Science and Justice 52 (2012) 191–198

Contents lists available at SciVerse ScienceDirect

Science and Justice

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

Learning about Bayesian networks for forensic interpretation: An example based on the 'the problem of multiple propositions'

A. Biedermann *, R. Voisard, F. Taroni

University of Lausanne, School of Criminal Justice, Institute of Forensic Science, Switzerland

ARTICLE INFO

Article history:
Received 6 July 2011
Received in revised form 7 May 2012
Accepted 8 May 2012

Keywords:
Bayesian networks
Likelihood ratio
Multiple propositions
Handwriting examination
Teaching

ABSTRACT

Both, Bayesian networks and probabilistic evaluation are gaining more and more widespread use within many professional branches, including forensic science. Notwithstanding, they constitute subtle topics with definitional details that require careful study. While many sophisticated developments of probabilistic approaches to evaluation of forensic findings may readily be found in published literature, there remains a gap with respect to writings that focus on foundational aspects and on how these may be acquired by interested scientists new to these topics. This paper takes this as a starting point to report on the learning about Bayesian networks for likelihood ratio based, probabilistic inference procedures in a class of master students in forensic science. The presentation uses an example that relies on a casework scenario drawn from published literature, involving a questioned signature. A complicating aspect of that case study – proposed to students in a teaching scenario – is due to the need of considering multiple competing propositions, which is an outset that may not readily be approached within a likelihood ratio based framework without drawing attention to some additional technical details. Using generic Bayesian networks fragments from existing literature on the topic, course participants were able to track the probabilistic underpinnings of the proposed scenario correctly both in terms of likelihood ratios and of posterior probabilities. In addition, further study of the example by students allowed them to derive an alternative Bayesian network structure with a computational output that is equivalent to existing probabilistic solutions. This practical experience underlines the potential of Bayesian networks to support and clarify foundational principles of probabilistic procedures for forensic evaluation.

1. Introduction

Probability theory now is widely accepted as the reference concept for dealing with uncertainty associated with the evaluation of forensic findings. Many textbooks are currently available that provide a broad scope of applications (e.g., [1–4]). Some resources are even dedicated exclusively to selected items of scientific evidence, such as DNA (e.g., [5,6]). It has however been recognised that, in many situations, the application of probability theory to real-world problems may become increasingly difficult because of the number of variables that need to be accounted for, as well as the degree of complication in their dependency structure. This outset has led researchers and practitioners to focus their attention to a graphical implementation of probability theory, known as Bayesian networks [7]. Developed in the area of artificial intelligence in the early 1980s, Bayesian networks subsequently attracted interest also among legal scholars (e.g., [8,9]). Although there now is a substantial body of literature on reported applications of Bayesian networks, it appears that they still are used by only a small group of scientists and researchers. In fact, practitioners often mention that applications reported in published literature – despite being of great interest – are difficult to apprehend without solid prior knowledge of the concept of Bayesian networks. It addition, the relations between particular Bayesian network structures and formal procedures for probabilistic evidence evaluation (e.g., in the form of likelihood ratio formulae) also present a recurrent instance of discussion.

An initial knowledge and understanding of the essential underpinnings of Bayesian networks may, however, not be easily acquired. Several textbooks in artificial intelligence and computer science are available, but practical experience shows that learners may find it difficult to make a step forward from a formal statement of the definition of Bayesian networks to an informed understanding of the detailed properties of the concept, as well as their mastering for modelling practical inference problems. This is a challenge that is not only encountered in forensic science, but is also reported in more specialist areas such as artificial intelligence and computer science (e.g., [10]). Indeed, there appears to be a breach in literature between a great abundance of reported applications on the one side, and very sparse contributions on how the concept may be usefully taught to interested practitioners. The paper here chooses this

* Corresponding author.
E-mail addresses: alex.biedermann@unil.ch (A. Biedermann), romain.voisard@unil.ch (R. Voisard), franco.taroni@unil.ch (F. Taroni).

1355-0306/$ – see front matter © 2012 Forensic Science Society. Published by Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.scijus.2012.05.004

outset as a starting point and reports on a practical experience of teaching Bayesian networks and likelihood ratio assignment in a class of master students.

In the context of forensic science, there is literature on how basic inference problems may be translated into appropriate Bayesian network representations (e.g., [7,11]). The paper [11], for example, contains a series of practical examples that illustrate the basic connections in Bayesian networks, as well as their properties (in terms of dependencies and independencies). There are, however, a series of additional aspects that users new to Bayesian networks may encounter, and these may require further discussion and provision of answers. Typical examples for this are the handling of settings involving multiple propositions, the formulation of likelihood ratios (or, their components) as well as the assessment of particular Bayesian network structures (i.e., similarities and differences between distinct network models, and their connections to existing likelihood ratio procedures).

The purpose of the paper here is to discuss this outset in some further detail through an example drawn from work with a class of master students in forensic science. Section 2 provides a description of the principal components of the teaching scenario designed for this task. Section 2.1 provides a formal description of a handwriting case adapted from existing literature, and proposed to students as a case study. Subsequent parts focus on further aspects such as tools and background information provided (Section 2.2), as well as the learning intentions (Section 2.3). Section 3 will continue by presenting different approaches on how the proposed handwriting case may be modelled through Bayesian networks. The suggested network approaches are compared against each other and set in relation to the formulaic approach described in Section 2.2.1. Besides, a discussion is included on how course participants approached the scenario with an alternative solution. This provides an indication about achieved learning outcomes during the practical learning session.

2. Components of the teaching scenario

2.1. Description of casework example and preliminary notes

The casework scenario initially proposed to students involves a questioned signature and is based on a casework example described in Köller et al. [12, p. 132]:

> "The material received included a letter dated 20 October 1997 containing an illegible abbreviated signature […], which, according to the letterhead, allegedly came from the letter-writer Martin Schmidt. […] Also received were 20 spontaneously produced comparative abbreviated signatures contained in various letters by the letter-writer Martin Schmidt during the period 1992 to 2002. […] the question posed was whether the illegible abbreviated signature under the letter of 20 October 1997 is authentic, i.e., written by Martin Schmidt, or not."

Note that students did not conduct any comparative handwriting examinations. They only read a written evaluative report on such examinations as given in [12]. One of the main reasons for choosing this scenario was that it is one of the few published examples that describe a full quantitative approach to handwriting evidence evaluation. More generally, handwriting is also chosen here as an exemplary area of application because it is not regularly approached in a genuine probabilistic perspective (such as DNA evidence, for example). In fact, besides some formative discussion on a Bayesian approach presented in [13], there is rather sparse further literature on this topic.

The various numerical assignments of probability in [12] will not be made an issue as such throughout this paper. It is entirely recognised here that they may be assessed differently according to the situation at hand as well as the circumstantial information available to the scientist. The main point with these numerical assignments is that they allow one to fully specify a probabilistic evaluation procedure, to do computations and to keep track of numerical results. This is an important aspect for the comparative study and discussion of likelihood ratios and Bayesian networks.

2.2. Provision of tools and background information

2.2.1. Probabilistic approach for results of comparative handwriting examinations

Part of the tools and starting information provided to learners consists of a formal description of a probabilistic approach for results of comparative handwriting examinations, outlined in some further detail here below. This approach specifies a pair of competing propositions which, in the setting of interest in this paper, are framed as follows:

- H_p: Writer Martin Schmidt wrote the questioned signature.
- H_d: The questioned signature was not written by Martin Schmidt, but by another person.

A complicating aspect of the proposed scenario (Section 2.1) is that the questioned signature may have been written under varying circumstances, which, in some very general sense, may be described as 'ordinary' or 'unusual' (still according to [12]). The latter situation is one that is thought to account for the true writer's intent to disguise, to neutralize, or to imitate. Consideration of these two categories of circumstances led Köller et al. [12] to extend the competing propositions to the following[1]:

- H_{p1}: Writer Martin Schmidt wrote the questioned signature under ordinary conditions.
- H_{p2}: Writer Martin Schmidt wrote the questioned signature under unusual conditions.
- H_{d1}: The questioned signature was not written by Martin Schmidt, but by another person, under normal conditions.
- H_{d2}: The questioned signature was not written by Martin Schmidt, but by another person, under unusual conditions.

The initial probabilities proposed for these extended propositions are as follows [12]:

$$Pr(H_{p1}|I) = 0.45, \; Pr(H_{p2}|I) = 0.05,$$
$$Pr(H_{d1}|I) = 0.02, \; Pr(H_{d2}|I) = 0.48.$$

In this notation, the term I represents circumstantial information that is used as a conditioning for the assignment of these probabilities.

The results of the comparative handwriting examinations are described as follows [12, p. 135]:

> "The comparative examinations performed on the samples from the comparison writer revealed the presence of extremely close correspondence of handwriting characteristics. The graphic characteristics of the disputed signature fit well within the observable range of variations in the comparative samples. No noteworthy differences in characteristics were detected."

Let these findings be summarised by a binary variable E. Conditional probabilities proposed in [12] are summarised in Table 1. The probabilities for the negation of E, written $-E$, are also included here for the sake of completeness and because they are an integral part of the Bayesian networks described in later sections of this paper.

[1] As will become clear as discussion proceeds throughout this paper, the introduction of these additional levels of considerations make this example particularly instructive for pointing out the value that Bayesian networks may add with respect to the purely formal approaches based, for example, on likelihood ratio formulae alone.

Table 1
Conditional probabilities for findings E of comparative handwriting examinations given each of the specified competing propositions H.

H:	H_{p1}	H_{p2}	H_{d1}	H_{d2}
E:	0.95	0.06	0.01	0.95
$-E$:	0.05	0.94	0.99	0.05

Based on the probability assignments outlined above, the posterior probability for the proposition according to which Martin Schmidt is the writer of the questioned signature under normal conditions is given by:

$$Pr(H_{p1}|E,I) = \frac{Pr(E|H_{p1},I)Pr(H_{p1}|I)}{\sum_i Pr(E|H_i,I)Pr(H_i|I)}, \ for \ \ i = \{p1, p2, d1, d2\}. \tag{1}$$

For the case considered here, this leads to a posterior probability[2] $Pr(H_{p1}|E) = 0.4821$. In an analogous way, Eq. (1) can also be formulated for the other propositions and this leads to:

$$Pr(H_{p2}|E) = 0.0034, \ Pr(H_{d1}|E) = 0.0002, \ Pr(H_{d2}|E) = 0.5143.$$

For the time being, the reader will be asked to consider this formal approach as a running example. It is acknowledged at this point that the description and definition of the findings E as well as the set of competing target propositions H may be framed in a different way, according to the scientist's preferences.[3] This is not, however, made a principal issue here. The main issue here will be that of analysing a given formal framework through Bayesian networks (more formally introduced in Section 2.2.2), once one can agree on a given initial framework. The resulting Bayesian networks can then serve as a starting point for studying further extensions and refinements of modeling assumptions that readers may wish to change.

2.2.2. Preliminaries on Bayesian networks

Prior to addressing the scenario outlined in Section 2.1, it is often useful to assure that participants are well acquainted with the main constituting elements of Bayesian networks as well as the software program that is going to be used for their practical manipulation. This thus represents a second main aspect of the tools and background information provided to students.

A convenient way of initiating discussion about the concept of Bayesian networks is by means of a basic network involving only two variables, using notation with which learners are familiar from previous lessons on probabilistic evaluation of scientific results (or, findings). This allows them to recognise that Bayesian networks are merely a graphical implementation of the principal elements of Bayes' theorem, both on a static (i.e., purely representational) and a dynamic level (i.e., in terms of computations).

Initially, learners may thus be asked to imagine a rational evaluator who considers a scientific finding E with the aim of drawing an inference to some proposition of interest, denoted H. For the time being, it will be assumed that the variables have binary states. The generic scientific observation (finding) is described by the state E whereas the negation of this kind of observation is denoted $-E$. In turn, the variable H covers the propositions H_p and H_d. These represent positions put forward by, respectively, the prosecution and the defense.

In terms of a Bayesian network, one can thus consider a node E with an entering arc from another node, denoted H. This results in a network fragment of the kind $H \rightarrow E$. This is an expression of the view that the probability of the observation of interest depends on the actual state of the node H. This latter node is in exactly one of its states. That is, either H_p is true, or H_d is true, although it will be imagined that one may not be able to tell which of these two settings actually applies. This is why probabilities $Pr(H_p)$ and $Pr(H_d)$ are assigned to these two target states. Given the arc from H to E, one will be required to specify conditional probabilities for E given H. Because findings and observations as encountered in the real world usually are imperfect to some degree, one should expect particular findings not only to occur when what one is trying to prove is true, but also when what one is trying to prove is not true [3]. A consequence of the latter view is that $Pr(E|H_d) \neq 0$. Whenever the evidence is believed, for example, more likely to occur when what one is trying to prove is true, than when the specified alternative is true, one would assign probabilities so that $Pr(E|H_p) > Pr(E|H_d)$. This is a qualitative expression of the likelihood ratio [14].

Fig. 1 provides a pictorial summary of the graphical relationship between the variables E and H as well as the respective node probabilities. Using one of the various commercially or academically available systems, a further step will then consist in implementing this Bayesian network in a computerised format. The Bayesian network thus constructed will allow one to illustrate the following:

- Initially (i.e., before any instantiation is made), the node H displays the probabilities defined for each of the two states of this node. In turn, the node E will display probabilities that depend on the uncertainty about the actual state of the node H. This is captured in terms of the so-called rule of the extension of conversation [15]:

$$Pr(E) = Pr(E|H_p)Pr(H_p) + Pr(E|H_d)Pr(H_d) \tag{2}$$

- When node H is set to one of its states (i.e., H_p or H_d), then the node E will display the probability specified for that event, conditional on the actual state H_p or H_d. For example, if $H = H_p$, then the node E will display the probability $Pr(E|H_p)$ for the state E. This can immediately be seen from Eq. (2), which reduces to $Pr(E|H_p)$ when $Pr(H_p) = 1$ and $Pr(H_d) = 0$. This is also called a 'predictive' line of reasoning because knowledge about the actual state of the variable H is put forward (i.e., in the direction of the arc) to the node E.
- When node E is set to one of its states (i.e., to E or $-E$), then the node H will display posterior probabilities calculated according to Bayes' theorem. For example, if E is observed, then the posterior probability for H is obtained as follows:

$$Pr(H_p|E) = \frac{Pr(E|H_p)Pr(H_p)}{Pr(E|H_p)Pr(H_p) + Pr(E|H_d)Pr(H_d)}. \tag{3}$$

This is an inference against the direction of the network's arc. In the context, it is also sometimes called 'inductive' inference.

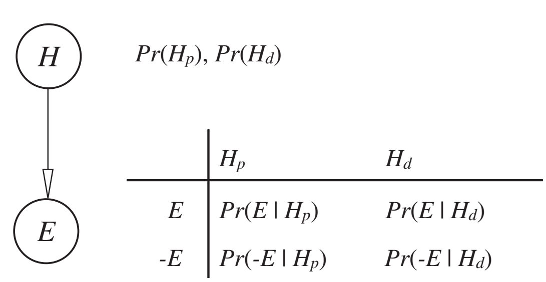

Fig. 1. Bayesian network for evaluating a single item of information (i.e., a scientific finding E) that is relevant for inference about a target proposition H.

[2] For the ease of presentation, the conditioning on circumstantial information, I, will be omitted from notation in the remainder of this paper.
[3] In [13], for instance, more detailed categories termed 'some', 'few' and 'many similarities' are used as descriptors of the outcomes of comparative handwriting examinations.

These general Bayesian network manipulations allow one to illustrate the nature of what – in the context – is called 'evidence propagation', that is an operation that is concerned with evaluating the conditional probability of nodes of interest given the observed values for one or several other nodes. Stated otherwise, during the process of informing a Bayesian network about what observations have become available, the states of certain nodes, also called 'evidence nodes', are fixed to particular values. The operations mentioned above may appear simple, if not to say banal. Notwithstanding, it is often useful to remind users of these aspects in order to point out that the structure, definition and output of Bayesian networks is by no means arbitrary, but the result of well-defined properties.

2.3. Description of teaching intentions

In essence, the probabilistic inference procedure presented in Section 2.2.1 amounts to a basic application of Bayes' theorem for multiple discrete propositions. Students may however feel uncomfortable with such an approach because it directs them to focus on a posterior probability of a principal proposition, given particular observations (i.e., findings).

In fact, a predominant part of current forensic evaluation procedures focus on the likelihood ratio, rather than on an opinion on an issue that is outside the scientists' area of competence. Therefore, the case as described so far provides a viable starting point for a series of reflective questions, such as:

- Can we conceive of a Bayesian network that allows one to capture Eq. (1)? Can different networks of this kind be constructed? If so, what are their similarities and differences?
- Instead of focusing on posterior probabilities, can we define a procedure that allows the coherent assignment of a likelihood ratio?
- Can Bayesian networks support likelihood ratio assignment, in particular when multiple competing propositions need to be accounted for? If so, how?

The purpose of using Bayesian networks in this context thus is to encourage students to approach a problem in a distinct and new way. This should help them to get acquainted with a tool that allows them to search for a solution which could be useful in their own casework. In addition, students are also encouraged to seek and develop alternative solutions, and to explain their choices.

2.4. Discussion of different Bayesian network approaches

The forthcoming Sections 3.1 and 3.2 will approach the reflective questions mentioned in Section 2.3 above through discussion and analysis of Bayesian network models from published literature. These model representations can be proposed to students who could not find a convenient or satisfactory network structure during the practical learning session. Alternatively, these structures can also be proposed to students for self-correction. The intention of this is that students examine these models in order to detect similarities and differences with respect to their own models.

Section 3.3 will present a particular Bayesian network that was developed by a group of master students in forensic science in a practical, computer-assisted learning lesson (at the authors' institution). This latter Bayesian network is entirely in agreement with the generic Bayesian networks outlined in the previous sections as well as with a more formally defined likelihood ratio procedure. It also provides additional insight into the evaluative questions that are associated with the casework example introduced in Section 1. In particular, the proposed network suggests that the scenario under study is not a genuine multiple proposition setting, but one that involves an amalgam of propositions. Their distinct representation in terms of a Bayesian network, as proposed by the students, helps to clarify this aspect and to set the evaluative task more appropriately in context.

3. Bayesian network modelling for the questioned signature case: outline of the various approaches

3.1. Approach 1

Modelling the handwriting scenario outlined earlier in Section 2.1 in terms of a Bayesian network is readily achieved if the preliminaries considered in Section 2.2.2 are well captured. On a structural account, there still are two variables, E (for the observations) and H (for the target proposition). The main difference between the generic Bayesian inference modelled in Section 2.2.2, and the inference approach to the handwriting case as presented in Section 2.1 consists in the number of partitions for the proposition H. This means that, instead of a node H with two propositions H_p and H_d, a node with four states H_{p1}, H_{p2}, H_{d1} and H_{d2} is needed. Accordingly, the probability table for the node E will also increase as defined earlier by Table 1.

As noted in Section 2.1, working with a set of multiple propositions $\{H_{p1}, H_{p2}, H_{d1}, H_{d2}\}$ may lead one to considering posterior probabilities for main propositions, given the findings E. A question of interest may thus be that of how to provide a likelihood ratio. The Bayesian network $H \rightarrow E$, with multiple states for node H is not, however, directly amenable for finding a likelihood ratio of the kind $Pr(E|H_p)/Pr(E|H_d)$. The reason for this is that, habitually, one is able to instantiate a single state of a given node to 'known', and this will result in setting the probability of the remaining states to zero. For the case considered here, however, there are two propositions for the numerator, H_{p1} and H_{p2}, and these cannot be set to 'true' simultaneously.

One way to avoid this complication is to regroup the propositions H_{p1} and H_{p2} in a single state H_p of a new node H' [16]. Analogously, one may regroup propositions H_{d1} and H_{d2} in a state H_d. That is, one can consider H_p to be established whenever either H_{p1} or H_{p2} holds, and H_d to established whenever either H_{d1} or H_{d2} is the case. Here, H_p and H_d refer to the authorship by, respectively, the suspect and an unknown person under *any* writing circumstance (i.e., normal or unusual). The binary node H' thus is adopted as a direct descendant of the node H, as shown in Fig. 2. The probability table for the node H' is completed using assignments of zero and one as shown by Table 2.

Using the numerical assignments defined in Section 2.1, the initial state of the Bayesian network is as shown in Fig. 2 (i). In order to obtain a value for the numerator of the likelihood ratio, that is $Pr(E|H_p)$, on needs to set the probability of the state H_p of the node H' to one. The effect of this is twofold:

- The node H displays probabilities 0.9 and 0.1 for, respectively, states H_{p1} and H_{p2}. The probabilities of the states H_{d1} and H_{d2} are zero. This is a direct consequence of the numerical specification provided in Table 2, which states that H_p is only 'compatible' with H_{p1} and H_{p2}. In an application of Bayes' theorem for inference about H, on the basis of H', this may lead to zero posterior probabilities. For example, $Pr(H_{d1}|H'=H_p)=0$ because the calculation $[Pr(H'=H_p|H_{d1})Pr(H_{d1})]/Pr(H'=H_p)$ contains a zero likelihood in the numerator. When probability is 'withdrawn' from the two states H_{d1} and H_{d2}, then it must be redistributed over the remaining two states H_{p1} and H_{p2}. This

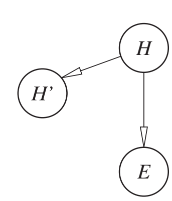

Fig. 2. Bayesian network for evaluating scientific findings E (with possible states E and $-E$) that is relevant for inference about the set of multiple target propositions $\{H_{p1}, H_{p2}, H_{d1}, H_{d2}\}$, represented in terms of the node H. The node H' is binary and regroups the pair of propositions H_{p1} and H_{p2} in a state H_p and the pair H_{d1} and H_{d2} in a state H_d.

Table 2
Logical assignments for the node H'.

	H:	H_{p1}	H_{p2}	H_{d1}	H_{d2}
H':	H_p	1	1	0	0
	H_d	0	0	1	1

redistribution is operated proportionally, that is the probabilities for H_{p1} and H_{p2} keep their relative magnitudes. Initially, they were 0.45 and 0.05, respectively, as shown in Fig. 2 (i). Given H_p, they became 0.9 and 0.1. This is a process of 'normalisation' obtained by dividing the probabilities H_{p1} and H_{p2} by their sum:

$$Pr(H = H_{p1}|H' = H_p) = \frac{Pr(H_{p1})}{Pr(H_{p1}) + Pr(H_{p2})}, \tag{4}$$

$$Pr(H = H_{p2}|H' = H_p) = \frac{Pr(H_{p2})}{Pr(H_{p1}) + Pr(H_{p2})}.$$

- These two values can be interpreted as probabilities for, respectively, normal and unusual writing conditions if it is the suspect who wrote the questioned signature. Notice that all of the above arguments apply analogously to the states H_{d1} and H_{d2} for a situation in which H_d is assumed to hold. As shown by Fig. 2 (iii), one can find $Pr(H_{d1}|H'=H_d)=0.04$ and $Pr(H_{d2}|H'=H_d)=0.96$.
- The node E displays the probability of E given H_p. This probability is still obtained by the rule of the extension of the conversation, defined earlier in Eq. (2). Applied to the setting here, one can write this rule as follows

$$Pr(E|H' = H_p) = Pr(E|H_{p1})Pr(H_{p1}|H_p) + Pr(E|H_{p2})Pr(H_{p2}|H_p). \tag{5}$$

- The extension to H_{d1} and H_{d2} is omitted from notation here because, given H_p, the probabilities of these two states are zero. As can be seen, for obtaining a value for the numerator of the likelihood ratio, it is correct to consider an extension to H_{p1} and H_{p2}, but it would not be correct to weight the conditional probabilities $Pr(E|H_{p1})$ and $Pr(E|H_{p2})$ only by $Pr(H_{p1})$ and $Pr(H_{p2})$, because the latter two probabilities do not sum to one. Instead, one needs to work with $Pr(H_{p1}|H'=H_p)$ and $Pr(H_{p2}|H'=H_p)$ as defined by Eq. (4).

On the basis of these considerations, one can find the following likelihood ratio:

$$LR = \frac{Pr(E|H_p)}{Pr(E|H_d)} = \frac{Pr(E|H_{p1})Pr(H_{p1}|H_p) + Pr(E|H_{p2})Pr(H_{p2}|H_p)}{Pr(E|H_{d1})Pr(H_{d1}|H_d) + Pr(E|H_{d2})Pr(H_{d2}|H_d)}$$

$$= \frac{0.95 \times 0.9 + 0.06 \times 0.1}{0.01 \times 0.04 + 0.95 \times 0.96} = \frac{0.861}{0.9124} \approx 0.94.$$

Table 3
Conditional probabilities for the node H, representing the full set of propositions $\{H_{p1}, H_{p2}, H_{d1}, H_{d2}\}$, given each of the overall propositions H_p and H_d.

	H':	H_p	H_d
H:	H_{p1}	0.9	0
	H_{p2}	0.1	0
	H_{d1}	0	0.04
	H_{p2}	0	0.96

3.2. Approach 2

An alternative approach to dealing with multiple propositions has been described by Buckleton et al. [17]. These authors represent the principal pair of competing propositions, H_p and H_d, in terms of a root note. Applied to the case studied here, this leads to a graphical representation as shown in Fig. 4. The main difference to the Bayesian network discussed in the previous section is the reversed arc between the two nodes H' and H. While the states of these two nodes remain the same, their probability tables change. Here, the node H' does not receive entering arcs from other nodes. Its node table thus contains unconditional probabilities $Pr(H_p)$ and $Pr(H_d)$. For purely technical reasons, probabilities 0.5 are assigned here to each of the latter two states. The reader may choose other values, but in essence these remain irrelevant as soon as one focuses on the evaluation of the probability of the finding E, given each of the competing target propositions. That is, instantiating H' to one of its states renders the probabilities initially assigned to this node ineffective.

More attention needs to be drawn to the probability table of the node H, which now contains conditional probabilities. Here one is directed to inquire, given H_p, about the probability of the suspect writing under normal (H_{p1}) and unusual (H_{p2}) conditions. The probabilities assigned to H_{d1} and H_{d2} are zero because they do not apply to situations in which H_p holds. Given H_d, the opposite holds and one specifies probabilities for H_{d1} and H_{d2} while zero probabilities are assigned to H_{p1} and H_{p2}. A summary of these assignments is given in Table 3.

As may be seen, the numerical values in Table 3 correspond to the posterior probabilities calculated in the previous section for the states $\{H_{p1}, H_{p2}, H_{d1}, H_{d2}\}$ of the node H, given H'. That is, the posterior probabilities for H shown in Fig. 3 (ii), given H_p, are now used in the lefthand-side of Table 3. In turn, the posterior probabilities for H shown in Fig. 3 (iii), given H_d, represent the probabilities in the far righthand-side of Table 3. The assignment of these probabilities can also be understood as a transcription of the initially defined prior probabilities in Section 2.1, that is

$$Pr(H_{p1}|I) = 0.45,\ Pr(H_{p2}|I) = 0.05,\ Pr(H_{d1}|I) = 0.02,\ Pr(H_{d2}|I) = 0.48,$$

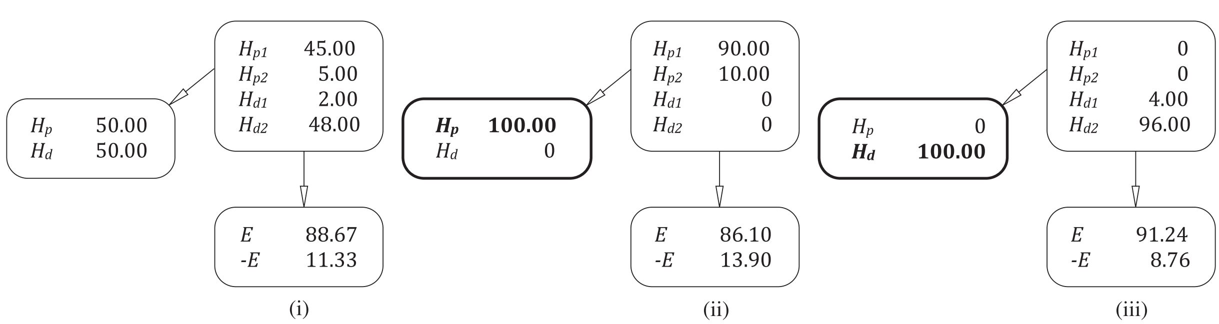

Fig. 3. Bayesian network defined earlier in Fig. 2. The initial state is shown in figure (i) whereas figures (ii) and (iii) show situations in which the node H is set to, respectively, H_p and H_d (instantiated nodes are indicated by a thicker border).

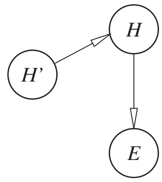

Fig. 4. Alternative Bayesian network structure for evaluating results E of comparative handwriting examinations. The node H covers the target propositions $\{H_{p1}, H_{p2}, H_{d1}, H_{d2}\}$. The node H' is binary and represents the pair of ultimate competing propositions H_p and H_d.

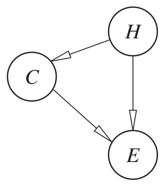

Fig. 5. Bayesian network structure for evaluating results E of comparative handwriting examinations. The node H covers the target propositions H_p and H_d whereas the node C models writing conditions defined as normal (C) and unusual ($-C$).

in terms of two nodes H and H', rather than one. Although $Pr(H_{p1}|H_p) = 0.9$ and $Pr(H_{p2}|H_p) = 0.1$, the node H' will display $Pr(H_{p1}) = 0.45$ and $Pr(H_{p2}) = 0.05$ when the node H is not instantiated, that is when H_p has only a probability of 0.5 to be true. Analogous argument applies for the states H_{d1} and H_{d2} given H_d.

The structural relationship between the nodes E and H is the same for the models shown in Figs. 2 and 4. The node table for E thus remains unchanged as well. The overall numerical output at the node E, given instantiations made at the node H, will also be in entire agreement with the findings of Section 3.1. In particular, the probability at the observational node E is still calculated according to the argument presented in bullet point 2 in Section 3.1 (Eq. (5)).

3.3. Approach 3

A further possibility for coping with the handwriting scenario outlined in Section 2.1 diverges slightly with respect to the two approaches discussed so far in this paper. As a main distinctive feature, it relies on a decomposition of the central set of propositions $\{H_{p1}, H_{p2}, H_{d1}, H_{d2}\}$ into two separate pairs of propositions:

- A first pair of propositions is represented by a node H, defined by two states H_p and H_d in the same way as done in the previous sections.
- A second pair of propositions focuses exclusively on the writing circumstances. This pair of propositions is represented by a node C with states referring to, respectively, normal (C) and unusual ($-C$) writing circumstances.

This separation represents a crucial insight for the currently studied scenario. It allows one to clarify that the initial set of propositions, $\{H_{p1}, H_{p2}, H_{d1}, H_{d2}\}$, is an amalgam of propositions relating to the writer of the questioned signature, and propositions relating to writing circumstances. In fact, in the original set of propositions, there is a 'duplication' of the variable relating to writing circumstances because a normal setting appears in both H_{p1} and H_{d1} whereas an unusual setting is associated with H_{p2} and H_{d2}.[4]

The nodes H and C can be associated with the observational variable E in terms of a Bayesian network structure as shown in Fig. 5. With regard to probability assignments, the following can be considered:

[4] In fact, it is generally advisable to avoid 'compound' or otherwise unnecessarily cumbersome propositions. They should be defined as simply as possible within the context of circumstances as it appears to the scientist (I. W. Evett, personal communication).

Table 4
Conditional probabilities for the node C, representing normal (C) and unusual ($-C$) writing conditions given, respectively, propositions H_p and H_d.

	H:	H_p	H_d
C:	C	0.9	0.04
	$-C$	0.1	0.96

Table 5
Conditional probabilities for findings E of comparative handwriting examinations given varying assumptions about the writer of the questioned signature (proposition H) and varying writing circumstances (proposition C).

	H:	H_p		H_d	
	C:	C	$-C$	C	$-C$
E:	E	0.95	0.06	0.01	0.95
	$-E$	0.05	0.94	0.99	0.05

- Node H: The probabilities assigned to the two states H_p and H_d are as chosen in the previous section. Again, their choice is not crucial since likelihood ratio assessment assumes instantiations for the node H.
- Node C: The probability table of this node contains values for normal (C) and unusual ($-C$) writing conditions given that the suspect is the author of the signature H_p as well as given a person other than the suspect, represented by H_d. As may be seen, the respective probability assignments correspond to those made earlier in Section 3.2 for the states H_{p1}, H_{p2}, H_{d1} and H_{d2}. A summary is given here in Table 4.
- Node E: The probability table associated with this node, Table 5, reflects the conditioning upon the two nodes H and C. The various combinations of the states of these two nodes correspond to the initial set of propositions $\{H_{p1}, H_{p2}, H_{d1}, H_{d2}\}$. For this reason, probabilities can be assigned as defined earlier in Table 1.

For the Bayesian network considered in this section, the likelihood ratio slightly diverges from that in the previous section, but only on a notational level. In order to obtain, for example, the numerator of the likelihood ratio, the variable E must be conditioned on H_p while allowing for uncertainty about C. This leads to the following:

$$Pr(E|H_p) = Pr(E|C, H_p)Pr(C|H_p) + Pr(E|\bar{C}, H_p)Pr(\bar{C}|H_p). \tag{6}$$

This expression is equivalent to that defined earlier in Eq. (5) essentially because the numerical assignments defined in Table 4 (used in Eq. (6)) comply with those defined in Table 3 (used in Eq. (5)). Stated otherwise, there is an equivalence of the conditional probabilities for C given H in Eq. (6) and the conditional probabilities for H given H' in Eq. (5). A development analogous to that of Eq. (6) applies for the denominator of the likelihood ratio.

The equivalence between the quantitative output of the Bayesian network considered in this Section and that of the networks discussed in Sections 3.1 and 3.2 is graphically presented in Fig. 6 in terms of a propagation to the ultimate proposition H, based on the observation at the node E.

4. Discussion and conclusions

4.1. Bayesian networks in forensic interpretation

Bayesian networks are a widely recognised method for probabilistic reasoning under uncertainty [18]. Initially developed in the area of artificial intelligence, they are now commonly used in virtually any branch of application in which it is important to handle sources of uncertainty coherently (e.g., risk analysis, strategy formulation, engineering, medical diagnosis, etc.) (e.g., [19]). In forensic science, too,

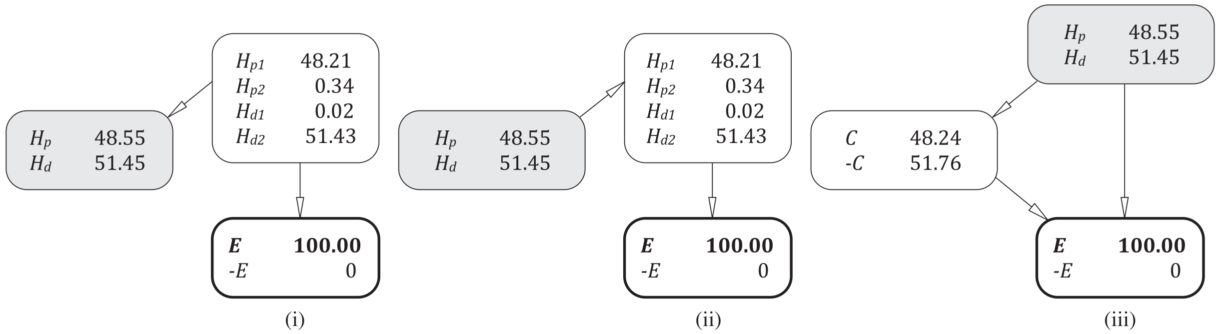

Fig. 6. Expanded representation of the Bayesian networks described in Section 3.1 (here: figure (i)), Section 3.2 (here: figure (ii)) and Section 3.3 (here: figure (iii)). The observational node E is instantiated (indicated by a thicker node border). The equivalence of the output at the node covering the main pair propositions H_p and H_d is shown in grey shading.

researchers and practitioners show a strong interest[5] in this modelling technique and this is illustrated through considerable research in published literature [7], in particular in the area of DNA evidence [21,22].

Despite substantial literature on applications of Bayesian networks for inference in forensic science, there is rather sparse literature on how interested scientists – unfamiliar with the method – can get acquainted with Bayesian networks. This represents a real difficulty essentially because published literature often presents networks that are already constructed. It may thus be difficult for general readers to understand how a given network structure was found and why it should be taken as acceptable. Moreover, because much of the computations are directly operated by algorithms of current Bayesian network software, it may be difficult for scientists new to the topic to keep track of the probabilistic calculations, yet to gain an awareness of how they are actually performed. Such aspects, network construction and probabilistic calculations, are – however – important requirements for an informed and meaningful use of Bayesian networks. Teaching Bayesian networks thus represents a topic of ongoing interest.

4.2. Problem based learning about Bayesian networks in forensic interpretation

In the experience of the authors here it was found useful and insightful to approach the challenge of learning about Bayesian networks in forensic interpretation by interacting with learners in terms of a practical tutorial, based on an example and accompanied by discussion. From a methodological point of view, learners are starting with a generic Bayesian network for representing the constituting elements of Bayes' theorem, as outlined in Section 2.2.2. The analysis is then extended in a stepwise manner to additional complications, notably to questions relating to multiple propositions. This topic, too, is one of continuing interest because practical cases regularly require the consideration of multiple potential sources. In the context of DNA evidence, for example, one may need to account for close relatives of a suspect, not only the suspect and an unrelated person [23]. The coherent handling of this issue within a probabilistic perspective is important in order to set the value of scientific findings appropriately into context.

While students may readily recognise the main proposition in an inference problem, and distinguish it from the available observations, it may be less obvious for them to derive a Bayesian network structure that appropriately reflects an argument that seeks to approach a given propositional level. In particular, reflective questions of interest may be:

- How can one 'know' whether a given network structure is appropriate?
- Can there be different Bayesian networks for the same inference problem? If so, what are the similarities and differences?
- How does particular numerical output from a Bayesian network come to be?

As pointed out through Sections 3.1 to 3.2, existing likelihood ratio formulae can serve as reference procedures and particular Bayesian network structures can be compared against these references. That comparison may involve the definition of the relevant variables, their relationships as well as numerical assignments. For the purpose of the discussion pursued here, the handwriting example taken from published literature [12] was found instructive because it contains both, formulations of posterior probabilities as well as numerical assignments. The study of this example is also rewarding because it allows one to point out that, with Bayesian networks, more can be done than just calculating posterior probabilities for main propositions. In particular, it is shown here that they can also support the derivation of a likelihood ratio (an aspect not formally addressed in the original publication [12]).

An interesting aspect of the proposed teaching scenario (Section 2), studied through Bayesian networks, is that it led learners to find an additional network structure that differs from existing network approaches for dealing with multiple propositions [16,17]. As outlined in Section 2.2.2, that structure is based on the observation that the example's original set of propositions is actually an amalgam of two propositions relating to, respectively, authorship of the questioned signature *and* writing conditions. Learners considered it more appropriate to separate these two aspects in terms of distinct node representations. The resulting Bayesian network thus contains two binary variables, one for the writing circumstances (specified as 'normal' and 'unusual') and one for the main competing propositions. The latter state the suspect (some other person) as the source of the questioned handwriting. A noteworthy insight from this analysis is that there are only two principal source level propositions, which means that the proposed handwriting scenario is not a genuine 'multiple proposition' case.[6]

The fact that course participants came up with this latter approach, along with a demonstration of the coherence of the numerical output, clearly demonstrates the utility of focusing – in a first

[5] This tendency is also reflected by The Royal Statistical Society's Statistics and the Law Working Group current works on a practitioner's guide on Bayesian networks and inferential reasoning. The group's first report, Guide No 1, on the fundamentals of probability and statistical evidence in criminal proceedings [20] was published in November 2010.

[6] A 'genuine' scenario with multiple propositions is understood here as one in which several explicitly mentioned potential sources appear, such as in a DNA evidence setting, where one may have the suspect, a sibling of the suspect as well as an unrelated individual from a relevant population of potential sources.

approach – on studying local network structures in some detail. The ability to provide an original solution for the proposed practical inference problem shows that learns did acquire more than knowledge and comprehension. Their learning outcomes represent intellectual skills in application, analysis, synthesis and evaluation. This provides argument in support of the view that the introduction of Bayesian networks in teaching probabilistic evidence evaluation is both feasible and beneficial.

References

[1] C.G.G. Aitken, D.A. Stoney, The Use of Statistics in Forensic Science, Ellis Horwood, New York, 1991.
[2] C.G.G. Aitken, Statistics and the Evaluation of Evidence for Forensic Scientists, John Wiley & Sons, Chichester, 1995.
[3] B. Robertson, G.A. Vignaux, Interpreting Evidence, Evaluating Forensic Science in the Courtroom, John Wiley & Sons, Chichester, 1995.
[4] C.G.G. Aitken, F. Taroni, Statistics and the Evaluation of Evidence for Forensic Scientists, 2nd edition John Wiley & Sons, Chichester, 2004.
[5] J.S. Buckleton, C.M. Triggs, S.J. Walsh, Forensic DNA Evidence Interpretation, CRC Press, Boca Raton, FL, 2005.
[6] D.J. Balding, Weight-of-Evidence for Forensic DNA Profiles, John Wiley & Sons, Chichester, 2005.
[7] F. Taroni, C.G.G. Aitken, G. Garbolino, A. Biedermann, Bayesian Networks and Probabilistic Inference in Forensic Science, John Wiley & Sons, Chichester, 2006.
[8] W. Edwards, Influence diagrams, Bayesian imperialism, and the *Collins* case: An appeal to reason, Cardozo Law Review 13 (1991) 1025–1074.
[9] B. Robertson, G.A. Vignaux, Taking fact analysis seriously, Michigan Law Review 91 (1993) 1442–1464.
[10] V. Hammond, E.L. Walker, Walker, Teaching Bayesian Networks through Interactive Exploration. FLAIRS-96, Florida Articial Intelligence Research Symposium: proceedings of the Ninth Florida Articial Intelligence Research Symposium, Key West, Florida, May 20–22, 1996, Florida AI Research Society (1996).
[11] 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.
[12] N. Köller, K. Niessen, M. Riess, E. Sadorf, Probability Conclusions in Expert Opinions on Handwriting, Substantiation and Standardization of Probability Statements in Expert Opinions, Luchterhand, München, 2004.
[13] A. Stockton, S. Day, Bayes, Handwriting and Science, Proceedings of the 59th Annual ASQDE Meeting – Handwriting & Technology: at the Crossroads, Des Moines, Iowa, U.S.A, 2001, pp. 1–10.
[14] P. Garbolino, Explaining relevance, Cardozo Law Review 22 (2001) 1503–1521.
[15] 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.
[16] F. Taroni, A. Biedermann, Inadequacies of posterior probabilities for the assessment of scientific evidence, Law, Probability and Risk 4 (2005) 89–114.
[17] J.S. Buckleton, C.M. Triggs, C. Champod, An extended likelihood ratio framework for interpreting evidence, Science & Justice 46 (2006) 69–78.
[18] O. Pourret, P. Naïm, B. Marcot, Bayesian networks: a practical guide to applications, John Wiley & Sons, Chichester, 2008.
[19] M. Goméz, 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.
[20] C.G.G. Aitken, P. Roberts, G. Jackson, Fundamentals of Probability and Statistical Evidence in Criminal Proceedings, Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses, Technical report, Royal Statistical SocietyÕs Working Group on Statistics and the Law, 2010.
[21] 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.
[22] A.P. Dawid, J. Mortera, P. Vicard, Object-oriented Bayesian networks for complex forensic DNA profiling problems, Forensic Science International 169 (2007) 195–205.
[23] J.S. Buckleton, C.M. Triggs, Relatedness and DNA: are we taking it seriously enough? Forensic Science International 152 (2005) 115–119.