Combining Uncertain Information of Differing Modalities

Fred Petry^{*1}, Paul Elmore¹, Ronald Yager²

¹ Geospatial Science and Technology Branch, Bldg. 1005

Naval Research Laboratory, Stennis Space Center, MS 39529

²Machine Intelligence Institute, Iona College, New Rochelle, NY

*Corresponding author: fpetry@nrlssc.navy.mil: Ph: 228-688-4948

Abstract: In this paper we consider approaches for combining separately possibilistic uncertainty, probabilistic uncertainty and situations where both forms of uncertainty appear. An approach to probability aggregation using rational consensus with equi-weighting is developed. This aggregation is analyzed with information measures as one way to assess combinations and understand the impact on uncertainty. The analysis is based on combinations of bounding cases of probability distributions. Measures of conflict and the effect on information are developed. Next possibility transformations are used and illustrated by three representative possibility cases. The resultant transformed probabilities are aggregated with general probability distributions and the result evaluated with information measures as before. Finally a general approach to combining possibility distributions directly using quality criteria is described. An example is provided to illustrate the basic possibility distribution aggregation fusion developed.

Keywords: probability distribution; possibility distribution; Gini index, possibilistic conditioning; possibility-probability transformation; conflict measures

1. Introduction

Uncertainty arising from multiple sources and of many forms appears in the everyday activities and decisions of humans. We want to examine approaches that can be used to combine these uncertainties into forms that can become useful for decision making. Effective decision-making should be able to make use of all the available, relevant information about such combined uncertainty. In this paper we consider new approaches for combining separately possibilistic uncertainty, probabilistic uncertainty and situations where both forms of uncertainty appear.

To formalize the discussion, let V be a discrete variable taking values in a space X that has both aleatory and epistemic sources of uncertainty [26]. Let P be a probability distribution $P: X \rightarrow [0, 1]$ such that $p_k \in [0,1]$, $\sum_{k=1}^n p_k = 1$ that models the aleatory uncertainty. Then the epistemic uncertainty can be modeled by a

possibility distribution [47] such that $\Pi : X \rightarrow [0, 1]$, where $\pi(x_k)$ gives the possibility that x_k is the value of V , $k=1...n$. A usual requirement here is the normality condition, $\text{Max}_x [\pi(x)] = 1$, that is at least one element in X must be a fully possible. Abbreviating our notation so that $p_k = p(x_k)$, ... and $\pi_k = \pi(x_k)$, ..., we have $P = \{p_1, p_2, \dots, p_n\}$ and $\Pi = \{\pi_1, \pi_2, \dots, \pi_n\}$.

We have been motivated to address issues of uncertainty combinations by our consideration of the following application. The Navy's piracy report available to the public, PAWW (Piracy Analysis and Warning Weekly) [24] involves making use of both aleatory and epistemic information [12]. The problem of piracy attacks on shipping is well recognized off of the Horn of Africa (Somalia), but is also a problem in several other areas globally as well (Nigeria and Indonesia) [16, 24].

Figure 1. PAWW Piracy Attack Reports for Horn of Africa Area

The modeling for PAWW developed at the Naval Research Laboratory [12] provides guidance for decision making relative to deployment of limited interdiction resources to prevent such attacks. There are several sources of information input to the system. Here we provide a simplified example. Typically there exists previously collected information that can be formulated by probability distributions. For example the previously observed number of pirate attack groups, n , could be

$$P(n) = \{.4, .3, .15, .15\}, n = 1, 2, 3, 4$$

An important factor for a response decision often involves various sources of human intelligence about current activity in the area of concern. Here sources might provide intelligence about the possibility of the number of such groups active in a certain area and specific time frame. Analysts can then represent this information as a possibility distribution such as

$$\Pi(n) = \{1.0, .7, .6, .3\}, n = 1, 2, 3, 4$$

A decision maker should be able take into account all such information in an informed manner such as by the aggregation of the distributions and assessments of the resulting aggregation.

In combining information from multiple sources, a basic issue that must be considered is whether the resultant combination is somehow "better" than the

original information. Some factors such as source reliability, effect of disinformation [23] and conflicting data might have to be considered. The approach we will utilize in this paper is the use of information measures such as Shannon entropy to assess if an aggregated result has increased the information content.

In this paper we first present in Section 2 background and related work on probability and possibility measures and aggregation approaches. An approach to aggregation of probability distributions based on consensual averaging is then considered in Section 3. This approach is evaluated using Shannon and Gini information measures with a range of probability distributions. An analysis of conflict is also developed and information values related to conflict determined. In the following section we consider the combination of possibility and probability distributions. A possibility distribution is converted to a probability distribution using a common transformation [8,9] and the probability distributions are then combined using the averaging aggregation approach of Section 3. Again information measures are used to evaluate the approach and show its effectiveness. Finally the combination of possibility distributions is introduced. We first consider simple conjunctions of possibility distribution and establish some basic criteria for their combinations. Then we present a more nuanced approach that will allow us to introduce more developed criteria for selection of the best aggregation.

2. Background and Related Work

2.1 Probability Measures

We want to be able to assess the utility of aggregated uncertainty through consideration of information measures including Shannon entropy, Gini index, and Renyi entropy, to gauge the aggregated uncertainty. In this section we will review measures that can be used to evaluate an uncertainty aggregation. Shannon's entropy has been a commonly accepted standard for information metrics; however, the concept of information is so rich and broad that multiple approaches to the quantification of information are desirable [21,36]. Thus, we will also examine other measures, such as the Gini index and Renyi entropy, in this section.

2.1.1. Shannon Entropy

Shannon entropy has been the most broadly applied measure of randomness or information content [35]. For a probability distribution $P = \{p_1, p_2, \dots, p_n\}$ this is given as

$$S(P) = - \sum_{i=1}^n p_i \ln(p_i) \quad (1)$$

The well-known minimum and maximum values for the Shannon entropy will be presented in the context of extreme probability cases in Section 3.

2.1.2 Gini Index

The Gini index, $G(P)$, also known as the Gini coefficient, is a measure of statistical dispersion developed by Gini [14], and is defined as

$$G(P) = 1 - \sum_{i=1}^n p_i^2 \quad (2)$$

Some practitioners use $G(P)$ versus $S(P)$ since it does not involve a logarithm, making analytic solutions simpler. The Gini index has been used in consideration of inequalities in various areas such as economics, ecology and engineering [1]. A very important application of the Gini index is as a splitting criterion for decision tree induction in machine learning and data mining [2].

It is accepted in practice for diagnostic test selection that the Shannon and Gini measures are interchangeable [32]. The specific relationship of Shannon entropy and the Gini index has been discussed in the literature [11]. Theoretical support for this practice is provided in Yager's independent consideration of alternative measures of entropy [41], where he derives the same form for an entropy measure as the Gini measure.

2.1.3. Renyi Entropy

Renyi introduced a parameterized family of entropies as a generalization of Shannon entropy [28,29]. The intention was to have the most general approach that preserved the additivity property and satisfied the probability axioms of Kolmogorov. Renyi entropy is

$$S_\alpha(P) = \frac{1}{1-\alpha} * \ln \left(\sum_{i=1}^n p_i^\alpha \right) \quad (3)$$

Cases of the parameter α :

$\alpha=0$: $S(P) = \ln |P|$ - Hartley Entropy [17]

$\lim \alpha \rightarrow 1$: $S_1(P) = - \sum_{i=1}^n p_i \cdot \ln(p_i)$ -- Shannon Entropy

$\alpha=2$: $S_2(P) = - \ln \left(\sum_{i=1}^n p_i^2 \right)$ -- Collision or quadratic entropy

$\alpha \rightarrow \infty$: $S_\infty(P) = \min_{i=1}^n (\ln p_i) = - \max_{i=1}^n (\ln p_i)$

This last case is the smallest entropy in the Renyi family and so is the strongest way to obtain an information content measure. It is never larger than the Shannon entropy. Thus, the possible ranges of α capture the following:

High α : high probability events

Low α : weight possible events more equally

$\alpha=0, 1 \rightarrow$ Hartley or Shannon, respectively

2.2. Possibility Measures

2.2.1 Specificity

The concept of specificity in the framework of possibility theory has a function analogous to the entropy measures for probability we have discussed above. Yager [37] gave a formal definition of the specificity of a fuzzy subset which was extended for possibility distributions as

$$Sp(\Pi) = \int_0^{\alpha_{\max}} \frac{1}{\text{card}(\Pi_\alpha)} d\alpha$$

where $\alpha_{\max} = \max_x \Pi(x)$ and is the α -possibility level of P , the subset of elements having possibility of at least α . A linear specificity measure provides an intuitive measure of specificity of a possibility distribution [27]. Let $\pi_m = \max_{k=1}^n \pi_k$ and then the specificity measure is this max value minus the average of the other possibility values [40]:

$$Sp(\Pi) = \pi_m - \left(\sum_{k=1, k \neq m}^n \pi_k \right) / n \quad (4)$$

2.2.2 Consistency

There have been a number of approaches to consistency measures of probability and possibility distributions that have been proposed [7,15]. In particular the Zadeh [47] consistency measure is $C_Z(\Pi, P)$,

$$C_Z(\Pi, P) = \sum_{i=1}^n \pi_i * p_i \quad (5)$$

This measure does not represent an inherent relationship but rather represents the intuition that a lowering of an event's possibility tends to lower its probability, but not the converse.

Another consistency measure that appears in the literature, $C_{DP}(\Pi, P)$, is due to Dubois and Prade [8,9]. Here for every subset A of the space X,

$$C_{DP}(\Pi, P) = 1 \text{ if } \Pi(A) \geq P(A)$$

and is 0 otherwise. This definition is based on the idea that possibility is the weaker representation for a situation than probability. So the stronger support for a situation cannot have a greater value for the situation than the weaker support.

For our purposes, we focus here on C_Z as it provides a range of values to evaluate the idea of consistency as it relates to the aggregation approach.

2.3 Aggregation Approaches

Aggregation functions of the form we will consider should satisfy some basic properties in general [31,42]. Let $\text{Agg}(y_1, y_2, \dots, y_q)$ be an aggregation function. The first desired property is idempotency where if all y_i are equal, $\forall i y_i = Y$, then

$$\text{Agg}(y_1, y_2, \dots, y_q) = Y$$

A second property often desired is commutativity which assures that indexing of the aggregation arguments does effect the result. Finally monotonicity is the property where, if $\forall i y_i \geq z_i$, then

$$\text{Agg}(y_1, y_2, \dots, y_q) \geq \text{Agg}(z_1, z_2, \dots, z_q)$$

An aggregation function obeying these properties is a mean function [42] and prototypically is an averaging function. It has been noted [13] that the arithmetic averaging process for probabilities produces a result that has a very natural probabilistic meaning.

There are some important classes of aggregation [3,44]. Two are the conjunctive and disjunctive functions which correspond to "and" and "or" type

functions. For conjunctive operators t-norms are a common way to represent this. Likewise for the disjunctive t-conorms are frequently used. Lastly we can consider the class of mean operators. Specific examples here include the simple average as well as median, Max and Min. A frequently used generalized family of mean operators are the Ordered Weighted Average (OWA) operators introduced by Yager [39].

Related work on the aggregation specifically of probability distributions can be found in the considerable research on consensus for decision making [6]. In particular we refer to the simplified averaging consensual probability as developed in Lehrer and Wagner's rational consensus research [22]. The traditional average that weights all distributions equally is commonly used as the base line for comparing consensus techniques [4]. The linear opinion pool involves taking a weighted linear average of the opinions [5].

There is some research related to our specific concern of aggregating probabilistic and possibilistic uncertainty. First a related approach is the possibilistic conditioning of probability distributions using the approach of Yager [45]. This form of aggregation makes it very amenable to apply the information measures [10].

In possibilistic conditioning, a function f dependent on both P and Π is used to find a new conditioned probability distribution such that

$$f(P, \Pi) \Rightarrow \text{new } \hat{P}$$

where $\hat{P} = \{\hat{p}_1, \hat{p}_2, \dots, \hat{p}_n\}$ with

$$\hat{p}_i = \frac{p_i \pi_i}{K}; \quad K = \sum_{i=1}^n p_i \pi_i \quad (6)$$

A strength of conditioned probability is that it also captures Zadeh's concept of consistency between the possibility and the original probability distribution. Note that the K is of the form of the consistency from eq. 5: $C_Z(\Pi, P) = K$.

Also Hunter and Liu [18] describe an Dempster-Shafer approach [34] to merging heterogeneous uncertainty information, in particular for our interest, probability and possibility. The basis of their approach is to view a probability as special case of

a mass function and a conversion of a possibility distribution into a mass function. Then they use Dempster's combination rule on the converted mass functions and illustrate their approach to merging structured news reports.

3. Aggregation of Probability Distributions

In this section we develop an approach to combination of probability distributions and information measures of the combined probability.

For our purpose here we utilize the simplest approach of equi-weighting but plan on investigating future applications using variable weights.

We note that averaging would not be strictly associative if we were calculating a sequence of averages. However if we calculate the average a single time over a set of probabilities, this can be viewed as a quasi-associative operation [38] and no prohibitive issue arises with such an averaging process.

3.1 Consensus Combination of Probability Distributions

Let E be a set of evidence consisting of q probability distributions:

$$E = \{ P_1, P_2, \dots, P_q \}$$

where

$$P_1 = \{p_{11}, p_{12}, \dots, p_{1n}\}, P_2 = \{p_{21}, p_{22}, \dots, p_{2n}\}, \dots, P_q = \{p_{q1}, p_{q2}, \dots, p_{qn}\}$$

As discussed above the aggregation approach we will investigate is given by the linear average distribution denoted $\hat{P}$

$$\hat{P} = \{ \hat{p}_1, \hat{p}_2, \dots, \hat{p}_n \} \text{ where } w_i \text{ is the weight for distribution } P_i$$

$$\hat{p}_k = \sum_{i=1}^q w_i * p_{ik} ; \sum_{i=1}^q w_i = 1 \quad (7)$$

Clearly $\hat{P}$ is normalized as $\sum_{k=1}^n \hat{p}_k = 1$.

In order to calculate evaluations of combined probability distributions using information measures, we must consider some representative cases. Thus we will examine the information content using the two bounding cases of extreme distributions [10], completely certain, P_{CC}

$$P_{CC} = \{ 1, 0, \dots, 0 \}$$

and completely uncertain, P_{CU}

$$P_{CU} = \{ 1/n, 1/n, \dots, 1/n \}.$$

Our approach is to first develop the general expressions for the average $\hat{P}$ of a set of probabilities. Then we later introduce restrictions on the general case to allow the computation of information measures and their comparison.

We begin by partitioning the q probability distributions in $\mathbf{E}$ into two groups:

$$A1 = \{P_1, P_2, \dots, P_r\} ; A2 = \{P_{r+1}, P_{r+2}, \dots, P_q\}$$

Later we will see this facilitates the restriction of $A1$ to only P_{CC} distributions and $A2$ to P_{CU} distributions.

Proceeding from eq. (7), we sum separately over $A1$ and $A2$ respectively.

$$\hat{p}_k = \sum_{i=1}^r w_i * p_{ik} + \sum_{i=r+1}^q w_i * p_{ik} \quad (8)$$

Denoting these summations as:

$$S1_k = \sum_{i=1}^r w_i * p_{ik}; \quad S2_k = \sum_{i=r+1}^q w_i * p_{ik},$$

then

$$\hat{p}_k = S1_k + S2_k$$

3.11. Constraints on partitions of $\mathbf{E}$

We next constrain the partitions of $\mathbf{E}$ to permit evaluations by letting the distributions and weights in $A1$ and $A2$ to become uniform. That is

$$p_{1k} = p_{2k} = \dots p_{rk}; p_{r+1k} = \dots p_{qk}; \quad w_1 = w_2 = \dots w_r; w_{r+1} = \dots w_q$$

For notational simplicity we use p_{1k} , w_1 and p_{qk} , w_q to denote these probabilities and weights. Now the summations above simplify to

$$S1'_k = p_{1k} \sum_{i=1}^r w_1 = w_1 p_{1k} * r$$

$$S2'_k = p_{qk} \sum_{i=r+1}^q w_q = w_q p_{qk} * (q-r)$$

The aggregated probability becomes

$$\hat{p}_k = w_1 p_{1k} * r + w_q p_{qk} * (q-r) \quad (9)$$

The distribution is still normalized as

$$\sum_{k=1}^n \hat{p}_k = [r w_1 \sum_{k=1}^n p_{1k}] + [(q-r)w_q \sum_{k=1}^n p_{qk}] = r w_1 + (q-r)w_q = 1$$

since in the two partitions the distributions are normal and the weights sum to 1.

3.2 Information Measures of Aggregated Probability

In this section we will use the two information measures Shannon entropy and the Gini index reviewed in Section 2.1. These are applied to the aggregated probability distributions for the P_{CC} and P_{CU} to provide comparisons to more general aggregation cases.

3.2.1 Shannon Entropy

Shannon entropy has been the most broadly applied measure of randomness or information content [30,35]. However, the concept of information is so rich and broad that multiple approaches to the quantification of information are desirable [21] and we will also focus on the Gini measure [14].

For a probability distribution $P = \{p_1, p_2, \dots, p_n\}$, the Shannon entropy is

$$S(P) = - \sum_{k=1}^n p_k \ln(p_k).$$

If all q distributions are P_{CC} and their weights are uniform in eq. (7) then by the idempotency property, $\hat{P}_{CC} = (1, 0, 0, \dots)$ and

$$S(\hat{P}_{CC}) = - (1 \ln(1) + \sum_{k=2}^n 0 \ln(0)) = 1 * 0 + \sum_{k=2}^n 0 = 0$$

That is when the probability distribution represents complete certainty, there is no uncertainty, i.e. maximum information.

Next for the case of aggregated complete uncertainty represented by the equiprobable distribution, P_{CU} , we obtain as above by averaging identical distributions, $\hat{P}_{CU} = (1/n, 1/n, \dots, 1/n)$ and

$$S(\hat{P}_{CU}) = - \sum_{k=1}^n \frac{1}{n} \ln\left(\frac{1}{n}\right) = - \frac{1}{n} \sum_{k=1}^n (\ln(1) - \ln(n)) = - n * \frac{1}{n} (0 - \ln(n)) = \ln(n).$$

That is, when all probabilities are equiprobable, this is the most unpredictable, uncertain situation and so represents the minimum information. So the range of Shannon's entropy is:

$$0 \leq S(\hat{P}) \leq \ln(n)$$

3.2.2 Gini Index

The Gini index, $G(P)$, also known as the Gini coefficient, as described in Section 2.1.2 is

$$G(P) \equiv 1 - \sum_{k=1}^n p_k^2$$

Often $G(P)$ is used versus $S(P)$ since it does not involve a logarithm, making analytic solutions tractable.

Specifically for $\hat{P}_{cc}$

$$G(\hat{P}_{cc}) = 1 - (\hat{p}_1^2 + \sum_{k=2}^n \hat{p}_k^2) = 1 - (1^2 + 0) = 0 \quad (10)$$

As for the Shannon entropy this corresponds to no uncertainty.

Then for the equi-probable distributions ,

$$G(\hat{P}_{cu}) = 1 - \sum_{k=1}^n \left(\frac{1}{n}\right)^2 = 1 - n\left(\frac{1}{n^2}\right) = 1 - \frac{1}{n} = \frac{n-1}{n} \quad (11)$$

and in general the range for $G(\hat{P})$ is

$$0 \leq G(\hat{P}) \leq \frac{n-1}{n} < 1$$

3.2.3 Evaluations of Combined Probability Distributions

We examine the Gini index first for the combined probability distribution $\hat{P}$, eq. (9), from the previous section.

$$G(\hat{P}) = 1 - \sum_{k=1}^n \hat{p}_k^2 = 1 - \sum_{k=1}^n (w_1 p_{1k} * r + w_q p_{qk} * (q-r))^2$$

Now we make several restrictions to the general expression to allow effective evaluations of related information measures. Here we will use the extreme distributions in A1 and A2, but first making the simplifying assumption that weights are uniform across both A1 and A2, i.e., $w_i = 1/q$, we obtain

$$G(\hat{P}) = 1 - \sum_{k=1}^n \left(p_{1k} \frac{r}{q} + p_{qk} \frac{q-r}{q}\right)^2 = 1 - \left[\left(p_{11} \frac{r}{q} + p_{q1} \frac{q-r}{q}\right)^2 + \sum_{k=2}^n \left(p_{1k} \frac{r}{q} + p_{qk} \frac{q-r}{q}\right)^2\right]$$

Next if we let the probability distributions in A1 be P_{cc} s and in A2 P_{cu} s then

$$G(\hat{P}) = 1 - \left[\left(1 * \frac{r}{q} + \frac{1}{n} * \frac{q-r}{q}\right)^2 + n-1 \left(\frac{1}{n} * \frac{q-r}{q}\right)^2\right]$$

$$G(\hat{P}) = 1 - [(\frac{r}{q} + z)^2 + n-1(z)^2] = 1 - [\frac{r^2}{q^2} + \frac{2r}{q} z + n z^2], \text{ where } z = \frac{q-r}{nq}$$

Finally we let $|A1| = |A2| = q/2$, $r = q/2$. Then $z = \frac{1}{2n}$, $z^2 = \frac{1}{4n^2}$

$$G(\hat{P}) = 1 - [\frac{q^2}{4q^2} + \frac{2q}{2q} \frac{1}{2n} + n \frac{1}{4n^2}] = 1 - (\frac{1}{4} + \frac{3}{4n}) = \frac{3}{4} (1 - \frac{1}{n})$$

Now we will assess the combined probability, $\hat{P}$, to determine if it is more informative than the original distributions in A1 and A2. First we compare to the Gini measure for $\hat{P}$ and clearly as expected

$$G(\hat{P}_{CC}) = 0 < G(\hat{P})$$

since $\hat{P}_{CC}$ was a combination of certain distributions (eq. 10). But now comparing to $G(\hat{P}_{CU})$, eq. 11, we would expect that $\hat{P}$ is more informative

$$G(\hat{P}_{CU}) = 1 - \frac{1}{n} > \frac{3}{4} (1 - \frac{1}{n}) = G(\hat{P})$$

as we have a combination of certain and uncertain distributions. Furthermore if we change the balance of the combination distributions, how will the resultant information content of $\hat{P}$ be affected? We see that if there are more P_{CU} s in the combination we expect that the information will be less than the case above. So let $r = q/4$, i.e. 1/4 of the distributions are P_{CC} and 3/4 are P_{CU} . Then we have $z = \frac{3}{4n}$,

$$z^2 = \frac{9}{16n^2} \text{ and}$$

$$G(\hat{P}) = 1 - (\frac{1}{16} + \frac{3}{8n} + \frac{9}{16n}) = \frac{15}{16} (1 - \frac{1}{n}) > \frac{3}{4} (1 - \frac{1}{n})$$

3.2.4 Evaluation Using Shannon Entropy

Now we examine the Shannon entropy measure as for the Gini index by using the extreme distributions, $w_i = 1/q$, and $r = q/2$. Then from eq. (9)

$$\hat{p}_1 = w_1 p_{11} * r + w_q p_{q1} * (q-r) = \frac{n+1}{2n}$$

$$\hat{p}_k = w_q p_{qk} * (q-r) = \frac{1}{2n} \quad k = 2..n$$

Since

$$S(\hat{P}) = \hat{p}_1 \ln(\hat{p}_1) + \sum_{k=1}^n \hat{p}_k \ln(\hat{p}_k)$$

we have to evaluate the expressions

$$\begin{aligned} & \frac{n+1}{2n} \ln\left(\frac{n+1}{2n}\right) + \sum_{i=2}^n \frac{1}{2n} \ln\left(\frac{1}{2n}\right) \\ &= \frac{n+1}{2n} (\ln(n+1) - \ln(2n)) + (n-1) \frac{1}{2n} (\ln 1 - \ln(2n)) \\ &= \frac{n+1}{2n} \ln(n+1) - 2n/2n (\ln(2n)) = \frac{n+1}{2n} \ln(n+1) - \ln(2n) = \frac{n+1}{2n} \ln(n+1) - (\ln(n) + \ln(2)) \end{aligned}$$

We have the negative of this expression for the entropy

$$H(\hat{P}) = \ln(n) + \left[\ln(2) - \frac{n+1}{2n} \ln(n+1) \right]$$

So if $\left[\ln(2) - \frac{n+1}{2n} \ln(n+1) \right] < 0$, then $H(\hat{P}) < \ln(n)$. First we note that for $n = 2$ we

have $\left[\ln(2) - \frac{3}{4} \ln(3) \right] = [0.69 - 0.82] < 0$. Then this will be negative in general if

$\frac{n+1}{2n} \ln(n+1)$ increases with n . To see this we use the first two terms of Taylor

series expansion for the log. The result is that for n and $n+1$ we must compare

$$1 + \frac{1}{2} \left(\frac{n}{n+1} \right) \text{ and } 1 + \frac{1}{2} \left(\frac{n+1}{n+2} \right)$$

So for $\frac{n}{n+1}$ to be less than $\frac{n+1}{n+2}$, we have for $n = 2$, $\frac{2}{3} < \frac{3}{4}$. Now if we cross multiply

and use induction for any $n > 2$:

$$n(n+2) < (n+1)^2 = n^2 + 2n + 1 < n^2 + 2n + 1$$

Thus we have again that the combined probability has lower entropy than $\hat{P}_{cu}$

$$H(\hat{P}) < \ln(n) = H(\hat{P}_{cu})$$

3.2.5 Summary of Information Measures

We have seen that the relative ranges for the measures are comparable

$$0 \leq S(\hat{P}) \leq \ln(n); 0 \leq G(\hat{P}) \leq \frac{n-1}{n} < 1$$

Furthermore we can see that similar conclusions about the extreme probability cases can be drawn using either Shannon or Gini measures. However as mentioned, we will continue to use the Gini index as its polynomial form is somewhat more straightforward to evaluate than the logarithmic products of the Shannon entropy.

3.3 Conflict Measures and Aggregated Probability

In the previous sections we have assumed that none of the completely certain probabilities were in conflict. Here we will provide a conflict measure and assess how conflict affects the aggregated probability $\hat{P}$ and its information content $G(\hat{P})$. We anticipate, as one might expect, that information content decreases with increasing conflict.

We shall say there is conflict in two completely certain distributions, P_u and P_v if for some i and j there is a disagreement in their values:

$$p_{ui} = 1; p_{vi} = 0 \text{ and } p_{uj} = 0; p_{vj} = 1$$

If we were to average these two distributions we would obtain a distribution

$$\hat{P} = (0, 0, \dots, \hat{p}_i = 1/2, \dots, \hat{p}_j = 1/2, \dots, 0)$$

There is a lowered information content due to such a conflict:

$$G(\hat{P}) = 1 - (0 + \dots + (1/2)^2 + 0 + \dots + (1/2)^2 + \dots 0) = 1 - (1/2) = 1/2$$

Now we will consider an approach to developing a measure of conflict. In a given set of probability distributions we must consider the pairwise agreement of all the distributions. These can then be summed to provide a measure of conflict, denoted CF. First we partition the q distributions into the subsets of those distributions that are non-conflicting, i.e. identical. We denote each subset as T_i , where there are $m \leq q$ such subsets. Then we can assess the overall conflict by the pairwise product of the cardinality of these subsets. Thus we obtain for CF:

$$CF = \sum_{i=1}^{m-1} \sum_{j=i+1}^m |T_i| * |T_j| \quad (12)$$

Note that for no conflict, $|T_1| = q$, and for $j > 1$, $|T_j| = 0$ and so $CF = 0$.

We can examine an example of all possible cases relative to conflict for four completely certain probability distributions (P_{cc}). In Table 1 we see these five cases and for each case, the conflict measure CF and the Gini information measure $G(\hat{P})$.

As can be seen from Table 1, as expected, more conflict means less information in the combined probability. In the terminology used in Sentz and Ferson [33], cases II, III and IV constitute "arbitrary" evidence where no element is common to all subsets of distributions but some may have elements in common. Also they term case V as "disjoint" evidence as there is no intersection of the subsets.

Table 1. Conflicting Distributions Cases Example

	I	II	III	IV	V
P ₁	(1,0,0,0)	(1,0,0,0)	(1,0,0,0)	(1,0,0,0)	(1,0,0,0)
P ₂	(1,0,0,0)	(1,0,0,0)	(1,0,0,0)	(1,0,0,0)	(0,1,0,0)
P ₃	(1,0,0,0)	(1,0,0,0)	(0,1,0,0)	(0,1,0,0)	(0,0,1,0)
P ₄	(1,0,0,0)	(0,1,0,0)	(0,1,0,0)	(0,0,1,0)	(0,0,0,1)
CF	0	3	4	5	6
G($\hat{P}$)	0.000	0.375	0.500	0.625	0.750

3.3.1 Minimal and Maximal Conflict

Examining Table 1 further, we observe that there are two extremes of conflict represented. Case II represents minimal conflict and case V maximal conflict. A specific conflict measure CF can be developed for both extreme cases. For minimal conflict we must have for some u,v, $|T_u| = q-1$ and $|T_v| = 1$ and for other i, $|T_i| = 0$. Then the CF is

$$CF_{\min} = \sum_{i=1}^{m-1} \sum_{j=i+1}^m |T_i| * |T_j| = |T_u| * |T_v| = (q-1)*1 = q-1$$

Then for maximal conflict we have for all i, $|T_i| = 1$. So CF here is

$$CF_{\max} = \sum_{i=1}^{m-1} \sum_{j=i+1}^m 1 * 1 = (m-1) + (m-2) + \dots + (m-(m-1)) = (m(m-1))/2$$

We note since there is exactly one S_i for each q, $m=q$ and CF is $\frac{q}{2} (q-1)$.

Next we can obtain the average, $\hat{P}$, for the case of conflicting P_{cc} s by considering the T_i subsets. For a particular i there are exactly $|T_i|$ 1's to be averaged. So $\hat{p}_i = |T_i|/q$. Then we have for the Gini index:

$$G(\hat{P}) = 1 - \sum_{i=1}^n (|T_i|/q)^2 = 1 - \frac{1}{q^2} \sum_{i=1}^n |T_i|^2 \quad (13)$$

Using this expression we can observe that given some conflict, the more probability distributions there are without conflict, the more information is provided. First for minimal conflict using eq. (13) we have

$$G(\hat{P}) = 1 - \frac{1}{q^2} [(q-1)^2 + 1^2] = 1 - \frac{1}{q^2} (q^2 - 2q + 2) = \frac{2(q-1)}{q^2}$$

So for increasing q we see that $G(\hat{P}) \rightarrow 0$ indicating more information content. Now for the case of maximal conflict we obtain for the q $|T_i| = 1$

$$G(\hat{P}) = 1 - \frac{1}{q^2} \sum_{i=1}^n 1^2 = 1 - \frac{1}{q^2} \sum_{i=1}^q 1^2 = 1 - \frac{q}{q^2} = 1 - \frac{1}{q} = \frac{q-1}{q}$$

But for increasing q we see that $G(\hat{P}) \rightarrow 1$ tending rapidly to the least information.

3.3.2 Conflict and Aggregations with Completely Uncertain Probabilities.

Finally we compare to our result at the end of section 3.2.3, where we combined an equal number ($q/2$) of P_{cc} s and P_{cu} s. Here we carry out the same computation for the case of minimal conflict of the P_{cc} s. In order to simplify the computations we will let $n = q/2$ and let $|T_1| = q-1$ and $|T_2| = 1$. Then for the completely uncertain distributions $p_{ij} = 1/(q/2) = 2/q$ and we have:

$$\hat{p}_1 = \frac{1}{q} [(\frac{q}{2} - 1) * 1 + \frac{q}{2} * \frac{2}{q}] = \frac{1}{q} [\frac{q}{2} - 1 + 1] = \frac{q}{2} * \frac{1}{q} = \frac{1}{2}$$

$$\hat{p}_2 = \frac{1}{q} [1 * 1 + \frac{q}{2} * \frac{2}{q}] = \frac{2}{q}$$

$$\hat{p}_i = \frac{1}{q} [\frac{q}{2} * \frac{2}{q}] = \frac{1}{q}, \text{ for } i = 3 \dots q/2$$

So the Gini value is

$$G_{mincf}(\hat{P}) = 1 - [(\frac{1}{2})^2 + (\frac{2}{q})^2 + \sum_{i=3}^{q/2} (\frac{1}{q})^2] = 1 - [\frac{1}{4} + \frac{4}{q^2} + (\frac{q}{2} - 2) * \frac{1}{q^2}]$$

$$G_{\mincf}(\hat{P}) = 3/4 - \frac{1}{q} \left(\frac{q+4}{2q} \right)$$

Again by letting $n = q/2$ in the previous result of section 3.2.3, we have for comparison:

$$G(\hat{P}) = 3/4 - \frac{3}{q}$$

Then noting that for increasing q , $\frac{3}{q} \approx 6 * \frac{1}{q} \left(\frac{q+4}{2q} \right)$, we have that

$$G_{\mincf}(\hat{P}) \gg G(\hat{P})$$

3.3.3 Summary - Conflict Results

We considered the concept of conflict for completely certain probability distributions and defined the conflict measure CF. Then assessments are made of how conflict affects the aggregated probability $\hat{P}$ and its information content $G(\hat{P})$. An example of conflict cases was provided in Table 1 which led to the idea of minimal and maximal conflict. Finally for the aggregation of completely certain and uncertain probability distributions, conflicts are analyzed showing that even minimal conflict significantly reduces the information content of the aggregated probability of the P_{CC} s and P_{CU} s.

4. Aggregation of Possibility - Probability Distributions

In this section we will consider the aggregation of uncertainty from heterogeneous sources that have both probabilistic and possibilistic representations. We have previously considered the related problem of possibilistic conditioned probability [45] and evaluation of the informativeness of such a probability [10] and will compare these results to the aggregation developed in this section.

Our approach here is to transform a possibility distribution to a probability distribution and then use the probability averaging combination developed in the previous section to provide an aggregated result.

4.1 Possibility- Probability Transformations

A number of possibility transformations have been considered [19,21,45] and here we will utilize one initially suggested by Dubois and Prade [8,9]. For this we must assume a possibility distribution where

$$\pi_1 \geq \pi_2 \geq \dots \geq \pi_n \text{ and } \pi_1 = 1.$$

Then we can obtain an associated probability distribution $P_1 = (p_{11}, \dots, p_{1n})$ (denoting $\pi_{n+1} = 0$) where:

$$p_{1j} = \sum_{k=j}^n (\pi_k - \pi_{k+1}) / k, \quad \text{for } j = 1 \text{ to } n \quad (14)$$

The distribution P_1 from the transformation will then be aggregated using our approach in the previous section with some other probability distribution P_2 . The aggregated distribution will then be evaluated using the Gini information measure by considering the two extreme probability distributions P_{CC} and P_{CU} for P_2 as described in section 3.

We can illustrate this transformation on the possibility distribution for the PARS example. Recall $\Pi_{\text{PARS}} = \{1.0, .7, .6, .3\}$. Then we obtain:

$$\begin{aligned} p_4 &= (.3-0)/4 = .075; & p_3 &= (.6-.3)/3 + .075 = .175 \\ p_2 &= (.7-.6)/2 + .175 = .225; & p_1 &= (1.0-.7)/1 + .225 = .525 \end{aligned}$$

We will later use this transformed distribution to aggregate with the PARS probability distribution and use an information measure to assess the aggregated result.

In this section we specifically examine three possibility distribution cases which are to be transformed into probability distributions using eq. (14). These cases are:

1. Complete Certainty: CC $\Pi_{CC} = \{1, 0, \dots, 0, 0\}$
2. Complete Uncertainty: CU $\Pi_{CU} = \{1, 1, \dots, 1, 1, 1\}$
3. Intermediate Uncertainty: CI $\Pi_{CI} = \{1, 1, \dots, 1, 0, 0 \dots 0\}$

4.1.1 Possibility - Probability Transformations for Three Possibility Distributions

Here we compute the probability distributions corresponding to three possibility distribution cases using the Dubois-Prade transformation of eq. (14).

Π_{CC} :

$$\begin{aligned}
p_1 &= (1-0)/1 + (0-0)/2 + \dots (0-0)/n = 1 \\
p_k &= (0-0)/k + (0-0)/k+1 + \dots (0-0)/n = 0 \quad k = 2, \dots, n \\
P_1 &= \{1, 0, \dots, 0, 0\}
\end{aligned} \tag{15a}$$

So P_1 is completely certain and from eq. (10), $G(P_1) = 0$.

Π_{CU} :

$$\begin{aligned}
p_k &= (1-1)/k + (1-1)/k+1 + \dots (1-0)/n = \frac{1}{n} \quad k = 1, \dots, n \\
P_1 &= \left\{ \frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n} \right\}
\end{aligned} \tag{15b}$$

In this case P_1 is completely uncertain and $G(P_1) = 1$ from eq. (11).

Π_{CI} :

$$\begin{aligned}
p_k &= (1-1)/k + (1-1)/k+1 + \dots (1-0)/m = \frac{1}{m} \quad k = 1, \dots, m \\
p_j &= (0-0)/k + (0-0)/k+1 + \dots (0-0)/n = 0 \quad k = m+1, \dots, n \\
P_1 &= \left\{ \frac{1}{m}, \frac{1}{m}, \dots, \left(p_m = \frac{1}{m}\right), 0, \dots, 0 \right\} \\
G(P_1) &= 1 - \left(\sum_{k=1}^m \left(\frac{1}{m}\right)^2 + \sum_{k=m+1}^n 0^2 \right) = 1 - m \frac{1}{m^2} = 1 - \frac{1}{m}
\end{aligned} \tag{15c}$$

We note that as $m \rightarrow n-1$, this distribution becomes similar to P_{CU} and $G(P_1) \rightarrow 1$.

4.2 Aggregated Probability Computations with Transformed Possibilities

To focus on the relationship between an existing probability distribution and a probability distribution obtained from a possibility transformation, we will only consider two distributions at a time in the aggregation and equally weight them. So from eq. (1)

$$\hat{p}_k = w_1 p_{1k} + w_2 p_{2k} = \frac{1}{2} (p_{1k} + p_{2k})$$

First we aggregate using for $P_1 = \{1, 0, \dots, 0, 0\}$ (eq. 15a) with the distribution P_2 .

$$\begin{aligned}
\hat{p}_1 &= \frac{1}{2} (1 + p_{21}) = \frac{1}{2} + p_{21}/2 \\
\hat{p}_k &= p_{2k}/2 \quad k = 2 \dots n \\
\hat{P}_{CC} &= \left\{ \frac{1}{2} + p_{21}/2, p_{22}/2, \dots, p_{2n}/2 \right\}
\end{aligned} \tag{16}$$

We can confirm this is a valid distribution as

$$\sum_{k=1}^n \hat{p}_k = \frac{1}{2} + p_{21}/2 + p_{22}/2 + p_{23}/2 \dots = \frac{1}{2} \left(\sum_{k=1}^n p_{2k} + 1 \right) = \frac{1}{2} (1 + 1) = 1$$

Next from eq. 15b, where $P_1 = \{ \frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n} \}$

$$\begin{aligned} \hat{p}_k &= \frac{1}{2} \left(\frac{1}{n} + p_{2k} \right) = \frac{1}{2n} + p_{2k}/2 \quad k = 1, \dots, n \\ \hat{P}_{CU} &= \left\{ \frac{1}{2n} + p_{21}/2, \frac{1}{2n} + p_{22}/2, \dots, \frac{1}{2n} + p_{2n}/2 \right\} \end{aligned} \quad (17)$$

and this is clearly a valid distribution

$$\sum_{k=1}^n \hat{p}_k = n \frac{1}{2n} + \left(\sum_{k=1}^n p_{2k} / 2 \right) = \frac{1}{2} + (1/2) = 1$$

Finally the intermediate uncertainty case (eq. 15c) where $P_1 = \{ \frac{1}{m}, \frac{1}{m}, \dots, (p_m = \frac{1}{m}), 0, \dots, 0 \}$ gives

$$\begin{aligned} \hat{p}_k &= \frac{1}{2} \left(\frac{1}{m} + p_{2k} \right) = \frac{1}{2m} + p_{2k}/2 \quad k = 1 \dots m; \quad \hat{p}_k = p_{2k}/2 \quad k = m+1 \dots n \\ \hat{P}_{CI} &= \left\{ \frac{1}{2m} + p_{21}/2, \dots, \frac{1}{2m} + p_{2m}/2, p_{2m+1}/2, \dots, p_{2n}/2 \right\} \end{aligned} \quad (18)$$

and as before these probabilities sum to 1

$$\sum_{k=1}^n \hat{p}_k = m \frac{1}{2m} + \left(\sum_{k=1}^m p_{2k} / 2 \right) + \sum_{k=m+1}^n p_{2k}/2 = \frac{1}{2} + \left(\sum_{k=1}^n p_{2k} / 2 \right) = \frac{1}{2} + (1/2) = 1$$

So now we illustrate the aggregation process by combining the transformed possibility and the probability distribution for the PARS example.

$$\hat{p}_4 = \frac{1}{2} (.075 + .15) = .1125; \quad \hat{p}_3 = \frac{1}{2} (.175 + .15) = .1625$$

$$\hat{p}_2 = \frac{1}{2} (.225 + .3) = .2625; \quad \hat{p}_1 = \frac{1}{2} (.525 + .4) = .4625$$

In the next section we assess this aggregation with Gini index.

4.3 Aggregated Probability Evaluations Using Gini Measures

Next we use the Gini information measure to assess the results of the aggregation. First we express the measure for each of the three aggregated probability cases. Then we evaluate these measures for the two cases: $P_2 = P_{CC}$; $P_2 = P_{CU}$.

We denote by S the frequently occurring term $S = \sum_{k=1}^n p_{2k}^2$ in the following computations of the Gini measure. From eqs. (16), (17) and (18) we obtain

$$G(\hat{P}_{CC}) = 1 - \left[\left(\frac{1}{2} + p_{21}/2 \right)^2 + \sum_{k=1}^n \left(p_{2k}/2 \right)^2 \right] = 1 - \frac{1}{4} (S + 2p_{21} + 1)$$

$$G(\hat{P}_{CU}) = 1 - \sum_{k=1}^n \left(\frac{1}{2n} + p_{2k}/2 \right)^2 = 1 - \frac{1}{4} \left(S + \frac{2}{n} \sum_{k=1}^n p_{2k} + \frac{n}{n^2} \right) = 1 - \frac{1}{4} \left(S + \frac{3}{n} \right)$$

$$G(\hat{P}_{CI}) = 1 - \left(\sum_{k=1}^m \left(\frac{1}{2m} + \frac{p_{2k}}{2} \right)^2 + \sum_{k=m+1}^n \left(\frac{p_{2k}}{2} \right)^2 \right) = 1 - \frac{1}{4} \left(S + \frac{2}{m} \sum_{k=1}^m p_{2k} + \frac{1}{m} \right)$$

4.3.1 Case 1: $P_2 = P_{CC}$

$$S = \sum_{k=1}^n p_{2k}^2 = (1^2 + 0^2 + \dots + 0^2) = 1$$

We will see that for each case, $G(\hat{P})$ is less than the worst of the measures for P_1 and P_2 .

$$G(\hat{P}_{CC}) = 1 - \frac{1}{4} (S + 2p_{21} + 1) = 1 - \frac{1}{4} (1 + 2 \cdot 1 + 1) = 1 - \frac{1}{4} (4) = 0$$

$$G(\hat{P}_{CU}) = 1 - \frac{1}{4} \left(S + \frac{3}{n} \right) = 1 - \frac{1}{4} \left(1 + \frac{3}{n} \right) = \frac{3}{4} \left(1 - \frac{1}{n} \right)$$

For the intermediate case we have

$$G(\hat{P}_{CI}) = 1 - \frac{1}{4} \left(S + \frac{2}{m} \sum_{k=1}^m p_{2k} + \frac{1}{m} \right) = 1 - \frac{1}{4} \left(1 + \frac{2}{m} \cdot 1 + \frac{1}{m} \right) = \frac{3}{4} \left(1 - \frac{1}{m} \right)$$

Table 2. Gini Measures: Case 1: $P_2 = P_{CC}$

Π	$G(P_1)$	$G(P_2)$	$G(\hat{P})$
CC	0	0	0
CU	$1 - \frac{1}{n}$	0	$\frac{3}{4} \left(1 - \frac{1}{n} \right)$
CI	$1 - \frac{1}{m}$	0	$\frac{3}{4} \left(1 - \frac{1}{m} \right)$

To summarize when P_1 is derived from a completely certain possibility and aggregated with a completely certain probability then as expected the measure is 0, i.e. no information has been added in the final combination. For the next two cases where $G(P_1) > 0$ we see that this aggregation shows that the completely certain probability P_2 has added information to the less informative derived probability as $G(\hat{P}) = \frac{3}{4} G(P_1)$ for each case.

4.3.2 Case 2: $P_2 = P_{CU}$

$$S = \sum_{k=1}^n p_{2k}^2 = \sum_{k=1}^n \left(\frac{1}{n}\right)^2 = n \frac{1}{n^2} = \frac{1}{n}$$

$$G(\hat{P}_{CC}) = 1 - \frac{1}{4} (S + 2p_{21} + 1) = 1 - \frac{1}{4} \left(\frac{1}{n} + 2\frac{1}{n} + 1\right) = \frac{3}{4} \left(1 - \frac{1}{n}\right)$$

Next

$$G(\hat{P}_{CU}) = 1 - \frac{1}{4} \left(S + \frac{3}{n}\right) = 1 - \frac{1}{4} \left(\frac{1}{n} + \frac{3}{n}\right) = 1 - \frac{1}{n}$$

Lastly

$$G(\hat{P}_{CI}) = 1 - \frac{1}{4} \left(S + \frac{2}{m} \sum_{k=1}^m p_{2k} + \frac{1}{m}\right) = 1 - \frac{1}{4} \left(\frac{1}{n} + \frac{2}{m} m^* \frac{1}{n} + \frac{1}{m}\right) = 1 - \frac{1}{4} \left(\frac{3}{n} + \frac{1}{m}\right)$$

Table 3. Gini Measures: Case 2: $P_2 = P_{CU}$

Π	$G(P_1)$	$G(P_2)$	$G(\hat{P})$
CC	0	$1 - \frac{1}{n}$	$\frac{3}{4} \left(1 - \frac{1}{n}\right)$
CU	$1 - \frac{1}{n}$	$1 - \frac{1}{n}$	$1 - \frac{1}{n}$
CI	$1 - \frac{1}{m}$	$1 - \frac{1}{n}$	$1 - \frac{1}{4} \left(\frac{3}{n} + \frac{1}{m}\right)$

From Table 3 we can see as expected that the completely certain aggregated probability P_1 is adding information to P_2 since $G(\hat{P}) = \frac{3}{4} G(P_2)$. In the second case where both distributions represent complete uncertainty, then the aggregated probability cannot have anymore information than in the original distributions.

Finally consider the intermediate possibility case. Recall that $m < n$ and so $\frac{1}{m} > \frac{1}{n}$.

Then $G(P_2) = 1 - \frac{1}{n} > G(P_1)$. Furthermore then

$$\frac{1}{4} \left(\frac{3}{n} + \frac{1}{m} \right) > \frac{1}{4} \left(\frac{3}{n} + \frac{1}{n} \right) = \frac{1}{n}$$

So in this case as well the aggregate probability is providing additional information

$$G(\hat{P}) = 1 - \frac{1}{4} \left(\frac{3}{n} + \frac{1}{m} \right) < 1 - \frac{1}{n} = G(P_2)$$

4.3.3 Summary - Aggregated Probability Measures

After developing the aggregated result of a probability and transformed possibility distribution, Gini measures were used to assess the results for the cases. Completely certain probability will add information to less than certain transformed possibility distributions. Likewise when a probability is completely uncertain, aggregating with the transformed probability will also add information.

Lastly to illustrate we use this measure to assess the aggregated probability for the PARS example. For the initial probability we have

$$G(P_{\text{PARS}}) = 1 - (.4^2 + .3^2 + .15^2 + .15^2) = 1 - .295 = .705$$

$$G(\hat{P}_{\text{PARS}}) = 1 - (.1125^2 + .1625^2 + .2625^2 + .4625^2) = 1 - .322 = .678$$

So we see that the aggregation has increased the information as

$$G(\hat{P}_{\text{PARS}}) < G(P_{\text{PARS}})$$

4.4 Ratio Scale Transformations

Another approach to transforming a possibility into a probability is the ratio scale (RS) transformation [25].

$$p_k = \pi_k / \left(\sum_{k=1}^n \pi_k \right) \quad (19)$$

This RS transformation does not produce different probabilities for the possibility distributions we have been considering. For Π_{CC} , $\sum_{k=1}^n \pi_k = 1$; Π_{CU} , $\sum_{k=1}^n \pi_k = n$;

Π_{CI} , $\sum_{k=1}^n \pi_k = m$. Using these we have

$$\Pi_{CC}: p_1 = 1/1 = 1; p_k = 0/1 = 0 \quad k = 2, \dots, n$$

$$\Pi_{CU}: p_k = 1/n = \frac{1}{n} \quad k = 1, \dots, n$$

$$\Pi_{CI}: p_k = 1/m = \frac{1}{m} \quad k = 1, \dots, m; p_k = 0/m = 0 \quad k = m+1, \dots, n$$

However we note that there is a difference in general [20] between these approaches. Consider the possibility distribution $\Pi = \{1, 0.8, 0.2\}$. For the RS

transformation $\sum_{k=1}^3 \pi_k = 2.0$. Then we obtain:

$$p_1 = 1/2.0 = 0.5; p_2 = 0.8 / 2.0 = 0.4; p_3 = 0.2/2.0 = 0.1.$$

From the DP transformation (eq. 14) we have:

$$p_3 = (.2 - 0)/3 = .067; p_2 = (.8 - .2)/2 + 0.3 + 0.067 = .367; p_1 = (1 - 0.8)/1 + 0.367 = .567$$

If we compute $\hat{P}$ for both and calculate the Gini index we have

$$RS: G(\hat{P}) = 1 - 0.605 = 0.395$$

$$DP: G(\hat{P}) = 1 - 0.6475 = 0.3525$$

So the DP transformation is providing more information when used in our probability combining approach.

4.5 Discussion of Combination Approach

First we consider the following example for possibility and probability distributions in order to illustrate that is not necessary that an aggregate probability $\hat{P}$ is more informative than an initial probability P_2 . We shall apply our previous information measures and see that these are consistent in their assessments. Let the possibility and probability distributions be:

$$\Pi = \{1.0, 0.1, 0.1, 0.1\}; P_2 = \{0.05, 0.8, 0.05, 0.05\}$$

We can observe a degree of inconsistency between Π and P_2 . For example for $\pi_1 = 1.0$, the corresponding probability is very small. For our purposes, we focus here on C_Z as it provides a range of values to evaluate the idea of consistency as it relates to the aggregation approach. We can note that the maximum value that $C_Z(\Pi, P)$

$= \sum_{i=1}^n \pi_i * p_i$ can have in general is 1 as $\sum_{i=1}^n p_i = 1$ and π_i is at most 1. Thus, the

range of C_Z is the interval $[0...1]$ where 0 can be considered as complete inconsistency and 1 complete consistency. In a more general sense we can relate this to the concept of conflict, of which consistency is only one aspect.

If we use the Zadeh [47] consistency measure, we have here

$$C_Z(\Pi, P_2) = (1 * .05 + .1 * .8 + .1 * .05 + .1 * .05) = 0.14$$

Since 0 can be considered as complete inconsistency the value of 0.14 indicates very limited consistency. In a more general sense this is related this to broader semantic issues such source reliability and trustworthiness.

Now we can apply the transformation (eq. 14) on Π and obtain

$$P_1 = \{ 0.925, 0.025, 0.025, 0.025 \}$$

and the combined probability distribution is

$$\hat{P} = \{ 0.4875, 0.0375, 0.0625, 0.4125 \}$$

We can assess this situation with the information measures. Starting with the Shannon entropy we have a smaller entropy for the initial probability distribution P_2 :

$$S(P_2) = 0.628 < S(\hat{P}) = 1.012$$

Likewise we can see that the Gini index measure also yields a similar result for this case involving a degree of inconsistency indicating that the aggregated probability $\hat{P}$ is less informative:

$$G(P_2) = 1 - 0.648 = 0.352 < G(\hat{P}) = 1 - 0.413 = 0.587$$

So the information measures are compatible with our intuitive assessment of the inconsistency of Π and P as well as the specific value of the consistency measure C_Z .

Finally we can compare this transformation approach to the possibilistic conditioning approach [10]. We will note that our new approach provides a more reasonable outcome in several ways.

For the cases of Π_{CC} and Π_{CU} the conditioning approach makes no distinction between the initial probabilities being conditioned. For Π_{CC} the result is P_{CC} and for Π_{CU} , it is P_{CU} regardless of the initial probability. The transformation approach is the same here in the cases where both the possibility and probability agree which is

certainly intuitively satisfactory. But rather than the default probabilities in the conditioned probability, the new approach gives for both (Π_{cc}, P_{cu}) and (Π_{cu}, P_{cc}) :

$$\hat{P} = \left\{ \frac{n+1}{2n}, \frac{1}{2n}, \dots, \frac{1}{2n} \right\}$$

which is more satisfactory and gives a symmetric solution for the symmetry of (Π_{cc}, P_{cu}) and (Π_{cu}, P_{cc}) . The differences are more effectively combined as opposed to the conditioning approach for which the possibility distribution totally drives the resulting probability.

5. Aggregation of Possibility Distributions

In this section we are considering a situation in which we have information represented by possibility distributions and we want to develop approaches to combining or aggregating these distributions. We shall first consider simple conjunctions of possibility distribution and establish some basic criteria for their combinations. Then we present a more nuanced approach that will allow us to introduce more developed criteria for selection of the best aggregation and fusion of the distributions.

5.1. Specificity Measure and Possibility Aggregation

Let V be a variable with domain $X = \{x_1, x_2, \dots, x_n\}$ and assume we have q pieces of information/evidence on the value of the variable V . We will have some observations of the value of V denoted by E

$$E = \{\Pi_1, \dots, \Pi_q\}$$

where each Π_i is a possibility distribution and the objective is to produce the combined result $\hat{\Pi}$.

A common approach to such an aggregation is by conjunctions of the possibility values [43]. So for aggregation of Π_i and Π_j , the k th value of a resulting possibility distribution $\hat{\Pi}$ is

$$\hat{\pi}_k = \pi_i(x_k) \wedge \pi_j(x_k)$$

The specificity measure is the max value minus the average of the other possibility values [40]:

$$Sp(\hat{\Pi}) = \hat{\pi}_m - \left(\sum_{k=1, \neq m}^n \hat{\pi}_k \right) / n$$

The specificity of a possibility distribution $\hat{\Pi}$ is decreased or lost in two ways:

- i) More possibility values loses specificity
- ii) Decreasing normality also loses specificity

Also the conjunction of possibility values in general decreases the resultant specificity value unless all values are equal.

This measure of specificity is one factor that can be used to guide the aggregation process. Consider all (non-empty) subsets $B_i \subseteq E$ and the aggregations of the distributions in each subset. Then we can preliminarily assess the quality of the aggregation of each subset with respect to two factors:

- a. specificity of each subset B_i
- b. degree to which we are satisfied or confident with the aggregation

Some criteria of satisfaction could be - a larger cardinality of subset B_i is better (monotonically increasing); - time - more recently obtained distributions are more valuable; if there is conflict then put greater weight on the more reliable subset B_i . We will introduce more specific criteria in section 5.3.

The specificity measure takes into account aspects of conflict and certainty (lack of uncertainty). Conflict in the sense of lack of good solutions (subnormality) reduces specificity. Uncertainty (in the sense of too many possible solutions) also reduces specificity.

5.2. Combination by Conjunction of Distributions

Here we show a more detailed approach to aggregation of possibility distribution for which we may allow domain elements to be fused to produce the final result [46]. In order to aggregate the possibility distributions in E , we must evaluate these distributions over their domain $X = \{x_1, x_2, \dots, x_n\}$. Specifically any subset of possibility distributions, B_i , must be instantiated with all permutations with repetition from X . For example for three domain values, $\{x_1, x_2, x_3\}$, some of such permutations might be:

$$(x_1 x_1 x_2), (x_1 x_2 x_1), (x_1 x_2 x_2), (x_1 x_3 x_2), (x_2 x_1 x_1) \dots$$

We will denote this set of permutations $\mathbf{P}(X)$ and let z_i be one of these permutations. To begin to formulate the aggregation we compute $R(z_i)$, $z_i \in \mathbf{P}(X)$:

$$R(z_i) = \bigwedge \Pi_k(z_i), \quad \text{for all } \Pi_k \in B_j \quad (20)$$

For the z_i 's, there will be a fusion function F that maps into the domain Y of the final aggregated possibility distribution for B_j , denoted $\hat{\Pi}(B_j)$:

$$F : z_i \rightarrow y_j \in Y$$

The specific F could be one of a number of functions depending on an application and the characteristics of the domain X . Often an averaging function will be used such as a mean or an OWA function.

Finally we compute an aggregated possibility distribution $\hat{\Pi}(B_j)$: $\{\hat{\pi}(y_1), \hat{\pi}(y_2), \dots\}$, as

$$\hat{\Pi}(B_j) = \text{Max } R(z_i), \quad \forall z_i \in \mathbf{P}(X) \text{ such that } y_j = F(z_i) \quad (21)$$

Now we consider an example to illustrate the above aggregation process. Assume an oil exploration company is evaluating an area for fracking drilling for oil/gas. There have been preliminary studies at sites relative to their geology, ease of equipment access and other factors. The company engages several experts to provide independent assessment of the area bounded by these sites.

Here for a specific example we assume there are three sites: $X = \{a, b, c\}$ and four possibility distributions representing the experts' assessments of the sites:

$$\Pi_1 = \{.7, 1.0, .3\}; \quad \Pi_2 = \{1.0, .8, .5\}; \quad \Pi_3 = \{.7, 1.0, .6\}; \quad \Pi_4 = \{.6, .9, 1.0\}$$

We will illustrate the aggregated possibility distribution for the subset, B , of the possibility distributions:

$$B = \{\Pi_1, \Pi_2, \Pi_4\}$$

The fusion function F for the z 's will be based on an average of the latitude and longitude co-ordinates of the specific three sites in a given permutation z of $\{a, b, c\}$:

$$\{(\text{lat}_a, \text{long}_a), (\text{lat}_b, \text{long}_b), (\text{lat}_c, \text{long}_c)\}$$

Note that z 's that correspond to a different permutation may still be mapped to the same y value. For example

$$y_{aab} = F(a,a,b) = F(a, b, a) = F(b,a,a) = ((\text{lat}_a + \text{lat}_a + \text{lat}_b)/3, (\text{long}_a + \text{long}_a + \text{long}_b)/3)$$

where for this example we denote the result as y_{aab} .

For this case we just show the aggregation for all z 's in $\mathbf{P}(X)$ starting with an initial "a". Table 5 shows the calculation of $R(z)$ for these z 's.

Table 4. Computation of $R(z)$

z	Π_1	Π_2	Π_4	$R(z)$	$y = F(z)$
aaa	.7	.1.0	.6	.6	y_{aaa}
aab	.7	1.0	.9	.7	y_{aab}
aac	.7	1.0	1.0	.7	y_{aac}
aba	.7	.8	.6	.6	y_{aab}
aca	.7	.5	.6	.5	y_{aac}
abb	.7	.8	.9	.7	y_{abb}
abc	.7	.8	1.0	.7	y_{abc}
acb	.7	.5	.9	.5	y_{abc}
acc	.7	.5	1.0	.5	y_{acc}

Considering just the values in Table 4, we have the partial calculations of the aggregated possibilities for the subset B:

$$\hat{\pi}(y_{aaa}) = \text{Max} \{ .6 \} = 0.6; \hat{\pi}(y_{aab}) = \text{Max} \{ .7, .6 \} = 0.7; \hat{\pi}(y_{aac}) = \text{Max} \{ .7, .5 \} = 0.7;$$

$$\hat{\pi}(y_{abb}) = \text{Max} \{ .7 \} = 0.7; \hat{\pi}(y_{abc}) = \text{Max} \{ .7, .5 \} = .7; \hat{\pi}(y_{acc}) = \text{Max} \{ .5 \} = 0.5;$$

These values are just for the z starting with an "a", so actually the final values for $\hat{\Pi}$ must include the those starting with "b" and "c" in the the permutation $\mathbf{P}(X)$ as well. We do not show the full development of $R(z)$ for all of these; but for this example we note that $z = baa$ maps to y_{aab} also and

$$R(baa) = \text{Min} (1.0, 1.0, .6) = 0.6$$

so

$$\hat{\pi}(y_{aab}) = \text{Max} \{ .7, .6, .6 \} = 0.7$$

The rest of the aggregated distribution follows in a similar fashion.

5.3 Quality Criterion Possibility Aggregation

Now we will develop a specific approach for selection of the most effective combination of possibility distributions. There are several factors we will investigate in developing a quality measure, including credibility, specificity and grade factors. Again denoting the aggregation of a subset B_i by $\hat{\Pi}(B_i)$, for each of the subsets of the information we describe measures to select the best $\hat{\Pi}(B_i)$.

A number of choices have been made relative to operators for combining factors in the aggregation evaluation. Our intention was to use appropriate default choice for these. Also our intention was to use familiar operations with the realization that other specific choices would be more suitable for particular applications. As we described in Section 2.3 Aggregation Approaches, the default choices we used are the most common representatives of the main classes of aggregation operations.

5.3.1 Combination Function

Here consider the concept of a combination function $\text{Comb}(x, y)$ to be used in conjunction with the possibility aggregation described in the previous section:

$$\text{Comb}: X \times X \rightarrow [0, 1]$$

where $\text{Comb}(x, y)$ represents the acceptability of combining x and y . We can extend this function by

$$\text{Comb}(x_1, \dots, x_r) = \text{Min}(\text{Comb}(x_i, x_j)) \text{ for every pair } (x_i, x_j) \in (x_1, \dots, x_r)$$

The function $\text{Comb}(x, y)$ has some of the characteristics of a similarity function but can include more sophisticated information as it is relating the acceptability of combining x and y , (note $\text{Comb}(x, y) \leq \text{Sim}(x, y)$). While clearly x and y must be similar (close) to combine, in some cases they should not be combined as semantic aspects of the information may have to be taken into account. For example if we are looking for the location of a house and if x and y are on different sides of a river here averaging them will end with the location of the house in the river.

Such a function can then be used to provide this sort of semantic insight for the computation of $R(z_i)$ when incorporated into eq. (20)

$$R(z_i) = \bigwedge \Pi_k(z_i) \bigwedge \text{Comb}(z_i), \forall z_i \in \mathbf{P}(X) \text{ such that } y_j = F(z_i) \quad (21)$$

5.3.2 Credibility Factor

Next we describe the credibility of subsets $B_i \subseteq E$. Let

$$\mu: B_i \Rightarrow [0, 1]$$

We will interpret $\mu(B_i)$ as indicating the credibility of the selection of the particular subset B_i of the pieces of evidence. Assigning credibility to sources can be considered at several levels. For example an overall global credibility can be associated with each of the sources. Another approach is where the credibility can be dependent on the type of variables being considered.

One possible factor involved in credibility is the confidence of the provider in the observation provided which we denote α_i . This is a value in the unit interval. The default value is 1, thus if a provider says nothing we let $\alpha_i = 1$.

One way of including these α_i in the process is to calculate for every subset B_i an aggregated value. Let $B_n = \{\Pi_i, \Pi_j, \dots \Pi_k\}$

$$\alpha_n(B_n) = \text{OWA Aggregation}(\alpha_i, \alpha_j, \dots \alpha_k)$$

This OWA [39] should be represented as an aggregation between the average and the min. OWA is a commonly used and well accepted aggregation approach. It provides a parameterized family that makes it convenient to use tailored OWAs for an application. So we note

$$\mu(E) = 1, \mu(\emptyset) = 0 \text{ and } \mu(B_i) \geq \mu(B_j) \text{ if } B_i \supseteq B_j$$

5.3.3 Specificity Factor

Now we can combine the specificity of the possibility distribution, $Sp(\hat{\Pi}(B_i))$, and the credibility for the subset B_i , $\mu(B_i)$, for a measure of the quality of this distribution.

$$\text{Qual}(\hat{\Pi}(B_i)) = \mu(B_i) * Sp(\hat{\Pi}(B_i))$$

Here the product operation (*) captures the basic aggregation operation of conjunction - i.e. "anding". As a default choice here it is more sensitive and easier for assessing the properties of the aggregation. So this provides a criterion for selection of the best combined value, $\hat{\Pi}(B_j)$

$$\text{Max Qual}(\hat{\Pi}(B_i)).$$

We can present to the decision maker both the aggregated value $\hat{\Pi}(B_i)$ and its associated quality.

5.3.4 Grade Factor

Another aspect of the information provided can be thought of intuitively as the "grade" of the information, denoted Gr_i . This will enable us to address our previous concerns about two wide pieces of information combining to an acceptable value. For each piece of information Π_i we can obtain its grade as

$$Gr_i(\Pi_i) = g(\text{Sp}(\Pi_i))$$

The actual formulation of g is subjective. Here the $Gr_i \in [0, 1]$ and g is monotonic; the smaller the specificity value, the lower the grade. Using this grade Gr_i we can calculate for any subset B_i of observations

$$Gr(B_i) = \text{the average of the grades of the information in } B_i.$$

Average is another one of the basic aggregation operators that is commonly used and provides easy intuition. Also it is an unbiased estimator. Now this aspect can be added as another factor in determining the quality

$$\text{Qual}(\hat{\Pi}(B_i)) = \mu(B_i) * \text{Sp}(\hat{\Pi}(B_i)) * Gr(B_i) \quad (22)$$

We now observe that a piece of provided information, $\mathcal{E}_i$, can be seen as a 4 tuple

$$\mathcal{E}_i = \langle \Pi_i, \mu_i, S_i, Gr_i \rangle$$

The pieces of information provided, denoted $\mathcal{E}_i$, have at least the two aspects that the observation provided, Π_i and the source S_i . More formally the measure μ is over the space S of sources. The sources, confidences and grades of the information in a subset B_i do not effect the determination of the aggregated value $\hat{\Pi}(B_i)$. The value $\hat{\Pi}(B_i)$ associated with the subset B_i of information is determined by the observations in B_i . The other factors are concerned with determining the quality of $\hat{\Pi}(B_i)$

6. Summary and Conclusions

In this paper we considered new approaches for combining separately possibilistic uncertainty, probabilistic uncertainty and situations where both forms of uncertainty appear. Many decision making applications of interest have multiple sources of information with different forms of uncertainty which makes it critical for such applications to be able to effectively utilize all such information.

An approach to probability aggregation using the rational consensus with equi-weighting was developed. The aggregation is analyzed with information measures to show it's usefulness and with considerations of conflict issues. Then possibility to probability transformations are used and the resultant probability is aggregated and the result evaluated with information measures. Finally a general approach to combining possibility distributions directly using a quality criteria.

We are considering the extensions of the quality criteria approach for possibility aggregations to our other aggregation techniques and how to effectively present such quality criteria to human decision makers. Also we are developing environments to carry out Monte Carlo simulations to test the combination approaches and the information evaluation measures. We are investigating ways to apply such simulations to actual decision making and assess if more effective outcomes result when the evaluation measures have indicated that the combined probability or possibility is more informative.

Acknowledgements

We would like to thank the Naval Research Laboratory's Base Program, Program Element No. 0602435N for sponsoring this research. Ronald Yager has been in part supported by ONR grant award Number N00014-13-1-0626.

References

1. O. Aristondo, J. Garcia-Lparesa, C. de la Vega, R. Pereira (2012) The Gini Index, the dual decomposition of aggregation functions and the consistent measurement of inequality. *Int. J of Intelligent Systems* 27: 132-152.
2. Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1984) *Classification and regression trees*. Wadsworth & Brooks/Cole, Monterey, CA.

3. Beliakov G, Pradera A and Calvo T (2007) *Aggregation functions: a Guide for practitioners*. Springer, Heidelberg DG.
4. Carvalho, A and Larson K. (2012) A consensual linear opinion pool. Proc. 23rd Int. Joint Conf. on Artificial Intelligence. 2518-2524.
5. Clemen R and Winkler R (1999) Combining probability distributions from experts in risk analysis. Risk Analysis 19: 187-203.
6. DeGroot, M. (1974) Reaching a consensus. Jour of Am. Stat. Assoc. 69: 118-121.
7. Delgado M, Moral S (1987) On the concept of possibility-probability consistency. Fuzzy Sets and Systems 21: 311-318.
8. Dubois D, Prade H (1982) On several representations of an uncertain body of evidence. In Gupta M and Sanchez E (eds) *Fuzzy Information and Decision Processes*. North Holland, Amsterdam, pp 167-182
9. Dubois D, Prade H (1983) Unfair coins and necessity measures: towards a possibilistic interpretations of histograms. Fuzzy Sets and Systems 10: 15-27.
10. Elmore P, Petry, F and Yager R (2014) Comparative measures of aggregated uncertainty representations. Journal of Ambient Intelligence and Humanized Computing, 5#6, 809-819.
11. Eliazar I, Sokolov I (2010) Maximization of statistical heterogeneity: from Shannon's Entropy to Gini's Index. Physica A 389: 3023-3038.
12. Esher L, Hall S, Regnier E, Sanchez P, Hansen J and Singham D (2010) Simulating pirate behavior to exploit environmental information. Proceedings of 2010 Winter Simulation Conference. 1330-1335.
13. Ferson S and Kreinovich V (2001) Representation, elicitation, and aggregation of uncertainty in risk analysis - from traditional probabilistic techniques to more general, more realistic approaches: A survey. Univ. of Texas at El Paso Computer Science Tech Report #11-1-2001.
14. Gini C (1912) Variabilita e mutabilita (Variability and Mutability), Tipografia di Paolo Cuppini, Bologna, Italy, 156 pages.
15. Gupta C (1993) A note on the transformation of possibilistic information into probabilistic information for investment decisions. Fuzzy Sets and Systems 56: 175-182.
16. Hansen S (2012) The dynamics of Somali piracy. Studies in Conflict and Terrorism 35:7-8, 523-530
17. Hartley, R. (1928) Transmission of Information. The Bell System Technical Journal. 7(3), 535-563.
18. Hunter A and Liu W (2006) Fusion rules for merging uncertain information. Information Fusion, 7:97-134.
19. Jumarie G (1994) Possibility-probability transformation: A new result via information theory of deterministic functions. Kybernetes, 23:56-59.
20. Klir G and Parviz B (1992) Probability-possibility transformations: a comparison. Int. J General Systems, 21: 291-310.
21. Klir G (2006) *Uncertainty and information*. John Wiley, Hoboken NJ.
22. Lehrer K and Wagner C (1981) *Rational Consensus in Science and Society*. D. Reidel, Dordrecht Holland.
23. Li D and Santos E (2011) Deception Detection in Human Reasoning. 2011 IEEE Int. Conf. on Systems, Man and Cybernetics, 165-171, Anchorage AK.
24. Office of Naval Intelligence, Piracy Analysis and Warning Weekly (PAWW) Report for 31 October - 6 November 2013, URL://www.oni.navy.mil/IntelligenceCommunity/piracy.htm. Downloaded from Homeland Security Digital Library: <https://www.hsdl.org/?view&did=746216>. Access date: 19 Feb 2015.

25. Oussalah M (2000) On the probability/possibility transformations: a comparative analysis. *Int. J General Systems*, 29: 671-718.
26. Parsons S (2001) *Qualitative Methods for Reasoning Under Uncertainty*. MIT Press, Cambridge MA
27. Pedrycz W and Gomide F (1996) *An Introduction to Fuzzy Sets: Analysis and Design*. MIT Press: Boston MA.
28. Renyi, A. (1961). On measures of information and entropy. *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability 1960*. pp. 547-561.
29. Renyi, A. (1970) *Probability Theory*. Amsterdam, ND: North-Holland.
30. Reza F (1961) *An Introduction to Information Theory*. McGraw Hill, New York
31. Rickard J and Aisbett J (2014) New classes of threshold aggregation functions based on the Tsallis q-exponential with applications to perceptual computing. *IEEE Trans on Fuzzy Systems*, 22 # 3, 672-684.
32. Sent, D. and van de Gaag, L. (2007) On the Behavior of Information Measures for Test Selection. In eds. J Carbonell and J. Siebnarm, *Lecture Notes in AI 4594*, Springer-Verlag, Berlin.
33. Sentz K and Ferson S (2002) Combination of evidence in Dempster-Shafer theory. <http://www.sandia.gov/epistemic/Reports/SAND2002-0835.pdf>
34. Shafer G (1976) *A Mathematical Theory of Evidence*. Princeton University Press: Princeton NJ.
35. Shannon C (1948) A mathematical theory of communication. *Bell Systems Tech. Jour.* 27; 379-423; 623-656.
36. Xu, D. and Erdogmus, D. (2010). Renyi's entropy, divergence and their nonparametric estimators. In J. Principe, ed. *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*. Springer, pp. 47-102.
37. Yager R (1982) Measuring tranquility and anxiety in decision making. *Int. Jour of General Systems*. 8: 139-146.
38. Yager R (1987) Quasi-associative operations in the combination of evidence. *Kybernetes* 16: 37-41.
39. Yager R (1988) On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Trans. on Systems, Man and Cybernetics*. 18: 183-190.
40. Yager R (1992) On the specificity of a possibility distribution. *Fuzzy Sets and Systems* 50 279-292.
41. Yager R (1995) Measures of entropy and fuzziness related to aggregation operators. *Information Sciences* 82: 147-166.
42. Yager R (1996) On mean type aggregation. *IEEE Trans. on Systems, Man and Cybernetics*. 26: 209-221.
43. Yager R (1997) A general approach to the fusion of imprecise information. *Int. Jour. of Intelligent Systems*. 12:1-29.
44. Yager R (2011) A measure based approach to the fusion of possibilistic and probabilistic uncertainty. *Fuzzy Optimization and Decision Making*. 10:91-113.
45. Yager R (2012) Conditional approach to possibility-probability fusion. *IEEE Trans on Fuzzy Systems* 20: 46-56.
46. Yager R (2011) On possibilistic and probabilistic information fusion. *Int. Jour. of Fuzzy System Applications*. 1:1-14.
47. Zadeh L (1978) Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems* 1: 3-28.

Figure 1. PAWW Piracy Attack Reports for Horn of Africa Area