



Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig



Review

Bayesian networks for evaluating forensic DNA profiling evidence: A review and guide to literature

A. Biedermann*, F. Taroni

University of Lausanne, School of Criminal Justice, Institute of Forensic Science, le batochime, 1015 Lausanne-Dorigny, Switzerland

ARTICLE INFO

Article history:

Received 22 February 2011

Received in revised form 7 June 2011

Accepted 24 June 2011

Keywords:

DNA evidence

Bayesian networks

Object-oriented Bayesian networks

ABSTRACT

Almost 30 years ago, Bayesian networks (BNs) were developed in the field of artificial intelligence as a framework that should assist researchers and practitioners in applying the theory of probability to inference problems of more substantive size and, thus, to more realistic and practical problems. Since the late 1980s, Bayesian networks have also attracted researchers in forensic science and this tendency has considerably intensified throughout the last decade. This review article provides an overview of the scientific literature that describes research on Bayesian networks as a tool that can be used to study, develop and implement probabilistic procedures for evaluating the probative value of particular items of scientific evidence in forensic science. Primary attention is drawn here to evaluative issues that pertain to forensic DNA profiling evidence because this is one of the main categories of evidence whose assessment has been studied through Bayesian networks. The scope of topics is large and includes almost any aspect that relates to forensic DNA profiling. Typical examples are inference of source (or, 'criminal identification'), relatedness testing, database searching and special trace evidence evaluation (such as mixed DNA stains or stains with low quantities of DNA). The perspective of the review presented here is not exclusively restricted to DNA evidence, but also includes relevant references and discussion on both, the concept of Bayesian networks as well as its general usage in legal sciences as one among several different graphical approaches to evidence evaluation.

© 2011 Elsevier Ireland Ltd. All rights reserved.

Contents

1. Introduction	000
2. Bayesian networks	000
3. Representational schemes for reasoning about evidence in legal sciences	000
3.1. General legal applications	000
3.2. 'Early' uses of Bayesian networks in forensic science and formative approaches for DNA evidence	000
4. Bayesian networks and the analysis of genetic markers: advanced modelling approaches	000
4.1. Network structures focusing on individual genes and genotypes	000
4.2. Object-oriented Bayesian networks	000
4.3. Miscellaneous applications	000
5. Discussion and conclusions	000
Acknowledgements	000
Appendix A	000
Appendix B	000
References	000

1. Introduction

This review aims at compiling, summarizing and discussing literature, published mainly in forensic science journals, on the

construction and implementation of graphical probabilistic models, that is Bayesian networks, for the assessment of the probative value of forensic DNA profiling evidence. The presentation is placed into a more general context by including also relevant references on the concept and definition of Bayesian networks as well as their study as a framework for reasoning about evidence in legal settings. Section 2 provides a general introduction to the concept of Bayesian networks as an approach developed in

* Corresponding author.

E-mail address: alex.biedermann@unil.ch (A. Biedermann).

the early 1980s for coherent reasoning in situations of uncertainty, which presents a major characterizing feature of many areas of theoretical and applied research (e.g., artificial intelligence, clinical diagnosis). More technical aspects and definitional issues are confined to [Appendix A](#) and [Appendix B](#). On the basis of these general considerations on the principal aspects and properties of Bayesian networks, [Section 3](#) will allow to pursue the discussion of Bayesian networks as a framework for evaluating scientific evidence in legal settings. It is in this part that a detailed overview is provided of the research and applications reported over the last two decades. The various topics and references to literature are organised in a broad thematical form rather than a strict chronological order. The aim is to provide an informed guide to further reading and to point out the general rationale behind Bayesian network guided probabilistic inference modelling for forensic DNA profiling evidence. This is further pointed out in a discussion presented in a concluding [Section 5](#). Not covered in this review are more distinct topics in their own right, such as available software, additional analyses (e.g., parameter sensitivity analyses) to investigate the performance of Bayesian network models and ways to obtain required probability assignments (e.g., by expert elicitation).

2. Bayesian networks

Based on elements of graph and probability theory, Bayesian networks can roughly be defined as a pictorial representation of the dependencies and influences (represented by arcs) among variables (represented by nodes) deemed to be relevant for a particular probabilistic inference problem. Since the early 1980s, Bayesian networks have gained increased acceptance in the field of expert system technology.¹ Notably through their ability to coordinate bidirectional² probabilistic inferences, Bayesian networks are now considered to be a general representation scheme for uncertain knowledge [[2–7](#)]. Although graphical approaches to represent probabilistic information have already been discussed by [Wright \[8,9\]](#),³ for example, it is generally agreed that the works of [Pearl \[11\]](#) have initiated the development of the formalism known today as Bayesian networks.

The name “Bayesian networks” (BNs) is among the most frequently encountered designations and will be used throughout this paper. One of the earliest uses of this term can be found in [Pearl \[12\]](#). According to the field of application, a variety of other terms – although some with possible nuances in definitional details – may be encountered. Among these terms are ‘Bayes nets’, ‘Bayesian belief networks’, ‘Bayesian expert systems’, ‘graphical probabilistic

networks’, ‘probabilistic influence diagrams’, ‘probabilistic network models’, ‘causal networks’ or, more generally, ‘knowledge maps’. For example, the term ‘belief’ in ‘Bayesian belief networks’ is sometimes emphasised in order to clarify that the probabilistic assignments used in a given model reflect degrees of personal belief of the individual that conducts a probabilistic modelling and inference process. In turn, the descriptor ‘influence diagram’ makes reference to an extension of Bayesian networks to elements of decision theory,⁴ in particular nodes for representing decisions and utilities [[4,17](#)]. Broadly speaking, utility refers to an individual’s valuation of a consequence.⁵ In yet other contexts, the notion of ‘causality’ is sometimes used. This stems from the fact that the links between the nodes of a Bayesian network can be interpreted as causal relationships, even though the definition of Bayesian networks does not refer to causality and there is no requirement that the links represent causal impact [[4,18,19](#)]. More generally, ‘causality’ is a philosophically subtle and tricky concept [[20](#)] and will be avoided in this paper.

Scientific literature reports a vast number of applications for Bayesian networks in virtually any field where probabilistic data analysis plays a central role [[21](#)]. Areas of application range, for example, from such fields as information retrieval [[22](#)] to meteorology [[23](#)]. A review by [Gómez \[24\]](#), covering more than one hundred references, lists fields such as agriculture and livestock management, economy, environment impact and natural resources management, industry, medicine, risk analysis, software development (information systems design, user interfaces, multi-agent systems, web applications, etc.) and strategy formulation. A common aspect of these latter areas of application is their reliance upon probability theory as a measure of uncertainty and the use of domain experts’ knowledge for model specification. Although there may be many situations in which the number of variables involved is rather limited so that the resulting degree of technicality remains manageable, this does not typically hold for applications that aim to address real-world (and possibly large scale) inference problems. When applying the probability apparatus to real-world situations, the number of parameters may increase rapidly and the required probability calculus may become very difficult to cope with. This was one of the reasons, for example, why the feasibility of developing substantial applications with probability theory in fields such as artificial intelligence has long been regarded with skepticism [[1](#)].

Similar observations have been made in discussions about the relevance of probabilistic reasoning in legal contexts. In his Chapter “Probability” in [Aitken and Stoney \[25\]](#), [Lindley \[26, p. 37\]](#), for example, notes that “sometimes the calculations are horrendous and cannot at the moment be done (...)”. This was not intended, however, as an advice against probability because the same author⁶ went on to argue that this “(...) is a technical difficulty that adequate research will, one hopes, overcome” [[26, p. 37](#)]. Almost fifteen years later, [Lindley](#) changed his argument slightly. Although he still notes that “(...) the accumulation of

¹ According to [Cowell et al. \[1\]](#), an expert system consists of a knowledge base and an inference engine. The knowledge base encodes domain-specific knowledge about a problem whereas the inference engine provides a means for processing the contents of the knowledge base.

² The term ‘bidirectional’ refers to the arcs (directed edges, see also [Appendix A](#)) of a graph and is interpreted here as the capacity to deal with the processes of, on the one hand, evaluating the probability of obtaining particular evidence given the truth of certain propositions of interest, and, on the other hand, of drawing inferences about propositions of interest, based on particular evidence. This property stimulated particular interest in Bayesian networks in areas that study deduction and induction through probability. Further details on such probability calculus in Bayesian networks are given later in [Appendix B](#).

³ The main focus of these works are models for statistical data with an emphasis on continuous random variables. Bayesian networks can also cope with continuous variables, but not without several constraints. Among the principal constraints is that the present state of technical development only allows one to handle conditional Gaussian (i.e., Normal) distributions. Another constraint, a structural one, forbids the conditioning of a discrete variable by a continuous variable. Notice that when continuous variables are used in the same model along with discrete variables, the resulting networks are sometimes called ‘hybrid’ [[10](#)] or ‘mixed’ [[1](#)]. One possibility for avoiding limitations due to continuous variables is to make their range of values discrete (see also [Appendix A](#)).

⁴ Decision theory represents an emerging concept in forensic science [[13–16](#)], but it is beyond the scope of this paper.

⁵ A consequence is defined as the combination of a particular state of nature (which may typically be unknown to a reasoner or decision maker) and a given decision. As an example, consider a legal actor’s decision to ‘individualise’ a suspect as the source of a crime mark when the true state of nature is that the latter individual is in fact the source of that mark [[15,16](#)]. This is a particular consequence which may also be termed a ‘correct identification’.

⁶ A similar viewpoint is due to [Friedman \[27, p. 1818\]](#): “If applied to take into account all the information we have about a situation, Bayesian analysis requires unrealistically complex calculations, but this does not suggest a problem with the theory. On the contrary, the complexity is in the world surrounding us, and the theory would have limited value if it could not in principle represent that complexity. Probability is a flexible template. It can take into account as much complexity as its user is able to handle.”

simple rules can produce complicated procedures” [28, p. xxiv], he now acknowledges that “methods of handling sets of evidence have been developed; for example, Bayes nets (...)” [28, p. xxiv]. These models – *Bayesian networks* (defined in further detail in Appendix A) – have an underlying probabilistic architecture and reflect the dependencies among variables deemed to be relevant for a particular inference problem. The graphical representation expresses the conditional independence⁷ assumptions that one may meaningfully impose in a probabilistic calculus in order to reduce its dimension to more tractable levels. It was felt that the incorporation of such assumptions is a way ahead to make probabilistic manipulations in real-world applications technically feasible and substantively acceptable [1].

3. Representational schemes for reasoning about evidence in legal sciences

3.1. General legal applications

The study of representational schemes for assisting reasoning about evidence in legal contexts has a remarkably long history. According to scholars in the field of fact analysis, such as Schum [29] or Robertson and Vignaux [30], the charting method developed by Wigmore [31,32] can be taken to be a predecessor of modern network approaches to inference, and ultimately, decision analyses. Wigmore developed a detailed, however essentially non-probabilistic, graphical apparatus for structuring large and complex masses of evidence. A reworked presentation of this approach can be found, for example, in the monographs of Anderson, Twining and Schum [33,34]. More generally, a special issue of *Law, Probability and Risk* (Volume 6, Numbers 1–4, 2007) contains a rich collection of papers (some of which include sections with historical reviews) on the topic of graphic and visual representations of evidence and inference in legal settings.

Graphical approaches that incorporate genuine means for coping with uncertainty, notably through probability theory, began to stimulate legal researchers only about two decades ago. Examples include decision trees⁸ and a modification of these known as ‘route diagrams’ [35,36]. Despite the rigour of their representational capacity, these concepts were not developed to substantive practical applications. Notably for problems involving an increased number of (possibly multiply valued) variables and a complex dependency structure, the compact and economical representation, with a thorough computational architecture, of Bayesian networks was found more advantageous. Subsequently, Bayesian networks advanced to a preferred technique of researchers and practitioners engaged in the joint study of probability and evidence in judicial contexts.

Since the early 1990s, both lawyers and forensic scientists have shown increased interest in Bayesian networks for studying issues that relate to evidence evaluation. While lawyers have been more concerned with structuring cases as a whole, forensic scientists focused primarily on the evaluation of selected items of scientific evidence. Many of the studies with an emphasis on general legal applications rely on Bayesian networks as a method for the retrospective analysis of complex and historically important *causes célèbres*. Edwards [37], for example, provided an alternative analysis of descriptive elements, such as car and hair colour,

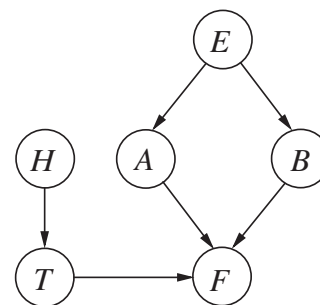


Fig. 1. Bayesian network proposed by Aitken and Gammernan [46]. The nodes are defined as follows: A: ‘X’ committed the murder; B: ‘Y’ committed the murder; E: eyewitness evidence of a row between ‘X’, ‘Y’ and the victim some time before the crime was committed; F: fibres from a jacket similar to one found in the possession of ‘X’ are found at the scene of the crime; H: ‘Y’ drives the car of ‘X’ regularly; T: ‘Y’ picks up fibres from jacket of ‘X’. [Reproduced with permission from Elsevier Limited (*Journal of the Forensic Science Society*)].

presented in the Collins case.⁹ Schum [29] and Kadane and Schum [38] worked on a probabilistic analysis of the Sacco and Vanzetti case¹⁰ with an emphasis on the credibility and relevance of evidence given by human sources, that is, testimony. Similarly, probabilistic analyses have been proposed for the Omar Raddad case¹¹ [39] and, more recently, for the O.J. Simpson case¹² [41].

An important contribution to the discussion of Bayesian networks for structuring and analysing legal arguments has been provided by Robertson and Vignaux [30]. Although not working on a *cause célèbre*, these authors present Bayesian networks as a new development in the field of fact analysis and point out the advantages of Bayesian networks over previously proposed methods, such as Wigmore charts or route diagrams.

Another contribution to the discussion of Bayesian networks in a judicial context has been provided by Fenton and Neil [42]. Besides their primary field of work, that is software reliability, these authors have used Bayesian networks to describe a previously unreported pitfall of intuition, called ‘jury fallacy’. This kind of fallacy applies to situations in which a prior similar conviction by the defendant is revealed after the jury returns a not guilty verdict. Using Bayesian networks with rather conservative assumptions, the authors show that it is fallacious to argue that a prior similar conviction should decrease the belief that jury was correct with their not guilty verdict. This analysis has later been discussed in the legal literature by Jowett [43,44].

3.2. ‘Early’ uses of Bayesian networks in forensic science and formative approaches for DNA evidence

According to Dawid and Evett [45], Aitken and Gammernan [46] were the first to suggest the use of directed acyclic graphs (i.e., Bayesian networks) for probabilistic reasoning in the assessment of forensic evidence. Based on a hypothetical murder scenario (see Fig. 1), these authors have shown how a network approach might be applied to cases involving several, possibly complicated, interrelated issues. They provide a detailed discussion on how (i) relevant propositions can be extracted from a scenario, (ii)

⁷ Conditional independence describes, broadly speaking, a setting in which the truth or otherwise of a proposition would not affect one’s belief in another proposition, given that a third proposition is already known.

⁸ Broadly speaking, decision trees are a kind of graphical model in which separate nodes are used for representing the different outcomes of the same variable. Probabilities, useable for various calculations, are assigned to edges that link nodes of different variables.

⁹ *People v. Collins*, 68 Cal. 2d 319, 438 P. 2d 33, 66 Cal. Rptr 497 (1968).

¹⁰ In this case, Bartolomeo Vanzetti and Nicola Sacco were found guilty of payroll robbery resulting in the murder of Frederick Parmenter and Alessandro Berardelli in South Braintree, MA, on April 15, 1920.

¹¹ The Raddad trial was held in Nice, France. Omar Raddad was charged and convicted in 1994 with murder and robbery of his employer, Mrs. Ghislaine Marchal.

¹² O.J. Simpson was accused of the murder of his ex-wife, Nicole Brown Simpson, and her friend, Ron Goldman. Elements of the O.J. Simpson case, in particular a mixed DNA crime stain, have also been discussed – using Bayesian networks – in Mortera et al. [40] (see also Section 4.1).

relationships between propositions are represented qualitatively in terms of a directed graph, and (iii) subjective beliefs are incorporated as probabilities and used for inference.

Aitken and Gammerman [46] have also mentioned that graphical probabilistic models, besides being used for representing evidence and propagating its effect, could also be an excellent tool for discussing disagreements over probability values that different experts may have. Their opinions, possibly diverging, can be modelled separately and their effects can be compared. This argument now is probably more timely than ever, since many academically and commercially available software tools allow Bayesian networks to be readily implemented in a computerised format.

In a further study, Aitken et al. [47,48] have investigated the potential of Bayesian networks for specific case analysis, also known as “offender profiling”. This approach differs from the previously discussed study (i.e., [46]) in the sense that it is not directly concerned with scientific evidence (e.g., fibres found on the crime scene), but merely with aspects of the offender, the victim and characteristics of the crime. Based on a dataset covering the details of several hundred cases¹³ of sexually motivated child murders and abductions, the authors propose different graphical models to relate the key parameters of a case. These models may be used to revise the probability of offender characteristics, given information about the victim and the crime. The study proposes practical examples on how investigators should revise their belief in the proposition that the offender is (i) living with a partner, and (ii) known to the victim, given information on the victim (e.g., age and sex) and the crime (e.g., method of killing). A second major difference between the studies of Aitken and Gammerman [46] and Aitken et al. [47,48] can be observed when comparing the nature of the knowledge incorporated into the graphical models. The numerical specification of the Bayesian networks described in Ref. [46] is largely based on personal assignments of probability, whereas more data driven assessments are used in the models described in Refs. [47,48].

The discussion on the use of graphical probabilistic models for evaluating forensic evidence, initiated in Ref. [46], has been continued in 1997 by Dawid and Evett [45]. These authors discuss an example which is more complex than the one in Ref. [46] because it involves two kinds of scientific evidence (i.e., fibres and blood), as well as the testimonies of two persons. A particular emphasis is made on the various dependence and independence properties implied by a graph structure. It is also emphasised how these relationships may assist the evaluation of different items of evidence within a complex framework of circumstances. Furthermore, it is shown how a graphical structure can aid in reducing the number of variables necessary for deriving relevant likelihood ratio formulae. From a more general point of view, this paper can be considered as the first that is focusing on Bayesian networks and DNA evidence, because it presents an example involving blood staining.

The previously discussed studies by Aitken and Gammerman [46] and Dawid and Evett [45] both incorporate either one or several items that may be referred to as ‘scientific evidence’,¹⁴ such as fibres or blood. In both studies, these items of scientific evidence are discussed in the context of individual case scenarios. As a consequence, the described graphical procedures involve considerable structural differences that reflect the particular case circumstances.

¹³ The cases relate to incidents reported in Great Britain since 1960.

¹⁴ This designation stems from the nature of these items, as they would typically involve a forensic scientist performing some sort of analysis, such as the microscopic comparison of fibres or the examination of blood staining.

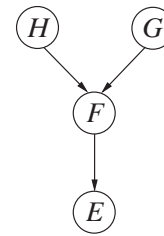


Fig. 2. Bayesian network for ‘one-trace’ cases involving transfer from the offender [49]. The definition of the variables is as follows: *H*: the suspect is the offender; *G*: the crime stain came from the offender; *F*: the crime stain came from the suspect; *E*: the suspect’s blood sample matches the crime stain. [Reproduced with permission from Elsevier Limited (*Forensic Science International*)].

A different approach has been followed in Garbolino and Taroni [49]. These authors do not primarily focus on a particular scenario. On the contrary, they use Bayesian networks to approach what they call a “(...) standard analysis of patterns of inference concerning scientific evidence (...)” [49]. The authors address some more generic and fundamental issues that forensic scientists should account for if they seek to evaluate their evidence in the light of propositions that are of judicial interest. For example, one of the proposed Bayesian networks can be used for evaluating ‘one-trace’ cases involving transfer¹⁵ (see Fig. 2). It incorporates parameters such as:

- the random match probability of the compared characteristics among members of the relevant population,
- the possibility that the stain would have been left by the suspect even though he was innocent of the offence, or
- the relevance¹⁶ of the crime stain for the case.

The model shown in Fig. 2 is applicable for DNA evidence (i.e., transfer material left on a crime scene) and thus represents, after the study of Dawid and Evett [45], a further instance of the use of Bayesian networks for evaluating DNA evidence.

In order to underline the appropriateness of the graph structure and the associated probability assessments, the authors in Ref. [49] show that the properties of their network are such that they agree with Evett’s [52] formula for the relevance term. This demonstration is essential for the discussion of Bayesian networks for evaluating scientific evidence because it allows one to argue that one can, through Bayesian networks, accurately represent existing and accepted probabilistic solutions for forensic inference problems. As such, the proposed generic graphical models in Ref. [49] can serve as building blocks, useable in combination with other network fragments in order to provide elements of a solution for larger problems.

At about the same time as Garbolino and Taroni [49], an application of Bayesian networks to cases involving small quantities of DNA was reported by Evett et al. [53]. To a certain extent, this study can be taken as a continuation of the study by Dawid and Evett [45]. While the study in Ref. [45] focused primarily on the discussion of Bayesian networks as a concept itself, the more recent study by Evett et al. [53] addresses issues more specifically related to the evaluation of DNA evidence. As these authors have observed, the considerable increase in the sensitivity of DNA analyses has initiated a tendency of courts to shift from questions of the kind ‘whose DNA is this?’ to ‘how did

¹⁵ Note that the kind of transfer considered here only concerns one direction. A detailed account on the use of Bayesian networks for approaching cross-transfer (i.e., from the victim to the criminal and vice versa) is given in Aitken et al. [50].

¹⁶ According to Stoney [51], material collected on a crime scene is ‘relevant’ if it has a true connection with the offender. The relevancy of evidence is probabilistic in nature, as it “(...) may range from very likely to practically nil (...)” [51].

this DNA get there?'. With regard to such challenges, the authors show that Bayesian networks are a powerful tool that usefully supplements the evaluation of evidence based on an already established approach, known as the 'hierarchy of propositions'.¹⁷ The paper discusses two case work examples, the first of which involves the outcomes of DNA analyses on a cigarette end recovered near the point of entry of a burgled home. The proposed Bayesian network accounts for questions of the following kind: Is the suspect the person who smoked the cigarette? Did the person who smoked the cigarette leave sufficient DNA to give a profile? Did DNA from the suspect enter the process by innocent means? Did DNA from some third person enter the process? These propositions are modelled as independent binary nodes that point towards a common descendant, termed 'outcome', with states 'match', 'mixture/match', 'different' and 'no profile'.

In order to study the sensitivity of the outcome to changes in the truthstate of the parental variables (that is the uncertain propositions), Beta probability distributions have been used to generate values at random for these parameters. From all the studies on forensic applications of Bayesian networks discussed until now, the paper of Evett et al. [53] is the first to suggest simulations of this kind. The authors have shown that these operations can be used as part of case pre-assessments.¹⁸ Evett et al. [53] provide a second example for which they propose an analysis using Bayesian networks. The scenario pertains to outcomes of DNA analyses performed on samples taken from different areas of a watch (suspected to be worn by the offender) found on the crime scene. This case is somewhat more complex than the first one, as more intermediate propositions are considered. In addition, the hypotheses according to which the suspect is or is not the offender have been included. On the whole, the example clearly illustrates that Bayesian networks allow probabilistic analyses to be made over a large number of variables along with different interrelated issues. The authors conclude that Bayesian networks thus provide valuable support in cases where a full algebraic solution would appear to be extremely difficult [53].

4. Bayesian networks and the analysis of genetic markers: advanced modelling approaches

4.1. Network structures focusing on individual genes and genotypes

In publications quoted so far in this review, a single discrete node is generally used for representing the outcomes of DNA analyses (e.g., the event that the crime and suspect sample 'correspond' with respect to their DNA profile). In the simplest case, this perspective amounts to a two-node network fragment in which knowledge about the state of an observational variable, call it E (short for 'DNA match' as, for example, in Fig. 2), is used for drawing an inference on propositions at the source level (i.e., 'the 'matching' suspect is (is not) the source of the crime stain'). If the latter propositions are represented by F , the corresponding network fragment is $F \rightarrow E$.

Following these formative and rather coarse modelling approaches, considerable research has been devoted to the application of Bayesian networks to inference problems involving the results of DNA analyses. An important contribution to this area

¹⁷ The 'hierarchy of propositions' describes an interpretative framework according to which evidence is evaluated by considering at least two competing hypotheses [54–57]. Usually, the propositions that are of interest to the courts are referred to as the 'Offence level' (e.g., 'The defendant killed Mister Y'). Other levels are the 'Activity level' (e.g., 'The defendant is the person who smashed the window') and the 'Source level' (e.g., 'The crime stain comes from the victim').

¹⁸ A case pre-assessment involves the evaluation of the probability distributions of the likelihood ratio given each of the competing hypotheses [55]. Usually, this is done before performing any analyses.

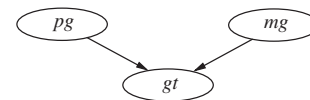


Fig. 3. Representation of a genotype, node gt , with pg and mg denoting the alleles inherited from the father, and the mother, respectively.

of application has been provided by Dawid et al. [58,59]. These authors have shown how appropriate graphical structures for Bayesian networks can be derived from initial pedigree representations of forensic identification problems. Part of their approach are basic representational schemes, useable as repeatable modules in analogous situations. These sub-models are more fine-grained network fragments focusing on individual genes, rather than full genotype representations. An example of this is shown in Fig. 3. Here the node gt , representing a genotype, is modelled as a logical combination of the alleles inherited from the mother and father respectively. These parentally inherited genes are represented by the nodes mg and pg , to be read 'maternal gene' and 'paternal gene', respectively.

A gene, A for instance, can take one of several different forms, also called alleles. Suppose there are n alleles at gene A . These alleles may be denoted A_1, A_2, \dots, A_n . In the Bayesian network shown in Fig. 3 the states A_1, A_2 and A_x are assumed for the nodes mg and pg . The third state, A_x , is an aggregation of all remaining alleles A_3, \dots, A_n . The possible states of the genotype node gt may then be defined as $A_1A_1, A_1A_2, A_1A_x, A_2A_2, A_2A_x$, and A_xA_x .

In order to illustrate this modelling approach for evaluating the results of analyses of genetic markers, consider a typical case of disputed paternity [59]. A certain male, denoted as the putative father pf , is supposed to be the father of a certain child c . DNA profiles are available on the mother m , the putative father pf and the child c . The parameter of interest is the likelihood ratio for the proposition that the putative father pf is the true father tf , noted $tf = pf?$, given knowledge about the genotypes of the child (cgt), the putative father ($pfgt$) and the mother (mgt). The paternity pedigree¹⁹ corresponding to this inference problem is shown in Fig. 4(i). Dawid et al. [59] have proposed a representation of this disputed paternity case in terms of a Bayesian network, which is shown in Fig. 4(ii). There would be a Bayesian network of this kind for each genetic marker that has been analysed. These networks can be analysed separately, and the resulting likelihood ratios combined subsequently. An alternative way of proceeding could consist in combining the single-locus networks within a single 'top-level' network, in the context also called a 'super-network'.

The paternity network depicted in Fig. 4(ii) allows one to obtain the same results as with classic arithmetic calculus of Essen-Möller [60] under the Hardy–Weinberg assumption of independence. One might thus be tempted to conclude that there is no real gain by working with Bayesian networks. As noted by Dawid et al. [59], there is indeed no need to develop new methods for problems that can be solved by simple algebra. These authors add, however, that the aim of proposing the network shown in Fig. 4(ii) is different: it provides a solid basis for evaluating DNA evidence and serves as a starting point for approaching more complex problems that accompany the evaluation of forensic DNA evidence. In order to illustrate this point, Dawid et al. [59] pursue an analysis and discussion of the following topics:

- Missing data: In some cases, the profiles of one or more target individuals may not be available. In such scenarios, indirect evidence (i.e., information on close relatives) may be used to infer something about the allelic configuration of the target

¹⁹ Note that squares represent males and circles females.

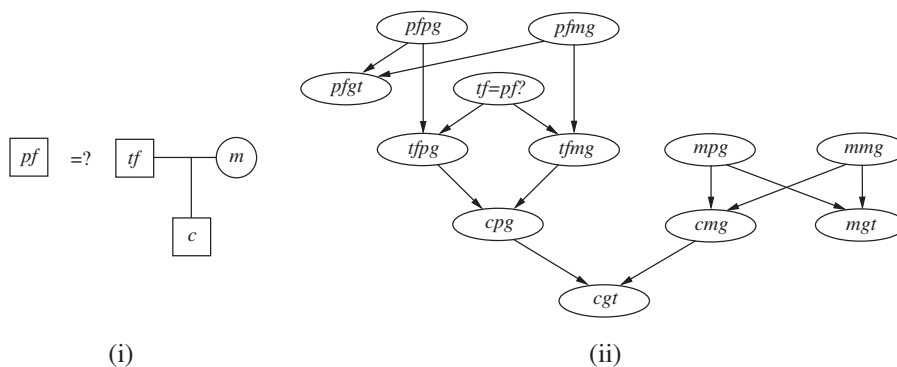


Fig. 4. Different representations of a case of disputed paternity [59]: (i) paternity pedigree (with squares representing males and circles females, *pf* denotes the putative father, *tf* the true father, *m* the mother and *c* the child) and (ii) a Bayesian network. For the Bayesian network, *tftp*, *tfmg*, *mpg* and *mmg* denote the paternal *p* and maternal *m* (in second place) genes of the true father *tf* and the mother *m* (in first place); *cpg* and *cmg* denote the child's paternal and maternal genes, respectively; *pfpt*, *mgt* and *cgt* denote the genotypes of the putative father, the mother and the child, respectively, and *tf = pf?* takes two values in answer: 'yes' or 'no' as to whether the true father is the putative father. [Reproduced with permission from John Wiley and Sons (Scandinavian Journal of Statistics)].

individuals. This analysis is discussed on the basis of two paternity cases.

- Mutation: This problem occurs when there are profiling results that are 'incompatible' with a particular proposition of a familial link. Again, this complication is discussed on the basis of a disputed paternity case.
- Inference of identity: Based on a fictional criminal case, the authors model the inference of identity of a mutilated murdered body on the basis of profiles available from living individuals.

The general modelling approach of Dawid et al. [59] has subsequently been used in further works that concentrated on aspects of the assessment of forensic DNA evidence. Mortera et al. [40,61], for example, have studied Bayesian network models for evaluating mixed DNA traces, including a discussion of issues such as missing individuals and silent alleles, along with a reference to the O.J. Simpson case. The book by Taroni et al. [62] proposes models for situations in which (i) the alternative proposition is that a sibling of the suspect left the crime stain (in agreement with a probabilistic approach previously described by Evett [63]), or (ii) multiple propositions need to be considered (e.g., that the crime stain comes from a brother of the suspect or an unrelated member of the suspect population). A further topic approached in Taroni et al. [62] through Bayesian networks is that of partial matches, that is a situation in which a suspect matches a crime stain only partially and when a proposition of interest is that a close relative of the suspect, such as a brother, is the source of the crime stain.

More generally, Bayesian network models can be used as an integral part of larger IT-environments, including, for instance, connections to operational DNA databases. Bruijning-van Dongen et al. [64], for example, describe a system architecture in which Bayesian networks are used for kinship analyses based on DNA profiles. A primary feature of such an implementation consists of its capacity to deal with the analyses of a large set of cases. This may be of interest for processes such as victim identification in case of a large disaster.

4.2. Object-oriented Bayesian networks

Despite the various compelling capacities of the Bayesian network formalism, the manual model construction may often be a painstaking process, in particular for larger applications. For example, it may be that a model may need to incorporate certain repetitive sub-models. These may be implemented by 'copy and paste' procedures, but whenever the specification of the sub-model of interest requires changes, all of the sub-models of the same type

would need to be reviewed as well. This may thus hinder an efficient model construction, maintenance and reconfiguration.

As a means to overcome such difficulties, Bayesian networks have been extended to 'object-oriented Bayesian networks'. The idea behind this approach is to define generic 'classes' of networks, particular nodes of which (so-called 'instances') can be used, as required, in place of nodes in other networks. The object-oriented Bayesian network language thereby allows one to describe inference problems in terms of inter-related objects. Without entering into any further details, it is solely noted here that an 'object' may either be a standard random variable (as defined in Appendix A), or consist of a set of attributes, each of which is an object. More detailed theory of object-oriented Bayesian networks is given, for example, in Koller and Pfeffer [65], Laskey and Mahoney [66], Bangsø and Wuillemin [67], Neil et al. [68] and Kjærulff and Madsen [17]. The object-oriented approach to Bayesian networks allows one to structure problems hierarchically and with different levels of abstraction. This is often regarded as a convenient property because humans tend to think naturally in terms of hierarchies of abstractions, in particular where it is difficult to mentally capture all aspects of a problem simultaneously.

This practical complication in probabilistic modelling is also encountered in forensic contexts [69], and DNA profiling results provide an illustrative example for this. In fact, in recent years, considerable research has been devoted to the application of object-oriented Bayesian networks for approaching various evaluative aspects relating to DNA profiling results. The main contributions are the following:

- Dawid et al. [70,71] present object-oriented Bayesian networks as an improved probabilistic modelling approach for the broad scope of evaluative situations (e.g., criminal identification or relatedness testing) and problems (e.g., mutation or silent alleles) they have addressed in their earlier paper in 2002 [59]. In the particular area of paternity testing, Hepler and Weir [72] provide an object-oriented extension that allows one to account for allelic dependencies.²⁰
- The papers by Dawid [73] and Vicard et al. [74] rely on object-oriented Bayesian networks for estimating mutation rates.
- The work by Cowell et al. [75] uses object-oriented Bayesian networks for developing a more refined approach to interpreting DNA-mixture evidence. While previous research on Bayesian

²⁰ A more general discussion on conventional (i.e., not object-oriented) Bayesian networks and sub-population effects is given in Taroni et al. [62].

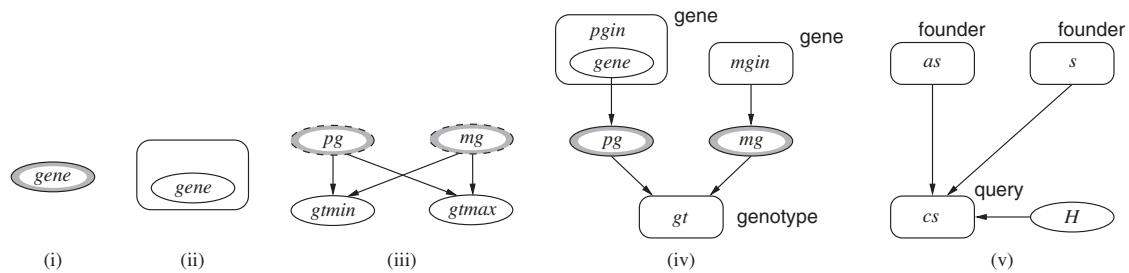


Fig. 5. (i): Representation of a gene at a given locus in terms of a class network 'gene' consisting of a single output node. (ii): Expanded representation of an instance of the class network 'gene'. (iii): Class network 'genotype' with nodes *gtmin* and *gtmax* representing, respectively, the minimum and maximum of the two input gene nodes *pg* ('paternal gene') and *mg* ('maternal gene'). (iv): Class network 'founder' containing instances of the class networks 'gene' and 'genotype'. (v): Object-oriented Bayesian network for a one-stain one offender scenario. Further definitional details are given in the text. [Reproduced with permission from Elsevier Limited (Forensic Science International)].

networks for DNA mixture analysis did not take into account the quantity of DNA [40], the methodology proposed in Ref. [75] focuses more specifically on the use of quantitative peak area information. In Ref. [76], Cowell et al. extend their application to two independent mixture samples that have DNA from the same set of contributors in order to evaluate the improvement in correct genotype assignment. More recently, they have reported an extension of their approach in order to account for artifacts such as allelic dropout, stutter bands and silent alleles [77].

- The study by Cavallini and Corradi [78] describes object-oriented Bayesian networks for database search settings,²¹ including an extension to propositions that consider that the source of a crime stain is a member of the families of the analysed database members.
- Green and Mortera [81] developed object-oriented Bayesian networks for studying the effect of assumptions about founding genes (e.g., uncertainty about allele probabilities).

As an illustration of the object-oriented Bayesian network approach proposed by Dawid et al. [71], let us start by considering a model for a gene at a given locus, typically inherited either paternally or maternally (Fig. 5(i)). That gene is represented by a single node *gene* which contains the full repertory of allele values (for a given locus) along with their population proportions. The bold gray border and the solid outer contour line indicate that this node is defined as a so-called 'output' node. This makes this one-node network a 'class network', which means that it can be invoked (possibly repeatedly) as an instance in some other network which may, depending on the level of hierarchy, also be referred to as the 'master' network.

Fig. 5(ii) provides an expanded representation of how the class network, called 'gene' here, would appear in some other network. This is further illustrated in Fig. 5(iv), which contains two instances of the network class 'gene': one is shown, in expanded form, as the top left-hand network fragment (i.e., node *pgin*) and another one is shown as the top right-hand network fragment (i.e., node *mgin*, in contracted form). These nodes, *pgin* and *mgin*, represent, respectively, an individual's paternally and maternally inherited genes at a given locus. The bottom node *gt* in Fig. 5(iv) is an instance of the class network 'genotype', whose internal structure is shown in Fig. 5(iii). This latter class network contains the two nodes *gtmin* and *gtmax*, which are defined as the minimum and maximum of the two input gene nodes *pg* (short for 'paternal gene') and *mg* (short for 'maternal gene'). This represents an alternative way for specifying an individual's genotype. That is, instead of having a single node *gt* (as shown earlier in Fig. 3) which contains the collection of possible genotypes at a given locus (as a function of

the specified set of alleles), distinct nodes are used in Fig. 5(iii) for representing an individual's allelic configuration. This latter representation may be preferable because the size of the probability table of the node *gt* in Fig. 3 may grow rapidly whenever the specified set of alleles increases.²² Let us further mention that the bold gray border and dashed outer line of the nodes *pg* and *mg* indicate that they are interface nodes of type 'input'. This means that they act as placeholders (or, stated otherwise, instances) of identical nodes in another network (e.g., in a 'master' network). This is illustrated in Fig. 5(iv), where the nodes *pg* and *mg* connect to the input nodes *pg* and *mg* of *gt*. The direct connection is not visible, however, because *gt* is shown in contracted form.

The network shown in Fig. 5(iv) is a class called 'founder', which is repeatedly invoked in the object-oriented Bayesian network shown in Fig. 5(v). This network models a situation in which one seeks to evaluate whether a suspect *s* or some alternative source *as* is the source of a crime stain *cs*. This latter node is an instance of the class 'query' [71], whose internal structure is not shown in further detail here. The main proposition at the source level is represented by a regular Boolean node *H*.

4.3. Miscellaneous applications

Besides the main applications in DNA evidence evaluation, that is 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 DNA analyses.

One such topic is that of DNA cross-transfer evidence. In fact, all works on stain 'identification' evidence quoted so far in this review consider only so-called unidirectionally transferred DNA stains (e.g., a stain from the victim transferred to an offender, or a stain left by an offender on a crime scene). In practice, however, there may well be cases in which a cross-transfer of DNA evidence occurred. For example, DNA evidence may be collected on both, a victim and a suspect in an assault case. Evaluating DNA evidence meaningfully in such situations may require a discussion of probabilities of transfer, persistence and recovery, innocent acquisition, relevance and innocent presence. Aitken et al. [50] studied these aspects through Bayesian networks.

A further topic largely unconsidered in DNA evidence evaluation is that of error rates. Usually, attention is solely drawn to the conditional profile probability. A probabilistic approach to account for the potential of error is given, for example, by Thompson et al.

²¹ Taroni et al. [62,79] discuss a Bayesian network without object-orientation for approaching database search settings in agreement with the likelihood ratio procedure described in Evett and Weir [80].

²² For example, when considering 8 alleles for the locus TH01 (considering data from, for example, Butler et al. [82]), this would imply $(8(8+1))/2 = 36$ distinct genotypes for the node *gt*. Along with 8 states for each of the two parental gene nodes *pg* and *mg*, the probability table for the node *gt* would then have $8 \times 8 \times 36 = 2304$ entries.

[83]. An implementation of this approach in terms of Bayesian networks is described in Taroni et al. [62,79].

The applications described so far also have in common that they almost exclusively focus on autosomal DNA profiling results. The use of Bayesian networks for evaluating profiling results of X- and Y-chromosome short tandem repeats is presented in, respectively, Hartsch et al. [84] and Taroni et al. [62].

Besides, applications for Bayesian networks have also been described for addressing the issue of a wrongful designation of a heterozygous genotype as a homozygote, which may lead to a false exclusion [85], as well as the evaluation of likelihood ratios when multiple propositions need to be considered [86].

A different approach for constructing Bayesian networks for inference from genetic markers has been described by Cowell [87]. This author has developed a software tool, called finex (Forensic Identification by Network EXpert systems), where inference problems based on DNA evidence can be expressed through the syntax of a graphical specification language. The program finex then uses an algorithm for automatically constructing an appropriate representation in terms of a Bayesian network. Compared to general purpose software, finex allows one to save time in setting up networks and reduces the potential for error while completing large probability tables. In addition, evidence from several markers to evaluate likelihoods can readily be combined.

5. Discussion and conclusions

Despite some technical and representational differences, most current practices for constructing Bayesian networks share some basic assumptions. Among these are, notably, independence within and across markers. Furthermore, it is assumed that all founders in a pedigree, including unrepresented individuals, are considered to be drawn randomly from the same homogenous population. It has been reported that future works in this area will aim to relax these assumptions [59].

A particular aspect of Bayesian networks for inference from genetic markers concerns the rationale behind their construction. Building appropriate graphical structures can largely be guided by Mendelian laws of inheritance and the logical relationships between genes and genotypes. Basic inferential structures, as shown in Fig. 3, can thus be combined in a clearly defined way. This may in part explain why graphical structures with large topologies are more common in the context of DNA evidence. Indeed, the Bayesian networks for more classic kinds of evidence (e.g., fibres [62,79] or toolmarks [88]) have rather local structures as they involve fewer variables. A main reason for this seems to lie in limitations of available knowledge about the structural dependencies among the different aspects of the inferential problems at hand. Usually, there is no generally applicable law, such as the law of inheritance in the context of DNA evidence, that may provide a clear indication as to how nodes ought to be combined. Therefore, the construction must essentially be based on logical considerations, which is a difficult task whenever the number of variables increases.

But there are also differences between the Bayesian network approaches for DNA evidence as discussed throughout this paper. This may become apparent when comparing the nature of the target questions these models seek to approach. The networks proposed by researchers such as Dawid et al. [59] are essentially concerned with drawing inferences on the source of a certain item of DNA evidence, or with identifying individuals in a certain pedigree. On the other hand, the more general networks proposed by Garbolino and Taroni [49], for example, address the question of whether the source of a certain crime stain (e.g., a suspect) could be the offender. This latter inference step requires further consider-

ations of general patterns of reasoning, such as evidential relevance [51].

In summary, Bayesian networks represent one of the more recent graphical approaches that have been proposed for the evaluation evidence in judicial contexts. Compared with most previously developed graphical methods, the use of Bayesian networks offers the additional advantage of incorporating probability theory as a coherent measure of uncertainty. Moreover, computerised systems currently exist that can perform calculations over a number of variables with varying dependency structure. More recent developments, in particular the extension to object-orientation, have further increased the capacity of Bayesian networks to deal with the level of complication that is associated with real-case settings involving DNA evidence. Bayesian networks thus are a valuable tool in the hands of forensic scientists. They represent a complementary contribution to the body of analytical techniques that are needed to approach inference problems in accordance with probability theory, both conceptually and practically.

Acknowledgement

The authors thank the anonymous referees for their comments that helped improve this paper.

Appendix A. Definition and properties of Bayesian networks

In Bayesian networks, graph theory is used to provide a qualitative model structure, whereas probability theory is used to characterise the nature and strength of the relations that reign within a model. This has concisely been formulated in the Preface of [89], where graphical models have been described as a marriage between graph theory and probability theory. More formally, a Bayesian network covers the following elements [4, e.g.]:

- A finite collection of random variables that are represented by nodes. Each of these nodes has a finite set of mutually exclusive states (sometimes also called 'outcomes').
- A set of directed edges that connect pairs of nodes.
- The set of variables and the set of directed edges are combined in such a way that a directed acyclic graph is obtained, that is, a graph where no loops are permitted.
- Node probability tables are associated with each variable of the network: the probability table of a variable A that receives entering edges from variables B_1, \dots, B_n contains conditional probabilities $\Pr(A|B_1, \dots, B_n)$, whereas a variable A with no entering edges from other variables contains unconditional²³ probabilities $\Pr(A)$.

The nodes of a Bayesian network represent propositional variables of interest, that is – very generally speaking – statements or assertions that such-and-such is the case (e.g., an outcome or a state of nature). It is assumed that personal degrees of belief can be assigned to them. Propositions are basic (intellectual) attributes formed by an individual during the course of a reasoning task. A proposition can be thought of as referring to states of affairs, although, most often, the actual state may not be known with certainty. For example, there may be uncertainty about the truth or otherwise of the proposition

²³ Let us notice that the term 'unconditional' refers here only to the absence of an explicit conditioning upon other variables (nodes) in the network. Strictly speaking, one should also consider a probability of the kind $\Pr(A)$ as conditional because there is always contextual information, habitually denoted by I , which is used when quantifying $\Pr(A)$. This would also imply that $\Pr(A)$ should be written more correctly as $\Pr(A|I)$, but I is often omitted in order to reduce notational burden.

according to which a crime stain has been left by the offender. Within a Bayesian network, such a proposition is conceptualised in terms of a node, whose states represent, respectively, the truth and the falsity of that proposition. The degree of belief maintained in each of these states is expressed numerically, that is, in terms of probabilities. These probabilities are organised in that node's probability table.

The mutually exclusive states of a variable are also referred to as the 'domain' of the variable. The domain of a variable may take one of different forms which will determine that variable's subtype. Examples include {red, green, blue} for a labeled variable, {T, F} for a Boolean variable (with 'T' and 'F' denoting 'true' and 'false', respectively), {−2, 1, 4, 5, 7.34, 8} for a numbered variable or {−∞; 0], [0; 10], [10; 100]} for an interval node. Note that the latter node type is used for specifying the intervals over which a continuous quantity can be discretised.

It is common to say that, if there is a link from a node *A* to a node *B* (i.e., a network fragment of the kind $A \rightarrow B$), that *A* is a 'parent' of *B*, whereas *B* will be said to be a 'child' of *A*. In analogy to the notions of 'parents' and 'children', terminology of kinship is used to denote various other relationships in a graph [90]. The meaning of 'ancestor' and 'descendant' is thus intuitively understandable. Some awareness is required, however, when modelling inferences from genetic markers (e.g., in forensic contexts of disputed paternity). Here, terms such as 'parent' and 'child' denote real familial relationships. As proposed by Cowell [87], the terms 'graphical parent' and 'graphical child' should thus be used to clarify that a discussion is focusing on a structural property of a Bayesian network. A similar distinction is made, for instance, by Lauritzen and Sheehan [91], who refer to 'graph parent' and 'bio parent'. Variables that have no entering edges from other nodes are sometimes referred to as 'root' or 'source' nodes. They can be used, for example, to represent a proposition (or hypothesis) that one seeks to establish. In judicial contexts, propositions of this kind have also been termed 'ultimate probanda' [30].

The arrows in a Bayesian network represent relationships that correspond to a property that a reasoner assumes to hold within the context of the inference problem at hand. If a network is properly constructed, then a directed edge from a node *A* to a node *B* signifies that *A* has a direct influence on *B*. As an intuitively appealing way of presentation, the links between nodes are sometimes interpreted as 'causal relationships', but, as mentioned earlier in Section 2, it is important to note that the definition of Bayesian networks does not refer to causality and there is no requirement that the links represent causal impact.

Bayesian networks make use of nodes and edges in much the same way as other kinds of graphical models. A more distinctive feature of Bayesian networks are, however, node probability tables. They allow probability theory to be implemented as a formalism for interpreting the nature and the strengths of the relationships between different graphical constituents. The node tables can be allowed to accommodate probabilities from a variety of different sources. Among the most common are subjective estimates from human experts, (statistical) data (e.g., from databases) or literature. Node probability tables can thus be considered as a means of interfacing a model to data [89]. Notice also that probability tables can be completed through the use of mathematical expressions by exploiting various variable subtypes as described above.

As mentioned so far in this Section, an important property of Bayesian networks is the encoding of collections of conditional

independence assumptions. Apart from this, there is a further important property of the Bayesian network formalism. It consists in its capacity of facilitating the representation of probability distributions. For the purpose of illustration, consider a distribution Pr defined on n discrete variables, ordered arbitrarily as A_1, A_2, \dots, A_n . Using the product rule, the joint distribution $Pr(A_1, \dots, A_n)$ can be decomposed as follows (also known as the 'chain rule'):

$$Pr(A_1, \dots, A_n) = \left[\prod_{i=2}^n Pr(A_i | A_1, \dots, A_{i-1}) \right] Pr(A_1) . \quad (A.1)$$

A more economic representation of a joint probability distribution can be obtained if one is capable of specifying variables that are not 'sensitive' to all predecessors but only a certain subset of those predecessors. Stated otherwise, a variable A_i may be independent of all other predecessors once a selected group of predecessors of A_i , called 'parents' (*par*) of A_i , is known. In such a case the product of Eq. (A.1) can be rewritten in a shorter form: $Pr(A_i | par(A_i))$.

Now, if the conditional independencies in a Bayesian network hold for the collection of variables A_1, \dots, A_n , then the Bayesian network provides a representation of the joint probability distribution $Pr(A_1, \dots, A_n)$ in terms of the product of all specified potentials, that is,

$$Pr(A_1, \dots, A_n) = \prod_{i=1}^n Pr(A_i | par(A_i)) \quad (A.2)$$

where $par(A_i)$ represents the set of parental variables of A_i . Eq. (A.2) is called the 'chain rule for Bayesian networks' and formally defines what a Bayesian network means: a representation of the joint probability distribution for all the variables.

Many of the notions introduced in this Section may also be formulated in more technical, mathematical notation. Detailed instances of such approaches can be found, for example, in Castillo et al. [3], Cowell et al. [1], Jensen and Nielsen [4] and Nam et al. [92].

Appendix B. 'Evidence propagation' in Bayesian networks

Once a Bayesian network is properly constructed, it can be consulted to process newly acquired information, that is, calculating the conditional probabilities of the nodes in the network given that the values of some of the nodes have been observed. It is customary to denote the latter as 'evidence variables' and the former as 'query variables'. The term 'evidence' is used here in a broad sense. It is taken as information that is available to a particular individual and that is thought to be incorporated in the system of beliefs maintained by that individual. Such 'information' may consist, in forensic contexts, of one or more propositions whose truthstate becomes certain for a given reasoner (e.g., when a scientist decides that there is a 'match' between evidential and reference material). This would then provide the basis for reasoning about other propositions that bear on an inference problem of interest. There may also be occasions on which 'information', or 'evidence', may be of a more technical nature, such as numerical data generated by an experimental procedure (e.g., a number n of glass fragments collected on a suspect's clothing, or, in the context of DNA evidence, detected alleles). Such numerical data may directly serve as input data that can be accommodated by a Bayesian network (e.g., by means of a numbered node). In either case, it is assumed that what is

considered as so-called ‘hard’ evidence is something that corresponds to the instantiation²⁴ of a network’s node.

The processing of information is a key task handled by Bayesian networks, and this represents one of the main points of interest for the study of probabilistic inference in forensic science. Actually, the core feature of Bayesian networks is, as their name says, the re-evaluation of probabilities, given particular evidence, which is a task carried out according to Bayes’ theorem. A convenient graphical illustration of this property amounts to a two-node network, involving the two binary variables A and B with the factorised form $\Pr(A)\Pr(B|A)$, that is, the Bayesian network $A \rightarrow B$.²⁵ Upon observing that B holds, one seeks to calculate $\Pr(A|B)$. To apply Bayes’ theorem for this task, the following calculations are needed: (i) $\Pr(B) = \Pr(B|A)\Pr(A) + \Pr(B|\bar{A})\Pr(\bar{A})$, (ii) $\Pr(A, B) = \Pr(A)\Pr(B|A)$, and (iii) $\Pr(A|B) = \Pr(A, B)/\Pr(B)$. Notice that calculation (i) invokes the so-called extension of the conversation rule. In this computation, evidence on B is ‘forwarded’ to A against the direction of the arc that holds between these two nodes. It is important to note that within Bayesian networks, propagation is also possible along the direction of the arcs. To continue the example introduced above, a propagation along the direction of the arc $A \rightarrow B$ means to calculate the probability of B given A , or, $\Pr(B|A)$ for short. What happens during the passage from $\Pr(B)$ to $\Pr(B|A)$ can be seen in the above mentioned Eq. (i). This relationship states that for assessing the uncertainty of B one needs to take into account uncertainty in relation to A . When A becomes ‘known’ this means that $\Pr(A) = 1$ and $\Pr(\bar{A}) = 0$. Consequently, (i) reduces to $\Pr(B|A)$ and this is just the respective value that has been specified in the conditional probability table of node B .

This description of probabilistic calculations that can be handled within Bayesian networks is informal and based solely on a local network fragment. Since the beginning of the development of Bayesian networks, researchers and practitioners have sought ways to automate these calculations, notably through computerised implementations. This was felt to be an essential step to be achieved if the approach is to be of use for real-world applications. A collection of algorithms is now available that allow calculations to be made in an efficient manner. The most frequently referenced method in this context is due to Lauritzen and Spiegelhalter [94]. This approach has later been implemented in academically and commercially available Bayesian network software. A more detailed discussion of algorithms for inference in Bayesian networks is not pursued here because it requires some length and further complexity. For the purpose of the current paper, attention is confined to structural issues of Bayesian networks. It is assumed that calculations will be confined, where needed, to a computer program. The assumption is made that the results provided by such systems comply with the definition of the basic concepts, that is, probability and Bayesian networks.

²⁴ Evidence on a variable corresponds to a statement of the certainties of that variable’s states. If it is known with certainty in which state a variable is, then evidence is called ‘hard’ and the node is said to be ‘instantiated’. Alternatively, an instantiated variable may also be referred to as an ‘observed’ variable. Generally, evidence that is not ‘hard’ is called ‘soft’. One can also talk about evidence in terms of a function. For a variable with discrete states, for example, an evidence function that assigns a zero probability to all but one state may be said to provide hard evidence.

²⁵ A common, general instance that can be represented in this way is the evaluation of a screening test [93] where B refers to the result of the test (which may either be positive or negative) and A is the proposition about which one is uncertain, for example, whether or not a sample is human blood, or whether or not an unknown powder contains something illegal.

References

- [1] R.G. Cowell, A.P. Dawid, S.L. Lauritzen, D.J. Spiegelhalter, Probabilistic Networks and Expert Systems, Springer, New York, 1999.
- [2] K.B. Korb, A.E. Nicholson, Bayesian Artificial Intelligence, Chapman & Hall/CRC, Boca Raton, 2004.
- [3] E. Castillo, J.M. Gutiérrez, S. Hadi, Expert Systems and Probabilistic Network Models, Springer, New York, 1997.
- [4] F.V. Jensen, T.D. Nielsen, Bayesian Networks and Decision Graphs, 2nd Edition, Springer, New York, 2007.
- [5] S. Wright, Probabilistic Reasoning in Expert Systems, John Wiley & Sons, Inc., New York, 1990.
- [6] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers, Inc., San Mateo, 1988.
- [7] G. Shafer, J. Pearl (Eds.), Readings in Uncertain Reasoning, Morgan Kaufmann, San Mateo, 1990.
- [8] S. Wright, Correlation and causation, Journal of Agricultural Research 20 (1921) 557–585.
- [9] S. Wright, The method of path coefficients, Annals of Mathematical Statistics 5 (1934) 161–215.
- [10] S. Russell, P. Norvig, Artificial Intelligence. A Modern Approach, 2nd Edition, Prentice Hall, New Jersey, 2003.
- [11] J. Pearl, Reverend Bayes on inference engines: a distributed hierarchical approach, in: Proceedings of the National Conference on Artificial Intelligence (AAAI’82), Morgan Kaufmann, Pittsburgh, 1982, pp. 133–136.
- [12] J. Pearl, Bayesian networks: a model of self-activated memory for evidential reasoning, in: Proceedings of the Cognitive Science Society, Ablex, Greenwich, 1985, pp. 329–334.
- [13] F. Taroni, S. Bozza, C.G.G. Aitken, Decision analysis in forensic science, Journal of Forensic Sciences 50 (2005) 894–905.
- [14] F. Taroni, S. Bozza, A. Biedermann, Two items of evidence, no putative source: an inference problem in forensic intelligence, Journal of Forensic Sciences 51 (2006) 1350–1361.
- [15] A. Biedermann, S. Bozza, F. Taroni, Decision theoretic properties of forensic identification: underlying logic and argumentative implications, Forensic Science International 177 (2008) 120–132.
- [16] F. Taroni, S. Bozza, A. Biedermann, G. Garbolino, C.G.G. Aitken, Data Analysis in Forensic Science: A Bayesian Decision Perspective, John Wiley & Sons, Chichester, 2010.
- [17] U.B. Kjærulff, A.L. Madsen, Bayesian networks and Influence Diagrams, A Guide to Construction and Analysis, Springer, New York, 2008.
- [18] E.A. Bender, Mathematical Methods in Artificial Intelligence, IEEE Computer Society Press, Los Alamitos, 1996.
- [19] H.R. Keshavan, J. Barnett, D. Geiger, T. Verma, Introduction to the special section on probabilistic reasoning, IEEE Transactions on Pattern Analysis and Machine Intelligence 15 (1993) 193–195.
- [20] I. Hacking, An Introduction to Probability and Inductive Logic, Cambridge University Press, Cambridge, 2001.
- [21] O. Pourret, P. Naïm, B. Marcot, Bayesian Networks: A Practical Guide to Applications, John Wiley & Sons, Chichester, 2008.
- [22] L.M. de Campos, J.M. Fernández-Luna, J.F. Huete, Fast propagation algorithms for singly connected networks and their applications to information retrieval, in: J.A. Gámez, S. Moral, A. Salmerón (Eds.), Advances in Bayesian networks, Springer, Berlin, 2004, pp. 271–288.
- [23] R. Cano, C. Sordo, J.M. Gutiérrez, Applications of Bayesian networks in meteorology, in: J.A. Gámez, S. Moral, A. Salmerón (Eds.), Advances in Bayesian networks, Springer, Berlin, 2004, pp. 309–328.
- [24] M. Gómez, Real-world applications of influence diagrams, in: J.A. Gámez, S. Moral, A. Salmerón (Eds.), Advances in Bayesian networks, Springer, Berlin, 2004, pp. 161–180.
- [25] C.G.G. Aitken, D.A. Stoney, The Use of Statistics in Forensic Science, Ellis Horwood, New York, 1991.
- [26] D.V. Lindley, Probability, in: C.G.G. Aitken, D.A. Stoney (Eds.), The Use of Statistics in Forensic Science, Ellis Horwood, New York, 1991, pp. 27–50.
- [27] R.D. Friedman, Assessing evidence, Michigan Law Review 94 (1996) 1810–1838.
- [28] D.V. Lindley, Foreword, in: C.G.G. Aitken, F. Taroni (Eds.), Statistics and the Evaluation of Evidence for Forensic Scientists, 2nd Edition, John Wiley & Sons, New York, 2004.
- [29] D.A. Schum, Evidential Foundations of Probabilistic Reasoning, John Wiley & Sons, Inc., New York, 1994.
- [30] B. Robertson, G.A. Vignaux, Taking fact analysis seriously, Michigan Law Review 91 (1993) 1442–1464.
- [31] J.H. Wigmore, The problem of proof, Illinois Law Review 8 (1913) 77–103.
- [32] J.H. Wigmore, The Science of Judicial Proof: As Given by Logic, Psychology and General Experience and Illustrated in Judicial Trials, 3rd Edition, Little, Brown, Boston, 1937.
- [33] T. Anderson, W. Twining, Analysis of Evidence: How to do Things with Facts Based on Wigmore’s Science of Judicial Proof, Northwestern University Press, Evanston, 1998.
- [34] T. Anderson, D.A. Schum, W. Twining, Analysis of evidence, in: Law in Context, 2nd Edition, Cambridge University Press, Cambridge, 2005.
- [35] R.D. Friedman, A close look at probative value, Boston University Law Review 66 (1986) 733–759.
- [36] R.D. Friedman, A diagrammatic approach to evidence, Boston University Law Review 66 (1986) 571–622.

- [37] W. Edwards, Influence diagrams, Bayesian imperialism, and the Collins case: an appeal to reason, *Cardozo Law Review* 13 (1991) 1025–1074.
- [38] J.B. Kadane, D.A. Schum, *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*, John Wiley & Sons, New York, 1996.
- [39] T.S. Levitt, K. Blackmond Laskey, Computational inference for evidential reasoning in support of judicial proof, *Cardozo Law Review* 22 (2001) 1691–1731.
- [40] J. Mortera, A.P. Dawid, S.L. Lauritzen, Probabilistic expert systems for DNA mixture profiling, *Theoretical Population Biology* 63 (2003) 191–205.
- [41] P. Thagart, Why wasn't O.J. convicted? Emotional coherence and legal inference, *Cognition and Emotion* 17 (2003) 361–383.
- [42] N. Fenton, M. Neil, The 'jury observation fallacy' and the use of Bayesian networks to present probabilistic legal arguments, *Mathematics Today - Bulletin of the IMA* 36 (2000) 180–187.
- [43] C. Jowett, Lies, damned lies, and DNA statistics: DNA match testing, Bayes' theorem, and the criminal courts, *Medicine, Science & Law* 41 (2001) 194–205.
- [44] C. Jowett, Sittin' in the dock with the Bayes, *New Law Journal Expert Witness Supplement* 151 (2001) 201–212.
- [45] A.P. Dawid, I.W. Evett, Using a graphical method to assist the evaluation of complicated patterns of evidence, *Journal of Forensic Sciences* 42 (1997) 226–231.
- [46] C.G.G. Aitken, A. Gammerman, Probabilistic reasoning in evidential assessment, *Journal of the Forensic Science Society* 29 (1989) 303–316.
- [47] C.G.G. Aitken, T. Connolly, A. Gammerman, G. Zhang, D. Bailey, R. Gordon, R. Oldfield, Statistical modelling in specific case analysis, *Science & Justice* 36 (1996) 245–255.
- [48] C.G.G. Aitken, A. Gammerman, G. Zhang, T. Connolly, D. Bailey, R. Gordon, R. Oldfield, Bayesian belief networks with an application in specific case analysis, in: A. Gammerman (Ed.), *Computational Learning and Probabilistic Reasoning*, John Wiley and Sons, Chichester, 1996, pp. 169–184.
- [49] P. Garbolino, F. Taroni, Evaluation of scientific evidence using Bayesian networks, *Forensic Science International* 125 (2002) 149–155.
- [50] C.G.G. Aitken, F. Taroni, P. Garbolino, A graphical model for the evaluation of cross-transfer evidence in DNA profiles, *Theoretical Population Biology* 63 (2003) 179–190.
- [51] D.A. Stoney, Relaxation of the assumption of relevance and an application to one-trace and two-trace problems, *Journal of the Forensic Science Society* 34 (1994) 17–21.
- [52] I.W. Evett, Establishing the evidential value of a small quantity of material found at a crime scene, *Journal of the Forensic Science Society* 33 (1993) 83–86.
- [53] I.W. Evett, P.D. Gill, G. Jackson, J. Whitaker, C. Champod, Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks, *Journal of Forensic Sciences* 47 (2002) 520–530.
- [54] R. Cook, I.W. Evett, G. Jackson, P.J. Jones, J.A. Lambert, A hierarchy of propositions: deciding which level to address in casework, *Science & Justice* 38 (1998) 231–239.
- [55] R. Cook, I.W. Evett, G. Jackson, P.J. Jones, J.A. Lambert, A model for case assessment and interpretation, *Science & Justice* 38 (1998) 151–156.
- [56] R. Cook, I.W. Evett, G. Jackson, P.J. Jones, J.A. Lambert, Case pre-assessment and review in a two-way transfer case, *Science & Justice* 39 (1999) 103–111.
- [57] I.W. Evett, G. Jackson, J.A. Lambert, More on the hierarchy of propositions: exploring the distinction between explanations and propositions, *Science & Justice* 40 (2000) 3–10.
- [58] A.P. Dawid, D.W. van Boxel, J. Mortera, V.L. Pascali, Inference about disputed paternity from an incomplete pedigree using a probabilistic expert system, *Bulletin of the International Statistical Institute* 58 (Book 1) (1999) 241–242.
- [59] A.P. Dawid, J. Mortera, V.L. Pascali, D. van Boxel, Probabilistic expert systems for forensic inference from genetic markers, *Scandinavian Journal of Statistics* 29 (2002) 577–595.
- [60] E. Essen-Möller, Die Beweiskraft der Ähnlichkeit im Vaterschaftsnachweis. Theoretische Grundlagen, Mitteilungen der Anthropologischen Gesellschaft (Wien) 68 (1938) 9–53.
- [61] J. Mortera, Analysis of DNA mixtures using probabilistic expert systems, in: P.L. Green, N.L. Hjort, S. Richardson (Eds.), *Highly Structured Stochastic Systems*, Oxford University Press, Oxford, 2003.
- [62] F. Taroni, C.G.G. Aitken, G. Garbolino, A. Biedermann, *Bayesian Networks and Probabilistic Inference in Forensic Science*, John Wiley & Sons, Chichester, 2006.
- [63] I.W. Evett, Evaluating DNA profiles in a case where the defence is "It was my brother", *Journal of the Forensic Science Society* 32 (1992) 5–14.
- [64] C.J. Bruijning-van Dongen, K. Slooten, W. Burgers, W. Wiegerinck, Bayesian networks for victim identification on the basis of DNA profiles, *Forensic Science International: Genetics Supplement Series 2* (2009) 466–468, *Progress in Forensic Genetics 13 - Proceedings of the 23rd International ISFG Congress*.
- [65] D. Koller, A. Pfeffer, Object-oriented Bayesian networks, in: *Proceedings of the Thirteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-97)*, 1997, pp. 302–313.
- [66] K.B. Laskey, S.M. Mahoney, Network fragments: representing knowledge for constructing probabilistic models, in: *Proceedings of the Thirteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-97)*, 1997, pp. 334–341.
- [67] O. Bangsø, P.H. Wuillemin, Object oriented Bayesian networks: a framework for top-down specification of large Bayesian networks with repetitive structures, *Tech. rep.*, Hewlett-Packard Laboratory for Normative Systems, Aalborg University, Aalborg, (2000).
- [68] M. Neil, N. Fenton, L. Nielsen, Building large-scale Bayesian networks, *The Knowledge Engineering Review* 15 (2000) 257–284.
- [69] A. Hepler, A.P. Dawid, V. Leucari, Object-oriented graphical representations of complex patterns of evidence, *Law Probability and Risk* 6 (2007) 275–293.
- [70] A.P. Dawid, J. Mortera, P. Vicard, Representing and solving complex DNA identification cases using Bayesian networks, *International Congress Series* 1288 (2006) 484–491.
- [71] A.P. Dawid, J. Mortera, P. Vicard, Object-oriented Bayesian networks for complex forensic DNA profiling problems, *Forensic Science International* 169 (2007) 195–205.
- [72] A. Hepler, B.S. Weir, Object-oriented Bayesian networks for paternity cases with allelic dependencies, *Forensic Science International: Genetics* 2 (2008) 166–175.
- [73] A.P. Dawid, An object-oriented Bayesian network for estimating mutation rates, in: *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, 2003.
- [74] P. Vicard, A.P. Dawid, J. Mortera, S.L. Lauritzen, Estimating mutation rates from paternity casework, *Forensic Science International: Genetics* 2 (2008) 9–18.
- [75] R.G. Cowell, S.L. Lauritzen, J. Mortera, Identification and separation of DNA mixtures using peak area information, *Forensic Science International* 166 (2007) 28–34.
- [76] R.G. Cowell, S.L. Lauritzen, J. Mortera, Probabilistic modelling for DNA mixture analysis, *Forensic Science International: Genetics Supplement Series 1* (2008) 640–642, *Progress in Forensic Genetics 12 - Proceedings of the 22nd International ISFG Congress*.
- [77] R.G. Cowell, Validation of an STR peak area model, *Forensic Science International: Genetics* 3 (2009) 193–199.
- [78] D. Cavallini, F. Corradi, Forensic identification of individuals included in a database of DNA profiles, *Biometrika* 93 (2006) 525–536.
- [79] F. Taroni, A. Biedermann, P. Garbolino, C.G.G. Aitken, A general approach to Bayesian networks for the interpretation of evidence, *Forensic Science International* 139 (2004) 5–16.
- [80] I.W. Evett, B.S. Weir, *Interpreting DNA Evidence*, Sinauer Associates Inc., Sunderland, 1998.
- [81] P.J. Green, J. Mortera, Sensitivity of inferences in forensic genetics to assumptions about funding genes, *The Annals of Applied Statistics* 3 (2009) 731–763.
- [82] J.M. Butler, R. Schoske, P.M. Vallone, J.W. Redman, M.C. Kline, Allele frequencies of 15 autosomal STR loci on U.S. Caucasian, African American and Hispanic populations, *Journal of Forensic Sciences* 48 (2003) 908–911.
- [83] W.C. Thompson, F. Taroni, C.G.G. Aitken, How the probability of a false positive affects the value of DNA evidence, *Journal of Forensic Sciences* 48 (2003) 47–54.
- [84] D. Hatsch, C. Keyser, R. Hienne, L. Bertrand, Resolving paternity relationships using X-chromosome STRs and Bayesian networks, *Journal of Forensic Sciences* 52 (2007) 895–897.
- [85] P. Gill, R. Puch-Solis, J. Curran, The *low-template-DNA* (stochastic) threshold - its determination relative to risk analysis for national DNA databases, *Forensic Science International: Genetics* 3 (2009) 104–111.
- [86] J.S. Buckleton, C.M. Triggs, C. Champod, An extended likelihood ratio framework for interpreting evidence, *Science & Justice* 46 (2006) 69–78.
- [87] R.G. Cowell, FINEX: a probabilistic expert system for forensic identification, *Forensic Science International* 134 (2003) 196–206.
- [88] A. Biedermann, F. Taroni, A probabilistic approach to the joint evaluation of firearm evidence and gunshot residues, *Forensic Science International* 163 (2006) 18–33.
- [89] M.I. Jordan (Ed.), *Learning in Graphical Models*, The MIT Press, London, 1999.
- [90] J. Pearl, *Causality: Models, Reasoning and Inference*, Cambridge University Press, Cambridge, 2000.
- [91] S.L. Lauritzen, N.A. Sheehan, Graphical models for genetic analyses, *Statistical Science* 18 (2003) 489–514.
- [92] P. Nam, P.H. Wuillemin, P. Leray, O. Pourret, A. Becker, *Réseaux Bayésiens*, 3rd Edition, Eyrolles, Paris, 2007.
- [93] D.H. Kaye, The validity of tests: Caveant omnes, *Jurimetrics Journal* (1987) 349–361.
- [94] S.L. Lauritzen, D.J. Spiegelhalter, Local computations with probabilities on graphical structures and their application to expert systems, *Journal of the Royal Statistical Society, Series B* 50 (1988) 157–224.