$, as against $H_0$) engendered by the evidence $E$. It is noteworthy that only the ratio of these terms enters, their absolute values being otherwise irrelevant.

To complete (1.1) 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 $E$ 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 (1.1).

We can express (1.1) in words as:

POSTERIOR ODDS = PRIOR ODDS $\times$ LIKELIHOOD RATIO.
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 $E$):

$$L(H) \propto \Pr(E|H).$$  \hfill (1.2)

The proportionality sign in (1.2) indicates that we have omitted a factor that does not depend on $H$, although it can depend on $E$. 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:

$$\Pr(H|E) \propto \Pr(H) \times L(H).$$  \hfill (1.3)

Again the omitted proportionality factor in (1.3) does not depend on $H$, although it might depend on $E$. It can be recovered, if desired, as the unique such factor for which the law of total probability, $\sum_H \Pr(H|E) = 1$, is satisfied.

When $E$ denotes all the evidence in the case, all the probabilities in (1.1) or (1.2) are unconditional; in particular, the prior probabilities 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 $E$ denotes a piece of evidence presented in mid-process, all the probabilities in (1.1) or (1.2) must be further conditioned on the evidence previously presented: in particular, the “prior” probabilities could themselves have been calculated using (1.1) or (1.2), as posterior probabilities based on earlier evidence.

In many cases the relationships between the various relevant items of evidence will be complex and subtle, and taking these relationships properly into account can be far from straightforward. It can then be helpful to structure these relationships in a more detailed way. The theory and technology of Bayesian Networks (BNs) (Cowell et al. 1999) can be useful to assist such structuring. In particular, for handling a variety of complex cases in forensic science, Bayesian networks, together with their associated computational methodology and technology, have been found extremely valuable, particularly in their “object-oriented” (OObN) form, as implemented in commercial software such as Hugin. BNs can be used to assist police in their investigations, to aid a trial lawyer in developing an argument, and (in principle, if as yet rarely in practice) to assist the trier of fact in coming to a conclusion. Bayesian networks for evaluating DNA evidence were introduced by Dawid.

\footnote{Obtainable from \url{https://www.hugin.com}}
For some illustrative cases, we describe below how we can construct a suitable BN or OOBN representation of a complex forensic identification problem, incorporating all the individuals involved and the relationships between them.

1.3 Simple Bayesian Networks for Forensic Problems

A brief formal description of Bayesian Networks is provided in the Appendix. Here we introduce their main characteristics by means of examples.

Example 1 (Eyewitness) We illustrate the simplest non-trivial Bayesian network with an example of a hit-and-run accident.

A cab is involved in a hit-and-run accident at night. Only white and blue cabs operate in the city. 62% of the cabs in the city are white and 38% are blue. An elderly witness identifies the cab as blue. The reliability of the witness is tested under similar conditions: the witness correctly identifies a white cab 80% of the time, and correctly identifies a blue cab 70% of the time.

Qualitative BN

At this point we consider only qualitative, structural aspects, which are well displayed in a BN. As we shall see below, a BN can also incorporate quantitative information on probabilities, and provide a powerful machine for manipulating these.

The outline of this story is represented by the simple Bayesian network of Figure 1.1. There is a node for each of the relevant variables: (the actual) cab colour and (the report of the) eye witness. These each have two possible states, white or blue for cab colour, and “white” or “blue” for eye witness. The arrow indicates that the eye witness report is influenced,
albeit non-deterministically, by the actual colour of the cab. We use a pseudo-
genetic language to describe the dependencies between nodes: cab colour is a “parent” of eye witness, and eye witness is a “child” of cab colour.

An important aspect of this, as of any, BN representation is that the direction of an arrow indicates a stochastic relationship that can be interpreted as, in some intuitive sense, causal in nature. One useful way of thinking about this is that the nature of the dependence should be stable across different ways and contexts in which it might arise. Thus in this example the quoted success rates of the witness are assumed to be essentially the same, both in the test cases presented, and in the actual case at hand.

We will also be interested in inferential relationships; in this particular example, having heard the report of the eye witness, we would like to know what can be inferred about the actual colour of the cab? Some alternative representations, such as Wigmore charts [Wigmore 1937; Anderson et al. 2005; Dawid et al. 2011], would reverse the direction of the arrow in Figure 1.1 so directly representing an inferential, non-causal, relationship. However, inferential relationships are typically not stable across different contexts, which is why we do not use them as basic ingredients. Nonetheless, as we shall see below, a BN with only causal arrows can readily compute such desired inferences.

Quantitative BN

We associate, with each node of the BN, a table describing the conditional probability distribution over its states, conditional on the state(s) of its parent node(s). In Figure 1.1 cab colour is a “founder” node, with no parents, so its table will just contain the unconditional (prior) probabilities for its states, as in Table 1.1.

<table>
<thead>
<tr>
<th>cab colour</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>white</td>
<td>0.62</td>
</tr>
<tr>
<td>blue</td>
<td>0.38</td>
</tr>
</tbody>
</table>

TABLE 1.1: Probability table for cab colour

As for the node eye witness, we need to specify the conditional probability for each of its states, given each state of its parent variable cab colour. This information is presented in Table 1.2 using the conditional probabilities as specified in the story.

The specification of the BN is now complete.

Computation and inference

There exist powerful general algorithms [Lauritzen and Spiegelhalter 1988; Cowell et al. 1999] for performing computations on a Bayesian network, which
Bayesian Networks in Forensic Science

<table>
<thead>
<tr>
<th>cab colour</th>
<th>eye witness</th>
<th>white</th>
<th>blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>“white”</td>
<td>0.8</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>“blue”</td>
<td>0.2</td>
<td>0.7</td>
<td></td>
</tr>
</tbody>
</table>

TABLE 1.2: Conditional probability table for eye witness

are embodied in a number of specialist software systems, such as HUGIN\(^2\), GeNIe\(^3\) and the R software gRain\(^4\) among others. Here we have used HUGIN 8.3. The basic computational routine finds the marginal distribution at every node. The resulting probabilities are as displayed in Figure 1.2. For

![Figure 1.2: Bayesian network for hit-and-run accident showing prior marginal distributions.](https://www.hugin.com/)

---

**eye witness** this simply repeats the input prior probabilities, whereas the figures for **eye witness** show that, even without knowing the actual colour of the cab, we would expect the witness to report “blue” with probability 0.39. This computation automates the formula for “extension of the conversation” (generalised addition law):

\[
\Pr(\text{eye witness} = \text{“blue”}) = \Pr(\text{eye witness} = \text{“blue”} | \text{cab colour} = \text{white}) \\
\times \Pr(\text{cab colour} = \text{white}) \\
+ \Pr(\text{eye witness} = \text{“blue”} | \text{cab colour} = \text{blue}) \\
\times \Pr(\text{cab colour} = \text{blue}) \\
= 0.2 \times 0.62 + 0.7 \times 0.38 \\
= 0.39,
\]

which yields the marginal probability.

By a straightforward variation on the basic computational algorithm, it is also possible to take account of new evidence about the state of one or more variables, thus executing “inference”. In our example, we have obtained the evidence **eye witness** = “blue”. Entering this evidence and “propagating” using the software yields the results shown in Figure 1.3. Taking account of

---

\(^2\)https://www.hugin.com/
\(^3\)https://www.bayesfusion.com/genie/
\(^4\)https://cran.r-project.org/web/packages/gRain/index.html
FIGURE 1.3: Bayesian network for hit-and-run accident, updated with eye witness evidence.

this fallible eye-witness’s evidence, the posterior probability of \textit{cab colour} = \textit{blue} is 68\%, as compared with its prior (unconditional) probability of 38\%. The system has automated Bayes’s Theorem:

\[
\Pr(\text{cab colour} = \text{blue} \mid \text{eye witness} = \text{“blue”}) = \frac{\Pr(\text{cab colour} = \text{blue}) \times \Pr(\text{eye witness} = \text{“blue”} \mid \text{cab colour} = \text{blue})}{\Pr(\text{eye witness} = \text{“blue”})} \tag{1.4}
\]

\[
= \frac{0.38 \times 0.7}{0.39} = 0.68. \tag{1.5}
\]

Suppose however that the prior probabilities in Table 1.1 had been different, for example as in Table 1.3, while the conditional probabilities of Table 1.2, on account of their “causal” interpretation, were preserved. We would then have computed different values:

\[
\Pr(\text{eye witness} = \text{“blue”}) = 0.45 \\
\Pr(\text{cab colour} = \text{blue} \mid \text{eye witness} = \text{“blue”}) = 0.78. \tag{1.6}
\]

In particular, the difference between (1.5) and (1.6) demonstrates that, unlike “causal” probabilities, inferential probabilities are volatile and context-dependent—which is why we insist on using causal arrows and causal conditional probabilities as the basic ingredients when specifying a BN.

\textit{Likelihood ratio}

Even though the actual prior probabilities at the founder node \textit{cab colour} are not equal, as in Table 1.3 it can be useful to proceed, formally, as if they
Bayesian Networks in Forensic Science

were. For then we can use the odds form (1.1) of Bayes’s Theorem:

\[
\begin{align*}
\frac{\Pr(\text{cab colour} = \text{blue} \mid \text{eye witness} = \text{“blue”})}{\Pr(\text{cab colour} = \text{white} \mid \text{eye witness} = \text{“blue”})} &= \frac{\Pr(\text{cab colour} = \text{blue}) \times \Pr(\text{eye witness} = \text{“blue”} \mid \text{cab colour} = \text{blue})}{\Pr(\text{cab colour} = \text{white}) \times \Pr(\text{eye witness} = \text{“blue”} \mid \text{cab colour} = \text{white})} \\
&= \frac{\Pr(\text{cab colour} = \text{blue})}{\Pr(\text{cab colour} = \text{white})} \times \frac{0.7}{0.2}.
\end{align*}
\] (1.7)

The final term in (1.7), \(0.7/0.2 = 3.5\), is the likelihood ratio, constructed entirely of causal probabilities, and so relevant across different contexts.

We see from (1.7) that if we insert, purely formally, equal prior probabilities, \(\Pr(\text{cab colour} = \text{blue}) = \Pr(\text{cab colour} = \text{white}) = 0.5\), the ratio of the resulting formal posteriors output by the software,

\[
\frac{\Pr(\text{cab colour} = \text{blue} \mid \text{eye witness} = \text{“blue”})}{\Pr(\text{cab colour} = \text{white} \mid \text{eye witness} = \text{“blue”})},
\]

will deliver the likelihood ratio. And once this stable element has been extracted, it can be used offline to construct a true posterior, again using (1.7), this time in conjunction with appropriately assessed prior odds,

\[
\frac{\Pr(\text{cab colour} = \text{blue})}{\Pr(\text{cab colour} = \text{white})}.
\]

**Incorporating additional evidence**

Now suppose that there is a further item of relevant evidence: a flake of paint was found at the scene, that can be assumed to have come from the offending cab. The flake is blue, but since the primarily blue cabs have some white areas, and the white cabs have some blue areas, this is not definitive evidence that the cab was blue. Rather, the conditional probabilities for the colour of the paint flake, given the predominant colour of the cab, are as in Table 1.4.

<table>
<thead>
<tr>
<th>cab colour</th>
<th>paint flake</th>
<th>white</th>
<th>blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>white</td>
<td>0.9</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>blue</td>
<td>0.1</td>
<td>0.9</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 1.4: Conditional probability table for paint flake**

A BN incorporating the additional variable paint flake is shown in Figure 1.4. The absence of any arrow between eye witness and paint flake encodes an assumption that, conditional on the state of their common parent cab colour, these two variables are independent. Marginal prior probabilities for this extended BN are as in Figure 1.5, while Figure 1.6 shows
FIGURE 1.4: Bayesian network for hit-and-run accident, including paint flake evidence. The two evidence items are independent, given the actual cab colour.

FIGURE 1.5: Bayesian network for hit-and-run accident, including paint flake, showing prior marginal distributions.

FIGURE 1.6: Bayesian network for hit-and-run accident, updated with eye witness and paint flake evidence.
the effect of incorporating both items of evidence, \textit{eye witness} = “blue”, \textit{paint flake} = blue. The posterior probability that the cab was blue has now gone up to 95%.

\textit{Which fleet?}

We now elaborate the story still further. There are two taxi fleets in the city, called \textit{BLUE} and \textit{WHITE}. The \textit{BLUE} fleet operates 20\% of the cabs, while the \textit{WHITE} fleet has 80\%, as in Table 1.5.

\begin{table}[h]
\centering
\begin{tabular}{lc}
\hline
\textbf{fleet} & \textbf{probability} \\
\hline
\textit{WHITE} & 0.8 \\
\textit{BLUE} & 0.2 \\
\hline
\end{tabular}
\caption{Prior probability table for \textit{fleet}}
\end{table}

In spite of their names, each fleet comprises only 70\% cabs of the named colour, the remaining 30\% being the other colour, as shown in Table 1.6.

\begin{table}[h]
\centering
\begin{tabular}{lcc}
\hline
\textbf{fleet} & \textbf{cab colour} & \textbf{WHITE} & \textbf{BLUE} \\
\hline
\textit{white} & 0.7 & 0.3 \\
\textit{blue} & 0.3 & 0.7 \\
\hline
\end{tabular}
\caption{Conditional probability table for \textit{cab colour}, given \textit{fleet}}
\end{table}

A BN extended to incorporate information on the fleet owning the cab, showing the marginal prior probabilities, is shown in Figure 1.7. We have chosen the numbers in Table 1.5 and Table 1.6 to make this extended story consistent with the more limited description in Figure 1.5. The absence of any arrow from \textit{fleet} to \textit{eye witness} or \textit{paint flake} in Figure 1.7 encodes the
assumption that these two evidence variables are influenced only by the actual colour of the cab, and not further by which fleet it is owned by.

On again incorporating both items of evidence, we obtain the posterior distributions shown in Figure 1.8. The posterior probabilities that cab colour = blue is 95%, as before, but we now obtain additional information: the probability is 65% that the offending cab belongs to the WHITE fleet.

Example 2 (General forensic identification) The general problem of identification of a suspect from forensic trace evidence gathered at the scene of a crime is extremely common, in a variety of forms. Examples include eyewitness evidence (as in Example 1), handwriting, rifling marks on bullets, glass fragments, fibres, footprints, fingerprints, bitemarks, and, of especial importance and power, DNA profiles—this last is considered in detail in §1.5 below.

We wish to establish the evidential value of a comparison between the crime trace and a similar characteristic observed on a suspect. The comparison might be whether or not these are identical, a match, as in simple DNA profiling. In other cases, such as when comparing samples of handwriting, or a bullet with a gun, or DNA mixture samples, the comparison may be less clearcut. An early use of probabilistic reasoning for forensic identification, dating back to 1865, was the Howland will case (anon 1870), which revolved around a possibly forged signature on a will. In another celebrated case, flawed probabilistic reasoning about handwriting identification resulted in the wrongful conviction of Alfred Dreyfus in 1894 (Reinach 1901).

A generic BN for any simple forensic identification problem is displayed in Figure 1.9. Node evidence represents the evidence found at the crime scene, node suspect the characteristic of the suspect, and node other the characteristic of some alternative donor of the trace. The other individual could be an unidentified random individual in some specified population, or a specific known individual. Both suspect and other are initially populated with an appropriate prior distribution, typically estimated from a relevant data-base.
Bayesian Networks in Forensic Science

The hypothesis variable, **suspect left the evidence?**, indicates whether the suspect or the other individual left the trace. It is populated with a formal prior distribution having equal probabilities for *true* and *false*. In a simple matching problem, such as for DNA evidence, **evidence** is logically determined by its parents, being equal to **suspect** or to **other**, according as **suspect left the evidence?** is *true* or *false*; other cases will require context-specific consideration.

One enters the observed trace evidence at **evidence**, the suspect’s characteristic at **suspect**, and, if available, the alternative donor’s characteristic at **other**. As seen in Example 1, after propagating this evidence through the system, the ratio of the formal posterior probabilities of *true* and *false* at node **suspect left the evidence?** will supply the likelihood ratio in favour of the suspect, rather than the other individual, having left the trace.

**FIGURE 1.9: Bayesian network for general forensic identification.** Was it the **suspect**, or some **other** person, who left the crime scene **evidence**?

1.4 Object-Oriented Bayesian Networks

A useful extension of the idea of a Bayesian Network is the Object-Oriented Bayesian Network (OOBN). This organises the nodes into a hierarchy of sub-networks, which can greatly simplify specification and interpretation. (However, computation still takes place at the level of the basic nodes). A particularly valuable application of OOBNs is based on generic network modules, also termed classes, fragments, or idioms, that can be reused, both within and across higher-level networks [Laskey and Mahoney 1997] [Neil et al. 2000] [Hepler et al. 2007] [Fenton et al. 2013]; we will typically use the description "(network) class". Thus the generic identification network of Figure 1.9 could form part of one or many other, more complex networks.

In what follows, we use **bold face** to indicate a network class, and **teletype face** to indicate a simple node or instance of a class. A class network is like a regular network, except that it can have interface — *input* and *output* — nodes, as well as internal nodes. Interface nodes are indicated by a
grey outer ring, an input node having a dotted outline, and an output node a solid outline. Any network can have nodes that are themselves instances of other networks, in addition to regular nodes. Each instance of a class network within another network is displayed as a rounded rectangle, which can be expanded if desired to display its interface nodes. All instances of a class have identical probabilistic structure, except that an input node can be identified, by an incoming arrow, with a node in another network. A full description of forensic OOBN networks can be found in Dawid et al. (2006); Dawid et al. (2007). We merely illustrate selected features here.

**Example 3 (Testimony)** The generic network class *testimony* integrates several key attributes of eyewitness testimony (Schum and Morris 2007). This is conceived as involving three stages: *sensation*, *objectivity*, and *veracity*. These are combined in the network of Figure 1.10, which itself builds on the network classes shown in Figure 1.11.

![Network class testimony](image)

**FIGURE 1.10:** Network class *testimony*, showing the chain of error-prone processes intervening between an actual event and a witness’s testimony of the event.

The attribute *sensation* models the possibility of mistakes in the witness’s perception of the event, due either to his sensory and general physical condition (leading to possible disagreement between the actual and the perceived features of the event), or to the conditions under which the observation is made. The latter aspect is termed *competence* (for example, if the witness was hiding under a table, he might not have been in a position to observe what was happening). These two processes are integrated in the node *sensation* of Figure 1.10, which is an instance of network class *sensation* shown in Figure 1.11a. Here the node *agreement* is itself an instance of the class *accuracy* of Figure 1.11b which uses a random *Error* to determine whether or not its output reproduces its input. Only if this is so, and also *competent = true*, will the attribute *sensation* agree with the actual *event*.

The attribute *objectivity* relates to whether or not the witness’s belief is
a correct interpretation of the evidence of his senses, while veracity relates
to whether or not he truthfully reports his belief. The nodes objectivity
and veracity in Figure 1.10 are likewise modelled as instances of the class
accuracy.

\[ \text{(a) sensation} \quad \text{(b) accuracy} \]

FIGURE 1.11: Testimony classes for use in Figure 1.10 (see text for details).

1.5 Forensic Genetics

Forensic DNA evidence has some 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 son 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?

Relationships and mixed trace: Is \( A \) the son of a contributor to a mix-
ture?

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

In the simplest disputed paternity problem, the evidence \( E \) 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.

1.5.1 Bayesian networks for simple criminal identification

In a simple criminal DNA identification case, the evidence is that the suspect’s DNA profile matches a trace found at the scene of the crime. We are interested in comparing two mutually exclusive hypotheses: the *prosecution hypothesis* \( H_p \): “the crime trace belongs to the suspect \( s \)” (loosely, “the suspect is guilty”), and the *defence hypothesis* \( H_d \): “the crime trace belongs to another actor, \( o \), randomly drawn from the population”. Representation of such problems as BNs was first introduced by Dawid et al. (2002).

In the current STR (short tandem repeat) technology, each DNA profile comprises around 16 to 20 forensic markers, with a genotype (unordered pair of alleles) measured for each marker. There is a fairly small repertory of allele values for each marker; the frequencies of these alleles in various reference populations can be estimated from research databases collected by forensic laboratories.

The analysis proceeds separately for each marker \( m \). The relevant OOBN is shown in Figure 1.12a. Figure 1.12b is an expanded version, showing the identification of nodes across the instance nodes. Nodes \( s \) and \( o \) in Figure 1.12a are each instances of a class *founder*, as pictured in Figure 1.13. This contains

(a) OOBN, showing network instances for alternative contributors \( s \) (suspect) and \( o \) (other) and crime trace nodes

(b) Expanded network, showing identification of internal nodes across instance nodes

*FIGURE 1.12: Network for criminal identification*

nodes for paternal gene \( pg \), maternal gene \( mg \), and genotype \( gt \). Each of the
Bayesian Networks in Forensic Science

FIGURE 1.13: Network class **founder**. An individual’s genotype $g_t$ is formed as the unordered pair of the individual’s maternal gene $m_g$ and paternal gene $p_g$. Nodes $p_g$ and $m_g$ is identified with the sole node of an instance ($p_{gin}$ and $m_{gin}$, respectively) of a simple class **gene**, that has as states the alleles for marker $m$, populated with their frequencies in a relevant reference population; while node $g_t$ of **founder** is an instance of a network class **genotype** (not shown) that simply forms the unordered combination of its input nodes $p_g$ and $m_g$ (since we cannot distinguish the paternal and maternal gene in a genotype).

Node $trace$ in Figure 1.12a is an instance of the **identification** network class shown in Figure 1.14 which is just a version of Figure 1.9: its output $trace$ is modelled as equal to $s_{gt}$ or $o_{gt}$, according as $S guilty?$ is true or false, respectively. The full OOBN of Figure 1.12a elaborates this to account for the specific features of DNA profile identification, in particular the way in which a paternal and a maternal gene combine to form a genotype. As will be shown in §1.7, such an OOBN representation also allows us to include additional features, such as uncertain allele frequencies and heterogeneous populations.

Node $S guilty?$ is assigned a formal prior probability 0.5 for true. The observed matching genotype at marker $m$ is entered as evidence at $g_t$ in $s$, and again at $crgt$ in $trace$, and propagated through the network. The resulting computed odds on true at $S guilty?$ can then be interpreted as the likelihood ratio in favour of $H_p$, based on the evidence of a match at marker $m$. Finally, under the assumption of independence across markers, multiplying these values across all markers delivers the overall likelihood ratio based on the full DNA evidence.

1.5.2 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 $p_f$. On the basis of these data, we wish to assess the likelihood ratio in favour of the hypothesis of paternity: $H_1$: $tf = p_f$, the true father is the putative father; as
FIGURE 1.14: Network class identification: trace is identical to sgt or ogt according as S guilty? is true or false.

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 a suitable specified population.

FIGURE 1.15: Pedigree for simple disputed paternity. Child c is the offspring of mother m and true father tf, who is either the putative father pf or an alternative father af.

Assuming independence across markers, we can again analyse them one at a time, before multiplying their associated likelihood ratio values together to obtain the overall likelihood ratio—which can finally be combined with the prior odds of paternity based on external background evidence, using (1.1), to obtain the posterior odds for paternity.

An OOBN representing the disputed pedigree, for a single forensic marker, is shown in Figure 1.15. We now describe the basic constituents of this OOBN. These will also be used in some of the more complex examples of forensic DNA inference that we shall consider below.

Each of m, pf and af is an instance of the class founder of Figure 1.13. Node c is an instance of a class child, whose structure is displayed in Figure 1.16a.

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 network class meiosis,
Bayesian Networks in Forensic Science

FIGURE 1.16: Network classes child and meiosis. In meiosis the gene $cg$ a parent transmits to a child is chosen at random from that parent’s paternal gene $pg$ and maternal gene $mg$. In child this process is repeated for both parents, the transmitted genes then being combined into the child’s genotype $gt$.

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 the network class genotype, which forgets the information on parental origin. 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, shown in Figure 1.17 to implement this selection.

FIGURE 1.17: Network class query for disputed paternity testing. The query node $tf=f1?$ determines, simultaneously for both genes $tfpg$ and $tfmg$ of the true father $tf$, whether these come from possible father $f1$ or from $f2$.

We initially formally set both hypotheses as equally probable, so that, after propagation of evidence, the ratio of their posterior probabilities yields the likelihood ratio based on this marker. By entering the data for each marker into the appropriate Bayesian network, we can thus easily calculate the overall likelihood ratio in favour of paternity.
As an illustrative example, suppose that the data, for marker D7S820, are: child’s genotype \( cgt = \{12, 12\} \), mother’s genotype \( mgt = \{10, 12\} \), putative father’s genotype \( pfgt = \{10, 12\} \). The node \texttt{gene} in class \texttt{gene} is populated with the estimated population frequencies of the various alleles for this marker. In particular, the frequencies of the observed alleles 10 and 12 for the US Caucasian population (Butler et al. 2003) are, respectively, 0.243 and 0.166. Node \texttt{tf=f1?} in \texttt{query} is assigned formal equal prior probabilities for \texttt{yes} and \texttt{no}. On inserting the genotype evidence and propagating, the posterior probability of \texttt{yes} is 0.751, so that the likelihood ratio in favour of paternity, based on marker D7S820 alone, is 0.751/0.249 = 3.01.

The above procedure for the simple disputed paternity problem might be regarded as overkill, since it is easy to solve from first principles. Conditioning on the genotypes of mother and putative father (which will make 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. Using Mendel’s law of segregation, this event has probability \( 0.5 \times 0.5 = 0.25 \) if the true father is the putative father, and probability \( 0.5 \times 0.166 = 0.083 \) if the true father is, instead, some unrelated individual from the population. Thus the likelihood ratio in favour of paternity, based on marker D7S820 alone, is 0.25/0.083 = 3.01. The OOBN analysis has merely automated this straightforward computation. The real the power of the OOBN approach lies in its versatility and generalisability. Once supplied with the basic building blocks \texttt{founder}, \texttt{child} and \texttt{query}, we can connect them together in different ways, much like a child’s construction set, to represent a wide range of problems that are not easy to solve by other means. Some illustrations are now given in \S1.5.3.

1.5.3 Bayesian networks for complex criminal cases involving family relationships

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 Example 4 and Example 5 we consider two examples of such a case. The analysis of all the data is clearly now much more complex.

Example 4 (Complex disputed paternity for an inheritance case) The pedigree in Figure 1.18 represents a real case of disputed inheritance. It is essentially a more complicated uncertain paternity dispute.

An Australian woman \( m1 \) was on holiday in Venice when she began to suffer from severe toothache. She went to the dentist a few times and they began an affair. When she returned to Australia she learned that the dentist had died. Some months later she gave birth to a boy \( c1 \) and claimed that her son was the dentist’s son and requested her son’s share of the dentist’s
substantial inheritance. The dentist \( d \) was married to \( m_2 \) and had two children \( c_{21} \) and \( c_{22} \). The case was brought as a civil law suit.

DNA profiles were taken of the undisputed family \( m_2, c_{21}, c_{22} \) and the disputed family members \( m_1, c_1 \), and it is questioned whether the deceased father \( d \) is also the true father \( tf_1 \) of child \( c_1 \) of mother \( m_1 \). The hypothesis of interest is represented by node \( tf_1 = d? \) which represents the two hypotheses: that \( d \) is the father of \( c_1 \), or some other unspecified man \( af \) of Australian origin is the father. Here \( c_1, c_{21} \) and \( c_{22} \) are all instances of class \textit{child}, \( tf_1 \) is an instance of \textit{query}, and \( af, m_1, m_2 \) and \( d \) are all instances of class \textit{founder}.

Example 5 (Two brothers) Figure 1.19 is a OOBN representation of a disputed paternity case where we have DNA profiles from a disputed child \( c_1 \) and from its mother \( m_1 \), but not from the putative father \( pf \). We do however have DNA from \( c_2 \), an undisputed child of \( pf \) by a different, observed, mother \( m_2 \), as well as from two undisputed full brothers \( b_1 \) and \( b_2 \) of \( pf \). The sibling relationship is made explicit by the incorporation of the unobserved grandfather \( gf \) and grandmother \( gm \), parents of \( pf, b_1 \) and \( b_2 \). 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 \( gf, gm, m_1, m_2 \) and \( af \) are all instances of class \textit{founder}; \( pf, b_1, b_2, c_1 \) and \( c_2 \) are instances of class \textit{child}; \( tf \) is an instance of class \textit{query}.

For this particular case the overall likelihood 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 \textit{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
FIGURE 1.19: Pedigree for incomplete paternity case. DNA from the putative father $pf$ of $c1$ is unavailable, but indirect information on his genotype is provided by his two brothers $b1$ and $b2$, as well as his undisputed child $c2$ by $m2$.

these ingredients need to be combined appropriately, using Bayes’s theorem, to produce the final probability of paternity.

Example 6 (Hanratty) Another complex case is that of James Hanratty who, found guilty of rape and murder, was in 1962 the last person to be executed in the United Kingdom. In 1998, in an attempt to obtain a posthumous rehabilitation of Hanratty, his mother and brother underwent DNA testing. Their DNA profiles were compared with a DNA sample left at the scene of the crime. Figure 1.20 shows the OOBN for this case. Nodes father, mother and other are all instances of class founder, brother and Hanratty are instances of class child, and crime sample is an instance of class query.

Based on this familial evidence, the likelihood ratio in favour of Hanratty
Bayesian Networks in Forensic Science

having been the source of the crime scene DNA was computed to be 440. This led to the exhumation of Hanratty’s body and the extraction of his full DNA profile, which proved to be a full match with the crime sample, yielding a likelihood ratio of 2.5 million. Although the defence attempted to attribute the match to contamination during the many years for which the crime items had been stored, the Court reaffirmed the original guilty verdict.

1.5.4 Mutation

Yet another advantage of the OOBN approach is that it is easy to modify the component networks to incorporate a variety of additional complications. One such is the possibility of mutation of genes in transmission from parent to child, which, in a simple case of disputed paternity, 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. 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 network class meiosis of Figure 1.16b as shown in Figure 1.21 by passing its output cog through an instance cag of a new network class 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 §1.5.2 or §1.5.3 will now automatically incorporate the additional possibility of mutation. Similar modifications to component networks can readily account for other compli-
cations, such as “silent” alleles, that fail to be picked up by the measuring apparatus.

1.6 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. DNA samples that are found on crime scenes are often complex, as for example when the samples contain DNA from several individuals, or when the DNA samples are degraded. A sample found at a crime scene may contain biological material from the victim but also from individuals who might be involved in the crime. Identifying their DNA is fundamental to help solve the criminal investigation. Advanced DNA technology now extracts genetic material from a huge variety of surfaces and objects which may have been handled by several individuals, only some of whom are related to the specific crime. DNA mixtures with DNA from many individuals occur frequently in multiple-rape cases or in traces left by groups of perpetrators touching the same objects such as crowbars, guns, cigarette butts, etc. In many cases such mixed samples could contain DNA from a victim and a perpetrator. 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 $u_1$ instead of the victim, with hypotheses $H_0: u_1 \& s$ versus $H_1: u_1 \& u_2$. Here we will firstly discuss the BNs used for analysing the discrete information from the alleles in the mixture and then briefly present the analysis for the continuous peak height/area.

Discrete features

Figure 1.22 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 network class founder, 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 with similar structure to the identification class, 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 $p2gt$, $vgt$, $u2gt$ and $p2=v?$. 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. $A_{\text{mix}}$ determines whether allele $A$ is in the mixture: this will be so if at least one $A$ allele is present in either $p_{1\text{gt}}$ or $p_{2\text{gt}}$. Similarly for $B_{\text{mix}}$, $C_{\text{mix}}$, $D_{\text{mix}}$, and $x_{\text{mix}}$ (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.

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 Lauritzen and Mortera (2002) show that it is often possible to set a relatively low upper limit to the number it is reasonable to consider. 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
Bayesian Networks for Analysing Mixed DNA Profiles

missing individuals, silent alleles and a mixed crime trace simply by piecing together instances of appropriate network classes.

Continuous features

So far we have only used discrete information, namely which allele values are present in the mixture and the other profiles from the measured individuals. A more sensitive analysis additionally uses measured “peak areas” or “peak heights”, which give quantitative information on the amounts of DNA involved. This requires much more detailed modelling, but again this can be effected by means of a Bayesian network (Cowell et al. 2007b). Figure 1.23 shows the top level network for two contributors, involving six markers, each an instance of a lower level network class marker as shown in Figure 1.24. Because the mixture proportion $\text{frac}$ 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”. This network is an extended version of the one shown in Figure 1.22 incorporating additional structure to model the quantitative peak area information. In particular, the nodes $\text{Aweight}$ etc. in marker are instances of a network class that models the quantitative information on the peak weight.

FIGURE 1.23: 6-marker OOBN for a DNA mixture of 2 contributors, using peak areas (reproduced from Cowell et al. (2007b)). The node $\text{frac}$ represents the mixture proportion, while nodes D18, FGA, etc. are instances of the network class marker of Figure 1.24.

Cowell et al. (2007b) analyse the data shown in Table 1.7, taken from Evett et al. (1998) involving a 6-marker mixed profile with between 2 and 4 distinct
Bayesian Networks in Forensic Science

FIGURE 1.24: Network class marker with three observed allele peaks. For a description of the top three layers see Figure 1.22. The bottom two layers model the continuous peak area measurements.

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.

<table>
<thead>
<tr>
<th>Marker</th>
<th>D8</th>
<th>D18</th>
<th>D21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alleles</td>
<td>10*</td>
<td>11</td>
<td>14*</td>
</tr>
<tr>
<td>Peak area</td>
<td>6416</td>
<td>383</td>
<td>5659</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Marker</th>
<th>FGA</th>
<th>THO1</th>
<th>VWA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alleles</td>
<td>21*</td>
<td>22*</td>
<td>23</td>
</tr>
<tr>
<td>Peak area</td>
<td>16099</td>
<td>10538</td>
<td>1014</td>
</tr>
</tbody>
</table>

TABLE 1.7: Data for mixed trace with two contributors. The starred values are the suspect’s alleles.

The appropriate extensions of the mixture models become relatively complex when the number of potential contributors to the mixture becomes large, or we want to allow for uncertainty in allele frequencies and/or population substructure.

Cowell et al. (2007a), Cowell et al. (2011), Cowell et al. (2015) extend the statistical model in Cowell et al. (2007b) for the quantitative peak information obtained from an electropherogram of a forensic DNA sample. A gamma model
Analysis of Sensitivity to Assumptions on Founder Genes

is used for the peak heights and the model further develops the modelling of various artefacts that can occur in the DNA amplification process.

The model can both find likelihood ratios for evidential calculations, and deconvolve a DNA mixture for the purpose of finding likely profiles of one or more unknown contributors to the mixture. Computation from this model relies on an efficient implementation of Bayesian network techniques. This allows for readily extension to simultaneous analysis of more than one mixture trace. This modelling of peak height information provides for a very efficient mixture analysis.

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, which had been buried for over 20 years. This DNA was contaminated and appeared to be a mixture of at least two individuals. This case, as well as a complex criminal DNA mixture case, was further analysed in Green and Mortera (2017), which presents general methods for inference about relationships between contributors to a DNA mixture and other individuals of known genotype. The model for relationship inference builds on the approach in Cowell et al. (2015), but makes more explicit use of the Bayesian networks in the modelling.

For further details on quantitative analysis of DNA mixtures we refer to the chapter “Mixture analysis and probabilistic genotyping” in Section 4 of this book.

1.7 Analysis of Sensitivity to Assumptions on Founder Genes

Many forensic genetics problems, as we have shown, can be handled using structured systems of variables, for which Bayesian networks offer an appealing practical modelling framework, and allow inferences to be computed by probability propagation methods. However, when standard assumptions are violated—for example when allele frequencies are unknown, there is identity by descent or the population is heterogeneous—dependence is generated among founding genes, which makes exact calculation of conditional probabilities by propagation methods less straightforward. The standard assumptions that the allele frequencies are fixed and known, that the individual actors in the model are independent and that the allele frequency database is homogeneous can all be questioned (Green and Mortera 2009). We now illustrate a couple of these issues.
1.7.1 Uncertainty in allele frequencies

In reality, the allele frequencies assumed when conducting probabilistic forensic inference are not known probabilities, but estimates based on empirical frequencies in a database.

For the criminal case of §1.5.1 the joint distribution of the founding genes is

\[ \prod_m \{p(\text{spg}_m)p(\text{smg}_m)p(\text{opg}_m)p(\text{omg}_m)\}, \]

and all questions about sensitivity can be expressed through modifications to (1.8). Some generate dependence between founding genes. Following Green and Mortera (2009), assuming the idealisation of a Dirichlet prior and multinomial sampling, the posterior distribution of a set of probabilities is Dirichlet\((M\rho(1), M\rho(2), \ldots, M\rho(k))\), where \(M\) is the (posterior) sample size and the \(\rho\)'s are essentially the database allele frequencies (posterior means). The founding genes \((\text{spg}, \text{smg}, \text{opg}, \text{omg})\) are drawn from this distribution, (conditionally) independently and identically across alleles. This corresponds to the standard set-up for a Dirichlet process model which, by marginalising over the Dirichlet distribution, can be represented in a BN using a Pólya urn scheme. This is represented by the network class \(\text{UGF}\) (“uncertain gene frequencies”) shown in Figure 1.25 for further details see Green and Mortera (2009). For efficiency of the probability propagation, in order to create smaller clique tables this network is set up so that all choices are binary, following the “divorcing” procedure (Jensen 1996), whereby auxiliary nodes are introduced in order to reduce the number of incoming edges of a selected node. An instance of this network class can then be incorporated as a building block in a higher level network that computes inference, for example, about a criminal identification case, a simple or complex paternity testing or a DNA mixture problem. Thus Figure 1.26 shows a network for criminal identification that integrates the network of Figure 1.12a with that of Figure 1.25. Similarly instances of \(\text{UFG}\) of Figure 1.25, representing uncertain allele frequencies, can be integrated into the networks described in §1.5.2 §1.5.3 §1.6. In this way, we can introduce uncertain allele frequencies for the reference population into any forensic identification problem.

1.7.2 Heterogeneous reference population

The assumption that the DNA reference population is homogeneous is questionable. The population is typically a mixture of subgroups.

Population heterogeneity raises two kinds of issues in the modelling. First, since unobserved actors are assumed to have genes drawn from a population, results can depend on which population (and correspondingly which allele frequency database) is used. Secondly, when there is uncertainty about which population is relevant, this can induce dependence between actors, observed or
**FIGURE 1.25:** Network class **UGF** for modelling uncertain allele frequencies with the Pólya urn scheme (adapted with permission from Green and Mortera (2009) where details can be found.)

**FIGURE 1.26:** Network for criminal case with uncertain allele frequencies represented through the Pólya urn scheme. The parameter $M$ can be interpreted as the effective size of the database.
not. Additionally, when uncertainty about subpopulation relates to untyped actors, dependence between markers is induced.

The upper level network for sensitivity of inferences to population structure for criminal identification, based on a synthetic population that is a mixture of Afro-Caribbean, Hispanic and Caucasian subpopulations is shown in Figure 1.27.

**FIGURE 1.27:** Network for two markers in a criminal identification allowing for subpopulation effect, where $S_1$ and $S_2$ are variables identifying the subpopulation. Nodes $sD13$ [resp., $sD3$] for the suspect and $oD13$ [resp., $oD3$] for an alternative suspect are instances of the network class shown in Figure 1.28, where the subpopulation nodes are populated with their associated allele frequencies for marker D13 [resp., D3].

Such problems are easily set up as Bayesian networks incorporating the structure shown in Figure 1.28. The variable $\mathbf{S}$ identifies the subpopulation, which may be dependent or independent between actors depending on the scenario of interest. Crucially, for each actor, $\mathbf{S}$ is the same for both genes for all markers, so that mixing across subpopulations is not the same as averaging the allele frequencies and assuming an undivided subpopulation. Note that, conditionally on subpopulation $\mathbf{S}$, every gene at every marker is drawn independently from the appropriate subpopulation gene pool.

As before, instances of appropriate network classes, like that of Figure 1.28 can be integrated into the networks described in §1.5.2, §1.5.3, §1.6. In this way, we can introduce both uncertain allele frequencies and heterogeneity into any forensic genetic identification problems.

### 1.8 Conclusions

We hope to have shown how useful BNs are for representing and solving a wide variety of complex forensic problems. The modularity and flexibility of BNs allows a complex problem to be broken down into simpler sub-networks.
that can then be pieced back together, so allowing one to build up a complex problem from simpler subproblems. Both genetic and non-genetic information can be represented in the same network. BNs can be used for a criminal case, both as a guide in the investigative phase, and to query the identification of an accused as the perpetrator. They can also be useful for combining different pieces of evidence pertaining to a specific case. The simplicity of their graphical representation make them a useful tool for any legal scholar, for a judge, for the jury or for an investigator to understand the dependencies between different pieces of evidence. In particular, using OOBNS we can construct a flexible computational toolkit, and use it to analyse complex cases involving DNA profiles as well as other types of evidence. In this way one can address a wide range of forensic queries.

BNs and OOBNs can be helpfully applied in many branches of forensic analysis beyond those illustrated here. We hope we have stimulated the reader’s interest in the use of BNs for modelling complex problems in forensic science.

### 1.1 Qualitative structure

A Bayesian network (BN) is a form of directed acyclic graph (DAG), comprising a finite set $V$ of nodes, with arrows between some of the nodes, in such a way that it is not possible, by following the arrows, to return to one’s starting
point. An example is shown in Figure 29. The nodes having arrows pointing out from them and into a node \( v \in V \) are called its parents, and denoted by \( \text{pa}(v) \); those with arrows pointing into them out from \( v \) are its children. This terminology is extended, in an obvious way, to ancestor, descendant, etc. Thus in Figure 29 the black node has the horizontally striped node as its only child, and the two grey nodes as its parents.

### 1.2 Independence properties

Each node \( v \in V \) is identified with a random variable \( X_v \), and for \( S \subseteq V \) we write \( X_S \) for \( (X_v : v \in S) \). A BN represents a joint distribution \( P \) over the variables \( (X_v : v \in V) \) when it is the case that, under \( P \), each node in \( V \) is independent of all its non-descendants, conditional on its parents: in the conditional independence notation of Dawid (1979),

\[
X_v \independent X_{\text{nd}(v)} \mid X_{\text{pa}(v)}.
\]

For example, in a distribution represented by Figure 29 the black node is independent of the vertically striped nodes, conditional on the grey nodes.

For any such distribution \( P \), we can deduce further implied probabilistic conditional independence properties, using the following entirely graphical routine.

**Algorithm 1 (Moralisation)**

Suppose we wish to query the conditional independence property \( X_A \independent X_B \mid X_C \). We proceed by the following steps.

![Directed acyclic graph (DAG)](image)
Appendix: Bayesian Network Basics

Step 1: Ancestral graph
Delete from the DAG all nodes that are not in $A$, $B$, or $C$, or any of their ancestors.

Step 2: Moralisation
Connect by an undirected arc any parents of a common child that are not already connected by an arrow. Then convert all arrows to undirected arcs.

Step 3: Separation
In the resulting undirected graph, look for a path from a node in $A$ to one in $B$ that does not enter $C$.

If there is no such path, deduce that, under $P$, $X_A \perp \perp X_B \mid X_C$.

Using Algorithm 1 we see that, for any distribution represented by Figure 30, the black node is independent of the vertically striped nodes, conditional on the grey nodes.

FIGURE 30: Another DAG. The black node is independent of the vertically striped nodes, conditional on the grey nodes.

.1.3 Quantitative structure
Let $\mathcal{X}_v$ denote the set of states (possible values) of $X_v$. For $S \subseteq V$ we write $\mathcal{X}_S$ for $\bigotimes_{i \in S} \mathcal{X}_i$. For $x = (x_i : i \in V) \in \mathcal{X}_V$, its projection (restriction), $(x_i : i \in S)$, to $\mathcal{X}_S$ is denoted by $x_S$.

We specify, for each node $v \in V$, a distribution over $\mathcal{X}_v$, conditional on each configuration $x_{\text{pa}(v)} \in \mathcal{X}_{\text{pa}(v)}$ of the states of its parents: the associated (discrete or continuous) density is denoted by $p(x_v \mid x_{\text{pa}(v)})$. For a distribution $P$ represented by the BN, the joint density of all variables factorises as the
product of its parent-child conditional densities:

\[
p(x) = \prod_{i \in V} p(x_i | x_{pa(i)}).
\]

Conversely any distribution with a density of this form will be represented by the BN.

When the variables are discrete, the parent-child conditional probability distributions can be represented as tables of the conditional probabilities \( p(x_v | x_{pa(v)}) \). The overall size of each table is the product of the sizes of the state-spaces of the variables concerned.

### 1.4 Computation

The structure of a BN incorporates a degree of modularity, which makes it possible to execute complex computations by dividing them up into a sequence of “local” computations, each involving a subset of the variables, known as a clique. This is particularly useful when all cliques are relatively small. There exist elegant algorithms to identify the cliques and streamline the computations, and these have been implemented in a number of software packages. In particular, for a discrete distribution specified by its parent-child conditional probability tables, the Lauritzen–Spiegelhalter “probability propagation” algorithm (Lauritzen and Spiegelhalter 1988) enables efficient computation of all marginal densities \( \{p(x_i) : i \in V\} \). Moreover, for any \( S \subseteq V \) and evidence \( X_S = x_S^* \), essentially the same algorithm computes the prior probability of obtaining that evidence, \( P(X_S = x_S^*) \), and the marginal posterior densities \( \{p(x_i | X_S = x_S^*) : i \in V\} \). It is also straightforward to compute revised probabilities after incorporating external likelihood evidence relating to a subset of the variables.
Bibliography


Bibliography


