

Comparative measures of aggregated uncertainty representations

Paul Elmore · Fred Petry · Ronald Yager

Received: 17 November 2013 / Accepted: 20 February 2014 / Published online: 16 March 2014
© Springer-Verlag Berlin (outside the USA) 2014

Abstract Uncertainty must be taken into account in all aspects of ambient intelligence and human decisions and activities. We investigate how to utilize both probabilistic and possibilistic sources of information for use in humanized decision-making. In particular we examine aspects of the possibilistic conditioning of probability developed by Yager. To provide bounding of the resulting probability analysis of the cases of completely certain and uncertain probability and possibility distribution are carried out. Additionally the cases of intermediate uncertainty and a general case of possibilities are analyzed. The Zadeh consistency measure is also used to assess these cases. To consider whether the conditioned probability is more informative for decision-making, three measures, Shannon entropy, Gini index and Renyi entropy are used to compare the original probability distributions and the conditioned distribution for the cases described.

Keywords Information theory · Decision making · Possibility theory · Shannon entropy · Gini index · Renyi entropy · Possibilistic conditioning · Consistency measures

1 Introduction

Uncertainty is pervasive in the ordinary, everyday activities and decisions of humans. Fuzzy set techniques

have been widely recognized for dealing with uncertainty in ambient intelligence (Acampora and Loia 2008) and human-centric systems (Pedrycz 2010). In this paper we are interested in a deeper understanding of such uncertainties and how they can be quantified for human decision makers.

One aspect that must be considered in particular is how to deal with the inherent uncertainty involved when information is aggregated in order to become useful for decision making. Effective decision-making should be able to make use of all the available, relevant information about such aggregated uncertainty. In this paper we investigate quantitative measures that can be used to guide the use of aggregated uncertainty. While there are a number of possible approaches to aggregate the uncertainty information that has been gathered, this paper will examine uncertainty aggregation by the soft computing approach of possibilistic conditioning of probability distribution representations using the approach of Yager (2012). This form of aggregation makes it very amenable to apply the information measures we consider in this paper.

To formalize the problem, let V be a discrete variable taking values in a space X that has both aleatory and epistemic sources of uncertainty (Parsons 2001). Let there be a probability distribution $P: X \rightarrow [0, 1]$ such that $p_i \in [0, 1]$, $\sum_{i=1}^n p_i = 1$ that models the aleatory uncertainty. Then the epistemic uncertainty can be modeled by a possibility distribution (Zadeh 1978) such that $\Pi: X \rightarrow [0, 1]$, where $\pi(x_i)$ gives the possibility that x_i is the value of V , $i = 1, 2, \dots, n$. A usual requirement here is the normality condition, $\max_x [\pi(x)] = 1$, that is at least one element in X must be a fully possible. Abbreviating our notation so that $p_i = p(x_i)$, etc. and $\pi_i = \pi(x_i)$, etc., we have $P = \{p_1, p_2, \dots, p_n\}$ and $\Pi = \{\pi_1, \pi_2, \dots, \pi_n\}$.

P. Elmore · F. Petry (✉)
Naval Research Laboratory, Stennis Space Center,
Mississippi, MS, USA
e-mail: fpetry@nrlssc.navy.mil

R. Yager
Machine Intelligence Institute, Iona College,
New Rochelle, NY, USA

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

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

A strength of this approach using conditioned probability is that it also captures Zadeh's concept of consistency between the possibility and the original probability distribution. Consistency provides an intuition of concurrence between the possibility and probability distributions being aggregated. In Eq. (1), K is identical to Zadeh's possibility-probability consistency measure (Zadeh 1978), $C_Z(\Pi, P)$; i.e. $C_Z(\Pi, P) = K$.

As an example of a conditioned probability distribution that could be used to provide guidance to a decision maker, consider the following military problem. Over the first decade of the 21st century, a major cause of casualties in both Iraq and Afghanistan combat zones has been from improvised explosive devices (IEDs). Prevention/avoidance of IED attacks is a critical decision and should be based on assessment of the most probable placements of IEDs (Benigni and Furrer 2012). One approach is to consider historical probability distributions characterizing typical placements sites. Let the placement sites considered be X_1, X_2, X_3 , and X_4 . The variable V_{IED} takes values from the space $X = \{X_1, X_2, X_3, X_4\}$. For this example, let the probability distribution for past IED placements be denoted as $p(V_{\text{IEDhistorical}}) \equiv P_{\text{IED}}$. So we have the distribution

$$P_{\text{IED}} = \left\{ \frac{X_1}{0.3}, \frac{X_2}{0.2}, \frac{X_3}{0.4}, \frac{X_4}{0.1} \right\},$$

where the upper halves indicate locations and the lower the corresponding probabilities.

Typically there may be additional or more current information based on intelligence reports that are subjective in nature. A possibility distribution could be used to represent such subjective information. If the intelligence officials provide their assessment the possibility distribution for this will be denoted as $\Pi(V_{\text{IEDintelligence}}) \equiv \Pi_{\text{IED}}$ and we have

$$\Pi_{\text{IED}} = \left\{ \frac{X_1}{1}, \frac{X_2}{0.6}, \frac{X_3}{0.8}, \frac{X_4}{0.2} \right\}$$

We can now combine these by the possibilistic conditioning approach. Using Eq. (1) we have first

$$K = 0.3 \times 1 + 0.2 \times 0.6 + 0.4 \times 0.8 + 0.1 \times 0.2 = 0.76,$$

Then,

$$\hat{p}_1 = 0.3 \times 1 / 0.76 = 0.39; \dots; \hat{p}_4 = 0.1 \times 0.2 / 0.76 = 0.03$$

The conditioned probability distribution for IED locations is then

$$\hat{P}_{\text{IED}} = \left\{ \frac{X_1}{0.39}, \frac{X_2}{0.16}, \frac{X_3}{0.42}, \frac{X_4}{0.03} \right\}.$$

The issue at hand is does \hat{P}_{IED} represent an improved estimate of the IED locations? In order to provide intuition and tools to assess this question, this paper will provide the following discussions, reusing the IED distributions above as an ongoing numerical example. Section 2 begins by providing theorems for extreme case of Π , one of absolute certainty and the other of complete uncertainty. These theorems provide simplifications, check results and characterize the approach. The section then continues with combination of two more general Π distributions with for four different classes of Π distributions. In Sect. 3, we assess the utility of an aggregated uncertainty and to decide if this aggregation provides more effective information through consideration of information measures; including Shannon entropy, Gini index, and Renyi entropy; to gauge the aggregated uncertainty. For our on going IED numerical example, the Shannon entropy (Reza 1961),

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

yields for P_{IED} and \hat{P}_{IED}

$$S(\hat{P}_{\text{IED}}) = 1.13 < S(P_{\text{IED}}) = 1.28 \quad (3)$$

These measures for the more generalized analytic cases are presented here. Section 4 then discusses consistency and shows that it provides an additional measure that is compatible with the information measures from the previous section. The paper then provides a summary and discussion of future research in Sect. 5.

2 Aggregation of possibility and probability by conditioning

To examine the conditioning approach further we formulate four distinct cases for the possibility distributions. The first two, complete certainty, complete uncertainty, represent the extreme cases of possibility distributions. Then two intermediate cases, partial certainty and a generalized

Table 1 Possibility and probability distribution cases

Possibility distribution cases

Case 1: complete certainty $\Pi = \{0, 0 \dots, \pi_k = 1, \dots, 0, 0\}$ Case 2: complete uncertainty $\Pi = \{1, 1, \dots, 1, 1, 1\}$ Case 3: intermediate uncertainty $\Pi = \{1, 1, \dots, 1, 0, 0 \dots, 0\}$

Case 4: generalized possibility

$$\Pi = \{1, w_2, w_3, \dots, w_n\}, 0 < w_i < 1$$

Probability distribution cases

Case A: complete certainty $P_{cc} = \{0, 0 \dots, p_k = 1, \dots, 0, 0\}$ Case B: complete uncertainty $P_{cu} = \{1/n, 1/n, \dots, 1/n, 1/n\}$

certainty case will be discussed. For each case we provide instantiations of the cases based on the two extreme probability distributions, completely certain, P_{cc} , and completely uncertain, P_{cu} . These cases are shown in Table 1. Additional measures discussed in Sects. 3 and 4 provide guidance for use of the result.

2.1 Case 1: complete certainty

A possibility distribution with exactly one possibility value equal to 1 and all other values equal 0 represents a completely certain distribution. Now we will prove the relationship between such a distribution and the conditioned probability.

Theorem 1 If a possibility distribution Π is completely certain, then its conditioned probability \hat{P} is completely certain.

Proof Π is completely certain if $\exists k$ such that $\pi_k = 1$, and $\pi_i = 0, \forall i \neq k$. To obtain the conditioned probability we first calculate K using Eq. (1):

$$K = p_k \pi_k + \sum_{i \neq k} p_i \pi_i = p_k \times 1 + \sum_{i \neq k} p_i \times 0 = p_k$$

So now we find the conditioned probabilities

$$\hat{p}_k = p_k \pi_k / K = p_k \times 1 / p_k = 1$$

$$\hat{p}_i = p_i \pi_i / K = p_i \times 0 / p_k = 0, i \neq k$$

Thus the conditioned probability distribution \hat{P} is

$$\hat{P} = \{0, \dots, \hat{p}_k = 1, \dots, 0\}$$

which is a completely certain probability distribution. \square

Some of the issues relative to the interpretation of this result with respect to consistency and conflict will be discussed in Sect. 4.

2.2 Case 2: complete uncertainty

If there is no distinction that is made on the values of the variable V by the possibility distribution, we say this implies

complete uncertainty. This is then represented in the distribution by all values equaling 1 as shown in Table 1.

Theorem 2 If a possibility distribution Π is completely uncertain, then its conditioned probability \hat{P} is identical to the original probability P .

Proof Π is completely uncertain if $\forall i, \pi_i = 1$. To obtain the conditioned probability we first calculate K using Eq. (1)

$$K = \sum_{i=1}^n p_i \pi_i = \sum_{i \neq k} p_i \times 1 = 1$$

since $\sum_{i=1}^n p_i = 1$ for any probability distribution. So now we find the conditioned probabilities

$$\hat{P}_i = p_i * \pi_i / K = p_i * 1 / 1 = p_i$$

Thus the conditioned probability distribution \hat{P} is

$$\hat{P} = \{p_1, p_2, \dots, p_n\} = P,$$

which is the original probability distribution. \square

The interpretation of this result is that the possibility distribution shows no preference for any specific value and so the default is that the information to be used in a decision should be that represented by the original probability distribution. So for the two extreme probability cases (Table 1) we have respectively:

$$(a) \quad P_{cc} := \hat{P}\{0, 0, \dots, 0, p_k = 1, \dots, 0\}$$

$$(b) \quad P_{cu} : \hat{P} = \left\{ \frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n} \right\}$$

Clearly these are valid as for both distributions: $\sum_{i=1}^n \hat{P}_i = 1$.

2.3 Case 3: intermediate uncertainty

Here we examine the case that falls between complete certainty and complete uncertainty for a possibility distribution. To represent this we allow m values of $\Pi = 1$, such that $1 < m < n$. For convenience we index these values from $i = 1$, so we have for the distribution:

$$\Pi = \{1, 1, \dots, 1, 0, 0 \dots, 0\} : \pi_i = 1;$$

$$i = 1 \dots m, \pi_j = 0; j = m + 1 \dots n$$

Then clearly $K = p_1 + p_2 + \dots + p_m$ and

$$\hat{p}_i = p_i \times 1 / (p_1 + p_2 + \dots + p_m);$$

$$i = 1 \dots m; \hat{p}_{m+1} = \dots \hat{p}_n = 0$$

In order to understand what happens in this intermediate uncertainty situation, we will examine the two extreme

probability distributions being conditioned by this possibility. First for P_{cc} we have to consider two subcases.

2.3.1 P_{cc} Subcase (1); $p_t = 1$; $t \leq m$

$$K = 0 \times 1 + 0 \times 1 + \dots + (p_t = 1) \times (\pi_t = 1) + \dots \times 0 \times 0 = 1$$

$$\hat{p}_t = \frac{1 \times 1}{K} = \frac{1}{1} = 1; \hat{p}_{j \neq t} = 0$$

$$\text{So } \hat{P} = \{0, 0, \dots, 0, \hat{p}_t = 1, \dots, 0\} = P_{cc}$$

2.3.2 P_{cc} Subcase (2); $p_t = 1$; $t > m$

For this subcase, however, there is a problem since $\pi_t = 0$, but $p_t = 1$. This case will be discussed further in Sect. 4.

2.3.3 P_{cu} Complete uncertainty

Next for the completely uncertain probability distribution P_{cu} we find

$$K = \sum_{i=1}^m p_i \pi_i + \sum_{i=m+1}^n p_i \pi_i = m \times \left(\frac{1}{n}\right) \times 1 + (n - (m + 1)) \times \left(\frac{1}{n}\right) \times 0 = \frac{m}{n}$$

Then the conditioned probability values are

$$\hat{p}_i = \frac{1}{n} \times 1 / \left(\frac{m}{n}\right) = \frac{1}{m}; i = 1 \dots m$$

$$\hat{p}_{m+1} = \dots \hat{p}_n = \frac{1}{n} \times 0 / \left(\frac{m}{n}\right) = 0$$

$$\hat{P} = \left\{ \frac{1}{m}, \frac{1}{m}, \dots, \frac{1}{m}, \hat{p}_{m+1} = 0, 0, \dots, 0 \right\}$$

Therefore, we have obtained a subset of equally distributed conditioned probabilities corresponding to the possibilities that are 1. Note that these equally distributed probabilities are greater than the $\frac{1}{n}$ values for the initial P_{cu} . Again, this is clearly a valid distribution as $\sum_{i=1}^n \hat{p}_i = m \times \left(\frac{1}{m}\right) = (n - (m + 1)) \times 0 = 1$

2.4 Case 4: Generalized possibility distribution

This is a general case for which we index $\pi_1 = 1$ and to capture the situation between complete certainty and uncertainty we use the weights, $0 < w_i < 1$, for the $n - 1$ arbitrary possibility values. So from Table 1 this possibility distribution is:

$$\Pi = \{1, w_2, w_3, \dots, w_n\}$$

and for the conditioned probabilities we obtain

$$K = p_1 \times 1 + \sum_{i=2}^n p_i w_i = p_1 + K', \text{ where } K' = \sum_{i=2}^n p_i w_i$$

$$\hat{p}_1 = p_1 \times 1 + (p_1 + K');$$

$$\hat{p}_i = p_i \times w_i + (p_1 + K'); i = 2 \dots n$$

Again we will examine the conditioning of the extreme probabilities and once more have to consider the subcases of P_{cc}

2.4.1 P_{cc} subcase (1); $t = 1$, $p_1 = 1$

$$K' = \sum_{i=2}^n 0 \times w_i = 0;$$

$$\hat{p}_1 = 1 \times 1 / (1 + 0) = 1; \hat{p}_i = 0 \times w_i / (1 + 0) = 0; i = 2 \dots n$$

$$\hat{P} = \{1, 0, \dots, 0\} = P_{cc}$$

2.4.2 P_{cc} subcase (2); $t > 1$, $p_t = 1$

We find the conditioned probability here as:

$$K = 0 \times 1 + p_t \times w_t + \sum_{i=2, i \neq t}^n 0 \times w_i = w_t$$

$$\hat{p}_1 = 0 \times 1 / w_t = 0; \hat{p}_t = p_t \times w_t / w_t = 1 \times w_t / w_t = 1$$

$$\hat{p}_i = 0 \times w_i / w_t = 0; i = 2 \dots n, i \neq t$$

$$\hat{P} = \{0, 0, \dots, \hat{p}_t = 1, \dots, 0\} = P_{cc}$$

2.4.3 P_{cu} complete uncertainty

Finally for the completely uncertain probability P_{cu}

$$K = \frac{1}{n} + \sum_{i=2}^n 1/n \times w_i = \left(1 + \sum_{i=2}^n w_i\right) / n$$

$$\hat{p}_1 = \frac{1}{n} \times 1 / \left(1 + \sum_{i=2}^n w_i\right) / n = 1 / \left(1 + \sum_{i=2}^n w_i\right)$$

$$\hat{p}_i = \frac{1}{n} \times w_i / \left(1 + \sum_{i=2}^n w_i\right) / n = w_i / \left(1 + \sum_{i=2}^n w_i\right) \quad i = 2 \dots n$$

Here we can see since $0 < w_i < 1$, then $\hat{p}_1 < 1$ and $\hat{p}_i < \hat{p}_1$. Also these conditioned probabilities still sum to 1.

$$\begin{aligned} \sum_{i=2}^n \hat{p}_i &= 1 / \left(1 + \sum_{i=2}^n w_i\right) + \sum_{i=2}^n w_i / \left(1 + \sum_{i=2}^n w_i\right) \\ &= \left(1 + \sum_{i=2}^n w_i\right) / \left(1 + \sum_{i=2}^n w_i\right) = 1 \end{aligned}$$

To consider specific cases for the weights here we look at an equal distribution of the weights values. In a sense this

is a default choice. After deciding on which possibility value to choose as 1, then if there is no preference for the other values, a default of equal values is reasonable. We first consider that since there are $n-1$ weights to be assigned we use the weight values as

$$w_1 = 1 \text{ and } w_i = \frac{1}{n-1}, i = 2, \dots, n.$$

First

$$\left(1 + \sum_{i=2}^n w_i\right) = \left(1 + (n-1) \times \frac{1}{n-1}\right) = 2$$

Then

$$\begin{aligned} \hat{p}_1 &= 1 / \left(1 + \sum_{i=2}^n w_i\right) = \frac{1}{2}; \text{ and } \hat{p}_i = w_i / \left(1 + \sum_{i=2}^n w_i\right) \\ &= \left(\frac{1}{n-1}\right) / 2 = \frac{1}{2(n-1)} \end{aligned}$$

So $\hat{P} = \left\{\frac{1}{2}, \frac{1}{2(n-1)}, \frac{1}{2(n-1)}, \dots, \frac{1}{2(n-1)}\right\}$ and clearly this is a valid distribution as

$$\begin{aligned} \sum_{i=1}^n \hat{p}_i &= \frac{1}{2} \sum_{i=2}^n \frac{1}{2(n-1)} = \frac{1}{2} + (n-1) \frac{1}{2(n-1)} = \frac{1}{2} + \frac{1}{2} \\ &= 1 \end{aligned}$$

3 Information measures

In this section we will consider measures that can be used to evaluate a conditioned probability distribution relative to the original probability. 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 (Klir 2006; Xu and Erdogmus 2010). Thus, we will also examine other measures, such as the Gini index and Renyi entropy, in this section.

3.1 Shannon entropy

Shannon entropy has been the most broadly applied measure of randomness or information content (Shannon 1948). For a probability distribution $P = \{p_1, p_2, \dots, p_n\}$ as was discussed previously in Eq. (2), $S(P) = -\sum_{i=1}^n p_i \ln(p_i)$. The well-known minimum and maximum values for the Shannon entropy are presented in the context of our two extreme probability cases.

First for complete certainty, P_{cc} , we recall that here, for some t , $p_t = 1$, and so

$$\begin{aligned} S(P_{cc}) &= -\left(1 \ln(1) + \sum_{i=1, \neq t}^n 0 \ln(0)\right) \\ &= 1 \times 0 + \sum_{i=1, \neq t}^n 0 = 0 \end{aligned}$$

Note this follows as $\lim_{p \rightarrow 0^+} p \ln p = 0$. That is when a probability distribution represents complete certainty, then we have no uncertainty, i.e. maximum information.

Then for the case of complete uncertainty represented by the equi-probable distribution, P_{cu} , where $\forall i \ p_i = 1/n$

$$\begin{aligned} S(P_{cu}) &= -\sum_{i=1}^n \frac{1}{n} \times \ln\left(\frac{1}{n}\right) \\ &= -\frac{1}{n} \sum_{i=1}^n (\ln(1) - \ln(n)) \\ &= -n \times \frac{1}{n} (0 - \ln(n)) = \ln(n). \end{aligned}$$

That is, when all probabilities are equi-probable, this is the most unpredictable, uncertain situation and so represents the minimum information. In summary the range of Shannon's entropy for a given probability distribution is:

$$0 \leq S(P) \leq \ln(n) \quad (4)$$

3.2 Gini Index

The Gini index, $G(P)$, also known as the Gini coefficient, is a measure of statistical dispersion developed by Gini (1912), and is defined as

$$G(P) \equiv 1 - \sum_{i=1}^n p_i^2 \quad (5)$$

Some practitioners use $G(P)$ versus $S(P)$ since it does not involve a logarithm, making analytic solutions simpler. Gini index is used in consideration of inequalities in various areas such as economics, ecology and engineering (Aristondo et al. 2012). A very important application of the Gini index is as a splitting criterion for decision tree induction in machine learning and data mining (Breiman et al. 1984).

It is accepted in practice for diagnostic test selection that the Shannon and Gini measures are interchangeable (Sent and van de Gaag 2007). The specific relationship of Shannon entropy and the Gini index has been discussed in the literature (Eliazar and Sokolov 2010). Theoretical support for this practice is provided in Yager's independent consideration of alternative measures of entropy (Yager 1995), where he derives the same form for an entropy measure as the Gini measure.

Now as done for Shannon entropy, we consider the maximum and minimum values for $G(P)$. Letting $R = \sum_{i=1}^n p_i^2$, then since $0 \leq p_i \leq 1$ ($0 \leq p_i^2 \leq 1$) and at least one $p_i > 0$, $0 < R \leq 1$. $R = 1$ only if for some t , $p_t = 1$. Thus, $G(P) > 0$ unless $p_t = 1$ where $G(P) = 0$. This is the case for the distribution P_{cc} , since $p_t = 1$, $p_i = 0$, $i \neq t$. Specifically

$$G(P_{cc}) = 1 - (p_t^2 + \sum_{i \neq t} p_i^2) = 1 - (1^2 + 0) = 0$$

As for the Shannon entropy this corresponds to no uncertainty and has the same value of 0.

Next we examine the index for the equi-probable distribution, P_{cu} , where $p_i = 1/n$ for all i .

$$G(P_{cu}) = 1 - \sum_{i=1}^n \frac{1}{(n)^2} = 1 - n \frac{1}{(n^2)} = 1 - \frac{1}{n} = \frac{n-1}{n}$$

Consider the behavior of $G(P_{cu})$ as n increases. For $n = 2$ ($p_i = \frac{1}{2}$; $p_j = \frac{1}{2}$)

$$G(P_{cu}) = \frac{2-1}{2} = 1/2$$

Then for $n = 10$ ($p_i = 0.1, \dots, p_j = 0.1$) we have

$$G(P_{cu}) = \frac{10-1}{10} = 9/10$$

Clearly then for $n \rightarrow \infty$, $G(P) \rightarrow 1$. Thus, in the case of an equiprobable distribution, we have increasing values for $G(P_{cu})$ with n , and in general the range for $G(P)$ is

$$0 \leq G(P) \leq \frac{n-1}{n} < 1 \quad (6)$$

Now we can use this measure for evaluating our IEDs' example and compare $G(P)$ for the original and the conditioned probability distributions. First

$$G(P_{IED}) = 1 - (0.3^2 + 0.2^2 + 0.4^2 + 0.1^2) = 1 - 0.3 = 0.7$$

So for \hat{P}

$$\begin{aligned} G(\hat{P}_{IED}) &= 1 - (0.39^2 + 0.16^2 + 0.42^2 + 0.03^2) \\ &= 1 - 0.354 = 0.646 \end{aligned}$$

Thus we see that as with the Shannon measure result (3), based on the Gini index \hat{P}_{IED} appears again to be more informative than P_{IED} .

3.3 Application of measures to the four cases

In this section we apply the Shannon and Gini measures to the original and conditioned probability distributions for the four possibility distribution cases of Sect. 2.2 and

compare the measures' values. As both measures have increasing values with increasing uncertainty, the conditioned probability will be more informative for decision-making if its measure value is less than for the original probability. We shall see that both measures basically agree for the cases considered although their specific values are in different ranges.

3.3.1 Case 1

For the completely certain possibility, we consider only where there is no conflict and the conditioned probability is

$$\hat{P} = \{1, 0, \dots, 0\}$$

Then we have first for both measures with the distribution P_{cc}

$$S(\hat{P}) = G(\hat{P}) = 0 = S(P_{cc}) = G(P_{cc})$$

But for the equi-probable initial distribution P_{cu}

$$S(P_{cu}) = \ln(n) > S(\hat{P}) = 0$$

$$G(P_{cu}) = \frac{n-1}{n} > G(\hat{P}) = 0$$

So the conditioned probability distribution is more informative in the second case for the probability P_{cu} .

3.3.2 Case 2

Next for the case of complete possibilistic uncertainty, we had $\hat{P} = P$ for all the probability distributions and so we have

$$S(\hat{P}) = S(P) \text{ and } G(\hat{P}) = G(P)$$

We can conclude that the conditioned probability distribution \hat{P} is no more informative than the original probability P since the possibility distribution Π does not contribute any information as it represents complete uncertainty.

3.3.3 Case 3

Recall this is the intermediate possibility case and here we consider the probability, P_{cc} , first for the Shannon measure and then the Gini index. Since for no conflict $\hat{P} = \{0, 0, \dots, 0, p_t = 1, \dots, 0\}$ then as before for this distribution

$$S(\hat{P}) = S(P_{cc}) = 0 \text{ and } G(\hat{P}) = G(P_{cc}) = 0$$

Next for the equi-probable distribution P_{cu} , the Shannon measure is

$$\begin{aligned}
S(\hat{P}) &= - \left(\sum_{i=1}^m \left(\frac{1}{m} \right) \ln \left(\frac{1}{m} \right) + \sum_{i=m+1}^n 0 \ln(0) \right) \\
&= - \frac{1}{m} \sum_{i=1}^m (\ln(1) - \ln(m)) \\
&= - \frac{1}{m} \times (-m \ln(m)) = \ln(m)
\end{aligned}$$

Now since P_{cu} is an equi-probable distribution and $n > m$
 $S(P_{cu}) = \ln(n) > \ln(m) = S(\hat{P})$

Next for the Gini measure

$$\begin{aligned}
G(\hat{P}) &= 1 - \left(\sum_{i=1}^m \hat{p}_i^2 + \sum_{i=m+1}^n \hat{p}_i^2 \right) \\
&= 1 - \left(\sum_{i=1}^m \left(\frac{1}{m^2} \right) + \sum_{i=m+1}^n 0 \right) = 1 - m \times \frac{1}{m^2}
\end{aligned}$$

Recall $G(P_{cu}) = 1 - 1/n$ and since $1 < m < n$, $1/n < 1/m$

$$G(\hat{P}) = 1 - \frac{1}{m} < 1 - \frac{1}{n} = G(P_{cu})$$

Thus we see that by both measures the conditioned probability is more informative in this case.

3.3.4 Case 4

This is the case of the generalized possibility distribution in which for P_{cc} we saw that $\hat{P} = P_{cc}$. So again we have
 $S(\hat{P}) = G(\hat{P}) = 0 = S(P_{cc}) = G(P_{cc})$

Next for the other probability distribution, P_{cu} , we had obtained for \hat{P} a general expression in terms of the weights w_i . Here we will consider the special case we examined for the equal distribution of the weights where we had for \hat{P}

$$\hat{P} = \left\{ \frac{1}{2}, \frac{1}{2(n-1)}, \frac{1}{2(n-1)}, \dots, \frac{1}{2(n-1)} \right\}$$

Now we can apply our measures to this conditioned distribution. First for $S(\hat{P})$:

$$\begin{aligned}
S(\hat{P}) &= - \left[\frac{1}{2} \ln \left(\frac{1}{2} \right) + \sum_{i=2}^n \frac{1}{2(n-1)} \ln \left(\frac{1}{2(n-1)} \right) \right] \\
&= \left[\frac{1}{2} (\ln 1 - \ln 2) + (n-1) \frac{1}{2(n-1)} (\ln 1 - \ln(2(n-1))) \right] \\
&= - \left[-\frac{\ln 2}{2} - \frac{1}{2} \ln(2(n-1)) \right] \\
&= \frac{1}{2} [\ln 2 + \ln(2(n-1))]
\end{aligned}$$

$$\begin{aligned}
\text{For } n = 2, S(\hat{P}) &= \frac{1}{2} [\ln 2 + \ln(2(1))] = \ln 2 \\
&= S(P), \text{ but for } n > 2
\end{aligned}$$

$$S(\hat{P}) < \ln n = S(P)$$

Next for the Gini index:

$$\begin{aligned}
G(\hat{P}) &= 1 - \left(\left(\frac{1}{2} \right)^2 + \sum_{i=1}^n \left(\frac{1}{2(n-1)} \right)^2 \right) \\
&= 1 - \left(\frac{1}{4} + (n-1) \times \left(\frac{1}{2(n-1)} \right)^2 \right) \\
&= 1 - \left(\frac{1}{4} + \frac{1}{4(n-1)} \right) \\
&= \frac{3n-4}{3n-4}
\end{aligned}$$

Similar to the Shannon measure for $n = 2$, the Gini measure is the same for P and \hat{P}

$$G(\hat{P}) = 1 - \frac{2}{4} = \frac{1}{2} \quad \text{and} \quad G(P) = 1 - \frac{1}{n} = \frac{1}{2}$$

Finally for $n > 2$, $G(\hat{P}) = \frac{3n-4}{4n-4} < G(P) = \frac{n-1}{n}$ since for $n = 3$, $G(\hat{P}) = \frac{5}{8} < G(P) = \frac{2}{3}$, and as $n \rightarrow \infty$, $G(\hat{P}) \rightarrow \frac{3}{4}$, but $G(P) \rightarrow 1$.

To consider this last case more generally, we examine the effects of the range of equi-distributed weights. When $w_i \rightarrow 0$, $1 < i \leq n$, $\Pi \rightarrow \{1, 0, \dots, 0\}$, complete certainty, and we recall for which case we have seen that the conditioned distribution is more informative. Then if $w_i \rightarrow 1$, $1 < i \leq n$, $\Pi \rightarrow \{1, 1, \dots, 1\}$, the case of complete uncertainty. So the conditioned probability distribution \hat{P} is no more informative than the original probability P for either measure.

3.4 Renyi entropy

Renyi (1961,1970) introduced a parameterized family of entropies as a generalization of Shannon entropy. The intention was to have the most general approach that preserved the additivity property and satisfied the probability axioms of Kolmogorov. Renyi entropy is

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

3.5 Cases of the parameter α

$\alpha = 0$: $S(P) = \ln|P|$ —Hartley Entropy (Hartley 1928)

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

$$\begin{aligned}\alpha = 2 : S_2(P) \\ = -\ln\left(\sum_{i=1}^n p_i^2\right) - \text{Collision or quadratic entropy}\end{aligned}\quad (7)$$

$$\alpha \rightarrow \infty : S_\infty(P) = \text{Min}_{i=1}^n (-\ln p_i) = -\text{Max}_{i=1}^n (\ln p_i) \quad (8)$$

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

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

The Hartley entropy is not of great interest here as for all of our cases here $|P| = |\hat{P}|$, and we have already considered the Shannon entropy. Now we consider the values of the S_2 measure, Eq. 7, for our two characteristic probabilities. For P_{cc}

$$\begin{aligned}S_2(P_{cc}) &= -\ln\left(p_1 + \sum_{i=2}^n p_i^2\right) = -\ln\left(1 + \sum_{i=2}^n 0^2\right) \\ &= -\ln 1 = 0\end{aligned}$$

and for

$$\begin{aligned}S_2(P_{cu}) &= -\ln\left(\sum_{i=1}^n p_i^2\right) = -\ln\left(\sum_{i=1}^n \frac{1}{n^2}\right) = -\ln\left(n \times \frac{1}{n^2}\right) \\ &= -(\ln 1 - \ln(n)) = \ln(n)\end{aligned}$$

These are the same as the results for the Shannon entropy. To continue we calculate S_2 for our IED example as we have done for the Shannon entropy and Gini index. So we have

$$\begin{aligned}S_2(P_{IED}) &= -\ln(0.3^2 + 0.2^2 + 0.4^2 + 0.1^2) = -\ln(0.3) \\ &= 1.20\end{aligned}$$

and for the conditioned probability

$$\begin{aligned}S_2(\hat{P}_{IED}) &= -\ln(0.39^2 + 0.16^2 + 0.42^2 + 0.03^2) \\ &= -\ln(0.356) = 1.03\end{aligned}$$

Again as for the other two measures, the resulting value for \hat{P} is less than for P . We want to consider only briefly the effect of larger values for the parameter α . For example from Eq. 8 for S_∞ we have

$$S_\infty(P_{IED}) = 0.92 > S_\infty(\hat{P}_{IED}) = 0.87$$

This continues the evaluation of \hat{P}_{IED} being more informative but we note the difference is somewhat smaller.

Next we can utilize the already determined sum of the squared probabilities from the Gini measure to evaluate for the first three possibility cases for S_2 .

3.5.1 Case 1

For the probability distribution P_{cc} we see

$$S_2(\hat{P}) = -\ln\left(1 + \sum_{i=2}^n 0^2\right) = -\ln 1 = 0 = S_2(P_{cc})$$

but for P_{cu}

$$S_2(\hat{P}) = 0 < S_2(P_{cu}) = \ln(n)$$

3.5.2 Case 2

Since $\hat{P} = P$,

$$S_2(\hat{P}) = 0 = S_2(P_{cc}) \text{ and } S_2(\hat{P}) = \ln(n) = S_2(P_{cu})$$

3.5.3 Case 3

For the completely certain probability as before,

$$S_2(\hat{P}) = 0 = S_2(P_{cc})$$

and for P_{cu}

$$\begin{aligned}S_2(\hat{P}) &= -\ln\left(\sum_{i=1}^m \hat{p}_i^2 + \sum_{i=m+1}^n \hat{p}_i^2\right) \\ &= -\ln\left(\sum_{i=1}^m \left(\frac{1}{m}\right)^2 + \sum_{i=m+1}^n 0\right) \ln(m)\end{aligned}$$

Again since $m < n$,

$$S_2(\hat{P}) = \ln(m) < \ln(n) = S_2(P_{cu}).$$

So we note that values of S_2 in these specific cases are the same as for the Shannon entropy measure; however, the exact numeric value obtained for example in Eq. 3 is not identical so we can conclude there is a close but not exact relationship between them.

3.6 Example: less informative conditioned probability

Next consider the following example for possibility and probability distributions in order to illustrate that not all \hat{P} 's are more informative than an initial probability P . We shall apply our previous information measures and see that these are consistent in their assessments. So let the possibility and probability distributions be:

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

As before we can compute \hat{P}

$$\begin{aligned}K &= 0.8 \times 0.1 + 0.1 \times 0.1 + 1.0 \times 0.05 + 0.1 \times 0.05 \\ &= 0.08 + 0.01 + 0.05 + 0.005 = 0.145\end{aligned}$$

and

$$\hat{p}_1 = 0.8 \times 0.1/0.145 = 0.552; \dots \hat{p}_4 = 0.1 \times 0.05/0.145 = 0.034$$

$$\hat{P} = \{0.552, 0.069, 0.345, 0.034\}$$

We can see intuitively that there is some degree of conflict or lack of consistency between the possibility and probability distributions. For example for the largest probability, $p_1 = 0.8$, the possibility is quite low, $\pi_1 = 0.1$. Then where $\pi_3 = 1.0$, we observe the corresponding probability, p_3 , is one of two lowest probability values, 0.05.

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

$$S(P) = 0.708 < S(\hat{P}) = 0.995$$

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

$$G(P) = 1 - 0.655 = 0.345 < G(\hat{P}) = 1 - 0.43 = 0.57$$

Finally we obtain similar results for the Renyi entropies, S_2 and S_∞ :

$$S_2(P) = -\ln(0.655) = 0.408 < S_2(\hat{P}) = -\ln(0.430) = 0.844$$

$$S_\infty(P) = 0.22 < S_\infty(\hat{P}) = 0.59$$

So the information measures are compatible with our intuitive assessment of the conflict between between Π and P . In the next section when we discuss in some detail Zadeh's consistency measure, we can observe this measure is also indicative of a lower consistency with these distributions.

4 Consistency evaluations of distributions

In this section, we use Zadeh's consistency measure as another approach to assess the integration of uncertainty representations as a supplement to the information measures of the previous section. We shall see that the measure yields evaluations compatible with the information measures of the previous section.

As noted by Sudkamp (1992), a probability–possibility transformation is a “purely mechanical manipulation of the distribution without regard to the underlying problem domain or evidence”. It does not by itself provide guidance of the usefulness of the outcome.

For example, reconsider the result of Theorem 1 with respect to the initial probability distribution. Let p_k be a very low probability; i.e. represent a “rare” event, $0 < p_k \ll 1$;

however, as we have seen $\hat{p}_k = 1$, which indicates that although the probability was very small, the corresponding event did actually occur in this particular instance based on the possibility distribution. Furthermore, if the initial probability p_k was actually 0, then $K = 0$ and the conditional probability is ill defined as we have an indeterminate result: $0/0$. Clearly, these results by themselves are unhelpful to decision making and we will see that the consistency measure reflects this.

There have been a number of approaches to consistency measures of probability and possibility distributions that have been proposed (Delgado and Moral 1987; Gupta 1993). As we discussed in the introduction, Zadeh's approach,

$$C_Z(\Pi, P) = \sum_{i=1}^n \pi_i \times p_i \quad (9)$$

is identical to the expression for K in the conditioned probability approach. 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 (1982, 1983). Here for every subset A of the space X ,

$$C_{DP}(\Pi, P) = 1 \text{ if } \Pi(A) \geq P(A) \quad (10)$$

and is 0 otherwise. This definition is based on the idea that possibility is the weaker representation for a situation than probability.

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 possibilistic conditioning approach. We can note that the maximum value that $C_Z(\Pi, P) = \sum_{i=1}^n \pi_i \times p_i$ 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 $[0 \dots 1]$ 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. Conflict generally is thought of as involving broader semantic issues such source reliability and trustworthiness.

Thus, for the case in Theorem 1 where $p_k = 0$ when $\pi_k = 1$, evaluation of C_Z yields

$$\begin{aligned} C_Z(\Pi, P) &= \pi_k \times p_k + \sum_{i \neq k}^n \pi_i \times p_i \\ &= 1 \times 0 + \sum_{i \neq k}^n 0 \times p_i \\ &= 0 \end{aligned}$$

This result implies that these distributions are indeed inconsistent, i.e. a total conflict, and we should not expect a

valid conditional probability distribution to be produced for such a situation. Resolution of such a conflict can be managed by considerations of semantic issues such as the reliability of the underlying information sources.

4.1 Zadeh's consistency measure, C_Z , for four possibility cases

We will consider comparative evaluation of the consistency cases in a following section some of which show a conflict. Note that for validity we can check to see that the conditioned distribution does indeed sum to 1 where there is no conflict.

Case 1:

For $\Pi(1,0,\dots,0)$;

$$C_Z(\Pi, P_{cc}) = 1 \times 1 + \sum_{i=2}^n 0 \times 0 = 1;$$

$$C_Z(\Pi, P_{cu}) = 1 \times \frac{1}{n} + \sum_{i=2}^n 0 \times \frac{1}{n} = \frac{1}{n}$$

For P_{cc} , this result shows complete consistency that since only one probability, p_1 , was considered as possible. For P_{cu} , the measure indicates that there is some inconsistency with P_{cu} .

Case 2:

For case 2, $\Pi(1, 1, \dots, 1)$, which is complete uncertainty, no distinctions are made relative to the probabilities and so both P_{cc} and P_{cu} are consistent with the possibility distribution.

$$C_Z(\Pi, P_{cc}) = 1 \times 1 + \sum_{i=2}^n 1 \times 0 = 1;$$

$$C_Z(\Pi, P_{cu}) = \sum_{i=1}^n 1 \times \frac{1}{n} = n \times \frac{1}{n} = 1.$$

Case 3, P_{cc} subcase 1; $p_t = 1$; $t \leq m$

With the intermediate possibility case 3 for P_{cc} where $p_t = 1$ and $t \leq m$ we have

$$C_Z(\Pi, P_{cc}) = 1 \times 1 + \sum_{i=2}^m 1 \times 0 + \sum_{i=m+1}^n 0 \times 0 = 1;$$

Case 3, P_{cc} subcase 2; $p_t = 1$; $t > m$

As noted in Sect. 2, there is a problem since $\pi_t = 0$, but $p_t = 1$. As a result, we have zero for the consistency measure,

$$C_Z(\Pi, P) = \sum_{i=1}^m 1 \times 0 + (p_t = 1) \times (\pi_t = 0) + 0 \times 0 = 0.$$

This result implies that these distributions are completely inconsistent or in conflict. Thus, no valid conditional probability distribution can be produced for such a situation.

Case 3, P_{cu} Complete uncertainty

$$C_Z(\Pi, P_{cu}) = \sum_{i=1}^m 1 \times \frac{1}{n} + \frac{n}{i=m+1} 0 \times \frac{1}{n} = \frac{m}{n} < 1$$

Similar to subcase 1 for P_{cu} , $n - m$ of the original probabilities are not compatible with the possibility distribution as reflected in the consistency measure. That is the inconsistency here is due to the contrast in the $n-m$ values of Π and P_{cu} .

Case 4, P_{cc} Subcase 1; $t = 1$, $p_t = 1$:

Finally for the general possibility case, Case 4, where $P = \{1, 0, \dots, 0\}$,

$$C_Z = 1 \times 1 + \sum_{i=2}^n w_i \times 0 = 1.$$

Case 4, P_{cc} Subcase 2; $t > 1$, $p_t = 1$:

Since $0 \leq w_i < 1$, we know not all probabilities are fully represented. Here, all $w_i > 0$ so we do not have a conflict as in Case 3, Subcase 2 above since $C_Z(\Pi, P) = w_t \times p_t = w_t$; $0 < w_t < 1$

Case 4, P_{cu} complete uncertainty

$$C_Z = 1 \times \frac{1}{n} + \sum_{i=2}^n w_i \times \frac{1}{n} < \frac{1}{n} + \frac{n-1}{n} = 1$$

4.2 Consistency for example distributions

Next let us consider the consistency for the IED example. For these distributions, if we recall the value of K then we have

$$C_Z(\Pi_{IED}, P_{IED}) = K = 0.3 + 0.12 + 0.32 + 0.02 = 0.76.$$

We have seen that the values of each of the three information measures we have evaluated for \hat{P}_{IED} are less than their values for P_{IED} . At issue is how specific values for consistency are related to the information measure values. Relative to the range of C_Z , 0.76 is reasonably large. We can see next how this consistency value compares to the example of distributions for which the information assessments were shown to be less informative.

Consider again the possibility and probability distributions of Sect. 3.5 above. For these we observed that all the information measures indicated the conditioned probability \hat{P} was less informative than the initial probability P . For these distributions the consistency measure is

$$C_Z(\Pi, P) = K = 0.08 + 0.01 + 0.05 + 0.005 = 0.145$$

Clearly this consistency value is quite low compared to $C_Z(\Pi_{IED}, P_{IED})$. So we can observe that higher consistency values are generally correlated with more informative conditioned probabilities.

Situations like this can occur in many applications. For example with web assistant agents, uncertainty aggregation

appears in the integration of information from sources such as user profiles, proximity-based fuzzy clustering and knowledge-based discovery (Loia et al. 2006).

5 Summary

Decision makers are constantly faced with making choices in complex situations for which they have imperfect and often conflicting information. They have difficult decisions in making effective use of this information. Typically such a mix of information has a variety of associated uncertainty, but ultimately the decision maker must come to specific conclusions or actions based on this. Our research here has developed preliminary approaches to assist in this process by providing information theory based quantitative evaluations to guide decisions.

We have developed exact expressions for conditioned probability based on the extreme cases, completely certain and uncertain. For these cases three information measures were applied and yielded compatible results for comparing the informativeness of the original versus the conditioned probability. As well, we carried out the possibilistic conditioning and information evaluations for numeric examples. Additionally we used the Zadeh consistency measure and have seen it correlates well with the evaluation results.

We are currently doing research on the aggregations of both multiple possibility distributions and probability distributions. This will allow us to potentially take advantage of such additional information sources before the computing the conditioned probability. Also we are developing environments to carry out Monte Carlo simulations to test the conditioning approach and the 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 conditioned probability is more informative.

Acknowledgments We would like to thank the Naval Research Laboratory's Base Program, Program Element No. 0602435 N for sponsoring this research. Ronald Yager has been in part supported by ARO MURI grant Number W911NF-09-1-0392.

References

- Acampora G, Loia V (2008) A proposal of ubiquitous fuzzy computing for Ambient Intelligence. *Inf Sci* 178:631–646. doi:10.1016/j.ins.2007.08.023
- Aristondo O, Garcia-Lparesta J, de la Vega C, Pereira R (2012) The Gini Index, the dual decomposition of aggregation functions and

- the consistent measurement of inequality. *Int J Intell Syst* 27:132–152. doi:10.1002/int.21517
- Benigni M, Furrer R (2012) Spatio-temporal improvised explosive device monitoring: improving detection to minimise attacks. *J Appl Stat* 39:2493–2508. doi:10.1080/02664763.2012.719222
- Breiman L, Friedman J, Olshen R, Stone C (1984) Classification and regression trees. Wadsworth & Brooks/Cole, Monterey
- Delgado M, Moral S (1987) On the concept of possibility-probability consistency. *Fuzzy Sets Syst* 21:311–318
- Dubois D, Prade H (1982) On several representations of an uncertain body of evidence. In: Gupta M, Sanchez E (eds) *Fuzzy information and decision processes*. North Holland, Amsterdam, pp 167–182
- Dubois D, Prade H (1983) Unfair coins and necessity measures: towards a possibilistic interpretations of histograms. *Fuzzy Sets Syst* 10:15–20
- Eliazar I, Sokolov I (2010) Maximization of statistical heterogeneity: from Shannon's Entropy to Gini's Index. *Phys A* 389:3023–3038. doi:10.1016/j.physa.2010.03.045
- Gini C (1912) Variabilita e mutabilita (Variability and Mutability). *Tipografia di Paolo Cuppini*, Bologna, Italy
- Gupta C (1993) A note on the transformation of possibilistic information into probabilistic information for investment decisions. *Fuzzy Sets Syst* 56:175–182
- Hartley R (1928) Transmission of information. *Bell Syst Tech J* 7:535–563
- Klir G (2006) *Uncertainty and information*. Wiley, Hoboken
- Loia V, Pedrycz W, Senatore S, Sessa M (2006) Web navigation support by means of cognitive proximity-driven assistant agents. *J Am Soc Inf Sci Technol* 57:512–527. doi:10.1002/asi.20306
- Parsons S (2001) *Qualitative methods for reasoning under uncertainty*. MIT Press, Cambridge
- Pedrycz W (2010) Human centricity in computing with fuzzy sets. *J Ambient Intell Humaniz Comput* 1:65–74. doi:10.1007/s12652-009-0008-0
- Renyi A (1961) On measures of information and entropy. In: *Proceedings of the 4th Berkeley symposium on mathematics, statistics and probability 1960*, pp 547–561
- Renyi A (1970) *Probability theory*. North-Holland, Amsterdam
- Reza F (1961) *An introduction to information theory*. McGraw Hill, New York
- Sent D, van de Gaag L (2007) On the behavior of information measures for test selection. In: Carbonell J, Siebarn J (eds) *Lecture Notes in AI 4594*. Springer, Berlin, pp 271–284
- Shannon C (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(379–423):623–656
- Sudkamp T (1992) On probability-possibility transformations. *Fuzzy Sets Syst* 5:73–81
- Xu D, Erdogmus D (2010) Renyi's entropy, divergence and their nonparametric estimators. In: Principe J (ed) *Information theoretic learning: Renyi's entropy and kernel perspectives*. Springer, Berlin, pp 47–102
- Yager R (1995) Measures of entropy and fuzziness related to aggregation operators. *Inf Sci* 82:147–166. SSDI: 0020-0255 (94)00030-F
- Yager R (2012) Conditional approach to possibility-probability fusion. *IEEE Trans Fuzzy Syst* 20:46–56. doi:10.1109/TFUZZ.2011.2165847
- Zadeh L (1978) Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst* 1:3–28