On Distinct Belief Functions in the Dempster-Shafer Theory

Prakash P. Shenoy

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

July 11, 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 in Dempster's definition is no double-counting of non-idempotent knowledge.
- 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

Conditional Independence

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

Definition (Conditional Independence (Shenoy 1994))

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.
- The definition of CI above is similar to CI in probability theory [Dawid 1979].
- The definition of CI above satisfies the graphoid axioms [Paz and Pearl 1987]

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, 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.
- 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 2 reflects this fact.

Non-informative BPAs

Definition (Non-informative BPAs)

Suppose m_1 and m_2 are two 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 the domain of m_2 and vice-versa.
- If s_1 and s_2 are disjoint, then m_1 and m_2 are trivially mutually non-informative.

Definition (A non-informative set of BPAs)

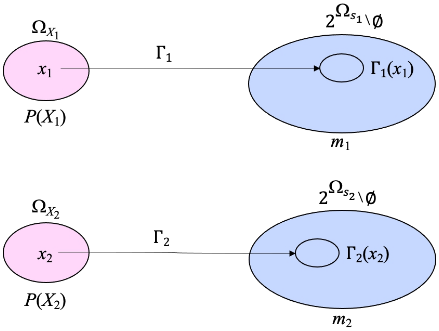
A set of BPAs is non-informative if every pair of BPA from the set is mutually non-informative.

- The definition of a non-informative set of BPAs gives us a sufficient condition for distinctness of BPAs in an undirected graphical model.

Outline

Dempster's Definition

Dempster's multi-valued semantics for belief functions:



- We have X_1 for which we have a PMF $P(X_1)$ and multivalued mapping $\Gamma_1 : \Omega_{X_1} \to 2^{S_1} \setminus \emptyset$ that results in a BPA m_1 for S_1.
- We have space X_2 for which we have a PMF $P(X_2)$ and multivalued mapping $\Gamma_2 : \Omega_{X_2} \to 2^{S_2} \setminus \emptyset$ 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.

Dempster's Definition

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)

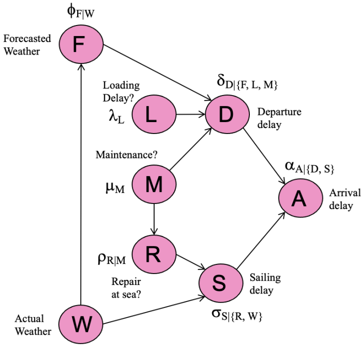
Suppose we have a directed acyclic graph $G = (\mathcal{V}, \mathcal{E})$ with n nodes in $\mathcal{V}$. A belief-function directed graphical model (BFDGM) is a pair $(G, \{m_1, \ldots, m_n\})$ such that BPA m_i associated with node X_i is a conditional BPA for X_i given $Pa_G(X_i)$, for $i = 1, \ldots,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

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

1. The assumption in the definition that all conditionals are distinct allows the combination in Eq. (1).
2. Given m, the joint BPA for $\mathcal{V}$ as defined in Eq. (1), 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, $i = 2, \ldots, n$, $X_i \perp\!\!\!\perp_m (\{X_1, \ldots X_{i-1}\} \setminus Pa_G(X_i)) | Pa_G(X_i)$.
3. $\{m_1, \ldots m_n\}$ are distinct if and only if all CI assumptions in the model are valid.

Directed Graphical Models

An Example: Captain's Problem

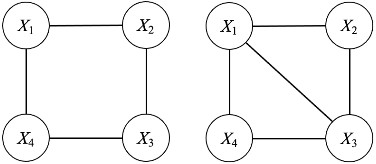


- A topological sequence: (W, F, L, M, D, R, S, A).
- CI assumptions: $L \perp\!\!\!\perp_m \{W, F\}$, $M \perp\!\!\!\perp_m \{W, F, L\}$, $D \perp\!\!\!\perp_m W \mid \{F, L, M\}$, $R \perp\!\!\!\perp_m \{W, F, L, D\} \mid M$, etc.

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

Undirected Graphical Models

Definition (Belief-function undirected graphical model)

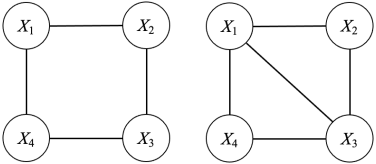
Suppose we have an undirected acyclic graph $G = (\mathcal{V}, \mathcal{E})$ with n nodes in $\mathcal{V}$ with cliques $r_1, \ldots r_k$. A belief-function undirected graphical model (BFUGM) is a pair $(G, \{m_1, \ldots, m_k\})$ such that m_i is a BPA for clique r_i. A fundamental assumption of a BFUGM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

$$m = \bigoplus_{i=1}^{k} m_i. \tag{2}$$

1. The assumption that the BPAs in the model are all distinct allows the Dempster's combination in Eq. (2).
2. Given m, the joint BPA for $\mathcal{V}$, it follows from Eq. (2) and the definition of CI that the following CI relations hold. For each $X_i \in \mathcal{V}$, $X_i \perp\!\!\!\perp_m \mathcal{V} \setminus (\{X_i\} \cup MB_G(X_i)) \,|\, MB_G(X_i)$.
3. BPAs $m_1, \ldots, m_k$ are distinct if $\{m_1, \ldots, m_k\}$ is non-informative and all CI assumptions in the model are valid. This condition is sufficient but not necessary.

Undirected Graphical Models

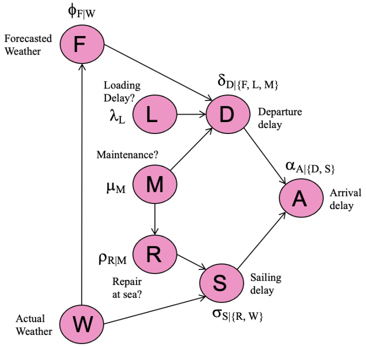
CI assumptions in BFUGMs:

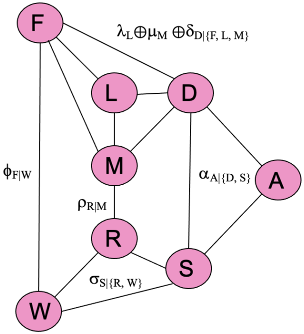


- For the BFUGM on the left: $m = m_{12} \oplus m_{23} \oplus m_{34} \oplus m_{14}$. This BFUGM has two CI assumptions: $X_1 \perp\!\!\!\perp_m X_3 \,|\, \{X_2, X_4\}$, and $X_2 \perp\!\!\!\perp_m X_4 \,|\, \{X_1, X_3\}$. The first one follows from $m = (m_{12} \oplus m_{14}) \oplus (m_{23} \oplus m_{34})$.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

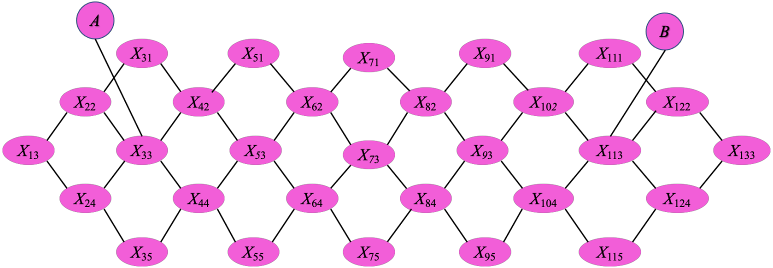
- One source of undirected graphical models is the moralization of a directed graphical model (where we marry parents and drop directions) [Lauritzen & Spiegelhalter 1988].
- The BPAs associated with the cliques are the same as the conditionals associated with each variable or some combination.
- So,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

- 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 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 if the CI assumptions of the model are valid.

Questions

Questions? Come to my poster...