# Higher Order Fuzzy Entropy and Hybrid Entropy of a Set

N. R. PAL*

and

S. K. PAL†

*Indian Statistical Institute, 203 B.T. Road, Calcutta 700 035, India*

## ABSTRACT

New entropy measures such as higher order fuzzy entropy and hybrid entropy of a set have been introduced. Various properties along with their proofs have been included. Applicability of these new measures to various problems has been highlighted.

## 1. INTRODUCTION

The present work consists of three parts. The existing fuzzy entropy measures have been critically analyzed in the first part. In the second part, a need for a higher order entropy $H^r$ ($r \geq 1$), which will give a measure of average uncertainty associated with any arbitrary subset with $r$-supports, has been established. The new definition has been introduced in such a way that when $r = 1$, the definition boils down to that of the existing entropy of a fuzzy set. Various attempts, available in the literature, to combine the probabilistic and fuzzy (possibilistic) entropy have been critically reviewed and a new definition of hybrid entropy ($H_{hy}$) has been introduced in the third part of the paper. This measure can be regarded as a generalized entropy of a set such that when the fuzziness is removed, the measure turns to be the classical entropy of the set.

---

*With the Computer Science Unit.
†With the Electronics and Communication Sciences Unit.

Various properties of these measures have been stated and proved. The higher order entropy $H'$ is found to possess some interesting properties which can be used in feature evaluation and image segmentation problems. As expected, $H'$ conveys more information regarding actual structure of the set than $H^1$ (conventional fuzzy entropy) does. $H_{hy}$, on the other hand, can be considered an objective function for proper defuzzification (enhancement) of a set. These characteristics have been further demonstrated with examples.

## 2.  CLASSICAL ENTROPY

Shannon [1, 2] defined the term entropy to represent a quantitative measure of information of a probabilistic system. He quantified information as a logarithmic function of the probability distribution. Let $p_i$ be the probability of occurrence of the $i$th state of an $n$-state system, $0 \leqslant p_i \leqslant 1$, $\Sigma p_i = 1$. The gain in information from the occurrence of the $i$th state of such a system is defined as $\log(1/p_i)$, and the expected value of the gain function is called the entropy of the system. Thus the entropy is

$$H = \sum_{i=1}^{n} p_i \log(1/p_i).$$
$$(1)$$

Recently, Pal and Pal [3] have pointed out some justification in favor of taking $\exp(1 - p_i)$ as the gain in information from the occurrence of the $i$th state of a probabilistic system. Therefore, a new definition of entropy takes the form

$$H = \sum_{i=1}^{n} p_i \exp(1 - p_i).$$
$$(2)$$

It should be mentioned here that both of the above measures give the average amount of information that would be gained from the occurrence of any arbitrary state of a probabilistic system.

## 3.  ENTROPY MEASURES OF FUZZY SETS

Zadeh [4] introduced the concept of fuzzy sets in which imprecise knowledge can be used to define an event. A fuzzy set $A$ is represented as

$$A = \{ x_i / \mu_A(x_i), i = 1, 2, \ldots, n \},$$
$$(3)$$

where $\mu_A(x_i)$ gives the degree of belongingness of the element $x_i$ to the set $A$.

In the case of an ordinary set an element either belongs to or does not belong to a set. Thus an ordinary set $A$ is defined by the characteristic function $C_A: R^n \to \{0,1\}$. On the other hand, a fuzzy set $A$ is defined by a characteristic function

$$\mu_A: R^n \to [0,1].$$

The function $\mu_A$ associates with each $x \in R^n$ a grade of membership to the set $A$ and is known as the membership function.

Similarly, a property $P$ defined on an event $x_i$ is a function $P(x_i)$, which can have values only in the interval $[0,1]$. A set of these functions which assigns the degree of possessing some property $P$ by the event $x_i$ constitutes what is called a property set.

In the areas of pattern recognition, image processing, speech recognition etc., it is often required to get some idea about the degree of ambiguity (fuzziness) present in a fuzzy set. A measure of fuzziness is expected to give the average amount of difficulty that is involved in taking a decision whether an element belongs to the set or not [5–7]. It is to be noted that this is not a property of the individual element of the set but a property of the set as a whole. Therefore, if every element of the set has a membership value of 0.5, then the fuzziness in the set should be maximum. There have been different definitions of entropy for a fuzzy set to provide measures of fuzziness.

Zadeh [4] defined the entropy of a fuzzy subset $A$ for the finite set $\{x_1, x_2, \ldots, x_n\}$ with respect to the probability distribution $\{p_1, p_2, \ldots, p_n\}$ as

$$H^P = - \sum_{i=1}^{n} \mu_A(x_i) p_i \log(p_i), \qquad (4)$$

where $\mu_A$ is the membership function of $A$, and $p_i$ is the probability of occurrence of $x_i$. Zadeh in [4] did not clarify the physical significance of $H^P$; even he did not mention whether $H^P$ can be used as a measure of fuzziness or not.

This $H^P$ can be viewed as a weighted Shannon entropy where the membership values are taken as the weight. Let us now see how the use of $H^P$, as a measure of fuzziness, leads to an undesirable situation. Consider a fuzzy set with two elements $x_1, x_2$ with probabilities $p_1, p_2$ as shown in Table 1.

Since probabilities are same for both the cases, the fuzziness is expected to be higher for case 1; but this is not reflected in $H^P$. This example therefore reveals the difficulty in using $H^P$ as a measure of fuzziness.

TABLE 1

Value of Zadeh's Entropy

| Case | $\mu_i$ | $p_i$ | $H^P$ |
|------|---------|-------|-------|
| 1    | 0.5     | 0.01  | 0.028007 |
|      | 0.5     | 0.99  |          |
| 2    | 0.8     | 0.01  | 0.44800  |
|      | 0.9     | 0.99  |          |

Kaufmann [5] defined the entropy of a fuzzy set with $n$ supports as

$$H^K = \{-1/\log(n)\} \sum_{i=1}^{n} \phi_i \log(\phi_i), \qquad (5)$$

where $\phi_i = \mu_i / \sum_{i=1}^{n} \mu_i$, $i = 1, 2, \ldots, n$.

The drawback of this measure is that it does not depend on the absolute values of $\mu_i$, but on their relative values. Thus a set with $\mu_i = 0.1$ or 0.8 or 1 for all of its elements would have same entropy (equal to 1). This is intuitively unappealing.

Deluca and Termini [6] have used a different expression. based on Shannon's function, to define the entropy of a fuzzy set as follows.

$$H = -k \sum_{i=1}^{n} \{\mu_i \log(\mu_i) + (1 - \mu_i)\log(1 - \mu_i)\}, \qquad (6)$$

where $k$ is a normalizing constant. Equation (6) is claimed to express an average amount of fuzziness (ambiguity) present in a set $A$ and it has the following desirable properties.

$P$ 1: $H$ is minimum iff $\mu_i = 0$ or 1 for all $i$.

$P$ 2: $H$ is maximum iff $\mu_i = 0.5$ for all $i$.

$P$ 3: $H \geqslant H^*$, where $H^*$ is the entropy of $A^*$, a sharpened version of $A$. ($A^*$ is a sharpened version of $A$ if $\mu^* \leqslant \mu$ for $\mu$ in $[0, 0.5]$ and $\mu^* \geqslant \mu$ for $\mu$ in $[0.5, 1]$.)

$P$ 4: $H = H'$, where $H'$ is the entropy of the complement set [for the sake of notational simplicity $\mu(x_i)$ has been written as $\mu_i$].

Any measure of fuzziness should possess these four properties. Kosko [7] defined the entropy of a fuzzy set as the ratio of distances between the fuzzy set and its nearest and furthest nonfuzzy neighbours. This measure also satisfies properties $P$ 1–$P$ 4.

Pal and Pal [3] have also given a measure of fuzziness in a set as

$$H = k \sum_{i=1}^{n} \{\mu_i \exp(1 - \mu_i) + (1 - \mu_i)\exp(\mu_i)\}. \qquad (7)$$

It has been proved [3] that Equation (7) satisfies properties $P\ 1 - P\ 4$.

It is to be noted here that the meaning of the entropy of a fuzzy set is quite different from that of the probabilistic entropy [Equations (1) and (2)]. The former gives, as mentioned before, the average amount of ambiguity (difficulty) in deciding whether an element belongs to a set or not, while the later gives the average gain in information from the occurrence of an event. Since an ordinary set is a special case of a fuzzy set, a relationship between these two types of information is expected. There have been several attempts in this regard to combine the probabilistic (Shannon's) and possibilistic (fuzzy) entropy. These are explained below.

Deluca and Termini [6] attempted to do so in the following manner. Consider an experiment in which the elements $x_1, x_2, \ldots, x_n$ may occur, once and only once in each trial with probabilities $p_1, p_2, \ldots, p_n$ ($p_i \geqslant 0$, $\Sigma_i p_i = 1$).

Shannon's entropy of the probability distribution is given by

$$H(p_1, p_2, \ldots, p_n) = - \sum_{i=1}^{n} p_i \log(p_i). \qquad (8)$$

This gives the average amount of information gained from the knowledge of the occurrence of an element. Suppose there exists a difficulty in the interpretation of $x_i$, the outcome of a trial, as 0 or 1. The amount of ambiguity involved in the interpretation of $x_i$ is given by [6]

$$S(\mu_i) = - \mu_i \log(\mu_i) - (1 - \mu_i)\log(1 - \mu_i). \qquad (9)$$

The statistical average $(m)$ of $S(\mu_i)$ is given by

$$m = \sum_{i=1}^{n} p_i S(\mu_i). \qquad (10)$$

This $m$ is claimed to give the average amount of difficulty in taking a decision (0 or 1) on the elements $x_i$, $i = 1, 2, \ldots, n$. Combining Equations (8) and (10), the total entropy is defined as follows:

$$H_{tot} = H(p_1, p_2, \ldots, p_n) + m(\mu, p_1, p_2, \ldots, p_n). \qquad (11)$$

According to Deluca and Termini, $H_{tot}$ gives the total average uncertainty that one may have in making a prevision about the elements of $\{x_1, x_2, \ldots, x_n\}$ which occur as a result of the experiment, and in taking a decision 0 or 1 on their values.

Xie and Bedrosian [8] defined the total entropy of a fuzzy set in a little different way. Consider a set $A^c$ containing only 0 and 1 with probabilities $p_0$ and $p_1$ ($p_0 + p_1 = 1$, $0 \leqslant p_0$, $p_1 \leqslant 1$). Suppose that, due to some reason, the sharpness in the set has changed and resulted in a fuzzy set $A$. The membership value of an element is changed to an arbitrary value in the range $[0, 0.5]$ from 0 and in $[0.5, 1]$ from 1. In this way the ordinary set $A^c$ has been changed to a fuzzy set $A$. Thus the fuzzy set $A$ has two types of uncertainties: One is due to the random uncertainty in the ordinary set, and the other is the fuzzy uncertainty arising due to the fuzziness in the set. They defined the total entropy of a fuzzy set as follows:

$$H_{tot} = H(p_0, p_1) + (1/n) \sum_{i=1}^{n} S(\mu_i). \tag{12}$$

This entropy reduces to Shannon's entropy of a two-state system when the fuzziness is removed (i.e., when the second part vanishes).

## 4. JUSTIFICATION FOR NEW DEFINITIONS

In this section we shall be justifying the need for two new definitions namely, hybrid and higher order entropy of fuzzy sets. Before explaining their necessity, let us, first of all, critically analyze the aforementioned measures.

Regarding Equation (11) the following points are in order. Deluca and Termini [6] presented Equation (11) as if $m$ in Equation (11) is different from Equation (6). But this is not the case. A critical analysis of the aforesaid example considered to derive Equation (11) shows that $m$ is, in fact, equal to $H$ in Equation (6). Suppose that the said experiment is repeated $N$ times; the entropy of the resulting fuzzy set, as given by Equation (6), would then be

$$H = (-1/N) \sum_{i=1}^{N} \{\mu_i \log(\mu_i) + (1 - \mu_i)\log(1 - \mu_i)\}. \tag{13}$$

Let $n_i$ be the number of times $x_i$ has occurred; then $\sum_{i=1}^{n} n_i = N$. Thus,

$$H = (-1/N) \sum_{i=1}^{n} n_i \{\mu_i \log(\mu_i) + (1 - \mu_i)\log(1 - \mu_i)\}$$

or

$$H = - \sum_{i=1}^{n} (n_i/N)\{\mu_i \log(\mu_i) + (1 - \mu_i)\log(1 - \mu_i)\}$$

or
$$H = - \sum_{i=1}^{n} p_i \{ \mu_i \log(\mu_i) + (1 - \mu_i) \log(1 - \mu_i) \}$$

or
$$H = \sum_{i=1}^{n} p_i S(\mu_i)$$

or
$$H = m(\mu, p_1, p_2, \ldots, p_n).$$

Therefore, $m$ in Equation (11) is not different form $H$ in Equation (6).

Referring back to the same experiment of Deluca and Termini, there will not be any difficulty in interpreting an outcome $x_i$ as either 0 or 1 when the associated fuzziness is removed. In other words, the entropy of the system should boil down to that of a two-state system $H(p_0, p_1)$, $p_0 + p_1 = 1$. But this is not supported by Equation (11), which, in such a case. reduces to $H(p_1, p_2, \ldots, p_n)$. This has also been pointed out by Xie and Bedrosian [8], whose measure [Equation (12)] is free from this drawback.

However, Equation (12) has the following unappealing implications. They established an equivalence between fuzzy information and Shannon information because both of the information measures have the same mathematical form. If $p_i = \mu_i$, they inferred that the average amount of fuzzy information yielded by a fuzzy set with $n$ elements is "equivalent" to the average amount of Shannon information yielded by $n$ independent binary Shannon information sources. Based on this, it has been concluded that fuzzy information can be transferred to Shannon information and inversely [8]. This type of equivalence is physically meaningless except that both of the measures yield same numerical value.

Moreover, fuzzy information is conceptually different from the probabilistic information. Their arithmatic sum may not yield any meaningful quantity. In that sense it is difficult to interpret $H_{tot}$. Furthermore, if fuzziness is removed, Equation (12) always reduces to $H(p_0, p_1)$ irrespective of the defuzzification process. For example, consider the following two cases. In the first case, the symbols in the range $[0, 0.5]$ are converted to zero and the remaining to unity. In the second case, some of the symbols in $[0, 0.5]$ are wrongly mapped to 1. In both cases Equation (12) will yield the same entropy. This is not at all a desirable property.

It is to be mentioned here that since a fuzzy set is a generalized version of an ordinary set, the entropy of a fuzzy set deserves to be a generalized version of classical entropy by taking into account not only the fuzziness of the set but also the underlying probability structure. In other words, it should be such that the classical entropy becomes its special case when fuzziness is removed. It is also not necessary that, with the removal of fuzziness, the value of generalized entropy decreases. Furthermore, considering the example of 0 and 1 in

providing entropy measure by Xie and Bedrosian, there would be one and only one type of uncertainty (difficulty), which is associated with the interpretation of an incoming symbol as 0 or 1. Of course, this difficulty depends on two factors, namely, the probability distribution and the possibility distribution; but this cannot be the sum of two uncertainties (e.g., probabilistic and fuzzy) as done in Equation (12). Such a measure may be called *hybrid entropy*.

The entropy of a fuzzy set as given by either Equation (6) or (7) gives the average ambiguity in taking the decision whether an element belongs to the set or not. Consider a fuzzy set "good football players" with 20 members. For any player $x$, $\mu(x)$ gives the degree of goodness of the player $x$. The entropy as given by Equation (6) or (7) for such a set gives an average amount of difficulty in taking a decision whether an individual player is good or not. This type of information is not always sufficient. Often we are interested to know if a team of, say, 11 players is formed by selecting any 11 players from the 20, to what extent, on an average, the team can be called "good." This raises two important issues: first, how to measure the degree to which a collection of objects, as a whole, possesses the property of "goodness"; second, how to get a measure of average amount of uncertainty related to such collections.

Let us now consider the first issue. Suppose $x_1, x_2, \ldots, x_n$ are $n$ supporting elements with membership values $\mu_1, \mu_2, \ldots, \mu_n$ with respect to some property $P$. Then to what extent does $\{x_1, x_2, \ldots, x_n\}$ as a whole (collectively) possess the property $P$? This obviously depends on the problem at hand. For example, in a quiz team, if $\mu_i$ is the ability of the $i$th member, then the ability of the team as a whole would be $\max(\mu_i)$, because if one member succeeds the team succeeds. On the other hand, suppose a group of acrobats are standing in such a fashion that all of them will fall if any one of them falls. Under this situation, if $\mu_i$ is the stability of the $i$th member, then the stability of the team as a whole would be $\min(\mu_i)$. Allen [9] has addressed this issue of properties of sets.

In order to get an answer to the second problem, a new definition of entropy of a fuzzy set is required which will give a measure of average uncertainty associated with any arbitrary subset with $r$ supports. Such an entropy may be called the *rth order entropy* of a fuzzy set. The definition should be such that for $r = 1$ it corresponds to a measure satisfying the properties $P\,1-P\,4$ of Equation (6) or (7).

## 5.  NEW DEFINITIONS

In this section we shall define the higher order entropy and the hybrid entropy of a fuzzy set.

## 5.1. HIGHER ORDER ENTROPY

Let $P$ be a fuzzy property set with a finite number of supports $n$, i.e., $P = \{\mu_i / x_i, i = 1, 2, \ldots, n\}$, where $\mu_i$ denotes the degree to which $x_i$ possesses the property $P$. Out of $n$ elements consider a combination of $r$ elements. Let $S_i^r$ denote the $i$th such combination and $\mu(S_i^r)$ denote the degree to which the combination $S_i^r$, as a whole, possesses the property $P$. There are $\left\{ \binom{n}{r} \right\}$ such combinations. The entropy of order $r$ of the fuzzy set $A$ is defined as

$$H^r = \left( 1 / \binom{n}{r} \right) \sum_{i=1}^{\binom{n}{r}} \{ \mu(S_i^r) \exp\{ 1 - \mu(S_i^r) \}$$

$$+ \{ 1 - \mu(S_i^r) \} \exp\{ \mu(S_i^r) \} \}. \tag{14}$$

A similar definition using the logarithmic function can also be given as follows:

$$H^r = \left( -1 / \binom{n}{r} \right) \sum_{i=1}^{\binom{n}{r}} \{ \mu(S_i^r) \log\{ \mu(S_i^r) \}$$

$$+ \{ 1 - \mu(S_i^r) \} \log\{ 1 - \mu(S_i^r) \} \}. \tag{15}$$

In our subsequent discussion, unless stated explicitly, $H^r$ will refer to Equation (14). Therefore, $H^r$ will give a measure of the average amount of difficulty in taking a decision on any subset of size $r$ with respect to the property $P$. In the example of acrobats, $H^r$ will denote an average amount of ambiguity (difficulty) in deciding a random team of $r$ acrobats as "stable." If $r = 1$, $H^r$ in Equations (14) and (15) reduces to Equations (7) and (6), respectively. $H^r$ has the following properties.

*Pr* 1. $H^r$ attains a maximum if $\mu_i = 0.5$ for all $i = 1, 2, \ldots, n$.

*Pr* 2. $H^r$ attains a minimum if $\mu_i = 0$ or $1$ for all $i = 1, 2, \ldots, n$.

*Pr* 3. $H^r \geqslant H^{*r}$, where $H^{*r}$ is the $r$th-order entropy of a sharpened version of the fuzzy set.

*Pr* 4. $H^r$ is, in general, not equal to $H^{r'}$, where $H^{r'}$ is the $r$th-order entropy of the complement set.

*Pr* 5. $H^r \geqslant H^{r+1}$ when all $\mu_i \in [0, 0.5]$ and $H^r \leqslant H^{r+1}$ when all $\mu_i \in [0.5, 1]$.

Note that the property $P$ 4 of Equation (6) or (7) is not, in general, valid here. The additional property *Pr* 5 implies that $H^r$ is a monotonically nonincreasing function of $r$ for $\mu_i \in [0, 0.5]$ and a monotonically nondecreasing

function of $r$ for $\mu_i \in [0.5, 1]$ (when min operator has been used to get the group membership value). Proof of $Pr\ 1-Pr\ 3$ for $H^1$ (i.e., for $r=1$) are given in [3] when exponential behavior of gain function is considered in the entropy measure. Their proofs for $H^r$ ($r>1$) are straightforward. The proof of $Pr\ 5$ is given below.

Let the property set $P$ be defined from the set of supports $S_i^r$ as

$$P_i^r = \left\{ \left( \mu_j^i / x_j^i \right), j = 1, 2, \ldots, r, i = 1, 2, \ldots, \binom{n}{r} \right\},$$

where $x_j^i \in S_i^r$ and $\mu_j^i$ denotes the degree of possessing the property $P$ by the $j$th element of $S_i^r$. Let $\mu(S_i^r) = \min_j \{ \mu_j^i, j = 1, 2, \ldots, r \}$. Now there are $(n-r)$ subsets of size $(r+1)$ such that

$$S_i^r \subset S_i^{r+1}.$$

Let us denote these subsets by $S_{i,j}^{r+1}$, $j = 1, 2, \ldots, n-r$; i.e., $S_{i,j}^{r+1}$ is the $j$th subset of $P$ of size $(r+1)$ such that

$$S_i^r \subset S_{i,j}^{r+1}.$$

Now

$$S_i^r \subset S_{i,j}^{r+1} \Rightarrow \min\{ \mu_1^i, \mu_2^i, \ldots, \mu_r^i \} \geq \min\{ \mu_1^{i,j}, \mu_2^{i,j}, \ldots, \mu_{r+1}^{i,j} \}$$

$$\Rightarrow \mu(S_i^r) \geq \mu(S_{i,j}^{r+1}), \qquad j = 1, 2, \ldots, n-r. \qquad (16)$$

Thus, for every $S_i^r$, $i = 1, 2, \ldots, \binom{n}{r}$, there are $(n-r)$ subsets of size $(r+1)$ such that condition (16) is true. Thus for every $\mu(S_i^r)$, there are $(n-r)$ times $\mu(s_i^{r+1})$ such that

$$\mu(S_i^r) \geq \mu(S_{i,j}^{r+1}), \qquad j = 1, 2, \ldots, (n-r).$$

Since from a set of size $(r+1)$ we can generate exactly $(r+1)$ distinct set of size $r$, we can infer that in the sequence

$$\mu(S_{i,j}^{r+1}), \qquad i = 1, 2, \ldots, \binom{n}{r}, \qquad j = 1, 2, \ldots, (n-r),$$

each of

$$\mu(S_i^{r+1}), \qquad i = 1, 2, \ldots, \binom{n}{r+1}$$

has occurred exactly $(r+1)$ times.

Thus,

$$(n-r) \sum_{i=1}^{\binom{n}{r}} \mu(S_i^r) \geqslant (r+1) \sum_{i=1}^{\binom{n}{r+1}} \mu(S_i^{r+1})$$

$$\Rightarrow \frac{(n-r) \sum_{i=1}^{\binom{n}{r}} \mu(S_i^r)}{\binom{n}{r}} \geqslant \frac{(r+1) \sum_{i=1}^{\binom{n}{r+1}} \mu(S_i^{r+1})}{\binom{n}{r}}$$

$$\Rightarrow (n-r)\overline{\mu_r} \geqslant (r+1) \frac{\sum_{i=1}^{\binom{n}{r+1}} \mu(S_i^{r+1})}{\binom{n}{r}}$$

$$\Rightarrow \overline{\mu_r} \geqslant \frac{\sum_{i=1}^{\binom{n}{r+1}} \mu(S_i^{r+1})}{\binom{n}{r+1}}$$

$$\Rightarrow \overline{\mu_r} \geqslant \overline{\mu_{r+1}}. \tag{17}$$

Thus, the average value of all $\mu(S_i^r)$ is not less than that of $\mu(S_i^{r-1})$.

In order to prove *Pr* 5, let us investigate the behavior of an entropic function with respect to its membership values. Consider the entropic function [Equation (7)]

$$H = (1/n) \sum_{i=1}^{n} \{\mu_i \exp(1-\mu_i) + (1-\mu_i)\exp(\mu_i)\}$$

or

$$H = (1/n) \sum_{i=1}^{n} C(\mu_i), \quad \text{where} \quad C(\mu_i) = \mu_i \exp(1-\mu_i) + (1-\mu_i)\exp(\mu_i).$$

Differentiating $C(\mu_i)$ with respect to $\mu_i$, we get

$$C'(\mu_i) = (1 - \mu_i)\exp(1 - \mu_i) - \mu_i\exp(\mu_i)$$

$$\Rightarrow C'(\mu_i) > 0 \qquad \text{for } 0 \leqslant \mu_i < 0.5$$

$$< 0 \qquad \text{for } 0.5 < \mu_i \leqslant 1$$

$$= 0 \qquad \text{for } \mu_i = 0.5.$$

Thus, $C(\mu_i)$ increases monotonically with $\mu_i$ over $[0, 0.5)$, attains a maximum at 0.5, and then monotonically decreases over $(0.5, 1]$.

This fact and inequality (17) prove the property *Pr* 5.

*Illustration*

Let us consider the problem of selecting a team of 5 acrobats from a group of 10. Define a fuzzy set $A$ called "stable acrobats" with 10 supports, i.e., $A = \{\mu_i / x_i, i = 1, 2, \ldots, 10\}$, where $\mu_i$ denotes the stability of the $i$th individual $x_i$. The fuzzy entropy of order 5 gives the ambiguity (difficulty) in deciding any arbitrary team of 5 players from the group of 10 as stable or unstable. The values of $H^1 - H^6$ presented in Table 2 highlight the key features of higher order entropy as compared to $H^1$. Based on Table 2 the following observations can be made.

When all the $\mu_i$ values are same, $H^1 - H^6$, as expected become the same. This is because of the fact that the difficulty in taking a decision regarding possession of a property on an individual is same as that of a group selected therefrom. The value of $H^r$ would, of course, be dependent on the $\mu_i$ values. Since all the $\mu_i$ in case 2 are 0.5 (most ambiguous), the corresponding $H^r$, value is higher than that of case 1.

Consider cases 3 and 4. In case 3 the difference between $H^1$ and $H^2$ is very high, indicating a higher dissimilarity within the singletons of the fuzzy set with respect to the possession of property $P$. On the other hand, the case 4 for which $H^1$ is very close to $H^2$ reflects that the fuzzy singletons of the set are more alike (similar) with respect to property $P$. It is also seen that the higher the similarity among singletons, the quicker is the convergence to the limiting value of $H^r$.

Based on the above observation, let us define an index of similarity of supports of a fuzzy set as $S = H^1 / H^2$ (when $H^2 = 0$, $H^1$ is also zero and $S$ is taken as 1). Obviously, when $\mu_i \in [0.5, 1]$ and the min operator is used to assign the degree of possession of the property by a collection of supports, $S$

TABLE 2

Higher Order Entropy

| Case | $\mu_A$ | $H^1$ | $H^2$ | $H^3$ | $H^4$ | $H^5$ | $H^6$ | $S$ |
|------|---------|-------|-------|-------|-------|-------|-------|-----|
| 1 | $\{1,1,1,1,1,1,1,1,1,1\}$ | 0 | 0 | 0 | 0 | 0 | 0 | 1.0 |
| 2 | $\{0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5\}$ | 1 | 1 | 1 | 1 | 1 | 1 | 1.0 |
| 3 | $\{1,1,1,1,1,0.5,0.5,0.5,0.5,0.5\}$ | 0.5 | 0.7777778 | 0.9166667 | 0.9761905 | 0.9960318 | 1 | 0.6428570 |
| 4 | $\{0.5,0.5,0.5,0.5,0.5,0.6,0.6,0.6,0.6,0.6\}$ | 0.9809018 | 0.9915119 | 0.9968170 | 0.9990906 | 0.9998485 | 1 | 0.9892983 |
| 5 | $\{0.6,0.6,0.65,0.9,0.9,0.9,0.9,0.9,0.9,0.915\}$ | 0.5384138 | 0.6785914 | 0.7810661 | 0.8553280 | 0.9059016 | 0.9373114 | 0.7934287 |
| 6 | $\{0.8,0.8,0.8,0.8,0.8,0.8,0.9,0.9,0.9,0.9\}$ | 0.5388815 | 0.6135760 | 0.6415336 | 0.6495325 | 0.6508657 | 0.6508657 | 0.8782636 |
| 7 | $\{0.5,0.5,0.5,0.5,0.5,0.5,0.9,0.9,0.9,0.9\}$ | 0.7483621 | 0.9161205 | 0.9790324 | 0.9970038 | 1 | 1 | 0.8168817 |
| 8 | $\{0.7,0.7,0.7,0.7,0.7,0.8,0.8,0.8,0.8,0.8\}$ | 0.7485938 | 0.8028872 | 0.8300339 | 0.8416682 | 0.8455463 | 0.8463219 | 0.9323772 |

will lie in $[0, 1]$, as $H^r \leqslant H^{r+1}$. Similarly, when $\mu_i \in [0, 0.5]$, $S$ may be defined as $H^2/H^1$ so that $S$ lies in $[0, 1]$. The higher the value of $S$, the more alike (similar) are the supports of the fuzzy set with respect to the property $P$. This index of similarity can therefore be regarded as a measure of the degree to which the members of a fuzzy set are alike.

Let us now consider cases 5 and 6. In both cases the $H^1$ values are almost same but the index of similarity is lower for case 5 than for case 6. This indicates that case 6 has more similar supports than case 5.

Note that cases 7 and 8 also have almost identical values for $H^1$, but these values are much higher than those of cases 5 and 6. Based on this observation, we can infer that fuzzy sets corresponding to cases 7 and 8 have a larger number of supports with a lower degree of possession in $[0.5, 1]$ of the property $P$ than those corresponding to cases 5 and 6. Again the index of similarity for case 8 is higher than that for case 7, indicating that the members of the fuzzy set corresponding to case 8 are more alike than those corresponding to case 7.

Therefore, the value of conventional fuzzy entropy $(H^1)$ can only indicate whether the fuzziness in a set is low or high. In addition to this, the value of $H^r$ also enables one to infer whether the fuzzy set contains similar supports (or elements) or not. The similarity index thus defined can be successfully used for measuring interclass and intraclass ambiguity (i.e., class homogenity and contrast) in pattern recognition and image processing problems.

Let us consider the work of Pal and Chakraborty [10], who have used the measure $H^1$ for evaluating feature importance in the pattern recognition problem. They used the $\pi$-type membership function to represent a pattern class. They made the membership values lie in the range $[0.5, 1]$ with 0.5 corresponding to the boundary elements of the pattern class and 1 at the central point (the point corresponding to the average value of the feature). This type of representation is a very natural one. They decided a feature to be of high importance if it has low $H^1$ value. We have already shown (e.g., cases 5 and 6, or 7 and 8 of Table 2) that it is possible to have two fuzzy sets with significantly different compactness, but almost identical values for entropy. It is therefore not appropriate to evaluate the importance of a feature just on the basis of $H^1$. A better evaluation may possibly be done by considering $H^1$ and $S$ together. Obviously, a low value of $H^1$ together with a high value of $S$ will indicate that the feature is more important.

Table 2 also reveals the validity of the property $Pr$ 5 i.e., $H^r \leqslant H^{r+1}$ when $\mu_i \in [0.5, 1]$.

## 5.2. HYBRID ENTROPY

In Section 3 we have discussed the various attempts that have been made to combine the probabilistic and fuzzy uncertainties and their associated prob-

lems. Since an ordinary set is a special case of a fuzzy set, it is logical to think of a generalized definition of entropy of a fuzzy set which reduces to the probabilistic entropy in absence of fuzziness. Let us refer back to the example (mentioned in Section 3) of digital communication over a noisy channel. For this type of example, Xie and Bedrosian [8] assumed two types of uncertainties. But one can visualize that there is one and only one type of difficulty and that is in the interpretation of an incoming symbol as 0 or 1. Of course, this is dependent on two different factors: the probability of generation of 0 and 1 by the source and the transfer function (channel noise), which makes them fuzzy.

Let $p_0$ and $p_1$ be the probabilities of occurrence of the 0 and 1 symbols, respectively, and let $\mu_i$ denote the membership for the fuzzy set "symbol close to 1."

Let us consider an expression

$$E_1 = (1/n) \sum_{i=1}^{n} \mu_i \exp(1 - \mu_i). \tag{18}$$

Differentiating $E_1$ with respect to $\mu_i$, we get

$$E_1'(\mu_i) = (1/n)(1 - \mu_i)\exp(1 - \mu_i), \qquad i = 1, 2, \ldots, n,$$

$$\Rightarrow E_1'(\mu_i) \geq 0 \qquad \text{for} \quad 0 \leq \mu_i \leq 1$$

(i.e., the rate of change of $E_1$ is seen to be nonnegative).

Thus $E_1$ is a monotonically increasing function of $\mu_i$ for $\mu_i \in [0, 1]$. In other words, as $\mu_i$ increases from 0 to 1, $E_1$ also increases. Now, for an incoming zero (0) symbol, if $\mu_i$ increases, the difficulty in the correct interpretation of the symbol increases and hence the difficulty in the interpretation of a "0" as "1" decreases; i.e., a wrong interpretation of a "0" symbol becomes favorable. On the other hand, for an incoming 1 symbol, if $\mu_i$ increases, the difficulty in its correct interpretation decreases. The higher the value of $E_1$, the more favorable will be the interpretation of an arbitrary incoming symbol as 1. Therefore, $E_1$ can be taken as the average likeliness (possibility) of interpreting a received symbol as "1."

Similarly,

$$E_0 = (1/n) \sum_{i=1}^{n} (1 - \mu_i)\exp(\mu_i) \tag{19}$$

can be viewed as the average likeliness of interpreting a received symbol as 0.

Since $p_0$ and $p_1$ are, respectively, the probabilities of occurrence of 0 and 1, the *hybrid entropy* of the fuzzy set $A$ (symbols close to 1) may be defined as

$$H_{hy} = -p_0 \log(E_0) - p_1 \log(E_1) \qquad (20)$$

(using logarithmic behavior)

or

$$H_{hy} = p_0 \exp(1 - E_0) + p_1 \exp(1 - E_1) \qquad (21)$$

(using exponential behavior)

$H_{hy}$ has the following properties.

PROPERTY 1. *In the absence of fuzziness when $np_0$ received symbols take the value zero (0) and $np_1$ symbols take the value of one (1), then $E_0$ and $E_1$ reduce to $p_0$ and $p_1$, respectively, and Equation (20) or (21) boils down to classical entropy. Thus we see that $H_{hy}$ boils down to classical entropy when a proper defuzzification process is applied to detect (restore) correctly the incoming symbols.*

This property has significant application in automatic image enhancement, segmentation, and the noise reduction process where the difference between the classical entropy and $H_{hy}$ can act as an objective criterion. The lower the difference, the lesser is the fuzziness associated with the individual symbol and the higher will be the accuracy in classifying them as 0 or 1.

PROPERTY 2. *If $\mu_i = 0.5$ for all $i = 1, 2, \ldots, n$, then $E_0 = E_1$ and Equation (20) takes the form*

$$H_{hy} = -p_0 \log\left\{ (1/n) \sum_{i=1}^{n} 0.5 * \exp(0.5) \right\} - p_1 \log\left\{ (1/n) \sum_{i=1}^{n} 0.5 * \exp(0.5) \right\}$$

or

$$H_{hy} = -\log\{0.5 * \exp(0.5)\},$$

*i.e.*, $H_{hy}$ *takes a constant value and becomes independent of $p_0$ and $p_1$.*

PROPERTY 3. *$E_0$ and $E_1$ cannot take any arbitrary values. They satisfy the following conditions:*

$$0.5 * \exp(0.5) * p_1 \leqslant E_1 \leqslant 0.5 * \exp(0.5) + (1 - 0.5 * \exp(0.5)) * p_1$$

*and*

$$(1-p_1)*0.5*\exp(0.5) \leqslant E_0 \leqslant 1-p_1*((1-0.5*\exp(0.5)))$$

*The proof of this property is given below.*

*Proof.* Let us consider Equation (19), which is

$$E_1 = (1/n) \sum_{i=1}^{n} \mu_i \exp(1-\mu_i)$$

$$\Rightarrow E_{1\max} = (1/n)[np_0*(0.5)*\exp(0.5)+np_1]$$

[since $E_1$ is found to be monotonically increasing function of $\mu_i$ and the expected number of symbol 0 (1) is $np_0$ ($np_1$) and each of them can take a maximum value of 0.5 (1)]

$$= p_1 + (1-p_1)*0.5*\exp(0.5)$$

$$= 0.5*\exp(0.5)+(1-0.5*\exp(0.5))p_1 \tag{i}$$

and

$$\Rightarrow E_{1\min} = (1/n)[np_0*0+np_1*0.5*\exp(0.5)]$$

$$= p_1*0.5*\exp(0.5) \tag{ii}$$

Combining (i) and (ii), one gets

$$0.5*\exp(0.5)*p_1 \leqslant E_1 \leqslant 0.5*\exp(0.5)+(1-0.5*\exp(0.5))p_1 \tag{c1}$$

$$\Rightarrow 0.82436*p_1 \leqslant E_1 \leqslant 0.82436+0.17564*p_1.$$

Similarly, from Equation (20) we have

$$E_{0\max} = (1/n)\{np_0*1+np_1*0.5*\exp(0.5)\}$$

$$= 1-p_1\{1-0.5\exp(0.5)\} \tag{iii}$$

and

$$E_{0\min} = (1/n)\{np_0*0.5*\exp(0.5)+np_1*0\}$$

$$= (1-p_1)*0.5*\exp(0.5). \tag{iv}$$

Combining (iii) and (iv), one can write the following condition:

$$(1 - p_1) * 0.5 * \exp(0.5) \leqslant E_0 \leqslant 1 - p_1(1 - 0.5 * \exp(0.5)) \qquad (c2)$$

$$\Rightarrow (1 - p_1) * 0.82436 \leqslant E_0 \leqslant 1 - 0.17564 * p_1.$$

This completes the proof.                                                                  ∎

PROPERTY 4. $E_0 + E_1$ attains the maximum value of $\exp(0.5)$ when all $\mu_i$ assume a value of 0.5. Thus,

$$E_0 + E_1 \leqslant \exp(0.5).$$

Proof of this property can be found in [3].

As mentioned in Property 1, when there is a proper defuzzification process applied to result in the $np_0$ times 0 symbol and the $np_1$ times 1 symbol, then $E_0$ and $E_1$ reduce to $p_0$ and $p_1$, respectively, to make $E_0 + E_1 = 1$.

*Interpretation to Image Processing*

Let us consider an example of a digital image in which, say, 70% pixels look white, while the remaining 30% look dark. Thus the probability of a white pixel $p_w$ is 0.7 and that of a dark pixel $p_b$ is 0.3. Suppose the whiteness of the pixels is not constant, i.e., there is a variation (grayness) and similar is the case with the black pixels. The hybrid entropy of such an image can be written as follows:

$$H_{hy} = - p_w \log\left\{ (1/n) \sum_i \mu_i \exp(1 - \mu_i) \right\} - p_b \log\left\{ (1/n) \sum_i (1 - \mu_i)\exp(\mu_i) \right\}$$

where $\mu_i$ gives the degree of whiteness of the $i$th pixel. Now $H_{hy}$ will be close to the classical entropy if there is not much variation in the grayness within the white and dark portions. Table 3 demonstrates the use of $H_{hy}$ in acting as an objective criterion for a proper defuzzification process.

Each fuzzy set of the Table 3 is a sharpened (defuzzified) version of the previous one such that $\mu$ value increases (decreases) in the sharpened version for those $\mu > 0.5$ ($< 0.5$). One can observe that with proper defuzzification $E_1$ approaches 0.7 and $E_0$ approaches 0.3; in fact, they monotonically decrease to 0.7 and 0.3, respectively, from their initial values. As a result, $|H - H_{hy}|$ decreases to zero.

Let us now consider the effect of improper defuzzification on the pattern shown in case 1 of the Table 4. Two types of defuzzifications are considered here. In cases 2-4 all the symbols with $\mu = 0.5$ are transformed to zero when some of them were actually generated from symbol "1." In cases 5 and 6 of

TABLE 3

Hybrid Entropy (with $p_0 = 0.3$ and $p_1 = 0.7$)

| Case | $\mu_A$ | $E_0$ | $E_1$ | $H_{hy}$ | $|H - H_{hy}|$ | $H_{tot}$ | $|H - H_{tot}|$ |
|---|---|---|---|---|---|---|---|
| 1 | {0.9, 0.8, 0.8, 0.8, 0.8, 0.7, 0.6, 0.45, 0.4, 0.3} | 0.6061934 | 0.8856076 | 0.2352039 | 0.3756607 | 1.169123 | 0.55825830 |
| 2 | {0.99, 0.9, 0.9, 0.9, 0.8, 0.79, 0.7, 0.2, 0.1, 0.05} | 0.5247224 | 0.7700894 | 0.3763399 | 0.2345245 | 0.9789118 | 0.36804750 |
| 3 | {0.999, 0.99, 0.99, 0.99, 0.98, 0.9, 0.9, 0.1, 0.05, 0} | 0.3622024 | 0.73642 | 0.5188341 | 0.0920302 | 0.7556359 | 0.1447716 |
| 4 | {1, 1, 1, 1, 1, 1, 1, 0.0004, 0, 0} | 0.30 | 0.7001086 | 0.6107557 | 0.00010866 | 0.6112173 | 0.00035295 |
| 5 | {1, 1, 1, 1, 1, 1, 1, 0, 0, 0} | 0.30 | 0.70 | 0.6108643 | 0.0 | 0.6108643 | 0.0 |

TABLE 4

Effect of Wrong Defuzzification (with $p_0 = 0.3$ and $p_1 = 0.7$)

| Case | $\mu_A$ | $E_1$ | $E_0$ | $H_{hy}$ | $\|H - H_{hy}\|$ | $H_{tot}$ | $\|H - H_{tot}\|$ |
|---|---|---|---|---|---|---|---|
| 1 | {0.9,0.9,0.8,0.8,0.7,0.6,0.5,0.5,0.4,0.3} | 0.6203827 | 0.8765243 | 0.2354792 | 0.3753851 | 1.171366 | 0.5605017 |
| 2 | {0.999,0.999,0.9,0.8,0.7,0.7,0.3,0.3,0.2,0.1} | 0.5766330 | 0.7760898 | 0.3426057 | 0.2682586 | 1.021888 | 0.4110242 |
| 3 | .{1,1,1,0.99,0.9,0.9,0.1,0.1,0,0} | 0.4508141 | 0.6481178 | 0.5425881 | 0.0682762 | 0.7464977 | 0.1356333 |
| 4 | {1,1,1,1,1,1,0,0,0,0} | 0.40 | 0.60 | 0.6324652 | 0.0216008 | 0.6108643 | 0.00 |
| 5 | {0.99,0.99,0.1,0.1,0.9,0.8,0.7,0.2,0.1,0.1} | 0.6304756 | 0.6345526 | 0.4567688 | 0.1540955 | 0.945773 | 0.3349087 |
| 6 | {1,1,0,0,1,1,1,0,0,0} | 0.5 | 0.5 | 0.6931472 | 0.08228284 | 0.6108643 | 0.00 |

Table 4 some of the $\mu$ values greater than 0.5 which were generated from symbol 1 (or belong to the white portion of the image) are wrongly defuzzified and brought down towards zero (instead of 1). In both situations, it is to be noted that $|H - H_{hy}|$ does not reduce to zero.

Let us now consider column no. 7 in Tables 3 and 4, which indicates that the measure $H_{tot}$ of Xie and Bedrosian does not have the aforesaid properties. Now, in image processing the process of defuzzification can be viewed as a contrast enhancement operation. Therefore, the measure $|H - H_{hy}|$ can be regarded as an objective criterion for appropriate enhancement of an image.

## 6. CONCLUSIONS

New entropy measures such as higher-order fuzzy entropy and hybrid entropy of a set have been introduced. The higher order entropy ($H^r$) is found to possess some interesting properties which can be used in feature evaluation and image segmentation problem. It leads to define a measure, called the "index of similarity" of supports of a set. As expected, $H^r$ conveys more information about the actual structure of a set than $H^1$ does. $H_{hy}$, on the other hand, can be used as an objective measure for proper defuzzification (enhancement) of a set.

## REFERENCES

1. C. E. Shannon, A mathematical theory of communication, *Bell System Tech. J.* 27:379–423 (Jul. 1948).
2. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, 1949.
3. N. R. Pal and S. K. Pal, Object background segmentation using a new definition of entropy, *IEE Proc. E*, 136:284–295 (1989).
4. L. A. Zadeh, Probability measures of fuzzy events, *J. Math. Anal. Appl.* 23:421–427 (1968).
5. A. Kaufmann, *Introduction to the Theory of Fuzzy Subsets — Fundamental Theoretical Elements*, Academic, New York, 1975, Vol. 1.
6. A. Deluca and S. Termini, A definition of nonprobabilistic entropy in the setting of fuzzy sets theory, *Inform. Control* 20:301–312 (1972).
7. B. Kosko, Fuzzy entropy and conditioning, *Inform. Sci.* 40:165–174 (1986).
8. W. X. Xie and S. D. Bedrosian, An information measure for fuzzy sets, *IEEE Trans. Syst. Man Cybernet.* SMC-14 (1):151–156 (Jan./Feb. 1984).
9. A. D. Allen, Measuring the emperical properties of sets, *IEEE Trans. Syst. Man Cybernet.* SMC-4 (1):66–73 (1974).
10. S. K. Pal and B. Chakraborty, Fuzzy set theoretic measure for automatic feature evaluation, *IEEE Trans. Syst. Man Cybernet.* SMC-16 (5):754–760 (Sept./Oct. 1986).