

2016

# An intelligent quality-based approach to fusing multi-source probabilistic information

Ronald R. Yager  
*Machine Intelligence Institute*

Fred Petry  
*Naval Research Laboratory*

Follow this and additional works at: <http://digitalcommons.unl.edu/usnavyresearch>

---

Yager, Ronald R. and Petry, Fred, "An intelligent quality-based approach to fusing multi-source probabilistic information" (2016). *U.S. Navy Research*. 103.  
<http://digitalcommons.unl.edu/usnavyresearch/103>

This Article is brought to you for free and open access by the U.S. Department of Defense at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in U.S. Navy Research by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



# An intelligent quality-based approach to fusing multi-source probabilistic information



Ronald R. Yager<sup>a,\*</sup>, Fred Petry<sup>b</sup>

<sup>a</sup> Machine Intelligence Institute, Iona College, New Rochelle, NY 10801, United States

<sup>b</sup> Marine Geosciences Division, Geospatial Sciences & Technology Branch, Naval Research Laboratory, Stennis Space Center, MS 39529, United States

## ARTICLE INFO

### Article history:

Received 9 July 2015

Revised 3 February 2016

Accepted 7 February 2016

Available online 16 February 2016

### Keywords:

Fusion

Entropy

Credibility

Quality-based

## ABSTRACT

Our objective here is to obtain quality-fused values from multiple sources of probabilistic distributions, where quality is related to the lack of uncertainty in the fused value and the use of credible sources. We first introduce a vector representation for a probability distribution. With the aid of the Gini formulation of entropy, we show how the norm of the vector provides a measure of the certainty, i.e., information, associated with a probability distribution. We look at two special cases of fusion for source inputs those that are maximally uncertain and certain. We provide a measure of credibility associated with subsets of sources. We look at the issue of finding the highest quality fused value from the weighted aggregations of source provided probability distributions.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

The use of fusion to combine data provided by multiple sources about the value of a variable is common in many applications [1]. One rational for fusing probabilistic distributions provided by multiple sources is to improve the quality of the information to decision makers [2]. Our interest here is looking at the problem of obtaining high quality fused values. One aspect of this quality is a reduction in the uncertainty of the information. Unfortunately, combining probability distributions information does not always result in a probability distribution with less uncertainty, this particularly is the case when the data that are being fused is conflicting. In order to formally quantify the uncertainty associated with a probability distribution we will use the concept of entropy. A second contributing factor to the association of quality with a fused value is that we have used quality sources of information, the more of these sources used, the more credible the results of the fusion process. In order to capture this criterion of a quality fusion we introduce a measure of credibility associated with use of various subsets of the sources. Here we provide a quantification of the notion of a quality fusion based on the objective of providing fused values having little uncertainty based on a credible subset of the sources.

## 2. Vector representation of probability distributions

Assume  $P_i$  is a probability distribution on the space  $X = \{x_1, \dots, x_n\}$  where  $p_{ij}$  is the probability of the occurrence of  $x_j$ . Here, each  $p_{ij} \in [0, 1]$  and  $\sum_{j=1}^n p_{ij} = 1$ . For our purposes in the following we shall find it useful, at times, to represent a probability distribution as an  $n$ -dimensional vector  $P_i = [p_{i1}, p_{i2}, \dots, p_{in}]$ . Here the vector has the special properties that all its components lie in the unit interval and their sum is one.

If  $P_i$  for  $i=1$  to  $q$  are a collection of probability distribution vectors then their weighed sum,  $P = \sum_{i=1}^q w_i P_i$ , is another vector whose components are  $p_j = \sum_{i=1}^q w_i p_{ij}$ . Furthermore, if the weights are standard weights,  $w_i \in [0, 1]$  and  $\sum_{i=1}^q w_i = 1$ , then  $P$  is also a probability distribution vector.

Another operation on vectors is the dot or inner product, see Bustince and Burillo [3]. If  $P_i$  and  $P_k$  are two probability vectors on the space  $X$  then their dot product is

$$P_i \cdot P_k = \sum_{j=1}^n p_{ij} p_{kj}.$$

We emphasize that the dot product is a scalar value. Furthermore, in the case where  $P_i$  and  $P_k$  are probability distributions then  $0 \leq P_i \cdot P_k \leq 1$ . A special case of dot product is where  $P_i$  and  $P_k$  are the same then  $P_i \cdot P_k = \sum_{j=1}^n (p_{ij})^2$ . For notational simplicity at times when it causes no confusion, we shall simply use  $P_i P_k$  for the dot product.

An important concept that is associated with this self dot product is the idea of the norm of the vector. In particular then norm

\* Corresponding author. Tel.: +1 212 249 2047.

E-mail addresses: [yager@panix.com](mailto:yager@panix.com) (R.R. Yager), [fred.petry@nrlssc.navy.mil](mailto:fred.petry@nrlssc.navy.mil) (F. Petry).

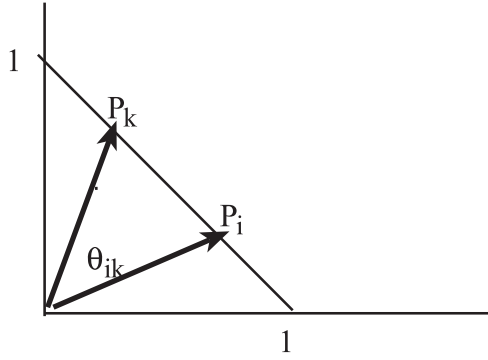


Fig. 1. Angle between probabilistic vectors.

$\|P_i\| = \sqrt{P_i P_i} = (\sum_{j=1}^n (p_{ij})^2)^{1/2}$ . The norm is referred to as the Euclidean length of a vector. Because of the special properties of the probability distribution vector,  $p_{ij} \in [0, 1]$  and  $\sum_i p_{ij} = 1$ , it can be easily shown that the maximal value of  $\|P_i\|$  occurs when one  $p_{ij} = 1$  and all other  $p_{ij} = 0$ . In this case,  $\|P_i\| = 1$ . Furthermore, in this case of a probability distribution vector the minimum value of  $\|P_i\|$  occurs when all  $p_{ij} = 1/n$  and this has the value  $\|P_i\| = (\sum_{i=1}^n (\frac{1}{n})^2)^{1/2} = (\frac{1}{n})^{1/2} = \frac{1}{\sqrt{n}}$ . We note for the self dot product,  $P_i P_i = \|P_i\|^2$  we have a maximal value of one and minimal value of  $\frac{1}{n}$  when all  $p_{ij} = \frac{1}{n}$ .

In the following we shall benefit from the use of an illustration of the probability vector in the two-dimensional case as shown in Fig. 1.

If  $P_i$  and  $P_k$  are two probability vectors it is known [4] that the Cosine of the angle between them denoted  $\theta_{ik}$  is expressed as

$$\cos(\theta_{ik}) = \frac{P_i P_k}{\|P_i\| \|P_k\|}$$

We note  $\cos(\theta_{ik})$  is the dot product of  $P_i$  and  $P_k$  divided by their respective norms. It is well known that if  $\cos(\theta_{ik}) \in [0, 1]$ , as is the case when  $P_i$  and  $P_k$  are probability distribution vectors, that  $\theta_{ik} \in [0, \frac{\pi}{2}]$ .

We further see that if  $P_i = P_k$  then  $\cos(\theta_{ik}) = \frac{P_i P_k}{\|P_i\| \|P_k\|} = \frac{P_i^2}{\|P_i\|^2} = \frac{P_i^2}{P_i^2} = 1$ . Thus if  $P_i$  and  $P_k$  are the same, coincident, then  $\cos(\theta_{ik}) = 1$ . Furthermore it is known in this case that  $\theta_{ik} = 0$ . At the other extreme is the case where  $P_i$  and  $P_k$  are orthogonal,  $P_i P_k = \sum_{j=1}^n p_{ij} p_{kj} = 0$  where  $\cos(\theta_{ik}) = \frac{P_i P_k}{\|P_i\| \|P_k\|} = 0$ . We get in this case that  $\theta_{ik} = \frac{\pi}{2}$ . We note that in the case where  $P_i$  and  $P_k$  are orthogonal then  $p_{ij} = 0$  when  $p_{ik} \neq 0$  and  $p_{ik} = 0$  when  $p_{ij} \neq 0$ .

We illustrate these extremes of coincident and orthogonal distributions for the two dimensional case in Fig. 2.

We note in the  $n$ -dimensional case a prototype example of orthogonality occurs when  $P_i$  has  $p_{ij1} = 1$  and  $P_k$  is  $p_{kj2} = 1$ . Here they each completely support different outcomes.

In [5] we suggested that  $\cos(\theta_{ik})$  can be used as measure of the degree of compatibility, Comp, between the two probability distributions, thus

$$\text{Comp}(P_i, P_k) = \frac{P_i P_k}{\|P_i\| \|P_k\|}$$

Here  $\text{Comp}(P_i, P_k) \in [0, 1]$  and the closer to one the more compatible the probability distributions. Furthermore  $1 - \text{Comp}(P_i, P_k)$ , denoted  $\text{Conf}(P_i, P_k)$ , can be seen as the degree of conflict between the two probability distributions. We note that if  $P_i$  and  $P_k$  are orthogonal then  $\text{Comp}(P_i, P_k) = 0$  that  $\text{Conf}(P_i, P_k) = 1$ . On the other hand if  $P_i$  and  $P_k$  are coincident, the same, then  $\text{Comp}(P_i, P_k) = 1$  and  $\text{Conf}(P_i, P_k) = 0$ .

An interesting special case occurs when one of the distributions,  $P_i$ , has  $p_{ij} = \frac{1}{n}$  for all  $j$ . Here we previously noted  $\|P_i\| = (\frac{1}{n})^{1/2}$ . Consider now  $\text{Comp}(P_i, P_k)$  where  $P_i$  is this uniform probability distribution. Here  $\text{Comp}(P_i, P_k) = \frac{P_i P_k}{\|P_i\| \|P_k\|}$ . However in this case

$$P_i P_k = \sum_{j=1}^n p_{ij} p_{kj} = \frac{1}{n} \sum_{i=1}^n p_{kj} = \frac{1}{n}$$

and thus  $\text{Comp}(P_i, P_k) = \frac{\frac{1}{n}}{\|P_k\| (\frac{1}{n})^{1/2}} = \frac{(\frac{1}{n})^{1/2}}{\|P_k\|} = \frac{1}{\sqrt{n}} = \frac{1}{\|P_k\|}$ . Two special cases of  $P_k$  are worth commenting on. If  $P_k$  is a certain distribution, it has  $p_{kj} = 1$  for one element, then  $\|P_k\| = 1$  and  $\text{Comp}(P_i, P_k) = \frac{1}{\sqrt{n}}$ . If  $P_k$  is also a uniform probability distribution, all  $p_{kj} = \frac{1}{n}$ , here then  $\|P_k\| = \frac{1}{\sqrt{n}}$  and we get  $\text{Comp}(P_i, P_k) = 1$ .

### 3. Entropy, certainty and information

An important concept associated with a probability distribution on the space  $X = \{x_1, \dots, x_n\}$  is the idea of entropy [6,7]. The most common measure of entropy is the Shannon entropy. Here if  $P$  is a probability distribution on the space with  $p_j$  the probability associated with  $x_j$ , then the Shannon entropy is  $H(P) = -\sum_{j=1}^n p_j \ln(p_j)$ . It is well known that the maximal entropy occurs when all  $p_i = \frac{1}{n}$  in which case  $H(P) = \ln(n)$ . The minimal entropy occurs for the case when one  $p_j = 1$  and all other  $p_j = 0$ , in this case  $H(P) = 0$ . What is clear is that the entropy is measuring the uncertainty associated with the probability distribution, the more uncertainty the more entropy. The complimentary idea of entropy is certainty (or

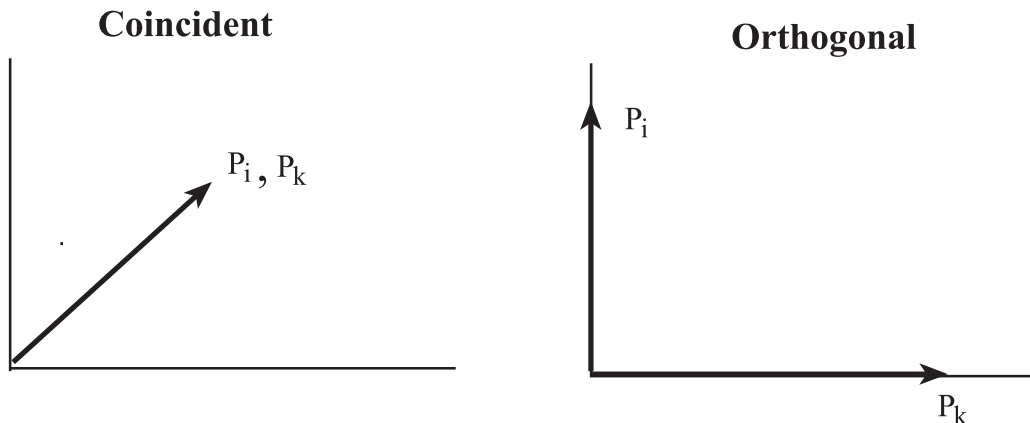


Fig. 2. Different relationships between probabilistic distributions.

information). The smaller the entropy the more information conveyed by a probability distribution. What should be clear is that for decision-making purposes we prefer distributions with smaller entropy as we have less uncertainty, more information.

While the Shannon entropy is the most well known formulation of entropy other formulations have been suggested [8–11], particularly in an attempt to overcome the complexity involved in working with the ln. One successful alternative formulation is the Gini entropy [12] which is much simpler to work with. The Gini entropy of the probability distribution  $P$  on  $X$  is

$$G(P) = 1 - \sum_{j=1}^n p_j^2$$

The Gini entropy is coincidental with the Shannon entropy in the sense the  $G(P)$  assumes its maximal value when all  $p_j = \frac{1}{n}$  and assumes its minimum value when one  $p_j = 1$ . While the Shannon entropy is based on the term  $p_j \ln(p_j)$  the Gini entropy is based on the term  $p_j^2$ .

In the following we shall find it convenient to use the Gini entropy. Again we see that bigger the value of  $G(P)$  the more uncertainty in the knowledge provided by the probability distribution. On the other hand the smaller  $G(P)$  the more certainty in the knowledge provided by the probability distribution, it is common to refer to this as being more information. Thus to increase certainty or information we decrease the entropy.

What is very notable about the Gini formulation for entropy is that it very clearly indicates what is needed to decrease entropy, increase the certainty or information. In particular, by increasing  $\sum_{j=1}^n p_j^2$  we decrease the entropy or uncertainty, we increase information.

Furthermore, in using the vector representation of the probability distribution  $P$  the norm of  $P$ ,  $\|P\| = (\sum_{j=1}^n p_j^2)^{1/2}$ . Thus we see that increasing the norm of the distribution serves to decrease the entropy, increase its information content. We note that  $\|P\|$  takes its maximal value of 1 with  $p_j = 1$  for some  $x_j$  and all other  $p_j = 0$ . In addition the Min value of  $\|P\|$  under the constraint that all  $p_j \in [0, 1]$  and  $\sum_{j=1}^n p_j = 1$  occurs when all  $p_j = 1/n$ .

Since our interest will be on obtaining probability distributions with more information we shall focus on how the term  $\sum_{j=1}^n p_j^2$  or its norm is affected by various operations. We shall refer to term  $\|P\|^2$  as the NegEnt and semantically see it as a measure of information or certainty. Here again we note that  $\frac{1}{n} \leq \|P\|^2 \leq 1$  and  $\frac{1}{\sqrt{n}} \leq \|P\| \leq 1$ .

Assume  $P$  and  $Q$  are two probability vectors on the space  $X$  and that the relation between these is

$$\begin{aligned} q_1 &= p_1 - \alpha \\ q_2 &= p_2 + \alpha \\ q_j &= p_j \quad \text{for } j = 3 \text{ to } n \end{aligned}$$

for  $\alpha \geq 0$ . We see here

$$\sum_{j=1}^n q_j^2 = (p_1 - \alpha)^2 + (p_2 + \alpha)^2 + \sum_{j=3}^n p_j^2$$

We note that since  $(p_1 - \alpha)^2 = p_1^2 - 2\alpha p_1 + \alpha^2$  and  $(p_2 + \alpha)^2 = p_2^2 + 2\alpha p_2 + \alpha^2$  we get

$$\sum_{j=1}^n q_j^2 - \sum_{j=1}^n p_j^2 = 2\alpha_1(p_2 - p_1) + 2\alpha^2 = 2\alpha((p_2 - p_1) + \alpha)$$

We observe that if  $p_2 > p_1$  then  $Q^2 > P^2$  and we have decreased the entropy, increased the certainty. Thus moving some amount of probability from an element with less probability to one of greater probability increases the norm, the certainty. We also observe that even if  $p_1 = p_2$  then  $Q^2 > P^2$ . Thus if two elements have the same

probability moving some probability from one element to the other increases the norm, increases the certainty. If  $p_2 < p_1$  the situation is more complex. If  $\alpha < |p_1 - p_2|$  then we decrease certainty but if  $\alpha > |p_1 - p_2|$  we increase the certainty, in this case we have moved enough probability to overcome the initial difference and essentially reversed the relationship.

We consider another related situation. Assume  $P = [p_1, p_2, \dots, p_n]$  is a probability distribution such that two elements,  $p_1$  and  $p_2$  share an amount of probability  $\Delta$ . We ask what is the allocation of the probability  $\Delta$  between  $p_1$  and  $p_2$  that results in the largest norm, the most certainty, smallest entropy. We see  $P^2 = \sum_{j=1}^n p_j^2 = p_1^2 + p_2^2 + \sum_{j=3}^n p_j^2$ . Assume we assign  $a$  to one of  $p_1$  or  $p_2$  and  $\Delta - a$  to the other. In this case

$$\begin{aligned} P^2 &= a^2 + (\Delta - a)^2 + \sum_{j=3}^n p_j^2 = a^2 + \Delta^2 - 2a\Delta + a^2 + \\ &= 2a^2 - 2a\Delta + \Delta^2 + \sum_{j=3}^n p_j^2 \end{aligned}$$

Taking the derivative of  $P$  with respect to  $a$  we get  $\frac{\partial P^2}{\partial a} = 4a - 2\Delta$ . Setting this to zero gives  $a = \frac{\Delta}{2}$ . To find the maximum of  $P^2$  we must evaluate  $P^2$  at  $a = \frac{\Delta}{2}$  and the two end points,  $a=0$  and  $a=\Delta$  thus

$$\begin{aligned} \mathbf{a} \quad \mathbf{P}^2 \\ 0 \quad &\Delta^2 + \sum_{j=3}^n p_j^2 \\ \frac{\Delta}{2} \quad &= \frac{\Delta^2}{2} + \sum_{j=3}^n p_j^2 \\ \Delta \quad &\Delta^2 + 0 + \sum_{j=3}^n p_j^2 \end{aligned}$$

Thus we see that the maximal value of  $P^2$  occurs when we give all the  $\Delta$  to one of the components.

A concept closely related to entropy is the idea of cross entropy [13,14]. If  $P$  and  $Q$  are two probability distributions on the space  $X = \{x_1, \dots, x_n\}$  then the standard definition of cross entropy is  $H(P, Q) = -\sum_{j=1}^n p_j \ln(q_j)$ . We see this definition is in the spirit of the Shannon measure of entropy. Here we can define a related measure in the spirit of the Gini index. In particular  $G(P, Q) = 1 - \sum_{j=1}^n p_j q_j$ . The larger  $G(P, Q)$  the more different the probability distributions, it is a kind of measure of difference between the distributions. We note here that using the vector representation we can express  $\sum_{j=1}^n p_j q_j$  as the dot product  $PQ$ . Furthermore we see that  $PQ$  is a kind of measure of relatedness of the two distributions,  $PQ$  is the negation of Gini measure of cross entropy. We note that as opposed to the Shannon type measure the Gini type measure is symmetric,  $G(P, Q) = G(Q, P)$ .

#### 4. Information in maximally certain and uncertain distribution

Consider now two probability distributions on  $X$ ,  $P = [p_1, \dots, p_n]$  and  $Q = [q_1, \dots, q_n]$ . Their linear aggregation  $R = w_1 P + w_2 Q$  is a probability distribution when  $w_1 + w_2 = 1$ . Here for each component  $r_j$  of  $R$  we have,  $r_j = (w_1 p_j + w_2 q_j)$ . To calculate the information associated with  $R$  we calculate

$$\begin{aligned} \|R\|^2 &= \sum_{j=1}^n (w_1 p_j + w_2 q_j)^2 = \sum_{j=1}^n (w_1^2 p_j^2 + w_2^2 q_j^2 + 2w_1 w_2 p_j q_j) \\ \|R\|^2 &= w_1^2 \sum_{j=1}^n p_j^2 + w_2^2 \sum_{j=1}^n q_j^2 + 2w_1 w_2 \sum_{j=1}^n p_j q_j \\ \|R\|^2 &= w_1^2 \|P\|^2 + w_2^2 \|Q\|^2 + 2w_1 w_2 PQ \end{aligned}$$

where  $PQ$  denotes the dot product of  $P$  and  $Q$ .

In passing we note that since  $w_2 = 1 - w_1$  we have

$$w_1^2 + w_2^2 + 2(w_1 w_2) = w_1^2 + (1 - w_1)^2 + 2(w_1)(1 - w_1) = w_1^2 + (1 - 2w_1 + w_1^2) + 2w_1 - 2w_1^2 = 1$$

Thus  $w_1^2 + w_2^2 + 2w_1 w_2 = 1$  and hence  $\|R\|^2$  is a weighted average of components  $\|P\|^2$ ,  $\|Q\|^2$  and  $PQ$ .

We also observe that since  $\cos(P, Q) = \frac{PQ}{\|P\| \|Q\|}$  we have

$$\|R\|^2 = w_1^2 \|P\|^2 + w_2^2 \|Q\|^2 + 2w_1 w_2 \|P\| \|Q\| \cos(P, Q)$$

where  $\cos(P, Q)$  is the compatibility of  $P$  and  $Q$ . Thus the more compatible, the less conflict, the bigger the NegEnt, more information, contained in  $R$ .

In the special case where  $w_1 = w_2 = \frac{1}{2}$  we have

$$\|R\|^2 = \frac{1}{4} \|P\|^2 + \frac{1}{4} \|Q\|^2 + \frac{1}{2} PQ = \frac{1}{4} \|P\|^2 + \frac{1}{4} \|Q\|^2 + \frac{1}{2} \|P\| \|Q\| \cos(P, Q)$$

Let us focus on this case where  $w_1 = w_2 = \frac{1}{2}$ .

If  $P$  and  $Q$  are two certain probability distributions, then  $\|P\|$  and  $\|Q\| = 1$  and there are two cases of interest. The first case is when they are completely compatible,  $p_j = q_j = 1$  for some  $j$ . Here  $\cos(P, Q) = 1$  and  $\|R\|^2 = \frac{1}{4}(1) + \frac{1}{4}(1) + \frac{1}{2}(1) = 1$ . On the other hand if they are completely conflicting,  $p_j = 1$  and  $q_k = 1$ , then  $\cos(P, Q) = 0$  and  $\|R\|^2 = \frac{1}{4} + \frac{1}{4} + 0 = \frac{1}{2}$ .

Assume  $P$  and  $Q$  are both maximally uncertain distributions, they have  $p_j = q_j = 1/n$  for all  $j$ . Here  $\|P\| = \|Q\| = (1/n)^{1/2}$  and since we have shown in this case that  $\cos(P, Q) = 1$  then

$$\|R\|^2 = \left(\frac{1}{4}\right) \frac{1}{n} + \left(\frac{1}{4}\right) \frac{1}{n} + \frac{1}{2} \left(\frac{1}{n}\right)^{1/2} \left(\frac{1}{n}\right)^{1/2} = \frac{1}{n}$$

If  $P$  is a completely certain distribution,  $\|P\| = 1$ , and  $Q$  is a completely uncertain distribution,  $\|Q\| = (1/n)^{1/2}$ , then  $PQ = \frac{1}{n}$  and  $\|R\|^2 = \frac{1}{4}(1) + \frac{1}{4} \frac{1}{n} + \frac{1}{2} \frac{1}{n} = \frac{n+3}{4n}$ .

Let us now consider the case where we have  $t$  probability distributions  $P_1, \dots, P_t$  where  $P_i = [p_{i1}, \dots, p_{in}]$ . Assume now  $R = \frac{1}{t} \sum_{i=1}^t P_i$ . Here then each component of  $R$ ,  $r_j = \frac{1}{t} \sum_{i=1}^t p_{ij}$ . In this case

$$\|R\|^2 = \sum_{j=1}^n r_j^2 = \sum_{j=1}^n \left(\frac{1}{t^2}\right) (p_{1j} + p_{2j} + \dots + p_{tj})^2$$

$$\|R\|^2 = \sum_{i=1}^t \frac{1}{t^2} (\|P_i\|^2) + \sum_{i=1}^t \sum_{\substack{k=1 \\ k \neq i}}^t \frac{1}{t^2} P_i \cdot P_k$$

$$\|R\|^2 = \sum_{i=1}^t \frac{1}{t^2} \|P_i\|^2 + \frac{2}{t^2} \sum_{i=1}^{t-1} \sum_{k=i+1}^t P_i \cdot P_k$$

To understand how we obtained this last formula from  $\sum_{j=1}^n \left(\frac{1}{t^2}\right) (p_{1j} + p_{2j} + \dots + p_{tj})^2$  we illustrate the situation for  $t=4$ . Consider the term,  $(p_{1j} + p_{2j} + p_{3j} + p_{4j})^2$ . The value of the squared sum can be best viewed in using the following matrix

$$\begin{matrix} & P_{1j} & P_{2j} & P_{3j} & P_{4j} \\ P_{1j} & \left[ \begin{matrix} p_{1j}p_{1j} & p_{1j}p_{2j} & p_{1j}p_{3j} & p_{1j}p_{4j} \\ p_{2j}p_{1j} & p_{2j}p_{2j} & p_{2j}p_{3j} & p_{2j}p_{4j} \\ p_{3j}p_{1j} & p_{3j}p_{2j} & p_{3j}p_{3j} & p_{3j}p_{4j} \\ p_{4j}p_{1j} & p_{4j}p_{2j} & p_{4j}p_{3j} & p_{4j}p_{4j} \end{matrix} \right] \end{matrix}$$

The value of the squared sum is equal to the sum of all the  $4 \times 4 = 16$  terms in the matrix. We can consider the matrix as consisting of three parts, the main diagonal and the upper and lower triangles. The sum of the term on the main diagonal is  $p_{1j}p_{1j} + p_{2j}p_{2j} + p_{3j}p_{3j} + p_{4j}p_{4j}$ . If we calculate this sum of these

overall  $j=1$  to  $n$  we get

$$\sum_{j=1}^n p_{1j}^2 + \sum_{j=1}^n p_{2j}^2 + \sum_{j=1}^n p_{3j}^2 + \sum_{j=1}^n p_{4j}^2 = \sum_{i=1}^4 \|P_i\|^2$$

Consider now the upper and lower triangle. First note that they consist of exactly the same elements. So to find the sum of elements in the upper and lower triangle we need just calculate the sum of one of them and multiply by two. Consider the upper triangle

$$p_{1j}p_{2j} + p_{1j}p_{3j} + p_{1j}p_{4j} + p_{2j}p_{3j} + p_{2j}p_{4j} + p_{3j}p_{4j}$$

With a little thought we see that we can express this sum as  $\sum_{i=1}^3 (\sum_{k=i+1}^4 p_{ij}p_{kj})$ . If we now sum these over  $j=1$  to  $n$  we get  $\sum_{j=1}^n (\sum_{i=1}^3 (\sum_{k=i+1}^4 p_{ij}p_{kj})) = \sum_{i=1}^3 (\sum_{k=i+1}^4 (\sum_{j=1}^n p_{ij}p_{kj}))$ . However we observe that  $\sum_{j=1}^n p_{ij}p_{kj}$  to the dot product  $P_i \cdot P_k$ . Thus the sum of the elements in the upper triangle is  $\sum_{i=1}^3 \sum_{k=i+1}^4 P_i \cdot P_k$ . Doubling this we get the sum of upper and lower triangles,  $2 \sum_{i=1}^3 \sum_{k=i+1}^4 P_i \cdot P_k$ . Thus we get  $\|R\|^2 = \frac{1}{t^2} \sum_{i=1}^t \|P_i\|^2 + \frac{2}{t^2} \sum_{i=1}^{t-1} \sum_{k=i+1}^t P_i \cdot P_k$

We now consider the calculation of  $\|R\|^2 = \frac{1}{t^2} \sum_{i=1}^t \|P_i\|^2 + \frac{2}{t^2} \sum_{i=1}^{t-1} \sum_{k=i+1}^t P_i \cdot P_k$  for some special cases. First we consider the case where all  $t$  probability distributions are certain distributions but may be conflicting in that they may be focused on different  $x_j$ . In this case  $\|P_i\|^2 = 1$  for all  $i$ . We now partition the probability distributions into groups of agreement. For simplicity we let  $g_j$  be the number of probability distributions that are focused on  $x_j$ , these distributions are compatible. We note that  $\sum_{i=1}^n g_j = t$ . We observe that if a pair of certain distributions are in agreement then  $P_i \cdot P_k = 1$  otherwise  $P_i \cdot P_k = 0$ . We see that  $\frac{2}{t^2} \sum_{i=1}^t \sum_{k=i+1}^t P_i \cdot P_k$  can be seen as being equal to the  $\frac{2}{t^2}$  times number of pairs of distribution that are compatible. If  $g_j$  are focused on  $x_j$  then there are  $\frac{g_j!}{2!(g_j-2)!} = \frac{g_j(g_j-1)}{2} = \frac{g_j^2 - g_j}{2}$  pairs of elements in agreement. Thus in this case of all certain probability distributions we have

$$\|R\|^2 = \frac{1}{t^2} t + \frac{2}{t^2} \sum_{j=1}^n \frac{g_j(g_j-1)}{2} = \frac{1}{t} + \frac{1}{t^2} \sum_{j=1}^n g_j(g_j-1)$$

If all the elements agree on the same value,  $x_1$ , then  $g_1 = t$  and all other  $g_j = 0$  and we get

$$\|R\|^2 = \frac{1}{t} + \frac{1}{t^2} (t)(t-1) = \frac{1}{t} + \frac{t^2 - t}{t^2} = \frac{1}{t} + 1 - \frac{1}{t} = 1$$

Assume half the distributions agree on one value and the other half agree on a second value. Here we have  $g_1 = t/2$  and  $g_2 = t/2$  (we assume  $t$  even for simplicity) then we get

$$\|R\|^2 = \frac{1}{t} + \frac{1}{t^2} (2) \left(\frac{t}{2}\right) \left(\frac{t}{2} - 1\right) = \frac{1}{t} + \frac{1}{t} \left(\frac{t}{2} - 1\right) = \frac{1}{t} (1 + t/2 - 1) = 0.5$$

If one third of the distributions agree on different values,  $g_1 = g_2 = g_3 = t/3$  then

$$\|R\|^2 = \frac{1}{t} + \frac{1}{t^2} (3) \left(\frac{t}{3}\right) \left(\frac{t}{3} - 1\right) = \frac{1}{t} + \frac{1}{t} \left(\frac{t}{3} - 1\right) = \frac{1}{t} \left[1 + \frac{t}{3} - 1\right] = \frac{1}{3}$$

The value of  $\|R\|^2$  for complex allocations of the  $g_j$  can be calculated using our formula  $\|R\|^2 = \frac{1}{t} + \frac{1}{t^2} \sum_{j=1}^n g_j(g_j-1)$ .

Let us return to our formula  $\|R\|^2 = \sum_{i=1}^t \frac{1}{t^2} \|P_i\|^2 + \frac{2}{t^2} \sum_{i=1}^t \sum_{k=i+1}^t P_i \cdot P_k$  and consider the situation where we have two classes of probability distributions. One being as in the preceding a certainty distribution, one of its components is one. The other is a

pure uncertainty distribution here all elements are  $1/n$ . First we note for  $P_i$  that has certainty, then  $\|P_i\|^2=1$ . On the other hand for any  $P_i$  that is pure uncertainty we have shown that  $\|P_i\|^2=1/n$ . Consider now the dot products. As in the preceding if both  $P_i$  and  $P_k$  are certainty then  $P_i P_k=1$  if they agree on the certainty element and  $P_i P_k=0$  if they disagree. Consider the situation when one of  $P_i P_k$  is pure uncertainty, for example  $P_i$ . If the other distribution  $P_k$  is certainty, we have  $P_i P_k = \sum_{j=1}^n P_{ij} P_{kj} = 1 \cdot \frac{1}{n} = \frac{1}{n}$ . If the other distribution is all pure uncertainty we get  $P_i P_k = \sum_{j=1}^n P_{ij} P_{kj} = \frac{n}{n^2} = 1/n$ . Thus we see independent of the second probability distribution if one of  $P_i$  or  $P_k$  is pure uncertainty then  $P_i P_k = \frac{1}{n}$ .

Let us now consider the calculation of  $\|R\|^2$  in mixed case. Assume  $t_1$  of the distributions are pure certainty and  $t_2$  are pure uncertainty. Furthermore, assume for  $j=1$  to  $n$  that  $g_j$  of the pure certainty distributions agree on the same value,  $x_j$ , here  $\sum_{j=1}^n g_j = t_1$ . In this case we get

$$\|R\|^2 = \frac{1}{t^2} \left( t_1 + \frac{t-t_1}{n} \right) + \frac{2}{t^2} \frac{1}{n} S1 + \frac{2}{t^2} S2$$

where  $S1 = \#$  of pairs containing at least a pure uncertainty distribution and  $S2 = (\#$  of pairs of two pure certainty distributions in agreement). We have already calculated the last term in the preceding, that is

$$\frac{2}{t^2} S2 = \frac{1}{t^2} \sum_{j=1}^n g_j (g_j - 1) = \frac{1}{t^2} \sum_{j=1}^n (g_j^2 - g_j)$$

We must now calculate the number containing a pure uncertainty,  $S1$ .

First we note given  $t$  distributions there are  $\frac{(t)(t-1)}{2}$  possible pairs. Given that there are  $t_1$  distributions with pure certainty then there are  $\frac{(t_1)(t_1-1)}{2}$  pairs of elements consisting of two pure certain distributions. From this we can conclude that  $S1$ , # of pairs containing at least one pure uncertainty distribution is

$$\begin{aligned} S1 &= \frac{(t)(t-1)}{2} - \frac{(t_1)(t_1-1)}{2} = \frac{(t^2 - t_1^2) - (t - t_1)}{2} \\ &= \frac{(t-t_1)(t+t_1) - (t-t_1)}{2} \\ S1 &= \frac{(t-t_1)(t+t_1-1)}{2} \end{aligned}$$

Combining all of these we get

$$\begin{aligned} \|R\|^2 &= \frac{1}{t^2} \left[ t_1 + (t-t_1) \frac{1}{n} + \frac{1}{n} ((t-t_1)(t+t_1-1)) \right. \\ &\quad \left. + \sum_{j=1}^n g_j (g_j - 1) \right] \\ \|R\|^2 &= \frac{1}{t^2} \left[ t_1 + \sum_{j=1}^n g_j (g_j - 1) + \frac{1}{n} ((t-t_1)(t+t_1-1)) \right] \\ \|R\|^2 &= \frac{1}{t^2} \left[ t_1 + \sum_{j=1}^n g_j (g_j - 1) + \frac{1}{n} (t^2 - t_1^2) \right] \end{aligned}$$

We note in the special case where all the pure certain distributions agree,  $g_1=t_1$  and all other  $g_j=0$  we get

$$\begin{aligned} \|R\|^2 &= \frac{1}{t^2} \left[ t_1 + t_1(t_1 - 1) + \frac{1}{n} (t^2 - t_1^2) \right] \\ \|R\|^2 &= \frac{1}{t^2} \left[ t_1^2 + \frac{1}{n} (t^2 - t_1^2) \right] \end{aligned}$$

### 5. Fusion of probability distributions

We now turn to our major interest, the fusion of multi-source probabilistic information. Assume  $V$  is a variable that takes its

value in the space  $X=\{x_1, \dots, x_n\}$ . In the following we let  $P_i$  be a probability distribution on  $X$  indicating the information provided by source  $i$  regarding the value of  $V$ . Here we let each  $P_i = [p_{i1}, \dots, p_{ij}, \dots, p_{in}]$ .

If we have  $t$  probability distributions,  $P_i$  for  $i=1$  to  $t$ , then the distribution basic uniform fusion of these is a probability distribution  $R = \frac{1}{t} \sum_{i=1}^t P_i$ . Each component of  $R$  is  $r_j = \frac{1}{t} \sum_{i=1}^t P_{ij}$ . We have previously shown the associated NegEnt is

$$\|R\|^2 = \frac{1}{t^2} \left[ \sum_{i=1}^t \|P_i\|^2 + 2 \sum_{i=1}^{t-1} \sum_{k=i+1}^t P_i P_k \right]$$

The larger  $\|R\|^2$  the more information provided by the fusion.

We note  $\sum_{i=1}^{t-1} \sum_{k=i+1}^t P_i P_k$  consists of  $\frac{(t)(t-1)}{2}$  terms. Thus we see total number of terms being combined is  $t + \frac{2(t)(t-1)}{2} = t^2$ . Thus, this is a simple weighted average. Since each term is contained in  $[0, 1]$  and since this a simple weighted average then  $\|R\|^2 \in [0, 1]$ .

For this aggregation we can calculate the average conflict between the components. We first recall the degree of conflict between  $P_i$  and  $P_k$  is

$$\text{Conf}(P_i, P_k) = 1 - \cos(\theta_{ik}) = 1 - \frac{P_i P_k}{\|P_i\| \|P_k\|}$$

Given that there are  $t$  distributions being combined then there are  $\frac{(t)(t-1)}{2}$  distinct pairs of distributions. Thus the average conflict in this fusion is

$$\begin{aligned} \text{Ave Conf}(P_i, P_k) &= \frac{2}{(t)(t-1)} \sum_{i=1}^{t-1} \sum_{k=i+1}^t \left( 1 - \frac{P_i P_k}{\|P_i\| \|P_k\|} \right) \\ &= 1 - \frac{2}{(t)(t-1)} \sum_{i=1}^{t-1} \sum_{k=i+1}^t \frac{P_i P_k}{\|P_i\| \|P_k\|} \end{aligned}$$

We note here since each  $\text{Conf}(P_j, P_k) \in [0, 1]$  then  $\text{AveConf}(P) \in [0, 1]$ .

Let  $P_i$  for  $i=1$  to  $t$ , be a collection of probability distributions provided by multiple sources. Assume our purpose in combining these is to obtain a fused estimate for the value of  $V$  that gives us the most information about  $V$ . That is, we want our fused value to have a high NegEnt value,  $\|R\|^2$ . We can make some observations about  $\|R\|^2$ . Pairs of probability distributions that are non-conflicting,  $P_i P_k$  large, tend to increase the NegEnt, the information, supplied by the aggregation. On the other hand those pairs with small compatibility,  $P_i P_k$  small, can tend to decrease the NegEnt. This reduction results from the fact that while pairs with small  $P_i P_k$  may add a little to the sum of the  $P_i P_k$ , they affect the value  $\frac{1}{t^2}$ . We note that  $t^2 = t + t(t-1)$  it is the number of probability distributions plus the number of pairs. Thus while a conflicting pair does not much affect the sum  $\sum_{i=1}^{t-1} \sum_{k=i+1}^t P_i P_k$  it can reduce the  $\|R\|^2$  because it is counted in the  $t^2$ .

It appears that one approach to obtaining fused values that have high NegEnt is to only fuse the probability distributions that have a high compatibility. However, by just looking at any fusion consisting of a subset of the probability distributions we are clearly losing credibility, persuasiveness, in the fusion. In order to take this into account we shall introduce a set measure that indicates the credibility of a fusion based on a simple weighted average of a subset of probability distributions. If  $Z = \{P_1, \dots, P_t\}$  the set of available probability distributions then we define the set measure  $\text{Cred}: 2^Z \rightarrow [0, 1]$  such that for any subset  $B$  of  $Z$ ,  $\text{Cred}(B)$  indicates the credibility of a fusion based on only the distributions in  $B$ . We see that natural properties to require of such a set measure are the following: 1.  $\text{Cred}(\emptyset) = 0$ , 2.  $\text{Cred}(Z) = 1$  and 3. If  $A \subseteq B$  then  $\text{Cred}(A) \leq \text{Cred}(B)$ .

Actually, more precisely, the credibility measure should be over the space of sources. However since there is a one to one

correspondence between a source and its provided probability distribution for simplicity of notation we shall continue to refer to this credibility as being over the set of probability distributions.

Let us look at some notable examples of credibility functions. One fundamental example is the case where  $Cred(Z)=1$  and  $Cred(B)=0$  for  $B \neq Z$ . Here the only fusion that has any credibility is the one using all of the supplied probability distributions. Another type of credibility function can be based on the requirement that “at least  $\lambda\%$ ” of the probability distributions are included in the fusion. Here  $Cred(B)=1$  if  $\frac{Cred(B)}{t} \geq \lambda$  and  $Cred(B)=0$  otherwise. Closely related is a credibility function that requires that “most” of the probability distributions are included in the fusion. Here we can represent “most” as a fuzzy subset  $M$  on the unit interval [15]. In this case for any subset  $B$  of  $Z$  we can obtain  $Cred(B)=M(\frac{Cred(B)}{Z})$  the membership grade of  $\frac{Cred(B)}{Z}$  in the fuzzy subset  $M$ .

The preceding examples of credibility functions have not taken into account any distinction between the sources providing the probability distributions. Another type of credibility function can be obtained if we associate with each  $P_i$  a value  $\alpha_i \in [0, 1]$  so that  $\sum_{i=1}^t \alpha_i = 1$ . Here  $\alpha_i$  can be seen as some indication of the importance of  $P_i$ . Using these importance we can obtain  $Cred(B) = \sum_{P_i \in B} \alpha_i$ .

Another class of credibility functions can be obtained as follows. Let  $F_i$  for  $i=1$  to  $q$  be a collection of subsets of  $Z$ . We note that formally this collection need not satisfy any special requirements, that is, they do not have to be disjoint or cover the whole space  $Z$ . Further we associate with each  $F_i$  a value  $\alpha_i \in [0, 1]$  so that  $\sum_{i=1}^q \alpha_i = 1$ . We now can use this collection of subsets of  $Z$  to construct various kinds of credibility functions. Let  $Poss(F_i/B)=1$  if  $F_i \cap B \neq \emptyset$  and  $Poss(F_i/B)=0$  if  $F_i \cap B = \emptyset$ . Using this we can obtain  $Cred(B) = \sum_{i=1}^q \alpha_i Poss(F_i/B)$ . Here we have associated with the collection of sources  $q$  categories and if a fusion contains an element from category  $F_i$  it get  $\alpha_i$  points.

Related to this is another credibility function using  $Cert(F_i/B)$  which is defined so that  $Cert(F_i/B)=1$  if  $F_i \subseteq B$  and  $Cert(F_i/B)=0$  if  $F_i \not\subseteq B$ . Using this we can obtain  $Cred(B) = \sum_{i=1}^q \alpha_i Cert(F_i/B)$ . Here again we have  $q$  categories and a fusion using  $B$  distributions however we get  $\alpha_i$  credibility points if the fusion contains all the elements in a category  $F_i$ .

Many different types of functions can be constructed to reflect various complex relations regarding the credibility of subsets of probability distribution.

We note that we can use the Takagi-Sugeno [16] approach to fuzzy rule based modeling to build credibility functions.

As an example of this fusion approach we consider a scenario of probability distributions of spatial locations in a search and rescue mission. Information from differing sources is common, but it is very important to make good decisions as to how to fuse such information for most likely locations due to both the need for timely rescue and associated search costs.

Specifically we assume there are three distributions,  $P_1, P_2$  and  $P_3$  that have source information relative to four potential spatial locations  $(x_1, x_2, x_3, x_4)$  for the search. For example the first two distributions might have been obtained from UAVs or a search plane. The third which differs somewhat was obtained from local officials who, from their previous rescue experiences, provide what they believe are the probabilities for the four search locations.

Now we examine these distributions determining their conflicts and the information (NegEnt) provided in the fusions. The distributions are:

$$P_1 : (.5, .2, .2, .1); \quad P_2 : (.4, .3, .2, .1); \quad P_3 : (.1, .2, .1, .6).$$

Then

$$||P_1|| = (.34)1/2 = .583; ||P_1||^2 = .34$$

$$||P_2|| = (.3)1/2 = .547; ||P_2||^2 = .3$$

$$||P_3|| = (.42)1/2 = .648; ||P_3||^2 = .42$$

$$P_1 \cdot P_2 = .31; P_1 \cdot P_3 = .17; P_2 \cdot P_3 = .18$$

So now we can calculate the conflict  $Conf(i,j) = 1 - (P_i \cdot P_j) / (||P_i|| \times ||P_j||)$

$$Conf(1,2) = 1 - .31 / (.583 \times .548) = 1 - .972 = .028$$

$$Conf(1,3) = 1 - .17 / (.583 \times .648) = 1 - .450 = .550$$

$$Conf(2,3) = 1 - .18 / (.548 \times .648) = 1 - .507 = .493$$

These conflicts are compatible with our intuitions by examining the differences in the location probabilities particularly with respect to  $P_3$ .

Next we can examine the possible pairwise fusions of the distributions denoted by a distribution  $R(i,j)$ :

$$R(1,2) = (.45, .25, .2, .1); ||R(1,2)||^2 = .316$$

$$R(1,3) = (.3, .2, .15, .35); ||R(1,3)||^2 = .275$$

$$R(2,3) = (.25, .25, .15, .35); ||R(2,3)||^2 = .27$$

So the distribution  $R(1,2)$  fusing the distributions with the least conflict provides the most information. The range for NegEnt values of fusions depends on the characteristics of the particular distributions. If we would fuse two identical distributions, the fusion information content would be the same as the original distribution. So a better comparison for fused distributions is the ratio of the NegEnt to the average of the NegEnt of the original distributions. If we do this we have

$$||R(1,2)||^2 = .316 / .32 = .988$$

$$||R(1,3)||^2 = .275 / .38 = .724$$

$$||R(2,3)||^2 = .27 / .36 = .75$$

This scaling reflects better the comparisons between the fused distributions. Also it reflects more appropriately the value for  $R(2,3)$  which has less conflict than  $R(1,3)$ .

Finally we can fuse all three distributions,  $R(1,2,3)$

$$R(1, 2, 3) = (.333, .233, .166, .266); ||R(1, 2, 3)||^2 = .264$$

and the ratio is

$$||R(1, 2, 3)||^2 = .264 / .353 = .748$$

So the ratio indicates that fusion of all three is roughly the same as  $R(2,3)$  but improved over  $R(1,3)$ .

The distributions and their information content provides an initial basis for decisions on selections of search areas but in the next Section, 6, we discuss how to use credibility along with NegEnt for a final fusion. So in that section we will revisit the above example using credibility.

## 6. On weighted average fusion

Let us now look further at the multi-source fusion problem. Assume  $V$  is a variable taking its value in the space  $X = \{x_1, \dots, x_n\}$  and we have a collection  $Z = \{P_1, \dots, P_t\}$  of probability distribution type information about the value of  $V$ . In addition we assume a credibility function  $Cred$  is providing information about the credibility of fused values using different subsets of  $Z$ .

Given a subset  $B$  of probability distributions from  $Z$  we can calculate the associated fused value,  $P_B$ . In particular if  $|B|$  is the number of distributions in  $B$  then  $P_B = \frac{1}{|B|} \sum_{P_i \in B} P_i$ . Our  $Cred$  measure now provides the credibility associated with the fusion based on the subset  $B$ ,  $Cred(B)$ . Here  $Cred(B) \in [0, 1]$ . In addition we can calculate the NegEnt value of the information associated with the

fused value  $P_B$

$$\|P_B\|^2 = \frac{1}{|B|^2} \left[ \sum_{P_i \in B} \|P_i\|^2 + 2 \sum_{i=1}^{t-1} \sum_{\substack{k=i+1 \\ P_k \in B}}^t P_i P_k \right]$$

Here also  $\|P_B\|^2 \in [0, 1]$ .

Our objective now is to obtain a fused value  $P_B$  that has both high values for  $\text{Cred}(B)$  and  $\|P_B\|^2$ . Since the number of possible subsets of  $Z$  is not prohibitive we can do an exhaustive search to find the best fusion. We first calculate for each subset  $B$  of  $Z$  its fused value  $P_B$  and its associated credibility  $\text{Cred}(B)$  and  $\text{NegEnt}(B)$ ,  $\|P_B\|^2$ . We must now compare the  $P_B$  based on their  $\text{Cred}(B)$  and  $\|P_B\|^2$ . Here we are here faced with a multi-criteria decision problem.

We now introduce the idea of dominance. We say that a fusion based on subset  $B_1$  dominates  $B_2$  if  $\text{Cred}(B_1) \geq \text{Cred}(B_2)$  and  $\text{NegEnt}(B_1) \geq \text{NegEnt}(B_2)$  and at least one of these is a strictly greater then. We now remove all subsets that are dominated by some other subset. Our preferred fusion will be one of the non-dominated fusions. We point out here the collection of non-dominate fusions subsets have at least have one member with credibility equal one. This is true because the credibility of the fusion based on the whole set  $Z$  has credibility one and any subset  $B$  that dominates  $Z$  must have credibility one.

At this point we have a collection of non-dominated fusions where each fusion is determined from a subset of space  $Z$ , we denote these as  $B_1, \dots, B_r$ . For each subset  $B_j$  we have its associated fusion  $P_{B_j}$  and its  $\text{NegEnt}$  value,  $\|P_{B_j}\|^2$  and its credibility,  $\text{Cred}(B_j)$ . Our objective is to use the information about the  $\text{NegEnt}$  value and credibility to select among these possible fusions, the  $P_{B_j}$ .

If there is only one non-dominated fusion then this is our selected fusion. If there is more than one non-dominated fusion our procedure for adjudicating between these must involve the introduction of some subjective preference type information from the responsible decision-maker. Here rather than dictate a best way, we shall suggest some ways a decision-maker can to choose among the non-dominated fusions. We note other possibilities exist.

Here we consider how to choose the final distribution based both on the information measure and credibility in our search and rescue example. We will use two of the possible credibility functions from the previous section,  $C_1$  based on the number of distributions in a subset, and  $C_2$  based on the idea of "most" distributions included in a subset.

The collection of relevant probability distributions is  $Z = \{P_1, P_2, P_3\}$ . So there are seven subsets of  $Z$  to consider:  $B_1 = \{P_1\}$ ,  $B_2 = \{P_2\}$ ,  $B_3 = \{P_3\}$ ,  $B_4 = \{P_1, P_2\}$ ,  $B_5 = \{P_1, P_3\}$ ,  $B_6 = \{P_2, P_3\}$ ,  $B_7 = \{P_1, P_2, P_3\}$ .

For the first credibility function,  $C_1$ , we use a threshold of at least two distributions included in a subset so we have

$$C_1(B_1) = C_1(B_2) = C_1(B_3) = 0 \text{ and } C_1(B_4) = C_1(B_5) = C_1(B_6) = C_1(B_7) = 1$$

Then for the second we use for the idea of "most" as the fuzzy value  $M$

$$M = \begin{cases} 0 & |B| = 1 \\ .7 & |B| = 2 \\ 1.0 & |B| = 3 \end{cases}$$

and so using  $M$  for the second credibility we have

$$C_2(B_1) = C_2(B_2) = C_2(B_3) = 0; C_2(B_4) = C_2(B_5) = C_2(B_6) = 0.7; C_2(B_7) = 1$$

Now to make our selection of which distribution to use we need to use the concept of dominance related to the credibility

and information. We can express this again as the predicate  $\text{Dom}(B_i, B_j)$ :

$$\text{Dom}(B_i, B_j) = [(\text{Cred}(B_i) \geq \text{Cred}(B_j)) \wedge (\|B_i\|^2 \geq \|B_j\|^2) \wedge \text{Cond}(>)]$$

where  $\text{Cond}(>)$  is true only if at least one of the " $\geq$ " is " $>$ ".

If  $\text{Dom}(B_i, B_j)$  is true then  $B_i$  dominates  $B_j$  and so  $B_j$  can be removed from consideration.

In order to help working through the possible dominance combinations for our example we provide the summarizing table:

Subset	NegEnt	C1	C2
B1	.34	0	0
B2	.3	0	0
B3	.42	0	0
B4	.316	1	0.7
B5	.275	1	0.7
B6	.27	1	0.7
B7	.264	1	1.0

We can assess  $\text{Dom}$  for each of the two credibility functions above to obtain the collection,  $\text{ND}$ , of non-dominated subsets. For the first credibility function we see that  $B_3$  dominates  $B_1$  and  $B_2$  as all three have  $C_1 = 0$  and  $B_3$  has a strictly greater  $\text{NegEnt}$ . Then  $B_1, B_2 \notin \text{ND}$ . Likewise  $B_4$  dominates  $B_5, B_6$  and  $B_7$  so  $B_5, B_6, B_7 \notin \text{ND}$ . Finally we have to consider the dominance relationship of  $B_3$  and  $B_4$ . Since  $(C_1(B_3) = 0) < C_1(B_4) = 1$ ,  $B_3$  cannot dominate  $B_4$ . Also  $\|B_4\|^2 < \|B_3\|^2$ , so  $B_4$  cannot dominate  $B_3$  and we have two non-dominated subsets

$$\text{ND} = \{B_3, B_4\}$$

For this case of more than one non-dominating fusion we must utilize some subjective considerations to select the most appropriate one. A possible consideration is that the distribution  $P_3$  obtained from the local officials might be thought of as more reliable than the fusion,  $B_4$ , of the less effective sensor obtained probabilities  $P_1$  and  $P_2$ .

Finally we examine dominance relations for the second credibility function  $C_2$ . As for  $C_1$ ,  $B_3$  dominates  $B_1$  and  $B_2$  and  $B_4$  dominates  $B_5$  and  $B_6$  but not  $B_7$  since  $[C_2(B_4) = .7] < [C_2(B_7) = 1]$ . So again the dominated subsets are eliminated:  $B_1, B_2, B_4, B_5 \notin \text{ND}$ . Now  $B_3$  cannot dominate  $B_4$  or  $B_7$  due to credibility since  $[C_2(B_3) = 0] < [(C_2(B_4) = .7) \text{ and } (C_2(B_7) = 1.0)]$ . Also  $B_4$  cannot dominate  $B_3$  since  $\|B_4\|^2 < \|B_3\|^2$  and  $B_7$  cannot dominate either  $B_3$  or  $B_4$  since its  $\text{NegEnt}$  is less than either one. So for the credibility  $C_2$

$$\text{ND} = \{B_3, B_4, B_7\}$$

and again we must discuss how select one of these. The considerations discussed for  $C_1$  can apply again to  $B_3$  and  $B_4$  but now  $B_7$  has to be considered also. A criteria here might dictate that since  $B_7$  has the highest credibility due to being the fusion of all three probabilities it should be chosen.

One observation we would like to make is that there appears to be some asymmetry between the two criteria, credibility and information content. In particular with regard to information content it would appear that we would like to obtain as much of this as possible, while with regard to credibility it may be that we want to have at least a certain level of credibility. Thus if we do not have some minimal degree of credibility any amount of information may not compensate. There appears a kind of priority here [17-19].

One approach here is for the responsible decision maker to provide some minimal level of credibility,  $\lambda$  and the select the fusion with the maximal  $\text{NegEnt}$  value having at least this level of credibility. In anticipation of generalizing this idea we could look at this approach a little more formally. Let  $S: [0, 1] \rightarrow [0, 1]$  be a function such that for any degree of credibility  $y$ ,  $S(y)$  indicates the decision makers degree of satisfaction with this level  $y$  of credibility. Here



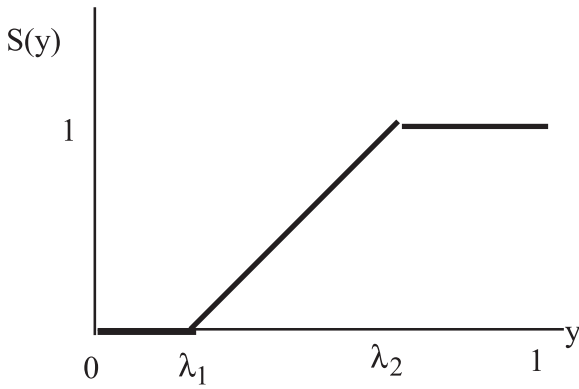


Fig. 3. Degree of satisfaction for credibility values.

$S$  should be monotonic and have  $S(1)=1$  and  $S(0)=0$ . Using this function we can associate with the fusion based on a subset  $B$  of  $Z$  a value,  $Qual(B)=S(Cred(B)) \ ||P_B||^2$ . Here then we select as our fusion value the  $P_B$  so that its  $Qual(B)$  is maximal.

If we consider the function  $S$  so that  $S(y)=0$  for  $y < \lambda$  and  $S(y)=1$  for  $y \geq \lambda$  we get the previous method for selecting the preferred fused value. In this case

$$Qual(B) = ||P_B||^2 \text{ for } Cred(B) \geq \lambda$$

$$Qual(B) = 0 \text{ for } Cred(B) < \lambda$$

Once having introduced this type of  $S$  function we can consider other formulas for  $S$ . One formula is shown in Fig. 3. Here if  $\lambda_1 = \lambda_2 = \lambda$  then we get the previous case. If  $\lambda_1 = 0$  and  $\lambda_2 = 1$  then we get  $Qual(B) = Cred(B) \ ||P_B||^2$ , the product of credibility and information content.

It is also possible to use Zadeh's idea of computing with words [20] to determine the form of  $S$ . Here we can linguistically specify some requirements for our requested degree of satisfaction by the credibility. We can represent this as a fuzzy set of the unit interval and then represent  $S$  as this fuzzy set.

Another approach for obtaining the quality of a fusion is to use the idea of prioritized aggregation introduced by Yager [19]. Here the score associated with a fusion is a weighted sum of its credibility and NegEnt, however here the weight associated with the criteria having the lower priority depends on the degree of satisfaction of higher order criteria. As a result of this relationship, lack of satisfaction to the higher priority criteria cannot be compensated for by very high satisfaction to the lower priority criteria. More formally the score of the fusion based on the subset  $B$  is determined as  $Score(B) = w_1 Cred(B) + w_2 ||P_B||^2$ . However here  $\frac{w_2}{w_1} = \frac{Cred(B)}{1}$ . Normalizing the weights so that  $w_1 + w_2 = 1$  we get

$$\begin{aligned} Score(B) &= \frac{1}{1 + Cred(B)} Cred(B) + \frac{Cred(B)}{1 + Cred(B)} ||P_B||^2 \\ &= \frac{Cred(B)}{1 + Cred(B)} (1 + ||P_B||^2) \end{aligned}$$

Here we see that if

Cred(B)	Score(B)
0	0
1/2	1/3(1 +   PB   <sup>2</sup> )
1	1/2(1 +   PB   <sup>2</sup> )

More generally if  $Cred(B) = n/d$  then  $Score(B) = \frac{n}{d+n} (1 + ||P_B||^2)$ . Thus if  $Cred(B) = 1/4$  then  $Score(B) = \frac{1}{5} (1 + ||P_B||^2)$  and if  $Cred(B) = 3/4$  then  $Score(B) = \frac{3}{7} (1 + ||P_B||^2)$ . If the credibility is 90% then  $\frac{n}{d} = \frac{9}{10}$  and  $Score(B) = \frac{9}{19} (1 + ||P_B||^2)$ .

### 7. Unequally weighted fusions

In the preceding we assumed that starting with the set  $Z = \{P_1, \dots, P_t\}$  we selected a subset  $B$  of  $Z$  and found our fusion by taking an equally weighted aggregation of the probability distributions in  $B$ . Here we suggest a more general approach based on a non-uniform weighted average of the elements in  $Z$ . So here we let  $W = [w_1, \dots, w_t]$  be a set of weights where  $w_i \in [0, 1]$  and  $\sum_{i=1}^t w_i = 1$  and we obtain a fusion  $P_W = \sum_{i=1}^t w_i P_i$ . Here  $P_W$  is a probability distribution whose  $j$ th component  $P_W(j) = \sum_{i=1}^t w_i P_{ij}$ .

We note here that the type of exhaustive search through all possible fusions to find the best fusion is not feasible here as there are too many possibilities. The preceding fusions obtained by using a simple average of the probability distributions in the subset  $B$  is a special case of using  $W$  obtained by assigning the weights as  $w_i = \frac{1}{|B|}$  for  $P_i \in B$  and  $w_i = 0$  for  $P_i \notin B$ .

In this new situation we see that

$$||PW||^2 = \sum_{j=1}^n \left( \sum_{i=1}^t w_i P_{ij} \right)^2 = \sum_{i=1}^t w_i^2 ||P_i||^2 + 2 \sum_{i=1}^{t-1} \sum_{k=i+1}^t w_i w_k P_i P_k$$

It can be shown that  $\sum_{i=1}^t w_i^2 + 2 \sum_{i=1}^{t-1} \sum_{k=i+1}^t w_i w_k = 1$  and thus  $||P_W||^2$  is a weighted average of the complements of the Gini entropies and the Gini cross entropies.

Thus  $||P_W||^2$  provides an indication of the information in a fusion based on using the weighted function  $W = [w_1, \dots, w_t]$ . Here  $||P_W||^2 \in [0, 1]$  and the bigger the value the more information, less uncertainty.

A second aspect in determining the quality of the fusion obtained by using the weighted value  $W = [w_1, \dots, w_t]$  is the credibility of this fusion. In order to determine the credibility associated with the weight vector  $W$ ,  $Cred(W)$ , our point of departure will be the given credibility measure  $Cred: 2^Z \rightarrow [0, 1]$  which associates with each subset  $B \subseteq Z$  a value  $Cred(B) \in [0, 1]$ . Using some ideas provided by Wang and Klir [21] about extending set measures we can obtain from this a credibility function,  $\widetilde{Cred}(W)$ , which associates with each vector  $W$  a value in the unit interval indicating a credibility of a fusion using a weighting of  $P_i$  based on the weights in  $W$ . In order to accomplish this we take the Choquet integral of a function  $f$  on  $Z$  with respect to the measure  $Cred$  on  $Z$  [22–24]. Here  $f$  is defined as  $f(P_i) = \frac{w_i}{\text{Max}_i[w_j]}$ . Thus  $f(P_i)$  is the weight assigned to  $P_i$  divided by the Max weight in  $W$ . Parenthetically,  $f$  can be viewed as the membership function of a fuzzy subset of  $Z$ .

We obtain the Choquet integral of  $f$  with respect to the measure  $Cred$  as follows [21]. Let  $\text{ind}$  be an index function so that  $\text{ind}(j)$  is the index of element in  $Z$  with the  $j$ th largest value for  $f$ , it is essentially the element in  $Z$  with the  $j$ th largest value for  $w_i$ . Here then  $f(P_{\text{ind}(j)}) = \frac{w_{\text{ind}(j)}}{\text{Max}_i[w_j]}$  is the  $j$ th largest value for  $f(P_i)$ . We now let  $H_j = \{P_{\text{ind}(k)} \mid k=1 \text{ to } j\}$ , it is the subset of  $Z$  with the  $j$  largest values of  $f$ . We note it is also the subset of  $Z$  with the  $j$  largest weights. Using this we obtain via the Choquet integral

$$\begin{aligned} \widetilde{Cred}(W) &= \sum_{j=1}^t (Cred(H_j) - Cred(H_{j-1})) f(P_{\text{ind}(j)}) \\ \widetilde{Cred}(W) &= \sum_{j=1}^t (Cred(H_j) - Cred(H_{j-1})) \frac{w_{\text{ind}(j)}}{\text{Max}_i[w_j]} \\ \widetilde{Cred}(W) &= \frac{1}{\text{Max}_i[w_j]} \sum_{j=1}^t (Cred(H_j) - Cred(H_{j-1})) w_{\text{ind}(j)} \end{aligned}$$

Consider now our original situation in which we used a subset  $B$  of  $Z$  with uniformly weighted components to get our fused value. Here the Credibility of this is  $Cred(B)$ . Let us show that we get this same value in the more general framework of using of vector  $W$ . In this case we have as previously noted that  $W$  is such

that  $w_i = \frac{1}{|B|}$  for  $P_i \in B$  and  $w_i=0$  for  $P_i \notin B$ . In this case we get  $\text{Max}[W]=\text{Max}_i|w_i|=\frac{1}{|B|}$ . Here we see that  $\widetilde{\text{Cred}}(W)$  is equal to

$$\frac{1}{\text{Max}[W]} \left[ \sum_{j=1}^{|B|} (\text{Cred}(H_j) - \text{Cred}(H_{j-1})) w_{\text{ind}(j)} + \sum_{j=|B|+1}^t (\text{Cred}(H_j) - \text{Cred}(H_{j-1})) w_{\text{ind}(j)} \right]$$

Since  $w_{\text{ind}(j)} = \frac{1}{|B|}$  for  $j=1$  to  $|B|$  and  $w_{\text{ind}(j)}=0$  for  $j=|B|+1$  to  $t$  we get

$$\widetilde{\text{Cred}}(W) = \frac{1}{\text{Max}[W]} \left[ \sum_{j=1}^{|B|} (\text{Cred}(H_j) - \text{Cred}(H_{j-1})) \frac{1}{|B|} \right]$$

with  $\text{Max}[W]=\frac{1}{|B|}$  we see that  $\widetilde{\text{Cred}}(W)=\sum_{j=1}^{|B|} (\text{Cred}(H_j) - \text{Cred}(H_{j-1}))$ . Further we see that

$$\begin{aligned} \widetilde{\text{Cred}}(W) &= \sum_{j=1}^{|B|} (\text{Cred}(H_j) - \text{Cred}(H_{j-1})) = \text{Cred}(H_{|B|}) \\ &\quad - \text{Cred}(\emptyset) = \text{Cred}(H_{|B|}). \end{aligned}$$

Since  $H_{|B|}$  is the set of elements in  $Z$  with that  $|B|$  largest weights,  $H_{|B|}=B$  we get as desired that  $\widetilde{\text{Cred}}(W)=\text{Cred}(B)$  in this case.

We can express  $\widetilde{\text{Cred}}(W)=\text{ChoqCred}(f_W)$  where  $f_W$  is the function  $f_W(P_i)=\frac{w_i}{\text{Max}[W]}$ . We shall now look at little more carefully at this  $\widetilde{\text{Cred}}$  function.

First let us make the following notation comment. If  $B$  is any non-empty subset of  $Z$  then we shall denote  $W_B$  as its weighting vector with all  $w_j = \frac{1}{|B|}$  for  $P_j \in B$  and  $w_j=0$  for all  $P_j \notin B$ . We note here that  $f(P_j)=1$  for all  $P_j \in B$  and  $f(P_j)=0$  for  $P_j \notin B$ .

**Theorem.** Assume  $W_B$  is any weighting vector that has weights of zero for any element not in  $B$ , then  $\widetilde{\text{Cred}}(W_B) \geq \text{Cred}(W_B)$ .

**Proof.** Let  $f_B$  and  $f_B$  be the associated function for each of these weighted vectors. Here we have

$$\begin{aligned} f_B(P_j) &= 1 \quad \text{for } P_j \in B \\ f_B(P_j) &= 0 \quad \text{for } P_j \notin B \end{aligned}$$

and

$$\begin{aligned} f^B(P_j) &\leq 1 \quad \text{for } P_j \in B \\ f^B(P_j) &= 0 \quad \text{for } P_j \notin B \end{aligned}$$

From this we see that  $f_B(P_j) \geq f^B(P_j)$  for all  $P_j$ . From the monotonicity of the Choquet integral [22] we get  $\widetilde{\text{Cred}}(W_B)=\text{ChoqCred}(f_B) \geq \text{ChoqCred}(f^B) \geq \text{Cred}(W^B)$ .

More generally we see the following

**Theorem.** Let  $D \subseteq B$  be two subsets of  $Z$ . Let  $W^D$  be a weighting vector that has zero weights for any element not in  $D$ . Then  $\widetilde{\text{Cred}}(W_B) \geq \widetilde{\text{Cred}}(W^D)$

**Proof.** Here again we have  $f_B(P_j)=1$  for  $P_j \in B$  and  $f_B(P_j)=0$  for  $P_j \notin B$  while  $f^D(P_j) \leq 1$  for  $P_j \in D$  and  $f^D(P_j)=0$  for  $P_j \notin D, P_j \in \bar{D}$  where  $\bar{D} \cap B \neq \emptyset$ .

From this we get  $f_B(P_j) \geq f^D(P_j)$  for all  $P_j$ . The result again follows from the monotonicity

A corollary of the preceding is that if  $D \subseteq B$  then  $\widetilde{\text{Cred}}(W_B) \geq \widetilde{\text{Cred}}(W^D)$ . This is of course the same as the monotonicity of the measure  $\text{Cred}, \text{Cred}(B) \geq \text{Cred}(D)$  if  $D \subseteq B$ .

If  $W$  is a weighting vector such that the components in  $B$  are non-zero while those that are not in  $B$  are zero we say that  $W$  is based on  $B$ .

So we have shown for any weight based on  $B$  the most credibility occurs if we uniformly assign the weights to the elements in  $B$ .

More generally from the monotonicity of the Choquet integral we see if  $W=[w_1, \dots, w_t]$  and  $\hat{W}=[\hat{w}_1, \dots, \hat{w}_t]$  are two sets of weighting vectors so that  $\frac{w_j}{\text{Max}[W]} \geq \frac{\hat{w}_j}{\text{Max}[\hat{W}]}$  for all  $j$  then  $\widetilde{\text{Cred}}(W) \geq \widetilde{\text{Cred}}(\hat{W})$ .

We note that while there exists some type of relationship between the choice of weighting vector  $W$  and  $\widetilde{\text{Cred}}(W)$ , the relationship between  $\|P_W\|^2$  and the choice of the weights is more complex since  $\|P_W\|^2 = \sum_{j=1}^t w_j^2 \|P_j\|^2 + 2 \sum_{i=1}^{t-1} \sum_{k=i+1}^t w_i w_k P_i P_k$ .

An interesting characterizing feature we can associate with each  $P_i$  is  $R_i = \frac{1}{t-1} \sum_{\substack{k=1 \\ k \neq i}}^t P_i P_k$ . We see  $R_i$  is the average cross-information.

Clearly the bigger this value the more desirable the probability distribution  $P_i$  is for including in the weighted fusion. Similarly, the larger  $\|P_i\|^2$  the more desirable the  $P_i$  is for including in the weighted fusion.

Here then given a set of probability distributions,  $Z=\{P_1, \dots, P_t\}$  and a credibility measure  $\text{Cred}$  on  $Z$ , the question we are faced with is the problem of how to select the appropriate weights for the fusion given our interest in obtaining a fused value that is both informative and credible. While we have provided formulations for calculating the information,  $\|P_W\|^2$  and credibility,  $\widetilde{\text{Cred}}(W)$  given a weighting vector  $W$ , the process of obtaining the optimal weighted vector  $W$  is a difficult multi-criteria optimization problem. Given the apparent difficulty of solving this optimization problem we shall look for some satisficing solution to this problem.

Here again we shall assume some minimal required level of credibility  $\alpha$  and look for solutions with a large value of  $\|P_W\|^2$  that have at least this minimal level of credibility. Let us look at the formulation for  $\|P_W\|^2$

$$\|P_W\|^2 = \sum_{i=1}^t w_i^2 P_i^2 + 2 \sum_{i=1}^{t-1} \sum_{k=i+1}^t w_i w_k P_i P_k$$

We note that terms  $P_i P_k$  and  $P_i^2$  can be calculated off line and are independent of choice of  $W$ .

Let us denote these as follows  $P_i P_k = m_{ik}$  and  $P_i^2 = m_{ii}$ . Thus here then

$$\|P_W\|^2 = \sum_{i=1}^t w_i^2 m_{ii} + 2 \sum_{i=1}^{t-1} \sum_{k=i+1}^t w_i w_j m_{ik}$$

and we have the constraints that  $\sum_{i=1}^t w_i = 1$  and all  $w_i \in [0, 1]$ . Here our objective is to try to get  $\|P_W\|^2$  as large as possible. In addition  $\widetilde{\text{Cred}}(W)=\sum_{j=1}^t (\text{Cred}(H_j) - \text{Cred}(H_{j-1})) \cdot \frac{w_{\text{ind}(j)}}{\text{Max}(W)}$  which we require to have a value of at least  $\alpha$ .

A useful characterizing feature we can associate with the basic  $\text{Cred}$  measure on the space  $Z=\{P_1, \dots, P_2\}$  is its Shapely index [25-27]. For any  $P_j \in Z$  we define its Shapely index  $S_j$  as

$$S_j = \sum_{k=0}^{t-1} \left( \gamma_k \sum_{\substack{K \subset F_j \\ |K|=k}} (\text{Cred}(K \cup \{P_j\}) - \text{Cred}(K)) \right)$$

In the above  $K$  is a subset of cardinality  $|k|, F_j = Z - \{P_j\}$  and  $\gamma_k = \frac{(n-k-1)!k!}{n!}$ . It can be shown that  $S_j \in [0, 1]$  and  $\sum_{j=1}^t S_j = 1$  [27]. This index can be seen as the average increase in ‘‘credibility’’ obtained by adding the element  $P_j$  to a set that does not contain it. We note that it can be shown that if  $\text{Cred}$  is a simple additive measure with  $\text{Cred}(\{P_j\})=\alpha_j$  then  $S_j = \alpha_j$  and if  $\text{Cred}$  is a cardinality based measure, then  $S_j = 1/t$  for all  $j$  [27].

Using these Shapely index values we can obtain an approximation to the credibility of a subset associated with a weighting vector  $W=[w_1, \dots, w_t]$  in particular

$$\widetilde{\text{Cred}}(W) = \sum_{j=1}^t S_j \frac{w_j}{\text{Max}_i[w_i]}$$

We recall that for the case where  $W$  is related to a crisp subset  $B$  of  $X$ ,  $w_j=1/|B|$  for all  $P_j \in B$  and  $w_j=0$  for all  $P_j \notin B$ . Here then  $\text{Max}_i[w_i]=1/|B|$  and we have  $\frac{w_j}{\text{Max}_i[w_i]}=1$  for  $P_j \in B$  and  $\frac{w_j}{\text{Max}_i[w_i]}=0$  for  $P_j \notin B$ . From this we get  $\widetilde{\text{Cred}}(B)=\sum_{x_j \in B} S_j = \text{Cred}(B)$ . This is  $\text{Cred}(B)$  for the special case of uniform weights.

Using this definition for  $\widetilde{\text{Cred}}(W)$  we now formulate our problem of finding the most informative weighted fusion of the elements of  $Z$  given that we want a minimal credibility of  $\alpha$ . In particular the problem becomes: Find  $w_1, \dots, w_t$  to

$$\text{Maximize : } \|P_W\|^2 = \sum_{i=1}^t w_i^2 m_{ii} + 2 \sum_{i=1}^{t-1} \sum_{k=i+1}^t w_i w_k m_{ik}$$

such that

- (1)  $\frac{1}{\text{Max}_i[w_i]} \sum_{i=1}^t w_i S_i, N \geq \alpha$
- (2)  $w_i \in [0, 1]$
- (3)  $\sum_{i=1}^t w_i = 1$

This is a nonlinear mathematical programming problem.

## 8. Conclusion

Our objective here was to obtain quality-fused values about the value of a variable from information provided by multiple sources in the form of probabilistic distributions. Here quality was measured by a lack uncertainty in the fused value, more informative fused values and the use of credible sources. We introduced a vector representation for a probability distribution and using the Gini formulation for entropy we showed how the norm of the vector provides a measure of the certainty, information, associated with a probability distribution. We looked at special cases of fusion for source inputs that were maximally uncertain and certain. We provided a measure of credibility associated with subsets of sources. We looked at the issue of finding the highest quality fused value from the weighted aggregations of source provided probability distributions.

## Acknowledgments

Ronald Yager was supported in part by an ONR grant award. Fred Petry would like to thank the Naval Research Labora-

tory's Base Program, Program Element No. 0602435N for their sponsorship.

## References

- [1] D.L. Hall, C.Y. Chong, J. Llinas, M. Liggins, *Distributed Data Fusion for Network-Centric Operations*, CRC Press, Boca Raton, 2012.
- [2] P. Elmore, F.E. Petry, R.R. Yager, Comparative measures of aggregated uncertainty representations, *J. Ambient Intell. Human. Comput.* 5 (2014) 809–819.
- [3] H. Bustince, P. Burillo, Correlation of interval-valued intuitionistic fuzzy sets, *Fuzzy Sets Syst.* 74 (1995) 237–244.
- [4] M.R. Spiegel, D. Lipschutz, D. Spellman, *Vector Analysis*, John Wiley, Hoboken, 2009.
- [5] R.R. Yager, N. Alajlan, An intelligent interactive approach to group aggregation of subjective probabilities Technical Report MII-3502, Machine Intelligence Institute, Iona College, New Rochelle, NY, 2015.
- [6] S. Kullback, *Information Theory and Statistics*, John Wiley and Sons, New York, 1959.
- [7] B. Buck, *Maximum Entropy in Action: A Collection of Expository Essays*, Oxford University Press, NY, 1991.
- [8] J. Aczel, Z. Daroczy, *On Measures of Information and their Characterizations*, Academic, New York, 1975.
- [9] R.R. Yager, K. Engemann, Entropy measures in sports, *Int. J. Syst. Meas. Decis.* 1 (1981) 67–72.
- [10] R.R. Yager, V. Kreinovich, Entropy conserving transforms and the entailment principle, *Fuzzy Sets Syst.* 158 (2007) 1397–1405.
- [11] Luenberger, *Information Science*, Princeton University Press, Princeton, NJ, 2006.
- [12] C. Gini, Variabilità e mutabilità, in: E. Pizetti, T. Salvemini (Eds.), *Reprinted in Memorie di Metodologica Statistica*, Libreria Eredi Virgilio Veschi, Rome, 1955.
- [13] P.-T. De Boer, D.P. Kroese, S. Mannor, R.Y. Rubinstein, A tutorial on the cross-entropy method, *Ann. Oper. Res.* 134 (2005) 19–67.
- [14] R.Y. Rubinstein, D.P. Kroese, *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation, and Machine Learning*, Springer-Verlag, New York, 2004.
- [15] L.A. Zadeh, A computational approach to fuzzy quantifiers in natural languages, *Comput. Math. Applic.* 9 (1983) 149–184.
- [16] T. Takagi, M. Sugeno, Fuzzy identification of systems and its application to modeling and control, *IEEE Trans. Syst. Man Cybernet.* 15 (1985) 116–132.
- [17] R.R. Yager, Modeling prioritized multi-criteria decision making, *IEEE Trans. Syst. Man Cybernet. Part B* 34 (2004) 2396–2404.
- [18] R.R. Yager, Prioritized aggregation operators, *Int. J. Approx. Reason.* 48 (2008) 263–274.
- [19] R.R. Yager, On prioritized multiple criteria aggregation, *IEEE Trans. Syst. Man Cybernet. Part B* 42 (2012) 1297–1305.
- [20] L.A. Zadeh, Fuzzy logic—computing with words, *IEEE Trans. Fuzzy Syst.* 4 (1996) 103–111.
- [21] Z. Wang, G.J. Klir, *Generalized Measure Theory*, Springer, New York, 2009.
- [22] G. Beliakov, A. Pradera, T. Calvo, *Aggregation Functions: A Guide for Practitioners*, Springer, Heidelberg, 2007.
- [23] E.P. Klement, R. Mesiar, E. Pap, A universal integral as common frame for Choquet and Sugeno, *IEEE Trans. Fuzzy Syst.* 18 (2010) 178–187.
- [24] Z. Wang, R. Yang, K.-S. Leung, *Nonlinear Integrals and Their Applications in Data Mining*, World Scientific, Singapore, 2010.
- [25] J.L. Marichal, M. Roubens, Entropy of discrete fuzzy measures, in: *Proceedings of Third International Workshop on Preferences and Decisions*, Trento, 2000, pp. 135–148.
- [26] T. Murofushi, A technique for reading fuzzy measures (i): the Shapely value with respect to a fuzzy measure, in: *Proceedings Second Fuzzy Workshop*, Nagaoka, Japan, 1992, pp. 39–48.
- [27] R.R. Yager, On the entropy of fuzzy measures, *IEEE Trans. Fuzzy Sets Syst.* 8 (2000) 453–461.