On Distinct Belief Functions in the Dempster-Shafer Theory

Prakash P. Shenoy

University of Kansas, School of Business, Department of Analytics, Information, and Operations

April 21, 2023

Outline

Outline

Introduction

- 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 used to combine distinct belief functions.
- What constitutes distinct belief functions?
- We have an answer in Dempster's multi-valued semantics for belief functions.
- In practice, we don't associate multi-valued functions with belief functions.
- The idea of distinct belief functions corresponds to no double-counting of uncertain knowledge semantics of conditional independence.
- Although we discuss distinct belief functions in the D-S theory, the discussion is generally valid in many uncertainty calculi, including probability theory, possibility theory, and Spohn's epistemic belief theory.

Outline

Basics of D-S Belief Function Theory

Static: We represent knowledge using either:

- basic probability assignment (BPA) m
- commonality function (CF) Q

Dynamic: We make inferences using the following operators:

- Marginalization rule
- Dempster's rule of combination
- Removal (inverse of combination)
- Inference: Given a set of distinct belief functions (BPA, BF, PF, CF) representing knowledge of the domain and all evidence, we would like to find the marginals of the joint belief function for some variables of interest.
- The joint belief function is obtained by combining all belief functions using Dempster's rule of combination.

Basic Probability Assignments

Notation:

- Suppose $\mathcal{V}$ denotes a finite set of variables.
- For each $X \in \mathcal{V}$, Ω_X denotes a finite set of states of X.
- Let r, s, t, etc., denote subsets of $\mathcal{V}$.
- For every non-empty subset $s \subseteq \mathcal{V}$,
$$\Omega_s = \times_{X \in s} \Omega_X$$
denotes the states of s.
- Let 2^{Ω_s} denote the set of all subsets of Ω_s

Basic Probability Assignments

Basic Probability Assignment:

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

$$m(\emptyset) = 0, \quad (1)$$

$$\sum_{\emptyset \neq \mathsf{a} \subseteq \Omega_s} m(\mathsf{a}) = 1. \quad (2)$$

- s is called the domain of m.
- Subsets $\mathsf{a} \subseteq \Omega_s$ such that $m(\mathsf{a}) > 0$ are called focal elements of m.

Basic Probability Assignments

- Consider probability mass function (PMF) for A: $P_A(a) = 0.01$, $P_A(\bar{a}) = 0.99$.
- P_A can be represented by BPA m_A for A as follows: $m_A(\{a\}) = 0.01$, $m_A(\{\bar{a}\}) = 0.99$
- If all focal elements of m are singleton subsets, then m is called Bayesian. BPA m_A for A is Bayesian.
- If we have only one focal element (with probability 1), then we say m is deterministic
- Consider E (either T or L). $E = e$ iff $T = t$ or $L = l$.
- Consider a BPA m for $\{T, L, E\}$ as follows:

$$m(\{(t, l, e), (t, \bar{l}, e), (\bar{t}, l, e), (\bar{t}, \bar{l}, \bar{e})\}) = 1$$

Then, m is deterministic.

Basics of D-S Belief Function Theory

- Suppose m is a BPA for s. We say m is vacuous if m is deterministic with focal set Ω_s, i.e., $m(\Omega_s) = 1$.
- Consider BPA m_A for A such that $m_A(\Omega_A) = 1$. m is vacuous.
- Consider BPA m_A for A such that $m_A(\{a\}) = 0.5$, $m_A(\{\bar{a}\}) = 0.5$. This is a Bayesian, non-vacuous BPA for A.
- In D-S belief function theory, we can distinguish between equally-likely states and vacuous knowledge.

Commonality Functions

Commonality Functions:

- A commonality Q_m for s corresponding to BPA m for s is a function $Q_m : 2^{\Omega_s} \to [0,1]$ such that

$$Q_m(\mathsf{a}) = \sum_{\mathsf{b} \subseteq \Omega_s : \mathsf{b} \supseteq \mathsf{a}} m(\mathsf{b}) \tag{3}$$

- $Q_m(\mathsf{a})$ represents the probability mass that could move to every state in a.
- It follows from Eq. (3) that $0 \leq Q_m(\mathsf{a}) \leq 1$.
- It follows from Eqs. (1)-(3) that $Q_m(\emptyset) = 1$.
- Example: Suppose BPA m for T is as follows:

$$m(\emptyset) = 0, m(\{t\}) = 0.0005, m(\{\bar{t}\}) = 0.9405, m(\Omega_T) = 0.0590.$$

Then,

$$Q_m(\emptyset) = 1, Q_m(\{t\}) = 0.0595, Q_m(\{\bar{t}\}) = 0.9995, Q_m(\Omega_T) = 0.0595.$$

Commonality Functions

- BPA m, CF Q_m has the same information.
- Given CF Q for s, we can recover corresponding m as follows:

$$m_Q(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_s} : \, \mathsf{b} \supseteq \mathsf{a}} (-1)^{|\mathsf{b} \setminus \mathsf{a}|} Q(\mathsf{b}). \tag{4}$$

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

$$Q(\emptyset) = 1, \tag{5}$$

$$\sum_{\mathsf{b} \in 2^{\Omega_s} : \, \mathsf{b} \supseteq \mathsf{a}} (-1)^{|\mathsf{b} \setminus \mathsf{a}|} Q(\mathsf{b}) \geq 0, \quad \text{and} \tag{6}$$

$$\sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_s}} (-1)^{|\mathsf{a}|+1} Q(\mathsf{a}) = 1. \tag{7}$$

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 $\sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_X}} m_Q(\mathsf{a})$. 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_X} by $K = \sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_s}} (-1)^{|\mathsf{a}|+1} Q_m(\mathsf{a})$, and the resulting function will then qualify as a CF.

Commonality Functions

- In some cases, we could have a CF that doesn't satisfy Eq. (6) but does satisfy Eqs. (5) and (7). In such cases, we will call such CFs pseudo-CFs. If we convert a pseudo-commonality function 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.
- 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 $\mathsf{a} \in 2^{\Omega_s}$.
- 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 $|\mathsf{a}| > 1$.

Marginalization Rule

Marginalization rule

- Marginalization in belief function theory is addition.
- Projection of states: If $\mathbf{x} \in \Omega_s$, and $X \in s$, then $\mathbf{x}^{\downarrow s \setminus \{X\}}$ is the state of $s \setminus \{X\}$ obtained from $\mathbf{x}$ by dropping the state of X.
- Projection of subset of states: If $\mathsf{a} \in 2^{\Omega_s}$, then $\mathsf{a}^{\downarrow s \setminus \{X\}}$ is

$$\mathsf{a}^{\downarrow s \setminus \{X\}} = \{\mathbf{x}^{\downarrow s \setminus \{X\}} \ : \ \mathbf{x} \in \mathsf{a}\}$$

Definition (Marginalization rule)

If m is a bpa for s, and $X \in s$, then $m^{\downarrow s \setminus \{X\}}$ is a bpa for $s \setminus \{X\}$ defined as follows:

$$m^{\downarrow s \setminus \{X\}}(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_s} \, : \, \mathsf{b}^{\downarrow s \setminus \{X\}} = \mathsf{a}} m(\mathsf{b})$$

for all $\mathsf{a} \in 2^{s \setminus \{X\}}$.

Marginalization Rule

- Consider the deterministic BPA m for $\{T, L, E\}$ as follows:

$$m(\{(t, l, e), (t, \bar{l}, e), (\bar{t}, l, e), (\bar{t}, \bar{l}, \bar{e})\}) = 1$$

  Then, $m^{\downarrow\{T,L\}}$ is the BPA for $\{T, L\}$ given by

$$m^{\downarrow\{T,L\}}(\{(t, l), (t, \bar{l}), (\bar{t}, l), (\bar{t}, \bar{l})\}) = m^{\downarrow\{T,L\}}(\Omega_{\{T,L\}}) = 1.$$

- Notice that $m^{\downarrow\{T,L\}}$ is the vacuous BPA for $\{T, L\}$.

Marginalization Rule

- The definition of marginalization of BPA functions has the following properties:
  - (Domain) If m is a BPA for s, and $X \in s$, then $m^{\downarrow s \setminus \{X\}}$ is a BPA for $s \setminus \{X\}$.
  - (Order does not matter) If m is a BPA for s, $X, Y \in s$, then

$$(m^{\downarrow s \setminus \{X\}})^{\downarrow s \setminus \{X,Y\}} = (m^{\downarrow s \setminus \{Y\}})^{\downarrow s \setminus \{X,Y\}}.$$

Dempster's Combination Rule

- The combination rule in the D-S theory of belief functions is Dempster's rule, which Dempster called the "product-intersection" rule.
- The product of the BPA masses is assigned to the intersection of the focal elements, any mass assigned to the empty set is discarded, and the remaining masses re-normalized.

Definition (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 $\mathsf{a} \in 2^{\Omega_{s_1 \cup s_2}}$

$$(m_1 \oplus m_2)(\mathsf{a}) = K^{-1} \sum_{\mathsf{a}_1 \in 2^{\Omega_{s_1}}, \mathsf{a}_2 \in 2^{\Omega_{s_2}} : (\mathsf{a}_1 \times \Omega_{s_2 \setminus s_1}) \cap (\mathsf{a}_2 \times \Omega_{s_1 \setminus s_2}) = \mathsf{a}} m_1(\mathsf{a}_1)\, m_2(\mathsf{a}_2), \tag{8}$$

where K is a normalization constant given by

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

Dempster's Combination Rule

- Dempster's combination rule can also be described using commonality functions.
- In terms of CFs, Dempster's rule is pointwise multiplication of commonality functions.

Theorem (Shafer 1976)

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. Let CF $Q_1 \oplus Q_2$ correspond to BPA $m_1 \oplus m_2$. Then, for all $\emptyset \neq \mathsf{a} \in 2^{\Omega_{s_1 \cup s_2}}$

$$(Q_1 \oplus Q_2)(\mathsf{a}) = K^{-1} Q_1(\mathsf{a}^{\downarrow s_1})\, Q_2(\mathsf{a}^{\downarrow s_2}), \tag{10}$$

where K is a normalization constant defined as follows:

$$K = \sum_{\emptyset \neq \mathsf{a} \in \Omega_{s_1 \cup s_2}} (-1)^{|\mathsf{a}|+1} Q_1(\mathsf{a}^{\downarrow s_1})\, Q_2(\mathsf{a}^{\downarrow s_2}). \tag{11}$$

The normalization constant in Eq. (11) is precisely the same as in Eq. (9).

Dempster's Combination Rule

- If $m \oplus m = m$, we say m is idempotent. For example, if m is deterministic, m is idempotent.
- In general $m \oplus m \neq m$.
- Thus, in combining, e.g., BPAs m_1 and m_2 by Dempster's rule, it is important that m_1 and m_2 are distinct pieces of evidence (to avoid double-counting of non-idempotent knowledge).
- Dempster's rule satisfies the following properties:
  - (Domain) If m_1 is a BPA for s_1 and m_2 is a BPA for s_2, then $m_1 \oplus m_2$ is a BPA for $s_1 \cup s_2$.
  - (Commutative) $m_1 \oplus m_2 = m_2 \oplus m_1$
  - (Associative) $m_1 \oplus (m_2 \oplus m_3) = (m_1 \oplus m_2) \oplus m_3$
- (Local computation) Marginalization and Dempster's rules satisfy the following property: If m_1 is a BPA for s_1, m_2 is a BPA for s_2, $X \in s_1$, and $X \notin s_2$, then:

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

Conditional Independence

- Conditional independence (CI) in the D-S theory is similar to CI in probability theory [Dawid 1979, Shenoy 1994].

Definition (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, denoted by $r \perp\!\!\!\perp_m s | t$, 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.

- The definition above uses factorization semantics of CI. This is useful in graphical models.

Conditional BPAs

- In directed graphical models, we have conditional BPAs.

Definition (Conditional BPAs)

Suppose r and s are disjoint subsets of variables and suppose $r' \subset r$. 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. Thus, $m_r \oplus m_{s|r'}$ is a BPA for $r \cup s$.

- We call s the head of the conditional $m_{s|r'}$, and r' the tail.
- Using the definition of CI, m_r and $m_{s|r'}$ are distinct if and only if $s \perp\!\!\!\perp_{m_r \oplus m_{s|r'}} (r \setminus r')|r'$.
- In a directed graphical 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.
- 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 5 reflects this fact.

Conditional BPAs

Where do conditional BPAs come from? One way is to use Smets' conditional embedding.

- Suppose we have some knowledge about Y in the context $X = x$ encoded as BPA m_{Y_x} for Y.
- The knowledge of Y encoded in BPA m_{Y_x} for Y is valid only in the case $X = x$.
- Using Smets' conditional embedding, we convert the BPA m_{Y_x} for Y to a BPA $m_{Y|x}$ for $\{X, Y\}$ as follows:

Definition (Smets' conditional embedding)

$m_{Y|x}$ for $\{X, Y\}$ is defined as follows:

$$m_{Y|x}((\{x\} \times \mathsf{b}) \cup ((\Omega_X \setminus \{x\}) \times \Omega_Y)) = m_{Y_x}(\mathsf{b})$$

for each focal element b of m_{Y_x}.

Conditional BPAs

An Example

- Suppose X and Y are variables with $\Omega_X = \{x, \bar{x}\}$ and $\Omega_Y = \{y, \bar{y}\}$.
- If $X = x$, assume m_{Y_x} is as follows:

$$\begin{aligned} m_{Y_x}(\{y\}) &= 0.8, \\ m_{Y_x}(\Omega_Y) &= 0.2. \end{aligned}$$

- Then $m_{Y|x}$ is a BPA for $\{X, Y\}$ as follows:

$$\begin{aligned} m_{Y|x}(\{(x, y), (\bar{x}, y), (\bar{x}, \bar{y})\}) &= 0.8, \\ m_{Y|x}(\Omega_{X,Y}) &= 0.2. \end{aligned}$$

Conditional BPAs

- The BPA $m_{Y|x}$ for $\{X, Y\}$ obtained from m_{Y_x} by Smets' conditional embedding has the following property.
  1. $(m_{Y|x})^{\downarrow X}$ is a vacuous bpa for X. Thus, it is a conditional BPA for Y given X.
  2. Suppose $m_{X=x}$ is a bpa for X as follows: $m_{X=x}(\{x\}) = 1$. Then,

$$(m_{Y|x} \oplus m_{X=x})^{\downarrow Y} = m_{Y_x}.$$

Conditional BPAs

- Consider $m_{Y|x}$:

$$\begin{aligned} m_{Y|x}(\{(x,y),(\bar{x},y),(\bar{x},\bar{y})\}) &= 0.8, \\ m_{Y|x}(\Omega_{X,Y}) &= 0.2. \end{aligned}$$

- It is clear that $m_{Y|x}^{\downarrow X}$ is vacuous for X.
- Consider $m_{X=x} \oplus m_{Y|x}$:

$m_{X=x} \oplus m_{Y\|x}$	$\{(x,y),(\bar{x},y),(\bar{x},\bar{y})\}$ 0.8	$\Omega_{\{X,Y\}}$ 0.2
$\{(x,y),(x,\bar{y})\}$ 1	$\{(x,y)\}$ 0.8	$\{(x,y),(x,\bar{y})\}$ 0.2

- Thus, $(m_{X=x} \oplus m_{Y|x})(\{(x,y)\}) = 0.8$, $(m_{X=x} \oplus m_{Y|x})(\{(x,y),(x,\bar{y})\}) = 0.2$.
- Thus, $(m_{X=x} \oplus m_{Y|x})^{\downarrow Y} = m_{Y_x}$.

Conditional BPAs

- Another source of conditionals is deterministic knowledge
- Consider the Chest Clinic example. $E = e$ if and only if $(T = t) \vee (L = l)$. This can be represented by a deterministic BPA m for $\{T, L, E\}$ as follows:

$$m(\{(t, l, e), (t, \bar{l}, e), (\bar{t}, l, e), (\bar{t}, \bar{l}, \bar{e})\}) = 1.$$

Notice that m is a conditional BPA for E given $\{T, L\}$ as $m^{\downarrow\{T,L\}}$ is vacuous.

Removal Operator

- Suppose we construct a joint BPA $m_{X,Y} = m_X \oplus m_{Y|X}$ for (X, Y).
- Notice that $(m_{X,Y})^{\downarrow X} = m_X$.
- Starting from the joint BPA $m_{X,Y}$, can we recover the conditional?
- Yes! Using the removal operator [Shenoy 1994]

Removal Operator

Definition (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:

$$(Q_{X,Y} \ominus Q_X)(\mathsf{a}) = K^{-1} Q_{X,Y}(\mathsf{a}) / Q(\mathsf{a}^{\downarrow X}) \tag{12}$$

for all $\mathsf{a} \in 2^{\Omega_{X,Y}}$, where K is a normalization constant defined by

$$K = \sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|+1} Q_{X,Y}(\mathsf{a}) / Q(\mathsf{a}^{\downarrow X}) \tag{13}$$

Removal Operator

- It follows from Eq. (12) that for all $\mathsf{a} \subseteq \Omega_{X,Y}$:

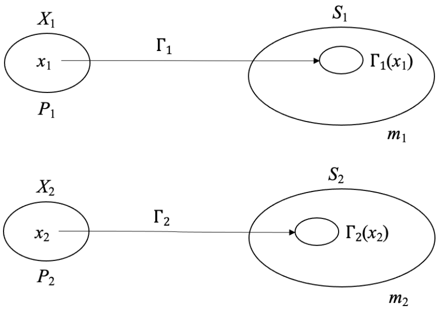
$$\begin{aligned}(Q_{X,Y} \ominus Q_X)(\mathsf{a}) &= ((Q_X \oplus Q_{Y|X}) \ominus Q_X)(\mathsf{a}) \\ &= Q_X(\mathsf{a}^{\downarrow X})\, Q_{Y|X}(\mathsf{a})/Q_X(\mathsf{a}^{\downarrow X}) \\ &= Q_{Y|X}(\mathsf{a})\end{aligned}$$

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

Outline

Distinct Belief Functions

Dempster's multi-valued semantics for belief functions:



- We have X_1 for which we have a PMF P_1 and multivalued mapping $\Gamma_1 : X_1 \to 2^{S_1} \setminus \emptyset$ that results in a BPA m_1 for S_1.
- We have space X_2 for which we have a PMF P_2 and multivalued mapping $\Gamma_2 : X_1 \to 2^{S_1} \setminus \emptyset$ that results in a BPA m_2 for S_2.
- m_1 and m_2 are distinct if and only if X_1 and X_2 are independent.

Distinct Belief Functions

Dempster's multi-valued semantics for belief functions:

- 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 cannot be used directly in practice.
- If we assume independence of variables X_1 and X_2 when they are not, then we double-count 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.

Directed Graphical Models

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 G 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 topology of the directed graph).

Directed Graphical Models

Definition (Belief-function directed graphical model)

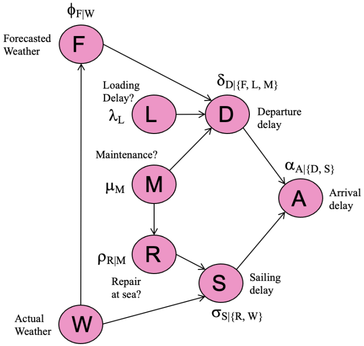
Suppose we have a directed acyclic graph $G = (\mathcal{V}, \mathcal{E})$ with n nodes in $\mathcal{V}$. A belief-function directed graphical model (BFDGM) is a pair $(G, \{m_1, \ldots, m_n\})$ such that BPA m_i associated with node X_i is a conditional BPA for X_i given $Pa_G(X_i)$, for $i = 1, \ldots, n$. 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

$$m = \bigoplus_{i=1}^{n} m_i. \tag{14}$$

1. The assumption in the definition that all conditionals are distinct allows the combination in Eq. (14).
2. Given m, the joint BPA for $\mathcal{V}$ as defined in Eq. (14), it follows from the definition of CI 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, $i = 2, \ldots, n$, given $Pa_G(X_i)$, X_i is conditionally independent of $\{X_1, \ldots X_{i-1}\} \setminus Pa_G(X_i)$.

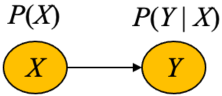
Directed Graphical Models

An Example: Captain's Problem



- A topological sequence: (W, F, L, M, D, R, S, A).
- CI assumptions: $L \perp\!\!\!\perp_m \{W, F\}$, $M \perp\!\!\!\perp_m \{W, F, L\}$, $D \perp\!\!\!\perp_m W \mid \{F, L, M\}$, $R \perp\!\!\!\perp_m \{W, F, L, D\} \mid M$, etc.

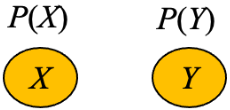
Directed Graphical Models

Consider the probabilistic graphical model: 

- The joint PMF $P(X,Y) = P(X) \otimes P(Y|X)$, where $\otimes$ is pointwise multiplication followed by normalization (Bayes' rule).
- $P(X,Y)^{\downarrow X} = (P(X) \otimes P(Y|X))^{\downarrow X} = P(X) \otimes P(Y|X))^{\downarrow X} = P(X)$. Notice that $P(Y|X))^{\downarrow X}$ is a potential with all ones, a vacuous potential.
- Assuming $P(X)$ has no zeroes, $P(X,Y) \oslash P(X) = (P(X) \otimes P(Y|X)) \oslash P(X) = P(Y|X)$, where $\oslash$ is pointwise division followed by normalization.
- There are no CI assumptions in this model.
- Thus, $P(X)$ and $P(Y|X)$ are always distinct.

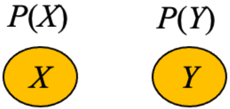
Directed Graphical Models

Consider the probabilistic graphical model:



- The joint PMF $P(X, Y) = P(X) \otimes P(Y)$.
- The model assumes $X \perp\!\!\!\perp_{P(X,Y)} Y$. Therefore $P(Y|X)(x, y) = P(Y)(y)$ for all $(x, y) \in \Omega_{X,Y}$.
- Thus, $P(X, Y) = P(X) \otimes P(Y|X) = P(X) \otimes P(Y)$.
- Assuming $X \perp\!\!\!\perp_{P(X,Y)} Y$, the potentials $P(X)$ and $P(Y)$ are distinct.

Directed Graphical Models

Consider the probabilistic graphical model: 

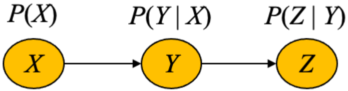
, where X and Y are not independent. Suppose $Y = X$.

Table: Comparing $P(X, Y)$ with $P(X) \otimes P(Y)$ (assuming $Y = X$)

$\Omega_{X,Y}$	$P(X)$	$P(Y \vert X)$	$P(X,Y)$	$P(Y)$	$P(X) \otimes P(Y)$
$(0,0)$	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

- The joint PMF $P(X,Y) = P(X) \otimes P(Y)$ as per the model is different from the true joint $P(X,Y)$.
- Without the CI assumption of the model, the potentials $P(X)$ and $P(Y)$ are not distinct. $P(X) \otimes P(Y)$ results in double-counting of non-idempotent knowledge.

Directed Graphical Models

Consider the probabilistic graphical model: 

- The joint PMF $P(X, Y, Z) = P(X) \otimes P(Y|X) \otimes P(Z|Y)$.
- The model assumes $Z \perp\!\!\!\perp_{P(X,Y,Z)} X \mid Y$.
- With this CI assumption, the three potentials are distinct.
- If the CI assumption is not valid, then $P(Z|Y)$ is not distinct from $P(X) \otimes P(Y|X)$.

Directed Graphical Models

In the case of a belief-function directed graphical model, we have a model similar to the probabilistic case

- The graphical model is associated with a set of CI assumptions.
- The definition of CI in the D-S theory is similar to the one for probability theory (Dawid 1979, Shenoy 1994).
- Associated with each variable X in the model, we have conditional for X given its parents.
- Unlike the probabilistic case, some conditionals may not be known.
- As in the probabilistic case, assuming the CI relations are valid, the BPAs in the model are distinct.

Directed Graphical Models

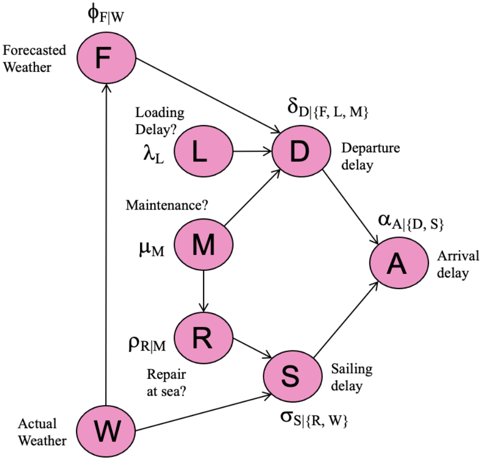
- Captain's Problem (R. Almond, *Graphical Belief Modeling*, Chapman and Hall, 1995)
  - A ship's captain is concerned about how many days his ship may be delayed before arrival at a destination.
  - The arrival delay is the sum of departure and sailing delays, $A = D + S$.
  - Delay in departure may be a result of maintenance (at most one day), delay in loading (at most one day), or due to forecast of bad weather (at most I day).
  - Delay in sailing may result from bad weather (at most one day) and whether repairs may be needed at sea (at most one day).
  - If maintenance is done before sailing, chances of repairs at sea are less likely.
  - Weather forecast says a small chance of bad weather (.2) and a good chance of good weather (0.6). The forecast is 80% reliable.
  - Captain has some knowledge of loading delay and whether maintenance is done before departure.

Directed Graphical Models

- Variables
  - A (arrival delay), $\Omega_A = \{0, 1, 2, 3, 4, 5\}$.
  - D (departure delay), $\Omega_D = \{0, 1, 2, 3\}$.
  - S (sailing delay), $\Omega_S = \{0, 1, 2\}$.
  - L (is loading delayed?), $\Omega_L = \{t, f\}$.
  - F (weather forecast), $\Omega_F = \{b, g\}$.
  - W (actual weather), $\Omega_W = \{b, g\}$.
  - M (is maintenance done before sailing?), $\Omega_M = \{t, f\}$.
  - R (is a repair at sea needed?), $\Omega_R = \{t, f\}$.

Directed Graphical Models

- The Captain problem can be described by a causal directed acyclic graph (DAG) as follows:



Directed Graphical Models

Conditional for F given W:

- Forecast is 80% reliable
- This piece of knowledge is represented by conditional BPA ϕ_1 for F given W such that

$$\begin{aligned} \phi_1(\{(b,b),(g,g)\}) &= 0.8, \\ \phi_1(\Omega_{\{W,F\}}) &= 0.2. \end{aligned}$$

Directed Graphical Models

Priors for L and M:

- Loading is delayed with chance 0.3 and on schedule with chance 0.5.
- This knowledge is modeled by BPA λ for $\{L\}$:

$$\begin{aligned}\lambda(\{t\}) &= 0.3,\\ \lambda(\{f\}) &= 0.5,\\ \lambda(\Omega_{\{L\}}) &= 0.2.\end{aligned}$$

- No maintenance was done on the ship before departure
- This piece of knowledge is represented by BPA μ for $\{M\}$ such that

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

Directed Graphical Models

Conditional for D given $\{L, F, M\}$:

- Loading delay, bad weather forecast, and maintenance each add one day to the departure delay
- We model this piece of knowledge by conditional BPA δ for D given $\{L, F, M\}$:

$$\delta(\{(f, g, f, 0), (t, g, f, 1), (f, b, f, 1), (f, g, t, 1), (f, b, t, 2), (t, g, t, 2), (t, g, f, 2), (t, b, t, 3)\}) = 1.$$

- Notice that $\delta^{\downarrow\{L,F,M\}}$ is vacuous for $\{L, F, M\}$.

Directed Graphical Models

Conditional for R given $M = t$:

- If maintenance was done before sailing, then the chances of repair at sea are between 10 and 30%. This is represented by BPA $\rho_{M=t}$ for R as follows:

$$\begin{aligned} \rho_{M=t}(\{t\}) &= 0.1, \\ \rho_{M=t}(\{f\}) &= 0.7, \\ \rho_{M=t}(\{t, f\}) &= 0.2. \end{aligned}$$

- After conditional embedding, ρ_1 is a conditional BPA for R given M as follows:

$$\begin{aligned} \rho_1(\{(t,t), (f,t), (f,f)\}) &= 0.1, \\ \rho_1(\{(t,f), (f,t), (f,f)\}) &= 0.7, \\ \rho_1(\Omega_{\{M,R\}}) &= 0.2. \end{aligned}$$

Directed Graphical Models

Conditional for R given $M = f$:

- If maintenance was not done before sailing, then the chances of repair at sea are between 20 and 80%. This is represented by conditional BPA $\rho_{M=f}$ for R as follows:

$$\begin{aligned}
\rho_{M=f}(\{t\}) &= 0.2, \\
\rho_{M=f}(\{f\}) &= 0.2, \\
\rho_{M=f}(\{t,f\}) &= 0.6.
\end{aligned}$$

- After conditional embedding, ρ_2 is a conditional BPA for R given M as follows:

$$\begin{aligned}
\rho_2(\{(f,t),(t,t),(t,f)\}) &= 0.2, \\
\rho_2(\{(f,f),(t,t),(t,f)\}) &= 0.2, \\
\rho_2(\Omega_{\{M,R\}}) &= 0.6.
\end{aligned}$$

- $\rho_1 \oplus \rho_2$ can be considered as a conditional BPA $m_{R|M}$ for R given M.

Directed Graphical Models

Conditional for S given $\{W, R\}$:

- At least 90% of the time, bad weather and repair at sea each add one day to the sailing delay
- We model this by conditional BPA σ for S given $\{W, R\}$ such that

$$\begin{aligned} \sigma(\{(0, g, f), (1, b, f), (1, g, t), (2, b, t)\}) &= 0.9, \\ \sigma(\Omega_{\{S,A,R\}}) &= 0.1 \end{aligned}$$

- Notice that $\sigma^{\downarrow\{W,R\}}$ is vacuous for $\{W, R\}$.

Directed Graphical Models

Conditional for A given $\{D, S\}$:

- Consider the piece of knowledge: Arrival delay is the sum of departure delay and sailing delay
- We model this piece of knowledge by a deterministic conditional BPA α for A given $\{D, S\}$ such that

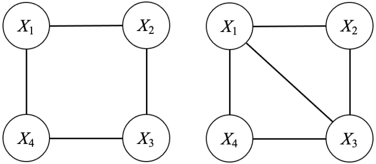
$$\begin{aligned}\alpha(\{&(0,0,0),(0,1,1),(0,2,2),(0,3,3),\\ &(1,0,1),(1,1,2),(1,2,3),(1,3,4),\\ &(2,0,2),(2,1,3),(2,2,4),(2,3,5),\\ &(3,0,3),(3,1,4),(3,2,5),(3,3,6)\}) = 1.\end{aligned}$$

- Notice that $\alpha^{\downarrow\{D,S\}}$ is vacuous for $\{D, S\}$.

Undirected Graphical Models

Notation:

- An undirected 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 undirected edges $\{X_i, X_j\}$ between two distinct variables in $\mathcal{V}$.
- A clique in G is a maximal completely connected subgraph of G.
- Given a variable $X \in \mathcal{V}$, the Markov blanket of X, denoted by $MB_G(X)$, is $\{Y \in \mathcal{V} : \{X, Y\} \in E\}$.



- The UG on the left has four cliques with node sets: $\{X_1, X_2\}$, $\{X_2, X_3\}$, $\{X_3, X_4\}$, $\{X_1, X_4\}$.
  $MB_G(X_1) = \{X_2, X_4\}$
- The UG on the right has two cliques with node sets: $\{X_1, X_2, X_3\}$, $\{X_1, X_3, X_4\}$.
  $MB_G(X_1) = \{X_2, X_3, X_4\}$.

Undirected Graphical Models

Definition (Belief-function undirected graphical model)

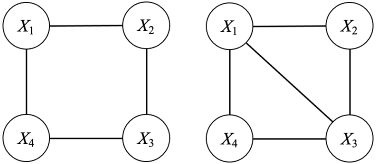
Suppose we have an undirected acyclic graph $G = (\mathcal{V}, \mathcal{E})$ with n nodes in $\mathcal{V}$ with cliques $r_1, \ldots r_k$. A belief-function undirected graphical model (BFUGM) is a pair $(G, \{m_1, \ldots, m_k\})$ such that m_i is a BPA for clique r_i. A fundamental assumption of a BFUGM 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

$$m = \bigoplus_{i=1}^{k} m_i. \tag{15}$$

1. The assumption in the definition that all conditionals are distinct allows the combination in Eq. (15).
2. Given m, the joint BPA for $\mathcal{V}$ as defined in Eq. (15), it follows from the definition of CI that the following CI relations hold. For each $X_i \in \mathcal{V}$, $X_i \perp\!\!\!\perp_m \mathcal{V} \setminus (\{X_i\} \cup MB_G(X_i)) \,|\, MB_G(X_i)$.

Undirected Graphical Models

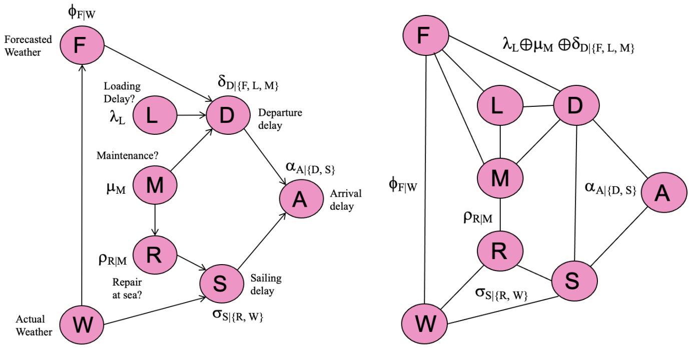
CI assumptions in BFUGMs:



- For the BFUGM on the left: $m = m_{12} \oplus m_{23} \oplus m_{34} \oplus m_{14}$. This BFUGM has two CI assumptions: $X_1 \perp\!\!\!\perp_m X_3 \,|\, \{X_2, X_4\}$, and $X_2 \perp\!\!\!\perp_m X_4 \,|\, \{X_1, X_3\}$. The first one follows from $m = (m_{12} \oplus m_{14}) \oplus (m_{23} \oplus m_{34})$. The second one follows from $m = (m_{12} \oplus m_{23}) \oplus (m_{34} \oplus m_{14})$.
- For the BFUGM on the right: $m = m_{123} \oplus m_{134}$ and 1 CI assumption: $X_2 \perp\!\!\!\perp_m X_4 \,|\, \{X_1, X_3\}$. This follows from $m = m_{123} \oplus m_{134}$.

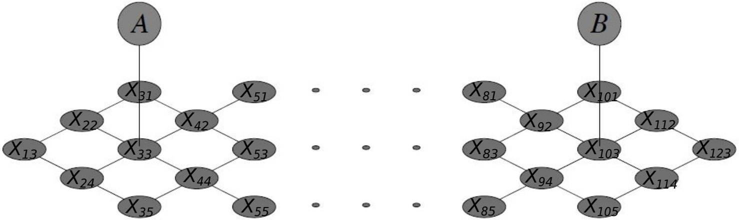
Undirected Graphical Models

- One source of undirected graphical models is the moralization of a directed graphical model (where we marry parents and drop directions) [Lauritzen & Spiegelhalter 1988].
- The BPAs associated with the cliques are the same as the conditionals associated with each variable or some combination.
- So, all BPAs associated with the cliques are distinct.



Undirected Graphical Models

- Communication network [Haenni-Lehmann 2002]
- We have a grid of $44 = 8 + 9 + 10 + 9 + 8$ communication nodes arranged in 12 columns and 5 rows
- There are 68 links, and each link has 90% reliability
- Nodes A and B are connected to the grid with links having 80% reliability
- What is the marginal of the joint for $\{A, B\}$?



Undirected Graphical Models

Definition (Non-informative BPAs)

Suppose m_1 and m_2 are two distinct BPAs for s_1 and s_2, respectively. We say m_1 and m_2 are mutually non-informative if $m_1^{\downarrow s_1 \cap s_2}$ and $m_2^{\downarrow s_1 \cap s_2}$ are vacuous BPAs for $s_1 \cap s_2$.

- Intuitively, m_1 doesn't tell us anything about m_2 and vice-versa.
- If s_1 and s_2 are disjoint, then they are trivially non-informative to each other.

Definition (A set of non-informative BPAs)

A set of BPAs is non-informative if every pair of BPA from the set is mutually non-informative.

Undirected Graphical Models

- Consider the variables in the grid with 19 columns and five rows. Let X_{13} denote the variable in column 1, row 3, and let X_{22} denote the variable in column 2 and row 2. Let $\Omega_{13} = \{t_{13}, f_{13}\}$, and and let $\Omega_{22} = \{t_{22}, f_{22}\}$.
- The BPA m_{13-22} associated with the edge between X_{13} and X_{22} is as follows:

$$
\begin{aligned}
m_{13-22}(\{(t_{13}, t_{22}), (f_{13}, f_{22})\}) &= 0.9, \\
m_{13-22}(\Omega_{13} \times \Omega_{24}) &= 0.1.
\end{aligned}
$$

- All BPAs in the model are similar to BPA m_{13-22}.
- BPAs m_{13-22} and m_{13-24} are mutually non-informative.
- The set of all BPAs in the communication network example is non-informative.
- Each BPA in this model models the reliability of the corresponding link between two nodes. Assuming the reliability of each link is independent of the reliabilities of other links, we can infer that all BPAs in the model are distinct.

Outline

Summary & Conclusions

- The main goal of this presentation is to discuss the notion of distinct belief functions in graphical models, both directed and undirected. We start with the definition given by Dempster in his multi-valued semantics of a BPA. In practice, this cannot be used as we don't associate a multi-valued function with each belief function in a model.
- We use heuristics of no double-counting of non-idempotent knowledge to define distinct belief functions.
- 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 if the conditional independence assumptions implied by the graphical model are valid.
- For a class of undirected graphical models, we have BPAs associated with each network clique with the same structure. For example, all BPAs have the same structure in the communication network example. Moreover, these BPAs are mutually non-informative. Thus, we can conclude that all BPAs in this example are distinct.

Questions

Questions?