On quantification of different facets of uncertainty

Nikhil R. Pal *

Electronics and Communication Science Unit, Indian Statistical Institute, 203, B.T. Road, Calcutta 700 035, India

Abstract

With a brief introduction to three major types of uncertainties, *randomness, nonspecificity and fuzziness* we discuss various attempts to quantify them. We also discuss several attempts to quantify total uncertainty in a system. We then talk about some new facets of uncertainty like, higher-order fuzzy entropy, hybrid entropy and conflict in a body of evidence. In conclusion, we indicate some other aspects of uncertainty that need to be modeled and quantified.

Keywords: Uncertainty; Randomness; Fuzziness; Nonspecificity; Entropy; Conflict

1. Introduction

Consider a simple die throwing experiment. The top face may have any one of the six numbers $X = \{1, 2, \ldots, 6\}$. If the top face is covered and you are asked to guess it, you face a type of uncertainty. If you have played with the die enough number of times, the best way you can express your answer (ignorance) is to state the *p*robability of occurrence of different faces. This type of uncertainty that arises because of randomness in the system is known as probabilistic uncertainty.

Now suppose an artificial vision system analyzes a digital image of the top face of the die and based on the evidence gathered suggests the value of the unknown top face. Let us call the information provided by the vision system, a body of evidence. Suppose the body of evidence assigns a confidence value of 0.67 to the top face being 1, 2 or 3 and a confidence value of 0.33 to the top face being 3, 4 or 5; that is, the evidence *focuses* on the sets $\{1, 2, 3\}$ and $\{3, 4, 5\}$. These sets are the *focal* elements. Clearly, there is uncertainty in this body of evidence. In fact, there are two types of uncertainties: which of the two hypothesis is true (random nature); and given a hypothesis which specific value represents the answer (lack of specification). Note that when the experiment is over there is no uncertainty about the outcome of the experiment. These two types of uncertainty arise usually due to limitations of the evidence gathering system. Systems like this can be modeled by Dempster–Shafer theory of evidence.

Finally, suppose the evidence gathered by the vision system is ambiguous and the system cannot come out with propositions like the previous ones but interprets the top face of the die as, say, *high* (or *low*). Here a third type of uncertainty appears due to linguistic imprecision or vagueness; this is called fuzzy uncertainty (FU). Fuzzy uncertainty differs from probabilistic uncertainty (PU) and Nonspecificity (NS) because it deals with situations where set boundaries are not sharply defined. Probabilistic and nonspecific uncertainties are not due to ambiguity about set-boundaries, but rather, about the belonging of elements or events to crisp sets. Sometimes uncertainty due to fuzziness is associated with probabilistic uncertainty also. For example, in the die experiment the occurrence of a 6 supports the fuzzy event HIGH more than a 3 does, but there is still an element of chance about the outcome of a throw, so the system contains both PU and FU.

The literature is well documented with these three types of uncertainty. The best known measure of uncertainty for a purely probabilistic system is the Shannon's entropy [27]. Several authors [15] have given different expressions for entropy (probabilistic uncertainty in a system). Almost all these definitions are based on some logarithmic functions of probabilities, and most of them reduce to Shannon's entropy under certain conditions. Pal and Pal [23] proposed a definition, which unlike the earlier attempts, is based on an exponential gain function.

Uncertainty in Dempster–Shafer framework has two facets: one arises due to randomness in the system and the other from nonspecificity in the evidence. Uncertainty due to nonspecificity arises if and only if, there is at least one focal set with more than one element. On the other hand, uncertainty due to randomness, arises if and only if, there is more than one focal element. When the evidence is Bayesian, that is, when there is complete lack of nonspecificity, (in other words, when the focal elements are singletons) uncertainty is usually characterized by Shannon's entropy.

Various authors have introduced different measures of uncertainty based on different views [14]. Yager [34, 35], Higashi and Klir [10], Dubois and Prade [2] and Lamata and Moral [19] have proposed measures of nonspecificity. Measures of uncertainty due to randomness, variously named, dissonance [35], confusion [7, 8], Discord [17], Strife [16] exist. Other measures of uncertainty due to randomness have been introduced by Smets [29] and Lamata and Moral [19]. Measures of total uncertainty for the Dempster–Shafer framework have been introduced by Lamata and Moral [19] and Klir and Ramer [17] as the algebraic sum of the measures of uncertainty due to randomness and nonspecificity. Pal et al. [21, 22] also introduced a new measure of total uncertainty.

Measures of fuzziness estimate the average ambiguity in fuzzy sets in some well-defined sense. Several authors have attempted to quantify fuzziness in a fuzzy set. Deluca and Termini [3] used Shannon's function for a two state system with membership values of fuzzy singletons in place of probabilities to define entropy of a fuzzy set. Kaufmann [11] introduced different indices of fuzziness like linear index of fuzziness and quadratic index of fuzziness. Knopfmacher [13] and Loo [18] showed that some earlier measures of fuzziness are special cases of a larger class of measures defined in terms of a set of axioms. Trillas and Riera [30] also proposed a general class of fuzziness measure. Yager [32, 33] associated fuzziness with the lack of distinction between a proposition and its negation and proposed a family of measures of fuzziness. Higashi and Klir [9] extended Yager's concept to a very general class of fuzzy complements. Kosko [12] defined a fuzziness measure as the ratio of the distances between the fuzzy set A, and its nearest and furthest nonfuzzy sets. Pal and Pal [23] used an information function with exponential gain to measure fuzziness in a set. Bhandari and Pal [1] defined a measure of fuzziness as the ratio of fuzzy divergence between A and its nearest and furthest nonfuzzy sets, where fuzzy divergence is an information measure for discrimination between two fuzzy sets. Bhandari and Pal [1] also used Renyi's entropy to define a measure of fuzziness.

These measures of uncertainty do not take into account all aspects of uncertainty. In this paper, with a brief review of some well known measures of uncertainty for each of the three kinds, we discuss some new facets of uncertainty and their quantification.

2. Some well-known measures of uncertainty

2.1. Probabilistic uncertainty

Let $P = \{p_1, \ldots, p_n\}$ be the set of probabilities associated with an n-state system $X = \{x_1, \ldots, x_n\}$ where p_i is the probability of x_i. Although, Shannon [27] axiomatically derived a unique expression for entropy, a measure of probabilistic uncertainty (or information) associated with the system, we can get the following intuitively plausible derivation. Let $\Delta I(p_i)$ be loss of ignorance or gain in information associated with the occurrence of the ith state of the system. Then $\Delta I(p_i)$ should be inversely related to p_i because with increase in probability the loss of ignorance decreases. For example, with the occurrence of a *certain* event we do not gain any information or lose any ignorance. Moreover, if there are two independent events x and y with probabilities p_x and p_y, the total gain in information from the joint occurrence of the two events should be the sum of information gains from the occurrences of the two events individually. In other words, since the probability of joint occurrence of x and y is $p_x.p_y$, we should have $\Delta I(p_x.p_y) = \Delta I(p_x) + \Delta I(p_y)$. These two desirable properties suggest that $\Delta I(p) = -k\log(p)$ is the choice, k is a positive constant. Therefore, entropy, the expected value of the gain in information (or loss of ignorance) associated with the system (assuming $k=1$) is

$$H_l(P) = -\sum_{i=1}^{n} p_i \log(p_i). \tag{1}$$

$H_l(P)$ has been successfully used in many applications in science, technology and economics. Eq. (1) has many interesting properties like:

Pl. 1: $H_l(P)$ is maximum iff $p_i = 1/n \ \forall i$.

Pl. 2: $H_l(P)$ is minimum iff $p_k = 1$ for some k and $p_i = 0,\ i \neq k$.

Pl. 3: Equalization of probability increases entropy. In other words, suppose $P1 = \{p_1, p_2, p_3, \ldots, p_n\}$ and $P2 = \{p_1 - \delta, p_2 + \delta, p_3, \ldots, p_n\}$ such that $p_1 > p_2$, $\delta > 0$ and $p_1 - \delta > p_2 + \delta$. Then $H_l(P1) \leqslant H_l(P2)$.

Viewing each $x_i \in X$ as a symbol, let $p(s_i)$ = probability of a sequence of symbols of length q. Then the *qth-order entropy* and entropy of the source are, respectively, defined as

$$H^{(q)} = -\sum_i p(s_i)\log(p(s_i))$$

and

$$H = \lim_{q\to\infty} \frac{1}{q} H^{(q)}.$$

A source is called a stationary source if for any integer r, $p(x_1/t_1 + r, x_2/t_2 + r, \ldots, x_n/t_n + r) = p(x_1/t_1, x_2/t_2, \ldots, x_n/t_n)$. For a stationary source it can be shown that [5]

Pl. 4: $H^{(k)} > H^{(n)}$ where $k > n$, and

Pl. 5: $H = \lim_{q\to\infty}(1/q)H^{(q)}$ exists.

For many other properties of $H_l(P)$, readers can refer to [5].

2.2. Uncertainty in Dempster–Shafer (DS) framework

Let $\mathscr{X} = \{x_1, x_2, \ldots, x_n\}$ be a finite set. In DS framework uncertainty is modeled by a *basic probability assignment* (BPA) – a function $m : \mathscr{P}(\mathscr{X}) \to [0,1]$, such that $m(\emptyset) = 0$ and $\sum_{A\subseteq\mathscr{X}} m(A) = 1$; $\mathscr{P}(\mathscr{X})$ is the power set of $\mathscr{X}$.

Let $\mathscr{F} = \{A \subseteq \mathscr{X} \mid m(A) > 0\}$. The element of $\mathscr{F}$ are called *focal elements* as mentioned in the introduction. The pair, $(\mathscr{F}, m)$, is called a *body of evidence* (BOE) and the set $\mathscr{X}$ as the *frame of discernment*.

We follow Shafer [28] in the following. Propositions of interest are of the form "*the true value of* θ *lies in* T", where $T \subseteq \mathscr{X}$ and θ is some quantity whose all possible values comprise the frame of discernment, $\mathscr{X}$. Thus, propositions are subsets and vice versa. The value $m(T)$ represents our confidence that the proposition corresponding to T is exactly true; in other words, it represents our confidence that "*the true value of* θ *lies exactly in* T, *and not in any proper subset of* T".

A BOE $(\mathscr{F}, m)$ is called consonant if the focal elements can be arranged in a nested sequence such that

$$A_i \subset A_2 \subset \cdots \subset A_k, \quad \text{when } |F| = k,\ A_i \in F.$$

Based on a BPA other measures of confidence are defined. A *belief measure* is a function $Bel : \mathscr{P}(\mathscr{X}) \to [0,1]$, defined by

$$Bel(A) = \sum_{B\subseteq A} m(B).$$

It represents our confidence that the value of θ lies in A or in any subset of A.

A *plausibility measure* is a function $Pl : \mathscr{P}(\mathscr{X}) \to [0,1]$, defined by

$$Pl(A) = 1 - Bel(\overline{A}).$$

Clearly $Pl(A)$ represents the extent to which we fail to disbelieve A and $Pl(A) \geqslant Bel(A)$.

A belief is called Bayesian, if $Bel(A \cup B) = Bel(A) + Bel(B)$ where $A \cap B = \emptyset$. In other words, belief is

Bayesian iff m focuses only on singletons, i.e., m is a probability distribution.

The *commonality function* for m is the function $Q : \mathscr{P}(\mathscr{X}) \to [0,1]$, defined by

$$Q(A) = \sum_{A \subseteq B} m(B).$$

It is well known [28] that all the above three measures of confidence and the BPA are equivalent, in the sense that, each of them can be expressed as a function of any one of the rest.

Various measures for uncertainty have been proposed by different authors for a BPA. Yager [35] proposed a measure of conflict, $E(m)$, called *dissonance* defined by

$$E(m) = -\sum_{A \in \mathscr{F}} m(A) \log_2 Pl(A).$$

This measure assumes that there is no conflict in evidence, if and only if, there is at least one point in common to all focal elements. Höhle [7, 8] proposed a measure of confusion

$$C(m) = -\sum_{A \in \mathscr{F}} m(A) \log_2 Bel(A).$$

C assumes that there is no conflict in evidence, if and only if, there is exactly one focal element.

Dissonance and confusion have interesting properties like, for a Bayesian belief $E(m)$ and $C(m)$ reduce to Shannon's entropy. Both $E(m)$ and $C(m)$ are additive for strongly independent BOEs. If m_x and m_y are two BPAs on X and Y, respectively, and m is a BPA on $X \times Y$ such that $m(A \times B) = m_x(A) * m_y(B)$, $A \subseteq X$, $B \subseteq Y$ then $E(m) = E(m_x) + E(m_y)$. Similar is the case for $C(m)$.

Klir and Ramer [17] suggest another measure of conflict, named Discord, defined by

$$D(m) = -\sum_{A \in \mathscr{F}} m(A) \log_2 \left[\sum_{B \in \mathscr{F}} m(B) |A \cap B| / |B| \right].$$

The motivation behind discord is that the degree of violation of $B \subseteq A$ should influence the value of the conflict measure. Later, Klir and Parviz [16] introduced another measure of conflict, viz., strife,

$$St(m) = -\sum_{A \in \mathscr{F}} m(A) \log_2 \left[\sum_{B \in \mathscr{F}} m(B)\, |A \cap B| / |A| \right].$$

Smets [29] defined a different measure of conflict, $L(m) = -\sum_{A \subseteq \mathscr{X}} \log_2 Q(A)$.

For nonspecificity also several indices are available. Specificity indicates how concentrated is the mass assignment on the elements of X. Yager's [34, 35] measure of specificity for belief structures is given by $S(m) = \sum_{A \in \mathscr{F}} m(A)/|A|$. $N(m) = 1 - S(m)$ is viewed as a measure of nonspecificity.

Higashi and Klir [10] derived a unique measure of nonspecificity for possibility distributions, called U-uncertainty, which has been extended to a general body of evidence by Dubois and Prade [2]. This, U-uncertainty, is defined by

$$I(m) = \sum_{A \in \mathscr{F}} m(A) \log_2 |A|.$$

$I(m)$ can be viewed as a generalization of Hartley's information measure because when $m(X) = 1$, $I(m) = \log_2 n =$ Hartley entropy.

Lamata and Moral [19] suggested a measure of imprecision $W(m) = \log(\sum_{A \subseteq X} m(A)|A|)$, where $W(m)$ is the logarithm of the average cardinality of focal sets; on the other hand, $I(m)$ is the average of logarithm of cardinalities of focal sets. Thus essentially, both $I(m)$ and $W(m)$ characterize the same type of information.

2.2.1. Total uncertainty (nonspecificity + randomness)

Several authors have tried to get an estimate of total uncertainty (TU) in a system (In the literature the word "total uncertainty" has been used to indicate measures which capture more than one facets of the uncertainty – it may not be an indicator of really the total uncertainty).

Klir and Ramer [17] defined total uncertainty as the algebraic sum of $D(m)$ and $I(m)$, i.e., $T(m) = D(m) + I(m)$. Lamata and Moral [19], on the other hand, added $E(m)$ and $I(m)$ to compute the total uncertainty $G(m) = E(m) + I(m)$.

Such composite measures of uncertainty are meaningful as long as the facets of uncertainty that are quantified by the elementary measures like $E(m)$, $I(m)$, etc., are additive in nature. These measures have interesting properties, but examples can be constructed where they exhibit intuitively unappealing behavior. This has been well analyzed in Pal et al. [21]. In [22]

Pal et al. derived a measure of total uncertainty from a set of intuitively appealing axioms. Their measure of total uncertainty in a body of evidence is

$$H(m)=\sum_{A\in F} m(A)\log_2\{|A|/m(A)\}.$$

This measure can be decomposed into two parts

$$\begin{aligned}H(m) &= -\sum_{A\in F} m(A)\log_2\{m(A)\}\\ &\quad+\sum_{A\in F} m(A)\log_2\{|A|\}\\ &= U(m)+I(m).\end{aligned}$$

$U(m)$ is responsible for the randomness in the system while $I(m)$ accounts for the non-specificity. $H(m)$ has several interesting properties like, $H(m)$ = Shannon's entropy for a Bayesian belief, $H(m)$ = Hartley entropy, when $m(X)=1$. The most interesting property is that $H(m)$ attains the unique maximum value when m focuses on all possible subsets of X in the most uniform manner, i.e., when $m(A)=|A|/K\ \forall A\subseteq X$, $K=\sum_{i=1}^{n} i\binom{n}{i}=2^{n-1}$.

This is the most chaotic assignment, mass is distributed over all possible subsets, and the mass is also uniformly distributed in the sense that $m(A)/|A|=m(B)/|B|=1/K\ \forall A,B\subseteq X$.

Note that the other two measures of TC attain the maximum value when either $m(X)=1$ or $m(\{x\})=1/|X|,\ x\in X$.

2.3. Fuzzy uncertainty

A measure of fuzziness estimates the *average* ambiguity present in a fuzzy set. Consider the properties that seem plausible for such a measure. The fuzziness of a crisp set using any measure should be zero, as there is no ambiguity about whether an element belongs to the set or not. If a set is maximally ambiguous ($\mu_A(x)=0.5\ \forall x$), then its fuzziness should be maximum. When a membership value approaches either 0 or 1, the ambiguity about belonging of its argument in the fuzzy set decreases. A fuzzy set A^* is called a sharpened version A if the following conditions are satisfied: $\mu_{A^*}(x)\leqslant\mu_A(x)$ if $\mu_A(x)\leqslant 0.5$; and $\mu_{A^*}(x)\geqslant\mu_A(x)$ if $\mu_A(x)\geqslant 0.5$.

For a sharpened version A^* of A the measure of fuzziness should decrease because sharpening reduces ambiguity. Another intuitively desirable property is that the fuzziness measure of a set and its complement be equal. For example, the ambiguity present in the sets TALL and NOT TALL (note that NOT TALL is not necessarily SHORT) should be the same.

We now formally define a measure of fuzziness for a discrete fuzzy set A as a mapping $H: P_n(X)\to R^+$ that quantifies the degree of fuzziness present in A where $P_n(X)$ is the set of all fuzzy subsets of X. A measure of fuzziness should satisfy at least the following five properties:

- *Sharpness* P1: $H(A)=0$ iff $\mu_A(x)=0$ or 1 $\forall x\in X$;
- *Maximality* P2: $H(A)$ is maximum iff $\mu_A(x)=0.5$ $\forall x\in X$;
- *Resolution* P3: $H(A)\geqslant H(A^*)$ where A^* is a sharpened version of A;
- *Symmetry* P4: $H(A)=H(1-A)$, where $\mu_{1-A}=\mu_A$;
- *Valuation* P5: $H(A\cup B)+H(A\cap B)=H(A)+H(B)$.

Ebanks [4] along with P1–P5, suggested a sixth requirement called generalized additivity, which is somewhat difficult to interpret. We already mentioned that several authors have quantified fuzziness. Some of these measures satisfy P1–P5 while others do not. Instead discussing all of them, we just discuss two new families, *additive* class and ***multiplicative*** class, of measures of fuzziness [20].

2.3.1. Multiplicative class

Let $f:[0,1]\to R^+$ be a concave-increasing function on [0, 1], i.e. $f'(t)>0\ \forall t\in[0,1]$ and $f''(t)<0$ $\forall t\in[0,1]$.

Now define

$$\begin{aligned}\hat{g}(t) &= f(t)f(1-t),\\ g(t) &= \hat{g}(t)-\min_{0\leqslant t\leqslant 1}\{\hat{g}(t)\}\end{aligned}\qquad(2)$$

and

$$H_*(A)=k\sum_{i=1}^{n} g(\mu_i),\quad k\in R^+.\qquad(3)$$

It can be easily seen that H_* satisfies P_1–P_5. Thus H_* is a measure of fuzziness.

Example 1. Let $f(t)=te^{1-t}$ then $H_*=H_{*QE}=k\sum\{\mu_i(1-\mu_i)\}$. The constant e has been absorbed in the constant k. It is easy to show directly that properties P1–P5 are satisfied by $H_{*\mathrm{QE}}$. We call $H_{*\mathrm{QE}}$ the *quadratic entropy* of the fuzzy set because of its similarity to Vajda's probabilistic quadratic entropy for a discrete probabilistic framework,

$$H_{V_{\mathrm{QE}}}(P)=\sum_{i=1}^{n} p_i(1-p_i). \quad (4)$$

2.3.2. Additive class

For the additive class f is less restricted than for the multiplicative case. We require only concavity, and we use addition (+) in (2) in place of multiplication. In other words, $f:[0,1]\to R^+$ and concave, i.e., $f''(t)<0\ \forall t\in[0,1]$. Thus, for the additive class the functions g and H_+ are defined as

$$\hat{g}(t)=f(t)+f(1-t),$$

$$g(t)=\hat{g}(t)-\min_{0\leqslant t\leqslant 1}\{\hat{g}(t)\},$$

and for $A\in P_n(X)$ we define

$$H_+=K\sum_{i=1}^{n} g(\mu_i), \quad K\in R^+. \quad (5)$$

It can be proved [20] that H_+ satisfies P1–P5 and hence a measure of fuzziness. Like the multiplicative class, there can be many expressions for the additive measure of fuzziness. We just give one example using the same f used in Example 1.

Example 2. Let $f(t)=t\mathrm{e}^{1-t}$ then $H_+=\sum_{i=1}^{n}\mu_i\mathrm{e}^{1-\mu_i}+(1-\mu_i)\mathrm{e}^{\mu_i}$ = the exponential fuzzy entropy of Pal and Pal [23] (subject to adjustment of constant).

It is easy to see that the additive and multiplicative classes are consistent with Yager's view of fuzziness. In other words, they can be interpreted as measuring the lack of distinction between A and its complement. Properties of these two measures are discussed in [20].

2.3.3. Total uncertainty (randomness + fuzziness)

Like DS framework some attempts have been made to quantify total uncertainty when a probabilistic system has fuzziness associcated with it. Like the attempts in DS framework, here also the total uncertainty cannot be viewed really as the TU in a system. These measures quantify TU in the sense they capture total uncertainty when the system does not have nonspecificity.

Let $X=\{x_1,\ldots,x_n\}$ be the universe of discourse and A be a fuzzy set defined on X, P be the probability distribution such that $p(x_i)=p_i$, $i=1,\ldots,n$, $\sum p_i=1$, i.e., each element of X has a probability of occurrence. With a view to measuring the total uncertainty (PU and FU) associated with such a system several expressions have been suggested [3, 31]. Let (X,P) be a discrete probability framework and $A\in P_n(X)$, $\mu_A(x_i)=\mu_i$, $i=1,\ldots,n$. Zadeh [36] defined the entropy of a fuzzy set with respect to the discrete probability framework P as the weighted Shannon entropy:

$$H_z(A,P)=-\sum_{i=1}^{n}\mu_i p_i \log p_i.$$

This entropy is a measure of uncertainty associated with a fuzzy event, and was the first composite measure of probabilistic and fuzzy uncertainties. Note that Zadeh did not call it to be a measure of TC. We included here as it attempts to account for both probability and fuzziness.

$H_z(A,P)\leqslant H_l(P)$, and this is counter-intuitive as uncertainty should not be reduced with addition of fuzziness. Moreover, if A is defuzzified, depending on the defuzzified output, H_z will change. For example, if $\mu_i\to 0$ (with $p_i\neq 0$), $H_z(A,P)\to 0$; again if $\mu_i\to 1$ $\forall_i$, $H_z(A,P)\to H_l(P)$.

Suppose we have a probabilistic framework (X,P), and there is some difficulty in interpreting x_i, the outcome of a trial, as 0 or 1. The average amount of ambiguity involved in the interpretation of such an outcome as suggested by Deluca and Termini [3] is

$$H_{\mathrm{DT}}(A,P)=-\sum_{i=1}^{n}p_i(\mu_i\log\mu_i+(1-\mu_i)\log(1-\mu_i)). \quad (6)$$

Deluca and Termini then defined the *total average uncertainty* in the system as the sum of Shannon's probabilistic entropy $H_l(P)$ and $H_{\mathrm{DT}}(A,P)$,

$$H_{\mathrm{DT}}^{\mathrm{tot}}(A,P)=H_l(P)+H_{\mathrm{DT}}(A,P). \quad (7)$$

$H_{\mathrm{DT}}^{\mathrm{tot}}(A,P)$ is interpreted as the total average uncertainty involved in making a prevision about the elements of X which appear as a result of the experiment, and in making a (0 or 1) decision about their values.

Note that (6) is essentially a measure of fuzziness. Suppose if the experiment is repeated N times and x_i occurs n_i times, $N = \sum_{i=1}^{n} n_i$, then the fuzziness in the resultant set with N elements is

$$\begin{aligned} H_{\mathrm{DT}} &= -K \sum_{i=1}^{N} \{\mu_i \log \mu_i + (1-\mu_i)\log(1-\mu_i)\} \\ &= -K \sum_{i=1}^{i=n} n_i \{\mu(x_i)\log \mu(x_i) \\ &\qquad + (1-\mu(x_i))\log(1-\mu(x_i))\} \\ &= -K' \sum_{i=1}^{i=n} \frac{n_i}{N} \{\mu(x_i)\log \mu(x_i) \\ &\qquad + (1-\mu(x_i))\log(1-\mu(x_i))\} \\ &= -K' \sum_{i=1}^{i=n} p_i \{\mu(x_i)\log \mu(x_i) \\ &\qquad + (1-\mu(x_i))\log(1-\mu(x_i))\} \\ &= H_{\mathrm{DT}}(A,P) \text{ in (6).} \end{aligned}$$

Xie and Bedrosian [31] defined the total uncertainty associated with a system in a slightly different framework. Consider a crisp set $\hat{A}$ with only two kinds of elements, 0 and 1, having probabilities $P = \{p_0, p_1\}$, where $p_1 = 1 - p_0$. Suppose the sharpness of $\hat{A}$ is impaired, so that 0 takes some value in the interval [0, 0.5] and 1 takes some value in [0.5, 1]. Then A becomes a fuzzy set. Xie and Bedrosian defined the total uncertainty associated with this system as the sum of Shannon's probabilistic entropy $H_l(P)$ and $H_{\mathrm{DT}}(A)$, where the fuzzy entropy of A defined by $H_{\mathrm{DT}}(A) = -\sum_{i=1}^{n}(\mu_i \log \mu_i + (1-\mu_i)\log(1-\mu_i))$. Thus,

$$H_{\mathrm{XB}}^{\mathrm{tot}}(A,P) = H_l(P) + H_{\mathrm{DT}}(A). \tag{8}$$

Eq. (8) simply adds together terms that measure each kind of uncertainty separately, and as explained earlier, it is exactly the same as in (7). Since fuzziness is conceptualy different from probabilistic uncertainty, such algebraic summation of elementary measures as in (7) and (8) is hard to justify. Referring back to the experiment of Deluca and Termini, in absence of fuzziness, the system for (7) reduces to a two state system. Should the total uncertainty reduce to the probabilistic uncertainty of a two state system or of an n-state system (because the original system is an n-state system). Some justification can be given to favor either view. Moreover, (7) and (8) reduce to Shannon's entropy (either of an n-state system or of a 2-state system) irrespective of the defuzzification process.

3. Some other aspects of uncertainty

3.1. Exponential probabilistic entropy

For Shannon's framework as $p_i \to 0$, $\Delta I(p_i) \to -\infty$ and $\Delta I(p_i = 0)$ is not defined; but as $p_i \to 1$, $\Delta I(p_i) \to 0$ and $\Delta I(p_i = 1) = 0$. We prefer to have a gain function which is finite for $p_i \in [0,1]$. Moreover, intuitively, ignorance associated with (or unlikeliness of) an event should better be a function of $(1 - p_i)$ rather than $1/p_i$. Examples can also be shown to suggest that some exponential function of $(1 - p_i)$ might be a good choice for the gain function ΔI [23, 24] at least for some problems. Thus, based on a set of desirable properties [23] the exponential gain function is defined as, $\Delta I(p_i) = e^{(1-p_i)}$.

And the expected gain in information associated with the system, i.e., the entropy of the system is defined as

$$H_e(P) = \sum_{i=1}^{n} p_i e^{(1-p_i)}. \tag{9}$$

Definition (9) bears many interesting properties like those of Shannon's entropy but it is not additive. We next present some of them:

Pe 1: $H_e(P)$ attains its unique maximum value at $p_i = 1/n$, $\forall i = 1,\ldots,n$.

Pe 2: $H_e(p)$ attains its minimum value when $p_k = 1$ for some k and $p_i = 0$, $i \neq k$.

Pe 3: Like $H_l(P)$, equalization of probability increases entropy. Thus, if $P1 = \{p_1, p_2, p_3, \ldots, p_n\}$ and $P2 = \{p_1 - \delta, p_2 + \delta, p_3, \ldots, p_n\}$ such that $p_1 > p_2$, $\delta > 0$ and $p_1 - \delta > p_2 + \delta$, then $H_e(P1) \leqslant H_e(P2)$.

$H_e(P)$ has interesting properties for compound experiments, see also [26]. Let there be two experiments A and B with n and m states, respectively. The experiment A is defined by the states $A = \{a_1, a_2, \ldots, a_n\}$ with probabilities $P = \{p_1, p_2, \ldots, p_n\}$ and the other experiment is defined by $B = \{b_1, b_2, \ldots, b_m\}$ with probabilities $Q = \{q_1, q_2, \ldots, q_m\}$. The compound

experiment is then defined as $A \times B = \{(a_k, b_l),\ k = 1,\dots,n;\ l = 1,\dots,m\}$. For $A \times B$ we have
Pe 4: $H_n(A/B) \leqslant H_n(A)$.
Pe 5: $H_m(B/A) \leqslant H_m(B)$.
Pe 6: $H_{nm}(A \times B) \leqslant H_n(A) + H_m(B/A)$.
Pe 7: $H_{nm}(A \times B) \leqslant H_n(A) + H_m(B)$.
Here $H_n(A/B)$ is the entropy of A conditioned by B and $H_{nm}(A \times B)$ is the entropy of the compound system $(A \times B)$.

3.2. Hybrid entropy and higher-order entropy

3.2.1. Hybrid entropy

Consider the experimental set up used by Xie and Bedrosian [31]. A source is generating symbols 0 and 1 with probabilities p_0 and p_1, respectively. Due to some noise in the system the sharpening in the set is impaired resulting in a fuzzy set A. Suppose μ_i denotes the membership of x_i to the fuzzy set $A \in P_n(X)$ defined as "*symbols close to* 1". Pal and Pal [25] defined the hybrid entropy of A as

$$H_{PP}^{hbl}(A,P) = -\ p_0 \log E_0 - p_1 \log E_1, \tag{10}$$

where E_0 and E_1 are defined as

$$E_0 = \frac{1}{n}\sum_{i=1}^{n}(1-\mu_i)\,e^{\mu_i} \tag{11}$$

and

$$E_1 = \frac{1}{n}\sum_{i=1}^{n}(\mu_i)\,e^{1-\mu_i}. \tag{12}$$

Expressions (11) and (12) are the two terms of the fuzzy entropy of Pal and Pal [23]. A detailed discussion on the justification for (10) can be found in [25]. Eq. (10) is suggested using Shannon's framework. The hybrid entropy can also be defined based on exponential gain function of Pal and Pal:

$$H_{PP}^{hbe}(A,P) = p_0 e^{1-E_0} + p_1 e^{1-E_1}, \tag{13}$$

with E_0 and E_1 as in (11) and (12).

Unlike (7) or (8) the hybrid entropies (10) and (13) reduce to the probabilistic (logarithmic or exponential) entropy of a 2-state system only with *proper* defuzzification. By proper defuzzification we mean the situation where the number of symbols (0 or 1) of each kind generated by the defuzzification process is the same as the number of symbols originally generated by the source. However, the hybrid entropies (10) and (13) and the total entropy (8) of Xie and Bedrosian cannot model the situation with an n-state system.

3.2.2. Higher-order entropy

Let X be a set of players with say, $n = 20$ members and A be the set of GOOD players defined on X. Suppose we form a team of size $r = 11 \leqslant n$. On an average, to what extent this team is good! It raises two issues: degree to which a collection possesses the property of "goodness" and how to get a measure of average ambiguity associated with such collections. Answer to the first depends on the problem at hand. Given a property characterized by a fuzzy set A, how can we extend the same property to a collection of elements from X. Let us denote the ith subset of X with r members as S_i^r. Let X be a set of acrobats, and a subset S_i^r of X are standing in such a manner that if one of the members falls the entire team falls. In this case the extent S_i^r is good may be defined as $\min_{x_k \in S_i^r}\{\mu_A(x_k)\}$. On the other hand, if X is a group of quiz experts, then degree to which the team S_i^r is good may be better computed as $\max_{x_k \in S_i^r}\{\mu_A(x_k)\}$ because in a quiz team, if one member succeeds, the team succeeds. If X is a group of football players or members of a Tug-of-war team, possibly some other aggregation operator (may be average) will be a better choice. Let $\mathscr{C}$ be the aggregation operator. The operator $\mathscr{C}$ can be interpreted to define a fuzzy set *A-team* (e.g., GOOD-*team*) on the universe containing all possible subsets (of size r) of X. Let $\mu(S_i^r) = \mathscr{C}\{\mu_A(x_i),\ x_i \in S_i^r\}$ be the degree to which S_i^r possesses the property *A-team*. Now to answer the second question, i.e., to get a measure of fuzziness of the fuzzy set *A-team*, we may use any one of the functional forms that can be used to quantify fuzziness in A. For example, using the quadratic fuzzy entropy, *the rth order fuzzy entropy* [25] of A can be defined as $H_{*\mathrm{QE}} = K\sum_i \mu(S_i^r)(1-\mu(S_i^r))$. Properties of some such measures can be found in [25].

3.3. Conflict in DS framework

Here our objective is to introduce a measure of uncertainty that properly accounts for *conflict* in a body of evidence. Most of the measures of uncertainty like discord, confusion, dissonance, etc., do not quantify the conflict aspect of uncertainty properly. Let us illustrate the concept of conflict with reference to the die throwing experiment, described in the introduction. Suppose we have two bodies of evidence $(\mathscr{F}_1, m_1)$ and $(\mathscr{F}_2, m_2)$ as follows:

$$m_1: m_1(\{1,2,3\}) = 0.67, \quad m_1(\{3,4,5\}) = 0.33,$$

and

$$m_2: m_2(\{1,2,3\}) = 0.67, \quad m_2(\{2,3,4\}) = 0.33.$$

Intuitively, the second body of evidence has less *conflict* and hence less uncertainty than the former. This is because, although, they both divide their assignments between two subsets in the same fashion, in the latter case, the focal elements have a greater overlap. We feel, the greater the overlap, the lesser should be the conflict. This facet of uncertainty which depends on the amount of overlap between focal elements can be intuitively understood to be the uncertainty arising from conflict. Note that conflict arises only when there is randomness, but it may not account for all of the uncertainty that may arise due to randomness.

We adopt a fresh approach to get a new measure of conflict. We view the dissimilarity between two propositions as a *metric distance* between them. The greater the distance between two propositions, the greater is the conflict between them. On this basis we suggest a set of desirable axioms for a measure of conflict between two propositions and derive a *unique* expression satisfying these axioms. The average of the conflict between propositions gives a measure of total conflict in a body of evidence.

3.3.1. Desirable axioms for a measure of conflict

Let $(\mathscr{F}, M)$ be *any* BOE and $A, B \in \mathscr{F}$. Let Π be a real-valued function whose value at (A, B) quantifies the conflict between propositions A and B. We think that the following axioms are essential for a measure of conflict.

1. *Nonnegativity*:
   $\Pi(A,B) \geqslant 0$, with equality iff $A = B$.
2. *Symmetry*:
   $\Pi(A,B) = \Pi(B,A)$.
3. *Triangle inequality*:
   $\Pi(A,B) \leqslant \Pi(A,C) + \Pi(B,C)\ \forall C \in \mathscr{F}$.
4. *Intermediate behavior*:
   If $\Pi(A,B) > \Pi(C,D)$ then $|A \cap B|/|A \cup B| < |C \cap D|/|C \cup D|$ and if $\Pi(A,B) < \Pi(C,D)$ then $|A \cap B|/|A \cup B| > |C \cap D|/|C \cup D|\ \forall C, D \in \mathscr{F}$.
5. *Normalization*:
   $\Pi(A,B) \leqslant 1$ with $\Pi(A,B) = 1$ iff $A \cap B = \emptyset$.
6. *Branching*:
   $\Pi(A,B) = \Pi(A, A \cup B) + \Pi(B, A \cup B)$.

Axioms 1–3 constitute the requirements for a metric. Axiom 4 says, the greater the conflict, the lesser is the fraction of the total part in agreement; similarly, the lesser the conflict the greater is the fraction of the total part in agreement. The normalization axiom asserts that when A and B are disjoint then they are in maximum conflict regardless of their sizes. Moreover, for any two sets, the maximum conflict is to be attained only when they are disjoint. The axiom of branching is introduced as an alternative way of computing $\Pi(A,B)$. It means that the conflict between A and B is the sum of the conflict between A and $A \cup B$, and that between B and $A \cup B$. This is intuitively reasonable as the conflict between A and $A \cup B$ arises due to the elements in $B - A$. Similarly, the conflict between B and $A \cup B$ arises because of the elements in $A - B$. In other words, Axiom 6 implies that conflict between A and B arises due to the elements of $A \cup B - A \cap B$; i.e., due to the elements which are not common to both A and B.

It can be proved [6] that the *unique* real valued function satisfying axioms 1–6 is given by

$$\Pi(A,B) = 1 - \frac{|A \cap B|}{|A \cup B|} \quad \forall A, B. \tag{14}$$

Thus, $\Pi(A,B)$ represents the conflict between propositions A and B,

$$\Pi(A,B) = 1 - \frac{|A \cap B|}{|A \cup B|} = \frac{|A \Delta B|}{|A \cup B|}, \tag{15}$$

where $A \Delta B = A \cup B - B \cap A$, is the symmetric difference of A and B.

We now give the necessary definitions leading to the total measure of conflict in a *BOE*.

Definition 1. Define $CON(A)$, representing the conflict of A with the rest of the propositions, as the average conflict of A with the rest. Thus,

$$CON(A) = \sum_{B \in \mathscr{F}} m(B) \Pi(A,B).$$

Definition 2. Define the *total conflict* in the BOE $(\mathscr{F}, m)$, $TC[(\mathscr{F}, m)]$, as the average of the conflict associated with each proposition. Thus,

$$\begin{aligned} TC[(\mathscr{F},m)] &= \sum_{A\in\mathscr{F}} m(A)CON(A) \\ &= \sum_{A\in\mathscr{F}} m(A) \sum_{B\in\mathscr{F}} m(B)\Pi(A,B). \qquad (16) \end{aligned}$$

This new measure of conflict has several interesting properties [6]. We discuss here some of them.

Pc. 1: Let $(\mathscr{F}, m)$ be a BOE, $A, B \in \mathscr{F}$ with $|B| \geqslant 2$. Let $B = B_1 \cup B_2$, $B_1 \cap B_2 = \emptyset$ and $B_1 \neq \emptyset$, $B_2 \neq \emptyset$. Then, $\Pi(A,B) < \Pi(A,B_1) + \Pi(A,B_2)$.

Pc. 2: Let $(\mathscr{F}, m)$ be a BOE, $B \in \mathscr{F}$, $B = \{b_1, \ldots, b_p\}$, $|B| = p \geqslant 2$. Then, $\Pi(A,B) \leqslant 1/p \sum_{j=1}^{p} \Pi(A, \{b_j\})\ \forall A \in \mathscr{F}$.

The next property describes the intermediate behavior of the total conflict in a body of evidence.

Pc. 3: Let $(\mathscr{F}, m)$ be a BOE where $\mathscr{F} = \{A_1, \ldots, A_n, B\}$ and $|B| = p \geqslant 2$. Let $(\mathscr{F}', m')$ be defined by $\mathscr{F}' = \{A_1, \ldots, A_n\} \cup \{\{b_1\}, \ldots, \{b_p\}\}$ and $m'(Z) = m_1(Z) + m_2(Z)$ where

$$m_1(Z) = \begin{cases} m(Z) & \text{if } Z = A_i,\ i = 1, \ldots, n, \\ 0 & \text{otherwise,} \end{cases}$$

$$m_2(Z) = \begin{cases} m(B)/p & \text{if } Z \in \{\{b_1\}, \ldots, \{b_p\}\}, \\ 0 & \text{otherwise.} \end{cases}$$

Then, $(\mathscr{F}', m')$ is a BOE, and $TC[(\mathscr{F}, m)] < TC[(\mathscr{F}', m')]$.

As a consequence of Pc. 3, we get Pc. 4, which gives the range of TC.

Pc. 4: $TC[(\mathscr{F}, m)]$ attains its maximum value of $1 - 1/|\mathscr{X}|$ iff the evidence is Bayesian with weights uniformly distributed over all singleton subsets of the frame of discernment, $\mathscr{X}$. That is for $m(\{x_i\}) = 1/|\mathscr{X}|\ \forall x_i \in \mathscr{X}$. TC attains its minimum value of zero iff $|\mathscr{F}| = 1$.

Pc. 5: When the evidence is Bayesian TC reduces to *Vajda's quadratic entropy* $\sum p_i(1 - p_i)$.

Note that most of the earlier measures like conflict, confusion, etc., reduce to Shannon's logarithmic entropy for a Bayesian belief, while TC becomes equal to Vajda's quadratic entropy under the same condition.

4. Conclusion and discussion

Primarily, there are three types of uncertainty, namely, probabilistic uncertainty, nonspecificity and fuzziness that may be associated with a system. We have reviewed some measures for each type of uncertainty. Note that these three are the major components of uncertainty, and by no means account for all aspects of uncertainty. Various attempts to quantify total uncertainty (PU + NS, PU + FU) have been presented. A critical analysis of different conceptual aspects of uncertainty reveals that each type of uncertainty may have different facets. For example, in the DS framework, in addition to uncertainty that is associated with the selection of one of several hypotheses, there is a facet that represents conflict between different hypotheses. When the universe of a fuzzy set is defined in terms of outcomes of a probabilistic experiment, interpretation of an outcome involves a different type of uncertainty where probability and fuzziness may interact in a complex manner. We discussed some models to quantify some such facets. Subjective evaluation of outcomes of a probabilistic system – experimenter's eagerness to realize some particular outcome – has a strong relation to the subjective assessment of uncertainty associated with a system. Shannon's weighted entropy is not an adequate model to capture this aspect of uncertainty. Investigation needs to be done in this area. So far, as to the knowledge of the author, no attempt has been made to estimate the total uncertainty in a system which involves all three major components of uncertainty. This requires to find a set of desirable axioms for characterizing total uncertainty. It will also require to find how different types of uncertainty interact with each other. These non-trivial tasks require further research.

References

[1] D. Bhandari, N.R. Pal, Some new information measures for fuzzy sets, Inform. Sci. 67 (3) (1993) 209–228.

[2] D. Dubois, H. Prade, A note on measures of specificity for fuzzy sets, Int. J. Gen. Systems 10 (1985) 279–283.

[3] A. Deluca, S. Termini, A definition of nonprobabilistic entropy in the setting of fuzzy sets theory, Inform. and Control 20 (1972) 301–312.

[4] B.R. Ebanks, On measures of fuzziness and their representations, J. Math. Anal. Appl. 94 (1983) 24–37.

[5] S. Guiasu, Information Theory with Applications, McGraw-Hill, New York, 1977.
[6] T. George, N.R. Pal, Quantification of Conflict in Dempster–Shafer framework: a new approach, Int. J. Gen. Systems 24 (4) (1996) 407–423.
[7] U. Höhle, Fuzzy plausibility measures, in: E.P. Klement (Ed.), Proc. 3rd Int. Seminar on Fuzzy Set Theory, Johannes Kepler University, Linz, 1981, pp. 7–30.
[8] U. Höhle, Entropy with respect to plausibility measures, Proc. 12th IEEE Int. Symp. on Multiple Valued Logic, Paris, 1982, pp. 167–169.
[9] M. Higashi, G.J. Klir, On measures of fuzziness and fuzzy complements, Int. J. Gen. Systems 8 (1982) 169–180.
[10] M. Higashi, G.J. Klir, Measures of uncertainty and information based on possibility distributions, Int. J. Gen. Systems 9 (1983) 43–58.
[11] A. Kaufmann, Introduction to the Theory of Fuzzy Subsets: vol. 1, Academic Press, New York, 1975.
[12] B. Kosko, Fuzzy entropy and conditioning, Inform. Sci. 40 (1986) 165–174.
[13] J. Knopfmacher, On measures of fuzziness, J. Math. Anal. and Appl. 49 (1975) 529–534.
[14] G.J. Klir, T.A. Folger, Fuzzy Sets, Uncertainty, and Information, Prentice-Hall, Englewood Cliffs, NJ, 1988.
[15] U. Kumar, V. Kumar, J.N. Kapoor, Normalized measures of entropy, Int. J. Gen. Systems 12 (1) (1986) 55–69.
[16] G.J. Klir, B. Parviz, A note on the measure of discord, in: D. Dubois et al. (Ed.), Proc. Eighth Conf. on Uncertainty in AI, Stanford, Morgan Kaufmann, San Mateo, Calif., 1992, pp. 138–141.
[17] G.J. Klir, A. Ramer, Uncertainty in Dempster–Shafer theory: a critical re-examination, Int. J. Gen. Systems 18 (1990) 155–166.
[18] S.G. Loo, Measures of fuzziness, Cybernetica 20 (1977) 201–210.
[19] M.T. Lamata, S. Moral, Measures of entropy in the theory of evidence, Int. J. Gen. Systems 14 (1987) 297–305.
[20] N.R. Pal, J.C. Bezdek, Measuring fuzzy uncertainty, IEEE Trans. Fuzzy Systems 2(2) (1994) 107–118.
[21] N.R. Pal, J.C. Bezdek, R. Hemasinha, Uncertainty measure for evidential reasoning I: a review, Int. J. Approximate Reasoning 7 (3/4) (1992) 165–183.
[22] N.R. Pal, J.C. Bezdek, R. Hemasinha, Uncertainty measure for evidential reasoning II: a new measure of total uncertainty, Int. J. Approximate Reasoning 8 (1) (1993) 1–16.
[23] N.R. Pal, S.K. Pal, Object-background segmentation using new definitions of entropy, IEEE Proc. 136, Part E (1989) 284–295.
[24] N.R. Pal, S.K. Pal, Entropy: a new definition and its applications, IEEE Trans. Systems Man and Cybern. 21 (5) (1991) 1260–1270.
[25] N.R. Pal, S.K. Pal, Higher order fuzzy entropy and hybrid entropy of a set, Inform. Sci. 61 (3) (1992) 211–231.
[26] N.R. Pal, S.K. Pal, Some properties of the exponential entropy, Inform. Sci. 66 (1–2) (1992) 119–137.
[27] C.E. Shannon, A mathematical theory of communication, Bell Systems. Tech. J. 27 (1948) 379–423.
[28] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.
[29] P. Smets, Information content of an evidence, Int. J. Man–Mach. Stud. 19 (1983) 33–43.
[30] E. Trillas, T. Riera, Entropies in finite fuzzy sets, Inform. Sci. 15 (1978) 159–168.
[31] W.X. Xie, S.D. Bedrosian, An information measure for fuzzy sets, IEEE Trans. Systems Man Cybern. SMC-14 (1) (1984) 151–156.
[32] R.R. Yager, On measures of fuzziness and negation, Part I: Membership in the unit interval, Int. J. Gen. Systems 5 (1979) 221–229.
[33] R.R. Yager, On measures of fuzziness and negation, Part II: Lattices, Inform. and Control 44 (1980) 236–260.
[34] R.R. Yager, Measuring tranquility and anxiety in decision making: an application of fuzzy sets, Int. J. Gen. Systems 8 (1982) 139–146.
[35] R.R. Yager, Entropy and specificity in a mathematical theory of evidence, Int. J. Gen. Systems 9 (1983) 249–260.
[36] L.A. Zadeh, Probability measures of fuzzy events, J. Math. Anal. Appl. 23 (1968) 421–427.