

# An explication of uncertain evidence in Bayesian networks: likelihood evidence and probabilistic evidence

## Uncertain evidence in Bayesian networks

Ali Ben Mrad<sup>1,2</sup> · Véronique Delcroix<sup>1</sup> · Sylvain Piechowiak<sup>1</sup> · Philip Leicester<sup>3</sup> · Mohamed Abid<sup>2</sup>

© Springer Science+Business Media New York 2015

**Abstract** This paper proposes a systematized presentation and a terminology for observations in a Bayesian network. It focuses on the three main concepts of *uncertain evidence*, namely *likelihood evidence* and *fixed and not-fixed probabilistic evidence*, using a review of previous literature. A probabilistic finding on a variable is specified by a local probability distribution and replaces any former belief in that variable. It is said to be fixed or not fixed regarding whether it has to be kept unchanged or not after the arrival of observation on other variables. Fixed probabilistic evidence

is defined by Valtorta et al. (*J Approx Reason* 29(1):71–106 2002) under the name *soft evidence*, whereas the concept of not-fixed probabilistic evidence has been discussed by Chan and Darwiche (*Artif Intell* 163(1):67–90 2005). Both concepts have to be clearly distinguished from *likelihood evidence* defined by Pearl (1988), also called *virtual evidence*, for which evidence is specified as a likelihood ratio, that often represents the unreliability of the evidence. Since these three concepts of uncertain evidence are not widely understood, and the terms used to describe these concepts are not well established, most Bayesian networks engines do not offer well defined propagation functions to handle them. Firstly, we present a review of uncertain evidence and the proposed terminology, definitions and concepts related to the use of uncertain evidence in Bayesian networks. Then we describe updating algorithms for the propagation of uncertain evidence. Finally, we propose several results where the use of fixed or not-fixed probabilistic evidence is required.

- ✉ Véronique Delcroix  
veronique.delcroix@univ-valenciennes.fr
- Ali Ben Mrad  
benmradaли2@gmail.com
- Sylvain Piechowiak  
Sylvain.Piechowiak@univ-valenciennes.fr
- Philip Leicester  
p.a.leicester@lboro.ac.uk
- Mohamed Abid  
mohamed.abid\_ces@yahoo.fr

**Keywords** Bayesian network · Uncertain evidence · Probabilistic evidence · Likelihood finding · Soft evidence · Virtual evidence

<sup>1</sup> LAMIH, UMR CNRS 8201 (Laboratory of Industrial and Human Automation control, Mechanical engineering and Computer Science), University of Valenciennes and Hainaut-Cambrésis, Le mont Houy, F59313 Valenciennes, CEDEX 9, France

<sup>2</sup> Computer & Embedded Systems (CES Lab.), National Engineering School of Sfax (ENIS), University of Sfax, 3038 Sfax, Tunisia

<sup>3</sup> Centre for Renewable Energy Systems Technology (CREST), Loughborough University Science and Enterprise Parks, LE11 3TU Leicestershire, UK

## 1 Introduction

Bayesian networks are probabilistic graphical models that provide a powerful way to embed knowledge and to update one's beliefs about target variables given new information about other variables. They are widely used for problems with inherent uncertainty such as classification, diagnosis and decision-making [65]. In a Bayesian network, prior knowledge is represented by a probability distribution  $P$  on the set of variables which define the problem, whereas

updated beliefs are represented by the posterior probability distribution  $P(\cdot | obs)$  where  $obs$  represents new information. Inference in Bayesian networks provides a means to explain given evidence e.g., Maximum A Posteriori Assignment (MAP), Most Probable Explanation (MPE) [61], and Most Relevant Explanation (MRE) [73]. Evidence is the starting point of these methods and refers to new information in a Bayesian network. A piece of evidence is also called a finding or an observation, and evidence refers to a set of findings. Figure 1 illustrates the propagation of evidence in belief updating.

A finding on a variable commonly refers to an instantiation of the variable. This can be represented by a vector with one element equal to 1, corresponding to the state the variable is in, and all other elements equal to zero. This type of evidence is usually referred to as hard evidence though other terms are sometimes used.

This paper focuses on another type of evidence that cannot be represented by such vectors: uncertain evidence. The objective of the paper is to clarify the term uncertain evidence and its underlying concepts. It is argued that three types of uncertain evidence need to be clearly distinguished, namely likelihood evidence, fixed probabilistic evidence and not-fixed probabilistic evidence.

Likelihood evidence is applicable when there is uncertainty about the veracity of an observation, such as, for example, the information given by an imperfect sensor. Ideally the Bayesian network should include two variables for each of two physical quantities: one for the unobserved real value, and one for the value observed by the sensor. Using likelihood evidence avoids the need to add both these variables in the model.

In contrast probabilistic evidence can be regarded as a new probability distribution on a variable arising from a new observation after creation of the model. This is dissimilar to likelihood evidence where the original probability distribution is not challenged, only ones belief in it and which may be amended with new likelihood evidence.

Not-fixed probabilistic evidence is typical of the situation where a variable is given new probability distribution as a result of the BN models application to a specific sub-population of the global population for which the model was built. Thus the conditional probabilities remain the same

and the variable with the new observed distribution can be updated in response to evidence at other nodes.

Fixed probabilistic evidence is conceptually similar to not-fixed but the new probability distribution is regarded as immutable, even after later evidence is applied to other nodes. A typical example is where probabilistic evidence is imparted to a subscriber from a publisher in a one-way communication in an agent encapsulated Bayesian network.

Below these concepts are further explicated and defined. The propagation of the various forms of evidence is discussed and examples of applications are given to further illustrate the differences between these types of evidence.

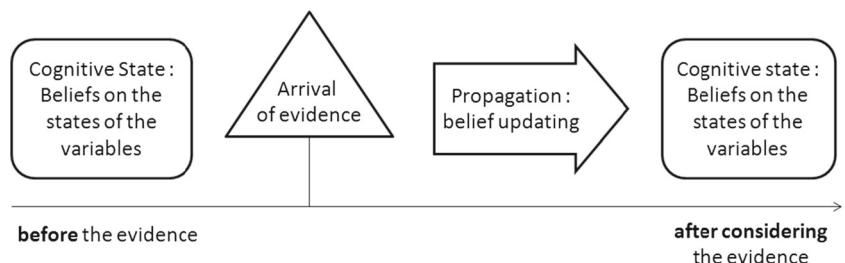
The rest of the paper is organized as follows. Section 2 presents definitions of a Bayesian network and hard evidence in a Bayesian network. Section 3 is a review of terminology and concepts about evidence in Bayesian networks, in the literature and in Bayesian network engines, followed by the proposed terminology for evidence in a Bayesian network. Section 4 proposes definitions, properties and examples of the three types of uncertain evidence. Section 5 describes updating algorithms for each type of uncertain evidence. Section 6 proposes and discusses three types of situation where the use of fixed probabilistic evidence is required. The first one is about the integration of Bayesian networks with Geographic Information Systems (GIS), the second one concerns the propagation of a continuous variable in a discrete Bayesian network, and the third one is about using fixed probabilistic evidence for a distributed Bayesian network. Section 7 offers our conclusions and presents our future research proposals.

## 2 Basics of a Bayesian network

This section concerns the definition of a Bayesian network and hard evidence.

### 2.1 Bayesian network definition

Bayesian networks [20, 35, 42, 61] are a class of probabilistic graphical models. After giving a definition of Bayesian networks, we provide a brief overview about learning and inference.



**Fig. 1** Propagation of evidence in belief updating

**Definition 1** (Bayesian network) A *Bayesian network* is a couple  $(G, P)$ , where  $G = (\mathbf{X}, \mathbf{E})$  is a directed acyclic graph with nodes  $\mathbf{X} = \{X_1, \dots, X_n\}$  and directed edges  $\mathbf{E}$  which represent conditional dependencies between nodes. The joint probability distribution for  $\mathbf{X} = \{X_1, X_2, \dots, X_n\}$  is given by the chain rule:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | pa(X_i))$$

where  $pa(X_i)$  represents the parents of  $X_i$  as defined by the presence of directed edge from a parent node to  $X_i$ .

The graph  $G$  represents the qualitative component of the Bayesian network; a quantitative component is given by  $P$ , which is defined by a set of conditional probability distributions associated with each node in  $\mathbf{X}$ . In this paper we consider only discrete random variables, and thus each node is associated with a conditional probability table (CPT). Both  $G$  and  $P$  can be obtained from human experts or learnt from available data.

Once the Bayesian network is defined, algorithms are used to propagate new evidence through it. Unfortunately, both exact and approximate methods of inference have been proved to be NP-hard [16, 17], rendering the propagation of evidence intractable in some cases. This intractability depends on the size of the model (number of nodes, degree of nodes, size of the sets of possible values for each node, and other graphical parameters, such as the treewidth). However, some approximate inference methods [72] have been shown to work very well in practice for large scale networks.

In the following, capital letters are used to represent random variables, and lower-case letters represent their values. Bold capital letters correspond to sets of variables. Here are some more notations used in the rest of the paper:

$X$  denotes a Bayesian network node  $X \in \mathbf{X}$  having its states (or values) in  $\mathcal{D}_X = \{x_1, \dots, x_m\}$ ,

$P(\mathbf{X})$  denotes  $P(X_1, \dots, X_n)$ ; it is the joint probability distribution that defines a Bayesian network on the set  $\mathbf{X}$ .

$P(x)$  denotes  $P(X = x)$ ,

$P(X)$  is the probability distribution ( $P(X = x_1), \dots, P(X = x_m)$ ).

$R(X_i)$  and  $R(X_j, X_k)$  are local probability distributions used to describe uncertain findings on  $X_i$  and  $(X_j, X_k)$ .

## 2.2 Hard evidence in a Bayesian network

Bayesian networks are commonly used to propagate observations represented by hard findings.

**Definition 2** (Hard finding) A hard finding  $e$  on a variable  $X$  in a Bayesian network with values in  $\mathcal{D}_X$  is defined by an observation vector of size  $m = |\mathcal{D}_X|$  containing a single 1, at the position corresponding to a state  $x \in \mathcal{D}_X$  and 0 for all other positions. This finding represents the instantiation of  $X$  to the value  $x$  and it is characterized by  $P(X = x | e) = 1$ . Hard evidence is a set of hard findings.

Let  $Q$  denote the probability distribution reflecting the belief state after taking into account hard evidence  $e$ . We have  $Q = P(\cdot | e)$ .

## 3 Review of uncertain evidence in the literature and Bayesian networks software

This section presents a seven point summary that establishes a picture of uncertain evidence in Bayesian networks. This leads us to the tabulation of both concepts and terminology presented at the end of this section. Detailed definitions of each concept are presented in the following section.

The seven points are:

1. A quasi-consensus on the definition of a finding in a Bayesian network;
2. Uncertain evidence: two main concepts;
3. Likelihood finding: a first concept of uncertain evidence well identified;
4. Soft evidence: a source of confusion due to terminology;
5. Uncertain evidence specified by a local probability distribution: no terminology, no consensus, and not supported by Bayesian network engines;
6. Confusion leads to debates about commutation of iterated belief revisions;
7. Uncertain evidence propagation in Bayesian network software products.

### 3.1 A quasi-consensus on the definition of a finding in a Bayesian network

The terms evidence and findings are synonyms, but only the term *finding* is used in the singular form. When no adjective qualifies the word evidence (or finding or observation), it generally represents hard evidence, that is to say the instantiation of a set of variables of the Bayesian network. In the literature, a hard finding on a variable is also called an observation, or a deterministic, specific, regular or positive finding. Despite the variety of terminology in the literature, the definition of hard evidence is clear and presents no particular problems (see Definition 2).

However, a generalization of the definition of finding is proposed by Jensen and Nielsen [35] in which a finding on a variable  $X$  is an  $m$ -dimensional table of zeros and ones, allowing several ones. We do not follow this proposition. The second point of this review identifies two main concepts of uncertain evidence in the context of belief updating in the probabilistic framework.

### 3.2 Two main concepts of uncertain evidence

Many real world problems require reasoning with uncertain inputs. Belief updating founded upon uncertain inputs may differ from belief updating founded on certain inputs. We cite the two possible meanings of the term “uncertain input” in the probabilistic framework, following Dubois, Moral and Prade [25]:

- “the uncertainty bears on the meaning of the input; the existence of the input itself is uncertain, due to, for instance, the unreliability of the source that supplies inputs.”
- “the input is a partial description of a probability measure; the uncertainty is part of the input and is taken as a constraint on the final cognitive state. The input is then a correction to the prior cognitive state.”

In the following, we use the term uncertain evidence as a generic term to refer to any of these two meanings. In the context of Bayesian networks, only one of these meanings is widely used. This is the next point of this review about uncertain evidence. Complete definitions, properties and examples pertaining to these two main meanings are given in Section 4.

### 3.3 The first concept of uncertain evidence: likelihood finding

The concept of *likelihood evidence* (or *virtual evidence*) [61] models the case where the observation *is* uncertain due to, for example, an unreliable source of information.

This type of uncertain evidence is well documented in the literature and its propagation is possible in some Bayesian network software, even if it is sometimes wrongly named. Table 1 provides the list of terms used to refer to likelihood evidence in several Bayesian network software applications. Four of them employ the term soft evidence instead of likelihood evidence or virtual evidence. In the next section we argue this term should not be used.

### 3.4 Soft evidence : a terminology with no consensus

The term *soft evidence* has been introduced in the context of Agent Encapsulated Bayesian networks (AEBN) [11, 69]. It refers to evidence specified by local probability distributions that define constraints on the posterior probability distribution and that cannot be changed by further information meaning that these probability distributions are fixed. Each observed local probability distribution on a subset of variables is different than the encoded prior probability distributions for these variables associated with the Bayesian network.

The term *soft evidence* is used with that meaning by Val-torta [40, 48, 49, 69] and other authors [43, 60, 63, 64, 68]. With that meaning, likelihood evidence can be interpreted as evidence **with** uncertainty, while *soft evidence* can be interpreted as evidence **of** uncertainty [64].

However, a review of the literature over the past ten years shows that the term *soft evidence* is sometimes used

**Table 1** Terminology used in several Bayesian networks software to name likelihood evidence in April 2014

Bayesian network software	Terminology used in Bayesian networks software to name likelihood findings
Netica [58]	likelihood finding
Hugin [51]	likelihood evidence
Genie [24]	virtual evidence
Infer.NET [56]	virtual evidence
gRain [33]	virtual / likelihood evidence
BayesiaLab [38]	likelihoods (soft evidence in the user’s manual)
BNT [57]	soft evidence
Bayes Server [66]	soft evidence
AgenaRisk [10]	soft evidence

**Table 2** Available features in Bayesian network engines to propagate uncertain evidence specified by a local probability distribution (April 2014)

Bayesian network software	Terminology used	Comments
BayesiaLab [38]	probability distribution (fixed)	Allows fixing the probabilities of the selected nodes to the current marginal probabilities.
	probability distribution (not fixed)	Allows entering a likelihood finding from a probability distribution.
Netica [58]	calibration	Ask for entering $P(X   \text{All observations})$
ProBT [1]	soft evidence	Additional variables named “coherence variables” are used to combine prior beliefs and new probabilistic information.

to refer to likelihood evidence [9, 13, 14, 18, 41, 44]. None of these articles refers to Valtorta’s use of the term; it is clear therefore, that their authors have developed a different contemporary use which has led to confusion. The same confusion occurs in Bayesian network software, where at least four products use soft evidence to refer to likelihood evidence (see Table 1). In another paper [12], the term was used to refer to a variable  $X$  which does not take a specific value  $x$ , that is to say  $X \neq x$ , which is usually called a negative finding. Moreover, the concept of uncertain evidence specified by a local probability distribution and that cannot be modified by other information is not always named soft evidence (Table 2).

Thus the term *soft evidence* is confusing since it is used ambiguously both in the literature and by Bayesian network software interfaces.

### 3.5 Uncertain evidence specified by a local probability distribution: no terminology, no consensus, rarely available in Bayesian network engines

Several papers concern both the specification of uncertain evidence and its propagation [15, 20, 60, 64, 71]. The term *uncertain evidence* is often employed in a generic way to name different types of non-hard evidence. They clearly distinguish between two types of uncertain evidence in Bayesian networks. The first concept is likelihood evidence, as proposed by Pearl in [61], and is clearly identified by the authors who mentioned it. The second concept of uncertain evidence is specified by a local probability distribution but is not so clearly defined. Thus two gaps need to be filled: (1) an adequate terminology should be defined to name uncertain evidence specified with a local probability distribution; (2) this second type of uncertain evidence requires a more precise definition to allow a clear identification of, and to make the distinction between two different sub-types. None of the above-cited papers focuses on terminology, nor identifies clearly the two

sub-types of uncertain evidence specified by a local probability distribution.<sup>1</sup>

Chan and Darwiche present and compare two methods of propagation of uncertain evidence [15]. They provide an interesting discussion about the specification of evidence in both cases. However, no terminology or definition is proposed in order to consider the local probability distribution as a particular type of uncertain evidence in a Bayesian network. Although the analysis of Chan and Darwiche [15] is referred to by Peng and Zhang [64], it does not lead to a clear distinction of both sub-types of uncertain evidence specified by a local probability distribution.

In Bayesian network software, very few products propose the propagation of uncertain evidence specified with a local probability distribution. Table 2 shows the available features of three Bayesian network software products.

### 3.6 Confusion leads to debates about commutation of belief revision

The confusion about uncertain evidence specified by a local probability distribution gave rise to debates between authors, particularly about the question of commutation. The question “Should, and do, iterated belief revisions commute?” [15] concerns the case of revision with several pieces of evidence, of which some are uncertain. Some authors claim that several pieces of evidence specified by a local probability distribution and carrying the “All things considered” interpretation must not be commutative [15]. Others argue that *soft evidence* is a true observation of the distributions of some events, and as such, they should all be preserved in the updated “posterior” distribution [64]. In the first case, the arriving information is susceptible to improvement by further evidence, whereas in the second case, the arriving information has to behave as hard evidence and can not be influenced by any other information.

<sup>1</sup>One article briefly mentions the three types of methods of propagation of uncertain evidence in a Bayesian network [7].

**Table 3** Proposed terminology about uncertain evidence in Bayesian network

Uncertain evidence		
Hard evidence [61]	Likelihood evidence [61]	Probabilistic evidence
		not-fixed [15]      fixed [64, 69]
No uncertainty	The observation is uncertain	The constraint specified by the probability distribution can be modified by other information ? YES      NO

The conclusion is that a third concept of uncertain evidence has to be more clearly defined in the context of Bayesian networks. More exactly, uncertain evidence can be divided into two main concepts, namely likelihood evidence and uncertain evidence specified by a partial measure of probability. This second concept has to be divided into two sub-concepts, according to whether the local probability distribution which specifies a constraint on the posterior belief state can be (or cannot be) modified by later arrival of new information.

### 3.7 Proposed terminology for uncertain evidence in a Bayesian network

In order to address the lack of an unambiguous terminology in Bayesian network theory and practice, we propose the use of the term *probabilistic evidence* for uncertain evidence specified by a local probability distribution. In contrast to hard evidence, the inconsistent use of terms for some type of uncertain evidence is problematic, in particular the term *soft evidence*, the misuse of which may cause real confusion. In order to make a clear distinction between the two sub-types of probabilistic evidence, we propose the terms *fixed probabilistic evidence* to refer to *soft evidence* such as defined by Valtorta [64, 69], and *not-fixed probabilistic evidence* to refer to the concept used in Jeffrey's rule and discussed in [15, 64]. The terms "likelihood" and "probabilistic" capture the ways the evidence are specified. The adjectives "fixed" and "not fixed" capture the expected behavior of the posterior probability distribution after further evidence is obtained. Table 3 presents the proposed terminology and the main associated characteristics.

Up to now, the two concepts of uncertain evidence specified by a probability distribution, namely fixed and not fixed probabilistic evidence, are poorly identified in the Bayesian network community. They are mostly absent from Bayesian network software products.

### 3.8 Uncertain evidence in Bayesian network software

Most available implementations of uncertain evidence propagation in Bayesian network engines concern Pearl's method of virtual evidence. Table 4 shows the features available for the updating of uncertain evidence among some available Bayesian network engines.<sup>2</sup>

This review of literature leads to the conclusion that two main types of uncertain evidence have been defined, but neither terminology nor concepts have been clearly defined by the Bayesian network user community. The next section presents the definitions and characteristics of likelihood evidence and probabilistic evidence.

## 4 Uncertain evidence in a Bayesian network

This section presents definitions and properties of the two main types of uncertain evidence, namely likelihood evidence and probabilistic evidence. Their definitions and properties are illustrated with some examples, and the main elements of propagation algorithms are given in the following section.

### 4.1 Likelihood evidence: definition and characteristics

Likelihood evidence corresponds to the cases where the observation *is* uncertain (Fig. 2). The uncertainty on the observation may come from the unreliability or imprecision of the source of the information.

<sup>2</sup>The software ProBT of Probayes proposes the use of a new variable named "coherence variable" to take into account new information specified by a local probability distribution [8]. Further studies remain to be carried out to compare this proposition with the propagation of likelihood evidence and probabilistic evidence.

**Table 4** Features of Bayesian network software about evidence updating

	Hard finding	Likelihood finding	Not fixed probabilistic finding	Fixed probabilistic finding
Elvira [26], Analytica [32], Samlam [21]	X			
AgenaRisk [10], Bayes Server [66], Infer.NET [56], BNT [57], Genie [24], Hugin [51], gRain [33]	X	X		
Netica [58]	X	X	X	
BayesiaLab [38]	X	X	X	X

**Definition 3** (Likelihood finding or virtual finding) A likelihood finding (or virtual finding) on a variable  $X$  of a Bayesian network is an observation with uncertainty of the variable. It is specified by a likelihood ratio<sup>3</sup>

$$L(X) = (L(X = x_1) : \dots : L(X = x_m)) = (P(obs | x_1) : \dots : P(obs | x_m))$$

where the  $L(X = x_i)$  are quantities relative to each other representing the probability of the observed event given  $X$  is in the state  $x_i$ . Likelihood evidence, also called *virtual evidence*, is a set of likelihood findings.

A particular case of likelihood finding occurs when the likelihood ratio is composed of only zeros and ones, in order to represent information meaning that only some values of the observed variables are possible (Definition 4). The zeros denote a *negative finding*, meaning that the corresponding states of  $X$  are impossible, whereas the ones denote a *disjunctive finding*, meaning that the variable is necessarily in one of the states corresponding to a one, but without specifying that some values are more probable than others.

**Definition 4** (Negative finding, disjunctive finding) A negative finding (or disjunctive finding) on a variable  $X$  with values in  $\mathcal{D}_X$  is defined by an observation vector of zeros and ones. It represents the information that  $X$  can be only in one of the states corresponding to the ones and that the other states are impossible.

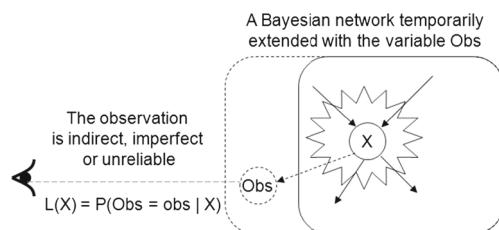
A negative finding whose observation vector contains a single one is a hard finding.

The next two properties describe how likelihood evidence interacts with beliefs before and after its propagation.

*Property 1* Likelihood evidence is specified “without a prior”, as a consequence, propagating likelihood evidence takes into account the beliefs in the variable before the evidence.

*Property 2* Belief in a variable after propagating a likelihood finding on it is not fixed: it can be modified by further evidence on other variables. In other words, let  $Q_1$  represent the beliefs after the propagation of a likelihood finding on  $X$  and  $Q_2$  represent the beliefs after a second piece of evidence on another variable; then it may occur that  $Q_1(X) \neq Q_2(X)$ , meaning that the belief in  $X$  has been modified by the second piece of evidence.

*Example 1* (Likelihood finding: optical character recognition (OCR)) A Bayesian network includes a variable  $X$  representing a letter of the alphabet that the writer wanted to draw. The set of values of  $X$  is the set of letters of the alphabet. A piece of uncertain information on  $X$  is received from a system of OCR. The input of this system is an image of a character and the output is a vector of similarity between the image of the character and each letter of the alphabet. Let  $o$  represent the observed image. Consider a case where,



**Fig. 2** A likelihood finding on  $X$ . The variable  $X$  is observed with uncertainty (e.g. via an imperfect sensor). The binary variable  $Obs$  represents the observation, whereas  $X$  represents the real value. The variable  $Obs$  is temporarily added to the graph in order to propagate the observation  $Obs = obs$ . The evidence is specified by the likelihood of the observed value  $obs$  with respect to each value of  $X$

<sup>3</sup>the terms of a likelihood ratio do not need to sum to one.

due to lack of clarity,  $o$  can be recognized as either the letter 'v', 'u' or 'n'. The OCR technology provides the indices such that  $P(Obs = o | X = 'v') = 0.8$ ,  $P(Obs = o | X = 'u') = 0.4$ ,  $P(Obs = o | X = 'n') = 0.1$  and  $P(Obs = o | X = x) = 0$  for any letter  $x$  other than 'u', 'v' or 'n'. This means that there is twice as much chance of observing  $o$  if the writer had wanted to draw the letter 'v' than if she had wanted to draw the letter 'u'. Such a finding on  $X$  is a likelihood finding on  $X$ , specified by  $L(X) = (0 : \dots : 0 : 0.1 : 0 : \dots 0 : 0.4 : 0.8 : 0 : 0 : 0 : 0 : 0)$ . Note that from the definition, the entries of  $L(X)$  need not add up to 1.

This example illustrates the two characteristics of likelihood finding: in the Bayesian network, the prior probability distribution  $P(X | pa(X))$  includes knowledge about the distribution of letters in the language of the text from which the character comes whereas the OCR technology does not integrate that knowledge. Thus it provides information about  $X$  without prior knowledge. In order to update the belief in the value of the character, the information provided by the OCR (the vector of similarity) has to be combined with the prior knowledge about the frequency of letters. Moreover, the result of propagation is not fixed since belief in  $X$  can be further modified by other information. For example, information about the neighboring characters could be taken into account.

**Example 2** (Negative finding: Example 1 continued) Consider now the negative (or disjunctive) finding given by the information that the observed image ( $o$ ) can be only the letter 'u' or 'v', and that all other letters are excluded. We have  $P(o | X = 'u') = P(o | X = 'v')$  and  $P(o | X = x) = 0$  for all  $x \in \mathcal{D}_X \setminus \{'u', 'v'\}$ . Thus,

$$P(X = x | o) = \begin{cases} \frac{P(X=x)}{P(X='u')+P(X='v')} & \text{for } x \in \{'u', 'v'\} \\ 0 & \text{for other letters.} \end{cases}$$

The posterior beliefs on the events  $X = 'u'$  and  $X = 'v'$  depend on the prior beliefs on these events and may differ from them.

## 4.2 Probabilistic evidence: definition and characteristics

Probabilistic evidence corresponds to another meaning of uncertain input where the evidence is specified by local probability distributions. Fixed probabilistic evidence is often called soft evidence [4–7, 48, 60, 64, 69].

The definition below is given in its simplest form. A more general version is given below (Definition 6).

**Definition 5** (Probabilistic finding, fixed or not-fixed) A *probabilistic finding* on a variable  $X \in \mathbf{X}$  is specified by a local probability distribution  $R(X)$  that defines a

constraint on the belief in  $X$  after this information has been propagated; it describes the state of beliefs in the variable  $X$  "all things considered". A probabilistic finding is *fixed* (or not) when the distribution  $R(X)$  can not be (or can be) modified by the propagation of other findings. *Probabilistic evidence* is a set of probabilistic findings.

The difference between fixed and not-fixed probabilistic evidence cannot be seen before the arrival of new information.

The next two properties describe how probabilistic evidence interacts with beliefs before and after its propagation.

**Property 3** A probabilistic finding  $R(X)$  on a variable  $X$  of a Bayesian network replaces any prior belief or knowledge on  $X$ . As a consequence, the prior  $P(X)$  is not used in the propagation of  $R(X)$ , and any previous finding or belief on  $X$  is lost.

Probabilistic evidence includes both the strength of the evidence and the state of beliefs before evidence.

**Property 4** A probabilistic finding  $R(X)$  on a variable  $X$  is preserved when updating belief. The beliefs after considering the probabilistic finding on  $X$  is represented by a probability distribution  $Q$  on  $\mathbf{X}$  such that  $Q(X) = R(X)$ .

**Property 5** A fixed probabilistic finding on  $X$  is not modified by further evidence on any other variables of the model, and a further finding on  $X$  is not possible, unless it overwrites the current evidence. Any kind of evidence received on other variables after fixed probabilistic evidence makes it necessary to re-propagate previous fixed probabilistic evidence together with the new evidence, in order to keep the former probabilistic evidence fixed. As a consequence, the propagation of several fixed probabilistic findings commutes: the result of propagation is independent of the order in which fixed probabilistic findings are received.

Fixed probabilistic evidence behaves as hard evidence in that the specified evidence remains unchanged after its propagation, and still remains unchanged after the arrival of other information on the same case.

**Property 6** A not-fixed probabilistic finding on  $X$  can be modified by further evidence on any variable in the model, including likelihood evidence on  $X$ . As a consequence, the propagation of several not-fixed probabilistic findings does not commute.

In order to illustrate these definitions and properties, we propose below three examples of probabilistic findings: the

first one concerns a not-fixed probabilistic finding in the example of character recognition, the second one presents fixed probabilistic evidence coming from the observation of a sub-population in the ASIA Bayesian network, and the third one shows a case of probabilistic evidence that first has to be fixed and not fixed afterwards.

*Example 3* (not-fixed probabilistic finding: Example 1 continued) Consider for the variable  $X$  in Example 1 that the language of the word from which the character comes, and the frequency of letters in that language are known. If the Bayesian network does not contain the variable “language of the text” ( $L$ ), this information can be applied as not fixed probabilistic evidence for the variable  $X$  representing the character:  $R(X) = (R(X = 'a'), R(X = 'b'), \dots, R(X = 'z'))$ , provided that  $R(X)$  satisfy the condition “all things considered”, meaning that no other prior belief has to be combined with it. This has to replace the prior belief in the event  $X = x$ . Since that information about  $X$  could be improved by further evidence such as a likelihood finding on  $X$  described in Example 1, it is a *not-fixed* probabilistic finding on  $X$ .

Remark: in that example, let us suppose now that the first information is the likelihood finding provided by the OCR, and the second information concerns the language of the text ( $L = \text{English}$ ), but  $L$  is not a variable of the Bayesian network. In that case, the probability distribution  $R(X)$  representing the frequency of each letters in English cannot be used to specify a not-fixed probabilistic finding on  $X$  since it does not consider “all things”, in particular, the information provided by the OCR technology is not taken into account in  $R(X)$ .

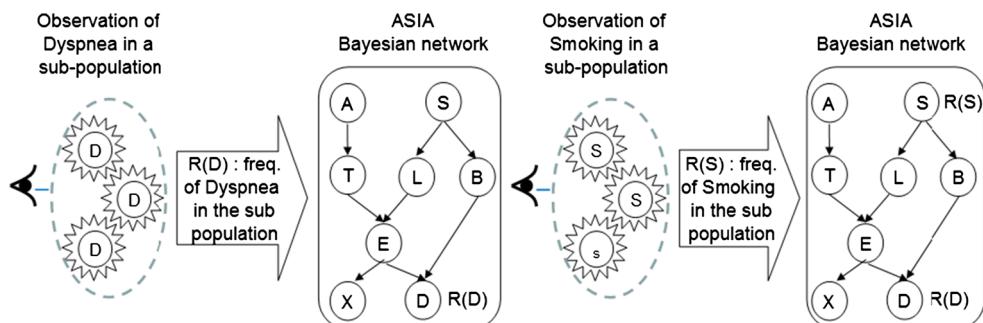
*Example 4* (Fixed probabilistic evidence in Asia Bayesian Network) Consider the Bayesian network Asia [52] which contains eight binary nodes, among which there is a (root) node *Smoking* and a (leaf) node *Dyspnea* (Fig. 3). Instead of having findings about a single person, consider findings coming from the data of a particular sub-population, such as the workers in a given factory *FunT*. Observing that half of them have dyspnea and a tenth of them smoke

constitutes fixed probabilistic findings on these variables such that:  $R(\text{Dyspnea}) = (0.5, 0.5)$  and  $R(\text{Smoking}) = (0.1, 0.9)$ . No other information about the factory *FunT* can modify these probability distributions. The first finding has to be preserved even after propagating the second probabilistic finding. Thus both findings have to behave as hard findings and must not to be modified by propagation. They are *fixed* probabilistic findings. When no more details are available, these findings cannot be considered as a single piece of probabilistic evidence on the two variables  $R(\text{Dyspnea}, \text{Smoking})$  as defined in the extended definition of probabilistic finding (Definition 6 below).

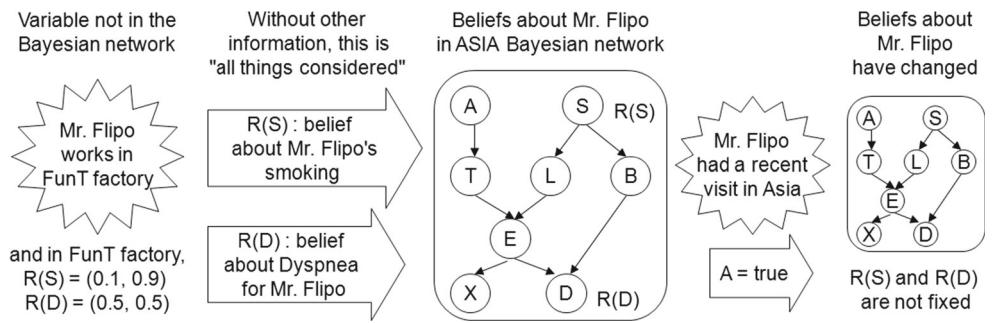
*Example 5* (Probabilistic findings in Asia Bayesian Network: fixed then not-fixed) Consider again the Bayesian network Asia and the case of Mr. Flipo who works in the factory *FunT* about which a recent survey has revealed that half of its workers suffer from dyspnea and only one in ten smoke (see Fig. 4). Without any more information on Mr. Flipo, the probability distributions  $R(D) = (0.5, 0.5)$  and  $R(S) = (0.1, 0.9)$  represent the posterior belief in the variables  $D$  and  $S$  for Mr. Flipo. The prior beliefs about  $D$  and  $S$  represented by  $P(D)$  and  $P(S)$  no longer have influence on our belief about Mr. Flipo thus  $R(D)$  and  $R(S)$  replace the prior beliefs, and they are given “all things considered”. Since the propagation of  $R(D)$  has to be preserved while  $R(S)$  is propagated (or *vice versa* regarding the order of propagation), the first probabilistic finding, say  $R(S) = (0.1, 0.9)$ , initially has to be fixed. If new information about Mr. Flipo arrives, such as a recent visit to Asia, it should modify our belief in both variables  $S$  and  $D$ . Thus,  $R(D) = (0.5, 0.5)$  and  $R(S) = (0.1, 0.9)$  become not-fixed probabilistic findings (the probabilistic finding on  $S$  is no longer kept fixed).

Consider now the case where the first information about Mr. Flipo concerns his recent visit to Asia ( $A = \text{true}$ ), and the second information is that Mr. Flipo works in the factory *FunT*, in which half its workers suffer from dyspnea and only one in ten smoke. The probability distributions  $R(D) = (0.5, 0.5)$  and  $R(S) = (0.1, 0.9)$  do not represent

**Fig. 3** Two fixed probabilistic findings about Dyspnea ( $D$ ) and smoking ( $S$ ) observed in a sub-population. The evidence is specified by the probability distributions  $R(D)$  and  $R(S)$  that represent the belief in  $D$  and  $S$  after the propagation



**Fig. 4** Two probabilistic findings about Mr. Flipo's case. The first is kept fixed while the second is propagated, but after, both probabilistic findings are *not fixed* since beliefs in variable  $S$  and  $D$  can be modified by further evidence about Mr. Flipo, such as his recent visit in Asia



the posterior belief in the variables  $D$  and  $S$  for Mr. Flipo, since they do not include the initial information about the visit to Asia. Since these probability distributions are not given with “all things considered”, they can not be used to specify probabilistic findings.

This example illustrates that the propagation of several probabilistic findings, such that all of them are preserved, requires that they be fixed until the last finding has been propagated. This has to be done even if the probabilistic findings can be later modified by other information. In that case, the initial probabilistic findings are no longer kept fixed.

In this example, a set of not fixed probabilistic findings are deduced from the same initial information. Since each of these findings defines a constraint on the posterior probability distribution, they have to be kept fixed until all of them are propagated. Afterwards, these probabilistic findings can be later modified by other information.

Example 4 illustrates that an observation from a sub-population constitutes a fixed probabilistic observation, whereas Example 5 illustrates that the information on a single instance furnished by knowledge of the sub-population to which it belongs, is not a fixed probabilistic observation since this information can be improved by further evidence. Two other kinds of probabilistic evidence are detailed below in Section 6.

There are two main differences between probabilistic evidence and likelihood evidence. Firstly the specification: for probabilistic evidence the distribution is specified “all things considered” whereas for likelihood evidence the likelihood ratio is without prior knowledge or belief. Secondly the propagation: while probabilistic evidence remains unchanged by updating the observed variables, likelihood evidence has to be combined with previous beliefs in order to update the belief in the observed variable(s).

Fixed probabilistic evidence on a variable  $X$  can be supplied by an expert on  $X$ , and her judgment on  $X$  cannot be improved by other evidence on any other variables of the model. This type of evidence can be obtained by the

precise observation of a variable on a sub-population. The difference between fixed and not-fixed probabilistic evidence is only visible when several pieces of evidence are received and propagated.

Definition 5 has been extended by Valtorta [69] in order to consider information about one or more variables of the model, specified by different forms of probabilistic evidence.

**Definition 6** (extended notion of probabilistic finding) A probabilistic finding on a subset of variables  $\mathbf{Y} \subset \mathbf{X}$  is a partial description of a probability measure that can be one of the following:

- (a) a joint probability distribution  $R(\mathbf{Y})$ ,
- (b) a conditional probability distribution  $R(\mathbf{Y} \mid \mathbf{Z})$  where  $\mathbf{Z} \subset \mathbf{X} \setminus \mathbf{Y}$ ,
- (c) probability assignments on arbitrary events on variables of  $\mathbf{Y}$ ,
- (d) probability assignments on arbitrary logic formulae on variables of  $\mathbf{Y}$ .

The extended notion of probabilistic evidence can be handled for evidence updating by the introduction of an observation node [69]. This technique allows the reformulation of extended probabilistic evidence into probabilistic evidence on a single new observation variable. In the following, we consider only probabilistic evidence involving a single variable.

#### 4.3 D-separation and uncertain evidence in a Bayesian network

The property of d-separation is central in Bayesian networks. It allows the identification of those variables in the network whose posterior probability could be modified by new information, regarding both previous observations and their relative position in the graph.

The property of d-separation between two variables requires the examination of all the paths between them to check whether they are *blocked* or not.

**Table 5** Classification of the status of the belief of a variable  $X$  regarding the kind of observation on that variable. This status holds as long as the observed case is considered

<b>Belief on <math>X</math> is fixed</b>	<b>Belief on <math>X</math> may change</b>
The belief on $X$ cannot be modified by other information (neither on itself, nor on other variable)	The belief on $X$ can be modified by any other information
The variable $X$ received - a hard finding - or a fixed probabilistic finding	The variable $X$ received - no evidence - or a likelihood finding - and / or a not fixed probabilistic finding

Usually, a path between two nodes  $X$  and  $Y$  is said to be blocked if there exists an intermediate node  $Z$  on the path such that one of the following conditions are true:

- there is a serial or a diverging connection on  $Z$  and  $Z$  is observed;
- there is a converging connection on  $Z$  and  $Z$  is not observed and none of its descendants is observed.

However, it is necessary to be more precise to explain what is meant by  $Z$  is observed (or not) regarding the different kinds of observations. Table 5 is a first step in that direction. It shows how each kind of evidence can be classified in two classes, regarding whether the belief on the observed variable is fixed or not, as long as the observed case is considered.

The classification given in Table 5 allows a more general characterization of a blocked path between two variables to be given (see Table 6 and Definition 7).

**Definition 7** (blocked path) A path between two nodes  $X$  and  $Y$  is *blocked* if there is an intermediate node  $Z$  on the path such that one of the following condition is true:

- there is a serial or a diverging connection on  $Z$  and  $Z$  received a hard finding or a fixed probabilistic finding;

- there is a converging connection on  $Z$  and  $Z$  received neither hard finding nor fixed probabilistic finding and the same occurs for its descendants.

The notion of d-separation can now be extended to fixed probabilistic evidence by using the definition above. Further studies would be required about d-separation and different types of uncertain evidence, particularly in the context of the propagation of new evidence in a Bayesian network including previous evidence.

#### 4.4 Synthesis of properties of all types of evidence in a Bayesian network

Table 7 summarizes the properties of different types of evidence in a Bayesian network. It is interesting to note that fixed probabilistic evidence has the same properties as hard evidence.

### 5 Uncertain evidence: propagation algorithms in Bayesian networks

This section briefly presents the main algorithms for the propagation of uncertain evidence, also called updating

**Table 6** Characterization of a blocked path between two variables  $X$  and  $Y$ , regarding the kind of evidence on an intermediate variable  $Z$  and the type of connection on  $Z$ 

	<b>No fixed observation on <math>Z</math> and its descendants</b> ( $Z$ and its descendants may receive likelihood findings and not fixed probabilistic findings)	<b><math>Z</math> is observed with certainty</b> ( $Z$ received a hard finding or a fixed probabilistic finding)
$X \dots \rightarrow Z \leftarrow \dots Y$	the path is blocked	new information can be transmitted from $X$ to $Y$
$X \dots \rightarrow Z \rightarrow \dots Y$	new information can be transmitted from $X$ to $Y$	the path is blocked
$X \dots \leftarrow Z \leftarrow \dots Y$		
$X \dots \leftarrow Z \rightarrow \dots Y$		

**Table 7** Synthesis of properties of different types of evidence in a Bayesian network

	Uncertain evidence			
	Hard evidence	Likelihood evidence	Probabilistic evidence not fixed	Probabilistic evidence fixed
The evidence defines a constraint on the posterior probability distribution	Yes	No	Yes	Yes
The evidence is specified “all things considered”	Yes	No	Yes	Yes
Several observations can be combined on the <b>same</b> variable	No	Yes	Yes	No
Posterior probability distribution is modified by later observations on <b>other</b> variables	No	Yes	Yes	No
The propagation requires the prior probability distribution	No	Yes	No	No
The propagation of several findings commutes	Yes	Yes	No	Yes

algorithms. First we present Pearl’s method of virtual evidence for propagating likelihood evidence. Then we present Jeffrey’s rule for probabilistic evidence and we explain why this method is restricted to the case where probabilistic evidence is not fixed. This leads us to discuss the commutativity of the propagation of several probabilistic observations. Next, recent algorithms to propagate probabilistic evidence respecting commutation are listed. Finally, we summarize propagation algorithms for uncertain evidence that are available in the main Bayesian network software.

### 5.1 Pearl’s method of virtual evidence

*Virtual evidence* refers to Pearl’s idea of interpreting a likelihood finding on an event as a hard finding on some virtual event that only depends on this event [61]. The virtual evidence method provides a convenient way of incorporating evidence *with* uncertainty in a Bayesian network.

Pearl’s method to propagate a likelihood finding on  $X$  extends the given Bayesian network by adding a binary virtual node which is a child of  $X$ . The uncertain evidence on  $X$  is replaced by hard evidence on the added node. The hard evidence on the added node is propagated using a classical inference algorithm in the Bayesian network. The uncertainty of the evidence is specified in the conditional probability table of the added virtual node.

The probability distribution  $Q$  representing beliefs after the propagation of a likelihood finding  $L(X)$  by the virtual evidence method is defined as follows. Consider a Bayesian network  $(G, P)$  with  $G = (\mathbf{X}, \mathbf{E})$ , and a likelihood finding on a variable  $X \in \mathbf{X}$ , specified by a likelihood ratio  $L(X)$ . Let  $O$  be the node added in the Bayesian network with the states  $\{o, \bar{o}\}$  where  $o$  is the observation. Let  $G' = (\mathbf{X} \cup \{O\}, \mathbf{E} \cup \{(X, O)\})$  be the augmented graph and  $(G', P')$  the augmented Bayesian network, where the

probability distribution  $P'$  is defined by  $P'(O = o \mid X) = L(X)$  and

$$P'(\mathbf{X} \cup \{O\}) = \prod_{X_i \in \mathbf{X}} P(X_i \mid pa(X_i)) \times P'(O \mid X).$$

With this notation, the posterior probability distribution  $Q$  can be defined by:

$$Q(\cdot) = P'(\cdot \mid O = o) \quad (1)$$

Equation 1 is directly linked with Property 1 stating that likelihood evidence on  $X$  has to be combined with prior belief in  $X$  to be propagated.

### 5.2 Jeffrey’s rule and conversion in likelihood evidence: propagating not-fixed probabilistic evidence

Jeffrey’s rule [34] specifies evidence using posterior probabilities. Propagating a probabilistic finding on  $X \in \mathbf{X}$  requires a revision of the probability distribution  $P$  on  $\mathbf{X}$  by a local probability distribution  $R(X)$ . The difficulty arises since Bayes’ rule cannot be applied because  $R(X)$  is not an event [64]. A probabilistic finding  $R(X)$  requires a reconsideration of the joint probability distribution  $P$  because it replaces the existing prior on the variable  $X$ .

The propagation of probabilistic evidence requires the replacement of the initial probability distribution  $P$  by another probability distribution  $Q$  that reflects the beliefs in the variables of the model after accepting the probabilistic evidence. This replacement is not definitive: it lasts as long as the specific observed case holds, whereas the Bayesian network applies to a larger population.

Jeffrey’s approach for this problem is known as “probability kinematics”, and it is based on the requirements that:

1. the posterior probability distribution on the observed variable  $X$   $Q(X)$  is unchanged:  $Q(X) = R(X)$ ,

2. the conditional probability distribution of other variables given  $X$  remains invariant under the observation:  

$$Q(\mathbf{X} \setminus \{X\} | X) = P(\mathbf{X} \setminus \{X\} | X).$$

Jeffrey's rule is given in (2): for a given local probability distribution  $R(X)$  and for  $Z \in \mathbf{X} \setminus \{X\}$ ,

$$Q(Z = z) = \sum_x P(Z = z | X = x)R(X = x) \quad (2)$$

In other words, even if  $P$  and  $Q$  disagree on  $X$ , they agree on the consequences of  $X$  on other variables.

However, Jeffrey's rule cannot be directly applied to Bayesian networks, because their operations are defined on full joint probability distributions. This can be overcome by converting a probabilistic finding to a likelihood finding:  $R(X)$  can be converted to a likelihood ratio

$$L(X) = \frac{R(x_1)}{P(x_1)} : \dots : \frac{R(x_n)}{P(x_n)}. \quad (3)$$

Propagating the likelihood finding  $L(X)$  with Pearl's method provides the same results as propagating  $R(X)$  by Jeffrey's rule [15, 64]. Thus, the posterior probability of  $X$  after propagating  $L(X)$  by Pearl's method, is equal to  $R(X)$ .

The propagation of a not-fixed probabilistic evidence with Jeffrey's rule is available in the Netica software [58] under the name of "calibration". It requires the user to input  $P(X = x_i | \text{"all observations"})$  for each value  $x_i \in \mathcal{D}_X$ . The term "all observations" means that the probabilistic finding integrates any information about the variable.

In case of several probabilistic findings, the method of converting probabilistic findings into likelihood findings does not preserve probabilistic findings. A simple example can be found in [15, 64]: let  $R_1(X_1)$  and  $R_2(X_2)$  be two pieces of probabilistic evidence, and  $Q$  be the probability distribution reflecting the state of beliefs after considering both findings by using either Jeffrey's rule or by the conversion into likelihood findings. Then  $Q(X_1) \neq R_1(X_1)$  or  $Q(X_2) \neq R_2(X_2)$  depending on the order of propagation. Results are not better when the second probabilistic finding is converted into likelihood finding using its probability revised after propagating the first finding. It therefore holds that the inclusion of several pieces of probabilistic evidence with Jeffrey's rule does not commute. In other words, final beliefs depend on the order of arrival of the probabilistic findings. The next section deals with the propagation of several pieces of fixed probabilistic findings, such that their order does not modify the final belief.

### 5.3 Fixed probabilistic evidence propagating

Propagating a single probabilistic finding can be done by its transformation into a likelihood finding as in (3). This

section concerns the propagation of several fixed probabilistic findings, that is to say that each of the specified probability distributions has to remain unchanged and the order of propagation should have no influence on the final result.

Several algorithms were recently proposed to propagate fixed probabilistic evidence in a Bayesian network. Most of them are based on the *Iterative Proportional Fitting Procedure* (IPFP) algorithm, which is an iterative method of revising a probability distribution to respect a set of given probability constraints in the form of posterior marginal probability distributions over a subset of variables. This algorithm first appeared in the literature in [45], and shortly after was used as a procedure to estimate cell frequencies in contingency tables under some marginal constraints [23]. More recently, a space-saving implementation of IPFP has been proposed [2, 36, 37]. However, the IPFP works on full joint distributions, and thus is not directly applicable to belief update in Bayesian networks. The algorithm could be applied for very small Bayesian networks, but would be infeasible for larger ones since it needs to literally modify each entry of the joint probability distribution table at each iteration.

The *big clique algorithm* is a variation of the junction tree algorithm, based on the IPFP. When constructing the junction tree, all variables involved in a soft finding are fully connected with each other by additional undirected edges. After triangulation, these nodes appear in a single clique (the big clique). The belief update is done by first updating the big clique by running IPFP to convergence and then propagating the resulting distribution of this clique to the rest of the junction tree.

The algorithms BN-IPFP1 and BN-IPFP2 [64] do not modify the junction tree and can work with any Bayesian network inference engine. Both algorithms utilize the IPFP, although in quite different ways. The iterations of BN-IPFP-1, BN-IPFP-2 and Big Clique algorithm all converge to the same distribution [64]. The BN-IPFP-1 algorithm first converts all pieces of probabilistic evidence to likelihood evidence and then iterates using the IPFP to update the Bayesian network until it settles down to a distribution that satisfies all given probabilistic evidence. The BN-IPFP-2 algorithm is more similar to the big clique algorithm, but without modifying the junction tree. BN-IPFP-2 can provide efficient computation when the number of variables involved in the probabilistic evidence is small.

The algorithm SMOOTH was developed by modifying the standard IPFP to support belief update with inconsistent evidence.

Table 8 provides a list of algorithms to propagate fixed probabilistic evidence [63, 64].

**Algorithm 1** BN-IPFP-1

---

**Require:** a Bayesian network  $(G, P)$  with  $G = (\mathbf{X}, \mathbf{E})$  prior distribution  $P(\mathbf{X})$ ,  
 $m$  consistent probabilistic findings on  $\{Y_f, f = 1, \dots, m\} \subset \mathbf{X}$ , specified by  
 $R(Y_1), \dots, R(Y_m)$ .

**Ensure:** the posterior probability distribution  $Q(\mathbf{X})$

$$Q_0(\mathbf{X}) = P(\mathbf{X})$$

$$k \leftarrow 1 \quad [k: \text{index of iterations}]$$

**repeat**

$$f \leftarrow 1 + (k - 1) \bmod m \quad [f: \text{index of current finding}]$$

$$l \leftarrow 1 + \lfloor (k - 1)/m \rfloor \quad [l: \text{index of iterations per finding}]$$

Construct the likelihood ratio

$$L_{f,l}(Y_f) \leftarrow \frac{R(y_f^1)}{Q_{k-1}(y_f^1)} : \dots : \frac{R(y_f^{f_s})}{Q_{k-1}(y_f^{f_s})}$$

where  $y_f^1, \dots, y_f^{f_s}$  are the states of  $Y_f$

Combine the previous likelihood ratios on  $Y_f$

$$L_k(Y_f) \leftarrow \prod_l L_{f,l}(Y_f)$$

Obtain  $Q_k(\mathbf{X})$  by updating  $Q_{k-1}(\mathbf{X})$  with  $L_k(Y_f)$  using Pearl's virtual evidence method

$$k \leftarrow k + 1$$

**until** convergence

---

### 5.3.1 The algorithm BN-IPFP-1

We present here the detail of one of the above algorithms. We choose BN-IPFP1 because of the initial results of the comparison of three algorithms (BN-IPFP1 and 2 and Big-clique) [48].

The BN-IPFP-1 algorithm [64] manages a set of consistent fixed probabilistic findings such that each fixed probabilistic finding  $R(Y_f)$  is dominated by the initial probability distribution ( $R(Y_f) << P(Y_f)$ ), meaning that there is no value  $y$  such that  $P(Y_f = y) = 0$  and  $R(Y_f = y) > 0$ .

The BN-IPFP-1 algorithm is independent of the inference algorithm and combines the IPFP and the conversion of probabilistic findings to likelihood findings. The BN-IPFP-1 algorithm converts separately each probabilistic finding to a likelihood finding and then iterates using IPFP to update the Bayesian network until it settles down to a distribution that satisfies all given probabilistic evidence. At each iteration, a new likelihood ratio is obtained by dividing a probabilistic finding (one at each iteration) by the marginal probability on that variable obtained in the previous step. This new likelihood ratio is then combined with all previous likelihood ratios on the same variable obtained in previous iterations (one for  $m$  iterations).

The proof of the convergence of the algorithm BN-IPFP-1 [64] is based on the convergence of the IPFP.

### 5.3.2 Dealing with the extended notion of fixed probabilistic evidence

The following procedure allows the handling of the extended notion of fixed probabilistic finding (see definition 6). This is done by adding an observation variable that is created as follows:

- First, an observation variable  $Obs$  is created for each piece of fixed probabilistic evidence received. Every

state of the observation variable corresponds to the possible outcomes of the probabilistic finding.

- Second, directed edges to  $Obs$  are added from all variables in the Bayesian network that have a direct influence on the observation, that is to say variables involved in the probabilistic finding (Fig. 5).
- Third, the dependence of the added nodes are modeled by specifying the conditional probability tables  $P(Obs | pa(Obs))$ .

Probabilistic evidence on the added observation node is propagated in the augmented Bayesian network thanks to one of the probabilistic evidence updating methods presented above.

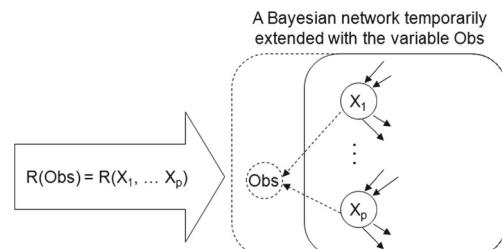
In the case of evidence on a set of observation variables  $E_1, \dots, E_p$  that are independent in the Bayesian network, the propagation can be done by considering a single piece of evidence  $R(E_1, \dots, E_p) = R(E_1) \times \dots \times R(E_p)$ .

## 6 Applications of fixed probabilistic evidence

Although the concept of probabilistic evidence in Bayesian networks was introduced in 1998 [11, 25], it remains little used by the Bayesian network user community. In this section we propose several examples of the use of fixed probabilistic evidence. Firstly we introduce the integration of Bayesian networks with Geographic Information Systems (GIS), secondly applications concerning the propagation of an observation on a continuous variable in a discrete Bayesian network, and thirdly using fixed probabilistic evidence for a distributed Bayesian network.

### 6.1 GIS applications

A range of applications of probabilistic evidence concerns the integration of a Geographic Information System (GIS) with a Bayesian network. The geographic area in focus furnishes the Bayesian network with the fixed probabilistic evidence derived from the GIS database held about a



**Fig. 5** A Bayesian network receives a fixed probabilistic finding on  $(X_1, \dots, X_p)$  (Definition 6). An observation variable  $Obs$  is created in order to consider the new piece of information as a fixed probabilistic finding on a single node

**Table 8** List of algorithms to propagate fixed probabilistic evidence

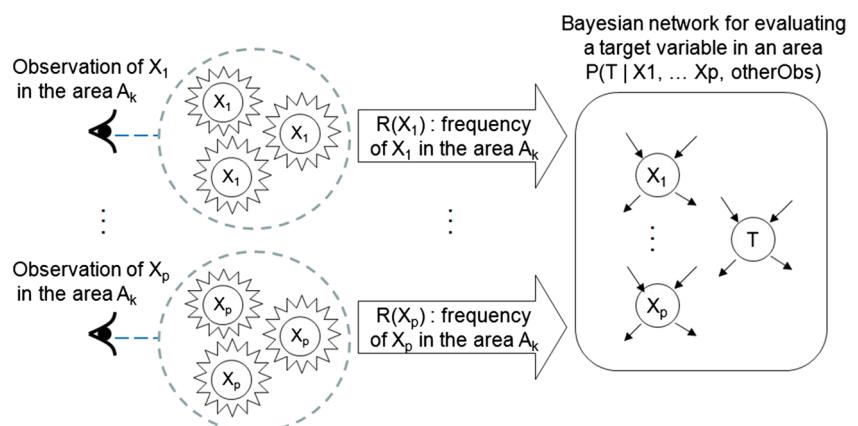
Algorithms	Characteristics
Big-Clique [11, 40, 69]	recoding of the junction tree algorithm based on IPFP
Lazy big clique [48]	modifies the updating algorithm [55]
BN-IPFP 1 and 2 [60, 64]	manages consistent evidence, independent of the inference algorithm, combines IPFP and the conversion to likelihood evidence, BN-IPFP-1: converts separately each probabilistic evidence very expensive if the Bayesian network is large. BN-IPFP-2: uses IPFP to calculate $R(X_1, \dots, X_p)$ from the set $R(X_1), \dots, R(X_p)$ then propagates by transforming to likelihood evidence
Soft updating algorithm [68]	context of hybrid Bayesian network
Approximate algorithms [62]	
IPFP [2, 36, 37]	A space-saving implementation for joint distributions, not for Bayesian networks
GEMA [70]	inconsistent fixed probabilistic evidence for general joint distributions
SMOOTH [64, 74]	propagation of inconsistent fixed probabilistic evidence (proof of convergence for two pieces of evidence.)

sub-population of the population used to derive the CPTs employed in the Bayesian network. For example, a Bayesian network is used to support conflict analysis for groundwater protection and observations about rainfall, groundwater salinity and land use are obtained from the GIS [31]. In a review of Bayesian network applications in ecosystem service modeling [46], the authors point out that in most of the studies which integrate a Bayesian network and a GIS, GIS is used as an input for the Bayesian network, providing probabilistic evidence for each geographical area.

Fixed probabilistic evidence can also be used in a Bayesian network to evaluate the social, economic and environmental impacts of community deployed renewable energy; each geographic area of interest furnishes the Bayesian network with a new fixed probability distribution to describe renewable energy resources, socio-economic parameters and the carbon intensity of displaced fossil fuels [53, 54]. The new probability distributions cannot replace

prior probability distributions in the Bayesian network since they have to be kept fixed; they have to be propagated as fixed probabilistic evidence.

Figure 6 illustrates the use of fixed probabilistic evidence from a GIS. New information about the geographic area of interest is represented by a set of variables  $X_1, \dots, X_p$ . An area is composed of several units, each of them having a specific value for each variable  $X_i$ . The unit can be a pixel, a house, etc. Each observation of a variable  $X_i$  in an area  $A_k$  is obtained from a source  $S_i$  that allows to compute the distribution of  $X_i$  in the area  $A_k$ . The source  $S_i$  can be a database, a GIS or any other source. Since information about an area may come from different sources, it is specified by a list of fixed probabilistic evidence rather than a single joint probability distribution on the variables  $X_1, \dots, X_p$ . Each local probability distribution has to be kept fixed since it represents the variability of the observed variable in the considered area.

**Fig. 6** Fixed probabilistic evidence on a geographic area of interest

The question of representing and propagating uncertainty in geospatial information has been discussed by Laskey and Wright [50]. In particular, the authors propose to represent statistical regularities and uncertain evidence with probabilistic ontologies. This consists of a generic model based on Multi-Entity Bayesian Network (MEBN) that allows the generating of a new Bayesian network for each pixel in the database.

However, propagating uncertain evidence in a Bayesian network remains a topic to explore in order to manage uncertainty in geospatial information.

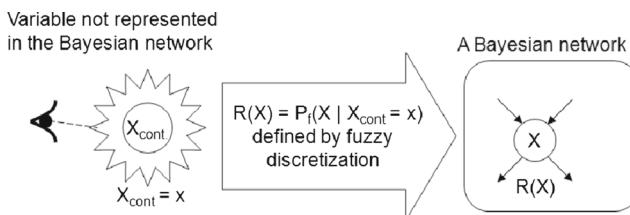
## 6.2 Observations on continuous variables

Here we present the principle of using fixed probabilistic evidence for propagating observations on continuous variables in a discrete Bayesian network, as proposed by Di Tomaso and Baldwin [68].

A common way to deal with a continuous variable in discrete Bayesian network is to discretize it. Despite the loss of information due to the discretization, this technique is broadly used within the framework of Bayesian networks. The impact of the choice of the discretization method in different Bayesian network classifiers has been studied, and it appears that it does not really have an effect when classifiers are being compared [29]. In contrast, the choice of the number of intervals influences the quality of the results. Each interval is considered as a specific discrete value and all the points within the interval are considered as if they were the same discrete value. Thus, they are treated in the same way wherever their position in the interval. For this reason different ways of defining the partition, i.e. with a differing number of intervals and thresholds can give very different results.

Discretization of continuous variables in a Bayesian network is a compromise between three criteria that do not always have the same importance:

(1) Information quality: discretization has to avoid or minimize any loss of information regarding the objectives of the model: to this end, intervals have to be defined such that, if different values for an evidence node deliver a different outcomes for target nodes, they must be contained



**Fig. 7** Fixed probabilistic evidence to manage observation on a continuous variable in a discrete Bayesian network

within different intervals. This criterion leads to the choice of smaller, and therefore a higher number of intervals.

(2) Statistical quality: with due regard for the available data, discretization has to ensure there are enough samples falling within each interval. This aspect is all the more important with insufficient data. This criterion leads to the choice of larger, and therefore a smaller number of intervals.

(3) Computational feasibility: discretization has to preserve the usability and the effectiveness of the model (spatial and temporal complexity of inference). This aspect is all the more important if the discretized variable has several parent nodes and/or child nodes, and if the overall size of the Bayesian Network is large. This criterion leads to the choice of a smaller number of (and therefore larger) intervals in order to create a model with a reasonable size. This is because the number of intervals into which a particular node is discretized, together with both the number of its parents, and their respective number of states, determine the size of its conditional probability table, and thus commensurately the size of the overall model. The size of the model is a limiting factor both for the learning of the Bayesian network and inference making.

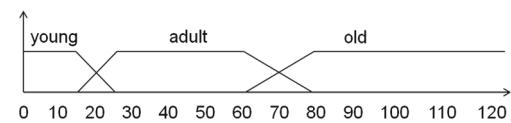
In situations where a very small number of intervals are thought necessary for any of the reasons outlined above, an observation on a continuous variable can be treated by using a small number of fuzzy partitions. Figure 7 shows how hard evidence is substituted by probabilistic evidence [68]. The probabilistic evidence is obtained by fuzzy discretization. This implies the computation of a probability distribution on the discretized variable's set of intervals such that each element of the probability distribution represents the degree of membership to the corresponding interval.

**Definition 8** A fuzzy partition on the universe  $\Omega$  is the set of fuzzy sets  $f_1, \dots, f_p$  such that

$$\forall x \in \Omega, \sum_{i=1}^p X_{f_i(x)} = 1$$

where  $X_{f_i}$  is the membership function of  $f_i$ , i.e. a function  $X_{f_i} : \Omega \mapsto [0, 1]$ .

Choosing a discretization with a small number of intervals reduces the size of the model, which facilitates inference making, and makes the learning step less data demand-



**Fig. 8** Membership functions of the fuzzy discretization of the variable age

**Table 9** Evidence vector obtained after discretization of two observed values of the age

	Discretization of the variable age		
	3 intervals	20 intervals (5 years slices)	fuzzy discretization with 3 states
$age = 19$	(1,0,0)	(0,0,0,1,0,...,0)	(0.6,0.4,0)
$age = 10$	(1,0,0)	(0,1,0,0,0,...,0)	(1,0,0)

ing. The loss of information may be partially compensated by fixed probabilistic evidence propagation.

The appropriateness of using probabilistic evidence rather than likelihood evidence is that likelihood evidence represents a subjective statement that can be improved by something observed later, while probabilistic evidence represents an observation that cannot be improved by anything observed later i.e. it is fixed. This ensures that the probability distribution of the observed variable will not be later influenced by further observations on any other variable. This is required since the initial hard evidence on the continuous variable means that we are certain of the observed value.

**Example 6 (Age)** Let us consider a variable  $age$  that represents the age of a person and a discretization in three intervals: *young* ( $age < 20$ ), *adult* ( $20 \leq age < 70$ ) and *old* ( $age \geq 70$ ). On the one hand, the proposed discretization does not distinguish between 1 year and 19 years whilst on the other, the values 19 years and 21 years are treated differently. An example of fuzzy discretization of the variable age is given in Fig. 8. Table 9 shows the evidence vector obtained after each kind of discretization of age for two observed values ( $age = 19$  and  $age = 10$ ). The probability obtained with fuzzy discretization is computed using the membership function shown in Fig. 8.

### 6.3 Probabilistic evidence for a distributed Bayesian network: AEBN framework

Large real world intelligent systems are often too complex or expensive to be built as centralized systems. Several solutions have been proposed to address this issue, including Agent Encapsulated Bayesian Network (AEBN) [11]. The principle behind this framework has been the prime motivation for the definition and use of fixed probabilistic evidence in a Bayesian network. Indeed, in an agent based model using AEBNs, the belief of a receiver agent is updated following the transmission of probabilistic evidence sent from a publisher agent.

Agent Encapsulated Bayesian Networks (AEBN) were first proposed by Bloemeke [11] and extended by Langevin and Valtorta [47, 49].

**Definition 9** An Agent Encapsulated Bayesian Networks (AEBN) is an agent equipped with a local Bayesian network ( $V, E, P$ ), that represents its internal knowledge base. The set of variables  $V = I \cup L \cup O$  is partitioned into three sets of variables:

- Input variables (I): variables for which other agents have better knowledge
- Output variables (O): variables for which this agent has the best knowledge (oracular knowledge)
- Local or hidden variables (L): variables which are private to this agent and not visible by the other agents.

The set  $E$  contains the edges in the model that define the causal relationships amongst the variables of  $V$ . A joint probability distribution  $P$  is defined over  $V$ . Variables in  $I \cup O$  are shared variables, while variables in  $L$  are private.

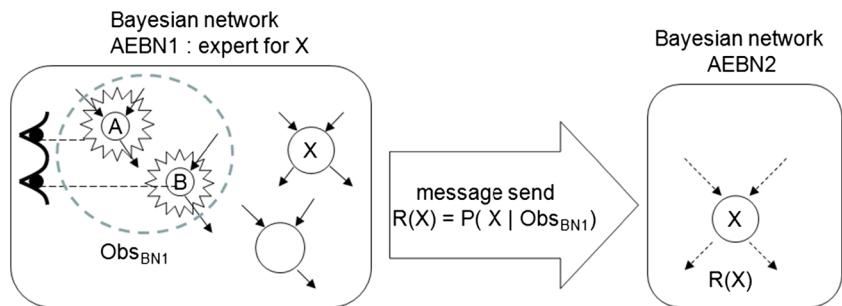
Agents are organized into a publisher/subscriber hierarchy, where agents are publishers of their output variables and subscribers to their input variables. They communicate by sending messages consisting of joint probability distributions over the subset of shared variables. Therefore, each agent that receives messages from other agents obtains fixed probabilistic evidence for one or more observation variables (see Fig. 9). Fixed probabilistic evidence on the added observation node is propagated in the augmented Bayesian network of the receiver agent, thanks to one of the probabilistic evidence update method presented in Table 8.

Fixed probabilistic evidence updating assures a kind of global consistency, since the belief in each shared variable, represented by the marginal posterior probability, is the same in every agent [69].

## 7 Discussion and conclusion

In this last section, we firstly discuss the difference between probabilistic evidence propagation and model revision. Then we examine a recently proposed Fuzzy/Bayesian network and explain why this is actually a case of probabilistic evidence. Finally, we conclude and present the perspectives of that work.

**Fig. 9** Message passing between AEBNs: the AEBN1 is publisher of the variable  $X$  and the AEBN2 is subscriber; it receives a fixed probabilistic finding on  $X$  from the AEBN1



### 7.1 Probabilistic evidence propagation versus model revision

Since probabilistic evidence is a constraint on the posterior probability distribution, it replaces the marginal probability distribution over the variables concerned and thus casts doubt on the joint probability distribution. Should we consider probabilistic evidence to be a knowledge contribution? In this case, why not change the model rather than propagate probabilistic evidence? One paper argues that probabilistic evidence propagation is suitable for model revision and not for updating [71], but this viewpoint is later abandoned by the author [40]. We present two scenarios which show clearly the choice between probabilistic evidence updating and model revision.

Firstly, consider evidence that has a temporary validity. Such evidence may result from the partial observation of a particular state of the modeled system that is valid as long as this state holds. Similarly the probabilistic evidence may come from the observation of a sub-population. A further example of temporary validity presents itself when an observation of a continuous variable is made and transformed via a fuzzy discretization. In those cases, probabilistic evidence can be considered as temporary hard evidence, and thus does not justify model revision.

In contrast the probability distribution  $P$  of a Bayesian network  $(G, P)$  to which the temporary probabilistic evidence is applied, represents *permanent* knowledge which does not need to be revised by short-term evidence.

The second argument deals with the fact that a model is often a compromise between efficiency and accuracy, meaning that the model never includes all parameters of the modeled system. This limit of the model can be partially overcome by using a probabilistic finding to consider information about a variable which is not included in the model. When the observer is able to translate the observed characteristic onto a variable of the model she may apply a probabilistic finding to that variable. While the knowledge embedded in the probability distribution is permanent knowledge this information is relevant only while the specific instance with that characteristic is considered. It can be concluded that information about a specific state of the

modeled system has to be taken into account via propagating evidence (hard, likelihood or probabilistic), even if it considers variables that are not in the model.

Revision of the model occurs when hypotheses associated with it are changed. Several methods for knowledge integration and Bayesian network revision have been proposed [63].

Some of the algorithms to propagate fixed probabilistic evidence (Table 8) have been adapted in order to revise the Bayesian network by the direct replacement of the initial probability distribution by the revised probability distribution. These adaptations are the E-IPFP, E-IPFP-SMOOTH and D-IPFP algorithms [63]. The E-IPFP algorithm integrates the constraints by only changing the conditional probability tables of the given Bayesian network while preserving the network structure. The E-IPFP-SMOOTH and D-IPFP algorithms are two variations of the E-IPFP algorithm. The first deals with the situation where the probabilistic constraints are inconsistent with each other or with the network structure of the given Bayesian network. The second reduces the computational cost by decomposing a global E-IPFP into a set of smaller local E-IPFP problems.

### 7.2 Probabilistic evidence and an example of fuzzy/Bayesian network

Despite a number of papers dealing with fuzzy Bayesian networks, for example [3, 28, 59, 67], there is still no commonly accepted definition of this concept. However, one of the objectives of these is to manage uncertainties in the input of the Bayesian Network. In this section, we examine a recently proposed Fuzzy/Bayesian network to solve a fault detection problem [19]. We explain why this is actually a case of probabilistic evidence.

The real system described in [19] aims to detect short circuits in stator-windings. The problem includes a set of continuous variables representing the difference in magnitude between currents. These variables do not belong to the proposed Bayesian network, but they are observed. Another set of Boolean variables are included in the Bayesian network, each of them associated with an observed continuous

variable. A set of rules allows the determination of the values of the Boolean variables given the continuous variables. If these rules are deterministic, the information on the continuous variables can be transformed into hard evidence on the Boolean variables. In the other cases, it provides uncertain evidence on the Boolean variables of the Bayesian network. In the presented application, these rules consist of a set of membership functions associated to each continuous variable. It provides a vector of membership degree for the states *True* and *False* of the associated Boolean variable. We suggest that each vector *specifies* a probabilistic finding on the Boolean variable. The equation proposed in [19] to define the posterior probability of a variable after such uncertain evidence would require both explanation and justification. The proposed equation is applied in a very simple example, including a single probabilistic finding for each considered case. We compared the results of this application with those obtained with Pearl's method of virtual evidence after converting probabilistic evidence into likelihood ratios<sup>4</sup> and we obtained the same results.<sup>5</sup> Having a single probabilistic finding does not allow the discussion of whether each probabilistic finding (given by the membership degree) has to be kept fixed or not.

The comparison between a fuzzy Bayesian network and probabilistic evidence suggests that further study of their differing properties is required to achieve a better characterization of these methods.

### 7.3 Conclusion and future work

In this paper, we have set out definitions and properties of hard and uncertain findings in a Bayesian network. Three kinds of uncertain evidence are distinguished: likelihood evidence, fixed and not-fixed probabilistic evidence. Evidence is a set of findings on the variables of the Bayesian network. (1) Likelihood evidence is unreliable, imprecise, or vague evidence; it is specified by a likelihood ratio and propagated by Pearl's method of virtual evidence. This method translates a likelihood finding on a variable  $X$  onto a hard evidence on a new virtual node, added as a child of the node  $X$ . (2) Fixed or not-fixed probabilistic evidence expresses a constraint on the state of some variables after this information has been propagated in the Bayesian network. A probabilistic finding on  $X$  is specified by a probability distribution  $R(X)$  that is given "all things considered", meaning that it replaces any former belief and knowledge on  $X$ . (2a) A not-fixed probabilistic finding can be propagated by converting it in likelihood evidence (see

<sup>4</sup>We use the Bayesian software Netica, Menu: Enter finding / function: calibration.

<sup>5</sup>There seems to be an error in the results displayed in Table 2 and 3 in [19] since they do not correspond to the results announced in the text and referring to these tables.

Eq. 3). It can be later modified by further evidence on any node of the Bayesian network, including  $X$  itself. (2b) Fixed probabilistic evidence is also known as soft evidence. A probabilistic finding on a variable  $X$  cannot be altered by any further information on variables in the model. Thus the propagation of several pieces of fixed probabilistic evidence is commutative. Fixed probabilistic evidence has to be propagated by specific algorithms such as big clique or BN-IPFP (Table 8).

We provide several examples where the application and propagation of fixed probabilistic evidence in a Bayesian network is of interest. In an agent organization based on an AEBN, a probabilistic finding on a shared variable  $X$  is used by a receiver agent to take into account the information sent by a publisher agent which is considered to be an expert on  $X$ . In a Bayesian network with discretized continuous variables, a fixed probabilistic finding on such a variable  $X$  is obtained when using fuzzy discretization, which for a coarse discretization maintains a greater fidelity to a more granular discretization. Finally, a fixed probabilistic finding can also summarize observations about a set of instances of a particular subgroup such as in Example 4.

This paper aims to contribute to the standardization and clarification of the definitions and properties of different types of evidence in a Bayesian network.

A number of terms for hard evidence are in use (specific finding, positive finding, regular finding, deterministic finding, observation), but there is no major semantic difference. The general understanding is that hard evidence is when there is a single one in the observation vector.

Negative evidence is characterized by observation vectors of zeros and ones which may include several ones. A negative finding is not a hard finding, except in case the observation vector contains a single one. It is a specific case of likelihood finding where the observation does not permit the making of a distinction between included values.

Terms to describe uncertain evidence have been made clearer: likelihood evidence, also called virtual evidence has to be clearly distinguished from probabilistic evidence, both in terms of specification and of propagation. This latter notion includes two sub-groups, fixed or not-fixed probabilistic evidence, regarding the impacts of future information on the constraint defined by the probabilistic evidence. Both classes are specified by a local probability distribution. The distinction between fixed and not-fixed probabilistic evidence, which is around the question of commutativity had already been raised and discussed in previous papers [15, 64], but it had not led to distinct definitions. Fixed probabilistic evidence was initially called soft evidence.

The problem addressed in this paper is one of belief updating and not the problem of model revision, which leads to a change of the probability distribution (or even the

graph) of the Bayesian network. In the case of probabilistic evidence, information is probabilistic in nature and leads to the replacement (temporarily) of the prior distribution extant in the defined model.

Currently, many Bayesian network engines allow the propagation of likelihood evidence, even though the terminology is not yet standardized. However, very few of them possess the ability to propagate fixed or not-fixed probabilistic evidence. This would be of a great interest and utility to the Bayesian network user community. The introduction of features in Bayesian network engines to allow the propagation of uncertain evidence in a Bayesian network would require clarity of terminology so that the user is confident about the type of uncertain evidence propagation being invoked. At least three new features are required in most Bayesian network engines. The first is a requirement to enter an observation on a subgroup of instances instead of a single instance. This would be useful for applications such as classification of sub-populations, as in [30] concerning the detection of atmospheric pollution, since each case represents both a sequence of images during a short period of time and a set of areas of interest in the image. Another useful application concerns the analysis of a body of text, such as automatic text summarization [27]. The use of probabilistic findings would allow the exploitation of features of the text both at the level of the sentence and at the level of a section, considered as a set of sentences. Secondly, software should facilitate the implementation of an agent organization based on AEBNs. This requires a subscriber agent to integrate the information from a corresponding publisher agent by propagating fixed probabilistic evidence in its local Bayesian network. This would be very useful in large applications based on probabilistic graphical models such as multi criteria decision aiding framework for recurrent problems [22]. Other set of applications that could benefit from AEBNs concerns forensic science. For example, in order to guide criminal investigations [39], the knowledge based systems elaborates a Bayesian network that regroups a set of plausible scenarios in order to determine which investigation strategies are likely to produce the most conclusive evidence. Such approaches could be augmented by considering some types of uncertain evidence, such as the confidence in a person's testimony, or the importance of a person's financial situation. The uncertainty of this evidence is not reducible and could be taken into account as probabilistic evidence. Such uncertain evidence could be obtained from experts, each of them using her own Bayesian network. The resulting probabilistic evaluation could be exploited by the main agent of the system, in an AEBN framework. The third feature of interest would allow the propagation of observations on continuous variables using a fuzzy discretization with degrees of membership of two or more coarse intervals.

Another feature of interest would allow the combination of several uncertain findings on the same variable when appropriate: for example when combining not-fixed probabilistic finding with a likelihood finding on the same variable, or several likelihood findings. At present, very few Bayesian network software applications allow such combinations. Another area of work to be done is a detailed comparison with the different types of uncertain evidence presented in this paper and the interesting proposition of Bessière [8] of using coherence variables to fuse prior belief with evidence specified by a probability distribution.

This research has been supported by the International Campus on Safety and Intermodality in Transportation; the Nord/Pas-de-Calais Region, the European Community, the Regional Delegation for Research and Technology, the Ministry of Higher Education and Research, and the National Center for Scientific Research. The authors gratefully acknowledge the support of these institutions.

## References

1. Ahuactzin JM, Bessière P, Mazer E, Mekhnacha K (2014) ProBT, Computer software. ProBayes, Grenoble. <http://www.probayes.com/>
2. Badsgaard JH, Malvestuto FM (2001) An implementation of the iterative proportional fitting procedure by propagation trees. *Comput Stat Data Anal* 37:297–322
3. Baldwin JF, Tomaso ED (2003) Inference and learning in fuzzy bayesian networks. In: The 12th IEEE International conference on fuzzy Systems, FUZZ-IEEE 2003, St. Louis, pp 630–635
4. Ben Mrad A, Delcroix V, Maalej MA, Piechowiak S, Abid M (2012) Uncertain evidence in Bayesian networks : Presentation and comparison on a simple example. In: 2012 Proceedings of the 14th conference on information processing and management of uncertainty in knowledge-based systems, IPMU, Catania, pp 39–48
5. Ben Mrad A, Delcroix V, Piechowiak S, Maalej MA, Abid M (2013) Understanding soft evidence as probabilistic evidence: Illustration with several use cases. In: 2013 5th international conference on modeling, simulation and applied optimization (ICMSAO), pp 1–6
6. Ben Mrad A, Maalej MA, Delcroix V, Piechowiak S., Abid M (2011) Fuzzy evidence in Bayesian networks. In: Proceedings of computing, soft, recognition, pattern, Dalian
7. Benferhat S, Tabia K (2012) Inference in possibilistic network classifiers under uncertain observations. *Ann Math Artif Intell* 64(2–3):269–309
8. Bessière P, Mazer E, Ahuactzin JM, Mekhnacha K (2013) Bayesian Programming. CRC Press
9. Bilmes J (2004) On soft evidence in Bayesian networks. Tech. Rep. UWEETR-2004-00016. Department of Electrical Engineering, University of Washington, Seattle
10. Birtles N, Fenton N, Neil M, Tranham E (2014) AgenaRisk manual (Version 6.1) Computer software. <http://www.agenaRisk.com/>
11. Bloemeke M (1998) Agent encapsulated Bayesian networks. Ph.D. thesis, Department of Computer Science, University of South Carolina

12. Butz CJ, Fang F (2005) Incorporating evidence in Bayesian networks with the select operator. In: Proceedings of the 18th Canadian conference on artificial intelligence. Springer-Verlag, pp 297–301
13. Chan H (2005) Sensitivity analysis of probabilistic graphical models. Ph.D. thesis. University of California, Los Angeles
14. Chan H, Darwiche A (2004) Sensitivity analysis in Bayesian networks: From single to multiple parameters. In: UAI, pp 67–75
15. Chan H, Darwiche A (2005) On the revision of probabilistic beliefs using uncertain evidence. *Artif Intell* 163(1):67–90
16. Cooper GF (1990) The computational complexity of probabilistic inference using Bayesian belief networks. *Artif Intell* 42:393–405
17. Dagum P, Luby M (1993) Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artif Intell* 60:141–153
18. D'Ambrosio B, Takikawa MDU (2000) Representation for dynamic situation modeling. Technical report, Information Extraction and Transport, Inc.
19. D'Angelo MFSV, Palhares RM, Cosme LB, Aguiar LA, Fonseca FS, Caminhas WM (2014) Fault detection in dynamic systems by a fuzzy/bayesian network formulation. *Appl Soft Comput* 21:647–653
20. Darwiche A (2009) Modeling and reasoning with Bayesian networks. Cambridge University Press, Cambridge
21. Darwiche A (2014) Samlam Computer software. University of California, Los Angeles. <http://reasoning.cs.ucla.edu/samiam>
22. Delcroix V, Sedki K, Lepoutre FX (2013) A Bayesian network for recurrent multi-criteria and multi-attribute decision problems, Choosing a manual wheelchair. *Expert Syst Appl* 40(7):2541–2551
23. Deming WE, Stephan FF (1940) On a least square adjustment of a sampled frequency table when the expected marginal totals are known. *Ann Math Stat* 11:427–444
24. Drudzel MJ (2014) Genie smile, Version 20. Computer software. Decision systems laboratory. University of Pittsburgh, Pennsylvania. <http://genie.sis.pitt.edu>
25. Dubois D, Moral S, Prade H (1998) Belief change rules in ordinal and numerical uncertainty theories. In: Gabbay D, Smets P, Dubois D, Prade H (eds) Belief change vol. 3 of the handbook of defeasible reasoning and uncertainty systems, management. Kluwer Academic Publishers, Dordrecht, pp 311–392
26. Elvira (2014) Elvira project, <http://leo.ugr.es/elvira/>
27. Fattah MA (2014) A hybrid machine learning model for multi-document summarization. *Appl Intell* 40(4):592–600
28. Ferreira L, Borenstein D (2012) A fuzzy-bayesian model for supplier selection. *Expert Syst Appl* 39(9):7834–7844
29. Flores MJ, Gámez JA, Martínez AM, Puerta JM (2011) Handling numeric attributes when comparing Bayesian network classifiers: Does the discretization method matter? *Appl Intell* 34(3):372–385
30. Gacquer D, Delcroix V, Delmotte F, Piechowiak S (2011) Comparative study of supervised classification algorithms for the detection of atmospheric pollution. *Eng Appl Artif Intell* 24(6):1070–1083
31. Giordano R, D'Agostino D, Apollonio C, Lamaddalena N, Vurro M (2013) Bayesian belief network to support conflict analysis for groundwater protection: The case of the Apulia region. *J Environ Manag* 115C:136–146
32. Henrion M (2014) Analytica, version computer software. Lumina decision systems, Los Gatos. <http://www.lumina.com/>
33. Höjsgaard S (2014) gRain, (Version 1.2-3) Computer software. Aalborg University, Denmark. <http://people.math.aau.dk/sorenh/software/gR/>
34. Jeffrey RC (1990) The logic of decision, 2nd edn. University of Chicago Press, Chicago, p 246
35. Jensen FV, Nielsen TD (2007) Bayesian networks and decision graphs, 2nd. Springer Publishing Company Incorporated
36. Jiroušek R (1991) Solution of the marginal problem and decomposable distributions. *Kybernetika* 27:403–412
37. Jiroušek R, Přeučil S (1995) On the effective implementation of the iterative proportional fitting procedure. *Comput Stat Data Anal* 19(2):177–189
38. Jouffe L, Munteanu P (2014) BayesiaLab, Laval. <http://www.bayesia.com>
39. Keppens J, Shen Q, Price C (2011) Compositional bayesian modelling for computation of evidence collection strategies. *Appl Intell* 35(1):134–161
40. Kim YG, Valtorta M, Vomlel J (2004) A prototypical system for soft evidential update. *Appl Intell* 21(1):81–97
41. Kjaerulff U, Madsen A (2013) Bayesian networks and influence diagrams: A guide to construction and analysis, vol 22, 2nd edn. Springer
42. Korb K, Nicholson A (2010) Bayesian artificial intelligence, 2nd. Chapman and Hall, london
43. Koski T, Noble J (2009) Bayesian networks: An introduction. Wiley series in probability and statistics. Wiley, Chichester
44. Krieg ML (2001) A tutorial on Bayesian belief networks. Technical Report DSTO-TN-0403, surveillance systems division electronics and surveillance research laboratory. Defense science and technology organisation, Edinburgh
45. Kruithof R (1937) Telefoonverkeersrekening. *De Ingenieur* 52:15–25
46. Landuyt D, Broekx S, D'hondt R, Engelen G, Aertsens J, Goethals P (2013) A review of Bayesian belief networks in ecosystem service modelling. *Environ Model Softw* 46:1–11
47. Langevin S (2011) Knowledge representation, communication, and update in probability-based multiagent systems. Ph.D. thesis. University of South Carolina, Columbia. AAI3454755
48. Langevin S, Valtorta M (2008) Performance evaluation of algorithms for soft evidential update in Bayesian networks: First results. In: SUM. 284–297
49. Langevin S, Valtorta M, Bloemeke M (2010) Agent-encapsulated Bayesian networks and the rumor problem. In: AAMAS '10 Proceedings of the 9th international conference on autonomous agents and multiagent systems, vol 1, pp 1553–1554
50. Laskey KB, Wright EJ, da Costa PCG (2010) Envisioning uncertainty in geospatial information. *Int J Approx Reason* 51(2):209–223
51. Lauritzen SL (2014) Hugin, Version 8.0 Computer software, Aalborg. <http://www.hugin.com>
52. Lauritzen SL, Spiegelhalter DJ (1988) Local computations with probabilities on graphical structures and their application to expert systems. *J R Stat Soc Ser B* 50:157–224. doi:[10.2307/2345762](https://doi.org/10.2307/2345762)
53. Leicester PA, Goodier CI, Rowley P (2012) Community energy delivers megawatts, pounds, carbon reductions, etcetera. In: Midlands energy graduate school (MEGS)
54. Leicester PA, Goodier CI, Ro wley P (2013) Using a Bayesian network to evaluate the social, economic and environmental impacts of community renewable energy. In: Clean technology for smart cities and buildings (CISB AT)
55. Madsen AL, Jensen FV (1999) Lazy propagation: A junction tree inference algorithm based on lazy evaluation. *Artif Intell* 113(1–2):203–245
56. Minka T, Winn J (2014) Infer.NET, (Version 205) Computer software. Microsoft Research, Cambridge. <http://research.microsoft.com/en-us/um/cambridge/projects/infernet/default.aspx>
57. Murphy K (2014) Bayesian Network Toolbox (BNT), (Version 1.0.7) Computer software. MIT AI lab, Cambridge. <http://www.cs.ubc.ca/murphyk/Software/BNT/bnt.html>
58. Norsys (2014) Netica application, (Version 5.12) Computer software. Norsys Software Corp, Vancouver. <http://www.norsys.com>

59. Pan H, Liu L (2000) Fuzzy bayesian networks - A general formalism for representation, inference and learning with hybrid bayesian networks, vol 14, pp 941–962
60. Pan R, Peng Y, Ding Z (2006) Belief update in Bayesian networks using uncertain evidence. In: ICTAI, pp 441–444
61. Pearl J (1988) Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann Publishers, San Mateo
62. Peng Y, Ding Z (2005) Modifying bayesian networks by probability constraints. In: Proceedings of the 21st Conference in uncertainty in artificial intelligence, Edinburg July 26-29 UAI '05, pp 459–466
63. Peng Y, Ding Z, Zhang S, Pan R (2012) Bayesian network revision with probabilistic constraints. *Int J Uncertainty Fuzziness Knowledge Based Syst* 20(3):317–337
64. Peng Y, Zhang S, Pan R (2010) Bayesian network reasoning with uncertain evidences. *Int J Uncertainty Fuzziness Knowledge Based Syst* 18(5):539–564
65. Pourret O, Naim P, Marcot B (2008) Bayesian networks: A practical guide to applications, statistics in practice, Wiley
66. Sandiford J (2014) Bayes Server, (Version 5.5) Computer software, East Preston. <http://www.bayesserver.com/>
67. Tang H, Liu S (2007) Basic theory of fuzzy bayesian networks and its application in machinery fault diagnosis. IEEE Computer Society, Washington
68. Tomaso ED, Baldwin JF (2008) An approach to hybrid probabilistic models. *Int J Approximate Reasoning* 47(2):202–218
69. Valtorta M, Kim YG, Vomlel J (2002) Soft evidential update for probabilistic multiagent systems. *Int J Approx Reason* 29(1):71–106
70. Vomlel J (2004) Integrating inconsistent data in a probabilistic model. *J Appl Non-Classical Log* 14(3):367–386
71. Vomlel J, Probabilistic reasoning with uncertain evidence. *Neural network world* (2004) *Int J Neural Mass-Parallel Comput Inf Syst* 14(5):453–465
72. Wang Y, Zhang NL, Chen T (2008) Latent tree models and approximate inference in Bayesian networks. *J Artif Intell Res* 32:879–900
73. Yuan C, Lim H, Lu TC (2011) Most relevant explanation in Bayesian networks. *J Artif Intell Res* 42(1):309–352
74. Zhang S, Peng Y, Wang X (2008) An Efficient Method for Probabilistic Knowledge Integration. In: Proceedings of The 20th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2008), November 3–5, 2008, vol 2. Dayton, pp 179–182



**Ali Ben Mrad** is a PhD student under cooperation between CES laboratory, University of Sfax (Tunisia) and LAMIH laboratory, University of Valenciennes (France). He received the master degree in Computer Science from University of Sfax in 2010.

His research is interesting in reasoning with uncertainty with probabilistic graphical models. He published in several international conferences, about different kind of Bayesian networks

applications.

Ali is also member of various research networks such as the reflection group “Bayesian Networks” and European Community for Researchers on Probabilistic Graphical Models (BN @work) and Executive Committee member of IEEE Engineering in Medicine and Biology Society (EMBS).



**Véronique Delcroix** after graduating from the National Institute of Applied Sciences (INSA) of Rouen, Véronique Delcroix received her PhD in Computer Science from the University of Paris VI (France). Since 1998, she is an associate professor at the university of Valenciennes, in the North of France. She teaches in an engineer school (ENSIAME) and belongs to a multidisciplinary laboratory (LAMIH) in the field of Artificial Intelligence.

Her research focuses on reasoning with uncertainty with probabilistic graphical models. She published in several international journals, about different kind of Bayesian networks applications. In particular, she proposed a system of recommendation based on Bayesian networks for recurrent multicriteria decision problems.



**Sylvain Piechowiak** is full professor in computer science at University of Valenciennes since 2003. He got his PhD in 1992 and his HDR in this university.

His research area concerns: constraint based reasoning (temporal, distributed, etc) and uncertainty based reasoning.

He uses these paradigms of reasoning to solve model-based diagnosis in transport systems or in health organizations, to support multi-agent systems in traffic simulation

and to elaborate timetables in university.

He is the author of many articles and papers in international journals and in international congresses or workshops.



**Mohamed Abid** is working now as a Professor at the Engineering National School of Sfax (ENIS), University of Sfax, Tunisia. He received the Ph.D. degree from the National Institute of Applied Sciences, Toulouse (France) in 1989 and the “thèse d'état” degree from the National School of Engineering of Tunis (Tunisia) in 2000 in the area of Computer Engineering & Microelectronics. He occupied the post of director of doctoral school “Sciences

& Technologies”, University of Sfax. He is founding member and Head of the research laboratory «Computer Embedded System» CES-ENIS, since 2006 (<http://www.ceslab.org>).



**Philip Leicester** is a Research Associate in Loughborough University working on the widespread penetration of solar PV and other low carbon interventions in to the energy system.

His research concerns the transition to a more sustainable, distributed, low-carbon energy infrastructure and key focus of this research is the evaluation of social, environmental and economic impacts using probabilistic approaches. In particular

integrated modelling in the urban context allows the assessment of carbon reductions, fuel affordability and returns on investments for both domestic and non-domestic technology adopters. This work has benefits from due consideration of probabilistic evidence in object oriented Bayesian networks. His contribution to this work has been conducted as part of the research project ‘PV2025 – Potential Costs and Benefits of Photovoltaic for UK Infrastructure and Society’ project which is funded by the RCUK’s Energy Programme (contract no: EP/K02227X/1).