International Journal of Approximate Reasoning 160 (2023) 108967

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

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

Making inferences in incomplete Bayesian networks: A Dempster-Shafer belief function approach

Prakash P. Shenoy

Analytics, Information, and Operations Area, School of Business, University of Kansas, 1654 Naismith Dr., Capitol Federal Hall, Lawrence, KS, 66045, USA

A R T I C L E   I N F O

Article history:
Received 30 November 2022
Received in revised form 13 May 2023
Accepted 14 June 2023
Available online 20 June 2023

Keywords:
Incomplete Bayesian networks
Dempster-Shafer belief function theory
Conditional belief functions
Smets' conditional embedding

A B S T R A C T

How do you make inferences from a Bayesian network (BN) model with missing information? For example, we may not have priors for some variables or may not have conditionals for some states of the parent variables. It is well-known that the Dempster-Shafer (D-S) belief function theory is a generalization of probability theory. So, a solution is to embed an incomplete BN model in a D-S belief function model, omit the missing data, and then make inferences from the belief function model. We will demonstrate this using an implementation of a local computation algorithm for D-S belief function models called the "Belief function machine." One advantage of this approach is that we get interval estimates of the probabilities of interest. Using Laplacian (equally likely) or maximum entropy priors or conditionals for missing data in a BN may lead to point estimates for the probabilities of interest, masking the uncertainty in these estimates. Bayesian reasoning cannot reason from an incomplete model. A Bayesian sensitivity analysis of the missing parameters is not a substitute for a belief-function analysis.

1. Introduction

Pearl [16, Ch. 9, p. 415] writes: "Pure Bayesian theory requires the specification of a complete probabilistic model before reasoning can commence. In causal modeling, for example, this means determining for each variable X, the conditional probabilities of the values of X given the factors perceived as causes of those values."

So, what steps can we take when we encounter missing information in a probability model? This article proposes using the Dempster-Shafer (D-S) theory of belief functions to draw inferences in Bayesian networks (BNs) with missing data. For instance, we may lack prior knowledge of certain variables or miss conditionals for some variables for some states of their parents. Our solution is to put an incomplete BN within a corresponding D-S belief function model, excluding the missing data, and then use the D-S theory to make inferences. We may only have partial knowledge about priors and conditionals, which standard Bayesian theory cannot accommodate, but D-S theory can. It is widely known that the D-S belief function theory is a more general version of the Bayesian probability theory. So, a corresponding D-S model will produce the same results if we possess complete information in a probability model.

One benefit of utilizing the D-S approach is obtaining interval estimates for the probabilities of interest. However, point estimates are produced when employing Laplacian or maximum entropy priors or conditionals for missing data in a BN,

E-mail address: pshenoy@ku.edu.
URL: https://pshenoy.ku.edu.

https://doi.org/10.1016/j.ijar.2023.108967

which mask the uncertainty in these estimates. Making inferences from a D-S belief function model is computationally more expensive than inferences from a BN model [15,22]. Nonetheless, local computational algorithms are available for making inferences from a D-S belief function model [28,14], and these algorithms have been implemented for solving large belief function graphical models.

A Bayesian sensitivity analysis of missing parameters may lead to probability intervals with narrower widths than those produced by a corresponding D-S analysis. This is because the pure Bayesian method cannot model complete or partial ignorance. Also, a Bayesian sensitivity analysis of several missing parameters is computationally intractable.

An outline of the remainder of the paper is as follows. In Section 2, we review the basics of the D-S theory, including conditional belief functions, conditional independence, and graphical models. Section 3 discusses representing a BN by an equivalent D-S belief function model, including a small example. Section 4 discusses making inferences from incomplete BNs and includes some new results about the nature of probability intervals resulting from a D-S analysis. Section 5 discusses using "Belief Function Machine," a set of routines in Matlab, for making inferences from large belief function graphical models. Finally, Section 6 concludes with a summary and comments on further work.

2. Basics of the D-S theory of belief functions

This section sketches the basics of the D-S theory of belief functions [7,18].

2.1. Representations

We represent knowledge using basic probability assignments, belief functions, plausibility functions, and credal sets.

Notation Let $\mathcal{V}$ denote the set of all variables. Let X, Y, Z, etc., denote elements of $\mathcal{V}$. Let r, s, t, etc., denote subsets of $\mathcal{V}$. Consider $s \subseteq \mathcal{V}$. For each $X \in s$, let Ω_X denote its finite state space, and let $\Omega_s = \times_{X \in s} \Omega_X$ denote the state space of s. Let 2^{Ω_s} denote the set of all subsets of Ω_s.

Basic probability assignment A *basic probability assignment* (BPA) m for s is a function $m : 2^{\Omega_s} \to [0, 1]$ such that

$$m(\emptyset) = 0, \text{ and} \tag{1}$$

$$\sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_s}} m(\mathsf{a}) = 1. \tag{2}$$

m represents some knowledge about the variables in s, and we say the *domain* of m is s. $m(\mathsf{a})$ is the probability assigned to the proposition represented by subset a of Ω_s. Subsets a such that $m(\mathsf{a}) > 0$ are called *focal elements* of m. If all the focal elements of m are singleton subsets of Ω_s, we say m is *Bayesian*. There is a 1-1 correspondence between a Bayesian BPA m and a corresponding probability mass function (PMF) P for s such that $P(a) = m(\{a\})$ for all $a \in \Omega_s$. If m has only one focal element (with probability 1), we say m is *deterministic*. If the focal element of a deterministic BPA is Ω_s, we say m is *vacuous*. A vacuous BPA for s is sometimes denoted by ι_s.

Belief function The knowledge encoded in a BPA m can be represented as a corresponding belief, plausibility, and credal function. The *belief function* Bel_m corresponding to BPA m is such that for all $\mathsf{a} \subseteq \Omega_s$,

$$Bel_m(\mathsf{a}) = \sum_{\mathsf{b} \subseteq \mathsf{a}} m(\mathsf{b}). \tag{3}$$

$Bel_m(\mathsf{a})$ represents the smallest degree of support for proposition a implied by BPA m.

Plausibility function The *plausibility function* Pl_m corresponding to BPA m is such that for all $\mathsf{a} \subseteq \Omega_s$,

$$Pl_m(\mathsf{a}) = \sum_{\mathsf{b} \subseteq \Omega_X : \mathsf{b} \cap \mathsf{a} \neq \emptyset} m(\mathsf{b}). \tag{4}$$

$Pl_m(\mathsf{a})$ represents the largest degree of support for proposition a implied by BPA m. It follows from Eqs. (3) and (4) that $0 \leq Bel_m(\mathsf{a}) \leq Pl_m(\mathsf{a}) \leq 1$, and $Pl(\mathsf{a}) = 1 - Bel(\bar{\mathsf{a}})$ and $Bel(\mathsf{a}) = 1 - Pl(\bar{\mathsf{a}})$, where $\bar{\mathsf{a}} = \Omega_s \setminus \mathsf{a}$ denotes the complement of a.

Credal sets A BPA can be represented as a 'credal' set of PMFs. Let $\mathcal{P}$ denote the set of all PMFs for s. Then the credal set corresponding to m, denoted by Cr_m, is defined as follows:

$$Cr_m = \{P \in \mathcal{P} : \text{for all } \mathsf{a} \subseteq \Omega_s,\, Bel_m(\mathsf{a}) \leq P(\mathsf{a}) \leq Pl_m(\mathsf{a})\} \tag{5}$$

In Eq. (5), only one inequality is needed—either one implies the other. For a vacuous BPA m, the credal set Cr_m consists of all PMFs $P \in \mathcal{P}$. For a Bayesian BPA corresponding to PMF P, the corresponding credal set consists of the single PMF P.

Bel_m, Pl_m, and Cr_m have exactly the same information as m. Given any one of these, we can recover the others.

2.2. Inference operators

There are two basic inference operators in the D-S theory, marginalization and combination. As the combination of two BPAs on different domains involves a vacuous extension, we also describe a vacuous extension of a BPA.

Marginalization Suppose m is a BPA for s and suppose $t \subseteq s$. The marginalization operator transforms a BPA m for s to a BPA $m^{\downarrow t}$ for t by eliminating variables in $s \setminus t$.

Projection of states means dropping some coordinates. If $(x, y) \in \Omega_{X,Y}$, then $(x, y)^{\downarrow X} = x$. The projection of a subset of states is achieved by projecting every state in the subset. Suppose $a \subseteq \Omega_{X,Y}$. Then,

$$a^{\downarrow X} = \{x \in \Omega_X : (x, y) \in \mathsf{a}\}.$$

Suppose m is a BPA for s, and $t \subseteq s$. Then, the marginal for m for t, denoted by $m^{\downarrow t}$, is a BPA for t such that for each $\mathsf{a} \subseteq \Omega_t$,

$$m^{\downarrow t}(\mathsf{a}) = \sum_{\mathsf{b} \subseteq \Omega_s : \mathsf{b}^{\downarrow t} = \mathsf{a}} m(\mathsf{b}). \tag{6}$$

The marginalization operator satisfies the following property. Suppose m is a BPA for s and suppose X_1 and X_2 are two distinct variables in s. Then

$$(m^{\downarrow s \setminus \{X_1\}})^{\downarrow s \setminus \{X_1, X_2\}} = (m^{\downarrow s \setminus \{X_2\}})^{\downarrow s \setminus \{X_1, X_2\}}. \tag{7}$$

Thus, the order in which variables are eliminated does not matter.

Vacuous extension Suppose m is a BPA for r, and suppose $r \subseteq t$. A vacuous extension of m to t, denoted by $m^{\uparrow t}$, is a BPA for t such that for all $\mathsf{a} \subseteq \Omega_r$

$$m^{\uparrow t}(\mathsf{a} \times \Omega_{t \setminus r}) = m(\mathsf{a}) \tag{8}$$

Thus, each focal element of $m^{\uparrow t}$ is a vacuous extension of a corresponding focal element of m with the same BPA mass. The vacuous extension is implicit in the definition of Dempster's combination. A numerical example of a vacuous extension is found in Example 3 in Section 3.

Dempster's combination rule Suppose m_1 is a BPA for s_1, m_2 is a BPA for s_2, and m_1 and m_2 are *distinct*. Then, $m_1 \oplus m_2$ is a BPA for $s_1 \cup s_2$ such that for all $\mathsf{a} \in 2^{\Omega_{s_1 \cup s_2}}$

$$(m_1 \oplus m_2)(\mathsf{a}) = K^{-1} \sum_{\mathsf{a}_1 \in 2^{\Omega_{s_1}}, \mathsf{a}_2 \in 2^{\Omega_{s_2}} : (\mathsf{a}_1 \times \Omega_{s_2 \setminus s_1}) \cap (\mathsf{a}_2 \times \Omega_{s_1 \setminus s_2}) = \mathsf{a}} m_1(\mathsf{a}_1)\, m_2(\mathsf{a}_2), \tag{9}$$

where K is a normalization constant given by

$$K = \sum_{\mathsf{a}_1 \in 2^{\Omega_{s_1}}, \mathsf{a}_2 \in 2^{\Omega_{s_2}} : (\mathsf{a}_1 \times \Omega_{s_2 \setminus s_1}) \cap (\mathsf{a}_2 \times \Omega_{s_1 \setminus s_2}) \neq \emptyset} m_1(\mathsf{a}_1)\, m_2(\mathsf{a}_2). \tag{10}$$

We assume $K > 0$. If $K = 0$, then m_1 and m_2 are said to be in total conflict and cannot be combined. If $K = 1$, we say m_1 and m_2 are *non-conflicting*.

Some comments on Dempster's combination rule:

1. Dempster called the combination rule "product-intersection" rule [7]. The product of the BPA values $m_1(\mathsf{a}_1)\, m_2(\mathsf{a}_2)$ is assigned to the intersection of the focal elements of m_1 and m_2.
2. If the BPAs being combined have different domains, we have to vacuously extend them to the union of the domains before we can intersect the focal elements.
3. If two or more intersections result in the same focal element of $m_1 \oplus m_2$, we add the products of the BPA masses. Hence the summation in Eq. (9).
4. If an intersection of focal elements of m_1 and m_2 results in the $\emptyset$, we discard the product of the BPA values and renormalize as in Eq. (10).
5. In general $m \oplus m \neq m$. Thus, Dempster's combination must be used to combine only distinct knowledge to avoid double-counting of non-idempotent knowledge [7]. Distinct belief functions are discussed further in Subsection 2.6.

It is easy to show that Dempster's combination is commutative and associative: $m_1 \oplus m_2 = m_2 \oplus m_1$, and $(m_1 \oplus m_2) \oplus m_3 = m_1 \oplus (m_2 \oplus m_3)$. Also, marginalization and Dempster's combination rule satisfy a vital property called the local computation property [28].

Local computation property Suppose m_1 is a BPA for s_1 and m_2 is a BPA for s_2. Suppose $X \in s_1$ and $X \notin s_2$. Then,

$$(m_1 \oplus m_2)^{\downarrow (s_1 \cup s_2) \setminus \{X\}} = (m_1)^{\downarrow s_1 \setminus \{X\}} \oplus m_2 \tag{11}$$

This property is the basis of computing marginals of joint belief functions. [10] describes an implementation of a local computation algorithm in Matlab called "Belief Function Machine" for computing the marginals of D-S belief function models.

2.3. Conditional independence

Shenoy [24] describes conditional independence relation in the framework of valuation-based systems using factorization semantics. Here, we describe it for the D-S theory of belief functions.

Definition 1 *(Conditional Independence (CI)).* Suppose $\mathcal{V}$ denotes the set of all variables, and suppose r, s, and t are disjoint subsets of $\mathcal{V}$. Suppose m is a joint BPA for $\mathcal{V}$. We say r and s are conditionally independent given t with respect to BPA m, written as $r \perp\!\!\!\perp_m s \mid t$ if and only if $m^{\downarrow r \cup s \cup t} = m_{r \cup t} \oplus m_{s \cup t}$, where $m_{r \cup t}$ is a BPA for $r \cup t$ and $m_{s \cup t}$ is a BPA for $s \cup t$, and $m_{r \cup t}$ and $m_{s \cup t}$ are distinct.

Some comments regarding Definition 1:

1. This definition generalizes the CI relation in probability theory [6].
2. The definition of CI in Definition 1 satisfies the 'graphoid' properties of probabilistic conditional independence [17]. See [24] for details.
3. Shenoy [26] discusses the semantics of CI in terms of no-double counting of non-idempotent knowledge.
4. There are other definitions of CI in the D-S theory [31,2,3], but they are not useful for belief-function graphical models.

2.4. Conditional belief functions

This subsection defines a conditional belief function similar to a conditional probability table in probability theory. The definition of a conditional belief function in this subsection is taken from [13].

Definition 2. Suppose r and s are disjoint subsets of variables, $r' \subseteq r$, and $m_{s|r'}$ is a BPA for $r' \cup s$. We say $m_{s|r'}$ is a conditional BPA for s given r' if and only if

1. $(m_{s|r'})^{\downarrow r'}$ is a vacuous BPA for r', and
2. for any BPA m_r for r, if m_r and $m_{s|r'}$ are distinct, then $m_r \oplus m_{s|r'}$ is a BPA for $r \cup s$.

We call s the *head* of the conditional $m_{s|r'}$, and r' the tail. In the second condition of Definition 2, we have BPA m_r for r and a conditional $m_{s|r'}$ for s given $r' \subseteq r$. If given r', s is conditionally independent (CI) of $r \setminus r'$, then in this case m_r and $m_{s|r'}$ are distinct. If the CI condition doesn't hold, then combining m_r and $m_{s|r'}$ may lead to double counting of non-idempotent knowledge and are therefore not distinct [27].

Subsection 2.5 discusses belief-function directed graphical models. In graphical models, the joint is constructed from the conditionals. We don't start with a joint. The definition of a conditional belief function in Definition 2 reflects this fact. Other definitions of conditional belief functions start from a joint and then factor the joint into a marginal and a conditional. These other definitions do not help in constructing graphical models. However, our definition is consistent with these other definitions for the joint implicitly defined by a graphical model [13].

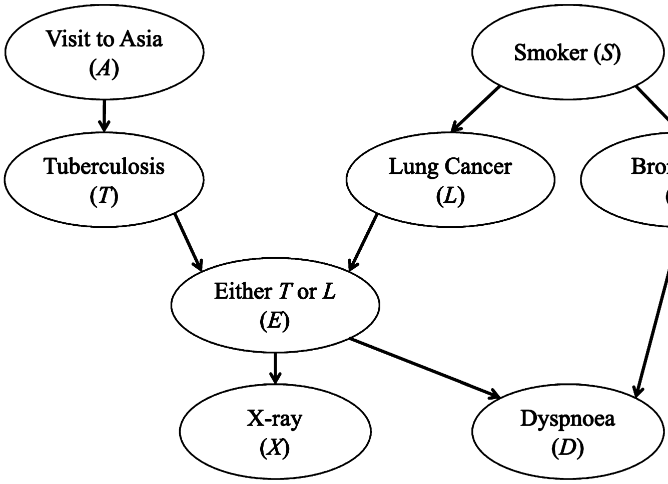
We will illustrate belief function conditionals using the Chest Clinic example [15].

Chest clinic Fig. 1 shows a Bayesian network. There are eight discrete binary variables; not all probabilities in the joint probability distribution are strictly positive. Fig. 1 also shows the conditional probability tables (CPTs).

Example 1 *(A deterministic belief function conditional).* In the chest clinic problem, consider the CPT $P(E|L,T)$ for E given (L,T) as shown in Fig. 1. All variables are discrete with two states. Thus, $\Omega_E = \{e, \bar{e}\}$, etc. Notice that this CPT is deterministic as all probabilities in the CPT are either 0 or 1. This CPT can be represented in the D-S theory by a deterministic BPA $m_{E|(L,T)}$ for $\{L,T,E\}$ as follows:

$$m_{E|(L,T)}(\{(l,t,e), (l,\bar{t},e), (\bar{l},t,e), (\bar{l},\bar{t},\bar{e})\}) = 1 \tag{12}$$

Notice that $m_{E|(L,T)}$ is a conditional for E given (L,T) because $(m_{E|(L,T)})^{\downarrow\{L,T\}}$ is a vacuous BPA for (L,T): $(m_{E|(L,T)})^{\downarrow\{L,T\}}(\Omega_{L,T}) = 1$. The other CPTs are all non-deterministic, and these will be encoded in the D-S theory using Smets' conditional embedding to be discussed next.



$P_A(a) = 0.01, P_A(\bar{a}) = 0.99$	$P_{E\|L,T}(l, t, e) = 1, P_{E\|L,T}(l, t, \bar{e}) = 0$
	$P_{E\|L,T}(l, \bar{t}, e) = 1, P_{E\|L,T}(l, \bar{t}, \bar{e}) = 0$
$P_S(s) = 0.50, P_A(\bar{s}) = 0.50$	$P_{E\|L,T}(\bar{l}, t, e) = 1, P_{E\|L,T}(\bar{l}, t, \bar{e}) = 0$
	$P_{E\|L,T}(\bar{l}, \bar{t}, e) = 0, P_{E\|L,T}(\bar{l}, \bar{t}, \bar{e}) = 1$
$P_{T\|A}(a, t) = 0.05, P_{T\|A}(a, \bar{t}) = 0.95$	$P_{X\|E}(e, x) = 0.98, P_{X\|E}(e\ \bar{x}) = 0.02$
$P_{T\|A}(\bar{a}, t) = 0.01, P_{T\|A}(\bar{a}, \bar{t}) = 0.99$	$P_{X\|E}(\bar{e}, x) = 0.05, P_{X\|E}(\bar{e}, \bar{x}) = 0.95$
$P_{L\|S}(s, l) = 0.10, P_{L\|S}(s, \bar{l}) = 0.90$	$P_{D\|E,B}(e, b, d) = 0.90, P_{D\|E,B}(e, b, \bar{d}) = 0.10$
$P_{L\|S}(\bar{s}, l) = 0.01, P_{L\|S}(\bar{s}, \bar{l}) = 0.99$	$P_{D\|E,B}(e, \bar{b}, d) = 0.70, P_{D\|E,B}(e, \bar{b}, \bar{d}) = 0.30$
$P_{B\|S}(s, b) = 0.80, P_{B\|S}(s, \bar{b}) = 0.20$	$P_{D\|E,B}(\bar{e}, b, d) = 0.80, P_{D\|E,B}(\bar{e}, b, \bar{d}) = 0.20$
$P_{B\|S}(\bar{s}, b) = 0.30, P_{B\|S}(\bar{s}, \bar{b}) = 0.70$	$P_{D\|E,B}(\bar{e}, \bar{b}, d) = 0.10, P_{D\|E,B}(\bar{e}, \bar{b}, \bar{d}) = 0.90$

Fig. 1. The directed acyclic graph and the CPTs for the Chest Clinic example.

Where do conditionals come from? A conditional BPA $m_{r|s}$ describes the relationship between the variables in r and s. One source of conditionals is Smets' conditional embedding [29]. To describe conditional embedding, consider the case of two variables, X and Y. To describe the dependency between X and Y, suppose that when $X = x$, our belief in Y is described by a BPA m_{Y_x} for Y.

The BPA m_{Y_x} for Y needs to be embedded into a BPA for $m_{Y|x}$ for (X, Y) such that

1. $m_{Y|x}$ is a conditional BPA for (X, Y), i.e., $m_{Y|x}^{\downarrow X}$ is the vacuous BPA for X, and
2. when we combine the belief that $X = x$ and marginalize the result to Y, we obtain m_{Y_x}.

One way to do this is to take each focal element $\mathsf{b} \in 2^{\Omega_Y}$ of m_Y and convert it to the corresponding focal element

$$(\{x\} \times \mathsf{b}) \cup ((\Omega_X \setminus \{x\}) \times \Omega_Y) \in 2^{\Omega_{X,Y}} \tag{13}$$

of BPA $m_{Y|x}$ for (X, Y) with the same mass. It is easy to confirm that this embedding method satisfies both conditions mentioned above.[1] Suppose we have several distinct conditionals, e.g., $m_{Y|x_1}$, $m_{Y|x_2}$, etc. obtained by conditional embedding, where x_1, and x_2 are distinct values of X. In this case, we combine the conditionals by Dempster's combination rule to obtain $m_{Y|X}$. An example of conditional embedding follows.

Example 2 *(Conditional embedding of a CPT).* Consider the CPT for T given A for the chest clinic example in Fig. 1. Given $A = a$, the condition probability distribution for T is $P(t|a) = 0.05$ and $P(\bar{t}|a) = 0.95$. This conditional distribution for T (given $A = a$) can be represented as a Bayesian BPA m_{T_a} for T as follows: $m_{T_a}(\{t\}) = 0.05$, $m_{T_a}(\{\bar{t}\}) = 0.95$. After conditional embedding, we have a BPA $m_{T|a}$ for (A, T) as follows:

$$m_{T|a}(\{(a, t), (\bar{a}, t), (\bar{a}, \bar{t})\}) = 0.05, m_{T|a}(\{(a, \bar{t}), (\bar{a}, t), (\bar{a}, \bar{t})\}) = 0.95 \tag{14}$$

The BPA $m_{T|a}$ for (A, T) is a conditional for T given A as $m_{T|a}^{\downarrow A}$ is a vacuous BPA for A: $m_{T|a}^{\downarrow A}(\Omega_A) = 1$.

[1] Smets [31] claims that the conditional embedding method satisfies the principle of minimal commitment: one should never give more support than is justified to any proposition.

Similarly, the conditional probability distribution for T given $A = \bar{a}$ can be embedded into a BPA $m_{T|\bar{a}}$ for (A, T) as follows:

$$m_{T|\bar{a}}(\{(a,t),(a,\bar{t}),(\bar{a},t)\}) = 0.01, m_{T|\bar{a}}(\{(a,t),(a,\bar{t}),(\bar{a},\bar{t})\}) = 0.99. \tag{15}$$

If we combine the two conditionals $m_{T|a}$ and $m_{T|\bar{a}}$ by Dempster's rule, we obtain the conditional BPA $m_{T|A} = m_{T|a} \oplus m_{T|\bar{a}}$ for T given A as follows:

$$\begin{aligned}
&m_{T|A}(\{(a,t),(\bar{a},t)\}) = P(t|a)\,P(t|\bar{a}) = 0.0005,\\
&m_{T|A}(\{(a,t),(\bar{a},\bar{t})\}) = P(t|a)\,P(\bar{t}|\bar{a}) = 0.0495,\\
&m_{T|A}(\{(a,\bar{t}),(\bar{a},t)\}) = P(\bar{t}|a)\,P(t|\bar{a}) = 0.0095,\\
&m_{T|A}(\{(a,\bar{t}),(\bar{a},\bar{t})\}) = P(\bar{t}|a)\,P(\bar{t}|\bar{a}) = 0.9405.
\end{aligned} \tag{16}$$

There is no conflict $(K = 1)$ in the combination $m_{T|a} \oplus m_{T|\bar{a}}$. Also, $m_{T|A}$ is not a Bayesian BPA for (A, T). $m_{T|A}$ has the following properties:

1. $(m_{T|A})^{\downarrow A}$ is the vacuous BPA for A;
2. If $m_{A=a}$ is a deterministic BPA for A such that $m_{A=a}(\{a\}) = 1$, then $(m_{A=a} \oplus m_{T|A})^{\downarrow T} = m_{T_a}$;
3. If $m_{A=\bar{a}}$ is a deterministic BPA for A such that $m_{A=\bar{a}}(\{\bar{a}\}) = 1$, then $(m_{A=\bar{a}} \oplus m_{T|A})^{\downarrow T} = m_{T_{\bar{a}}}$.

Thus, $m_{T|A}$ is a belief function analog of CPT $P(T|A)$.

2.5. BF directed graphical models

We start with some notation. A directed graph G is a pair $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{X_1, \ldots, X_n\}$ denotes the set of *nodes* and $\mathcal{E}$ denotes the set of *directed edges* (X_i, X_j) between two distinct variables in $\mathcal{V}$. For any node $X \in \mathcal{V}$, let $Pa_G(X)$ denote $\{Y \in \mathcal{V} : (Y, X) \in \mathcal{E}\}$. A directed graph is said to be *acyclic* if and only if there exists a sequence of the nodes of the graph, say $(X_1, \ldots, X_n)$ such that if there is a directed edge $(X_i, X_j) \in \mathcal{E}$ then X_i must precede X_j in the sequence. Such a sequence is called a *topological* sequence (as it depends only on the structure of the directed graph).

Definition 3 *(BF directed graphical model).* Suppose we have a directed acyclic graph $G = (\mathcal{V}, \mathcal{E})$ with n nodes in $\mathcal{V}$. A belief-function directed graphical model ($BFDGM$) is a pair $(G, \{m_1, \ldots, m_n\})$ such that BPA m_i associated with node X_i is a conditional BPA for X_i given $Pa_G(X_i)$, for $i = 1, \ldots, n$. A fundamental assumption of a BFDGM is that $m_1, \ldots, m_n$ are all distinct, and the joint BPA m for $\mathcal{V}$ associated with the model is given by

$$m = \bigoplus_{i=1}^{n} m_i. \tag{17}$$

Some comments about Definition 3:

1. The assumption in Definition 3 that all conditionals are distinct allows the combination in Eq. (17).
2. Given m, the joint BPA for $\mathcal{V}$ as defined in Eq. (17), it follows from Definition 1 that the following CI relations hold. Suppose $(X_1, \ldots, X_n)$ is a topological sequence associated with BFDGM $(G, \{m_1, \ldots, m_n\})$. Then for each X_i, $i = 2, \ldots, n$, $X_i \perp\!\!\!\perp_m (\{X_1, \ldots X_{i-1}\} \setminus Pa_G(X_i)) \mid Pa_G(X_i)$.
3. An example of a BFDGM is the Chest Clinic example discussed in Section 2.4.

2.6. Distinct belief functions

This material is taken from Shenoy [27]. Distinct belief functions are also called independent belief functions in the D-S literature.[2] Dempster's combination rule is only applicable to combining distinct BPAs. So, what are distinct BPAs? Dempster [7] provides a definition. Consider the multi-valued semantics of BPAs as shown in Fig. 2.

We have a probability mass function (PMF) $P(X_1)$ for X_1, a multivalued function $\Gamma_1 : \Omega_{X_1} \to 2^{S_1} \setminus \emptyset$ that defines the BPA m_1 for S_1. Similarly, we have a probability mass function (PMF) $P(X_2)$ for X_2, a multivalued function $\Gamma_2 : \Omega_{X_2} \to 2^{S_2} \setminus \emptyset$ that defines the BPA m_2 for S_2. m_1 and m_2 are distinct if and only if X_1 and X_2 are independent random variables, i.e., $P(X_1, X_2) = P(X_1) \otimes P(X_2)$, where $\otimes$ is the probabilistic combination operator, point-wise multiplication followed by normalization.

[2] The terminology of 'distinct' belief functions is due to Smets [30]. As independence is usually associated with random variables, we prefer the terminology of distinct belief functions.

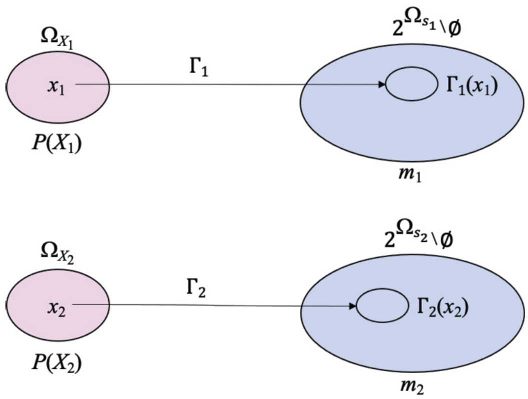


Fig. 2. Dempster's multi-valued semantics for BPAs.

Some comments about Dempster's definition.

1. As $P(X_1)$, and $P(X_2)$ are PMFs, and the two multi-valued mappings Γ_1 and Γ_2 map non-empty subsets of S_1 and S_2 respectively, it is clear that m_1 and m_2 are BPAs for S_1 and S_2, respectively.
2. In practice, not every belief function in a belief function model is associated with a multi-valued mapping. Thus, Dempster's definition cannot be used directly in practice.
3. We say BPA m is idempotent if $m \oplus m = m$. Idempotent knowledge is knowledge encoded in a BPA m that is idempotent. For example, if m is deterministic, then m is idempotent. Thus, double-counting idempotent knowledge is not a problem; double-counting non-idempotent knowledge is.
4. If we assume independence of variables X_1 and X_2 when they are not, then if we combine m_1 and m_2, we are double-counting non-idempotent knowledge. Thus, the spirit of Dempster's definition is that two belief functions are distinct if, when combining them using Dempster's combination rule, we are not double-counting non-idempotent knowledge. We will use this heuristic in discussing what constitutes distinct belief functions in practice.
5. The discussion of distinct belief functions is valid more broadly to many uncertainty calculi, including probability theory.

Using the no-double counting non-idempotent knowledge heuristic, we can argue that in the probabilistic graphical model $X \to Y$, potentials $P(X)$ and $P(Y|X)$ are always distinct [26,27]. Thus, $P(X, Y) = P(X) \otimes P(Y|X)$ is the joint PMF of (X, Y). There are no CI assumptions in this model.

If X and Y are independent with respect to the joint PMF $P(X, Y)$ represented by the graphical model with variables X and Y with no directed edges from X to Y or vice versa with conditionals $P(X)$ and $P(Y)$, then $P(X, Y) = P(X) \otimes P(Y|X) = P(X) \otimes P(Y)$ (as X and Y are independent, $P(Y|X)(x, y) = P(Y)(y)$). Thus, $P(X)$ and $P(Y)$ are distinct potentials.

Suppose X and Y are not independent, and the dependency of Y on X is described by conditional $P(Y \mid X)$. If we assume independence of X and Y, then $P(X) \otimes P(Y) = P(X) \otimes (P(X) \otimes P(Y|X))^{\downarrow Y}$ and we are double-counting the knowledge in $P(X)$ (assuming it is not idempotent). In this case, $P(X)$ and $P(Y)$ are not distinct potentials.

The concept of distinct belief functions in the D-S theory is similar to the discussion above in probability theory. Each belief-function directed graphical model is associated with a set of CI assumptions for the variables in the model. The definition of conditional independence in the D-S belief function theory is similar to that of probability theory [6,24]. Also, associated with each variable X in the model, we have a conditional for X given its parents. Unlike the Bayesian case, some conditionals may not be known. In this case, a vacuous BPA is associated with such variables [27]. If we have partial knowledge of the conditionals, we can represent it by a non-Bayesian BPA. As in the probabilistic case, assuming the CI relations are valid, the BPAs in the model are distinct.

3. Representing a BN by a belief function graphical model

This section will show that any BN can be represented as a directed graphical belief function model. The joint probability distribution defined in a BN is described as a Bayesian BPA in the corresponding D-S belief function model.

Theorem 1. *Consider a two-variable BN with variables X and Y with state spaces Ω_X and Ω_Y, respectively, as shown in Fig. 3. Let P_X and $P_{Y|X}$ denote a prior probability mass function (PMF) for X and a CPT for Y, respectively. Let $P_{X,Y}$ denote the joint PMF of (X, Y) such that $P_{X,Y}(x, y) = P_X(x)\, P_{Y|X}(x, y)$ for all $(x, y) \in \Omega_{X,Y}$. Let BPA m_X for X denote the Bayesian BPA for X corresponding to PMF P_X, and let BPA $m_{Y|X}$ for (X, Y) denote the conditional BPA for $Y|X$ obtained from the CPT $P_{Y|X}$ by Smets' conditional embedding. Then, $m_X \oplus m_{Y|X}$ is the Bayesian BPA for (X, Y) corresponding to PMF $P_{X,Y}$.*

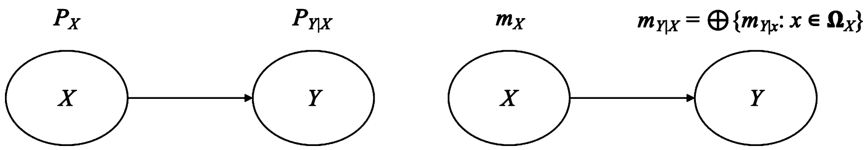


Fig. 3. A two-variable BN and a corresponding directed graphical belief function model.

Table 1
Details of $m_X \oplus m_{Y|x_1}$.

$m_{Y\|x_1} \oplus m_X$	$\{x_1\} \times \Omega_Y$ $P(x_1)$	$\{x_2\} \times \Omega_Y$ $P(x_2)$	$\ldots$ $\ldots$	$\{x_m\} \times \Omega_Y$ $P(x_m)$
$\{(x_1, y_1), \{x_2, \ldots x_m\} \times \Omega_Y$ $P(y_1\|x_1)$	$\{(x_1, y_1)\}$ $P(x_1, y_1)$	$\{x_2\} \times \Omega_Y$ $P(x_2)\, P(y_1\|x_1)$	$\ldots$ $\ldots$	$\{x_m\} \times \Omega_Y$ $P(x_m)\, P(y_1\|x_1)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\{(x_1, y_n), \{x_2, \ldots x_m\} \times \Omega_Y$ $P(y_n\|x_1)$	$\{(x_1, y_n)\}$ $P(x_1, y_n)$	$\{x_2\} \times \Omega_Y$ $P(x_2)\, P(y_n\|x_1)$	$\ldots$ $\ldots$	$\{x_m\} \times \Omega_Y$ $P(x_m)\, P(y_n\|x_1)$

Proof. This is a folk theorem: well-known (see, e.g., [1]), but no proof appears to be published anywhere. Therefore, we include a proof here. We will assume $|\Omega_X| = m$, and $|\Omega_Y| = n$. Let $\Omega_X = \{x_1, \ldots, x_m\}$, and let $\Omega_Y = \{y_1, \ldots, y_n\}$. Each conditional $P(Y|x_i)$ is represented as a conditional $m_{Y|x_i}$ using Smets' conditional embedding. Thus, we have m conditionals $m_{Y|x_i}$, one for each $x_i \in \Omega_X$.

We start by Dempster's combination $m_X \oplus m_{Y|x_1}$. As the domain of m_X is X and the domain of $m_{Y|x_1}$ is (X, Y), we need to vacuously extend m_X to (X, Y). Let $m_X^{\uparrow\{X,Y\}}$ denote such a vacuous extension. The focal elements of $m_X^{\uparrow\{X,Y\}}$ are $\{x_i\} \times \Omega_Y$ for $i = 1, \ldots, m$, with BPA values $P(X = x_i)$ Notice that after vacuous extension, $m_X^{\uparrow\{X,Y\}}$ is no longer a Bayesian BPA.

Now consider $m_{Y|x_1}$. Prior to conditional embedding, we have a Bayesian BPA $m_{Y_{x_1}}$ for Y whose focal elements are $\{y_1\}, \ldots, \{y_n\}$ with BPA values $P(y_1|x_1), \ldots, P(y_n|x_1)$. After conditional embedding, the focal elements of $m_{Y|x_1}$ are $\{(x_1, y_1), \{x_2, \ldots, x_m\} \times \Omega_Y\}$, …, $\{(x_1, y_n), \{x_2, \ldots, x_m\} \times \Omega_Y)\}$ (with BPA values $P(y_1|x_1), \ldots, P(y_n|x_1)$).

Next, we combine m_X and $m_{Y|x_1}$ using Dempster's rule. The details are shown in Table 1. Notice that some of the focal elements of $m_X \oplus m_{Y|x_1}$ are singleton subsets (in the second column of Table 1) whose BPA values are the same as the joint distribution of $P_{X,Y}$.

Next, we do a Dempster's combination of $(m_X \oplus m_{Y|x_1}) \oplus m_{Y|x_2}$. The singleton focal elements $\{(x_1, y_j)\}$ for $j = 1, \ldots, n$ remain singletons with values $P(x_1, y_j)\, P(y_j|x_2)$, and these values sum to $P(x_1, y_j)$. Also, the focal elements $\{x_2\} \times \Omega_Y$ of $m_X \oplus m_{Y|x_1}$ (in the third column of Table 1) become singleton subsets $\{(x_2, y_j)\}$ whose values sum to $P(x_2, y_j)$, the same as the joint distribution of $P(X, Y)$.

If we continue in this manner until we have combined all conditionals $m_{Y|x_i}$ for $i = 1, \ldots, m$, the joint BPA $m_X \oplus m_{Y|X}$ will be a Bayesian BPA whose values are the same as the Bayesian joint distribution $P_{X,Y}$.

What if Y is independent of X? This is a special case where the conditional distributions $P(Y|x_1) = \ldots = P(Y|x_m) = P(Y)$. Our proof remains valid for this case as the only fact about the conditionals we use in the proof is that $\sum_{j=1}^{n} P(Y = y_j|x_i) = 1$ for all $i = 1, \ldots, m$. Of course, this case can be handled without the use of conditionals $P(Y = y_j \mid x_i)$ by just using Bayesian BPA m_Y corresponding to $P(Y)$ without the need for Smets' conditional embedding. As for $P(X)$, we need to vacuously extend m_Y to (X, Y), and the theorem is true. □

Theorem 1 can be generalized to a complete BN with more than two variables.

Theorem 2. *Suppose we have a BN with n variables $X_1, \ldots X_n$. Suppose each CPT $P(X_i|pa(X_i))$ in the BN is represented as a conditional BPA $m_{X_i|pa(X_i)}$ obtained by conditional embedding of CPT $P(X_i|pa(X_i))$. Then, $\oplus_{i=1}^{n} m_{X_i|pa(X_i)}$ is a Bayesian BPA corresponding to $\otimes_{i=1}^{n} P_{X_i|pa(X_i)}$. Here, $\otimes$ denotes the probabilistic combination operator, pointwise multiplication followed by normalization [28].*

Proof. Without loss of generality, let $(X_1, \ldots, X_n)$ denote a topological sequence associated with the directed acyclic graph of the BN. First, consider the sub-BN consisting of only variables $\{X_1, X_2\}$. It follows from Theorem 1 that the joint BPA for (X_1, X_2) is Bayesian corresponding to $\otimes_{i=1}^{2} P_{X_i|pa(X_i)}$. Next, consider the sub-BN corresponding to $(X_1, \ldots, X_3)$. We can consider this as a BN with two variables (X_{12}, X_3), where X_{12} is a composite variable whose state space is Ω_{X_1,X_2} with a prior distribution $P(X_{12}) = P(X_1) \otimes P(X_2|X_1)$. Again, it follows from Theorem 1 that the joint BPA $\oplus_{i=1}^{3} m_{X_i|pa(X_i)}$ is a Bayesian BPA corresponding to $\otimes_{i=1}^{3} P_{X_i|pa(X_i)}$. Proceeding in this manner, it follows from induction that the result is true. □

Example 3 *(A BN consisting of A and T).* Consider the BN consisting of variables A and T with the corresponding conditionals as shown in Fig. 1. Let m_A denote the Bayesian BPA for A corresponding to P_A and let $m_{T|A}$ denote the conditional BPA

Table 2
The computation of $m_{A,T}$ in Example 3.

	$m_A \oplus m_{T|A}$	$m_A^{\uparrow\{A,T\}}$ $\{(a,t),(a,\bar{t})\}$ 0.01	$\{(\bar{a},t),(\bar{a},\bar{t})\}$ 0.99
$m_{T|A}$			
	$\{(a,t),(\bar{a},t)\}$ 0.0005	$\{(a,t)\}$ 0.000005	$(\{\bar{a},t)\}$ 0.000495
	$\{(a,t),(\bar{a},\bar{t})\}$ 0.0495	$\{(a,t)\}$ 0.000495	$(\{\bar{a},\bar{t})\}$ 0.049005
	$\{(a,\bar{t}),(\bar{a},t)\}$ 0.0095	$\{(a,\bar{t})\}$ 0.000095	$(\{\bar{a},t)\}$ 0.009405
	$\{(a,\bar{t}),(\bar{a},\bar{t})\}$ 0.9405	$\{(a,\bar{t})\}$ 0.009405	$(\{\bar{a},\bar{t})\}$ 0.931095

Table 3
The marginal BPA m_T for T in Example 4.

2^{Ω_T}	m_T	Bel_T	Pl_T
$\emptyset$			
$\{t\}$	0.0005	0.0005	0.0595
$\{\bar{t}\}$	0.9405	0.9405	0.9995
$\{t,\bar{t}\}$	0.0590	1	1

for $T|A$ after Smets' conditional embedding of $P_{T|A}$ as shown in Example 2. The computation of the joint BPA for (A,T) is shown in Table 2.

Thus, in this complete model for A and T, the joint BPA $m_{A,T}$ for $\{A,T\}$ is as follows:

$$m_{A,T}(\{(a,t)\}) = 0.000005 + 0.000495 = 0.0005$$
$$m_{A,T}(\{(a,\bar{t})\}) = 0.000095 + 0.009405 = 0.0095$$
$$m_{A,T}(\{(\bar{a},t)\}) = 0.000495 + 0.009405 = 0.0099$$
$$m_{A,T}(\{(\bar{a},\bar{t})\}) = 0.049005 + 0.931095 = 0.9801.$$

Notice that there is no conflict, i.e., $(K = 1)$, in the combination $m_A \oplus m_{T|A}$ as $m_{T|A}$ is a conditional, i.e., $m_{T|A}^{\downarrow T}$ is vacuous.

The joint $m_{A,T}$ is Bayesian. Its marginal for T is as follows: $m_T(\{t\}) = 0.0005 + 0.0099 = 0.0104$, and $m_T(\{\bar{t}\}) = 0.0095 + 0.9801 = 0.9896$, which is the same as in the Bayesian solution of this problem.

4. Making inferences from an incomplete BN

In this section, we examine the case of an incomplete BN where we are missing some priors, some conditionals, or both. In a corresponding belief function model, if we are missing the prior for a variable, we can omit the prior or use a vacuous BPA to represent the missing information. We do the same if we are missing a conditional for a variable.

Example 4 *(A BN model with a missing prior).* Consider the BN from Example 3 in which we are missing the prior for A. If we embed this incomplete BN to a corresponding D-S belief function model, we have just the conditional BPA $m_{T|A}$, which from Example 2 is as follows:

$$m_{T|A}(\{(a,t),(\bar{a},t)\}) = 0.0005,$$
$$m_{T|A}(\{(a,t),(\bar{a},\bar{t})\}) = 0.0495,$$
$$m_{T|A}(\{(a,\bar{t}),(\bar{a},t)\}) = 0.0095,$$
$$m_{T|A}(\{(a,\bar{t}),(\bar{a},\bar{t})\}) = 0.9405. \tag{18}$$

The conditional BPA $m_{T|A}$ for $\{A,T\}$ is not Bayesian. We can compute the marginal $m_T = (m_{T|A})^{\downarrow T}$ as shown in Table 3. The corresponding belief and plausibility functions are also shown in Table 3. Thus, the bounds on $P(t)$, the probability that a patient suffers from Tuberculosis, are as follows: $0.0005 \le P(t) \le 0.0595$. Similarly the bounds on $P(\bar{t})$ are: $0.9405 \le P(\bar{t}) \le 0.9995$. Notice that the bounds include the point estimates ($P(t) = 0.0104$, and $P(\bar{t}) = 0.9896$) from the complete model in Example 3.

If we model the missing prior for A by the Laplacian (equally likely) Bayesian distribution $P(a) = 0.5$, $P(\bar{a}) = 0.5$, and find the marginal for T, we get the point estimates $P(t) = 0.03$ and $P(\bar{t}) = 0.97$. This, of course, masks the uncertainty in the posterior marginal distribution for T.

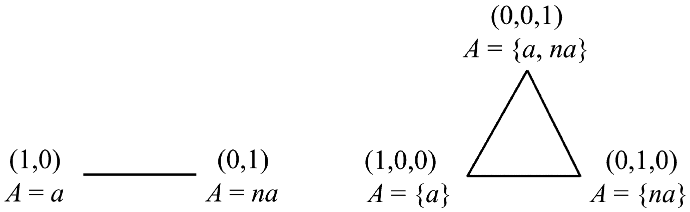


Fig. 4. Left: The possible values of a Bayesian prior for A. Right: The possible values for a prior BPA for A.

A question that arises is as follows: If we do a sensitivity analysis on a Bayesian analysis of this problem with $P(A) = (p, 1-p)$, where $p = P(A = a)$ and $1 - p = P(A = \bar{a})$, we vary $p \in [0, 1]$, the marginal posterior probability of $T = t$ varies from 0.01 when $p = 1$ to 0.05 when $p = 0$. The bounds on $P(T = t)$ from a D-S belief function model are much wider—from 0.0005 to 0.0595. Why?

1. A Bayesian sensitivity analysis is restricted to a one-dimensional region: $\{(p, 1-p, 0) : 0 \le p \le 1\}$ (see Fig. 4). Consider a BFDGM corresponding to an incomplete BN with a missing prior. A BFDGM is said to be *complete* if the BPA corresponding to the missing prior is Bayesian. If we assume *complete* knowledge of the prior of A, denoted by $\{(p, 1-p, 0) : 0 \le p \le 1\}$ (where $p = m_A(\{a\})$, $1 - p = m_A(\{\bar{a}\})$, and $0 = m_A(\{a, \bar{a}\})$, then the marginal $m_T = (m_A \oplus m_{T|A})^{\downarrow T}$ is a Bayesian BPA representing $P(t) = p\,0.05 + (1-p)\,0.01$ and $P(\bar{t}) = p\,0.95 + (1-p)\,0.99$. Under the *complete* knowledge assumption, varying $0 \le p \le 1$, we get a range of point estimates for $P(t)$, not intervals. A Bayesian model can only reason with complete information [16, Ch. 9, p. 415]. Taking the range of point estimates and constructing an interval is ad-hoc and may not be justified by the laws of probability.
2. A D-S belief-function analysis is more expressive than a Bayesian sensitivity analysis discussed in Comment 1. If we are completely ignorant of the prior distribution of A, we can model it using the vacuous BPA for A, ι_A (denoted by the point $(0, 0, 1)$ in Fig. 4). BPA ι_A for A corresponds to the credal set Cr_{ι_A} which consists of *all* PMFs of A. A D-S analysis of the model where we have BPA ι_A as the prior for A, and conditional BPA $m_{T|A}$ for (A, T) results in the interval $[0.0005, 0.0595]$ for $P(T = t)$ as discussed in Example 4 with width 0.0590. The wider interval here (compared to the Bayesian analysis discussed in Comment 1 where we get the point estimates) is a consequence of the ignorance of the prior for A. There is no ignorance in a Bayesian prior for A. Unlike the Bayesian case (discussed in Comment 1 above), there is no need for an ad-hoc sensitivity analysis. It is inappropriate to compare the interval $[0.01, 0.05]$ for $P(t)$ obtained from a Bayesian meta-analysis with the interval $[0.0005, 0.0595]$ obtained from a regular D-S analysis. A Bayesian analysis yields point estimates of $P(t)$ regardless of the value of $p = P(A = a)$ as there is no ignorance in the prior for A. A D-S analysis using the vacuous BPA for A results in an interval because of ignorance of the prior of A. There is no anomaly in these results.
3. A D-S analysis can also model partial knowledge of priors. Consider the point $(m_A(\{a\}), m_A(\{\bar{a}\}), m_A(\{a, \bar{a}\})) = (0.3, 0.3, 0.4)$ representing the knowledge that $0.3 \le P(A = a) \le 0.7$ and $0.3 \le P(A = \bar{a}) \le 0.7$. A D-S analysis results in the posterior marginal for T as follows:

   $$(m_T(\{t\}), m_T(\{\bar{t}\}), m_T(\{t, \bar{t}\}) = (0.0182, 0.9582, 0.0236).$$

   This can be interpreted as follows: posterior probability $0.0182 \le P(T = t) \le 0.0418$. Again, there is no need for a sensitivity analysis. See Table 5 for the posterior intervals and widths for some sample values of m_A.
4. The width of the probability interval for $T = t$ varies continuously from 0.0590 (for $(0, 0, 1)$) to 0 (for $(p, q, 0)$ as a function of $r = m_A(a, \bar{a})$.
5. The following lemma describes the probability intervals for t and $\bar{t}$ and their widths in a general 2-variable BN with binary variables.

Lemma 1 *(A non-Bayesian prior). Consider a BFDGM with two binary variables: A with $\Omega_A = \{a, \bar{a}\}$, and T with $\Omega_T = \{t, \bar{t}\}$. Assume we have a non-Bayesian prior BPA m_A for A with parameters $p = m_A(\{a\})$, $q = m_A(\{\bar{a}\})$, and $r = m_A(\{a, \bar{a}\})$, $p + q + r = 1$. Assume we have a conditional BPA $m_{T|A}$ for T given A that is derived from a Bayesian CPT for T with parameters $P(t|a)$, $P(\bar{t}|a)$, $P(t|\bar{a})$ and $P(\bar{t}|\bar{a})$, $P(t|a) + P(\bar{t}|a) = 1$, and $P(t|\bar{a}) + P(\bar{t}|\bar{a}) = 1$. For this model, the marginal BPA $m_T = (m_A \oplus m_{T|A})^{\downarrow T}$ is as follows:*

$$m_T(\{t\}) = p\,P(t|a) + q\,P(t|\bar{a}) + r\,P(t|a)\,P(t|\bar{a}),$$

$$m_T(\{\bar{t}\}) = p\,P(\bar{t}|a) + q\,P(\bar{t}|\bar{a}) + r\,P(\bar{t}|a)\,P(\bar{t}|\bar{a}),$$

$$m_T(\{t, \bar{t}\}) = r\,(P(t|a)\,P(\bar{t}|\bar{a}) + P(\bar{t}|a)\,P(t|\bar{a}))$$

Thus, the probability interval for $T = t$ is

$$[m_T(\{t\}), m_T(\{t\}) + m_T(\{t, \bar{t}\})],$$

Table 4
The computation of $m_A \oplus m_{T|A}$ in Lemma 1.

		$m_A^{\uparrow\{A,T\}}$		
	$m_A \oplus m_{T\mid A}$	$\{(a,t),(a,\bar{t})\}$ p	$\{(\bar{a},t),(\bar{a},\bar{t})\}$ q	$\Omega_{A,T}$ r
$m_{T\mid A}$				
	$\{(a,t),(\bar{a},t)\}$ $P(t\mid a)P(t\mid\bar{a})$	$\{(a,t)\}$ $p\,P(t\mid a)P(t\mid\bar{a})$	$\{(\bar{a},t)\}$ $q\,P(t\mid a)P(t\mid\bar{a})$	$\{(a,t),(\bar{a},t)\}$ $r\,P(t\mid a)P(t\mid\bar{a})$
	$\{(a,t),(\bar{a},\bar{t})\}$ $P(t\mid a)P(\bar{t}\mid\bar{a})$	$\{(a,t)\}$ $p\,P(t\mid a)P(\bar{t}\mid\bar{a})$	$\{(\bar{a},\bar{t})\}$ $q\,P(t\mid a)P(\bar{t}\mid\bar{a})$	$\{(a,t),(\bar{a},\bar{t})\}$ $r\,P(t\mid a)P(\bar{t}\mid\bar{a})$
	$\{(a,\bar{t}),(\bar{a},t)\}$ $P(\bar{t}\mid a)P(t\mid\bar{a})$	$\{(a,\bar{t})\}$ $pP(\bar{t}\mid a)P(t\mid\bar{a})$	$\{(\bar{a},t)\}$ $q\,P(\bar{t}\mid a)P(t\mid\bar{a})$	$\{(a,\bar{t}),(\bar{a},t)\}$ $r\,P(\bar{t}\mid a)P(t\mid\bar{a})$
	$\{(a,\bar{t}),(\bar{a},\bar{t})\}$ $P(\bar{t}\mid a)P(\bar{t}\mid\bar{a})$	$\{(a,\bar{t})\}$ $p\,P(\bar{t}\mid a)P(\bar{t}\mid\bar{a})$	$\{(\bar{a},\bar{t})\}$ $q\,P(\bar{t}\mid a)P(\bar{t}\mid\bar{a})$	$\{(a,\bar{t}),(\bar{a},\bar{t})\}$ $rP(\bar{t}\mid a)P(\bar{t}\mid\bar{a})$

Table 5
Posterior intervals and widths for T for some sample values of m_A.

m_A	Interval $[Bel_T(\{t\}), Pl_T(\{t\})]$	Width of interval $Pl_T(\{t\}) - Bel_T(\{t\})$
$(1,0,0)$	$[0.01, 0.01]$	0
$(0,1,0)$	$[0.05, 0.05]$	0
$(0,0,1)$	$[0.0005, 0.0595]$	0.0590
$(0.3, 0.3, 0.4)$	$[0.0182, 0.0418]$	0.0236
$(0.3, 0.6, 0.1)$	$[0.02105, 0.02695]$	0.0059
$(0.5, 0.4, 0.1)$	$[0.02905, 0.03495]$	0.0059

Table 6
The computation of $m_{A,T}$ in Example 5.

		$m_A^{\uparrow\{A,T\}}$	
	$m_A \oplus m_{T\mid a}$	$\{(a,t),(a,\bar{t})\}$ 0.01	$\{(\bar{a},t),(\bar{a},\bar{t})\}$ 0.99
$m_{T\mid a}$			
	$\{(a,t),(\bar{a},t),(\bar{a},\bar{t})\}$ 0.05	$\{(a,t)\}$ 0.0005	$\{(\bar{a},t),(\bar{a},\bar{t})\}$ 0.0495
	$\{(a,\bar{t}),(\bar{a},t),(\bar{a},\bar{t})\}$ 0.95	$\{(a,\bar{t})\}$ 0.0095	$(\{\bar{a},t),(\bar{a},\bar{t})\}$ 0.9405

with width $m_T(\{t,\bar{t}\})$ that depends only on $r = m_A(\{a,\bar{a}\})$ and not on p or q. Similarly, the probability interval for $T = \bar{t}$ is

$$[m_T(\{\bar{t}\}), m_T(\{\bar{t}\}) + m_T(\{t,\bar{t}\})],$$

with width $m_T(\{t,\bar{t}\})$ that depends only on $r = m_A(\{a,\bar{a}\})$ and not on p or q.

Proof. Dempster's combination $m_A \oplus m_{T|A} = m_{A,T}$ is shown in Table 4. Of the 12 cells, five have only t's, five have only $\bar{t}$'s, and two have both t and $\bar{t}$. When we compute the marginal $m_T = (m_{A,T})^{\downarrow T}$, $m_T(\{t\})$ is the sum of the masses associated with the five cells that have only t's. Similarly for $m_T(\{\bar{t}\})$, and $m_T(\{t,\bar{t}\})$. Thus, the result follows. □

Example 5 *(A BN model with a missing conditional).* Consider the BN from Example 3 in which we are missing a conditional for T given $A = \bar{a}$. If we embed this incomplete BN to a corresponding D-S belief function model, we have just two BPAs: m_A and $m_{T|a}$, the conditional for T given $A = a$. The belief function computations are shown in Table 6.

The joint BPA $m_{A,T}$ for $\{A,T\}$ is as follows:

$$m_{A,T}(\{(a,t)\}) = 0.0005$$
$$m_{A,T}(\{(a,\bar{t})\}) = 0.0095$$
$$m_{A,T}(\{(\bar{a},t),(\bar{a},\bar{t})\}) = 0.9900$$

The joint BPA $m_{A,T}$ for $\{A,T\}$ is not Bayesian. We can compute the marginal $m_T = (m_{A,T})^{\downarrow T}$ as shown in Table 7. The corresponding belief and plausibility functions are also shown in Table 7. Thus, the bounds on $P(t)$, the posterior probability that a patient suffers from Tuberculosis, are as follows: $0.0005 \le P(t) \le 0.9905$. Similarly the bounds on $P(\bar{t})$ are: $0.0095 \le P(\bar{t}) \le 0.9995$. Notice that the bounds include the point estimates ($P(t) = 0.0104$, and $P(\bar{t}) = 0.9896$) from the complete model in Example 3.

Table 7
The marginal BPA m_T for T in Example 5.

2^{Ω_T}	m_T	Bel_T	Pl_T
$\emptyset$			
$\{t\}$	0.0005	0.0005	0.9905
$\{\bar{t}\}$	0.0095	0.0095	0.9995
$\{t, \bar{t}\}$	0.9900	1	1

If we model the missing conditional for T given $A = \bar{a}$ by the Laplacian (equally likely) conditional distribution $P(t|\bar{a}) = 0.5$, $P(\bar{t}|\bar{a}) = 0.5$, and find the marginal for T, we get point estimates $P(t) = 0.4955$ and $P(\bar{t}) = 0.5045$. This, of course, masks the uncertainty in the posterior marginal distribution for T.

Some comments:

1. The posterior marginal BPA m_T for T has mass 0.9900 for the focal element $\{t, \bar{t}\}$, which is the value $m_A(\bar{a})$ of $\bar{a}$ for which we are missing a conditional distribution for T.
2. The width of the interval for posterior probability $P(T = t)$ and $P(T = \bar{t})$ is the same $(= 0.99 = m_T(\{t, \bar{t}\}))$.
3. A belief-function model allows for partial knowledge of conditionals. For example, suppose in Example 5, when $A = \bar{a}$, the conditional for T is not missing, but as follows: $m_{T_{\bar{a}}}(\{t\}) = 0.01$, $m_{T_{\bar{a}}}(\{\bar{t}\}) = 0.90$, $m_{T_{\bar{a}}}(\{t, \bar{t}\}) = 0.09$. Thus, $0.01 \le P(T = t|\bar{a}) \le 0.10$, and $0.90 \le P(T = \bar{t}|\bar{a}) \le 0.99$. We use Smets' conditional embedding to create a conditional BPA for T given A and combine all three BPAs to find the marginal BPA for T as follows: $m_T(\{t\}) = 0.0104$, $m_T(\{\bar{t}\}) = 0.9005$, and $m_T(\{t, \bar{t}\}) = 0.0891$. Thus, the posterior marginal for T is: $0.0104 \le P(T = t) \le 0.0995$, and $0.9005 \le P(T = \bar{t}) \le 0.9896$. The width of both these intervals is $0.0891 = m_A(\{\bar{a}\}) \cdot m_{T_{\bar{a}}}(\{t, \bar{t}\})$ = $0.99 \cdot 0.09$, which is a consequence of Dempster's product-intersection combination $m_A \oplus m_{T|A}$. A general result is as follows.

Lemma 2 *(A non-Bayesian conditional). Consider a BN with two binary variables: A with state space $\Omega_A = \{a, \bar{a}\}$, and T with state space $\Omega_T = \{t, \bar{t}\}$. Suppose we have a Bayesian prior for A represented by BPA $m_A(\{a\}) = P(a)$, and $m_A(\{\bar{a}\}) = P(\bar{a})$, where $P(a) + P(\bar{a}) = 1$. When $A = a$, we have a Bayesian BPA m_{T_a} for T as follows: $m_{T_a}(\{t\}) = P(t|a)$, and $m_{T_a}(\{\bar{t}\}) = P(\bar{t}|a)$ where $P(t|a) + P(\bar{t}|a) = 1$. When $A = \bar{a}$, suppose we have a non-Bayesian BPA $m_{T_{\bar{a}}}$ for T as follows: $m_{T_{\bar{a}}}(\{t\}) = p$, $m_{T_{\bar{a}}}(\{\bar{t}\}) = q$, $m_{T_{\bar{a}}}(\Omega_T) = r$, where $p + q + r = 1$. For this model, the marginal BPA for T is non-Bayesian. The probability of $T = t$ is in the interval*

$$[P(t|a)\,P(a) + p\,P(\bar{a}),\, P(t|a)\,P(a) + p\,P(\bar{a}) + r\,P(\bar{a})].$$

Thus, the width of the interval is $r\,P(\bar{a})$, where $r = m_{T_{\bar{a}}}(\Omega_T)$, and $P(\bar{a})$ is the probability of the state of A for which we have a non-Bayesian conditional.

Proof. After Smets' conditional embedding of m_{T_a} and $m_{T_{\bar{a}}}$, we have:

$$m_{T|a}(\{(a, t), (\bar{a}, t), (\bar{a}, \bar{t})\}) = P(t|a),$$
$$m_{T|a}(\{(a, \bar{t}), (\bar{a}, t), (\bar{a}, \bar{t})\}) = P(\bar{t}|a),$$

and

$$m_{T|\bar{a}}(\{(a, t), (a, \bar{t}), (\bar{a}, t)\}) = p,$$
$$m_{T|\bar{a}}(\{(a, t), (a, \bar{t}), (\bar{a}, \bar{t})\}) = q,$$
$$m_{T|\bar{a}}(\Omega_{A,T}) = r.$$

Let $m_{T|A} = m_{T|a} \oplus m_{T|\bar{a}}$. Then $m_{T|A}$ is as follows:

$$m_{T|A}(\{(a, t), (\bar{a}, t)\}) = P(t|a)\,p,$$
$$m_{T|A}(\{(a, t), (\bar{a}, \bar{t})\}) = P(t|a)\,q,$$
$$m_{T|A}(\{(a, t), (\bar{a}, t)), (\bar{a}, \bar{t})\}) = P(t|a)\,r,$$
$$m_{T|A}(\{(a, \bar{t}), (\bar{a}, t)\}) = P(\bar{t}|a)\,p,$$
$$m_{T|A}(\{(a, \bar{t}), (\bar{a}, \bar{t})\}) = P(\bar{t}|a)\,q,$$
$$m_{T|A}(\{(a, \bar{t}), (\bar{a}, t), (\bar{a}, \bar{t})\}) = P(\bar{t}|a)\,r.$$

Next, we compute $m_{A,T} = m_A \oplus m_{T|A}$. $m_{A,T}$ is as follows:

$$m_{A,T}(\{(a,t)\}) = P(t|a)\,P(a),$$
$$m_{A,T}(\{(a,\bar{t})\}) = P(\bar{t}|a)\,P(a),$$
$$m_{A,T}(\{(\bar{a},t)\}) = p\,P(\bar{a}),$$
$$m_{A,T}(\{(\bar{a},\bar{t})\}) = q\,P(\bar{a}),$$
$$m_{A,T}(\{(\bar{a},t),(\bar{a},\bar{t})\}) = r\,P(\bar{a}).$$

The marginal $m_T = (m_{A,T})^{\downarrow T}$ of the joint is as follows:

$$m_T(\{t\}) = P(t|a)\,P(a) + p\,P(\bar{a}),$$
$$m_T(\{\bar{t}\}) = P(\bar{t}|a)\,P(a) + q\,P(\bar{a}),$$
$$m_T(\{t,\bar{t}\}) = r\,P(\bar{a}).$$

Thus, the result follows. □

5. Belief function machine

Belief function machine (BFM) [10] is a set of routines in MatLab for constructing belief function graphical models and computing marginals of the joint for variables of interest using local computation. It was written in 2002 by Phan Hong Giang under the supervision of Thierry Denoeux, Prakash Shenoy, and Philippe Smets. It was further developed in 2003 by Sushila Shenoy by incorporating features described in [11] to enable the solution of large belief function models.

Some features of BFM are as follows. Belief function models are constructed using UIL (unified input language). UIL is based on the framework of the valuation-based system [23]. It can find marginals of belief function models containing hundreds of variables.

A UIL text file for the chest clinic example is as follows:

```
# Variables and their state space
DEFINE VARIABLE A {a ~a}; # Visit to Asia
DEFINE VARIABLE S {s ~s}; # Smoker
DEFINE VARIABLE T {t ~t}; # Tuberculosis
DEFINE VARIABLE L {l ~l}; # Lung Cancer
DEFINE VARIABLE B {b ~b}; # Bronchitis
DEFINE VARIABLE E {e ~e}; # Either T or L
DEFINE VARIABLE X {x ~x}; # X-ray
DEFINE VARIABLE D {d ~d}; # Dyspnoea

# CPT for T given A
DEFINE CONDITIONAL RELATION TA {T} GIVEN {A};
SET CONDITIONAL VALUATION TA GIVEN {a}{(~t)} 0.95 {(t)} 0.05;
SET CONDITIONAL VALUATION TA GIVEN {~a}{(t)} 0.01 {(~t)} 0.99;

# CPT for L given S
DEFINE CONDITIONAL RELATION LS {L} GIVEN {S};
SET CONDITIONAL VALUATION LS GIVEN {s}{(~l)} 0.9 {(l)} 0.1;
SET CONDITIONAL VALUATION LS GIVEN {~s}{(l)} 0.01 {(~l) 0.99;

# CPT for B given S
DEFINE CONDITIONAL RELATION BS {B} GIVEN {S};
SET CONDITIONAL VALUATION BS GIVEN {s}{(~b)} 0.4 {(b)} 0.6;
SET CONDITIONAL VALUATION BS GIVEN {~s}{(~b)} 0.7 {(b)} 0.3;

# Deterministic conditional for E given T, L
DEFINE RELATION ETL {E T L};
SET VALUATION ETL {(~e ~t ~l)(e ~t l)(e t ~l)(e t l)} 1.0;

# CPT for X given E
DEFINE CONDITIONAL RELATION XE {X} GIVEN {E};
SET CONDITIONAL VALUATION XE GIVEN {e} {(x)} 0.98 {(~ x)} 0.02;
SET CONDITIONAL VALUATION XE GIVEN {~e} {(x)} 0.05 {(~ x)} 0.95;
```

```
# CPT for D given B, E
DEFINE CONDITIONAL RELATION DBE {D} GIVEN {B E};
SET CONDITIONAL VALUATION DBE GIVEN {b e} {(d)} 0.9 {(~d)} 0.1;
SET CONDITIONAL VALUATION DBE GIVEN {b ~e} {(d)} 0.7 {(~d)} 0.3;
SET CONDITIONAL VALUATION DBE GIVEN {~b e} {(d)} 0.8 {(~ d)} 0.2;
SET CONDITIONAL VALUATION DBE GIVEN {~b ~e} {(d)} 0.1, {(~d)} 0.9;

#Prior distribution for S
DEFINE RELATION PRIORS {S};
SET VALUATION PRIORS {(~s)} 0.5 {(s)} 0.5;

#Prior distribution for A
DEFINE RELATION PRIORA {A};
SET VALUATION PRIORA {(a)} 0.01 {(~a)} 0.99;
```

Some comments.

1. All text to the right of # are comments. In the first set of statements, we define all variables by specifying their names and all possible states within curly parenthesis; a space separates each state. Thus, e.g., variable A has two states: a and $\sim a$.
2. In the subsequent statements, we define belief functions. Unconditional BPAs are described in two steps. In the first step, we have a DEFINE RELATION statement with the name of the BPA and its domain. In the second step, we have a SET VALUATION statement that specifies the details of the BPA (focal elements and values). Thus the deterministic conditional for E given T, L is defined as an unconditional deterministic belief function, whose name is ETL, and whose domain is $\{E, T, L\}$. This belief function has one focal element within curly parenthesis followed by its value. If there is more than one focal element, these are separated by space as in the definition of the prior distribution for A and S.
3. Conditional belief functions are specified using the DEFINE CONDITIONAL RELATION statement that has the name of the belief function followed by the head of the conditional and then the tail. In BNs, the head is a singleton variable, but the tail consisting of the parents of the head can have several variables. The details of the conditional belief function are specified for each state (or set of states) of the tail variables. Thus the conditional distribution of T given $A = a$, $P_{T|a}(t) = 0.95$, and $P_{T|a}(\sim t) = 0.05$ is specified by the SET CONDITIONAL VALUATION statement:

   ```
   SET CONDITIONAL VALUATION TA GIVEN {a}{(~t)} 0.95, {(t)} 0.05;
   ```

   Specifying a conditional after conditional embedding is unnecessary, as BFM has a command to do it. This is illustrated in Subsection 5.1.

5.1. Using BFM

BFM consists of a set of routines that run in MatLab. Suppose the UIL file for the chest clinic example is saved as a text file named ‘cclinic.txt.’ A short script for computing the marginal belief function for T is shown below. BFM can also build a binary join tree [25], propagate belief functions, and compute the marginals for every variable in a model.

```
uil2bm('cclinic.txt', 'bmcclinic')
# translate UIL file into BFM format

global BELIEF VARIABLE ATTRIBUTE STRUCTURE FRAME QUERY
global BELTRACE NODE BJTREE TRANSPROTOCOL
# declaration of global variables

load bmcclinic
# load file containing BFM data structure

belall = condiembed([BELIEF(:).number])
# convert all conditional valuations into conditional belief functions using Smet's conditional embedding, leaving the unconditional valuations unchanged.

keepbel(belall)
# discard unnecessary belief functions from memory.
```

Table 8
The marginal BPAs m_T for T, m_L for L, and m_B for B, in the complete chest clinic example. Empty cells have 0 values.

2^{Ω_T}	m_T	2^{Ω_L}	m_L	2^{Ω_B}	m_B
∅		∅		∅	
$\{t\}$	0.0104	$\{l\}$	0.0550	$\{b\}$	0.5500
$\{\sim t\}$	0.9896	$\{\sim l\}$	0.9450	$\{\sim b\}$	0.4500
$\{t, \sim t\}$		$\{l, \sim l\}$		$\{b, \sim b\}$	

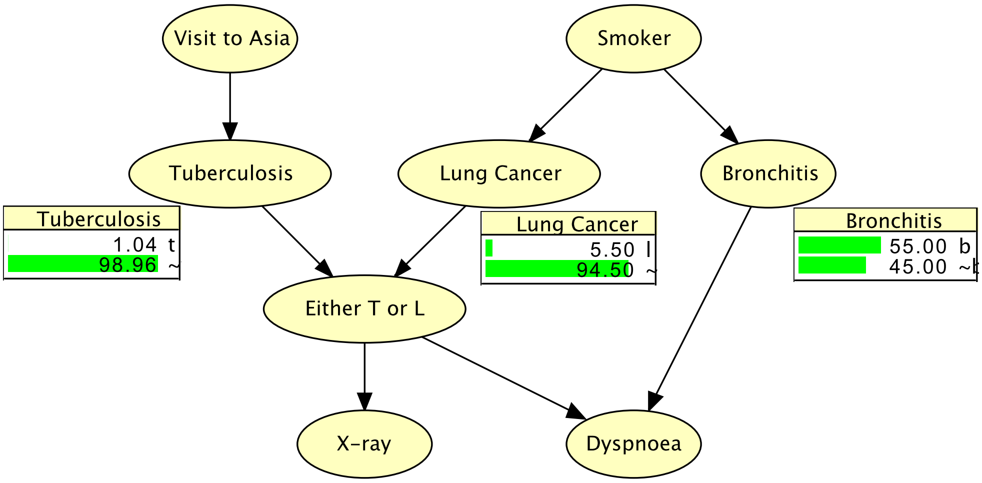


Fig. 5. The results from a Bayesian propagation for the complete chest clinic example.

showbel(solve(belall, 'T'))
# Find the marginal BPA of the joint for T and display its details.

bjtbuild('T', belall)
# build a binary join tree with T as the final node.

showbel(solvetreeall(1));
# Find the marginal for every variable using binary join tree BJTREE(1) created by the previous command and display the details.

The marginals for variables T, B, and L resulting from the last command are shown in Table 8. Fig. 5 shows the results from a Bayesian propagation in Hugin. Notice that the results are the same in both cases.

Now assume that we have an incomplete BN where we are missing the following pieces of information:

1. prior for A,
2. prior for S,
3. conditional distribution $P_{T|\sim a}$,
4. conditional distribution $P_{X|\sim e}$, and
5. conditional distribution $P_{D|e,b}$.

We comment out the corresponding SET VALUATION and SET CONDITIONAL VALUATION statements in the UIL text file and re-run the script. The results are shown in Table 9. Notice that the marginals are not Bayesian. Thus,

$$0 \leq P_T(t) \leq 1, 0.001 \leq P_L(l) \leq 0.109, 0.18 \leq P_B(b) \leq 0.72,$$

$$0 \leq P_T(\bar{t}) \leq 1, 0.891 \leq P_L(\bar{l}) \leq 0.999, 0.28 \leq P_B(\bar{b}) \leq 0.82$$

If we insert Laplacian priors/conditionals for the missing information and do a Bayesian propagation, we get the point estimates of the marginals as shown in Fig. 6. These point estimates lie in the intervals from a belief function analysis. A belief function model with interval estimates of the marginals is more useful for decision-making [8] than the point estimates from a Bayesian propagation that doesn't provide *any* information about the uncertainty of the estimates.

Table 9
The marginal BPAs m_T for T, m_L for L, and m_B for B in the incomplete chest clinic example. Empty cells have 0 values.

2^{Ω_T}	m_T	2^{Ω_L}	m_L	2^{Ω_B}	m_B
$\emptyset$		$\emptyset$		$\emptyset$	
$\{t\}$		$\{l\}$	0.0010	$\{b\}$	0.1800
$\{\sim t\}$		$\{\sim l\}$	0.8910	$\{\sim b\}$	0.2800
$\{t, \sim t\}$	1.0	$\{l, \sim l\}$	0.1080	$\{b, \sim b\}$	0.5400

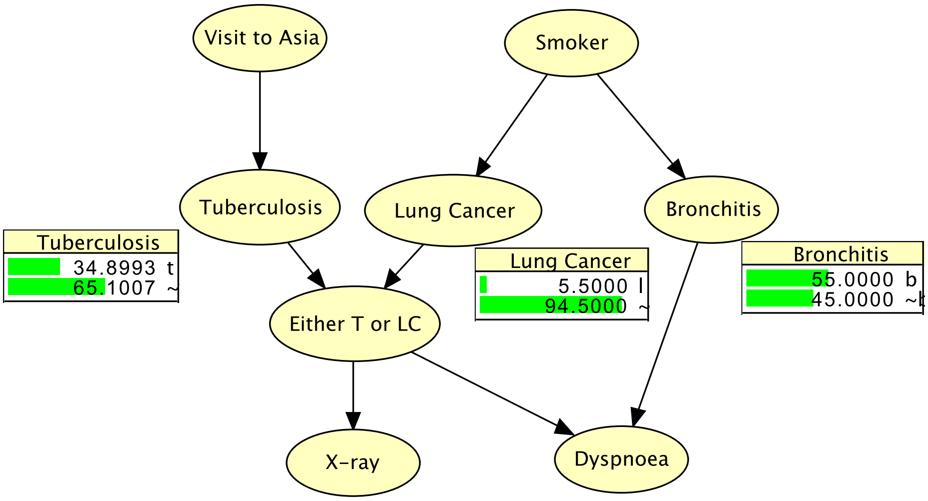


Fig. 6. The results from a Bayesian propagation for the incomplete chest clinic example.

6. Summary & conclusions

This article discusses the challenges of working with incomplete Bayesian Networks (BNs), where certain priors and conditional distributions are missing or partially known. When the parameters of a BN are estimated from data, some rare scenarios may not occur in the dataset. Similarly, if domain experts estimate the parameters, they may feel uncomfortable providing estimates for certain variables due to their lack of experience.

To address this issue, our proposed method is to embed an incomplete Bayesian network model in a corresponding D-S belief function model and then reason from such a model. D-S models are a generalization of Bayesian reasoning. We can express ignorance of missing information using vacuous belief functions in the D-S theory. We can also express partial knowledge of priors and conditional using belief functions. There is no analog of a vacuous belief function or partial information in a Bayesian representation. Consequently, Bayesian analysis cannot reason with incomplete/partial information [16, Ch. 9, p. 415]. The best we can do is a sensitivity analysis where we vary the parameters of a missing prior or conditional (assuming "complete" knowledge modeled as a Bayesian BPA). Such an analysis cannot replicate the results obtained from a belief-function analysis (without the necessity of sensitivity analysis). Also, such a sensitivity analysis (assuming complete knowledge) results in a confidence interval with a narrower width than a corresponding D-S analysis, which may be misleading. Finally, if we have several missing priors/conditionals, a sensitivity analysis is not a computationally tractable task.

Dempster's combination rule for aggregating evidence is more computationally demanding than Bayesian aggregation. For large BN models, the marginals of the joint probability distribution can be computed without explicitly computing the joint distribution using local computation. The same is true for the D-S theory, which satisfies the same axioms on which local computation is based [28]. There are implementations of local computation algorithms for the D-S theory. One such implementation is described in this paper.

We do not claim that the method described in this article is the only method for making inferences in incomplete Bayesian networks. For example, in the imprecise probability literature, there is an alternative theory of belief functions that use credal set semantics of belief functions, which considers belief functions as a set of PMFs such that the lower bound of the probability of a proposition a is equal to the value of Bel(a) for all propositions a $\subseteq \Omega_r$. Credal set semantics are incompatible with Dempster's combination rule [19–21,12].

Fagin and Halpern [9] propose another rule for updating beliefs, referred to as the Fagin-Halpern combination rule. Suppose we start with a credal set of PMFs characterized by BPA m for r and observe some proposition a $\subseteq \Omega_r$. In that case, one possible updating rule is to condition each PMF in the credal set on proposition a and then find the revised BPA m' corresponding to the lower envelope of the revised set of PMFs. The Fagin-Halpern rule does precisely this and differs from Dempster's rule of conditioning, a special case of Dempster's combination rule. It would be interesting to compare the two distinct theories of belief functions. Like the D-S theory, the Fagin-Halpern theory is also a generalization of probability theory. However, we must determine if the Fagin-Halpern combination rule satisfies the local computation property. Also, we must learn how to represent conditional knowledge as conditional belief functions (similar to Smets' conditional embedding). Finally, we must find out if the Fagin-Halpern belief function theory is implemented to compute the marginals of large belief function graphical models. Comparing these two belief function theories for making inferences from incomplete Bayesian networks is an important task that needs to be done.

Another alternative to using D-S theory is to use Cozman [4]'s credal networks. There are local computation algorithms for credal networks [5] and some implementations [32]. Comparison of credal networks with the D-S theory is another important task that needs to be done.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This article was written in response to an invitation from Antonio Salmerón and Rafael Rumí to serve as an invited speaker for the 11th International Conference on Probabilistic Graphical Models, PGM-2022, in Almeria, Spain, October 5–7, 2022. I am grateful to Antonio and Rafael for the invitation. I am also grateful to three reviewers of this journal who provided critical and constructive comments that have improved the content and exposition of this paper. The Ronald G. Harper Professorship supported this study at the University of Kansas School of Business.

References

[1] R.G. Almond, Graphical Belief Modeling, Chapman & Hall, London, UK, 1995.
[2] B. Ben-Yaghlane, P. Smets, K. Mellouli, Belief function independence: I. The marginal case, Int. J. Approx. Reason. 29 (2002) 47–70.
[3] B. Ben-Yaghlane, P. Smets, K. Mellouli, Belief function independence: II. The conditional case, Int. J. Approx. Reason. 31 (2002) 31–75.
[4] F.G. Cozman, Credal networks, Artif. Intell. 120 (2000) 199–233.
[5] F.G. Cozman, C.P. Campos, Local computation in credal networks, in: Proceedings of the Workshop on Local Computation for Logics and Uncertainty, ECAI, 2004, pp. 5–11.
[6] P. Dawid, Conditional independence in statistical theory, J. R. Stat. Soc., Ser. B 41 (1979) 1–15.
[7] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat. 38 (1967) 325–339.
[8] T. Denoeux, P.P. Shenoy, An interval-valued utility theory for decision making with Dempster-Shafer belief functions, Int. J. Approx. Reason. 124 (2020) 194–216.
[9] R. Fagin, J.Y. Halpern, A new approach to updating beliefs, in: P. Bonissone, M. Henrion, L. Kanal, J. Lemmer (Eds.), Uncertainty in Artificial Intelligence, vol. 6, North-Holland, 1991, pp. 347–374.
[10] P. Giang, S. Shenoy, The belief function machine: an environment for reasoning with belief functions in Matlab, Working Paper, University of Kansas School of Business, Lawrence, 2003, KS 66045, https://pshenoy.ku.edu/Papers/BFM072503.zip.
[11] R. Haenni, N. Lehmann, Resource bounded and anytime approximation of belief function computation, Int. J. Approx. Reason. 31 (2002) 103–154.
[12] J.Y. Halpern, R. Fagin, Two views of belief: belief as generalized probability and belief as evidence, Artif. Intell. 54 (1992) 275–317.
[13] R. Jiroušek, V. Kratochvíl, P.P. Shenoy, On conditional belief functions in directed graphical models the Dempster-Shafer theory, Working Paper 341, University of Kansas School of Business, Lawrence, 2022, KS 66045, https://pshenoy.ku.edu/Papers/WP341.pdf.
[14] S.L. Lauritzen, F.V. Jensen, Local computation with valuations from a commutative semigroup, Ann. Math. Artif. Intell. 21 (1997) 51–69.
[15] S.L. Lauritzen, D. Spiegelhalter, Local computations with probabilities on graphical structures and their application to expert systems, J. R. Stat. Soc., Ser. B 50 (1988) 157–224.
[16] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Mateo, CA, 1988.
[17] J. Pearl, A. Paz, Graphoids: graph-based logic for reasoning about relevance relations, in: B.D. Boulay, D. Hogg, L. Steele (Eds.), Advances in Artificial Intelligence II, North-Holland, Amsterdam, 1987, pp. 357–363.
[18] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
[19] G. Shafer, Constructive probability, Synthese 48 (1981) 1–60.
[20] G. Shafer, Perspectives on the theory and practice of belief functions, Int. J. Approx. Reason. 4 (1990) 323–362.
[21] G. Shafer, Rejoinders to comments on "Perspectives on the theory and practice of belief functions", Int. J. Approx. Reason. 6 (1992) 445–480.
[22] G. Shafer, P.P. Shenoy, Probability propagation, Ann. Math. Artif. Intell. 2 (1990) 327–352.
[23] P.P. Shenoy, A valuation-based language for expert systems, Int. J. Approx. Reason. 3 (1989) 383–411.
[24] P.P. Shenoy, Conditional independence in valuation-based systems, Int. J. Approx. Reason. 10 (1994) 203–234.
[25] P.P. Shenoy, Binary join trees for computing marginals in the Shenoy-Shafer architecture, Int. J. Approx. Reason. 17 (1997) 239–263.
[26] P.P. Shenoy, No double counting semantics for conditional independence, in: F.G. Cozman, R. Nau, T. Seidenfeld (Eds.), Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications (ISIPTA-05), Society for Imprecise Probabilities and Their Applications, 2005, pp. 306–314, http://hdl.handle.net/1808/519.
[27] P.P. Shenoy, On distinct belief functions in the Dempster-Shafer theory, in: E. Miranda, I. Montes, E. Quaeghebeur, B. Vantaggi (Eds.), Proceedings of ISIPTA-23, in: Proceedings of Machine Learning Research (PMLR), Vol. 215, ML Research Press, 2023, in press, https://pshenoy.ku.edu/Papers/ISIPTA23a.pdf.

[28] P.P. Shenoy, G. Shafer, Axioms for probability and belief-function propagation, in: R.D. Shachter, T. Levitt, J.F. Lemmer, L.N. Kanal (Eds.), Uncertainty in Artificial Intelligence 4, in: Machine Intelligence and Pattern Recognition Series, vol. 9, North-Holland, Amsterdam, Netherlands, 1990, pp. 169–198.
[29] P. Smets, Un modele mathematico-statistique simulant le processus du diagnostic medical, Ph.D. thesis, Free University of Brussels, 1978.
[30] P. Smets, The concept of distinct evidence, in: B. Bouchon-Meunier (Ed.), Proceedings of the 4th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'92), 1992, pp. 789–794.
[31] P. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, Int. J. Approx. Reason. 9 (1993) 1–35.
[32] S. Tolo, E. Patelli, M. Beer, An open toolbox for the reduction, inference computation and sensitivity analysis of credal networks, Adv. Eng. Softw. 115 (2018) 126–148.