On Distinct Belief Functions in the Dempster-Shafer Theory

Prakash P. Shenoy PSHENOY@KU.EDU
University of Kansas School of Business, Lawrence, KS 66045 USA

Abstract

Dempster's combination rule is the centerpiece of the Dempster-Shafer (D-S) theory of belief functions. In practice, Dempster's combination rule should only be applied to combine two distinct belief functions (in the belief function literature, distinct belief functions are also called independent belief functions). So, the question arises: what constitutes distinct belief functions? We have an answer in Dempster's multi-valued functions semantics for distinct belief functions. The probability functions on the two spaces associated with the multi-valued functions should be independent. In practice, however, we don't always associate a multi-valued function with belief functions in a model. In this article, we discuss the notion of distinct belief functions in graphical models, both directed and undirected. The idea of distinct belief functions corresponds to no double-counting of non-idempotent knowledge semantics of conditional independence. Although we discuss the notion of distinct belief functions in the context of the DS theory, the discussion is valid more broadly to many uncertainty calculi, including probability theory, possibility theory, and Spohn's epistemic belief theory.

Keywords: distinct belief functions, Dempster-Shafer belief function theory, belief-function directed graphical model, belief-function undirected graphical model

1. Introduction

The centerpiece of the Dempster-Shafer (DS) theory of belief functions is Dempster's combination rule [6, 22]. In practice, Dempster's combination rule should only be applied to combine two "distinct" belief functions. So, the question arises: what constitutes distinct belief functions[1]? We have an answer in Dempster [6]'s multi-valued functions semantics for belief functions. The probability functions on the two spaces that are the domains of the multi-valued functions (of the two belief functions) should be independent. In practice, however, we don't always associate a multi-valued function with every belief function in a belief function model. In this article, we discuss the notion of distinct belief functions in graphical models, both directed and undirected. The idea of distinct belief functions corresponds to no double-counting of non-idempotent knowledge semantics of conditional independence [26]. Although we discuss the notion of distinct belief functions in the context of the DS theory, the discussion is valid more broadly to many uncertainty calculi, including probability theory, possibility theory, and Spohn's epistemic belief theory.

One of the earliest to discuss the notion of distinct belief functions is Shafer [23][2]. There is no formal definition of distinct belief functions, and the discussion is about combining non-distinct belief functions. Shafer advocates sorting out the common knowledge among two non-distinct pieces of evidence by refining the state spaces of the pieces instead of seeking generalizations of Dempster's combination rule to combine non-distinct evidence.

Smets [30] discusses Dempster's combination rule as a (matrix) multiplication of two matrices called specializations. Given a specialization representation of a piece of evidence, say a basic probability assignment (BPA) m_A for X, he defines a canonical factorization of the matrix m_A into $Q_m \cdot \varDelta_A \cdot Q_m^{-1}$, where Q_m is a matrix consisting of 0's and 1's that converts a BPA into a corresponding commonality function (CF) Q_m, and $\varDelta_A$ is a diagonal matrix whose values are the CF values of m_A. If the matrix representations of two pieces of evidence, say $\varDelta_A$ and $\varDelta_B$, includes a common matrix m_0 that is vacuous, then m_A and m_B are defined to be distinct. m_0 is referred to as a *correlation* matrix. If m_0 is not vacuous, then m_A and m_B are non-distinct. The idea of distinct evidence is the same as in [23]. He writes: "The problem of recognizing distinctness become essentially a problem of acknowledging that there is a vacuous correlation ... It can not be achieved by only comparing m_A and m_B."

Several studies propose to deal with combining non-distinct evidence by modifying Dempster's combination rule by making some assumptions about the nature of the non-distinctness of the pieces of evidence being combined [33, 8, 19, 7, 20, 9, 4]. Like Shafer, we agree that sorting the dependence among pieces of evidence is a better strategy for combining non-distinct evidence than modifying Dempster's rule. Otherwise, we would need a meta-rule to

[1]The concept of distinct belief functions is also referred to as independent belief functions in the literature. The terminology of *distinct* belief functions is due to Smets [30]. As independence is usually associated with variables and not functions, we prefer the terminology of distinct belief functions.

[2][23] was published (almost verbatim) as [24]

decide which variant of Dempster's rule should be used to combine non-distinct evidence.

The main goal of this article is to discuss the notion of distinct belief functions, especially in belief-function graphical models, both directed and undirected. We start with the definition stated by Dempster [6] in his multi-valued semantics of a BPA. We provide heuristics suggested by Dempster's definition for determining whether the belief functions in a graphical model are distinct. Two or more belief functions are distinct if there is no double-counting of non-idempotent knowledge. In graphical models, this implies that the set of belief functions in a graphical model are distinct only if the conditional independence conditions implied by the factorization of the joint belief function are valid.

An outline of the remainder of the paper is as follows. Section 2 reviews the basics of D-S theory, including basic probability assignments and commonality functions, marginalization and Dempster's combination rule, conditional belief functions, the removal operator, conditional independence relations, and graphical models. Section 3 has Dempster [6]'s formal definition and a discussion of distinct belief functions in the context of directed and undirected belief function models. Finally, Section 4 concludes with a summary and comments on further work.

2. Basics of D-S theory of Belief Functions

This section sketches the basics of the D-S theory of belief functions [6, 22].

Representations We represent knowledge using basic probability assignments, belief functions, plausibility functions, and commonality functions. Here, we only define basic probability assignments and commonality functions.

Notation Let $\mathcal{V}$ denote the set of all variables. Let X, Y, Z, etc., denote elements of $\mathcal{V}$. Let r, s, t, v, 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. $\emptyset$ denotes the empty set.

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

$$\begin{aligned} m(\emptyset) &= 0, \text{ and} & (1) \\ \sum_{\emptyset \neq \mathsf{a} \subseteq \Omega_s} m(\mathsf{a}) &= 1. & (2) \end{aligned}$$

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 a 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*[3]. If the focal element of a deterministic BPA is Ω_s, we say m is *vacuous*. Sometimes, we denote the vacuous BPA for s by ι_s.

Commonality Function The *commonality function* (CF) Q_m corresponding to BPA m for s is such that for all $\mathsf{a} \subseteq \Omega_s$,

$$Q_m(\mathsf{a}) = \sum_{\mathsf{b} \supseteq \mathsf{a}} m(\mathsf{b}). \quad (3)$$

Some comments about the definition of Q_m in Eq. (3):

1. $Q_m(\mathsf{a})$ represents the probability mass that could move to every state in a.
2. It follows from Eq. (3) that $0 \leq Q_m(\mathsf{a}) \leq 1$.
3. It follows from Eqs. (1)–(2) that $Q_m(\emptyset) = 1$.
4. CFs are non-increasing in the sense that if $\mathsf{a} \subseteq \mathsf{b}$, then $Q(\mathsf{a}) \geq Q(\mathsf{b})$.
5. A CF has the same information as in a BPA. Given a CF Q for s, let m_Q denote the corresponding BPA. We can recover m_Q from Q as follows [22].

$$m_Q(\mathsf{a}) = \sum_{\mathsf{b} \subseteq \Omega_s : \mathsf{b} \supseteq \mathsf{a}} (-1)^{|\mathsf{b} \setminus \mathsf{a}|} Q(\mathsf{b}). \quad (4)$$

6. Thus, it follows that $Q : 2^{\Omega_s} \to [0, 1]$ is a well-defined CF iff for all $\emptyset \neq \mathsf{a} \subseteq \Omega_s$

$$\begin{aligned} Q(\emptyset) &= 1, & (5) \\ \sum_{\mathsf{b} \subseteq \Omega_s : \mathsf{b} \supseteq \mathsf{a}} (-1)^{|\mathsf{b} \setminus \mathsf{a}|} Q(\mathsf{b}) &\geq 0, \quad \text{and} & (6) \\ \sum_{\emptyset \neq \mathsf{a} \subseteq \Omega_s} (-1)^{|\mathsf{a}|+1} Q(\mathsf{a}) &= 1. & (7) \end{aligned}$$

The left-hand side of Eq. (6) is $m_Q(\mathsf{a})$, and the left-hand side of Eq. (7) can be shown to be $\sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_s}} m_Q(\mathsf{a})$. Eq. (7) can be regarded as a normalization condition for a CF. If we have a function $Q : 2^{\Omega_s} \to [0, 1]$ that satisfies Eqs. (5) and (6), but not (7), then we can divide each of the values of the function for non-empty subsets in 2^{Ω_s} by $K = \sum_{\emptyset \neq \mathsf{a} \subseteq \Omega_s} (-1)^{|\mathsf{a}|+1} Q_m(\mathsf{a})$, and the resulting function will then qualify as a CF.

[3] Deterministic BPAs are also called *categorical* or *logical* in the D-S literature.

7. In some cases, we could have a CF that doesn't satisfy Eq. (6) but satisfies Eqs. (5) and (7). In such cases, we call such CFs pseudo-CFs. If we convert a pseudo-CF to a BPA using Eq. (4), then such a BPA will have negative masses that add to 1. We will call such BPAs pseudo-BPAs. Pseudo-CFs have been studied in [16, 17].
8. For the vacuous BPA ι_s for s, the CF Q_{ι_s} corresponding to BPA ι_s is given by $Q_{\iota_s}(\mathsf{a}) = 1$ for all $\mathsf{a} \subseteq \Omega_s$.
9. If m is a Bayesian BPA for s, then Q_m is such that $Q_m(\mathsf{a}) = m(\mathsf{a})$ if $|\mathsf{a}| = 1$, and $Q_m(\mathsf{a}) = 0$ if $|\mathsf{a}| > 1$.

Inference Operators There are three basic inference operators in the D-S theory—marginalization, combination, and removal. The marginalization operator allows us to coarsen knowledge by removing variables. The combination operator enables us to combine distinct knowledge. The removal operator is an inverse of the combination operator and allows us to remove a marginal from 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,

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

Definition 1 (Marginalization) *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 $\mathsf{a} \subseteq \Omega_t$,*

$$m^{\downarrow t}(\mathsf{a}) = \sum_{\mathsf{b} \subseteq \Omega_s : \mathsf{b}^{\downarrow t} = \mathsf{a}} m(\mathsf{b}). \tag{8}$$

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

$$(m^{\downarrow s \setminus \{X_1\}})^{\downarrow s \setminus \{X_1, X_2\}} = (m^{\downarrow s \setminus \{X_2\}})^{\downarrow s \setminus \{X_1, X_2\}}. \tag{9}$$

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

Definition 2 (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[4]*. Then, $m_1 \oplus m_2$ is a BPA for $s_1 \cup s_2$ such that for all $\mathsf{a} \subseteq \Omega_{s_1 \cup s_2} = (\mathsf{a}_1 \times \Omega_{s_2 \setminus s_1}) \cap (\mathsf{a}_2 \times \Omega_{s_1 \setminus s_2}) : \mathsf{a} \neq \emptyset$ where $\mathsf{a}_1 \subseteq \Omega_{s_1}$ and $\mathsf{a}_2 \subseteq \Omega_{s_2}$,*

$$(m_1 \oplus m_2)(\mathsf{a}) = K^{-1} \sum_{\mathsf{a}_1, \mathsf{a}_2 : \mathsf{a} \neq \emptyset} m_1(\mathsf{a}_1)\, m_2(\mathsf{a}_2), \tag{10}$$

where K is a normalization constant given by

$$K = \sum_{\mathsf{a}_1, \mathsf{a}_2 : \mathsf{a} \neq \emptyset} m_1(\mathsf{a}_1)\, m_2(\mathsf{a}_2). \tag{11}$$

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.

Dempster's combination rule can also be described using commonality functions. Consider two distinct BPAs m_1 for s_1 and m_2 for s_2, and let Q_1 and Q_2 denote the corresponding commonality functions. Then, as showed in [22], for all $\emptyset \neq \mathsf{a} \subseteq \Omega_{s_1 \cup s_2}$,

$$(Q_1 \oplus Q_2)(\mathsf{a}) = K^{-1} Q_1(\mathsf{a}^{\downarrow s_1})\, Q_2(\mathsf{a}^{\downarrow s_2}), \tag{12}$$

where K is a normalization constant defined as follows:

$$K = \sum_{\emptyset \neq \mathsf{a} \in \Omega_{s_1 \cup s_2}} (-1)^{|\mathsf{a}|+1} Q_1(\mathsf{a}^{\downarrow s_1})\, Q_2(\mathsf{a}^{\downarrow s_2}). \tag{13}$$

The normalization constant in Eq. (13) is precisely the same as in Eq. (11).

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,

$$(m_1 \oplus m_2)^{\downarrow (s_1 \cup s_2) \setminus \{X\}} = (m_1)^{\downarrow s_1 \setminus \{X\}} \oplus m_2 \tag{14}$$

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

The removal operator is discussed in Subsection 2.3.

2.1. Conditional Independence

Shenoy [25] 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.

[4]The notion of distinct BPAs is discussed in Section 3. Intuitively, m_1 and m_2 are distinct if combination of m_1 and m_2 doesn't result in double-counting of non-idempotent knowledge.

Definition 3 (Conditional independence) *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 $r \cup t$, $m_{s \cup t}$ is a BPA for $s \cup t$, and $m_{r \cup t}$ and $m_{s \cup t}$ are distinct.*

This definition generalizes the CI relation in probability theory [5]. There are other definitions of conditional independence in the D-S theory (e.g., [31, 2, 3]) using the semantics of non-interactivity. Still, these are not useful in describing CI in belief-function graphical models.

The definition of CI in Def. 3 satisfies the graphoid properties of probabilistic conditional independence [21]. Specifically, suppose m is a BPA for $\mathcal{V}$, and r, s, t, v are disjoint subsets of $\mathcal{V}$.

1. $r \perp\!\!\!\perp_m s \mid t$ if and only if $s \perp\!\!\!\perp_m r \mid t$ (symmetry).
2. If $r \perp\!\!\!\perp_m (s \cup v) \mid t$, then $r \perp\!\!\!\perp_m s \mid t$ (decomposition).
3. If $r \perp\!\!\!\perp_m (s \cup v) \mid t$, then $r \perp\!\!\!\perp_m s \mid (t \cup v)$ (weak union).
4. If $r \perp\!\!\!\perp_m s \mid t$ and $r \perp\!\!\!\perp_m v \mid (t \cup s)$, then $r \perp\!\!\!\perp_m (s \cup v) \mid t$ (contraction).
5. If m is such that $Q_m(\mathsf{a}) > 0$ for all $a \subseteq \Omega_{\mathcal{V}}$, then $r \perp\!\!\!\perp_m s \mid (t \cup v)$ and $r \perp\!\!\!\perp_m v \mid (t \cup s)$, then $r \perp\!\!\!\perp_m (s \cup v) \mid t$ (intersection).

Proofs of these properties can be found in [25].

2.2. 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 [14].

Definition 4 (Conditionals) *Suppose r and s are disjoint subsets of variables and suppose $r' \subseteq r$. Suppose $m_{s|r'}$ 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, m_r and $m_{s|r'}$ are distinct[5]. Thus, $m_r \oplus m_{s|r'}$ is a BPA for $r \cup s$.*

We call s the head *of the conditional, and r the* tail.

In a directed graphical belief function model, we have a conditional associated with each variable X. The head of the conditional is X, and the tail consists of the parents of X. For variables with no parents, we have priors associated with such variables. For convenience, priors can be regarded as conditionals with empty tails. For such BPAs, the first condition in the definition is trivially true as the sum of the probability masses in a BPA is 1.

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 Def. 4 reflects this fact. Other definitions of conditional belief functions start from a joint and then factor the joint into a marginal and a conditional (see, e.g.,[1]). These other definitions do not help in constructing graphical models. Our definition, however, is consistent with these other definitions for the joint that a graphical belief function model implicitly defines [14].

Non-informative BPAs The notion of non-informative BPAs is taken from [13].

Definition 5 (Non-informative belief functions) *Suppose m_1 is a BPA for r_1 and m_2 is a BPA for r_2. We say m_1 and m_2 are mutually non-informative if $m_1^{\downarrow (r_1 \cap r_2)} = m_2^{\downarrow (r_1 \cap r_2)} = \iota_{r_1 \cap r_2}$. Also, given a set of BPAs, the set of BPA is non-informative if every pair of BPAs in the set are mutually non-informative.*

Some comments about non-informative belief functions:

- Suppose BPA m_1 for r_1 and m_2 for r_2 are mutually non-informative. Then, m_1 can be regarded as a conditional for $r_1 \setminus (r_1 \cap r_2)$ given $r_1 \cap r_2$, and m_2 can be regarded as a conditional for $r_2 \setminus (r_1 \cap r_2)$ given $r_1 \cap r_2$.
- Notice that if $r_1 \cap r_2 = \emptyset$, then m_1 and m_2 are mutually non-informative.

We will encounter mutually non-informative BPAs in the Haenni and Lehmann [11]'s *Communication Network* example discussed in Section 3.3.

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. Thus, $m_{Y_x} : 2^{\Omega_Y} \to [0, 1]$ such that $\sum_{\emptyset \neq \mathsf{a} \subseteq \Omega_Y} m_{Y_x}(\mathsf{a}) = 1$. 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}.

[5] The notion of distinct BPAs is discussed in Section 3. As we will see, m_r and $m_{s|r'}$ are distinct if and only if $s \perp\!\!\!\perp_{(m_r \oplus m_{s|r'})} (r \setminus r') \mid r'$.

One way to do this is to take each focal element $\mathsf{b} \subseteq \Omega_Y$ of m_{Y_x} and convert it to the corresponding focal element

$$(\{x\} \times \mathsf{b}) \cup ((\Omega_X \setminus \{x\}) \times \Omega_Y) \subseteq \Omega_{X,Y} \qquad (15)$$

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. 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 implicit assumption is that $m_{Y|x_1}$ and $m_{Y|x_2}$ are distinct BPAs for $\{X, Y\}$.

Other sources of belief function conditionals are described in [12, 14]. Conditionals can also be constructed using the removal operator, discussed in the following subsection.

2.3. Removal Operator

The removal operator (also called 'decombination' in [32]) allows us to remove knowledge [25]. Suppose we construct a joint belief function for X and Y using BPA m_X for X and a conditional $m_{Y|X}$ for Y given X. Thus, the joint BPA for (X, Y) is $m_{X,Y} = m_X \oplus m_{Y|X}$. Notice that the marginal of $m_{X,Y}$ for X is m_X, i.e., $(m_{X,Y})^{\downarrow X} = m_X$. If we are given the joint BPA $m_{X,Y}$ for (X, Y), can we recover the conditional $m_{Y|X}$? The answer is yes, using the removal operator.

Definition 6 (Removal) *Suppose $m_{X,Y}$ is a BPA for (X, Y) such that $m_{X,Y} = m_X \oplus m_{Y|X}$, where m_X is a BPA for X, and $m_{Y|X}$ is a conditional for Y given X. Notice that $(m_{X,Y})^{\downarrow X} = m_X$. Let $Q_{X,Y}$ and Q_X denote the CFs corresponding to $m_{X,Y}$ and m_X respectively. Then, the removal of Q_X from $Q_{X,Y}$, written as $Q_{X,Y} \ominus Q_X$, is defined as follows:*

$$(Q_{X,Y} \ominus Q_X)(\mathsf{a}) = K^{-1} Q_{X,Y}(\mathsf{a})/Q_X(\mathsf{a}^{\downarrow X}) \qquad (16)$$

for all $\mathsf{a} \subseteq \Omega_{X,Y}$, where K is a normalization constant defined by

$$K = \sum_{\emptyset \neq a \subseteq \Omega_{X,Y}} (-1)^{|\mathsf{a}|+1} Q_{X,Y}(\mathsf{a})/Q_X(\mathsf{a}^{\downarrow X}) \qquad (17)$$

In Eqs. (16) and (17), if $Q_{X,Y}(\mathsf{a}) = 0$, then $Q_X(\mathsf{a}^{\downarrow X}) = 0$, and $0/0$ is defined to be 0.

Some comments on Def. 6:

1. The definition of the removal operator in Def. 6 is restricted to the case where the CF Q_X being removed is explicitly included in $Q_{X,Y}$ in the sense that $Q_{X,Y} = Q_X \oplus Q_{Y|X}$. This guarantees that $Q_{X,Y} \ominus Q_X$ is a well-defined CF [12, 14].

2. It follows from Eq. (16) that

$$\begin{aligned}(Q_{X,Y} \ominus Q_X)(\mathsf{a}) &= ((Q_X \oplus Q_{Y|X}) \ominus Q_X)(\mathsf{a}) \\ &= Q_X(\mathsf{a}^{\downarrow X})\, Q_{Y|X}(\mathsf{a})/Q_X(\mathsf{a}^{\downarrow X}) \\ &= Q_{Y|X}(\mathsf{a})\end{aligned}$$

   Thus, the removal operator can recover the conditional from the joint.

3. Removal can be defined more generally where the marginal CF $Q_X = (Q_{X,Y})^{\downarrow X}$ being removed from $Q_{X,Y}$ is not explicitly included in $Q_{X,Y}$. In this case, removal will result in a pseudo-CF as Eq. (6) will be violated [12, 14]. Pseudo-CFs are useful in inference [17]. This is because $(Q_{X,Y} \ominus Q_X) \oplus Q_X = Q_{X,Y}$.

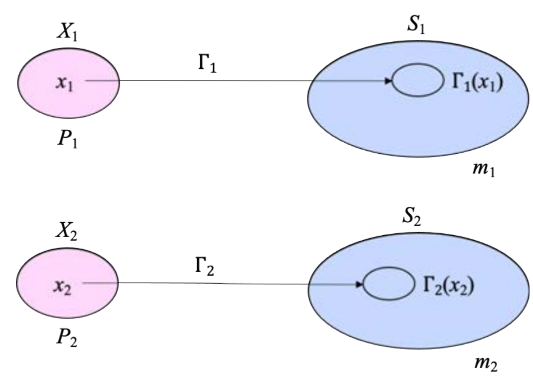
4. Some properties of the removal operator are as follows [25]:

   - Suppose Q is a CF for r and $s \subseteq r$. Then $Q \ominus Q^{\downarrow s}$ is a CF for r, assuming it is well-defined.
   - Suppose Q is a CF for r. Then $Q \ominus Q = \iota_r$, where ι_r is the vacuous CF for r.
   - Suppose Q_1, Q_2 are CFs for r and s, respectively, and suppose $t \subseteq s$. Then $(Q_1 \oplus Q_2) \ominus Q_2^{\downarrow t} = Q_1 \oplus (Q_2 \ominus Q_2^{\downarrow t})$

3. Distinct Belief Functions

This section discusses the notion of distinct belief functions. We start with Dempster [6]'s multi-valued mapping semantics associated with BPAs.

Figure 1: Dempster's multi-valued semantics for BPAs.



Definition 7 (Distinct belief functions) *Consider two discrete finite variables X_1 and S_1 with state spaces Ω_{X_1} and Ω_{S_1}. Assume that we have a probability mass function (PMF) P_1 on X_1. We have a multi-valued mapping*

$\Gamma_1 : X_1 \to 2^{\Omega_{S_1}}$ such that for each $x \in \Omega_{X_1}$, we associate a non-empty subset of S_1, $\Gamma_1(x) \in 2^{\Omega_{S_1}} \setminus \emptyset$. The multi-valued mapping Γ_1 defines the BPA m_1 for S_1 such that for all $\mathsf{a} \in 2^{\Omega_{S_1}} \setminus \emptyset$,

$$m_1(\mathsf{a}) = \sum_{x \in \Omega_{X_1}} \{P_1(x) : \Gamma_1(x) = \mathsf{a}\}. \tag{18}$$

Suppose we have another pair of discrete and finite variables X_2 and S_2 with PMF P_2 on X_2, and another multi-valued mapping $\Gamma_2 : X_2 \to 2^{\Omega_{S_2}} \setminus \emptyset$. The multi-valued mapping Γ_2 defines the BPA m_2 for S_2 such that for all $\mathsf{a} \in 2^{\Omega_{S_2}} \setminus \emptyset$,

$$m_2(\mathsf{a}) = \sum_{x \in \Omega_{X_2}} \{P_2(x) : \Gamma_2(x) = \mathsf{a}\}. \tag{19}$$

We say m_1 and m_2 are distinct if and only if the random variables X_1 (with PMF P_1) and X_2 (with PMF P_2) are independent.

Some comments on Def. 7:

1. As P_1, and P_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 the definition of distinct belief function in Def. 7 cannot be used directly in practice.

3. If we assume independence of variables X_1 and X_2 when they are not, then we are double-counting non-idempotent knowledge[6] [26]. Thus, the spirit of Def. 7 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.

3.1. Directed Graphical Models

In this subsection, we discuss the idea of distinct belief functions in a belief-function directed graphical model by incorporating ideas from probability theory.

Before we define a belief-function directed graphical model, we start with some notation. A directed graph G_d is a pair $G_d = (\mathcal{V}, \mathcal{E})$, 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 $Pa_{G_d}(X)$ 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).

[6]We say BPA m is idempotent if $m \oplus m = m$. For example, if m is deterministic, then m is idempotent. Idempotent knowledge is knowledge encoded in a BPA m that is idempotent. Thus, double-counting idempotent knowledge is not a problem; double-counting non-idempotent knowledge is.

Definition 8 (BF directed graphical model) *Suppose we have a directed acyclic graph $G_d = (\mathcal{V}, \mathcal{E})$ with n nodes in $\mathcal{V}$. A belief-function directed graphical model (BFDGM) is a pair $(G_d, \{m_1, \ldots, m_n\})$ such that BPA m_i associated with node X_i is a conditional BPA for X_i given $Pa_{G_d}(X_i)$, 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*

$$m = \bigoplus_{i=1}^{n} m_i. \tag{20}$$

Some comments about Def. 8:

1. The assumption in Def. 8 that all conditionals are distinct allows the combination in Eq. (20).

2. Given m, the joint BPA for $\mathcal{V}$ as defined in Eq. (20), it follows from Def. 3 that the following CI relations hold. Suppose $(X_1, \ldots, X_n)$ is a topological sequence associated with BFDGM $(G_d, \{m_1, \ldots, m_n\})$. Then for each X_i, $i = 2, \ldots, n$, given $Pa_{G_d}(X_i)$, X_i is conditionally independent of $\{X_1, \ldots X_{i-1}\} \setminus Pa_{G_d}(X_i)$.

3. An example of a BFDGM is given in Section 3.1.

Consider the probabilistic directed graphical model $X \to Y$, with potentials[7] $P(X)$ and $P(Y|X)$. $P(X)$ is a prior PMF for X, and $P(Y|X)$ is called a conditional probability table (CPT) for Y. The joint probability function of (X, Y) is the probabilistic combination of these two potentials, i.e., $P(X, Y) = P(X) \otimes P(Y|X)$. Here, $\otimes$ denotes the probabilistic combination operator, pointwise multiplication followed by normalization. Thus, $P(X, Y)(x, y) = P(X)(x) \cdot P(Y|X)(x, y)$. The directed graphical model $X \to Y$ makes no conditional independence assumptions. If we compute the marginal for X from $P(X, Y)$, we obtain $P(X)$, i.e.,

$$\begin{aligned} P(X) &= (P(X) \otimes P(Y|X))^{\downarrow X}, && (21)\\ &= P(X) \otimes P(Y|X)^{\downarrow X}, && (22)\\ &= P(X). && (23) \end{aligned}$$

[7]Potentials are unnormalized probability functions. A conditional probability table is not a probability distribution but can be considered a potential.

Eq. (22) follows from Eq. (21) using the local computation property of probabilistic combination. Eq. (23) follows from Eq. (22) utilizing the property of conditionals ($P(Y|X)^{\downarrow X}$ is a vacuous potential for X). Also, assuming the potential $P(X)$ has no zeroes, if we compute the conditional $P(X,Y) \div P(X,Y)^{\downarrow X}$ from the joint, we obtain $P(Y|X)$ (here, $\div$ denotes a pointwise division of the second potential from the first, the inverse of the $\otimes$ operator). Thus, we can conclude that the probabilistic combination of potential $P(X)$ and $P(Y|X)$ does not involve double counting of non-idempotent knowledge, i.e., the potentials $P(X)$ and $P(Y|X)$ are always distinct (regardless of the numeric values of these potentials).

Now, consider the probabilistic graphical model for X and Y without a directed edge from X to Y (or vice versa) with potentials $P(X)$ and $P(Y)$. This graphical model assumes X and Y are independent, and the joint PMF of (X,Y) is $P(X,Y) = P(X) \otimes P(Y)$. With the independence assumption, $P(Y|X)(x,y) = P(Y)(y)$ for all $(x,y) \in \Omega_{X,Y}$. Thus,

$$\begin{aligned} P(X,Y) &= P(X) \otimes P(Y|X), \\ &= P(X) \otimes P(Y) \end{aligned}$$

and there is no double counting of non-idempotent knowledge.

Next, consider the case where we have a model consisting of two probability potentials, PMFs $P(X)$ for X, and $P(Y)$ for Y, and suppose X and Y are *not* independent. In this case, the potentials $P(X)$ and $P(Y)$ are not distinct. Since X and Y are not independent, let $P(Y|X)$ denote the dependency of Y on X. Thus, $P(Y) = (P(X) \otimes P(Y|X))^{\downarrow Y}$. Thus,

$$P(X) \otimes P(Y) = P(X) \otimes (P(X) \otimes P(Y|X))^{\downarrow Y}. \quad (24)$$

Notice that in Eq. (24), $P(X)$ is counted twice, and if it is not idempotent, Eq. (24) will result in an incorrect joint distribution of (X,Y). We will illustrate this using an example.

Example 1 (Double-counting of knowledge) *Suppose X and Y are random variables with state spaces $\Omega_X = \Omega_Y = \{0,1\}$. Suppose $P(X)$ and $P(Y|X)$ are as shown in Table 1. $P(Y|X)$ represents the dependency $Y = X$. Notice that $P(X) \otimes P(Y)$ is different from the actual joint $P(X,Y)$.*

For yet another example, consider the directed graphical model $X \to Y \to Z$ with the potentials $P(X)$ for X, conditionals $P(Y|Z)$ for Y given X, and conditional $P(Z|Y)$ for Z given Y. This graphical model assumes that X and Z are conditionally independent (CI) given Y. With this CI assumption, the three potentials in the model are distinct. Without the CI assumption, the potentials are not distinct (similar to the previous example where X and Y are not independent).

Table 1: Comparing $P(X,Y)$ with $P(X) \otimes P(Y)$.

$\Omega_{\{X,Y\}}$	$P(X)$	$P(Y\|X)$	$P(X,Y)$	$P(Y)$	$P(X) \otimes P(Y)$
$\{(0,0)\}$	0.2	1	0.2	0.2	0.04
$\{(0,1)\}$	0.2	0	0	0.8	0.16
$\{(1,0)\}$	0.8	0	0	0.2	0.16
$\{(1,1)\}$	0.8	1	0.8	0.8	0.64

In the case of D-S belief-function directed graphical models, we have a situation similar to the probabilistic case. Each graphical model is associated with a set of conditional independence 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 [5, 25]. Also, associated with each variable X in the model, we have a conditional for X given its parents. Unlike the probabilistic case, some conditionals may not be known, so we have a vacuous BPA associated with such variables [27]. As in the probabilistic case, assuming the CI relations are valid, the BPAs in the model are distinct.

3.2. An Example of a BFDGM

Example 2 (Almond [1]'s Captain's Problem) *A ship's captain is concerned about how many days his ship may be delayed before arrival at a destination. The arrival delay is the sum of the departure delay and sailing delay. Departure delay may be a result of maintenance (at most one day), loading delay (at most one day), or a forecast of bad weather (at most one day). Sailing delays may result from bad weather (at most one day) and whether repairs are needed at sea (at most one day). If maintenance is done before sailing, chances of repairs at sea are less likely. The weather forecast says a slight chance of bad weather (0.2) and a good chance of good weather (0.6). The forecast is 80% reliable. The captain knows the loading delay and whether maintenance is done before departure. Figure 2 shows the directed acyclic graph associated with this problem.*

A topological ordering of the variables is as follows: (W,F,L,M,D,R,S,A). Let m denote the joint BPA of this model. The CI assumptions of this graphical model are as follows:

1. *$L \perp\!\!\!\perp_m \{W,F\}$;*
2. *$M \perp\!\!\!\perp_m \{W,F,L\}$;*
3. *$D \perp\!\!\!\perp_m W \mid \{F,L,M\}$;*
4. *$R \perp\!\!\!\perp_m \{W,F,L\} \mid M$;*
5. *$S \perp\!\!\!\perp_m \{F,L,M,D\} \mid \{W,R\}$; and*

6. $A \perp\!\!\!\perp_m \{W, F, L, M, R\} \mid \{D, S\}$.

Assuming the CI relations are all valid, and the BPAs in the model are all conditionals, the BPAs are distinct.

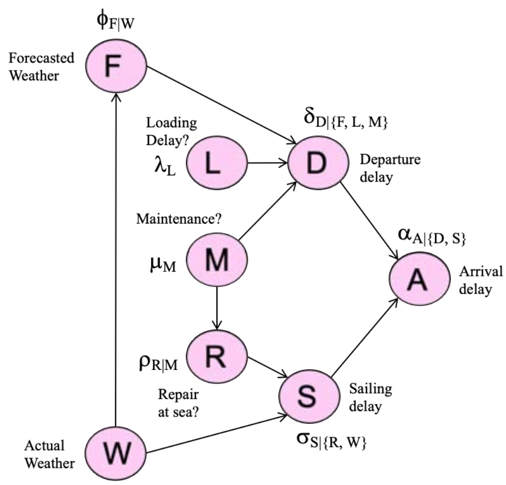


Figure 2: The directed acyclic graph for the *Captain's Problem*. The Greek alphabets adjacent to a variable denote the prior or conditional associated with the variable.

Table 2 shows the variables and their states. The conditional BPAs are as follows.

Table 2: The variables, their state spaces, and associated conditionals in the captain's problem.

Variable	Name	State Space	Assoc. Conditional
W	Actual weather	$\{g_w, b_w\}$	vacuous for W
F	Forecasted weather	$\{g_f, b_f\}$	$\phi_{F\mid W}$
L	Loading delay?	$\{t_l, f_l\}$	λ
M	Maint. done?	$\{t_m, f_m\}$	μ
R	Repair at sea?	$\{t_r, f_r\}$	ρ_1 and ρ_2
D	Dep. delay	$\{0, \ldots, 3\}$	$\delta_{D\mid\{F,L,M\}}$
S	Sailing delay	$\{0, \ldots, 3\}$	$\sigma_{S\mid\{W,R\}}$
A	Arrival delay	$\{0, \ldots, 6\}$	$\alpha_{A\mid\{D,S\}}$

1. *Weather forecast is 80% accurate. $\phi_{F|W}$ is a conditional for F given W.*

$$\begin{aligned}\phi_{F|W}(\{(g_w, g_f), (b_w, b_f)\}) &= 0.8,\\ \phi_{F|W}(\Omega_{W,F}) &= 0.2.\end{aligned}$$

2. *Loading is delayed with a chance of 0.3 and on schedule with a chance of 0.5. λ is a prior for L.*

$$\begin{aligned}\lambda(\{t_l)\}) &= 0.3,\\ \lambda(\{f_l)\}) &= 0.5,\\ \lambda(\Omega_L) &= 0.2.\end{aligned}$$

3. *Maintenance is not done. μ is a prior for M.*

$$\mu(\{f_m\}) = 1.$$

4. *If maintenance is done before sailing, the chances of repair at sea are between 10 and 30%. ρ_1 is a BPA for R given $M = t_m$.*

$$\begin{aligned}\rho_1(\{t_r\}) &= 0.1,\\ \rho_1(\{f_r\}) &= 0.7,\\ \rho_1(\Omega_R) &= 0.2.\end{aligned}$$

ρ_1 needs to conditionally embedded into a BPA for $\{R, M\}$ before it is considered as a conditional.

5. *If maintenance is not done before sailing, the chances of repair at sea are between 20 and 80%. ρ_2 is a BPA for R given $M = f_m$.*

$$\begin{aligned}\rho_2(\{t_r\}) &= 0.2,\\ \rho_1(\{f_r\}) &= 0.2,\\ \rho_1(\Omega_R) &= 0.6.\end{aligned}$$

ρ_2 needs to conditionally embedded into a BPA for $\{R, M\}$ before it is considered as a conditional.

6. *Bad weather and repair at sea each add a day to sailing delay. This proposition is true 90% of the time. $\sigma_{S|W,R}$ is a conditional for S given (W, R).*

$$\begin{aligned}&\sigma_{S|W,R}(\{(g_w, f_r, 0), (b_w, f_r, 1),\\ &\quad (g_w, t_r, 1), (b_w, t_r, 2)\}) = 0.9,\\ &\sigma_{S|W,R}(\Omega_{W,R,S}) = 0.1.\end{aligned}$$

7. *Departure delay may be a result of maintenance (at most one day), loading delay (at most one day), or a forecast of bad weather (at most one day). $\delta_{D|F,L,M}$ is a deterministic conditional for D given $\{F, L, M\}$.*

$$\begin{aligned}&\delta_{D|F,L,M}(\{(g_f, f_l, f_m, 0), (b_f, f_l, f_m, 1),\\ &(g_f, t_l, f_m, 1), (g_f, f_l, t_m, 1), (b_f, t_l, f_m, 2),\\ &(b_f, f_l, t_m, 2), (g_f, t_l, t_m, 2), (b_f, t_l, t_m, 3)\}) = 1.\end{aligned}$$

8. *The arrival delay is the sum of departure and sailing delays. $\alpha_{A|D,S}$ is a deterministic conditional for A given $\{D, S\}$.*

$$\alpha_{A|D,S}(\{(0,0,0),(0,1,1),(0,2,2),(0,3,3),\\(1,0,1),(1,1,2),(1,2,3),(1,3,4),\\(2,0,2),(2,1,3),(2,2,4),(2,3,5),\\(3,0,3),(3,1,4),(3,2,5),(3,3,6)\}) = 1.$$

Notice that all BPAs are conditionals.

3.3. Undirected Graphical Models

First, we start with some notation. Consider an undirected graph $G_u = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{X_1, \ldots, X_n\}$ denotes the set of nodes, and $\mathcal{E}$ denotes the set of undirected edges $\{X_i, X_j\}$ between two distinct variables in $\mathcal{V}$. A *clique* in G_u is a maximal completely connected subgraph of G. Given a variable $X \in \mathcal{V}$, the Markov blanket of X, denoted by $MB_{G_u}(X)$, is $\{Y \in \mathcal{V} : \{X, Y\} \in E\}$. The definition of a belief-function undirected graphical model below is taken from [13].

Definition 9 (BF undirected graphical model) *A* belief-function undirected graphical model *(BFUGM) is* $(G_u = (\mathcal{V}, \mathcal{E}), \{m_1, \ldots, m_k\})$*, where* G_u *is an undirected graph with cliques* $r_1, \ldots, r_k$*, and for each* $i = 1, \ldots, k$*,* m_i *is a BPA for* r_i*. A fundamental assumption of a BFUGM is that the BPAs are all distinct. Thus, a belief-function undirected graphical model corresponds to the joint BPA* m *for* $\mathcal{V}$ *defined as follows:*

$$m = \bigoplus_{i=1}^{k} m_i, \tag{25}$$

assuming that m *as defined in Eq. (25) is a well-defined BPA, i.e., the normalization constant* K *in Dempster's combination (Eq. (11)) is non-zero.*

Some comments about Def. 9.

1. The assumption in Def. 9 that the BPAs are all distinct allows the combination in Eq. (25).
2. Given m, the joint BPA for $\mathcal{V}$, it follows from Def. 9 that the following CI relations hold. For each $X \in \mathcal{V}$, $X \perp\!\!\!\perp_m (\mathcal{V} \setminus MB_{G_u}(X)) \mid MB_{G_u}(X)$.

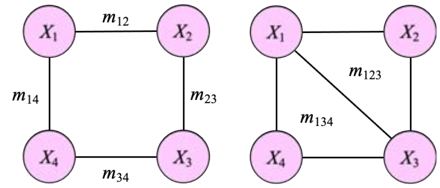
3.4. Examples of BFUGMs

In this subsection, we describe several examples of BFUGMs.

Example 3 (Two BFUGMs) *Consider the BFUGM on the left in Fig. 3. This UG has four cliques* $\{X_1, X_2\}$*,* $\{X_2, X_3\}$*,* $\{X_3, X_4\}$*,* $\{X_1, X_4\}$*. Suppose that the BPAs associated with the corresponding cliques are* m_{12}*,* m_{23}*,* m_{34}*, and* m_{14}*. Then, the joint BPA* m *associated with this BFUGM is:*

$$m = m_{12} \oplus m_{23} \oplus m_{34} \oplus m_{14}. \tag{26}$$

Figure 3: Two BFUGMs



This BFUGM has two CI assumptions: $X_1 \perp\!\!\!\perp_m X_3 \mid \{X_2, X_4\}$*, and* $X_2 \perp\!\!\!\perp_m X_4 \mid \{X_1, X_3\}$*. The first one follows from* $m = (m_{12} \oplus m_{14}) \oplus (m_{23} \oplus m_{34})$ *and Def. 3. The second one follows from* $m = (m_{12} \oplus m_{23}) \oplus (m_{34} \oplus m_{14})$ *and Def. 3.*

Consider the BFUGM on the right in Fig. 3. This UG has two cliques: $\{X_1, X_2, X_3\}$ *and* $\{X_1, X_3, X_4\}$*. Suppose the BPAs associated with the corresponding cliques are* m_{123} *and* m_{134}*. Then the joint BPA* m *associated with this BFUGM is:*

$$m = m_{123} \oplus m_{134} \tag{27}$$

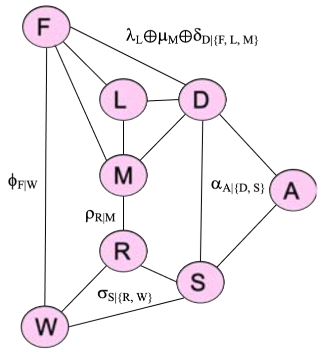
This BFUGM has one CI assumption: $X_2 \perp\!\!\!\perp_m X_4 \mid \{X_1, X_3\}$*. This follows directly from Eq. (27) and Def. 3.*

One source of undirected graphical models is the "moralization" of a directed graphical model (where we marry parents and drop directions) [18]. The BPAs in the undirected model are the same as (or some combination of) the BPAs in the corresponding directed model. Therefore, as the belief functions in a directed graphical model are distinct, the belief functions in the corresponding undirected graphical models are also distinct. For example, consider the directed graphical model $X \to Y \to Z$ with BPAs m_X for X, conditional BPA $m_{Y|X}$ for Y given X, and conditional $m_{Z|Y}$ for Z given Y. After moralization, we have an undirected graphical model $X - Y - Z$ with two BPAs $m_{X,Y} = m_X \oplus m_{Y|X}$ for $\{X, Y\}$ and conditional BPA $m_{Z|Y}$ for $\{Y, Z\}$. The conditional independence assumption associated with this model is: X is conditionally independent of Z given Y. Thus, we assume that the BPAs $m_{X,Y}$ and $m_{Z|Y}$ are distinct. We cannot take arbitrary BPAs $m_{X,Y}$ for (X, Y) and $m_{Y,Z}$ for (Y, Z) and claim that we have a model. We implicitly assume that the belief functions are distinct when using Dempster's combination rule. If the BPAs are not distinct, the result of Dempster's combination rule may lead to the double-counting of non-idempotent knowledge.

Fig. 4 shows the BFUGM obtained from the *Captain's Problem* (Fig. 2) by marrying parents and dropping directions. All the BPAs in this model are distinct.

Another source of undirected graphical models is where the clique belief functions all have the same structure for each clique. An example is Haenni and Lehmann [11]'s *Communication Network* example, where each clique consists of two linked variables, say X_1 and X_2, with state

Figure 4: The BFUGM obtained from the BFDGM in Fig. 2 by marrying parents and dropping directions.



spaces $\Omega_{X_1} = \{t_1, f_1\}$ and $\Omega_{X_2} = \{t_2, f_2\}$, respectively. The BPA m_{12} for $\{X_1, Y_1\}$ is as follows:

$$\begin{aligned} m_{12}(\{(t_1, t_2), (f_1, f_2)\}) &= 0.90, \\ m_{12}(\Omega_{\{X_1, X_2\}}) &= 0.1. \end{aligned} \tag{28}$$

In words, the reliability of the link $\{X_1, X_2\}$ is 90%. Figure 5 shows the undirected graph associated with this model. The reliabilities of the links $\{A, X_{33}\}$ and $\{B, X_{113}\}$ are 80%, and the reliabilities of all other links are 90%. The structure (focal elements) of all BPAs in the model is similar to the BPA m_{12} in Eq. (28).

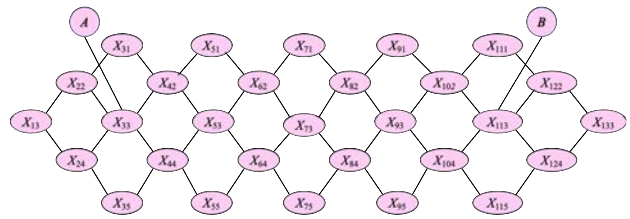
Notice that any two adjacent cliques will intersect at a single variable. Suppose m_{12} is a BPA for $\{X_1, X_2\}$, and m_{23} is a BPA for $\{X_2, X_3\}$. Notice that $(m_{12})^{\downarrow X_2}$ is a vacuous BPA for X_2. Similarly, $(m_{23})^{\downarrow X_2}$ is a vacuous BPA for X_2. Thus, $(m_{12} \oplus m_{23})^{\downarrow \{X_1, X_2\}} = m_{12}$ and $(m_{12} \oplus m_{23})^{\downarrow \{X_2, X_2\}} = m_{23}$. Thus, m_{12} and m_{23} are *mutually non-informative*. Also, the set of all BPAs in the communication network example is non-informative.

One consequence of this property is that m_{23} can be considered a conditional BPA for X_3 given X_2 (or for X_2 given X_3), and m_{12} can be considered a conditional BPA for X_1 given X_2 (or for X_2 given X_1). Thus, m_{12} and m_{23} are distinct BPAs using the logic of conditionals in Def. 4.

Each BPA in this model models the reliability of a link between two linked nodes. Suppose that the reliabilities of all the communication links are independent and the CI assumptions of the model are valid. In that case, we can infer that the BPAs in the undirected model are distinct.

Another argument for distinct belief functions in this example is as follows. As the set of all BPAs is non-informative, it seems intuitive that there is no double-counting of non-idempotent knowledge (assuming the CI assumptions are valid).

Figure 5: The *Communication Network* undirected graphical model. The variable X_{ij} is in the i^{th} column ($i = 1, \ldots, 13$), and j^{th} row ($j = 1, \ldots, 5$).



4. Summary & Conclusions

The main goal of this article is to discuss the notion of distinct belief functions in graphical models, both directed and undirected. We start with the definition given by Dempster [6] in his multi-valued semantics of a BPA. This cannot be used literally in practice as we don't associate a multi-valued function with each belief function in a model.

We provide heuristics for determining whether the belief functions in graphical models are distinct. The heuristics are based on Dempster's definition. For directed graphical models, we have conditionals associated with each variable in the model given its parents. The conditionals are all distinct if and only if *the conditional independence assumptions implied by the graphical model are valid*.

It is also straightforward for undirected graphical models derived from directed models by moralizing and dropping directions [18]. For a class of undirected graphical models, we have BPAs associated with each network clique with the same structure. For example, in the communication network example, all BPAs have the same structure, and each represents the reliability of the corresponding link in the communication network. Assuming that the reliabilities are independent, we can conclude that the BPAs in this example are distinct.

Unlike the case of directed graphical models, we do not have a general criterion for when the BPAs in an undirected graphical model are distinct. We have CI assumptions associated with an undirected graphical model that must be valid. The concept of a set of non-informative belief functions may be useful. This needs further investigation.

For learning belief-function graphical models from data, all existing structure learning algorithms in probability theory [15] should also apply to D-S belief functions theory as the definition of CI relations in D-S theory is the same as in probability theory. For parameter learning (BPAs), this remains to be done.

Acknowledgments

The Ronald G. Harper Professorship supported this study at the University of Kansas School of Business. The ideas discussed in this paper have benefitted from long email discussions with Radim Jiroušek and Václav Kratochvíl (I am not suggesting that Radim and Václav agree with this paper's discussions). I am also grateful to three anonymous ISIPTA-23 reviewers that provided constructive comments to improve the exposition.

References

[1] R. G. Almond. *Graphical Belief Modeling*. Chapman & Hall, London, UK, 1995.

[2] B. Ben-Yaghlane, P. Smets, and K. Mellouli. Belief function independence: I. The marginal case. *International Journal of Approximate Reasoning*, 29(1): 47–70, 2002.

[3] B. Ben-Yaghlane, P. Smets, and K. Mellouli. Belief function independence: II. The conditional case. *International Journal of Approximate Reasoning*, 31(1–2): 31–75, 2002.

[4] M. Chebbah, A. Martin, and B. Ben Yaghlane. Combining partially independent belief functions. *Decision Support Systems*, 73(5):37–46, 2015.

[5] Phil Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series B*, 41(1):1–15, 1979.

[6] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38(2):325–339, 1967.

[7] T. Denœux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artificial Intelligence*, 172(2–3):234–264, 2008.

[8] D. Dubois and H. Prade. Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4(3):244–264, 1988.

[9] C. Fu and S. Yang. The conjunctive combination of interval-valued belief structures from dependent sources. *International Journal of Approximate Reasoning*, 53(5):769–785, 2012.

[10] P.H. Giang and S. Shenoy. The belief function machine: An environment for reasoning with belief functions in Matlab. Working paper, University of Kansas School of Business, Lawrence, KS 66045, 2003. URL https://pshenoy.ku.edu/Papers/BFM072503.zip.

[11] R. Haenni and N. Lehmann. Resource bounded and anytime approximation of belief function computation. *International Journal of Approximate Reasoning*, 31 (1–2):103–154, 2002.

[12] R. Jiroušek, V. Kratochvíl, and P. P. Shenoy. On conditional belief functions in the Dempster-Shafer theory. In S. Le Hégarat-Mascle, I. Bloch, and E. Aldea, editors, *Belief Functions: Theory and Applications, 7th International Conference, BELIEF 2022*, volume 13506 of *Lecture Notes in Artificial Intelligence*, pages 207–218, Switzerland, 2022. Springer Nature.

[13] R. Jiroušek, V. Kratochvíl, and P. P. Shenoy. Computing the decomposable entropy of graphical belief function models. Working Paper 340, University of Kansas School of Business, Lawrence, KS 66045, January 2023. URL https://pshenoy.ku.edu/Papers/WP340.pdf.

[14] R. Jiroušek, V. Kratochvíl, and P. P. Shenoy. On conditional belief functions in directed graphical models in the Dempster-Shafer theory. Working Paper 341, University of Kansas School of Business, Lawrence, KS 66045, April 2023. URL https://pshenoy.ku.edu/Papers/WP341.pdf.

[15] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

[16] A. Kong. *Multivariate belief functions and graphical models*. PhD thesis, Department of Statistics, Harvard University, Cambridge, Massachusetts, 1986.

[17] S. L. Lauritzen and F. V. Jensen. Local computation with valuations from a commutative semigroup. *Annals of Mathematics and Artificial Intelligence*, 21(1): 51–69, 1997.

[18] S. L. Lauritzen and D. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, series B*, 50(2):157–224, 1988.

[19] E. Lefevre, O. Colot, and P. Vannoorenberghe. Belief function combination and conflict management. *Information Fusion*, 3(2):149–162, 2002.

[20] T. Nakama and E. Ruspini. Combining dependent evidential bodies that share common knowledge. *International Journal of Approximate Reasoning*, 55(5): 2109–2125, 2014.

[21] J. Pearl and A. Paz. Graphoids: Graph-based logic for reasoning about relevance relations. In B. D. Boulay, D. Hogg, and L. Steele, editors, *Advances in Artificial Intelligence II*, pages 357–363. North-Holland, Amsterdam, 1987.

[22] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[23] G. Shafer. The problem of dependent evidence. Working Paper 164, University of Kansas School of Business, Lawrence, KS 66045, October 1984.

[24] G. Shafer. The problem of dependent evidence. *International Journal of Approximate Reasoning*, 79(12): 41–44, 2016.

[25] P. P. Shenoy. Conditional independence in valuation-based systems. *International Journal of Approximate Reasoning*, 10(3):203–234, 1994.

[26] P. P. Shenoy. No double counting semantics for conditional independence. In F. G. Cozman, R. Nau, and T. Seidenfeld, editors, *Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications (ISIPTA-05)*, pages 306–314. Society for Imprecise Probabilities and Their Applications, 2005.

[27] P. P. Shenoy. Making inferences in incomplete Bayesian networks: A Dempster-Shafer belief function approach. Working Paper 343, University of Kansas School of Business, Lawrence, KS 66045, March 2023. URL https://pshenoy.ku.edu/Papers/WP343.pdf.

[28] P. P. Shenoy and G. Shafer. Axioms for probability and belief-function propagation. In R. D. Shachter, T. Levitt, J. F. Lemmer, and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 4*, volume 9 of *Machine Intelligence and Pattern Recognition Series*, pages 169–198. North-Holland, Amsterdam, Netherlands, 1990.

[29] P. Smets. *Un modele mathematico-statistique simulant le processus du diagnostic medical*. PhD thesis, Free University of Brussels, 1978.

[30] P. Smets. The concept of distinct evidence. In B. Bouchon-Meunier, editor, *Proceedings of the 4th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'92)*, pages 789–794, 1992.

[31] P. Smets. Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem. *International Journal of Approximate Reasoning*, 9 (1):1–35, 1993.

[32] P. Smets. The canonical decomposition of a weighted belief. In *Proceedings of the 1995 International Joint Conference on Artificial Intelligence (IJCAI)*, volume 95, pages 1896–1901, 1995.

[33] R. R. Yager. On the Dempster-Shafer framework and new combination rules. *Information Sciences*, 41(2): 93–137, 1987.