A Similarity Approach to Evidence Combination in Dempster-Shafer Theory

J. Lawry
Department of Engineering Mathematics,
University of Bristol,
Bristol BS8 1TR, UK
Email j.lawry@bris.ac.uk

J. Recasens
Secció Matemàtiques i Informàtica - ETSAV
Univ. Politècnica de Catalunya,
Pere Serra 1-15. Sant Cugat del Vallès.
Barcelona. Spain.
Email recasens@upc.edu

I. González Rodríguez
Departmento de Informática,
Universidad de Oviedo,
Ed. Departmental 1, Campus de Viesques,
33271 Gijon, Spain,
Email inesgr@uniovi.es

Abstract

Combination operators for fusing different sources of evidence are investigated from the perspective of maximising proximity or similarity to a consistent solution. Operators are defined as the composition of an extension function and a normalisation method where the former identifies a unique joint belief assignment from two marginals and the latter redistributes mass associated with the empty set to other pairs of focal sets. This definition motivates an initial study into normalisation methods followed by a related examination of extension functions. The normalisation of a joint belief assignment is taken to be the closest or most similar consistent assignment according to a predefined metric. A range of distance and similarity metrics are considered and their associated normalisation methods are then identified. Extension functions are based on identifying the joint belief assignment with the required marginals that are closest to a consistent assignment. In effect this means finding the assignments which are closest to their normalised assignment. Issues of uniqueness are discussed and a number of operators are identified on the basis of extension functions and normalisation methods derived from particular metrics. Finally, we attempt to justify this approach by arguing that the very decision to intersect two sources of evidence carries with it an implicit assumption of consistency, rather than an assumption of independence as in Dempster's rule.

Keywords Evidence combination, closest consistent solution, distance metric, similarity measure, normalisation method, extension function

1 Introduction

In Dempster-Shafer theory the fundamental mechanism for combining information from different sources is through Dempster's rule. In fact this operator plays a central role in the whole theory as it also underlies the definitions of conditional belief and plausibility, these notions being fundamental to belief updating. Dempster's rule has a number of advantages, these being mainly linked to a range of desirable algebraic properties. For example, commutativity and associativity mean that when fusing evidence from multiple sources the order in which the operation is performed does not change the outcome. Consequently, the combination of evidence from a large number of sources can be carried out recursively, greatly improving computational efficiency. Also, the computational cost of a single application of the rule is negligible since no optimisation is involved. These pragmatic considerations are emphasised by Haenni [7] in his critique of a recent paper proposing an alternative weighted combination operator [13]. However, despite such advantages Dempster's rule has faced sustained criticism over the years. For instance, Zadeh [22] provides a now famous example illustrating that if the two sources of information are highly conflicting the application of Dempster's rule can lead to counter intuitive results. Another well known problem relates to the failure of Dempster's rule to satisfy idempotence; i.e. combining two identical basic belief assignments can yield a different assignment. This would seem counter intuitive especially when it is observed that the independence assumption of Dempster's rule may have resulted in inconsistencies which were then subsequently removed. If the two sources of evidence provide identical information what could be the possible source of such inconsistency? Indeed more generally it is not necessarily the case when combining two (possibly) consistent sources of evidence that Dempster's rule will identify a consistent solution. The generation of such inconsistency when sources may be consistent clearly results from the independence assumption underlying the rule and we shall subsequently argue that the decision to intersect two pieces of information has associated with it an implicit assumption of consistency rather than one of independence. Certainly, as emphasised by Voorbraak [20] the independence assumption of Dempster's rule is very strong and may often be unrealistic in practice. These issues will be discussed in more detail in section 4.

A number of alternative operators have been proposed in the literature (e.g. [2], [4], [21]) in order to address some of the criticisms of Dempster's rule. Many of these provide alternative mechanisms for eliminating inconsistency to overcome the difficulties highlighted by Zadeh [22]. For example, Yager [21] suggests that mass associated with pairs of elements with empty intersection should be reallocated to the whole universe, whereas Dubois and Prade [4] propose that such mass should be allocated to the union of the focal sets involved. On the other hand, Baldwin [2] and more recently Cattaneo [5] have proposed alternatives to the independence assumption of Dempster's rule. More specifically, Baldwin suggests that the family of probability distributions induced by the belief and plausibility measures of the combined basic belief assignment should correspond to the intersection of the two families of probabilities generated by the two basic belief assignments that are being combined. This is a rather natural idea but unfortunately turns out to be problematic as a means of defining a combination operator since the intersection of two families of distributions both generated from basic belief assignments is not necessarily representable in terms of belief and plausibility measures. Cattaneo has proposed that the combination operator should return the least specific assignment that minimizes inconsistency between the two sources. This is justified with regard to a monotonicity principle whereby if possible the combined belief measure should exceed both belief measures from the two evidence sources. In this paper we shall propose an alternative perspective on combination operators based on measures of similarity to (or distance from) a consistent solution. To give a clearer explanation of the proposed methodology it is useful to consider in more detail the structure of Dempster's rule.

As an operator Dempster's rule can be viewed as the composition of two distinct but related operations. Given basic belief assignments m_1 and m_2, the first step is to apply an extension function in order to identify a unique joint belief assignment with m_1 and m_2 as marginals. In Dempster's rule the joint belief assignment selected is $m_1 \times m_2$. Given this joint belief assignment the second step is to remove inconsistencies through a normalisation process whereby joint mass allocated to pairs of focal sets with empty intersection is reallocated in some way to the focal set pairs with non-empty intersection. In Dempster's rule this normalisation process involves simply rescaling the pairs of focal sets so that they sum to one. Once a normalised joint belief assignment is obtained the combined belief assignment is formed by aggregating over all possible non-empty sets that result from intersecting focal sets of m_1 and m_2. Now clearly there are alternatives to both the extension function and the normalisation method of Dempster's rule. In this paper we investigate normalisation methods which identify the most similar consistent solution according to some distance metric (or similarity measure). Intuitively the idea is that normalisation is a form of error correction, the overall effects of which it is therefore desirable to minimize. On this basis we will argue that the extension function should identify a joint belief assignment for which normalisation has a minimal effect. Overall this is equivalent to the assumption that the extension function should select an assignment most similar to a consistent assignment. Given this approach it is natural for us to deal initially with the issue of normalisation and then consider extension functions. In this respect we shall investigate the types of normalization methods and subsequent extension functions resulting from a range of different distance metrics and similarity measures. Initially, however, we introduce the basic notation and concepts fundamental to this treatment of combination operators.

2 Notation and Fundamental Concepts

Following the terminology of Smets [18]. Let Ω denote the set of all possible states of the world. We then assume that there is a true but unknown state of the world $w^* \in \Omega$ about which beliefs can be formulated on the basis of evidence.

Definition 1. *Basic Belief Assignment*
A basic belief assignment is a function $m : 2^\Omega \to [0,1]$ *such that* $m(\emptyset) = 0$ *and* $\sum_{S \subseteq \Omega} m(S) = 1$

Intuitively $m(S)$ corresponds to the level of belief that can be associated with the constraint $w^* \in S$ but which cannot then be allocated to more precise constraints $w^* \in T$ for $T \subset S$. In other words, $m(S)$ is the level of belief that the exact constraint representing the current state of knowledge regarding w^* is '$w^* \in S$'. For any basic belief assignment we can restrict our attention to those subsets of Ω, referred to as focal sets, which have non-zero mass.

Definition 2. *Focal Sets*
The focal sets of a basic belief assignment m *is given by:*

$$\mathcal{F} = \{F \subseteq \Omega : m(F) > 0\}$$

Given a basic belief assignment we can define lower and upper measures quantifying the level of belief in the assertion that $w^* \in S$ for $S \subseteq \Omega$ as follows:

Definition 3. *Belief and Plausibility Measures*

$$Bel(S) = \sum_{F \in \mathcal{F} : F \subseteq S} m(F) \ \text{ and } Pl(S) = \sum_{F \in \mathcal{F} : F \cap S \neq \emptyset} m(F)$$

Now suppose that we have two sources of information regarding w^* represented by basic belief assignments m_1 and m_2. We then define a joint belief assignment with marginal assignments m_1 and m_2 as follows:

Definition 4. *Joint Belief Assignment with Marginals* m_1 *and* m_2
Let $m_1 : \mathcal{F} \to [0,1]$ *and* $m_2 : \mathcal{G} \to [0,1]$ *be basic belief assignments on* 2^Ω *with focal sets* $\mathcal{F} \subseteq 2^\Omega$ *and* $\mathcal{G} \subseteq 2^\Omega$ *respectively. Then a joint belief assignment with marginals* m_1 *and* m_2 *is a function* $\underline{m} : \mathcal{F} \times \mathcal{G} \to [0,1]$ *satisfying:*

$$\forall F \in \mathcal{F} \ \sum_{G \in \mathcal{G}} \underline{m}(F,G) = m_1(F) \ \text{ and }$$

$$\forall G \in \mathcal{G} \ \sum_{F \in \mathcal{F}} \underline{m}(F,G) = m_2(G)$$

Intuitively, $\underline{m}(F,G)$ quantifies the level of belief of some third agent (carrying out the combination) that the evidence from source one supports the exact state of knowledge being '$w^* \in F$' while that from source two supports the exact state of knowledge being '$w^* \in G$'. In the following definition we introduce a number of sets of joint belief assignments defined on the cross product space $\mathcal{F} \times \mathcal{G}$:

Definition 5. *We introduce the following sets of joint belief assignments:*

(i) Set of joint belief assignments on $\mathcal{F} \times \mathcal{G}$.

$$\mathcal{JM} = \left\{ \underline{m} : \mathcal{F} \times \mathcal{G} \to [0,1] : \sum_{F \in \mathcal{F}} \sum_{G \in \mathcal{G}} \underline{m}(F,G) = 1 \right\}$$

(ii) Set of joint belief assignments on $\mathcal{F} \times \mathcal{G}$ consistent with the intersection operation.

$$\mathcal{CJM} = \{ \underline{m} \in \mathcal{JM} : \forall (F,G) \in \mathcal{F} \times \mathcal{G} \textit{ where } F \cap G = \emptyset \textit{ then } \underline{m}(F,G) = 0 \}$$

(iii) Set of joint belief assignments on $\mathcal{F} \times \mathcal{G}$ with marginals m_1 and m_2.

$$\mathcal{JM}_{1,2} = \left\{ \underline{m} \in \mathcal{JM} : \forall F \in \mathcal{F} \sum_{G \in \mathcal{G}} \underline{m}(F,G) = m_1(F) \textit{ and } \forall G \in \mathcal{G} \sum_{F \in \mathcal{F}} \underline{m}(F,G) = m_2(G) \right\}$$

Definition 5 gives us a way of explicitly defining consistency between assignments whereby two assignments m_1 and m_2 are said to be consistent provided $\mathcal{JM}_{1,2} \cap \mathcal{CJM} \neq \emptyset$.

We can define volumes corresponding to convex subsets of $[0,1]^k$ where $k = |\mathcal{F}| \times |\mathcal{G}|$ to represent $\mathcal{JM}$, $\mathcal{CJM}$ and $\mathcal{JM}_{1,2}$ geometrically as follows:

Definition 6. *Volume Representations*
Let $\mathcal{F} \times \mathcal{G} = \{H_i : i = 1, \ldots, k\}$ be some enumeration of the joint focal elements. Then we define:

$$V(\mathcal{JM}) = \left\{ \langle p_1, \ldots, p_k \rangle \in [0,1]^k : \sum_{i=1}^{n} p_i = 1 \right\}$$

$$V(\mathcal{CJM}) = \{ \langle q_1, \ldots, q_k \rangle \in V(\mathcal{JM}) : q_i = 0 \textit{ if } H_i = (F,G) \textit{ and } F \cap G = \emptyset \}$$

W.l.o.g we assume that $H_i : i = 1, \ldots, k$ are enumerated such that $\{(F,G) : F \cap G = \emptyset\} = \{H_1, \ldots, H_c\}$ in which case:

$$V(\mathcal{CJM}) = \{ \langle q_1, \ldots, q_k \rangle \in V(\mathcal{JM}) : q_i = 0 \textit{ for } i = 1, \ldots, c \}$$

$$V(\mathcal{JM}_{1,2}) =$$
$$\left\{ \langle p_1, \ldots, p_k \rangle \in V(\mathcal{JM}) : \forall F \in \mathcal{F} \sum_{H_i = (F,G) : G \in \mathcal{G}} p_i = m_1(F), \forall G \in \mathcal{G} \sum_{H_i = (F,G) : F \in \mathcal{F}} p_i = m_2(G) \right\}$$

Given this notation we can define a normalisation method simply as a function mapping from the set of joint belief assignments to the set of consistent joint belief assignments in such away as already consistent assignments remain unchanged.

Definition 7. *Normalisation Method*
A normalisation method is a function $\nu : \mathcal{JM} \rightarrow \mathcal{CJM}$ *such that if* $\underline{m} \in \mathcal{CJM}$ *then* $\nu(\underline{m}) = \underline{m}$

An extension function can be defined as a function taking as arguments a basic belief assignment m_1 on $\mathcal{F}$ and m_2 on $\mathcal{G}$ and mapping to an element of $\mathcal{JM}_{1,2}$:

Definition 8. *Extension Function*
An extension function is a function mapping from a pair of basic belief assignments to a joint belief assignment:

$e : \mathcal{M}_{\mathcal{F}} \times \mathcal{M}_{\mathcal{G}} \rightarrow \mathcal{JM}$ *such that* $e(m_1, m_2) \in \mathcal{JM}_{1,2}$
and where

$$\mathcal{M}_{\mathcal{F}} = \left\{ m : \mathcal{F} \rightarrow [0,1] : \sum_{F \in \mathcal{F}} m(F) = 1 \right\} \text{ and } \mathcal{M}_{\mathcal{G}} = \left\{ m : \mathcal{G} \rightarrow [0,1] : \sum_{G \in \mathcal{G}} m(G) = 1 \right\}$$

We can then define any combination operator in terms of the composition of an extension function and a normalisation method as follows:

Definition 9. *Combination Operator*
Given an extension function e *and a normalisation method* ν *we define the combination operator* $\oplus_{e,\nu}$ *according to:*

$$\forall S \subseteq \Omega \; m_1 \oplus_{e,\nu} m_2(S) = \sum_{(F,G): F \cap G = S} \underline{m}(F,G) \text{ where}$$

$$\underline{m} = \nu \circ e(m_1, m_2)$$

In the following sections we will consider normalisation methods and extension functions based on distance metrics (and similarity measures) defined between joint belief assignments in $\mathcal{JM}$. For normalisation, given $\underline{m} \in \mathcal{JM}$ we shall attempt to identify the element $\underline{\hat{m}} \in \mathcal{CJM}$ closest (or most similar) to $\underline{m}$. For extension functions, when given two assignments $m_1 \in \mathcal{M}_{\mathcal{F}}$ and $m_2 \in \mathcal{M}_{\mathcal{G}}$ the aim will be to identify the joint assignment $\underline{m} \in \mathcal{JM}_{1,2}$ closest to an assignment in $\mathcal{CJM}$. Given that, according to our criterion, for any $\underline{m} \in \mathcal{JM}_{1,2}$ the closest member of $\mathcal{CJM}$ is its normalisation $\underline{\hat{m}} \in \mathcal{CJM}$ then this is equivalent to identifying the $\underline{m} \in \mathcal{JM}_{1,2}$ for which the distance from its normalised assignment $\underline{\hat{m}} \in \mathcal{CJM}$ is minimal (see figure 1). For mathematical convenience, we shall tend to work in the volume representations of $\mathcal{JM}$, $\mathcal{CJM}$ and $\mathcal{JM}_{1,2}$ (as given in definition 6) since the definition of distance metrics and similarity measures is more straightforward in these spaces.

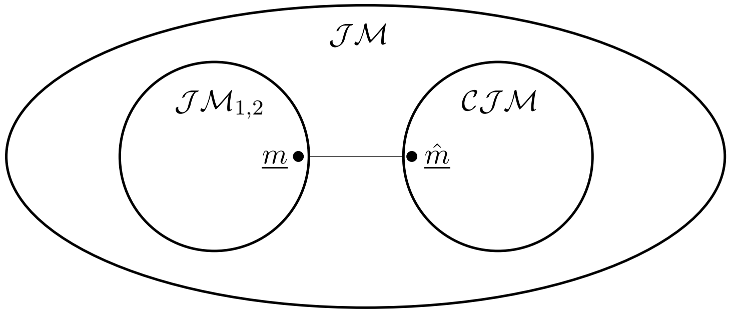

Figure 1: Diagram showing the joint belief assignment nearest to a consistent assignment

3 Similarity Based Normalisation Methods

In this paper we adopt an approach whereby the similarity between two joint belief assignments in $\mathcal{JM}$ is measured in terms of the similarity or distance between the corresponding elements in $V(\mathcal{JM})$. Specifically, we investigate quantifying this degree of similarity using three different types of metric. In the first instance the similarity between two elements $\vec{p},\ \vec{p}' \in V(\mathcal{JM})$ is taken to be $\|\vec{p}-\vec{p}'\|$ where $\|\bullet\|$ is a norm defined on the vector space $\mathbb{R}^k$. In fact we shall restrict our attention to p-norm distance metrics defined as follows:

Definition 10. *p-norm Distance Metrics* [1]

$$\left\|\vec{p}-\vec{p}'\right\|_s = \left(\sum_{i=1}^{k} \left|p_i - {p_i}'\right|^s\right)^{\frac{1}{s}} \quad \textit{where } s \geq 1$$

Probably, the most commonly used p-norm is the so-called Euclidean norm*, with* $s=2$*:*

$$\|\vec{p}\|_2 = \sqrt{\sum_{i=1}^{n} p_i^2}$$

Secondly we shall consider quantifying the distance between two elements of $V(\mathcal{JM})$ using T-indistinguishability operators (sometimes called fuzzy equivalence relations)[23] [3]. Such operations generalize the concept of an equivalence relation and provide a measure of the degree to which two elements cannot be distinguished from each other.

In the final case we shall view normalisation as a form of updating based on the assumption that the two sources of evidence are consistent. From this perspective we propose to quantify the distance between an element in $V(\mathcal{JM})$ and a corresponding element in $V(\mathcal{CJM})$ in terms of the relative change in information context resulting from updating the former to the latter. This information change will be measured using cross (relative) entropy, a choice that in itself requires a probabilistic interpretation of joint belief assignments. We shall subsequently return to discuss this issue in more detail.

[1] In this definition we shall depart from the standard notation and use the parameter s instead of p. This is merely to avoid any confusion with $\vec{p} \in V(\mathcal{JM})$

3.1 p-norm Normalisation Methods

In this subsection we shall investigate normalisation methods that can be justified on the basis of the p-norm family of distance metrics defined on $V(\mathcal{JM})$. Now for $\vec{p} \in V(\mathcal{JM})$ we shall denote its normalisation by $\vec{\hat{p}} \in V(\mathcal{CJM})$ (i.e. $\nu(\vec{p}) = \vec{\hat{p}}$). Notice that since for all $s \geq 1$ $\|\vec{p} - \vec{p}\| = 0$ it follows that if $\vec{p} \in V(\mathcal{CJM})$ then $\vec{\hat{p}} = \vec{p}$ as required by definition 7. For this reason the following results will focus on elements of $V(\mathcal{JM}) - V(\mathcal{CJM})$. The next theorem shows that for $s \geq 2$ p-norms support a normalisation method that redistributes the mass associated with the empty set uniformly to all remaining focal set pairs.

Initially we introduce the following definition of the set of positive normalisations for any $\vec{p}$. This is the set of normalisations where the only mass to be redistributed is from focal pairs with empty intersection to focal pairs with non-empty intersection. For such normalisations $\hat{p}_i \geq p_i$ for $i = c+1, \ldots, k$ or alternatively $\hat{p}_i = p_i + \epsilon_i$ where $\epsilon_i \geq 0$ for $i = c+1, \ldots, k$. This class of normalisation excludes those where mass is also reallocated between pairs of focal elements with non-empty intersection.

Definition 11. *Positive Normalisations of Joint Belief Assignments*
Let $\vec{p} \in V(\mathcal{JM}) - V(\mathcal{CJM})$ then the set of positive normalisations of $\vec{p}$ is defined as:

$$V(\mathcal{CJM})_{\geq \vec{p}} = \{\vec{q} \in V(\mathcal{CJM}) : q_i \geq p_i \text{ for } i = c+1, \ldots, k\}$$

Theorem 12. *Given $s \in \mathbb{N}$, $s \geq 2$, for any joint belief assignment $\vec{p} \in V(\mathcal{JM}) - V(\mathcal{CJM})$ there exists a unique joint belief assignment $\vec{q} \in V(\mathcal{CJM})$ for which the distance evaluated by the p-norm $\|\vec{p} - \vec{q}\|_s$ is minimal, where $\vec{q} = \vec{\hat{p}}$ given by:*

$$\hat{p}_i = \begin{cases} 0 & : i = 1, \ldots, c \\ p_i + \frac{\sum_{j=1}^{c} p_j}{k-c} & : i = c+1, \ldots, k \end{cases}$$

Proof. Clearly, minimizing

$\|\vec{p} - \vec{q}\|_s$ is equivalent to minimizing $D_s = (\|\vec{p} - \vec{q}\|_s)^s$

Hence, for any $\vec{p}$ we want to find $\vec{q}$ that minimizes:

$$D_s(\vec{p}, \vec{q}) = \sum_{j=1}^{c} p_j^s + \sum_{j=c+1}^{k} |p_j - q_j|^s$$

We now show that only $\vec{q} \in V(\mathcal{CJM})_{\geq \vec{p}}$ need be considered. To see this suppose $\vec{q} \in V(\mathcal{CJM})$ is such that $q_i < p_i$ for $i \in I$ and $q_i \geq p_i$ for $i \in J$ where $I,\ J$ forms a partition of the set of indicies $\{c+1, \ldots, k\}$. Then we can always find another $\vec{q'} \in V(\mathcal{CJM})_{\geq \vec{p}}$

with $q'_i = p_i$ for $i \in I$ and $q_i \geq q'_i \geq p_i$ for $i \in J$. In this case, for any $s \geq 2$ we have that:

$$D_s(\vec{p}, \vec{q}) = \sum_{i=1}^{c} p_i^s + \sum_{i \in I} (p_i - q_i)^s + \sum_{i \in J} (q_i - p_i)^s$$

$$\geq \sum_{i=1}^{c} p_i^s + \sum_{i \in J} \left(q'_i - p_i\right)^s = D_s\left(\vec{p}, \vec{q}'\right)$$

Hence we may substitute the absolute value in the above expression for ordinary brackets, thus obtaining:

$$D_s(\vec{p}, \vec{q}) = \sum_{j=1}^{c} p_j^s + \sum_{j=c+1}^{k} (q_j - p_j)^s$$

Since $\vec{q}$ satisfies the condition $g(\vec{q}) = \sum_{j=c+1}^{k} q_j - 1 = 0$, we can use Lagrange multipliers to find $\vec{q}$ minimizing D_s with the condition $g(\vec{q}) = 0$: Let us consider $F = D_s + \lambda g$.

$$\begin{cases} \frac{\partial F}{\partial q_j} = s\left(q_j - p_j\right)^{s-1} + \lambda = 0 \quad j = c+1, ...k \ (1) \\ \sum_{j=c+1}^{k} q_j = 1 \end{cases}$$

From (1),

$$q_j - p_j = \left(-\frac{\lambda}{s}\right)^{\frac{1}{s-1}} \quad (2)$$

$$\sum_{j=c+1}^{k} q_j - \sum_{j=c+1}^{k} p_j = 1 - \sum_{j=c+1}^{k} p_j = \left(-\frac{\lambda}{s}\right)^{\frac{1}{s-1}} (k-c)$$

$$\frac{\sum_{j=1}^{c} p_j}{k-c} = \left(-\frac{\lambda}{s}\right)^{\frac{1}{s-1}}$$

and from (2), $q_j = p_j + \frac{\sum_{j=1}^{c} p_j}{k-c}$.

In order to prove that $\vec{q}$ with $q_j = p_j + \frac{\sum_{j=1}^{c} p_j}{k-c}$ is a minimum, let us calculate the Hessian of F at $\vec{q}$.

$$\frac{\partial^2 F_s}{\partial q_j^2}(\vec{q}) = s(s-1)\left(q_j - p_j\right)^{s-2} > 0$$

$$\frac{\partial^2 D_s}{\partial q_i \partial q_j}(\vec{q}) = 0 \quad i \neq j$$

The Hessian matrix of F at $\vec{q}$ is diagonal with positive entries in it, which assures the minimality of $\vec{q}$.

□

Notice that theorem 12 also allows for the case when $\sum_{i=c+1}^{k} p_i = 0$. This means that the joint belief assignment is totally inconsistent in that it assigns zero mass to all focal set pairs (F, G) for which $F \cap G \neq \emptyset$. In this case, then according to theorem 12 the normalised assignment should assign equal mass to all such focal pairs so that:

$$q_i = \frac{1}{k-c} : i = c+1, \ldots, k$$

This would perhaps seem intuitive since, in the completely inconsistent case, there is no reason a priori to distinguish between focal pairs with non-empty intersection.

It is also interesting to note that theorem 12 does not hold when $s = 1$ according to which

$$\|\vec{p} - \vec{q}\|_{s=1} = \sum_{i=1}^{n} |p_i - q_i|$$

For this metric we have the following result:

Theorem 13. *Given $\vec{p} \in V(\mathcal{JM}) - V(\mathcal{CJM})$ then $\|\vec{p} - \vec{q}\|_1$ is minimal for $\vec{q} \in V(\mathcal{CJM})$ if and only if $\vec{q} = \vec{\hat{p}}$, where $\vec{\hat{p}}$ has the following form:*

$$\hat{p}_i = \begin{cases} 0 & : i = 1, \ldots, c \\ p_i + \epsilon_i \text{ where } \epsilon_i \geq 0 \text{ and } \sum_{i=c+1}^{k} \epsilon_i = \sum_{i=1}^{c} p_i & : i = c+1, \ldots, k \end{cases}$$

Proof. $(\Rightarrow)$

$\forall \vec{q} \in V(\mathcal{CJM})$ then $q_i = p_i + \epsilon$ where $\epsilon_i \in [-1, 1]$

Now since $\sum_{i=c+1}^{k} q_i = 1$ then $\sum_{i=c+1}^{k} \epsilon_i = \sum_{i=1}^{c} p_i$ and therefore

$$\|\vec{p} - \vec{q}\|_1 = \sum_{i=1}^{c} p_i + \sum_{i=c+1}^{k} |\epsilon_i| \geq \sum_{i=1}^{c} p_i + \sum_{i=c+1}^{k} \epsilon_i = 2 \sum_{i=1}^{c} p_i$$

In the case where $\epsilon_i \geq 0 : i = c+1, \ldots, k$ then

$$\|\vec{p} - \vec{q}\|_1 = \sum_{i=1}^{c} p_i + \sum_{i=c+1}^{k} \epsilon_i = 2 \sum_{i=1}^{c} p_i$$

and hence $\vec{q}$ is a minimum as required.
$(\Leftarrow)$
Suppose $\epsilon_j < 0$ for some $j \in \{c+1, \ldots, k\}$ then

$$\|\vec{p} - \vec{q}\|_1 = \sum_{i=1}^{c} p_i + \sum_{i=c+1}^{k} |\epsilon_i| > \sum_{i=1}^{c} p_i + \sum_{i=c+1}^{k} \epsilon_i = 2 \sum_{i=1}^{c} p_i$$

and hence $\vec{q}$ is not a minimum as required. ∎

Hence, from theorem 13 we have that for any $\vec{p} \in V(\mathcal{JM}) - V(\mathcal{CJM})$ the norm $\|\bullet\|_1$ restricts normalisations to the set $V(\mathcal{CJM})_{\geq \vec{p}}$ (definition 11) but does not discriminate between normalisations of this class.

We now consider a distance metric based on max which can be obtained by taking the limit of $\|\bullet\|_s$ as $s \to \infty$.

Definition 14. *The maximum norm* $\|\bullet\|_\infty$ *on* $\mathbb{R}^n$ *is defined as:*

$$\forall \vec{x} \in \mathbb{R}^n \ \ \|\vec{x}\|_\infty = \max\{|x_i| : i = 1, \ldots, n\}$$

Theorem 15. $\|\bullet\|_s$ *converges pointwise to* $\|\bullet\|_\infty$ *as s tends to infinity. That is:*

$$\forall \vec{x} \in \mathbb{R}^n, \ \lim_{s\to\infty} \|\vec{x}\|_s = \|\vec{x}\|_\infty$$

The normalisation result obtained for any s-norm with $s \geq 2$ (theorem 12), together with the pointwise convergence of p-norms (theorem 15), yield the following normalisation result for the maximum norm:

Theorem 16. *For any joint belief assignment* $\vec{p} \in V(\mathcal{JM}) - V(\mathcal{CJM})$*, then* $\|\vec{p} - \vec{q}\|_\infty$ *is minimal across* $\vec{q} \in V(\mathcal{CJM})$ *for* $\vec{q} = \vec{\hat{p}}$ *where:*

$$\hat{p}_i = \begin{cases} 0 & : i = 1, \ldots, c \\ p_i + \frac{\sum_{j=1}^{c} p_j}{k-c} & : i = c+1, \ldots, k \end{cases}$$

Proof. Let $\vec{\hat{p}}$ be the consistent normalisation of $\vec{p}$ given by:

$$\hat{p}_i = \begin{cases} 0 & : i = 1, \ldots, c \\ p_i + \frac{\sum_{j=1}^{c} p_j}{k-c} & : i = c+1, \ldots, k \end{cases}$$

and let $\vec{q}$ be an arbitrary but fixed vector in $V(\mathcal{CJM})$.

Now, for $s \in \mathbb{N}, \ s \geq 2$, let $a_s = \|\vec{p} - \vec{q}\|_s \in \mathbb{R}$ and let $b_s = \|\vec{p} - \vec{\hat{p}}\|_s \in \mathbb{R}$. From theorem 15 we know that the sequence $(a_s)_{s=2}^{\infty}$ converges to the real number $a = \|\vec{p} - \vec{q}\|_\infty$ (i.e. $\lim_{s\to\infty} a_s = a$). Similarly, the sequence $(b_s)_{s=2}^{\infty}$ converges to the real number $b = \left\|\vec{p} - \vec{\hat{p}}\right\|_\infty$ (i.e. $\lim_{s\to\infty} b_s = b$). Hence, the sequence $(c_s)_{s=2}^{\infty}$ defined by:

$$c_s = a_s - b_s = \|\vec{p} - \vec{q}\|_s - \|\vec{p} - \vec{\hat{p}}\|_s$$

converges such that:

$$\lim_{s\to\infty} c_s = \lim_{s\to\infty} (a_s - b_s) = \lim_{s\to\infty} a_s - \lim_{s\to\infty} b_s = a - b = \|\vec{p} - \vec{q}\|_\infty - \|\vec{p} - \vec{\hat{p}}\|_\infty$$

Furthermore, we know that for any $s \in \mathbb{N}, \ s \geq 2$ the distance $\|\vec{p} - \vec{\hat{p}}\|_s$ is minimal across $V(\mathcal{CJM})$, which means that all elements in the sequence $(c_s)_{s=2}^{\infty}$ are strictly positive:

$$0 < \|\vec{p} - \vec{q}\|_s - \|\vec{p} - \vec{\hat{p}}\|_s = c_s$$

Taking limits on both sides, we obtain the following inequality:

$$0 \leq lim_{s \to \infty} c_s = \|\vec{p} - \vec{q}\|_\infty - \|\vec{p} - \vec{\hat{p}}\|_\infty$$

Hence,

$$\|\vec{p} - \vec{\hat{p}}\|_\infty \leq \|\vec{p} - \vec{q}\|_\infty$$

as required. □

Notice that the uniqueness of the minimum for p-norms is not preserved when taking limits. In other words, the normalisation $\vec{\hat{p}}$ of $\vec{p}$ may not be the unique element in $V(\mathcal{CJM})$ minimizing the distance to $\vec{p}$ under the maximum norm. For instance, let us assume that $k = 4$, $c = 2$ and that $\vec{p}$, $\vec{\hat{p}}$ and q are as follows:

$$\vec{p} = \langle 0.3, 0.2, 0.25, 0.25 \rangle$$
$$\vec{\hat{p}} = \langle 0, 0, 0.5, 0.5 \rangle$$
$$\vec{q} = \langle 0, 0, 0.55, 0.45 \rangle$$

In this case:

$$\|\vec{p} - \vec{\hat{p}}\|_\infty = \|\vec{p} - \vec{q}\|_\infty = 0.3$$

Interestingly, the lack of uniqueness is due to the components $j = 1, \ldots, c$ and if we restrict ourselves to the components $j = c{+}1, \ldots, k$ then $\vec{\hat{p}}$ is indeed the unique element minimizing the distance defined by $\|\bullet\|$. More precisely, for any $\vec{q} \in V(\mathcal{CJM})$ the maximum distance from $\vec{p}$ can be expressed as follows:

$$\|\vec{p} - \vec{q}\|_\infty = \max\left(\| \langle p_1, \ldots, p_c \rangle \|_\infty, \|p_{c+1} - q_{c+1}, \ldots, p_k - q_k\|_\infty\right)$$

Obviously, only the second part of this expression $\|p_{c+1} - q_{c+1}, \ldots, p_k - q_k\|_\infty$ is dependent on $\vec{q}$. For the above example we have that:

$$\|\vec{p} - \vec{\hat{p}}\|_\infty = \|\vec{p} - \vec{q}\|_\infty = \| \langle p_1, p_2 \rangle \|_\infty = 0.3 \text{ whereas}$$
$$\| \langle p_3 - \hat{p}_3, p_4 - \hat{p}_4 \rangle \|_\infty = 0.25 < \| \langle p_3 - q_3, p_4 - q_4 \rangle \|_\infty = 0.3$$

Indeed if we redefine the distance metric so that it is only based on the components $j = c + 1, \ldots, k$ then uniqueness can be restored as follows:

Definition 17.

$$\forall \vec{p} \in V(\mathcal{JM}), \ \forall \vec{q} \in V(\mathcal{CJM}) \ \ \|\vec{p} - \vec{q}\|_\infty^{>c} = \|p_{c+1} - q_{c+1}, \ldots, p_k - q_k\|_\infty$$

Theorem 18. *For any joint belief assignment* $\vec{p} \in V(\mathcal{JM}) - V(\mathcal{CJM})$ *for which* $\sum_{j=c+1}^{c} p_j > 0$*, there is a unique joint belief assignment* $\vec{q} \in V(\mathcal{CJM})$ *for which* $\|\vec{p} - \vec{q}\|_{\infty}^{>c}$ *is minimal, given by* $\vec{q} = \vec{\hat{p}}$ *where:*

$$\hat{p}_i = \begin{cases} 0 & : i = 1, \ldots, c \\ p_i + \frac{\sum_{j=1}^{c} p_j}{k-c} & : i = c+1, \ldots, k \end{cases}$$

Proof. For $\vec{\hat{p}}$ the normalisation of $\vec{p}$ defined such that

$$\hat{p}_i = \begin{cases} 0 & : i = 1, \ldots, c \\ p_i + \frac{\sum_{j=1}^{c} p_j}{k-c} & : i = c+1, \ldots, k \end{cases}$$

we have that:

$$|p_j - \hat{p}_j| = |p_i - \hat{p}_i| = \frac{\sum_{j=1}^{c} p_j}{k-c} \ \forall i, j = c+1, \ldots k$$

Also since $\hat{p}_j > p_j$ then

$$|p_j - \hat{p}_j| = \hat{p}_j - p_j \ \text{ for } j = c+1, \ldots, k$$

Hence

$$\forall j = c+1, \ldots, k \ \|\vec{p} - \vec{\hat{p}}\|_{\infty}^{>c} = \hat{p}_j - p_j$$

Now for any other $\vec{q} \in V(\mathcal{CJM})$ for which $\vec{q} \neq \vec{\hat{p}}$, since

$$\sum_{j=c+1}^{k} \hat{p}_j = \sum_{j=c+1}^{k} q_j$$

there must exist a coordinate j such that $q_j > \hat{p}_j > p_j$. Therefore,

$$\|\vec{p} - \vec{q}\|_{\infty}^{>c} \geq q_j - p_j > \hat{p}_j - p_j = \|\vec{p} - \vec{\hat{p}}\|_{\infty}^{>c}$$

as required. □

3.2 T-indistinguishability Normalisation Methods

In this section we will use the theory of indistinguishability operators to define a range of similarity measures between elements of $V(\mathcal{JM})$. We will then show how these measures can be used to provide justification for a number of normalisation methods.

Indistinguishability operators are a generalization of the notion of equivalence to allow for a non-binary measure of similarity between elements. They are also referred to as fuzzy equalities and fuzzy equivalence relations depending on the context in which they are applied.

Definition 19. *A fuzzy relation $E : X \times X \rightarrow [0, 1]$ on a set X is a T-indistinguishability measure if and only if for all x, y, z of X satisfies the following properties*

- $E(x, x) = 1$ *(Reflexivity)*
- $E(x, y) = E(y, x)$ *(Symmetry)*
- $T(E(x, y), E(y, z)) \leq E(x, z)$ *(T-Transitivity)*

$E(x, y)$ can be viewed as the degree of similarity or indistinguishability between elements x and y. The function T in definition 19 models a conjunctive operation and is referred to as a t-norm [9]. Hence, the third axiom intuitively states that the degree to which x is similar to z exceeds the degree to which x is similar to y and y is similar to z, for any element y. Clearly, this is a generalization of the triangle inequality.

Definition 20. *Continuous t-norms*
A continuous t-norm is a function $T : [0, 1] \times [0, 1] \rightarrow [0, 1]$ satisfying the following axioms: $\forall a, b, c \in [0, 1]$

- *T is continuous*
- $T(a, 1) = a$ *(boundary condition)*
- *If $b \leq c$ then $T(a, b) \leq T(a, c)$ (monotonicity)*
- $T(a, b) = T(b, a)$ *(commutativity)*
- $T(a, T(b, c)) = T(T(a, b), c)$ *(associativity)*

T is Archimedean iff moreover

- $T(a, a) < a$, *if $a \neq 0, 1$*

Continuous Archimedean t-norms can be generated according to an additive generation function f as can be seen from the following theorem (see [9]):

Theorem 21. *Characterisation Theorem for t-norms*
A function $T : [0, 1] \times [0, 1] \rightarrow [0, 1]$ is a continuous archimedean t-norm iff there exists a strictly decreasing continuous function $f : [0, 1] \rightarrow [0, \infty]$ with $f(1) = 0$ such that

$$\forall a, b \in [0, 1] \;\; T(a, b) = f^{(-1)}(f(a) + f(b))$$

where

$$f^{(-1)}(a) = \begin{cases} 1 & : a < 0 \\ f^{-1}(a) & : a \in [0, f(0)] \\ 0 & : a > f(0) \end{cases}$$

If $f(0) = \infty$ then T is strict, otherwise T is non-strict.

Example 22. *Examples of Archimedean t-norms*

- **Lukasiewicz t-norm:** $T(x,y) = \max(0, x+y-1),\ f(x) = 1-x$
- **Product t-norm:** *If* $T(x,y) = x \times y,\ f(x) = -log(x)$
- **Schweizer and Sklar:** $T_\lambda(x,y) = \left[\max\left(0, x^\lambda + y^\lambda - 1\right)\right]^{\frac{1}{\lambda}},\ f_\lambda(x) = \frac{1-x^\lambda}{\lambda}$ *where* $\lambda \neq 0$

A t-norm T naturally generates a T-indistinguishability measure as follows:

Definition 23. *Given a continuous t-norm T we can define a T-indistinguishability operator according to:*

$E_T(x,y) = T(\overrightarrow{T}(x|y), \overrightarrow{T}(y|x))$ *where*
$\overrightarrow{T}(x|y) = \sup\{\alpha \in [0,1]\, |T(\alpha, x) \leq y\}$

The operator $\overrightarrow{T}(\bullet|\bullet)$ in definition 23 is referred to as the residuation of T and corresponds to the logical implication operator derived from T.

Theorem 24. *For T a continuous archimedean t-norm with additive generator f it holds that:*

$\forall x, y \in [0,1]\ \ E_T(x,y) = f^{(-1)}(|f(x) - f(y)|)$

Example 25. *T - indistinguishability operators based on t-norms*

- **Lukasiewicz t-norm:** *If* $T(x,y) = \max(0, x+y-1)$ *then* $E_T(x,y) = 1 - |x-y|$
- **Product t-norm:** *If* $T(x,y) = x \times y$ *then* $E_T(x,y) = \min\left(\frac{x}{y}, \frac{y}{x}\right)$
- **Schweizer and Sklar:** *If* $T_\lambda(x,y) = \left[\max\left(0, x^\lambda + y^\lambda - 1\right)\right]^{\frac{1}{\lambda}}$ *then* $E_T(x,y) = \left(1 - \left|x^\lambda - y^\lambda\right|\right)^{\frac{1}{\lambda}}$

Given a T-indistinguishability operator defined on $[0,1]$ we now consider how this can be extended to provide a measure of similarity between the elements of $V(\mathcal{JM})$. The first approach we consider is based on a quasi-arithmetic mean defined in terms of the generator function f as follows:

Definition 26. *Similarity Measures based on Quasi-Arithmetic Means*
Let T be a continuous Archimedean t-norm with generator function f then we define the following measure of similarity between joint belief assignments $\vec{p} \in V(\mathcal{JM})$ and $\vec{q} \in V(\mathcal{CJM})$:

$$AS(\vec{p}, \vec{q}) = f^{(-1)}\left(\frac{\sum_{i=1}^{k} f(E_T(p_i, q_i))}{k}\right)$$

Example 27. *For the Schweizer and Sklar family of t-norms*
$T_\lambda(a,b) = \left[\max\left(0, x^\lambda + y^\lambda - 1\right)\right]^{\frac{1}{\lambda}}$ *with* $\lambda > 0$ *then:*

$$AS_\lambda(\vec{p},\vec{q}) = f^{(-1)}\left(\frac{\sum_{i=1}^k f_\lambda\left(\left(1 - \left|p_i^\lambda + q_i^\lambda\right|\right)^{\frac{1}{\lambda}}\right)}{k}\right) = f_\lambda^{(-1)}\left(\frac{\sum_{i=1}^k \left|p_i^\lambda - q_i^\lambda\right|}{\lambda k}\right)$$

$$= \left[1 - \frac{\sum_{i=1}^k \left|p_i^\lambda - q_i^\lambda\right|}{k}\right]^{\frac{1}{\lambda}}$$

Lemma 28. *Let h be a decreasing and concave function $h : \mathbb{R} \to \mathbb{R}$ and consider the map $g(x_1,..,x_n) = \sum_{i=1}^n h(x_i)$. The maximum of g in the region defined by the restrictions $a_i \le x_i \le b$ $i = 1,...,n$ and $\sum_{i=1}^n x_i = b$ for fixed a_i, b is reached at the points with coordinates $x_i = a_i$ for all $i = 1,...,n$ except for exactly one x_j with corresponding $a_j = argmax\{a_i : i = 1,\ldots,n\}$ for which $x_j = c - \sum_{i=1,i\neq j}^n a_i$.*

Lemma 29. *Let h be a decreasing and convex function $h : \mathbb{R} \to \mathbb{R}$ and consider the map $g(x_1,..,x_n) = \sum_{i=1}^n h(x_i)$. The maximum of g in the region defined by the restrictions $a_i \le x_i \le b$ $i = 1,...,n$ and $\sum_{i=1}^n x_i = b$ for fixed a_i, b is reached at the points with coordinates $x_i = a_i$ for all $i = 1,...,n$ except for exactly one x_j with corresponding $a_j = argmin\{a_i : i = 1,\ldots,n\}$ for which $x_j = c - \sum_{i=1,i\neq j}^n a_i$.*

Proof of lemmas 28 and 29 can be found in Stromberg [19]

Theorem 30. *For $\vec{p} \in V(\mathcal{JM}) - V(\mathcal{CJM})$ then $AS_\lambda(\vec{p},\vec{q})$ where $0 < \lambda < 1$ $(\lambda > 1)$ is maximal for $\vec{q} \in V(\mathcal{CJM})$ such that $q_i = p_i$ for all $i = c+1,\ldots,k$ except exactly one q_j with corresponding $p_j = argmax\{p_i : i = c+1,\ldots,k\}$*
$(p_j = argmin\{p_i : i = c+1,\ldots,k\})$ for which $q_j = 1 - \sum_{i=c+1,i\neq j} p_i$.

Proof. We note that only $\vec{q} \in V(\mathcal{CJM})_{\ge\vec{p}}$ need be considered. To see this suppose $\vec{q} \in V(\mathcal{CJM})$ is such that $q_i < p_i$ for $i \in I$ and $q_i \ge p_i$ for $i \in J$ where $I,\ J$ forms a partition of the set of indices $\{c+1,\ldots,k\}$. Then we can always find another $\vec{q'} \in V(\mathcal{CJM})$ with $q'_i = p_i$ for $i \in I$ and $q_i \ge q'_i \ge p_i$ for $i \in J$. In this case for any t-norm T we have that:

$E_T\left(p_i, q'_i\right) = 1 \ge E_T(p_i, q_i)$ for $i \in I$ and
$E_T\left(p_i, q'_i\right) \ge E_T(p_i, q_i)$ for $i \in J$ since $q_i \ge q'_i \ge p_i$

Now, since quasi-arithmetic means are non-decreasing functions in each variable it follows that:

$$AS\left(\vec{p},\vec{q'}\right) \ge AS(\vec{p},\vec{q})$$

and therefore $AS(\vec{p},\bullet)$ is not maximal at $\vec{q}$. Now for $\vec{q} \in V(\mathcal{CJM})_{\ge\vec{p}}$ we have that:

$$AS_\lambda(\vec{p},\vec{q}) = f_\lambda^{(-1)}\left(\frac{\sum_{i=1}^k \left|p_i^\lambda - q_i^\lambda\right|}{\lambda k}\right)$$

Hence, since $f^{(-1)}$ is a decreasing function and $k > 0$ it follows that maximizing $AS_\lambda(\vec{p}, \vec{q})$ is equivalent to minimizing $\frac{|p_i^\lambda - q_i^\lambda|}{\lambda}$ subject to the constraints that $q_i \geq p_i : i = c+1, \ldots, k$. Now since $\vec{q} \in V(\mathcal{CJM})_{\geq \vec{p}}$

$$\frac{|p_i^\lambda - q_i^\lambda|}{\lambda} = \frac{\sum_{i=1}^{c} p_i^\lambda + \sum_{i=c+1}^{k}(q_i^\lambda - p_i^\lambda)}{\lambda} = \frac{\sum_{i=1}^{c} p_i^\lambda - \sum_{i=c+1}^{k} p_i^\lambda}{\lambda} + \frac{\sum_{i=c+1}^{k} q_i^\lambda}{\lambda}$$

which is minimal when $\sum_{i=c+1}^{k} q_i^\lambda$ is minimal. Now trivially minimizing $\sum_{i=c+1}^{k} q_i^\lambda$ is equivalent to maximizing:

$$k - c - \sum_{i=c+1}^{k} q_i^\lambda = \sum_{i=c+1}^{k} 1 - q_i^\lambda$$

and for $\lambda < 1$ $1 - q^\lambda$ is a decreasing concave function of q and hence the required result follows by lemma 28. For $\lambda > 1$ $1 - q^\lambda$ is a decreasing convex function of q and hence the required result follows by lemma 29 $\square$

We now introduce a non-additive measure of similarity based on min. Naturally, this measure tends to be more conservative than AS.

Definition 31. *Minimum based Similarity Measure*
Let T be a continuous t-norm then we define the following measure of similarity:

$$MinS(\vec{p}, \vec{q}) = \min(\{E_T(p_i, q_i) : i = c+1, \ldots, k\})$$

$MinS$ given in definition 31 is restricted to the components $i = c+1, \ldots, k$ for similar reasons to those discussed regarding the distance metric $\| \bullet \|_\infty^{>c}$ in section 3.1.

Theorem 32. *(a) For a continuous non-strict archimedean t-norm T with additive generator f, then for $\vec{p} \in V(\mathcal{JM}) - V(\mathcal{CJM})$ such that $\sum_{i=c+1}^{k} p_i > 0$, $MinS(\vec{p}, \vec{q})$ is maximal across $\vec{q} \in V(\mathcal{CJM})$ when all values of $f(p_i) - f(q_i)$ are equal for $i = c+1, \ldots, k$*
(b) For a continuous strict archimedean t-norm T with additive generator f, then for $\vec{p} \in V(\mathcal{JM}) - V(\mathcal{CJM})$ such that $\sum_{i=c+1}^{k} p_i > 0$, $MinS(\vec{p}, \vec{q})$ is maximal across $\vec{q} \in V(\mathcal{CJM})$ when $q_i = 0$ for $i \in I$ and $f(p_i) - f(q_i) = f(p_j) - f(q_j)$ for $i \neq j \in \{c+1, \ldots, k\} - I$ where $I = \{i > c : p_i = 0\}$.

Proof. As in theorem 30 we can restrict consideration to $\vec{q} \in V(\mathcal{CJM})_{\geq \vec{p}}$.
(a) Since both f and $f^{(-1)}$ are continuous it is possible to find $\vec{\hat{p}} \in V(\mathcal{CJM})_{\geq \vec{p}}$ such that:

$$f(p_i) - f(\hat{p}_i) = f(p_j) - f(\hat{p}_j) \;\; i, j = c+1, \ldots, k$$

By theorem 24 this trivially implies that

$$E_T(p_i, \hat{p}_i) = E_T(p_j, \hat{p}_j) \;\; i, j = c+1, \ldots, k$$

Now consider another $\vec{q} \in V(\mathcal{CJM})_{\geq \vec{p}}$ such that $\vec{q} \neq \vec{\hat{p}}$. Then since $\sum_{i=c+1}^{k} q_i = \sum_{i=c+1}^{k} \hat{p}_i = 1$ it must hold that for some $j \in \{c+1, \ldots, k\}$ $q_j > \hat{p}_j > p_j$.

Since f is a strictly decreasing function and $q_j > \hat{p}_j > p_j$ then

$$|f(p_j) - f(q_j)| = f(p_j) - f(q_j) > f(p_j) - f(\hat{p}_j) = |f(p_j) - f(\hat{p}_j)|$$
$$\Rightarrow f^{(-1)}(|f(p_j) - f(q_j)|) < f^{(-1)}(|f(p_j) - f(\hat{p}_j)|)$$
$$\Rightarrow E_T(p_j, q_j) < E_T(p_j, \hat{p}_j) \text{ (by theorem 24) } = MinS\left(\vec{p}, \vec{\hat{p}}\right)$$

Therefore

$$MinS(\vec{p}, \vec{q}) < MinS\left(\vec{p}, \vec{\hat{p}}\right)$$

as required.

(b) From theorem 24 we have that for strict archimedean t-norms that $E_T(0, q_i) = 0$ for $q_i > 0$ and $E_T(0,0) = 1$. Let $\vec{\hat{p}} \in V(\mathcal{CJM})_{\geq \vec{p}}$ be such that $\hat{p}_i = 0$ for all $i \in I$ and let $\vec{q} \in V(\mathcal{CJM})_{\geq \vec{p}}$ be such that $q_j > 0$ for some $j \in I$. Hence,

$$MinS\left(\vec{p}, \vec{\hat{p}}\right) \geq MinS(\vec{p}, \vec{q}) = E_T(p_j, q_j) = 0$$

Now suppose that $\vec{\hat{p}} \in V(\mathcal{CJM})_{\geq \vec{p}}$ is such that $\hat{p}_i = 0$ for all $i \in I$ and $f(p_i) - f(q_i) = f(p_j) - f(q_j)$ for $i \neq j \in \{c+1, \ldots, k\} - I$. Also let $\vec{q} \in V(\mathcal{CJM})_{\geq \vec{p}}$ be such that $q_i = 0$ for all $i \in I$ but where $\vec{q} \neq \vec{\hat{p}}$. Then by the argument given in part (a) we have that:

$$MinS\left(\vec{p}, \vec{\hat{p}}\right) \geq MinS(\vec{p}, \vec{q})$$

as required. □

Example 33. *Normalisations Generated from SMin*
Schweizer and Sklar t-norms: $f_\lambda(x) = \frac{1-x^\lambda}{\lambda}$ *for* $\lambda > 0$ *and hence* $f_\lambda(p_i) - f_\lambda(q_i) = \frac{q_i^\lambda - p_i^\lambda}{\lambda}$. *Therefore, from theorem 32 we have that* $MinS$ *is maximal when* $q_i = \left(\Delta + p_j^\lambda\right)^{\frac{1}{\lambda}}$: $i = c+1, \ldots, k$. *Now*

$$\sum_{i=c+1}^{k} q_i = \sum_{i=c+1}^{k} \left(\Delta + p_i^\lambda\right)^{\frac{1}{\lambda}} = 1 \Rightarrow \prod_{i=c+1}^{k} e^{\left(\Delta + p_i^\lambda\right)^{\frac{1}{\lambda}}} = e \Rightarrow \left[\prod_{i=c+1}^{k} e^{\left(\Delta + p_i^\lambda\right)}\right]^{\frac{1}{\lambda}} = e$$
$$\Rightarrow \prod_{i=c+1}^{k} e^{\left(\Delta + p_j^\lambda\right)} = e^\lambda \Rightarrow e^{\Delta(k-c)} \prod_{i=c+1}^{k} e^{p_i^\lambda} = e^\lambda \Rightarrow e^{\sum_{i=c+1}^{k} p_i^\lambda} = e^{\lambda - \Delta(k-c)}$$
$$\Rightarrow \Delta = \frac{\lambda - \sum_{i=c+1}^{k} p_i^\lambda}{k-c} \Rightarrow q_j = \left[\frac{\lambda - \sum_{i=c+1}^{k} p_i^\lambda}{k-c} + p_j\right]^{\frac{1}{\lambda}} : j = c+1, \ldots, k$$

Notice in the case that $\lambda = 1$, *(i.e.: Lukasiewicz t-norm), we have the, now familiar, normalisation:*

$$q_j = p_j + \frac{\sum_{i=1}^{c} p_i}{k-c} \quad j = c+1, \ldots, k$$

This is not surprising given theorem 18 since for $\lambda = 1$

$MinS(\vec{p}, \vec{q}) = \min\{1 - |p_i - q_i| : i = c+1, \ldots, k\}$

and maximizing this function is equivalent to minimizing:

$\max\{|p_i - q_i| : i = c+1, \ldots, k\} = \|\vec{p} - \vec{q}\|_\infty^{>c}$

Product t-norm: *For* $i \in I$ $q_i = 0 = \frac{p_i}{\sum_{i=c+1}^{k} p_i}$. *For* $i \in \{c+1, \ldots, k\} - I$ $f(x) = -\log(x)$ *and therefore* $f(p_i) - f(q_i) = \log(q_i) - \log(p_i) = \log\left(\frac{q_i}{p_i}\right)$. *Therefore,* $\log\left(\frac{q_i}{p_i}\right) = \Delta : i = c+1, \ldots, k \Rightarrow q_i = p_i e^{\Delta} : i = c+1, \ldots, k$ *and hence:*

$$\sum_{i=c+1}^{k} q_i = e^{\Delta} \sum_{i=c+1}^{k} p_i = 1 \Rightarrow e^{\Delta} = \frac{1}{\sum_{i=c+1}^{k} p_i} \Rightarrow q_j = \frac{p_j}{\sum_{i=c+1}^{k} p_i} : j = c+1, \ldots, k$$

3.3 Cross Entropy Normalisation Methods

Cross (relative) entropy, sometimes referred to as Kullback-Leibler distance [11], provides a measure of distance between probability distributions defined as follows:

Definition 34. *Cross (Relative) Entropy*
The cross entropy of a distribution P *relative to a distribution* Q *on universe* X *is given by:*

$$CE(P\|Q) = \sum_{x \in X} P(x) \log_2\left(\frac{P(x)}{Q(x)}\right)$$

Strictly speaking $CE(P\|Q)$ is not a distance metric because it is not symmetric and does not satisfy the triangle inequality. However, it does have the following desirable property:

Theorem 35. $CE(P\|Q) \geq 0$ *with equality if and only if* $P = Q$.

$CE(P\|Q)$ can be interpreted as the change in information when a prior distribution P is updated to a posterior distribution Q on the basis of new evidence. Since elements of $\mathcal{JM}$ are (technically) probability distributions on $\mathcal{F} \times \mathcal{G}$ we can evaluate the cross entropy $CE(\vec{q}\|\vec{p})$ for $\vec{p} \in V(\mathcal{JM}) - V(\mathcal{CJM})$ and $\vec{q} \in V(\mathcal{CJM})$. In this case $CE(\vec{q}\|\vec{p})$ quantifies the change in information when the assignment $\vec{p}$ is updated to $\vec{q}$ on the basis of the assumption that the two sources of evidence are consistent.

The following theorem shows that the cross entropy measure $CE(\vec{q}\|\vec{p})$ supports the standard normalisation method as used in Dempsters rule where masses of focal set pairs with non-empty intersection are simply renormalised so that they sum to one. This result is an extension of that given in [12] where only possibility distributions were considered.

Theorem 36. *For any joint belief assignment $\vec{p} \in V(\mathcal{JM}) - V(\mathcal{CJM})$ such that $p_j > 0$ for $j = c+1, \ldots, k$ there exists a unique consistent joint belief assignment $\vec{q} \in V(\mathcal{CM})$ where $\vec{q}$ has minimum cross entropy relative to $\vec{p}$ given by $\vec{q} = \vec{\hat{p}}$ where:*

$$\hat{p}_i = \begin{cases} 0 & : \ i = 1, \ldots, c \\ \frac{p_i}{1-\sum_{j=1}^{c} p_j} & : \ i = c+1, \ldots, k \end{cases}$$

Proof.

$CE(\vec{q}\|\vec{p}) =$

$$q_{c+1}\log_2\left(\frac{q_{c+1}}{p_{c+1}}\right) + \ldots + q_{k-1}\log_2\left(\frac{q_{k-1}}{p_{k-1}}\right) + (1 - \sum_{j=c+1}^{k-1} q_j)\log_2\left(\frac{(1-\sum_{j=c+1}^{k-1} q_j)}{p_k}\right)$$

Therefore,

$$\frac{\partial CE}{\partial q_i} = \log_2\left(\frac{q_i}{p_i}\right) - \log_2\left(\frac{(1-\sum_{j=c+1}^{k-1} q_j)}{p_k}\right)$$

Now CE is minimal when $\frac{\partial CE}{\partial q_i} = 0$ and hence when,

$$\frac{q_i}{p_i} = \frac{(1-\sum_{j=c+1}^{k-1} q_j)}{p_k} \ : \ i = c+1, \ldots, k-1$$

Therefore, $\frac{p_i}{q_i} = \frac{p_j}{q_j} \ : \ j \neq i$ and in particular, $q_i = \frac{q_{c+1}}{p_{c+1}} p_i \ : \ i = c+2, \ldots, k-1$
Also

$$\frac{q_{c+1}}{p_{c+1}} = \frac{(1 - q_{c+1} - \sum_{j=c+2}^{k-1} q_j)}{p_k}$$

$$\Rightarrow q_{c+1} = \frac{p_{c+1}(1-q_{c+1}) - q_{c+1}\sum_{j=c+2}^{k-1} p_j}{p_k} = \frac{p_{c+1} - q_{c+1}\sum_{j=c+1}^{k-1} p_j}{p_k}$$

$$\Rightarrow q_{c+1}p_k + q_{c+1}\sum_{j=c+1}^{k-1} p_j = p_{c+1} \Rightarrow q_{c+1}\left(1 - \sum_{j=1}^{k-1} p_j + \sum_{j=c+1}^{k-1} p_j\right) = p_{c+1}$$

$$\Rightarrow q_{c+1}\left(1 - \sum_{j=1}^{c} p_j\right) = p_{c+1} \Rightarrow q_{c+1} = \frac{p_{c+1}}{\left(1-\sum_{j=1}^{c} p_j\right)}$$

Therefore,

$$q_i = \frac{q_{c+1}}{p_{c+1}} p_i = \frac{p_{c+1}}{\left(1-\sum_{j=1}^{c} p_j\right)} \frac{p_i}{p_{c+1}} = \frac{p_i}{\left(1-\sum_{j=1}^{c} p_j\right)} \ : \ i = c+1, \ldots, k$$

as required. □

Joslyn [8] has argued against the application of entropy measures at the basic belief assignment level because their use implicitly assumes a probabilistic interpretation. While the latter is certainly true it may be the case that under certain circumstances such an interpretation is appropriate. One possibility is the use of a voting model (Gaines [6] and Baldwin [1]) to elicit belief assignments. In this interpretation, each member v of a population of voters V is asked to identify the set of worlds '$S(v) \subseteq \Omega$' for which the constraint $w^* \in S(v)$ is the most precise representation of their current knowledge. A frequentist model is then used to evaluate a basic belief assignment according to:

$$\forall F \subseteq \Omega \; m(T) = \frac{1}{|V|} |\{v \in V : S(v) = F\}|$$

In the case of evidence combination the voting model would be conceptually as follows: It is assumed that there are two distinct populations, V_1 and V_2 each with opinions regarding w^*. For example these might be populations of different types of medical expert. The agent carrying out evidence combination would then select a population $V_{1,2} \subseteq V_1 \times V_2$ of pairs of voters (v_i, v_j) where $v_i \in V_1$ and $v_j \in V_2$. For each pair both of the voters are asked to provide $S(v)$ from which a joint belief assignment would be obtained according to:

$$\underline{m}(F, G) = \frac{1}{|V_{1,2}|} |\{(v_i, v_j) \in V_{1,2} : S(v_i) = F, \; S(v_j) = G\}|$$

In this case we can view $S(v_i)$ and $S(v_j)$ as distinct random sets and $\underline{m}$ as a joint distribution on $(S(v_i), S(v_j))$. From this perspective we can clearly interpret $\underline{m}$ probabilistically which would provide some justification for the use of the cross entropy measure. Interestingly, cross entropy suggests that the normalisation minimizing information change corresponds, in the above voting model, to simply removing those pairs (v_i, v_j) for which $S(v_i) \cap S(v_j) = \emptyset$ from $V_{1,2}$ and then recalculating $\underline{m}$ accordingly.

Another probabilistic interpretation of Dempster-Shafer theory was proposed by Lemmer [14] where the elements of Ω correspond to labels for a set of objects $\mathcal{O}$. These labels are defined so that one and only one can be truthfully applied to any given object $o \in \mathcal{O}$. We now suppose that there are two sensor s_1 and s_2 which measure different characteristics of an object, each providing some evidence as to the correct label. Furthermore, suppose that this evidence restricts the label of $o \in \mathcal{O}$ to a subset $S(s_i, o) \subseteq \Omega$. We then define m_1, m_2 and $\underline{m}$ in terms of frequencies as follows:

$$m_1(F) = \frac{1}{|\mathcal{O}|} |\{o \in \mathcal{O} : S(s_1, o) = F\}|$$

$$m_2(G) = \frac{1}{|\mathcal{O}|} |\{o \in \mathcal{O} : S(s_2, o) = G\}|$$

$$\underline{m}(F, G) = \frac{1}{|\mathcal{O}|} |\{o \in \mathcal{O} : S(s_1, o) = F, \; S(s_2, o) = G\}|$$

We shall return to consider an example of this semantics in section 4.

Both the voting model and Lemmer's semantics can be view as special cases of a more general random set semantics described as follows: In this semantics an evidence source corresponds to a random set mapping from an underlying system state to a set of possible states of the world.

Definition 37. *A Source of Evidence*
Let $\mathcal{S} = \{x_i\}_i$ be a set of underlying system states (assumed to be finite for simplicity). Then a source of evidence about the true state of the world is a random set $E : \mathcal{S} \to 2^{\Omega}$.

For example, E might be a sensor with some built in error tolerance so that it outputs a range of values indicating the presence or absence of certain properties from a system when it is in a particular state. From this perspective then, the true state of the world is a functional mapping from $\mathcal{S}$ into Ω as follows:

Definition 38. *The true-state of the world*
The true state of the world is a function $w^* : \mathcal{S} \to \Omega$

This formulation allows us to state precisely what its means for an evidence source to be reliable regarding a particular state of the system, since in such a case we would expect $w * (x) \in E(x)$.

Definition 39. *Reliable Evidence*
A source of evidence E is said to be reliable at state x if $w^* (x) \in E (x)$.

Now given two sources of evidence $E_1,\ E_2 : \mathcal{S} \to 2^{\mathcal{W}}$ and a probability measure μ on $2^{\mathcal{S}}$ we can define assignments m_1, m_2 and $\underline{m}$ according to:

$$\forall F \subseteq \mathcal{W}\ m_1(F) = \mu(x \in \mathcal{S} : E_1(x) = F)$$
$$\forall G \subseteq \mathcal{W}\ m_2(G) = \mu(x \in \mathcal{S} : E_2(x) = G)$$
$$\underline{m}(F, G) = \mu(x \in \mathcal{S} : E_1(x) = F, E_2(x) = G)$$

3.4 Summary of Normalisation Methods

The above sub-sections have provided some justification for three main normalisation methods, on the basis of a range of distance metrics and similarity measures. These are summarised below:

Normalisation Method 1
$\nu (\underline{m}) = \underline{\hat{m}}$ where $\forall F \in \mathcal{F}, \forall G \in \mathcal{G}$:

$$\underline{\hat{m}}\,(F, G) = \begin{cases} 0 & : \text{ if } F \cap G = \emptyset \\ \underline{m}\,(F, G) + \frac{\sum_{(F,G):F\cap G=\emptyset} \underline{m}(F,G)}{|\{(F,G):F\cap G\neq\emptyset\}|} & : \text{ otherwise} \end{cases}$$

According to this method the mass assigned to pairs with empty intersection in reallocated uniformly to all other pairs of focal sets. Such a normalisation method is supported by all p-norms with $s \geq 2$ (theorem 12) including $\| \bullet \|_{\infty}^{>c}$ (theorem 18), and by default $SMin$

based on the Lukasiewicz t-norm (example 33).

Normalisation Method 2
$\nu(\underline{m}) = \hat{\underline{m}}$ where $\forall F \in \mathcal{F}, \forall G \in \mathcal{G}$:

$$\hat{\underline{m}}(F,G) = \begin{cases} 0 & : \text{ if } F \cap G = \emptyset \\ \frac{\underline{m}(F,G)}{\sum_{(F,G):F\cap G \neq \emptyset} \underline{m}(F,G)} & : \text{ otherwise} \end{cases}$$

This is the standard normalisation method as applied in Dempster's rule whereby the mass assigned to focal pairs with non-empty intersections is simply rescaled so that these values sum to one. This normalisation method arises from the cross entropy measure of distance (theorem 36) and also the $SMin$ similarity measure based on the product t-norm (example 33) .

Normalisation Method 3
$\nu(\underline{m}) = \hat{\underline{m}}$ where $\forall F \in \mathcal{F}, \forall G \in \mathcal{G}$:

$$\hat{\underline{m}}(F,G) = \begin{cases} 0 & : \text{ if } F \cap G = \emptyset \\ \underline{m}(F,G) + \sum_{(F,G):F\cap G = \emptyset} \underline{m}(F,G) & : \text{ if } \underline{m}(F,G) \text{ is maximal (minimal)} \\ \underline{m}(F,G) & : \text{ otherwise} \end{cases}$$

In the case that the maximum (minimum) of $\underline{m}$ is not unique then one of the values is selected arbitrarily.

This normalisation method simply allocates all the mass associated with the empty set to the focal set pair with maximum (minimum) mass under $\underline{m}$. All other values of $\underline{m}$ remain unchanged. The normalisation method results from the assumption of AS similarity measures based on the Schweizer and Sklar family of t-norms (theorem 30).

4 Similarity Based Extension Functions

In this section we will argue that the decision to apply a combination operator of the type given in definition 9 carries with it an implicit assumption of consistency, rather than an assumption of independence as underlies Dempster's rule. To make the case for this claim we begin by considering the situation where both sources of evidence are certain.

Suppose that source one asserts '$w^* \in F$' and source two asserts '$w^* \in G$' where, while certain, both F and G are highly imprecise (e.g. $|F| = |G| = \frac{|\Omega|}{2}$). In this situation it is perfectly possible that the combined evidence $F \cap G$ is very precise (e.g. $|F \cap G| = 1$). But what is the justification for inferring precision from imprecision in this way? Surely, the only possible justification lies in the belief of the agent carrying out the combination, that the true state of the world w^* is contained both in F and in G. Clearly, this can only be the case if F and G are consistent (i.e. $F \cap G \neq \emptyset$).

We now claim, in view of this argument and since a combination operator as given in definition 9 is a method of intersecting two uncertain sources of evidence, that when an agent makes a decision to apply such an operator they are assuming that both sources are reliable for most system states in the sense of definition 39 and hence as a consequence they are implicitly assuming at least a high level of consistency between the sources. If this were not the case they would adopt a different method of evidence combination, perhaps by taking union or by applying some kind of additive aggregation operator. In other words, it is only really meaningful to intersect two sources of evidence if they are dependent in the sense that they both relate to the true state of the world w^*. Hence, once the decision to intersect m_1 and m_2 has been made then the joint belief assignment $\underline{m}$ should be chosen to take account of the implicit assumption of consistency between the two sources. That is, $\underline{m}$ should be as close as possible (as measured by some metric) to a consistent assignment as is permitted by the constraint that its marginals are m_1 and m_2. In our notation this means that $\vec{p} \in V(\mathcal{JM}_{1,2})$ should be selected so that the distance from (or similarity of) $\vec{p}$ to the closest (most similar) element of $V(\mathcal{CM})$ should be minimal (maximal). Since, the closest element in $V(\mathcal{CM})$ to a given $\vec{p} \in V(\mathcal{JM}_{1,2})$ is its normalisation $\vec{\hat{p}}$ then the problem reduces to identifying that $\vec{p} \in V(\mathcal{JM}_{1,2})$ closest to its normalised assignment $\vec{\hat{p}}$. Notice that, if we view normalisation as a form of error correction, then selecting the joint belief assignment in this way means that the effect of normalisation is minimized.

In an attempt to clarify the above argument we now make a case for the maximal consistency assumption within the context of constructive probabilities introduced by Shafer [17] (see Voorbraak [20] for an exposition); this being the framework in which Shafer initially justifies Dempster's rule. Suppose we have a set of codes $\mathcal{C}$ and a message X. The codes are selected at random by the sender according to probability distribution P_1 which is also known to the receiver. The message X contains information regarding the true state of the world and applying code $c_i \in \mathcal{C}$, X decodes to the constraint '$w^* \in c_i(X)$' where $c_i(X) \subseteq \Omega$. The basic belief assignment $m_1(F)$ then corresponds to the receivers probability that the message sent was '$w^* \in F$' as given by:

$$m_1(F) = \sum_{c_i : c_i(X) = F} P_1(c_i)$$

Now further suppose that we also have a second set of codes $\mathcal{D}$ selected at random according to distribution P_2 in order to encode a second message Y. Y also relates to the true state of the world so that for $d_j \in \mathcal{D}$, $d_j(Y) \subseteq \Omega$. Let the basic belief assignment for Y be m_2.

Consider then the receivers knowledge regarding the actual pair of codes (c_i, d_j) used to encode messages X and Y respectively. Let $P(c_i, d_j | X, Y)$ be the receivers distribution on pairs (c_i, d_j) taking into account the two received messages X and Y. The joint belief assignment is then defined in terms of this distribution according to:

$$\underline{m}(F, G) = \sum_{(c_i, d_j) : c_i(X) = F, d_j(Y) = G} P(c_i, d_j | X, Y)$$

Now it would seem reasonable to argue that $\underline{m}$ should have marginals m_1 and m_2 as follows: Knowing X and Y does not give the receiver any additional information regarding c_i (d_j) beyond that which is already provided by the prior distribution P_1 (P_2). Hence,

$$\forall c_i \in \mathcal{C}\ P\left(c_i|X,Y\right) = \sum_{d_j \in \mathcal{D}} P\left(c_i, d_j|X,Y\right) = P_1\left(c_i\right) \text{ and}$$

$$\forall d_j \in \mathcal{D}\ P\left(d_j|X,Y\right) = \sum_{c_i \in \mathcal{C}} P\left(c_i, d_j|X,Y\right) = P_2\left(d_j\right)$$

Dempster's rule then results from the additional assumption that knowing X and Y does not provide any further information regarding pairs (c_i, d_j) than is provided separately by the prior distributions P_1 and P_2. That is:

$$P\left(c_i, d_j|X,Y\right) = P_1\left(c_i\right) P_2\left(d_j\right)$$

We argue that such an assumption is unrealistic given the receiver's decision to intersect messages X and Y. More specifically, we suggest that the decision to intersect X and Y is likely to have resulted from an assumption similar to the following: *Either the encoding, transmission or decoding process may result in some errors to either X or Y. However, these errors are not sufficient to invalidate the assumption that $w^* \in c_i(X) \cap d_j(Y)$ where c_i and d_j are the actual codes used.* But this reasoning suggests that the receiver would tend to give higher values to $P\left(c_i, d_j|X,Y\right)$ for those pairs (c_i, d_j) where $c_i(X) \cap d_j(Y) \neq \emptyset$, since it is only for such codes that $w^* \in c_i(X) \cap d_j(Y)$ can possibly be true. Consequently, X and Y do provide some additional information regarding the probability of pairs of codes.

Perhaps not surprisingly the criterion of maximal consistency does not always yield a unique joint belief assignment in which case we follow Cattaneo [5] in then selecting the $\underline{m}$ which results in the least specific combined belief assignment as identified by the following measure due to Klir and Wierman [10]:

Definition 40. *Nonspecificity*

$$N(m) = \sum_{S \neq \emptyset} m(S) \log_2\left(|S|\right)$$

Insisting on maximum nonspecificity in this manner has the advantage that when taken in conjunction with maximal consistency selection of joint assignments it ensures that the resulting combination operator is idempotent. Also, this criterion mimimizes the information context of the resulting belief function beyond that which can be justified by the maximum consistency assumption. In those cases where nonspecificity still does not yield a unique solution we propose (again following Cattaneo [5]) that the centre of mass of the volume representing the remaining assignments should then selected.

In the results presented below we restrict consideration to the following two subsets of $V\left(\mathcal{JM}_{1,2}\right)$:

Definition 41.

$$V(\mathcal{JM}_{1,2})^{+} = \left\{ \vec{p} \in V(\mathcal{JM}_{1,2}) : \sum_{i=c+1}^{k} p_i > 0 \right\}$$
$$V(\mathcal{JM}_{1,2})^{++} = \{ \vec{p} \in V(\mathcal{JM}_{1,2}) : p_i > 0 \ for\ i = c+1, \ldots, k \}$$

$V(\mathcal{JM}_{1,2})^{+}$ is simply the set of joint belief assignments that are not totally inconsistent (i.e. $\sum_{(F,G):F \cap G = \emptyset} \underline{m}(F,G) < 1$). While some distance metrics (e.g. p-norms) do provide a solution to the normalisation problem even in the case of total inconsistency it seems intuitively unreasonable to select such an assignment if other, (partially) consistent, assignments are available. In those cases where $V(\mathcal{JM}_{1,2})^{+} = \emptyset$ we would argue that it is unreasonable to consider combining m_1 and m_2 conjunctively as they cannot be consistent. $V(\mathcal{JM}_{1,2})^{++} \subset V(\mathcal{JM}_{1,2})^{+}$ further restricts the set of joint belief assignments to those that give non-zero mass to all pairs of focal elements with non-empty intersections. Such a restriction is more difficult to justify intuitively, although we might claim that no consistent focal pair should be ruled out a priori. Certainly, however, it is necessary to restrict ourselves in this way, when using the cross entropy measure which is undefined outside $V(\mathcal{JM}_{1,2})^{++}$.

Theorem 42. *For any $\vec{p} \in V(\mathcal{JM}_{1,2})^{+}$ let $\vec{\hat{p}}$ be the normalisation of $\vec{p}$ defined by:*

$$\hat{p}_i = \begin{cases} 0 & : i = 1, \ldots, c \\ p_i + \frac{\sum_{j=1}^{c} p_j}{k-c} & : i = c+1, \ldots, k \end{cases}$$

then $\|\vec{p} - \vec{\hat{p}}\|_s$, for $s \geq 2$ is minimal across $V(\mathcal{JM}_{1,2})^{+}$ if and only if the following function of $\vec{p}$ is minimal:

$$\sum_{j=1}^{c} p_j^s + \frac{\left(\sum_{j=1}^{c} p_j\right)^s}{(k-c)^{s-1}}$$

Proof. Minimizing $\|\vec{p} - \vec{\hat{p}}\|_s$ is equivalent to minimizing

$$\sum_{j=1}^{c} p_j^s + \left(\sum_{c+1}^{k} \|p_j - \hat{p}_j\|^s \right) = \sum_{j=1}^{c} p_j^s + \sum_{j=c+1}^{k} \frac{\left(\sum_{j=1}^{c} p_j\right)^s}{(k-c)^s}$$
$$= \sum_{j=1}^{c} p_j^s + \frac{\left(\sum_{j=1}^{c} p_j\right)^s}{(k-c)^{s-1}}$$

□

Theorem 43. *For any $\vec{p} \in V(\mathcal{JM}_{1,2})^{+}$ let $\vec{\hat{p}}$ be the normalisation of $\vec{p}$ defined by:*

$$\hat{p}_i = \begin{cases} 0 & : i = 1, \ldots, c \\ p_i + \frac{\sum_{j=1}^{c} p_j}{k-c} & : i = c+1, \ldots, k \end{cases}$$

then $\|\vec{p}-\vec{\hat{p}}\|_{\infty}^{>c}$ *is minimal across* $V\left(\mathcal{JM}_{1,2}\right)^{+}$ *if and only if* $\sum_{j=1}^{c} p_j$ *is minimal*

Proof. The result follows trivially since

$$\|\vec{p}-\vec{\hat{p}}\|_{\infty}^{>c} = \sum_{j=1}^{c} p_j$$

□

Theorem 44. *For any* $\vec{p} \in V\left(\mathcal{JM}_{1,2}\right)^{+}$ *let* $\vec{\hat{p}}$ *be the normalisation of* $\vec{p}$ *defined by:*

$$\hat{p}_i = \begin{cases} 0 & : i = 1,\ldots,c \\ \frac{p_i}{\sum_{i=c+1}^{k} p_i} & : i = c+1,\ldots,k \end{cases}$$

then $MinS\left(\vec{p},\vec{\hat{p}}\right)$ *based on the product t-norm is maximal across* $V\left(\mathcal{JM}_{1,2}\right)^{+}$ *if and only if* $\sum_{j=1}^{c} p_j$ *is minimal*

Proof.

$$MinS\left(\vec{p},\vec{\hat{p}}\right) = \min\left\{\min\left(\frac{p_i}{\hat{p}_i},\frac{\hat{p}_i}{p_i}\right) : i \in I\right\} = \frac{p_i}{\hat{p}_i} = \sum_{j=c+1}^{k} p_j$$

$$= 1 - \sum_{j=1}^{c} p_j$$

which is maximal when $\sum_{j=1}^{c} p_j$ is minimal. □

Theorem 45. *For any* $\vec{p} \in V\left(\mathcal{JM}_{1,2}\right)^{++}$ *let* $\vec{\hat{p}}$ *be the normalisation of* $\vec{p}$ *defined by:*

$$\hat{p}_i = \begin{cases} 0 & : i = 1,\ldots,c \\ \frac{p_i}{\sum_{i=c+1}^{k} p_i} & : i = c+1,\ldots,k \end{cases}$$

Then the cross entropy of $\vec{\hat{p}}$ *relative to* $\vec{p}$ *is minimal across* $V\left(\mathcal{JM}_{1,2}\right)^{++}$ *if and only if* $\sum_{j=1}^{c} p_j$ *is minimal.*

Proof.

$$CE = \sum_{i=c+1}^{k} \hat{p}_i \log_2\left(\frac{\hat{p}_i}{p_i}\right) = \sum_{i=c+1}^{k} \frac{p_i}{1-\sum_{j=1}^{c} p_j} log_2\left(\frac{p_i}{\left(1-\sum_{j=1}^{c} p_j\right) p_i}\right)$$

$$= \log_2\left(\frac{1}{1-\sum_{j=1}^{c} p_j}\right)\sum_{i=c+1}^{k}\frac{p_i}{1-\sum_{j=1}^{c} p_j} = \log_2\left(\frac{1}{1-\sum_{j=1}^{c} p_j}\right)\frac{\sum_{i=c+1}^{k} p_i}{1-\sum_{j=1}^{c} p_j}$$

$$\log_2\left(\frac{1}{1-\sum_{j=1}^{c} p_j}\right)\frac{1-\sum_{i=1}^{c} p_i}{1-\sum_{j=1}^{c} p_j} = \log_2\left(\frac{1}{1-\sum_{j=1}^{c} p_j}\right) = -\log_2\left(1-\sum_{j=1}^{c} p_j\right)$$

Now $\log_2(1-x)$ is strictly increasing for $x \in [0,1]$ and hence CE is minimal if and only if $\sum_{j=1}^{c} p_j$ is minimal. □

Theorems 42-45 provide constraints on the joint belief assignment that should be selected from $\mathcal{JM}_{1,2}$. These motivate the three combination operators outlined in the following section.

4.1 Summary of Combination operators

On the basis of theorems 42-45 and the results given in section 3 the metrics $\|\bullet\|_s : s \geq 2$, $\|\bullet\|_\infty^{\geq c}$ and CE together with $SMin$ based on the product t-norm, provide a level of justification for a number of different combinations of extension function and normalisation method. These can then be used to generate the three combination operators given below, by taking the composition as in definition 9.

Combination Operator 1: Based on $\|\bullet\|_s$ with $s \geq 2$ (see theorem 12 and theorem 42)
Extension Function: Select $\underline{m} \in \mathcal{JM}_{1,2}$ to be the centre of mass of the most nonspecific joint belief assignments minimizing the expression:

$$\sum_{(F,G):F\cap G=\emptyset} \underline{m}(F,G)^s + \frac{\left(\sum_{(F,G):F\cap G=\emptyset} \underline{m}(F,G)\right)^s}{|\{(F,G) : F\cap G \neq \emptyset\}|^{s-1}}$$

Normalisation Method: The selected joint assignment $\underline{m}$ is normalised according to:

$$\hat{\underline{m}}(F,G) = \begin{cases} 0 & : \text{ if } F\cap G=\emptyset \\ \underline{m}(F,G) + \frac{\sum_{(F,G):F\cap F=\emptyset} \underline{m}(F,G)}{|\{(F,G):F\cap G\neq\emptyset\}|} & : \text{ otherwise} \end{cases}$$

Combination Operator 2: Based on $\|\bullet\|_\infty^{\geq c}$ (see theorem 18 and theorem 43)
Extension Function: Select $\underline{m} \in \mathcal{JM}_{1,2}$ to be the centre of mass of the most nonspecific joint belief assignments minimizing the expression:

$$\sum_{(F,G):F\cap G=\emptyset} \underline{m}(F,G)$$

Normalisation Method: The selected joint assignment $\underline{m}$ is normalised according to:

$$\hat{\underline{m}}(F,G) = \begin{cases} 0 & : \text{ if } F\cap G=\emptyset \\ \underline{m}(F,G) + \frac{\sum_{(F,G):F\cap F=\emptyset} \underline{m}(F,G)}{|\{(F,G):F\cap G\neq\emptyset\}|} & : \text{ otherwise} \end{cases}$$

Combination Operator 3: Based on CE and $MinS$ with the product t-norm (see theorem 36 and theorem 45 or example 33 and theorem 44). This is identical to the operator proposed by Cattaneo [5].
Extension Function: Select $\underline{m} \in \mathcal{JM}_{1,2}$ to be the centre of mass of the most nonspecific joint belief assignments minimizing the expression:

$$\sum_{(F,G):F\cap G=\emptyset} \underline{m}(F,G)$$

Normalisation Method: The selected joint assignment $\underline{m}$ is normalised according to:

$$\hat{\underline{m}}(F,G) = \begin{cases} 0 & : \text{ if } F \cap G = \emptyset \\ \frac{\underline{m}(F,G)}{\sum_{(F,G):F\cap G\neq\emptyset} \underline{m}(F,G)} & : \text{ otherwise} \end{cases}$$

The following example illustrates the range of different solutions that can be obtained from these three operators.

Example 46. *Let m_1 be defined by*

$m_1(\{x_1, x_2, x_3\}) = 0.4, m_1(\{x_1, x_2\}) = 0.3, m_1(\{x_1\}) = 0.3$

Let m_2 be defined by

$m_2(\{x_2, x_3\}) = 0.5, m_2(\{x_3\}) = 0.5$

The set of joint belief assignments $\mathcal{JM}_{1,2}$ can then be represented by the following tableau:

$\mathcal{F} \times \mathcal{G}$	$\{x_2, x_3\} : 0.5$	$\{x_3\} : 0.5$
$\{x_1, x_2, x_3\} : 0.4$	$\{x_2, x_3\}$ p_6	$\{x_3\}$ p_5
$\{x_1, x_2\} : 0.3$	$\{x_2\}$ p_4	$\emptyset$ p_3
$\{x_1\} : 0.3$	$\emptyset$ p_2	$\emptyset$ p_1

Subject to the following constraints:

$p_5 + p_6 = 0.4, p_3 + p_4 = 0.3, p_1 + p_2 = 0.3, p_2 + p_4 + p_6 = 0.5, p_1 + p_3 + p_5 = 0.5$
$p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1$

Hence, we can define:

$$V(\mathcal{JM}_{1,2}) = \{\langle p_4 + p_6 - 0.2, 0.5 - p_4 - p_6, 0.3 - p_4, p_4, 0.4 - p_6, p_6 \rangle \in [0,1]^6\}$$

Also, in this case:

$$V(\mathcal{CJM}) = \left\{ \langle 0, 0, 0, q_4, q_5, q_6 \rangle \in [0,1]^6 : \sum_{i=4}^{6} q_i = 1 \right\}$$

Now the extension function based on $\|\bullet\|_2$ requires that we minimize the following function:

$$\left(\sum_{j=1}^{c} p_j^2 \right) + \frac{\left(\sum_{j=1}^{c} p_j \right)^2}{(k-c)} = (0.3 - p_4)^2 + (0.5 - p_4 - p_6)^2 + (p_4 + p_6 - 0.2)^2 + \frac{(0.6 - p_4)^2}{3}$$

This is minimal when $p_4 = 0.3$ and $p_6 = 0.05$ which generates the following (in this case) unique joint belief assignment:

$\mathcal{F}\times\mathcal{G}$	$\{x_2,x_3\}:0.5$	$\{x_3\}:0.5$
$\{x_1,x_2,x_3\}:0.4$	$\{x_2,x_3\}$ 0.05	$\{x_3\}$ 0.35
$\{x_1,x_2\}:0.3$	$\{x_2\}$ 0.3	$\emptyset$ 0
$\{x_1\}:0.3$	$\emptyset$ 0.15	$\emptyset$ 0.15

Normalizing according to $\hat{p}_i = p_i + \frac{\sum_{j=1}^{c} p_j}{k-c} : i = c+1,\ldots,k$ *as required by combination operator 1 then gives:*

$\mathcal{F}\times\mathcal{G}$	$\{x_2,x_3\}$	$\{x_3\}$
$\{x_1,x_2,x_3\}$	$\{x_2,x_3\}$ 0.15	$\{x_3\}$ 0.45
$\{x_1,x_2\}$	$\{x_2\}$ 0.4	$\emptyset$ 0
$\{x_1\}$	$\emptyset$ 0	$\emptyset$ 0

Hence:

$$m_1 \oplus_{e,\nu} m_2(\{x_2,x_3\}) = 0.15,\ \ m_1 \oplus_{e,\nu} m_2(\{x_2\}) = 0.4\ \ m_1 \oplus_{e,\nu} m_2(\{x_3\}) = 0.45$$

Alternatively, an extension function based on Cross Entropy, $MinS$ *with the product t-norm or* $\|\bullet\|_{\infty}^{>c}$*requires that we minimize:*

$$\sum_{i=1}^{c} p_i = 0.6 - p_4$$

This is minimal when $p_4 = 0.3$ *which generates the following set of joint belief assignments:*

$\mathcal{F}\times\mathcal{G}$	$\{x_2,x_3\}:0.5$	$\{x_3\}:0.5$
$\{x_1,x_2,x_3\}:0.4$	$\{x_2,x_3\}$ p_6	$\{x_3\}$ $0.4-p_6$
$\{x_1,x_2\}:0.3$	$\{x_2\}$ 0.3	$\emptyset$ 0
$\{x_1\}:0.3$	$\emptyset$ $0.2-p_6$	$\emptyset$ $0.1+p_6$

where $p_6 \leq 0.2$.

Hence, applying the normalisation $\hat{p}_i = \frac{p_i}{\sum_{i=c+1}^{k} p_i}$ *as required by combination operator 3 gives:*

$$m_1 \oplus_{e,\nu} m_2(\{x_2,x_3\}) = \frac{p_6}{0.7},\ \ m_1 \oplus_{e,\nu} m_2(\{x_2\}) = \frac{3}{7},\ \ m_1 \oplus_{e,\nu} m_2(\{x_3\}) = \frac{0.4-p_6}{0.7}$$

where $p_6 \leq 0.2$

The nonspecificity of this mass assignment is then given by:

$$N(m_1 \oplus_{e,\nu} m_2) = \sum_{S \neq \emptyset} m_1 \oplus_{e,\nu} m_2(S) \log_2(|S|)$$

$$= \frac{p_6}{0.7} \log_2(2) + \frac{0.4 - p_6}{0.7} \log_2(1) + \frac{3}{7} \log_2(1) = \frac{p_6}{0.7}$$

which is maximal when $p_6 = 0.2$ *so that:*

$$m_1 \oplus_{e,\nu} m_2(\{x_2, x_3\}) = \frac{2}{7},\ m_1 \oplus_{e,\nu} m_2(\{x_2\}) = \frac{3}{7},\ m_1 \oplus_{e,\nu} m_2(\{x_3\}) = \frac{2}{7}$$

Alternatively applying the normalisation $\hat{p}_i = p_i + \frac{\sum_{i=1}^{c} p_i}{k-c} : i = c+1, \ldots, k$ *as in combination operator 2 gives:*

$$m_1 \oplus_{e,\nu} m_2(\{x_2, x_3\}) = p_6 + 0.1,\ m_1 \oplus_{e,\nu} m_2(\{x_2\}) = 0.4,\ m_1 \oplus_{e,\nu} m_2(\{x_3\}) = 0.5 - p_6$$

where $p_6 \leq 0.2$

The nonspecificity of this mass assignment is then given by:

$$N(m_1 \oplus_{e,\nu} m_2) = \sum_{S \neq \emptyset} m_1 \oplus_{e,\nu} m_2(S) \log_2(|S|)$$

$$= (p_6 + 0.1) \log_2(2) + (0.5 - p_6) \log_2(1) + 0.4 \log_2(1) = p_6 + 0.1$$

which is maximal when $p_6 = 0.2$ *so that:*

$$m_1 \oplus_{e,\nu} m_2(\{x_2, x_3\}) = 0.3,\ m_1 \oplus_{e,\nu} m_2(\{x_2\}) = 0.4,\ m_1 \oplus_{e,\nu} m_2(\{x_3\}) = 0.3$$

We now consider an example proposed by Lemmer [14] (also discussed by Voorbraak [20]) based on his interpretation of Dempster-Shafer theory in terms of labelling objects (section 3.3) and use this to illustrate our argument that the decision to intersect different sources of evidence carries with it an implicit assumption of consistency. Clearly, in this example Dempster's rule cannot be applied since the sources of evidence are dependent in that they measure different attributes of the same object. In this sense we claim that it is typical of many sensor fusion problems.

Example 47. *Suppose we have an urn of balls* $\mathcal{O}$ *each labelled by with one and only one of the labels from* $\Omega = \{a, b, c\}$*. Each ball also has the following physical characteristics: Its weight is either light or heavy and its colour is either red or blue. Furthermore, light balls are known to be labelled* a *whereas heavy balls are labels either* b *or* c*, and blue balls are labelled* c *whereas red balls are either* a *or* b*. Let* s_1 *be a sensor classifying weight and* s_2 *be a sensor classifying colour. Also, let* $P(t)$ *be the fraction of balls in* $\mathcal{O}$ *with label* t*. Then according to Lemmer's model [14] as described in section 3.3:*

$$m_1(\{a\}) = P(a),\ m_1(\{b, c\}) = 1 - P(a)$$
$$m_2(\{c\}) = P(c),\ m_2(\{a, b\}) = 1 - P(c)$$

Now intuitively it seems reasonable to intersect the evidence from these two sensors since it is based on different characteristics of the same object *(i.e the same ball), which is then aggregated across* $\mathcal{O}$*. Hence, for any particular* $o \in \mathcal{O}$*, provided the sensors are functioning correctly, we would expect the true label of* o *to be contained in the set* $S(s_1, o) \cap S(s_2, o)$*. From this perspective, since* Ω *contains all possible labels, we would expect a high level of consistency between the two sources even allowing for a certain degree of error from the sensors. That is, on average we would expect that* $S(s_1, o) \cap S(s_2, o) \neq \emptyset$*.*

In the case that $P(a) + P(c) \leq 1$ *then combination operators 1-3 all give the result:*

$$m_1 \oplus_{e,\nu} m_2(\{a\}) = P(a), \; m_1 \oplus_{e,\nu} m_2(\{c\}) = P(c),$$
$$m_1 \oplus_{e,\nu} m_2(\{b\}) = 1 - P(a) - P(c)$$

This effectively corresponds to the assumption that, given no evidence to the contrary, both sensors are working correctly. It also yields the same result as a probabilistic analysis of the evidence. In contrast, Dempster's rule a priori assumes a level of error (inconsistency) corresponding to:

$$\sum_{(F,G):F \cap G = \emptyset} \underline{m}(F, G) = \underline{m}(\{a\}, \{c\}) = P(a) P(c)$$

In the case that $P(a) + P(c) > 1$ *then there must be some error in the sensors and hence some level of inconsistency. The extension functions from operators 1-3 all give:*

$$\sum_{(F,G):F \cap G = \emptyset} \underline{m}(F, G) = \underline{m}(\{a\}, \{c\}) = P(a) + P(c) - 1$$

The normalisation method from operators 1 and 2 then yields:

$$m_1 \oplus_{e,\nu} m_2(\{a\}) = 1 - P(c) + \frac{1}{3}(P(a) + P(c) - 1),$$
$$m_1 \oplus_{e,\nu} m_2(\{c\}) = 1 - P(a) + \frac{1}{3}(P(a) + P(c) - 1),$$
$$m_1 \oplus_{e,\nu} m_2(\{b\}) = \frac{1}{3}(P(a) + P(c) - 1)$$

Alternatively, the normalisation operator 3 gives:

$$m_1 \oplus_{e,\nu} m_2(\{a\}) = \frac{1 - P(c)}{2 - P(a) - P(c)},$$
$$m_1 \oplus_{e,\nu} m_2(\{c\}) = \frac{1 - P(a)}{2 - P(a) - P(c)},$$
$$m_1 \oplus_{e,\nu} m_2(\{b\}) = 0$$

In terms of random set semantics we have that:

$$\mathcal{S} = \{\langle light, red\rangle, \langle light, blue\rangle, \langle heavy, red\rangle, \langle heavy, blue\rangle\}$$

and we have two sources of evidence E_1 *and* E_2 *based on sensors* s_1 *and* s_2 *respectively, such that:*

$$E_1\left(\langle light, red\rangle\right) = E_1\left(\langle light, blue\rangle\right) = \{a\}$$
$$E_1\left(\langle heavy, red\rangle\right) = E_1\left(\langle heavy, blue\rangle\right) = \{b, c\}$$

and

$$E_2\left(\langle light, red\rangle\right) = E_2\left(\langle heavy, red\rangle\right) = \{a, b\}$$
$$E_2\left(\langle light, blue\rangle\right) = E_2\left(\langle heavy, blue\rangle\right) = \{c\}$$

The probability measure μ *on* $2^{\mathcal{S}}$ *is then defined by the following probability distribution on* $\mathcal{S}$*:*

$$\forall \langle x, y\rangle \in \mathcal{S}\ \mu\left(\langle x, y\rangle\right) = \frac{1}{|\mathcal{O}|}\left|\{o \in \mathcal{O} : size\ of\ o\ is\ x,\ and\ colour\ of\ o\ is\ y\}\right|$$

Let

$$\mu\left(\langle light, red\rangle\right) = \mu_1,\ \mu\left(\langle light, blue\rangle\right) = \mu_2,\ \mu\left(\langle heavy, blue\rangle\right) = \mu_3$$
$$\mu\left(\langle heavy, red\rangle\right) = 1 - \mu_1 - \mu_2 - \mu_3$$

then

$$m_1\left(\{a\}\right) = \mu_1 + \mu_2 = P\left(a\right),\ m_1\left(\{b, c\}\right) = 1 - \mu_1 - \mu_2 = 1 - P\left(a\right)$$
$$m_2\left(\{a, b\}\right) = \mu_2 + \mu_3 = P\left(c\right),\ m_2\left(\{c\}\right) = 1 - \mu_2 - \mu_3 = 1 - P\left(c\right)$$

and

$$\underline{m}\left(\{a\}, \{a, b\}\right) = \mu_1,\ \underline{m}\left(\{a\}, \{a, b\}\right) = \mu_2,\ \underline{m}\left(\{b, c\}, \{c\}\right) = \mu_3$$
$$\underline{m}\left(\{b, c\}, \{c\}\right) = 1 - \mu_1 - \mu_2 - \mu_3$$

From this perspective the selection functions for combination operators 1-3 identify the joint belief assignment $\underline{m}$ *for which* $\mu_2 = \mu\left(\langle light, blue\rangle\right)$ *takes the minimum value of* $\max\left(0, P\left(a\right) + P\left(c\right) - 1\right)$. *The maximizes the possible reliability of the two sources of evidence in the sense that:*

$$\mu\left(x \in \mathcal{S} : w^*\left(x\right) \in E_1\left(x\right) \cap E_2\left(x\right)\right) \leq \mu\left(x \in \mathcal{S} : E_1\left(x\right) \cap E_2\left(x\right) = \emptyset\right) = 1 - \mu_2$$

Hence, these operators assume the maximum level of reliability for the two sources of evidence that is permitted by the two observed belief assignments m_1 *and* m_2*.*

In addition, we can interpret the normalisation methods 1 and 2 at the random set level as follows: Suppose that $P(a)+P(c) > 1$ *then it follows that* $\mu\left(\langle light, blue\rangle\right) = \mu_2 > 0$ *and clearly both sources of evidence cannot be reliable at this state. Taking* $\mu_2 = P(a)+P(c)-1$ *then it follows that* $\mu_1 = 1 - P(c)$*, and* $\mu_3 = 1 - P(a)$ *so that* $\mu\left(\langle heavy, red\rangle\right) = 0$*. Now* *since there is no positive evidence of unreliability at any state except $\langle light, blue\rangle$ we assume that the sensors are reliable at all other states. Also since we have no knowledge of the nature of the error occurring at state $\langle light, blue\rangle$ we only know that the true set value for $E_1(\langle light, blue\rangle) \cap E_2(\langle light, blue\rangle) \subseteq \Omega - \emptyset$. Both normalisation methods then assume the actual values of E_1 and E_2 for this state are amongst those which have actually occurred in the data, so that $E_1(\langle light, blue\rangle) \cap E_2(\langle light, blue\rangle) \in \{F \cap G : F \in \mathcal{F},\ G \in \mathcal{G},\ F \cap G \neq \emptyset\}$. Beyond this, normalisation method 1 then assumes that all such set values should be equally likely giving an assignment:*

$$E_1(\langle light, blue\rangle) \cap E_2(\langle light, blue\rangle) = \{a\} : \frac{1}{3},\ \{b\} : \frac{1}{3},\ \{c\} : \frac{1}{3}$$

In this case:

$$m_1 \oplus_{e,\nu} m_2 (\{a\}) = \mu_1 + \frac{1}{3}\mu_2 = 1 - P(c) + \frac{1}{3}(P(a) + P(c) - 1),$$
$$m_1 \oplus_{e,\nu} m_2 (\{c\}) = \mu_3 + \frac{1}{3}\mu_2 = 1 - P(a) + \frac{1}{3}(P(a) + P(c) - 1),$$
$$m_1 \oplus_{e,\nu} m_2 (\{b\}) = 1 - \mu_1 - \mu_2 - \mu_3 + \frac{1}{3}\mu_2 = \frac{1}{3}(P(a) + P(c) - 1)$$

On the other hand, as with normalisation method 2 we might take the prior probability of each value $S \in \{F \cap G : F \in \mathcal{F},\ G \in \mathcal{G},\ F \cap G \neq \emptyset\}$ to be proportional to $\mu(x \in \mathcal{S} : E_1(x) \cap E_2(x) = S)$ so that:

$$E_1(\langle light, blue\rangle) \cap E_2(\langle light, blue\rangle) =$$
$$\{a\} : \frac{\mu_1}{1-\mu_2},\ \{b\} : \frac{\mu_3}{1-\mu_2},\ \{c\} : \frac{1-\mu_1-\mu_2-\mu_3}{1-\mu_2}$$

In this case:

$$m_1 \oplus_{e,\nu} m_2 (\{a\}) = \mu_1 + \frac{\mu_1}{1-\mu_2}\mu_2 = \frac{1-P(c)}{2-P(a)-P(c)},$$
$$m_1 \oplus_{e,\nu} m_2 (\{c\}) = \mu_3 + \frac{\mu_3}{1-\mu_2}\mu_2 = \frac{1-P(a)}{2-P(a)-P(c)},$$
$$m_1 \oplus_{e,\nu} m_2 (\{b\}) = 1 - \mu_1 - \mu_2 - \mu_3 + \frac{1-\mu_1-\mu_2-\mu_3}{1-\mu_2}\mu_2 = 0$$

For this example we might argue that the former solution is perhaps more intuitive than the latter for the following reason: Inconsistency and hence errors arise when sensor s_1 indicates 'light' while sensor s_2 indicates 'blue'. The first of the above solutions allows for the possibility that in such a case both sensors are in fact incorrect, so that the true reading should be (heavy, red). The second solution effectively rules out this possibility on the grounds of the high frequency of observed a and c balls (so high in fact that $P(a)+P(c) > 1$), thereby assuming that only one sensor is in error. However, this seems somewhat unjustified since there is no knowledge as to the nature of the sensor errors.

The following example is an adaptation of that given in [20] and subsequently discussed by Liu and Hong [15].

Example 48. *Suppose that Jim wants to know whether the street outside is slippery. Instead of observing this himself, he asks Fred. Fred tells him that 'it is slippery'. However, Jim knows that Fred is sometimes careless in answering questions. Based on his knowledge about Fred, Jim estimates that* $\alpha \times 100\%$ *of the time Fred reports what he knows and* $(1-\alpha) \times 100\%$ *he is careless.*

Furthermore, suppose Jim has some other evidence about whether the street is slippery: his trusty indoor-outdoor thermometer says that the temperature is $31^o F$*, and he knows that because of traffic, ice could not form on the street at this temperature. However, he knows that the thermometer could be wrong even though it has been very accurate in the past. Suppose that there is a* $\beta \times 100\%$ *chance that the thermometer is working properly. For this example* $\Omega = \{yes, no\}$ *and we have the following two basic belief assignments:*

$$m_1 = \{yes\} : \alpha,\ \{yes, no\} : 1-\alpha$$
$$m_2 = \{no\} : \beta,\ \{yes, no\} : 1-\beta$$

Now one might suppose that the two sources of evidence as represented by Fred and the thermometer are independent in that there is no direct connection between them. However, from the point of view of intersecting their output the two sources are dependent in the sense that they both refer to the slipperiness of the same road. Hence, even allowing for inaccuracy in the two sources we would expect a high level of consistency between them. On the other hand, Dempster's rule assumes that the two sources are independent resulting in the following joint belief assignment:

$$\underline{m}(\{yes\}, \{no\}) = \alpha \times \beta,\ \underline{m}(\{yes\}, \{yes, no\}) = \alpha \times (1-\beta),$$
$$\underline{m}(\{yes, no\}, \{no\}) = (1-\alpha) \times \beta,\ \underline{m}(\{yes, no\}, \{yes, no\}) = (1-\alpha) \times (1-\beta)$$

Hence, Dempster's rule assumes an a priori belief in the inconsistency of the two sources of $\alpha \times \beta$ *even in those case when* $\alpha + \beta \leq 1$ *for which* m_1 *and* m_2 *are consistent.*

Alternatively, in the case where $\alpha + \beta \leq 1$ *the selection functions for operators 1-3 identify the following joint belief assignment, where the a priori assumption is no inconsistency between the sources:*

$$\underline{m}(\{yes\}, \{no\}) = 0,\ \underline{m}(\{yes\}, \{yes, no\}) = \alpha,$$
$$\underline{m}(\{yes, no\}, \{no\}) = \beta,\ \underline{m}(\{yes, no\}, \{yes, no\}) = 1-\alpha-\beta$$

From this all three operators result in the following combined assignment:

$$m_1 \oplus_{e,\nu} m_2(\{yes\}) = \alpha,\ m_1 \oplus_{e,\nu} m_2(\{no\}) = \beta,\ m_1 \oplus_{e,\nu} m_2(\{yes, no\}) = 1-\alpha-\beta$$

If $\alpha+\beta>1$ then we are forced to assume the existence of some inconsistency between Fred and the thermometer. The extension function for operators 1-3 all then yield the following joint assignment, where a priori a minimum mass of $\alpha+\beta-1$:

$$\underline{m}\left(\{yes\},\{no\}\right)=\alpha+\beta-1,\ \underline{m}\left(\{yes\},\{yes,no\}\right)=1-\beta,$$
$$\underline{m}\left(\{yes,no\},\{no\}\right)=1-\alpha,\ \underline{m}\left(\{yes,no\},\{yes,no\}\right)=0$$

From this normalisation method 1 (as in operators 1 and 2) gives:

$$m_1\oplus_{e,\nu}m_2\left(\{yes\}\right)=1-\beta+\frac{1}{3}\left(\alpha+\beta-1\right),$$
$$m_1\oplus_{e,\nu}m_2\left(\{no\}\right)=1-\alpha+\frac{1}{3}\left(\alpha+\beta-1\right),$$
$$m_1\oplus_{e,\nu}m_2\left(\{yes,no\}\right)=\frac{1}{3}\left(\alpha+\beta-1\right)$$

Alternatively, normalisation method 3 (as in operator 3) gives:

$$m_1\oplus_{e,\nu}m_2\left(\{yes\}\right)=\frac{1-\beta}{2-\alpha-\beta},$$
$$m_1\oplus_{e,\nu}m_2\left(\{no\}\right)=\frac{1-\alpha}{2-\alpha-\beta},$$
$$m_1\oplus_{e,\nu}m_2\left(\{yes,no\}\right)=0$$

5 Conditional Belief Measures

In Dempster-Shafer theory there is a direct link between Dempster's rule and the definition of conditional belief and plausibility. Given a prior basic belief assignment m with focal sets $\mathcal{G}$ suppose it is then learnt that $w^*\in F$. This knowledge can be represented by a belief assignment m' for which $m'(F)=1$ and $m'(F')=0$ for $F'\neq F$. A conditional assignment is then defined according to:

$$\forall S\subseteq\Omega\ m\left(S|F\right)=m'\oplus m\left(S\right)\ \text{ where }\ \oplus\ \text{ is Dempster's rule.}$$

from which the associated conditional belief and plausibility measures can be obtained in the normal manner.

Now clearly this approach can be generalized to other combination operators so that:

$$\forall S\subseteq\Omega\ m\left(S|F\right)=m'\oplus_{e,\nu}m\left(S\right)$$

Interestingly given the definition of m' there is in fact only one joint belief assignment with marginals m' and m defined by: $\forall G\in\mathcal{G}$

$$\underline{m}\left(G,F\right)=m\left(G\right)$$
$$\underline{m}\left(G,F'\right)=0\text{ if }F'\neq F$$

Hence, the definition of $m(\bullet|F)$ is dependent only on the normalisation method ν. In view of this we can now investigate the conditional belief and plausibility measures generated by the three normalisation methods summarised in section 3.4.

In fact normalisation method 2 is the standard normalisation as applied in Dempster's rule and hence any combination operator employing this method will generate the standard Dempster-Shafer conditionals given by:

$$\forall S \subseteq \Omega \; Bel(S|F) = \frac{Bel(S \cup F^c) - Bel(F^c)}{1 - Bel(F^c)} \text{ and}$$

$$Pl(S|F) = \frac{Pl(S \cap F)}{Pl(F)}$$

Therefore, we need only focus on normalisation methods 1 and 3.

Now for all normalisation methods the focal elements of the conditional belief assignment are:

$$\mathcal{G}_F = \{F \cap G : G \in \mathcal{G}, \; F \cap G \neq \emptyset\}$$

By applying method 1 and uniformly reallocating the mass where $F \cap G = \emptyset$ it can easily be seen that $m(\bullet|F)$ is given by:

$$\forall T \subseteq \Omega \; m(T|F) =$$

$$\sum_{G \in \mathcal{G}: G \cap F = T} m(G) + \frac{|\{G \in \mathcal{G} : G \cap F = T\}|}{|\{G \in \mathcal{G} : G \cap F \neq \emptyset\}|} \times \left(\sum_{G \in \mathcal{G}: G \cap F = \emptyset} m(G) \right) \text{ if } T \in \mathcal{G}_F$$

$$= 0 \text{ otherwise}$$

In this case the corresponding belief and plausibility measures are as given in the following theorem:

Theorem 49. *Let $m : 2^\Omega \to [0,1]$ be a basic belief assignment with focal sets $\mathcal{G}$ and for $F \subseteq \Omega$. If the conditional mass assignment $m(\bullet|F)$ is generated from normalisation method 1 then the corresponding conditional belief and plausibility measures defined by $\forall S \subseteq \Omega \; Bel(S|F) = \sum_{T:T \subseteq S} m(T|F)$ and $Pl(S|F) = \sum_{T:T \cap S \neq \emptyset} m(T|F)$ respectively, are equivalent to:*

$$\forall S \subseteq \Omega \; Bel(S|F) = Bel(F^c \cup S) - \frac{|\{G : F \cap G \cap S^c \neq \emptyset\}|}{|\{G : F \cap G \neq \emptyset\}|} \times Bel(F^c)$$

$$Pl(S|F) = Pl(F \cap S) + \frac{|\{G : F \cap G \cap S \neq \emptyset\}|}{|\{G : F \cap G \neq \emptyset\}|} (1 - Pl(F))$$

Proof. Note that

$$\sum_{G \in \mathcal{G}: G \cap F = \emptyset} m(G) = \sum_{G \in \mathcal{G}: G \subseteq F^c} m(G) = Bel(F^c) \text{ and hence}$$

$$m(T|F) = \sum_{G \in \mathcal{G}: G \cap F = T} m(G) + \frac{|\{G \in \mathcal{G} : G \cap F = T\}|}{|\{G \in \mathcal{G} : G \cap F \neq \emptyset\}|} Bel(F^c)$$

Hence

$$Bel\left(S|F\right) =$$
$$\sum_{T\in\mathcal{G}_F:T\subseteq S}\sum_{G\in\mathcal{G}:G\cap F=T} m\left(G\right) + \frac{Bel\left(F^c\right)}{\left|\{G\in\mathcal{G}:G\cap F\neq\emptyset\}\right|}\sum_{T\in\mathcal{G}_F:T\subseteq S}\left|\{G\in\mathcal{G}:G\cap F=T\}\right|$$
$$= \sum_{G\in\mathcal{G}:G\cap F\neq\emptyset,\ G\cap F\subseteq S} m\left(G\right) + \frac{\left|\{G:F\cap G\neq\emptyset,\ F\cap G\subseteq S\}\right|}{\left|\{G\in\mathcal{G}:G\cap F\neq\emptyset\}\right|}\times Bel\left(F^c\right)$$

Now

$$\sum_{G\in\mathcal{G}:G\cap F\neq\emptyset,\ G\cap F\subseteq S} m\left(G\right) = \sum_{G\in\mathcal{G}:F\cap G\subseteq S} m\left(G\right) - \sum_{G\in\mathcal{G}:F\cap G=\emptyset} m\left(G\right) =$$
$$\sum_{G\in\mathcal{G}:G\cap F\cap S^c=\emptyset} m\left(G\right) - Bel\left(F^c\right) = 1 - \sum_{G\in\mathcal{G}:G\cap F\cap S^c\neq\emptyset} m\left(G\right) - Bel\left(F^c\right) =$$
$$1 - Pl(F\cap S^c) - Bel\left(F^c\right) = Bel\left(F^c\cup S\right) - Bel\left(F^c\right)$$

Therefore

$$Bel(S|F) = Bel\left(F^c\cup S\right) - \left(1 - \frac{\left|\{G:F\cap G\neq\emptyset,\ F\cap G\subseteq S\}\right|}{\left|\{G\in\mathcal{G}:G\cap F\neq\emptyset\}\right|}\right)\times Bel\left(F^c\right) =$$
$$Bel\left(F^c\cup S\right) - \left(\frac{\left|\{G\in\mathcal{G}:G\cap F\neq\emptyset\}\right| - \left|\{G:F\cap G\neq\emptyset,\ F\cap G\subseteq S\}\right|}{\left|\{G\in\mathcal{G}:G\cap F\neq\emptyset\}\right|}\right)\times Bel\left(F^c\right) =$$
$$Bel\left(F^c\cup S\right) - \frac{\left|\{G:F\cap G\cap S^c\neq\emptyset\}\right|}{\left|\{G\in\mathcal{G}:G\cap F\neq\emptyset\}\right|}\times Bel\left(F^c\right)\ \text{ as required}$$

The result for Pl then follows trivially from the relation $Pl\left(S|F\right) = 1 - Bel\left(S^c|F\right)$. □

For normalisation method three all the mass associated with the empty set is reallocated to the pair of focal elements with non-empty intersection for which $\underline{m}$ is maximal (minimal). This yields the following definition of $m\left(\bullet|F\right)$:

$$m\left(T|F\right) = \begin{cases} \sum_{G\in\mathcal{G}:G\cap F=T} m\left(G\right) + \sum_{G\in\mathcal{G}:G\cap F=\emptyset} m\left(G\right) & : \text{ if } F\cap G^0 = T \\ \sum_{G\in\mathcal{G}:G\cap F=T} m\left(G\right) & : \text{ if } T\in\mathcal{G}_F, T\neq F\cap G^0 \\ 0 & : \text{ otherwise} \end{cases}$$

where $G^0 = argmax\left\{m\left(G\right):G\cap F\neq\emptyset\right\}$ ($G^0 = argmin\left\{m\left(G\right):G\cap F\neq\emptyset\right\}$).

Theorem 50. *Let $m : 2^\Omega \to [0,1]$ be a basic belief assignment with focal sets $\mathcal{G}$ and for $F\subseteq\Omega$. If the conditional mass assignment $m\left(\bullet|F\right)$ is generated from normalisation method 3 then the corresponding conditional belief and plausibility measures defined by $\forall S\subseteq\Omega\ Bel\left(S|F\right) = \sum_{T:T\subseteq S} m\left(T|F\right)$ and $Pl\left(S|F\right) = \sum_{T:T\cap S\neq\emptyset} m\left(T|F\right)$ respectively, are equivalent to:*

$$Bel\left(S|F\right) = \begin{cases} Bel\left(F^c\cup S\right) - Bel\left(F^c\right) & : \ \textit{if } F\cap G^0 \not\subseteq S \\ Bel\left(F^c\cup S\right) & : \ \textit{if } F\cap G^0 \subseteq S \end{cases}$$

$$Pl(S|F) = \begin{cases} Pl(F \cap S) - Pl(F) + 1 & : \textit{if } F \cap G^0 \cap S \neq \emptyset \\ Pl(F \cap S) & : \textit{if } F \cap G^0 \cap S = \emptyset \end{cases}$$

Proof. Suppose $F \cap G^0 \not\subseteq S$ then:

$$Bel(S|F) = \sum_{T \in \mathcal{G}_F : T \subseteq S} m(S|F) = \sum_{G: \emptyset \neq G \cap F \subseteq S} m(G) = \sum_{G: G \cap F \subseteq S} m(G) - \sum_{G: G \cap F = \emptyset} m(G)$$

$$= \sum_{G: G \subseteq F^c \cup S} m(G) - \sum_{G: G \subseteq F^c} m(G) = Bel(F^c \cup S) - Bel(F^c)$$

Suppose $F \cap G^0 \subseteq S$ then:

$$Bel(S|F) = \sum_{T \in \mathcal{G}_F : T \subseteq S} m(S|F) = \sum_{G: \emptyset \neq G \cap F \subseteq S} m(G) + \sum_{G: G \cap F = \emptyset} m(G)$$

$$= \sum_{G: G \cap F \subseteq S} m(G) = \sum_{G: G \subseteq F^c \cup S} m(G) = Bel(F^c \cup S)$$

as required. The result for Pl then follows trivially from the relation $Pl(S|F) = 1 - Bel(S^c|F)$. □

6 Summary and Conclusions

In this paper we have investigated combination operators formed through the composition of an extension function and a normalisation method and based on the principle of maximum proximity or similarity to a consistent solution. We have argued that the latter is implicit in the decision to intersect the two sources of evidence. Different measures of distance and similarity between joint belief assignments result in different normalisation methods and subsequently different extension functions. For example, the choice of a p-norm $\| \bullet \|_s$ with $s \geq 2$ results in a normalisation method that uniformly redistributes mass from pairs of focal sets with empty intersection to those pairs with non-empty intersection. The closest consistent assignments according to these metrics are those which minimize a particular convex function of $\underline{m}(F,G) : F \cap G = \emptyset$. Alternatively, when proximity is measured by cross entropy or the $SMin$ similarity measure based on the product t-norm, the appropriate normalisation method redistributes mass proportionately as in Dempster's rule. In these cases the corresponding extension function directly minimizes inconsistency as quantified by $\sum_{(F,G):F \cap G = \emptyset} \underline{m}(F,G)$. This, together with the assumption of maximal nonspecificity, results in exactly the combination operator proposed by Cattaneo [5]. A hybrid of these two operators can be obtained from the limit metric $\| \bullet \|_{\infty}^{>c}$ where the normalisation method redistributes mass uniformly but where the extension function minimizes $\sum_{(F,G):F \cap G = \emptyset} \underline{m}(F,G)$.

For any normalisation method there is a naturally corresponding definition of conditional belief and plausibility measure. For the three normalisation methods described in section 3 the subsequent conditional belief and plausibility measures have been investigated in section 5.

In section 4 of this paper we have argued that the decision to intersect two sources of information carries with it an implicit assumption of consistency rather than one of independence as in Dempster's rule. Measuring consistency in terms of proximity to a consistent assignment results in (amongst other possibilities) the three combination operators summarised in section 4.1. These operators do unfortunately carry much higher computational costs than Dempster's rule. While it is beyond the scope of this paper to provide a detailed analysis of operator complexity we give a cursory treatment of the issue below.

The complexity of operators 1-3 stems from two sources. In the first case the extension function now requires us to solve a linear optimisation problem. However, while more costly than the independence solution this is not necessarily problematic where only two sources are being combined. The real difficulty with operators 1-3 lies in their failure to satisfy associativity (as can easily be seen from any number of counter examples). This means that multiple sources of evidence cannot be dealt with in a recursive manner but rather must be combined in a multi-dimensional joint space. However, although associativity is certainly a desirable property from a computational viewpoint it turns out to be such a strong assumption that it rules out other properties that are arguably just as intuitive. In particular Cattaneo [5] shows that there are no operators of the form given in definition 9 using either normalisation method that are both idempotent and associative no matter what extension function is employed. For completeness we reproduce Cattaneo's counter example below:

Example 51. *Let* $\Omega = \{A, \neg A\}$ *for some proposition* A *and suppose:*

$m_1(\{A\}) = \alpha,\ m_1(\{A, \neg A\}) = 1 - \alpha$ *and*
$m_2(\{\neg A\}) = \alpha,\ m_2(\{A, \neg A\}) = 1 - \alpha$
where $\alpha > 0.5$

Then for any combination operator using either normalisation method 1 or 2 [2] *we have that:*

$m_1 \oplus_{e,\nu} m_2(\{A\}) = \beta,\ m_1 \oplus_{e,\nu} m_2(\{\neg A\}) = \beta,\ m_1 \oplus_{e,\nu} m_2(\{A, \neg A\}) = 1 - 2\beta$
where $\beta < \alpha$

Now assuming both idempotence and associativity gives that:

$m_1 \oplus_{e,\nu} m_2 = (m_1 \oplus_{e,\nu} m_1) \oplus_{e,\nu} m_2 = m_1 \oplus_{e,\nu} (m_1 \oplus_{e,\nu} m_2)$

[2] Cattaneo [5] only considers normalisation method 2, however the result is easily extended to normalisation method 1

where the right hand side assignment must then be derived from a joint belief assignment of the form:

$$\underline{m}(\{A\},\{A\}) = \alpha,\ \underline{m}(\{A\},\{\neg A\}) = 0,\ \underline{m}(\{A\},\{A,\neg A\}) = 0,$$
$$\underline{m}(\{A,\neg A\},\{A\}) = \beta - \alpha,\ \underline{m}(\{A,\neg A\},\{\neg A\}) = \beta,\ \underline{m}(\{A,\neg A\},\{A,\neg A\}) = 1 - 2\beta$$

However, this is only a valid assignment if $\beta \geq \alpha$ which is an inconsistency. Hence the operator cannot be both idempotent and associative.

Clearly, then the case for a combination operator, cannot be made in terms of computational efficiency alone since properties such as associativity are likely to be inconsistent with other highly desirable logical properties. Above all there must be a strong epistemological justification for any operator and if such a case can be made then attempts to overcome computational complexity should take the form of identifying useful approximation algorithms. From this perspective, we hope to provide a more detailed investigation of complexity and its relation to the principle of maximal consistency in subsequent work.

References

[1] J.F. Baldwin, (1986), 'Support Logic Programming', *International Journal of Intelligent Systems*, Vol. 1, pp73-104

[2] J. F. Baldwin, (1991), 'Combining Evidences for Evidential reasoning'. *International Journal of Intelligent Systems*, Vol. 6, pp569-616

[3] D. Boixader, J. Jacas, J. Recasens, (1999), 'Fuzzy Equivalence Relations: Advanced Material', Fundamentals of Fuzzy Sets, (Eds. D. Dubois, H. Prade), Kluwer, pp261-290

[4] D. Dubois, H. Prade, (1986), 'On the Unicity of Dempster Rule of Combination', *International Journal of Intelligent Systems*, Vol.1, pp133-142

[5] M.E.G.V Cattaneo, (2003), 'Combining Belief Functions Issued from Dependent Sources', *Proceedings of 3rd International Symposium on Imprecise Probabilities and Their Applications*

[6] B.R. Gaines, (1978), 'Fuzzy and Probability Uncertainty Logics', *Journal of Information and Control*, Vol. 38, pp154-169

[7] R. Haenni, (2002), 'Are alternatives to Dempsters rule of combination real alternatives? Comments on 'About the belief function combination and the conflict management problem' Lefevre et al', *Information Fusion*, Vol.3, pp237-239

[8] C. Joslyn, (1994), *Possibilistic Processes for Complex Systems Modeling*, PhD Dissertation, State University of New York at Binghamton.

[9] E.P. Klement, R. Mesiar, E. Pap, (2000), *Triangular Norms*, Vol.8 of Trends in Logic, Kluwer Academic Publishers, Dordrecht.

[10] G. J. Klir, M. J.Wierman, (1999), 'Uncertainty-Based Information: Elements of Generalized Information Theory', *Studies in Fuzziness and Soft Computing*, Vol. 15, 2nd edition, Physica-Verlag.

[11] S. Kullback, R. A. Leibler, (1951), 'On information and sufficiency', *Annals of Mathematics and Statistics*, Vol. 22, pp79-86

[12] J. Lawry, (2001), 'Possibilistic Normalisation and Reasoning Under Partial Inconsistency', *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 9, No. 4, pp413-436

[13] E. Lefevre, O. Colot, P. Vannoorenberghe, (2002), 'Belief Function Combination and Conflict Management', *Information Fusion*, Vol. 3, pp149-162

[14] J. F. Lemmer, (1986), 'Confidence Factors, Empiricism and the Dempster-Shafer Theory of Evidence', *Uncertainty in Artificial Intelligence*, (Eds L.N. Kanal, J.F. Lemmer), North-Holland, Amsterdam, pp117-125

[15] W. Liu, J. Hong, (2000), 'Re-investigating Dempster's Idea on Evidence Combination', *Knowledge and Information Systems*, Vol.2, No.2

[16] G. Shafer, (1976), *A Mathematical Theory of Evidence*, Princeton University Press, Princeton NJ

[17] G. Shafer, (1981), 'Constructive Probability', *Synthese*, Vol. 48, pp1-60

[18] P. Smets, R. Kennes, (1994), 'The Transferable Belief Model', *Artificial Intelligence*, Vol. 66, pp191-234

[19] K.R. Stromberg, (1981), *An Introduction to Classical Analysis*, Wadsworth International Group, California

[20] F. Voorbraak, (1991), 'On the Justification of Dempster's rule of Combination', *Artificial Intelligence*, Vol. 48, pp171-197

[21] R.R. Yager, (1987), 'On the Dempster-Shafer Framework and New Combination Rules', *Information Science*, Vol. 41, pp93-138

[22] L. A Zadeh, (1986), 'A simple view of the DempsterShafer theory of evidence and its implication for the rule of combination', *AI Magazine* Vol. 7, pp85-90.

[23] L.A. Zadeh, (1971), 'Similarity Relations and Fuzzy Orderings', *Information Science*, Vol. 3, pp177-200