

Calculating Dempster-Shafer Plausibility

Jeffrey A. Barnett (Currently jbb@notatt.com)

Northrop Research and Technology Center
One Research Park
Palos Verdes Peninsula, CA 90274

Abstract

A sufficient condition is developed for the equality of the plausibility and commonality measures of the Dempster-Shafer belief calculus. When the condition is met, as it is in many applications, an efficient method to calculate relative plausibility is available. In particular, the method can be used to calculate the relative plausibility of atomic hypotheses and, therefore, it can be used to find the choice that maximizes this measure. The computation is efficient enough to make Dempster-Shafer practical in some domains where computational complexity would otherwise counter-indicate its use.

1 Summary

The Dempster-Shafer belief calculus [7] is a theory of plausible reasoning because it provides methods to represent and combine weights of evidence. The measures **Bel** and **Pl** are used to assist decision making and play roles similar to that played by distribution functions in probability theory.

Bel is called a belief or lower probability function and **Pl** is called a plausibility or upper probability function. **Bel** is often 0 on all or most of the atomic hypotheses in complex domains unless a large number of evidence

sources are available. However, **PI** generally provides some discrimination even when the evidence is sparse. Therefore, **PI** is a more robust guide to decision making than is **Bel**.

Bel and **PI** computations can be prohibitively expensive unless assumptions about the structure of the evidence are available. Orponen [4] shows that these computations are $\#P$ -complete when evidence sources are combined. Shafer and Logan [6] and Gordon and Shortliffe [2] develop efficient algorithms for cases where bodies of evidence illuminate hypotheses that can be arranged in strict hierarchies; and Barnett [1] deals only with cases where bodies of evidence focus on atomic hypotheses and their complements.

In the following, an efficient method is developed to find values of **PI**(θ) up to a multiplicative constant. This method relies only on the assumption that θ be an hypothesis that is atomic with respect to the evidence, a concept made precise below. In particular, the method can be used with the atomic hypotheses of any domain and, therefore, it can find the choice maximizing **PI**.

Basic concepts and notations of the Dempster-Shafer theory are introduced in the next section. That material can be skipped by readers familiar with Shafer's book [7]. Subsequent sections develop methods to efficiently compute values proportional to **PI**.

2 The Dempster-Shafer Theory

The basic concepts and mechanisms of the Dempster-Shafer theory are introduced in this section. The introduction parallels Shafer [7]. The interested reader is referred to Hummel and Landy [3] and Schocken and Kleindorfer [5]

for additional discussion and alternative interpretations of the theory.

2.1 Representing Evidence

The Dempster-Shafer theory represents a problem domain by a set, Θ , of mutually exclusive and exhaustive *atomic* hypotheses. A subset of Θ is interpreted as the disjunction of its elements. If the function $\mathbf{m}: 2^\Theta \rightarrow [0, 1]$ satisfies $\mathbf{m}(\emptyset) = 0$ and

$$\sum_{S \subseteq \Theta} \mathbf{m}(S) = 1,$$

then \mathbf{m} is called a *basic probability assignment* over Θ . One can imagine that a mass of weight $\mathbf{m}(S)$ is attached to each $S \subseteq \Theta$ and that it is free to move anywhere in S . The weight of the mass stands for the strength of our belief in the proposition that S represents.

The function, $\mathbf{Bel}: 2^\Theta \rightarrow [0, 1]$, is derived, from the basic probability assignment \mathbf{m} , by

$$\mathbf{Bel}(\theta) = \sum_{S \subseteq \theta} \mathbf{m}(S),$$

for all $\theta \subseteq \Theta$. This \mathbf{Bel} is called the *belief function*. Another function, $\mathbf{Pl}: 2^\Theta \rightarrow [0, 1]$, is defined by $\mathbf{Pl}(\theta) = 1 - \mathbf{Bel}(\bar{\theta})$, where $\bar{\theta}$ is the complement of θ , i.e., $\bar{\theta} = \Theta - \theta$, and \mathbf{Pl} is called the *plausibility function*. Thus,

$$(1) \quad \mathbf{Pl}(\theta) = \sum_{S \cap \theta \neq \emptyset} \mathbf{m}(S).$$

Therefore, $\mathbf{Bel}(\theta)$ measures the total mass that is constrained to move within θ while $\mathbf{Pl}(\theta)$ measures the total mass that can visit somewhere within θ but may move outside as well. Obviously, $\mathbf{Bel}(\theta) \leq \mathbf{Pl}(\theta)$ for all $\theta \subseteq \Theta$ and,

hence, **Bel** and **Pl** are sometimes referred to, respectively, as the *lower* and *upper probability measures*.

The function, $\mathbf{Q}: 2^\Theta \rightarrow [0, 1]$, defined by

$$(2) \quad \mathbf{Q}(\theta) = \sum_{\theta \subseteq S \subseteq \Theta} \mathbf{m}(S),$$

for all $\theta \subseteq \Theta$, is called the *commonality function*. Thus, $\mathbf{Q}(\theta)$ measures the total mass that is free to move to every element of θ .

The mass movement metaphor makes it easy to see that \mathbf{Q} is a *decreasing* function while **Pl** and **Bel** are both *increasing*: if $S \subseteq T$, then $\mathbf{Q}(S) \geq \mathbf{Q}(T)$ but $\mathbf{Pl}(S) \leq \mathbf{Pl}(T)$ and $\mathbf{Bel}(S) \leq \mathbf{Bel}(T)$.

The function, \mathbf{m} , is akin to the density function of standard probability theory. The difference is that masses must be attached to singleton sets and, hence, have no freedom of motion. Therefore, Dempster-Shafer provides a generalization that is useful to represent the uncertainty that results from lack of specificity, i.e., belief can be invested in disjunctive propositions without splitting that belief among the atomic elements comprising the disjunction.

Both **Bel** and **Pl** play roles similar to distribution functions in the standard theory. There is no analogy to \mathbf{Q} and it is generally used as a technical device to simplify proofs of computational theorems.

Shafer [7] shows that, given any one of \mathbf{m} , \mathbf{Q} , **Pl**, or **Bel**, the other three can be calculated: Formula 2 defines \mathbf{Q} in terms of \mathbf{m} and

$$(3) \quad \mathbf{Pl}(\theta) = \sum_{\substack{S \subseteq \theta \\ S \neq \emptyset}} (-1)^{|S|+1} \mathbf{Q}(S)$$

recovers **Pl** from \mathbf{Q} . Further, the fact that $\mathbf{Bel}(\theta) = 1 - \mathbf{Pl}(\bar{\theta})$ follows directly

from the definition of **PI** and

$$\mathbf{m}(\theta) = \sum_{S \subseteq \theta} (-1)^{|\theta-S|} \mathbf{Bel}(S)$$

recovers **m** from **Bel**. Therefore, the cycle from **m** to **Q** to **PI** to **Bel** and back to **m** is complete and, therefore, there is a 1-to-1 correspondence between members of the four function families.

A notational convention is used where functions with the same subscript represent the same underlying evidence, e.g., **m_i**, **Q_i**, **PI_i**, and **Bel_i** all are defined in terms of one another. Further, a body of evidence in any of the four forms may be called a belief function since **Bel** is uniquely determined and can be recovered.

2.2 Combining Evidence

A theory of plausible reasoning must provide a method to combine bodies of evidence in addition to ways to represent them. In the Dempster-Shafer theory, this role is played by *Dempster's rule of combination* which is also called the *orthogonal sum*. If the **m_i**, for $1 \leq i \leq m$, are basic probability assignments, their orthogonal sum, **m**, satisfies $\mathbf{m}(\emptyset) = 0$ and, for all non-empty $\theta \subseteq \Theta$,

$$(4) \quad \mathbf{m}(\theta) = K \sum_{\substack{S_1, \dots, S_m \\ \bigcap S_i = \theta}} \prod_{1 \leq i \leq m} \mathbf{m}_i(S_i),$$

where

$$(5) \quad K = \left(\sum_{\substack{S_1, \dots, S_m \\ \bigcap S_i \neq \emptyset}} \prod_{1 \leq i \leq m} \mathbf{m}_i(S_i) \right)^{-1}.$$

The orthogonal sum does not exist when the combined belief function attaches a zero-weight mass to every set. In this case, the separate bodies of

evidence are said to be *totally* or *flatly contradictory* and it is not possible to combine them. When the evidence is combinable, i.e., K exists, the orthogonal sum can be shown to be a belief function. That belief function is calculated by

$$(6) \quad \mathbf{Q}(\theta) = K \prod_{1 \leq i \leq m} \mathbf{Q}_i(\theta).$$

This alternative form of Dempster's rule is used below. N.B., In general, a computation that is exponential in $|\Theta|$ may result if formula 3 is composed with formula 6 to calculate **Pl**.

When the \mathbf{m}_i are combinable, the notation $\mathbf{m} = \mathbf{m}_1 \oplus \cdots \oplus \mathbf{m}_m$ is used for the orthogonal sum. The \oplus operator is commutative and associative. It also has an identity element, a basic probability assignment that attaches a mass of 1 to Θ and a mass of 0 to every proper subset of Θ .

2.3 Focal Element

The formulas above involving basic probability assignments can be rewritten, without effect, to restrict summations to only those sets for which $\mathbf{m}(S) \neq 0$ (or $\mathbf{m}_i(S_i) \neq 0$). That motivates defining $\mathcal{F}_i = \{F \mid \mathbf{m}_i(F) \neq 0\}$ and calling its elements *focal elements* of the belief function \mathbf{m}_i . The set

$$(7) \quad \mathcal{C} = \bigcap_{1 \leq i \leq m} \bigcup_{F \in \mathcal{F}_i} F,$$

is the *core* of the combined belief function. It is empty if and only if K (formula 5) does not exist, in which case the \mathbf{m}_i are flatly contradictory.

3 Decision-Making With **PI**

The typical problem-solving scenario using the Dempster-Shafer belief calculus is to gather bodies of evidence and represent them as belief functions, combine these belief functions with Dempster’s rule, and then select the hypothesis best supported by the combined evidence—**Bel** and **PI** are the measures provided by the theory to make the selection. Henceforth, the \mathbf{m}_i are the bodies of evidence and $\mathbf{m} = \mathbf{m}_1 \oplus \cdots \oplus \mathbf{m}_m$ is their combination.

If $h \in \Theta$, then $\mathbf{Bel}(\{h\}) = \mathbf{m}(\{h\})$ and formula 4 shows that $\mathbf{m}(\{h\}) = 0$ unless there exist a $F_i \in \mathcal{F}_i$, for each $1 \leq i \leq m$, such that $\cap F_i = \{h\}$. Thus, the possibility that $\mathbf{Bel}(\{h\}) \neq 0$ increases when (1) the evidence bodies are rich—the cardinality of the \mathcal{F}_i are large and/or (2) the number of evidence sources is large. The evidence must become richer and/or more plentiful as $|\Theta|$ increases, which rarely happens, or $\mathbf{Bel}(\{h\}) = 0$ for most $h \in \Theta$.

PI is not trivially zero as often as **Bel** because $\mathbf{PI}(\theta) \neq 0$ if and only if $\theta \cap \mathcal{C} \neq \emptyset$ and, therefore, $\mathbf{PI}(\{h\}) \neq 0$ for all $h \in \mathcal{C}$. However, $h \in \mathcal{C}$ is necessary but not sufficient to show that $\mathbf{Bel}(\{h\}) \neq 0$. Thus, **PI** is generally a more discriminating measure than is **Bel** to determine problem solution.

Another pragmatic reason to consider **PI** for decision making is computational complexity. Unless the evidence has a special structure that can be exploited, finding the $h \in \Theta$ that maximizes $\mathbf{Bel}(\{h\})$ necessitates doing $O(\prod |\mathcal{F}_i|)$ operations. On the other hand, the results developed below find the $h \in \Theta$ that maximizes $\mathbf{PI}(\{h\})$ with $O(|\Theta| \cdot \sum |\mathcal{F}_i|)$ operations and this is often a considerable improvement.

4 Atomic Hypotheses

Let $\mathcal{E} = \{\mathbf{m}_i \mid 1 \leq i \leq m\}$ be the evidence available to be combined. Define the predicate $\mathbf{At}_{\mathcal{E}}(\theta)$ to be true if and only if (1) $\theta \neq \emptyset$ and (2) for all $F \in \mathcal{F}_i$, where $1 \leq i \leq m$, either $\theta \subseteq F$ or $\theta \cap F = \emptyset$. When $\mathbf{At}_{\mathcal{E}}(\theta)$, we will say that θ is *atomic with respect to the evidence, \mathcal{E}* .

As an example, let $\Theta = \{abcde\}$ and assume that the focal elements of \mathbf{m}_1 are $\{ab\}$ and $\{abc\}$ and the focal elements of \mathbf{m}_2 are $\{ce\}$ and $\{abcde\}$. Then $\mathcal{E} = \{\mathbf{m}_1 \mathbf{m}_2\}$, and $\{a\}$, $\{b\}$, $\{c\}$, $\{d\}$, $\{e\}$, and $\{ab\}$ are the only subsets of Θ that are atomic with respect to \mathcal{E} . Atoms, (elements of Θ), as in this example, are always atomic with respect to the evidence.

Lemma 1 $\mathbf{At}_{\mathcal{E}}(\{h\})$ for all $h \in \Theta$.

Proof: This is true because a set that has only one element is either a subset of or disjoint from any other set. Since the $F \in \mathcal{F}_i$ are sets, the lemma follows from the definition of $\mathbf{At}_{\mathcal{E}}$. \square

Therefore, atomic hypotheses are always atomic with respect to the evidence; no additional assumptions about the structure of \mathcal{E} are necessary. The theorem to be proved below is that $\mathbf{Pl}(\theta) = \mathbf{Q}(\theta)$ when $\mathbf{At}_{\mathcal{E}}(\theta)$ and, consequently, an efficient polynomial-time version of formula 6 can be used to calculate relative values of \mathbf{Pl} .

The idea behind calling a hypothesis atomic is that it be minimal in some theory-related way. For example, Tarski [8] defines the atoms of a boolean algebra to be minimal in the sense that if a is atomic, no hypotheses except \emptyset and a , imply it. He then shows that a is atomic if and only if a has the property that for any element, b , of the algebra, either a implies b or a implies $\neg b$.

In the current investigation, with subsets of Θ interpreted as disjunctions of their elements, it is easy to see that $\mathbf{At}_{\mathcal{E}}(\theta)$ if and only if $\theta \models F$ or $\theta \models \neg F$, where F is any focal element of any member of \mathcal{E} , because $\theta \models F$ corresponds to $\theta \subseteq F$ and $\theta \models \neg F$ corresponds to $\theta \cap F = \emptyset$ which implies that $\theta \subseteq \overline{F}$. However, the concept of atomic with respect to the evidence includes hypotheses that can be interpreted as disjunctions while Tarski's idea of atomic does not.

Lemma 2 *If $\mathbf{At}_{\mathcal{E}}(\theta)$ and non-empty $S \subseteq \theta$, then $\mathbf{At}_{\mathcal{E}}(S)$.*

Proof: $\mathbf{At}_{\mathcal{E}}(\theta)$ means that either $\theta \subseteq F$ or $\theta \cap F = \emptyset$, where F is an arbitrary focal element of some \mathbf{m}_i . But $S \subseteq \theta$ implies that, respectively, $S \subseteq F$ or $S \cap F \neq \emptyset$ (because $S \neq \emptyset$) and, therefore, $\mathbf{At}_{\mathcal{E}}(S)$. \square

Below it is shown that $\mathbf{Pl}(S) = \mathbf{Pl}(\theta)$ when $S \subseteq \theta$ and $\mathbf{At}_{\mathcal{E}}(\theta)$. Thus, “ $\mathbf{At}_{\mathcal{E}}(\theta)$ ” means that the available evidence cannot distinguish the hypothesis θ , via \mathbf{Pl} , from the sharper ones that imply it.

Before passing on to the derivation of the main results, note that the predicate $\mathbf{At}_{\mathcal{E}}$ depends on \mathcal{E} , not just the orthogonal sum of its elements. Let $\mathcal{E} = \{\mathbf{m}_i \mid 1 \leq i \leq m\}$, $\mathbf{m} = \mathbf{m}_1 \oplus \cdots \oplus \mathbf{m}_m$, and $\mathcal{M} = \{\mathbf{m}\}$. It is straightforward to show that $\mathbf{Pl}(\theta) = \mathbf{Q}(\theta)$ if and only if $\mathbf{At}_{\mathcal{M}}(\theta)$. Further, $\mathbf{At}_{\mathcal{E}}(\theta)$ implies $\mathbf{At}_{\mathcal{M}}(\theta)$ but, in general, the converse is not true unless $\mathbf{Pl}(\theta) = 0$.

5 Pl Theorems

The computations developed in this section find values of \mathbf{Pl} up to a multiplicative constant. The main idea is to show that $\mathbf{Pl}(\theta) = \mathbf{Q}(\theta)$ for many

interesting $\theta \subseteq \Theta$. When this is possible, values proportional to \mathbf{PI} can be calculated by an elaboration of formula 6 that ignores K . Thus, if $\theta \subseteq \Theta$ is one of the sets for which this technique works,

$$\begin{aligned}
(8) \quad \mathbf{PI}(\theta) &= \mathbf{Q}(\theta) \\
&= K \times \prod_{1 \leq i \leq m} \mathbf{Q}_i(\theta) \\
&\propto \prod_{1 \leq i \leq m} \mathbf{Q}_i(\theta) \\
&\propto \prod_{1 \leq i \leq m} \sum_{\theta \subseteq S \subseteq \Theta} \mathbf{m}_i(S) \\
&\propto \prod_{1 \leq i \leq m} \sum_{\substack{F \in \mathcal{F}_i \\ \theta \subseteq F}} \mathbf{m}_i(F)
\end{aligned}$$

where “ \propto ” means proportional to and formula 2 has been substituted in formula 6. A slight specialization for the hypotheses $h \in \Theta$ is

$$(9) \quad \mathbf{PI}(\{h\}) \propto \prod_{1 \leq i \leq m} \sum_{\substack{F \in \mathcal{F}_i \\ h \in F}} \mathbf{m}_i(F).$$

Next, several results are developed that enable efficient computation of \mathbf{PI} up to a multiplicative constant. Theorem 1 shows the equality of the commonality and plausibility functions when θ is atomic with respect to \mathcal{E} . The rest of the results are simple corollaries of this theorem.

Theorem 1 *If $\mathbf{At}_{\mathcal{E}}(\theta)$, then $\mathbf{PI}(\theta) = \mathbf{Q}(\theta)$.*

Proof: The given, $\mathbf{At}_{\mathcal{E}}(\theta)$, along with $\mathbf{m}(S) \neq 0$ is shown to imply that $S \cap \theta \neq \emptyset$ if and only if $\theta \subseteq S$. Therefore, formulas 1 and 2 yield $\mathbf{PI}(\theta) = \mathbf{Q}(\theta)$ because the summations in both formulas add up the same set of non-zero values. If $\theta \subseteq S$, then $S \cap \theta = \theta \neq \emptyset$ follows because $\theta \neq \emptyset$ from $\mathbf{At}_{\mathcal{E}}(\theta)$.

The proof will be complete if it can be shown that $\theta \subseteq S$ when $S \cap \theta \neq \emptyset$, $\mathbf{m}(S) \neq 0$, and $\mathbf{At}_{\mathcal{E}}(\theta)$. Since $\mathbf{m}(S) \neq 0$, formula 4 implies that there must

exist a $F_i \in \mathcal{F}_i$ for each $1 \leq i \leq m$, such that $S = \cap F_i$ and, therefore, that $(\cap F_i) \cap \theta \neq \emptyset$ because $S \cap \theta \neq \emptyset$ is assumed. Thus, $F_i \cap \theta \neq \emptyset$ for each i . Since $\mathbf{At}_{\mathcal{E}}(\theta)$ means that either $\theta \subseteq F_i$ or $\theta \cap F_i = \emptyset$ —the latter is a contradiction—it follows that $\theta \subseteq F_i$. Therefore, $\theta \subseteq \cap F_i = S$. \square

Two simple corollaries are immediately available that justify the computational formulas introduced at the beginning of the section. The second result was noted by Shafer [7, page 222].

Corollary 1 *Formula 8 is a valid computation if $\mathbf{At}_{\mathcal{E}}(\theta)$.*

Proof: Formula 8 was derived by assuming that $\mathbf{PI}(\theta) = \mathbf{Q}(\theta)$, as is justified by the theorem, and that K is independent of θ , as is justified by inspection of formula 5. \square

Corollary 2 *Formula 9 is a valid computation for all $h \in \Theta$.*

Proof: Lemma 1 shows that $\mathbf{At}_{\mathcal{E}}(\{h\})$ and, therefore, formula 8 is valid. Formula 9 only assumes, in addition, that $|\{h\}| = 1$. \square

The next two corollaries show that Θ is partitioned into *maximal* sets that are atomic with respect to the evidence. If θ_1 and θ_2 are subsets of the same maximal set, then $\mathbf{PI}(\theta_1) = \mathbf{PI}(\theta_2)$. It is also easy to show that $\mathbf{PI}(\theta) = \mathbf{Q}(\theta)$ if and only if either $\mathbf{At}_{\mathcal{E}}(\theta)$ or $\mathbf{PI}(\theta) = 0$.

Corollary 3 *If non-empty $S \subseteq \theta$ and $\mathbf{At}_{\mathcal{E}}(\theta)$, then $\mathbf{PI}(S) = \mathbf{PI}(\theta)$.*

Proof: Since \mathbf{PI} is increasing and \mathbf{Q} is decreasing, $\mathbf{PI}(S) \leq \mathbf{PI}(\theta)$ and $\mathbf{Q}(S) \geq \mathbf{Q}(\theta)$. Further, $\mathbf{PI}(\theta) = \mathbf{Q}(\theta)$ since $\mathbf{At}_{\mathcal{E}}(\theta)$. Thus,

$$\mathbf{PI}(S) \leq \mathbf{PI}(\theta) = \mathbf{Q}(\theta) \leq \mathbf{Q}(S)$$

But Lemma 2 proves that $\mathbf{At}_{\mathcal{E}}(S)$ follows from this corollary's assumptions and, hence, $\mathbf{Pl}(S) = \mathbf{Q}(S)$. Therefore, $\mathbf{Pl}(S) = \mathbf{Pl}(\theta)$. \square

Corollary 4 *If $\mathbf{At}_{\mathcal{E}}(\theta_1)$, $\mathbf{At}_{\mathcal{E}}(\theta_2)$, and $\theta_1 \cap \theta_2 \neq \emptyset$, then $\mathbf{Pl}(\theta_1) = \mathbf{Pl}(\theta_2)$.*

Proof: Let $\theta = \theta_1 \cap \theta_2$ and note that $\theta \neq \emptyset$ by hypothesis. From the previous corollary, $\mathbf{Pl}(\theta) = \mathbf{Pl}(\theta_1)$ and $\mathbf{Pl}(\theta) = \mathbf{Pl}(\theta_2)$ because, respectively, $\theta \subseteq \theta_1$ and $\theta \subseteq \theta_2$. Therefore, $\mathbf{Pl}(\theta_1) = \mathbf{Pl}(\theta_2)$. \square

The next result justifies an efficient computation for relative plausibility when the elements of \mathcal{E} are simple support functions: \mathbf{m}_i is a *simple support function* if $\mathcal{F}_i = \{F_i, \Theta\}$ or $\mathcal{F}_i = \{F_i\}$ for some non-empty $F_i \subset \Theta$. Therefore, if \mathbf{m}_i is a simple support function, there exists an $0 < s_i \leq 1$ such that $\mathbf{m}_i(F_i) = s_i$ and $\mathbf{m}_i(\Theta) = 1 - s_i$. For the θ such that $\mathbf{At}_{\mathcal{E}}(\theta)$, relative values of \mathbf{Pl} are calculated by

$$(10) \quad \mathbf{Pl}(\theta) \propto \prod_{\substack{1 \leq i \leq m \\ \theta \not\subseteq F_i}} (1 - s_i).$$

Corollary 5 *Formula 10 is valid if the elements of \mathcal{E} are simple support functions and $\mathbf{At}_{\mathcal{E}}(\theta)$.*

Proof: Since $\mathbf{At}_{\mathcal{E}}(\theta)$, the use of formula 8 is justified by corollary 1. Further, for each i , either $\theta \subseteq F_i$ or $\theta \cap F_i \neq \emptyset$. In the first case, the sum of the \mathbf{m}_i over $F \in \mathcal{F}_i$, in formula 8 is 1 because θ is a subset of every focal element and, thus, a unit factor can be removed from the product. In the second case, $\theta \cap F_i = \emptyset$, and the sum of the \mathbf{m}_i is just $\mathbf{m}_i(\Theta) = 1 - s_i$. \square

The importance of this corollary lies in the fact that virtually every Dempster-Shafer application reported or analyzed in the literature uses simple support functions to represent the available bodies of evidence.

6 Maximizing \mathbf{PI}

Corollary 2 justifies the use of formula 9 as the basis of an efficient algorithm to find the $h \in \Theta$ that maximizes $\mathbf{PI}(\{h\})$. The algorithm, if necessary, checks to see if \mathcal{C} , computed by formula 7, is empty, and hence, if the evidence combination fails to exist.

The computational complexity and the best algorithms to calculate \mathcal{C} and find the maximizing h can depend on details of the representation chosen for belief functions and sets. In the following, simply assume that each \mathbf{m}_i is represented as a list of pairs: there is a pair of the form $\{F, \mathbf{m}_i(F)\}$ for each $F \in \mathcal{F}_i$.

Given this representation, it is clear that formula 7 can evaluate \mathcal{C} with $O(\sum |\mathcal{F}_i|)$ set union operations, $O(|\mathcal{E}|)$ set intersection operations, and some arithmetic. It is also clear that formula 9 can be used to calculate a value proportional to $\mathbf{PI}(\{h\})$ by performing $O(\sum |\mathcal{F}_i|)$ set membership tests and some arithmetic. Therefore, the maximizing $h \in \Theta$ can be found in $O(|\Theta| \cdot \sum |\mathcal{F}_i|)$ operations by doing $|\Theta|$ evaluations and selecting the best one.

7 Conclusion

The ability to use a theory is determined in part by our ability to do the computations that it entails. Since the Dempster-Shafer belief calculus is known to involve $\#P$ -complete calculations, general use of the theory is restricted to cases where structural information enables efficient short cuts or acceptable approximations are available.

The results here provide a simple and efficient method to make decisions with \mathbf{PI} even though the values of that function are only determined to within

a multiplicative constant. However, that is sufficient to enable the theory to be used in many applications.

Acknowledgment: I wish to thank D. W. Geiger and D. B. Hunter for many valuable suggestions that have improved the readability and accuracy of this paper.

References

- [1] Barnett, J.A., Computational methods for a mathematical theory of evidence, in: *Proceedings IJCAI-81*, Vancouver, BC (1981) 868–875.
- [2] Gordon, J., and Shortliffe, E.H., A method for managing evidential reasoning in a hierarchical hypothesis space, *Artificial Intelligence* **26** (1985) 323–357.
- [3] Hummel, R.A., and Landy, M.S., A statistical viewpoint on the theory of evidence, *IEEE PAMI* **10** (1988) 235–247.
- [4] Orponen, P., Dempster’s rule of combination is #P-complete, *Artificial Intelligence* **44** (1990) 245–253.
- [5] Schocken, S., and Kleindorfer, P.R., Artificial intelligence dialects of the Bayesian belief revision language, *IEEE SMC* **19** (1989) 1106–1121.
- [6] Shafer, G., Logan, R., Implementing Dempster’s rule for hierarchical evidence, *Artificial Intelligence* **33** (1987) 271–298.

- [7] Shafer, G., *A Mathematical Theory of Evidence* (Princeton University Press, Princeton, NJ, 1976).
- [8] Tarski, A., On the foundations of Boolean algebra, Chapter XI; in *Logic, Semantics, Meta-Mathematics* (Tr. J.H. Woodger) (Oxford at the Clarendon Press, 1969).