# Entropy of Credibility Distributions for Fuzzy Variables

Pingke Li and Baoding Liu, *Senior Member, IEEE*

***Abstract*—This paper deals with the degree of uncertainty associated with fuzzy variables. Based on the notion of credibility measure, a definition of entropy is formulated from an information theoretical view and its properties are investigated. Finally, some comments are given on the construction of alternative definitions of entropy and axiomatic characterization.**

***Index Terms*—Credibility, entropy, fuzzy variable, information, uncertainty.**

## I. Introduction

FUZZINESS, a feature of uncertainty, results from the lack of sharp distinction of the boundary of a set, i.e., an individual is neither definitely a member of the set nor definitely not a member of it. Fuzziness also arises in logic when a proposition can be treated as neither definitely true nor definitely false. Fuzzy entropy, a term used to represent the degree of uncertainty, was first mentioned in 1965 by Zadeh [50]. The first attempt to quantify the fuzziness was made in 1968 by Zadeh [51], who introduced a probabilistic framework and defined the entropy of a fuzzy event as a weighted Shannon entropy. The idea of measuring fuzziness without reference to probabilities began in 1972. De Luca and Termini [3] proposed four requirements with which a fuzzy entropy measure should comply and defined a kind of entropy of fuzzy sets using Shannon's functional form. However, Ebanks [7] proposed a fifth requirement for the definition of fuzzy entropy and gave the necessary and sufficient conditions on functions that satisfy the five requirements for discrete fuzzy sets. Bezdek [1] generalized the result of De Luca and Termini [3] in 1973 by defining the classification entropy. In 1975, Kaufmann [15] introduced an index of fuzziness based on the Minkowski distance from a fuzzy set to its nearest crisp set. Knopfmacher [19] and Loo [26] showed that the definitions given by De Luca and Termini [3] and Kaufmann [15] were special cases of a larger class of measures of fuzziness. Yager [41], [42] provided another way to view the degree of fuzziness in terms of a lack of distinction between a proposition and its negation. This concept was extended by Higashi and Klir [13] to a very general class of fuzzy complements. Besides, there is a lot of literature concerning the definition of fuzzy entropy and its applications such as Pal and Pal [30], [31], Yager [47], and Fan and Ma [8]. A good survey of fuzzy entropy for finite universal sets can be found in Pal and Bezdek [34], [35].

In 1978, Zadeh [52] proposed possibility theory as a mathematical framework of the calculus of fuzzy restrictions. A fuzzy restriction is a fuzzy set that represents a restriction on the possibility of an event given an imprecise or vague description of the event. Possibility theory had been developed by many researchers such as Nahmias [28], Yager [43], and Dubois and Prade [4], [5], and had become a powerful tool to deal with fuzziness. Besides, Liu and Liu [25] introduced a self-dual measure, *credibility measure*, to quantify the chance of occurrence of fuzzy events and Liu [24] provided credibility theory as an axiomatic foundation of fuzziness. A frequently used concept in possibility theory is that of fuzzy variable which was first used by Kaufmann [15]. The possibility distribution of a fuzzy variable can be induced by the membership function of a normal fuzzy set. However, although a possibility distribution and a membership function have a common mathematical expression, the underlying ideas are subtly different. As an example, the membership function of fuzzy set "young" specifies the compatibility of every possible value of ages with the concept "young," while possibility theory is concerned with the possibility of the occurrence of a particular age given that someone is known to be young. That is, the membership function represents uncertainty in presence of a concept described as a fuzzy set while the possibility distribution indicates the possibility of the occurrence of each event which is expressed by some values of a fuzzy variable. The uncertainty associated with possibility distributions is called nonspecificity in the literature and is measured as the degree of impreciseness of the outcomes of a variable restricted by a fuzzy set. Yager [44] initiated the concept of specificity of a possibility distribution which was interpreted by Dubois and Prade [6] as a measure of information attached to a possibility distribution. Higashi and Klir [14] also suggested a measure of imprecision, called U-uncertainty, for normal fuzzy sets and proposed three properties as the requirements for measuring nonspecificity. These requirements were further investigated by Ramer and Lander [37] and Ramer [36]. Kroupa [20], [21] studied the measures of nonspecificity in the framework of Choquet integral. Besides, some further research has been done by Yager [45], [46], Klir and Ramer [17], Pal *et al.* [32], [33], and Garmendia *et al.* [9]. Klir [16] provided a comprehensive and in-depth overview of these topics. The notion of nonspecificity is used as a measure of evaluation in fuzzy inference and linguistic analysis, and plays a similar role in possibility theory that the notion of entropy plays in probability theory. Yuan and

Manuscript received May 20, 2005; revised January 25, 2006, May 19, 2006, and October 14, 2006. This work was supported by the National Natural Science Foundation of China under Grant 60425309.

P. Li is with the Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC 27695 USA (e-mail: pli@ncsu.edu).

B. Liu is with the Department of Mathematical Sciences, Tsinghua University, Beijing 100084, China (e-mail: liu@tsinghua.edu.cn).

Digital Object Identifier 10.1109/TFUZZ.2007.894975

Shaw [49] proposed a method to construct fuzzy decision trees using nonspecificity as a measure of discrimination. Gebhardt and Kruse [10], [11] provided a foundation of a learning algorithm for possibilistic networks. See, also, Wu and Klir [40], Yeung and Wang [48], and Ricotta [38] for other applications.

When the uncertainty of a fuzzy variable is concerned, it depends on both the possibilities of all the values that the fuzzy variable could take and the amount of them, which leads to the difference between measuring the uncertainty of a fuzzy variable and that of a fuzzy set. From an information theoretical view, a desirable property in this case is that the degree of uncertainty is minimum when the fuzzy variable always takes a unique value, and is maximum when the fuzzy variable takes all the values with possibility 1. Here, we know that the uncertainty of a fuzzy variable does not express the difficulty in deciding whether an element belongs to a set or not but represents the extent of uncertainty resulted from information deficiency.

This paper aims at introducing a definition of entropy of credibility distributions to measure the uncertainty associated with fuzzy variables. The measure is called entropy because it measures information deficiency and because of its entropy-like functional form. The entropy of credibility distributions can be regarded to some extent as an alternative measure of nonspecificity. Unlike the measure of U-uncertainty proposed by Higashi and Klir [14], which was defined on normalized ordered possibility distributions, the entropy of credibility distributions is defined directly by the credibilities of focal elements and only requires the normality of the possibility distributions. In Section II, we introduce some basic concepts and results on fuzzy sets and fuzzy variables. In Section III, the definition of entropy of discrete credibility distributions is proposed and its properties are investigated in detail. The definition of entropy of continuous credibility distributions is provided and analyzed in Section IV.

## II. Preliminaries

In this section, we will briefly review some concepts and results on fuzzy sets and fuzzy variables.

### A. Fuzzy Sets and Their Uncertainty

A fuzzy set, characterized by its membership function, is a generalization of a classical set which is referred to as a crisp set for distinction. The membership function of a fuzzy set measures a degree of membership that can range from 0 to 1. The more often the object belongs to the fuzzy set, the higher the degree of membership. The domain of a membership function of a fuzzy set is called the universe of discourse. The universe may be any set, but generally, it is either a discrete set or the set of real numbers.

A fuzzy set $\tilde{A}$ over a universe $X$ is defined by a membership function $\mu_{\tilde{A}} : X \to [0, 1]$. The membership value $\mu_{\tilde{A}}(x)$ specifies the degree to which the element $x$ is in the fuzzy set $\tilde{A}$. Any element with a nonzero membership degree is called a *focal* element of $\tilde{A}$. Besides, the support of a fuzzy set $\tilde{A}$ is the crisp set that contains all $x \in X$ with $\mu_{\tilde{A}}(x) > 0$ and is denoted by $\mathrm{Supp}(\tilde{A})$. Generally, we only concentrate on all the focal elements of a fuzzy set $\tilde{A}$, i.e., $\mathrm{Supp}(\tilde{A})$. In addition, we call $\tilde{A}$ a discrete fuzzy set if the universe $X$ is a discrete set while we call $\tilde{A}$ a continuous fuzzy set if $X$ is the set of real numbers and $\mu_{\tilde{A}}$ is continuous on $X$.

The characterization and quantification of uncertainties associated with fuzzy sets are important issues that affect the management of uncertainty in system modeling and design. As mentioned previously, there are many measures of fuzziness which usually can be classified as entropy-like measures and distance-like measures. However, a main idea in these definitions is that the uncertainty about belongingness of a concerned element decreases when its membership value approaches either 0 or 1. A fuzzy set $\tilde{A}^*$ is called a sharpening of the fuzzy set $\tilde{A}$ if $\mu_{\tilde{A}^*}(x) \leq \mu_{\tilde{A}}(x)$ when $\mu_{\tilde{A}}(x) < 0.5$ and $\mu_{\tilde{A}^*}(x) \geq \mu_{\tilde{A}}(x)$ when $\mu_{\tilde{A}}(x) \geq 0.5$. The degree of fuzziness of $\tilde{A}^*$ should be lower than that of $\tilde{A}$ because sharpening reduces the uncertainty of belongingness. More formally, we have the following definition of entropy of finite discrete fuzzy sets, proposed by De Luca and Termini [3]:

$$H(\tilde{A}) = -\sum_{i=1}^{n} (\mu_i \ln \mu_i + (1 - \mu_i)\ln(1 - \mu_i))$$

where $n$ is the number of the focal elements in $X$. When dealing with the continuous fuzzy sets, we have the corresponding definition as follows:

$$H(\tilde{A}) = -\int_{-\infty}^{\infty} (\mu(x)\ln\mu(x) + (1 - \mu(x))\ln(1 - \mu(x)))dx.$$

From the definitions, De Luca and Termini [3] proved the following:

1) $H(\tilde{A}) = 0$ if and only if $\mu_{\tilde{A}}(x) = 0$ or $1, \forall\, x \in X$;
2) $H(\tilde{A})$ is maximum if and only if $\mu_{\tilde{A}} = 0.5, \forall\, x \in X$;
3) $H(\tilde{A}) \geq H(\tilde{A}^*)$, where $\tilde{A}^*$ is a sharpening of $\tilde{A}$;
4) $H(\tilde{A}) = H(\tilde{A}^c)$, where $\tilde{A}^c$ is the complement of $\tilde{A}$.

These properties of fuzzy entropy are usually treated as the basic requirements for measures of fuzziness sometimes with some variations; see, for instance, Klir and Yuan [18].

### B. Fuzzy Variables and Credibility Measures

A fuzzy variable, with a fuzzy restriction on its values, is closely related with the linguistic approach and is used to model humanistic systems such as in artificial intelligence, pattern recognition, medical diagnosis, and other related areas.

Let $\xi$ be a fuzzy variable with membership function $\mu(x)$ which satisfies the normalization condition, i.e., $\sup_x \mu(x) = 1$. In the setting of possibility theory, the possibility and necessity measures for the fuzzy event $\{\xi \in A\}$ deduced from $\mu(x)$ are given by

$$\mathrm{Pos}\{\xi \in A\} = \sup_{x \in A} \mu(x)$$
$$\mathrm{Nec}\{\xi \in A\} = 1 - \sup_{x \in A^c} \mu(x)$$

where $A$ is any subset of the real numbers $\mathcal{R}$. The set $\mathrm{Supp}\{\xi\} = \{x \in \mathcal{R} \mid \mu(x) > 0\}$ is called the support of $\xi$ and its elements are called *focal* elements. For each $x \in \mathrm{Supp}(\xi)$, the fuzzy event $\{\xi = x\}$ is called a basic fuzzy event and the function $\mathrm{Pos}\{\xi = x\}$, which is nothing else than $\mu(x)$, is also referred to as the possibility distribution of $\xi$.

As it is well known, a possibility measure only provides nonprobabilistic information on the occurrence of the fuzzy event $\{\xi \in A\}$ while the necessity measure quantifies the impossibility of its opposite event. Thus, the two measures together provide the total nonprobabilistic information about the given fuzzy event, which means that we should take both of them into account when measuring the uncertainty of fuzzy variables. A general idea to make it convenient is to construct a new measure which contains all the information characterized by possibility and necessity measures. Liu and Liu [25] proposed an alternative definition of measure, called *credibility measure*, to quantify the chance of occurrence of the fuzzy event $\{\xi \in A\}$ by

$$\mathrm{Cr}\{\xi \in A\} = \frac{1}{2}\left(\sup_{x\in A}\mu(x) + 1 - \sup_{x\in A^c}\mu(x)\right)$$

where $A$ is any subset of $\mathcal{R}$. This formula is also called *credibility inversion theorem*. Accordingly, the function $\mathrm{Cr}\{\xi = x\}$ is referred to as the credibility distribution of $\xi$.

Though a credibility measure is the arithmetic average of a possibility measure and a necessity measure, it indeed provides all the nonprobabilistic information about a fuzzy event. Both the possibility measure and the necessity measure can be derived, respectively, from a credibility measure in the following ways:

$$\mathrm{Pos}\{\xi \in A\} = \begin{cases} 2\mathrm{Cr}\{\xi \in A\}, & \text{if } \mathrm{Cr}\{\xi \in A\} \le 0.5 \\ 1, & \text{if } \mathrm{Cr}\{\xi \in A\} > 0.5 \end{cases}$$

and

$$\mathrm{Nec}\{\xi \in A\} = \begin{cases} 0, & \text{if } \mathrm{Cr}\{\xi \in A\} \le 0.5 \\ 2\mathrm{Cr}\{\xi \in A\} - 1, & \text{if } \mathrm{Cr}\{\xi \in A\} > 0.5. \end{cases}$$

In addition, we have $\mathrm{Cr}\{\xi \in A\} + \mathrm{Cr}\{\xi \in A^c\} = 1$, i.e., the credibility measure is self dual, which makes it useful and convenient in applications.

A fuzzy variable $\xi$ is said to be discrete if there exists a countable sequence $\{x_1, x_2, \ldots\}$ such that $\mathrm{Pos}\{\xi \neq x_1, \xi \neq x_2, \ldots\} = 0$, i.e., the membership function of $\xi$ takes nonzero values at a countable set of points. Furthermore, a fuzzy variable $\xi$ is said to be simple if there exists a finite sequence $\{x_1, x_2, \ldots, x_n\}$ such that $\mathrm{Pos}\{\xi \neq x_1, \xi \neq x_2, \ldots, \xi \neq x_n\} = 0$, i.e., the membership function of $\xi$ takes nonzero values at a finite number of points. Besides, a fuzzy variable $\xi$ is said to be continuous if $\mathrm{Pos}\{\xi = x\}$, equivalently, the membership function $\mu(x)$ is a continuous function of $x$.

The most frequently used continuous fuzzy variables in applications are trapezoidal fuzzy variables and triangular fuzzy variables. A trapezoidal fuzzy variable can be fully determined by a quadruplet $(r_1, r_2, r_3, r_4)$ of crisp numbers with $r_1 \le r_2 \le r_3 \le r_4$, whose membership function is defined as

$$\mu(x) = \begin{cases} \dfrac{x - r_1}{r_2 - r_1}, & \text{if } r_1 \le x \le r_2 \\ 1, & \text{if } r_2 \le x \le r_3 \\ \dfrac{x - r_4}{r_3 - r_4}, & \text{if } r_3 \le x \le r_4 \\ 0, & \text{otherwise} \end{cases}.$$

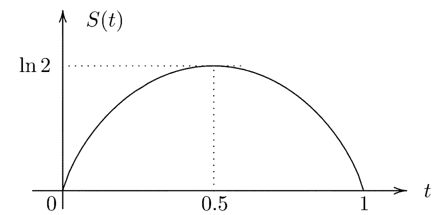



Fig. 1. Shape of function $S(t)$.

Similarly, a triangular fuzzy variable can be fully determined by a triplet $(r_1, r_2, r_3)$ of crisp numbers with $r_1 \le r_2 \le r_3$, whose membership function is defined as

$$\mu(x) = \begin{cases} \dfrac{x - r_1}{r_2 - r_1}, & \text{if } r_1 \le x \le r_2 \\ \dfrac{x - r_3}{r_2 - r_3}, & \text{if } r_2 \le x \le r_3 \\ 0, & \text{otherwise} \end{cases}.$$

For the up-to-date credibility theory, the reader may consult Liu [22].

## III. Entropy of Discrete Credibility Distributions

As we have pointed out, measuring the uncertainty of a fuzzy variable should concern the number of all the focal elements of the fuzzy variable and its credibility distribution. The more focal elements there are, the more values the fuzzy variable can take, thus the more uncertainty. On the other hand, the credibility distribution indicates the difficulty in predicting the specified value that the fuzzy variable takes. In this section, based on the notion of credibility distribution, we formulate a definition of entropy to quantify the uncertainty of fuzzy variables. The measure is called entropy due to its entropy-like functional form borrowed from the definition of De Luca and Termini, and in essence, its underlying meaning. The entropy of a credibility distribution, that has a certain resemblance to that of a probability distribution, characterizes the uncertainty resulting from information deficiency which is caused by the impossibility to predict the specified value that a fuzzy variable takes.

For convenience, we denote $S(t) = -t\ln t - (1-t)\ln(1-t)$ with the convention that $0 \cdot \ln 0 = 0$. It is well known that $S(t)$ is strictly concave on $[0, 1]$ and symmetrical about $t = 0.5$. The shape of $S(t)$ is shown in Fig. 1.

Using the function $S(t)$, we give the following definition of entropy of discrete credibility distributions.

*Definition 1:* Let $\xi$ be a discrete fuzzy variable taking values in $\{x_1, x_2, \ldots\}$. Then, the entropy of its credibility distribution is defined by

$$H(\xi) = \sum_{i=1}^{\infty} S(\mathrm{Cr}\{\xi = x_i\}).$$

Especially, if $\xi$ is a simple fuzzy variable taking values in $\{x_1, x_2, \ldots, x_n\}$, then the entropy of its credibility distribution is

$$H(\xi) = \sum_{i=1}^{n} S(\mathrm{Cr}\{\xi = x_i\})$$

where $S(t) = -t\ln t - (1-t)\ln(1-t)$.

*Property 1 (Symmetry):* Let $\xi$ be a simple fuzzy variable that takes values in $\{x_1, x_2, \ldots, x_n\}$ with possibilities $\{\mu_1, \mu_2, \ldots, \mu_n\}$, and let $\eta$ be a simple fuzzy variable that takes values in $\{x_1, x_2, \ldots, x_n\}$ with possibilities $\{\nu_1, \nu_2, \ldots, \nu_n\}$, where $\{\nu_1, \nu_2, \ldots, \nu_n\}$ is a rearrangement of the sequence $\{\mu_1, \mu_2, \ldots, \mu_n\}$. Then, we have $H(\xi) = H(\eta)$.

*Proof:* The property follows immediately from the definition of entropy.

The symmetry property states that the entropy is indifferent to the values of focal elements and invariant with respect to permutations of credibilities of a given distribution. It is the reason that the entropy can be essentially defined on the credibility distributions of fuzzy variables.

*Lemma 1:* Let $\xi$ be a discrete fuzzy variable taking values in $\{x_1, x_2, \ldots\}$. If there exists an index $k$ such that $\mathrm{Cr}\{\xi = x_k\} = 1$, then we have $\mathrm{Cr}\{\xi = x_i\} = 0, \forall\, i \neq k$.

*Proof:* Let $\mu_i$ be the possibility that $\xi$ takes the value $x_i$, i.e., $\mu_i = \mathrm{Pos}\{\xi = x_i\}$, $i = 1, 2, \ldots$, respectively. Since $\mathrm{Cr}\{\xi = x_k\} = (\mu_k + 1 - \max_{i \neq k} \mu_i)/2$, we have $\mu_k = 1$ and $\mu_i = 0$, $i \neq k$. Thus, $\mathrm{Cr}\{\xi = x_i\} = (\mu_i + 1 - \max_{j \neq i} \mu_j)/2 = 0$, $i \neq k$, which proves the lemma.

Lemma 1 states that the basic fuzzy event $\{\xi = x_k\}$ would always hold if $\mathrm{Cr}\{\xi = x_k\} = 1$. In this case, the fuzzy variable $\xi$ is essentially a crisp number and, therefore, contains no uncertainty. It follows directly from Lemma 1 that there do not exist two basic fuzzy events, say $\{\xi = x_i\}$ and $\{\xi = x_j\}$, such that $\mathrm{Cr}\{\xi = x_i\} = 1$ and $\mathrm{Cr}\{\xi = x_j\} > 0$, simultaneously.

*Property 2 (Decisivity):* Let $\xi$ be a discrete fuzzy variable taking values in $\{x_1, x_2, \ldots\}$. Then, the entropy

$$H(\xi) \geq 0$$

and equality holds if and only if there exists an index $k$ such that the membership function of $\xi$ is

$$\mu(x_i) = \begin{cases} 1, & \text{if } i = k \\ 0, & \text{otherwise} \end{cases}$$

i.e., $\xi$ is essentially a crisp number in this case.

*Proof:* It follows immediately from the definition of entropy that $H(\xi) \geq 0$. If $H(\xi) = 0$, we have $\mathrm{Cr}\{\xi = x_i\} = 0$ or 1 for each $i$. According to Lemma 1, there exists one and only one index $k$ with $\mathrm{Cr}\{\xi = x_k\} = 1$, i.e., the membership function of $\xi$ is

$$\mu(x_i) = \begin{cases} 1, & \text{if } i = k \\ 0, & \text{otherwise} \end{cases}.$$

On the other hand, if $\xi$ is essentially a crisp number, it has the membership function as the previous form. It is easy to verify that the entropy $H(\xi) = 0$.

*Property 3 (Maximality):* Let $\xi$ be a simple fuzzy variable taking values in $\{x_1, x_2, \ldots, x_n\}$. Then, the entropy

$$H(\xi) \leq n \ln 2$$

and equality holds if and only if its membership function $\mu(x_i) = 1$ for all $i = 1, 2, \ldots, n$.

*Proof:* Since the function $S(t)$ reaches its maximum $\ln 2$ at $t = 0.5$, we have

$$H(\xi) = \sum_{i=1}^{n} S(\mathrm{Cr}\{\xi = x_i\}) \leq n \ln 2$$

and equation holds if and only if $\mathrm{Cr}\{\xi = x_i\} = 0.5$, $i = 1, 2, \ldots, n$, which is equivalent to $\mu(x_i) = 1$ for all $i = 1, 2, \ldots, n$.

The maximality property states that the entropy of a credibility distribution reaches its maximum when all the basic fuzzy events have the equal possibility of occurrence. In this case, there is no preference between all the values that the fuzzy variable could take.

*Property 4 (Expansibility):* Let $\xi$ be a simple fuzzy variable that takes values in $\{x_1, x_2, \ldots, x_n\}$ with possibilities $\{\mu_1, \mu_2, \ldots, \mu_n\}$, and let $\eta$ be a simple fuzzy variable that takes values in $\{x_1, x_2, \ldots, x_n, x_{n+1}\}$ with possibilities $\{\mu_1, \mu_2, \ldots, \mu_n, \mu_{n+1}\}$. Then, we have $H(\xi) \leq H(\eta)$ with equality when $\mu_{n+1} = 0$.

*Proof:* The theorem follows directly from the definition of entropy and the fact that

$$\mathrm{Cr}\{\xi = x_{n+1}\} = \frac{1}{2}\left(\mu_{n+1} + 1 - \max_{1 \leq i \leq n} \mu_i\right) = \frac{\mu_{n+1}}{2} \geq 0$$

where the equality holds when $\mu_{n+1} = 0$.

The expansibility property states that the entropy only relates to the possibilities (or credibilities) of all the focal elements and increases with the number of the focal elements. Suppressing an outcome with zero possibility (or credibility) does not change the entropy. The following monotonicity property indicates that the entropy also increases with the possibility degrees of the focal elements.

*Property 5 (Monotonicity):* Let $\xi$ be a simple fuzzy variable that takes values in $\{x_1, x_2, \ldots, x_n\}$ with possibilities $\{\mu_1, \mu_2, \ldots, \mu_n\}$, and let $\eta$ be a simple fuzzy variable that takes values in $\{x_1, x_2, \ldots, x_n\}$ with possibilities $\{\nu_1, \nu_2, \ldots, \nu_n\}$. If $\mu_i \leq \nu_i$ for all $i = 1, 2, \ldots, n$, we have $H(\xi) \leq H(\eta)$. Furthermore, if there exists an index $k$ such that $\mu_k < \nu_k$, we have $H(\xi) < H(\eta)$.

*Proof:* Denote the credibilities $\mathrm{Cr}\{\xi = x_i\}$ and $\mathrm{Cr}\{\eta = x_i\}$ by

$$c_i = \frac{1}{2}\left(\mu_i + 1 - \max_{j \neq i} \mu_j\right) \quad \text{and} \quad c_i' = \frac{1}{2}\left(\nu_i + 1 - \max_{j \neq i} \nu_j\right)$$

for $i = 1, 2, \ldots, n$, respectively. The argument breaks down into the following two cases.

Case 1) $\mu_i = 1$. For this case, we have $\nu_i = 1$ and $\max_{j \neq i} \mu_j \leq \max_{j \neq i} \nu_j$. Thus, we have

$$1 - \max_{j \neq i} \mu_j \geq 1 - \max_{j \neq i} \nu_j$$

and

$$\begin{aligned} 0.5 &\leq \frac{1}{2}\left(\nu_i + 1 - \max_{j \neq i} \nu_j\right) \\ &\leq \frac{1}{2}\left(\mu_i + 1 - \max_{j \neq i} \mu_j\right). \end{aligned}$$

That is, $0.5 \le c_i' \le c_i$, and then, the inequality $S(c_i) \le S(c_i')$ holds.

Case 2) $\mu_i < 1$. For this case, there exists an index $k \neq i$ such that $\mu_k = 1$ and $\nu_k = 1$ due to normalization, i.e., $\max_{1\le i\le n} \mu_i = \max_{1\le i\le n} \nu_i = 1$. Thus, we have $c_i = \mu_i/2$ and

$$c_i' = \frac{1}{2}\left(\nu_i + 1 - \max_{j\neq i} \nu_j\right) = \frac{\nu_i}{2} \le 0.5.$$

That is, $c_i \le c_i' \le 0.5$, and then, the inequality $S(c_i) \le S(c_i')$ holds.

Now, we have $S(c_i) \le S(c_i')$ for all $i = 1, 2, \ldots, n$, and thus

$$H(\xi) = \sum_{i=1}^{n} S(c_i) \le \sum_{i=1}^{n} S(c_i') = H(\eta).$$

In addition, if there exists an index $k$ such that $\mu_k < \nu_k$ holds in either of the previous cases, we have $S(c_i) < S(c_i')$, and thus, $H(\xi) < H(\eta)$.

Generally, the set of credibility distributions over a given discrete support set does not form a convex set, i.e., the convex combination of two credibility distributions is not necessarily a credibility distribution. However, it is the case when the support set contains only two focal elements. Considering a convex subset of credibility distributions, we have the following property of strict concavity.

*Property 6 (Strict Concavity):* Let $\xi$ be a simple fuzzy variable that takes values in $\{x_1, x_2, \ldots, x_n\}$ with possibilities $\{\mu_1, \mu_2, \ldots, \mu_n\}$, and let $\eta$ be a simple fuzzy variable that takes values in $\{x_1, x_2, \ldots, x_n\}$ with possibilities $\{\nu_1, \nu_2, \ldots, \nu_n\}$. Then, we have

$$\lambda H(\xi) + (1-\lambda)H(\eta) < \sum_{i=1}^{n} S(\lambda c_i + (1-\lambda)c_i') \quad \forall\, \lambda \in (0,1)$$

where $c_i = \mathrm{Cr}\{\xi = x_i\}$ and $c_i' = \mathrm{Cr}\{\eta = x_i\}$ for $i = 1, 2, \ldots, n$, respectively.

*Proof:* The inequality follows directly from the strict concavity of the function $S(t)$.

An immediate consequence of the strict concavity property is that maximizing the entropy over a convex subset of discrete credibility distributions always leads to a unique maximum. It may give rise to the maximum entropy principle, i.e., one should choose a credibility distribution with maximum entropy when only partial information is available. However, to make sense, one should work on a convex subset of credibility distributions, which may be not necessarily easy identifiable in practice.

Until now, we have proposed the definition of entropy of discrete credibility distributions and investigated some properties of this definition. Now, we give a numerical example to illustrate the calculation of the entropy.

*Example 1:* Suppose that the age of Tom is about 20 and the term "about 20" in this case can be modeled as

$$\left\{\frac{17}{0.1}, \frac{18}{0.5}, \frac{19}{0.8}, \frac{20}{1}, \frac{21}{0.8}, \frac{22}{0.5}, \frac{23}{0.1}\right\}$$

where the denotation (17/0.1) represents that the integer 17 is a possible value of the occurrence and 0.1 is the membership that expresses the degree of compatibility between the age 17 and the term "about 20." Let $\xi$ be the age of Tom. Thus, $\xi$ is a simple fuzzy variable taking values in $\{17, 18, 19, 20, 21, 22, 23\}$.

Computing the credibility distribution of $\xi$, we get the result $\{0.05, 0.25, 0.4, 0.6, 0.4, 0.25, 0.05\}$. Thus, the entropy $H(\xi) = 3.4894$. According to Property 3, the maximum entropy of a credibility distribution with seven focal elements is $7\ln 2$. We can normalize the result as

$$H^*(\xi) = \frac{H(\xi)}{7\ln 2} = 0.7192.$$

## IV. Entropy of Continuous Credibility Distributions

In applications, membership functions of fuzzy variables are often continuous functions defined on the set of real numbers. To deal with this situation, we proposed a definition of entropy of continuous credibility distributions, which is a natural counterpart of the definition of entropy in the discrete case.

*Definition 2:* Let $\xi$ be a continuous fuzzy variable. Then, the entropy of its credibility distribution is defined by

$$H(\xi) = \int_{-\infty}^{\infty} S(\mathrm{Cr}\{\xi = x\})dx$$

if it exists, where $S(t) = -t\ln t - (1-t)\ln(1-t)$.

Here, we note that the credibility distribution $\mathrm{Cr}\{\xi = x\}$ is nonzero only in the set $\mathrm{Supp}(\xi)$, i.e., the entropy only relates to the support of a fuzzy variable.

*Property 7:* Let $\xi$ be a continuous fuzzy variable taking values on the interval $[a, b]$. We have

$$0 < H(\xi) \le (b-a)\ln 2.$$

*Proof:* The property follows from the fact that $\mathrm{Cr}\{\xi = x\} = \mu(x)/2$ and that the function $S(t)$ reaches its maximum $\ln 2$ at $t = 0.5$.

When a continuous fuzzy variable degenerates into a crisp number, its entropy also tends to the minimum 0. However, the crisp number is not a continuous fuzzy variable.

*Property 8:* Let $\xi$ and $\eta$ be two continuous fuzzy variables with membership functions $\mu(x)$ and $\nu(x)$, respectively. If $\mu(x) \le \nu(x)$ for all $x \in \mathcal{R}$, we have $H(\xi) \le H(\eta)$. Furthermore, if the entropies are finite and there exists $x_0 \in \mathcal{R}$ such that $\mu(x_0) < \nu(x_0)$, we have $H(\xi) < H(\eta)$.

*Proof:* Since $\mu(x) \le \nu(x)$ for all $x \in \mathcal{R}$, we have $\mathrm{Cr}\{\xi = x\} \le \mathrm{Cr}\{\eta = x\} \le 0.5$. Thus, $S(\mathrm{Cr}\{\xi = x\}) \le S(\mathrm{Cr}\{\eta = x\})$, and then, $H(\xi) \le H(\eta)$. In addition, if $H(\xi)$ and $H(\eta)$ are finite and there exists $x_0 \in \mathcal{R}$ such that $\mu(x_0) < \nu(x_0)$, there must exist an open interval $(a, b)$ containing $x_0$ such that $\mu(x) < \nu(x)$ for all $x \in (a, b)$. Thus, $S(\mathrm{Cr}\{\xi = x\}) < S(\mathrm{Cr}\{\eta = x\})$ for all $x \in (a, b)$ and then

$$\int_a^b S(\mathrm{Cr}\{\xi = x\})dx < \int_a^b S(\mathrm{Cr}\{\eta = x\})dx$$

which means $H(\xi) < H(\eta)$.

*Property 9:* Let $\xi$ be a continuous fuzzy variable with membership function $\mu(x)$. We have

$$H(p\xi + q) = |p|H(\xi)$$

where the fuzzy variable $p\xi + q$ is a linear transform of $\xi$, whose membership function is $\mu((x-q)/p)$, $p,\ q \in \Re$, and $p \neq 0$.

*Proof:* Straightforward.

Now, we have defined the entropy of continuous credibility distributions in a general way of extension and investigated some similar properties. However, the entropy in continuous case may exhibit its particular characteristics. Property 9 indicates that the entropy of a continuous credibility distribution changes after a linear transform while the entropy of a discrete credibility distribution is invariant to the linear transforms. Therefore, further investigation is necessary on the entropy in the continuous situation.

With the definition of entropy of continuous credibility distributions, we can calculate the entropies for trapezoidal fuzzy variables and triangular fuzzy variables.

*Example 2:* Let $\xi$ be a trapezoidal fuzzy variable $(r_1, r_2, r_3, r_4)$. Then, the entropy is

$$H(\xi) = \frac{1}{2}(r_4 - r_1) + \left(\ln 2 - \frac{1}{2}\right)(r_3 - r_2).$$

*Example 3:* Let $\xi$ be a triangular fuzzy variable $(r_1, r_2, r_3)$. Then, the entropy is

$$H(\xi) = \int_{r_1}^{r_3} S(\mathrm{Cr}\{\xi = x\})dx = \frac{1}{2}(r_3 - r_1).$$

According to the calculation, we note that if a sequence of trapezoidal fuzzy variables converges to a triangular fuzzy variable, the entropy will also converge to that of the corresponding triangular distribution.

## V. Conclusion

A definition of entropy of credibility distributions is proposed in this paper to measure the uncertainty associated with fuzzy variables. Some properties of the entropy are investigated in detail. The proposed definition of entropy shares some critical properties with the measure of U-uncertainty defined by Higashi and Klir [14], and therefore, can be considered as an alternative measure of nonspecificity. Compared with the measure of U-uncertainty which was defined on normalized ordered possibility distributions, the proposed definition of entropy can be calculated directly with the notion credibility measure. This might make the calculation of the entropy somewhat convenient. However, measures of uncertainty that satisfy the desired properties are not unique. A function with properties similar to the function $S(t)$ may serve as a substitute and provide a new definition for measuring the uncertainty of fuzzy variables. An example is that the quadratic function $Q(t) = t(1-t)$, $t \in [0,1]$ can be used to define the quadratic entropy of credibility distributions

$$H_Q(\xi) = \sum_{i=1}^{\infty} Q(\mathrm{Cr}\{\xi = x_i\}).$$

This definition is inspired by Havrda and Charvát's nonadditive entropy of order 2 [12] and has a similar form to Vajda's probabilistic quadratic entropy [39]. It can be verified that the quadratic entropy of credibility distributions satisfies all the properties mentioned in this paper. Therefore, an axiomatical approach is needed to characterize the proposed entropy. However, the construction of a suitable set of axioms would be difficult and some requirements could be inconsistent. Intuitively, besides the properties of symmetry and decisivity, a fundamental property is that the entropy should increase with the number of focal elements or with the increase of the possibility degrees, which has been illustrated respectively in Properties 4 and 5 in this paper. Also, the concept of information gain associated with the proposed definition of entropy should be well investigated since the properties of subadditivity and additivity do not hold in this case. Some related discussion can be found in Ramer [36]. On the other hand, it is necessary to investigate the computational performance of the proposed definition of entropy compared with other evaluation measures in applications. Some similar work can be found in Borgelt and Kruse [2] and Marsala and Bouchon-Meunier [27].

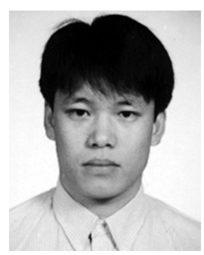
## Acknowledgment

The authors would like to thank the anonymous referees and the Editor-in-Chief for their valuable suggestions, which have made substantial improvements to this paper.

## References

[1] J. C. Bezdek, "Fuzzy mathematics in pattern classification," Ph.D. dissertation, Cntr. Appl. Math., Cornell Univ., Ithaca, NY, 1973.
[2] C. Borgelt and R. Kruse, "Some experimental results on learning probabilistic and possibilistic networks with different evaluation measures," in *Proc. 1st Int. Joint Conf. Qualitative Quant. Practical Reason.*, 1997, vol. 1244, pp. 71–85.
[3] A. De Luca and S. Termini, "A definition of nonprobabilistic entropy in the setting of fuzzy sets theory," *Inf. Control*, vol. 20, pp. 301–312, 1972.
[4] D. Dubois and H. Prade, "A class of fuzzy measures based on triangular norms," *Int. J. General Syst.*, vol. 8, pp. 43–61, 1982.
[5] D. Dubois and H. Prade, *Théorie des Possibilités*. Paris, France: Masson, 1985.
[6] D. Dubois and H. Prade, "A note on measures of specificity for fuzzy sets," *Int. J. General Syst.*, vol. 10, pp. 279–283, 1985.
[7] B. R. Ebanks, "On measures of fuzziness and their representations," *J. Math. Anal. Appl.*, vol. 98, pp. 24–37, 1983.
[8] J. L. Fan and Y. L. Ma, "Some new fuzzy entropy formulas," *Fuzzy Sets Syst.*, vol. 128, pp. 277–284, 2002.
[9] L. Garmendia, R. R. Yager, E. Trillas, and A. Salvador, "On t-norms based measures of specificity," *Fuzzy Sets Syst.*, vol. 133, pp. 237–248, 2003.
[10] J. Gebhardt and R. Kruse, "Learning possibilistic networks from data," in *Proc. 5th Int. Workshop Artif. Intell. Statist.*, Fort Lauderdale, FL, 1995, pp. 233–244.
[11] J. Gebhardt and R. Kruse, "Tightest hypertree decompositions of multivariate possibility distributions," in *Proc. Int. Conf. Inf. Process. Manage. Uncertainty Knowl.-Based Syst.*, Granada, Spain,, 1996, pp. 923–927.
[12] J. Havrda and F. Charvát, "Quatification method of classsification processes, the concept of structural $\alpha$-entropy," *Kybernetika*, vol. 3, pp. 30–35, 1967.
[13] M. Higashi and G. J. Klir, "On measures of fuzziness and fuzzy complements," *Int. J. General Syst.*, vol. 8, pp. 169–180, 1982.
[14] M. Higashi and G. J. Klir, "Measures of uncertainty and information based on possibility distributions," *Int. J. General Syst.*, vol. 9, pp. 43–58, 1983.
[15] A. Kaufmann, *Introduction to the Theory of Fuzzy Subsets*. New York: Academic, 1975.

[16] G. J. Klir, *Uncertainty and Information: Foundations of Generalized Information Theory*. New York: Wiley, 2006.
[17] G. J. Klir and A. Ramer, "Uncertainty in Dempster-Shafer theory: A critical re-examination," *Int. J. General Syst.*, vol. 18, pp. 155–166, 1990.
[18] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic, Theory and Applications*. Upper Saddle River, NJ: Prentice-Hall, 1995.
[19] J. Knopfmacher, "On measures of fuzziness," *J. Math. Anal. Appl.*, vol. 49, pp. 529–534, 1975.
[20] T. Kroupa, "On application of Choquet integral in possibilistic information theory," *Neural Netw. World*, vol. 13, pp. 541–548, 2003.
[21] T. Kroupa, "Application of the Choquet integral to measures of information in possibility theory," *Int. J. Intell. Syst.*, vol. 21, pp. 349–359, 2006.
[22] B. Liu, *Foundation of Uncertainty Theory*. Berlin, Germany: Springer-Verlag, 2007.
[23] B. Liu, "Toward fuzzy optimization without mathematical ambiguity," *Fuzzy Optim. Decision Making*, vol. 1, pp. 43–63, 2002.
[24] B. Liu, *Uncertainty Theory: An Introduction to Its Axiomatic Foundation*. Berlin, Germany: Springer-Verlag, 2004.
[25] B. Liu and Y. K. Liu, "Expected value of fuzzy variable and fuzzy expected value models," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 4, pp. 445–450, Aug. 2002.
[26] S. G. Loo, "Measures of fuzziness," *Cybernetica*, vol. 20, pp. 201–210, 1977.
[27] C. Marsala and B. Bouchon-Meunier, "Choice of a method for the construction of fuzzy decision trees," in *Proc. 12th IEEE Int. Conf. Fuzzy Syst.*, St. Louis, MI, 2003, pp. 584–589.
[28] S. Nahmias, "Fuzzy variables," *Fuzzy Sets Syst.*, vol. 1, pp. 97–110, 1978.
[29] H. T. Nguyen, "Toward a calculus of the mathematical notion of possibility," in *Advances in Fuzzy Set Theory and Applications*, M. M. Gupta, R. K. Ragade, and R. R. Yager, Eds. New York: North-Holland, 1979, pp. 235–246.
[30] N. R. Pal and S. K. Pal, "Object-background segmentation using new definitions of entropy," *Inst. Electr. Eng. Proc.—Comput. Digital Technol.*, vol. 136, pp. 284–295, 1989.
[31] N. R. Pal and S. K. Pal, "Higher order fuzzy entropy and hybrid entropy of a set," *Inf. Sci.*, vol. 61, pp. 211–231, 1992.
[32] N. R. Pal, J. C. Bezdek, and R. Hemasinha, "Uncertainty measures for evidential reasoning I: A review," *Int. J. Approx. Reason.*, vol. 7, pp. 165–183, 1992.
[33] N. R. Pal, J. C. Bezdek, and R. Hemasinha, "Uncertainty measures for evidential reasoning II: A new measure," *Int. J. Approx. Reason.*, vol. 8, pp. 1–16, 1993.
[34] N. R. Pal and J. C. Bezdek, "Measuring fuzzy uncertainty," *IEEE Trans. Fuzzy Syst.*, vol. 2, no. 1, pp. 107–118, Feb. 1994.
[35] N. R. Pal and J. C. Bezdek, "Quantifying different facets of fuzzy uncertainty," in *Fundamentals of Fuzzy Sets*, D. Dubois and H. Prade, Eds. London, U.K.: Kluwer, 2000, pp. 459–480.
[36] A. Ramer, "Axioms of uncertainty measures: Dependence and independence," *Fuzzy Sets Syst.*, vol. 35, pp. 185–196, 1990.
[37] A. Ramer and L. Lander, "Classification of possibilistic uncertainty and information functions," *Fuzzy Sets Syst.*, vol. 24, pp. 221–230, 1987.
[38] C. Ricotta, "On possible measures for evaluating the degree of uncertainty of fuzzy thematic maps," *Int. J. Remote Sens.*, vol. 26, pp. 5573–5583, 2005.
[39] I. Vajda, "Bounds on the minimal error probability and checking a finite or countable number of hypothsis," *Inf. Transmission Problems*, vol. 4, pp. 9–17, 1968.
[40] N. E. Wu and G. J. Klir, "Optimal redundancy management in reconfigurable control systems based on normalized nonspecificity," *Int. J. Syst. Sci.*, vol. 31, pp. 797–808, 2000.
[41] R. R. Yager, "On measures of fuzziness and negation, part I: Membership in the unit interval," *Int. J. General Syst.*, vol. 5, pp. 221–229, 1979.
[42] R. R. Yager, "On measures of fuzziness and negation, part II: Lattices," *Inf. Control*, vol. 44, pp. 236–260, 1980.
[43] R. R. Yager, "A foundation for a theory of possibility," *J. Cybern.*, vol. 10, pp. 177–204, 1980.
[44] R. R. Yager, "Measuring tranquility and anxiety in decision making: An application of fuzzy sets," *Int. J. General Syst.*, vol. 8, pp. 139–144, 1982.
[45] R. R. Yager, "Entropy and specificity in a mathematical theory of evidence," *Int. J. General Syst.*, vol. 9, pp. 249–260, 1983.
[46] R. R. Yager, "On the specificity of a possibility distribution," *Fuzzy Sets Syst.*, vol. 50, pp. 279–292, 1992.
[47] R. R. Yager, "On the entropy of fuzzy measures," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 4, pp. 453–461, Aug. 2000.
[48] D. S. Yeung and X. Z. Wang, "Improving performance of similarity-based clustering by feature weight learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 556–561, Apr. 2002.
[49] Y. Yuan and M. J. Shaw, "Induction of fuzzy decision trees," *Fuzzy Sets Syst.*, vol. 69, pp. 125–139, 1995.
[50] L. A. Zadeh, "Fuzzy sets and systems," in *Proc. Symp. Syst. Theory, Polytech. Inst. Brooklyn*, New York,, 1965, pp. 29–37.
[51] L. A. Zadeh, "Probability measures of fuzzy events," *J. Math. Anal. Appl.*, vol. 23, pp. 421–427, 1968.
[52] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets Syst.*, vol. 1, pp. 3–28, 1978.



**Pingke Li** was born in Zhejiang, China, in 1979. He received the B.S. degree in control science from the Department of Mathematics, Fudan University, Shanghai, China, in 2002 and the M.S. degree in operations research from the Department of Mathematical Sciences, Tsinghua University, Beijing, China, in 2005. Currently, he is working toward the Ph.D. degree at the North Carolina State University, Raleigh.

**Baoding Liu** (M'99–SM'00) graduated from the Department of Mathematics, Nankai University, Tianjin, China, in 1986, and received the M.S. and Ph.D. degrees in operations research and control theory from the Chinese Academy of Sciences, Beijing, China, in 1989 and 1993, respectively.

He joined Tsinghua University, Tsinghua, China, as an Associate Professor in 1996, and was appointed Professor of Mathematics in 1998. He has published over 100 papers, and authored seven books, including *Uncertainty Theory* (Springer: New York, 2004) and *Theory and Practice of Uncertain Programming* (Physica-Verlag: Heidelberg, Germany, 2002).

Dr. Liu is serving on the editorial boards of several international journals, including Associate Editor of the IEEE TRANSACTIONS ON FUZZY SYSTEMS and Book Review Editor of *Fuzzy Optimization and Decision Making*.