INFORMATION SCIENCES 66, 119–137 (1992)

Some Properties of the Exponential Entropy

NIKHIL R. PAL

and

SANKAR K. PAL

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

Communicated by Abraham Kandel

ABSTRACT

A new definition of probabilistic entropy based on the exponential behavior of gain function has recently been introduced by the authors [1]. It has also been extended to define various image entropy measures [2] and hybrid and higher order entropy of fuzzy sets [3]. The present work investigates some additional properties (in the light of both probability theory and fuzzy set theory) of this new entropy definition to ensure its greater range of applications.

1. INTRODUCTION

In an earlier work some limitations of Shannon's entropy have been discussed and a new definition of probabilistic entropy has been suggested by the authors [1]. This new definition assumes an exponential behavior for the information gain function. Several desirable properties of the new definition have also been established there. This definition has then been extended to define a nonprobabilistic entropy of a fuzzy set. In this context, higher order entropy and conditional entropy of an image have been defined and applied to develop image segmentation algorithms. In the context of image processing, the superiority of this exponential entropy over Shannon's logarithmic entropy has been established in several ways [1–5].

In this paper several other desirable properties of the exponential entropy are proved. Some interesting theorems regarding a compound probabilistic experiment are also established. It is found that for a station-

 0020-0255/92/$5.00

ary information source the expected information content of a sequence of symbols of length m is greater than that with a sequence of length q (where $m > q$). Moreover, for a stationary source the limiting value of information content per symbol (usually defined as the entropy of the source) is also found to exist.

2. DEFINITION

For the exponential entropy, the gain in information from the occurrence of the ith state of an n-state system is taken as [1]

$$\Delta I(p_i) = e^{1-p_i} \tag{1}$$

where $0 \leq p_i \leq 1$ and $\sum_{i=1}^{n} p_i = 1$. The entropy of such a system is thus defined as the expected value of ΔI; i.e.,

$$H = \sum_{i=1}^{n} p_i e^{1-p_i}. \tag{2}$$

The normalized entropy of the system can then be defined as

$$H_{\text{nor}} = (H - 1)/(e^{1-1/n} - 1). \tag{3}$$

On the other hand, Shannon's entropy [6] is given by

$$H^s = -\sum_{i=1}^{n} p_i \log p_i. \tag{4}$$

Figure 1 depicts the plot of the exponential entropy (normalized) and Shannon's entropy. It is quite interesting to observe that both curves are almost identical, thereby indicating that the basic nature of the entropy function remains the same under the new framework.

In [1] it has been proved that $H(p_1, \ldots, p_n) \leq H(1/n, \ldots, 1/n)$, i.e., entropy is maximum when all states are equally probable. It has also been established that for a binary source H monotonically increases for p in $[0, 0.5)$ and monotonically decreases for p in $(0.5, 1]$ with a maximum at $p = 0.5$. Some additional properties of the new entropy (in the light of both probability theory and fuzzy set theory) will be investigated in the following sections through theorems and proofs.

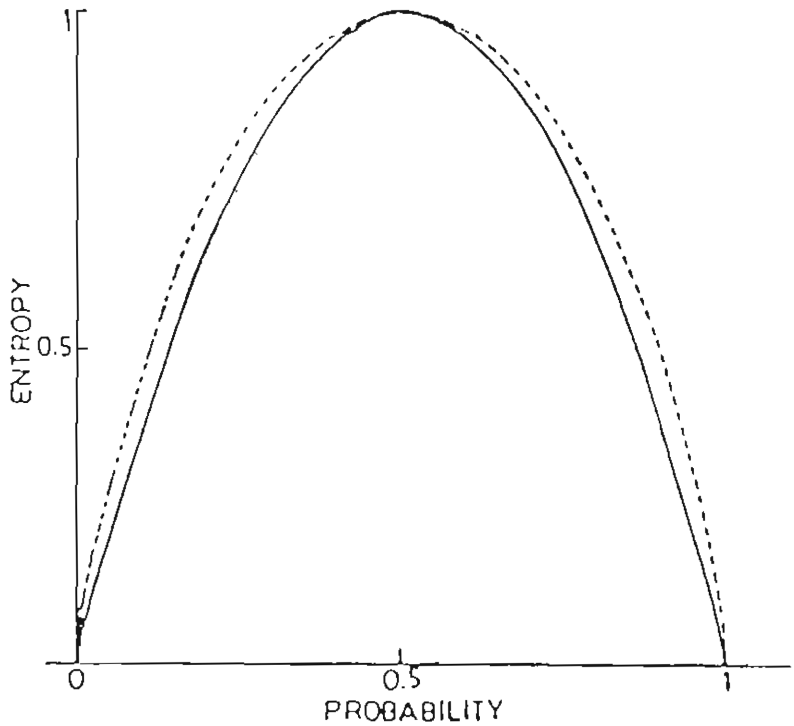

Fig. 1. Plot of Shannon's entropy and exponential entropy for a two-state system. (—— exponential, ---- logarithmic)

3. SOME PROPERTIES OF THE EXPONENTIAL ENTROPY

THEOREM 1.

$$H = \sum_{i=1}^{n} p_i e^{1-p_i}, \qquad 0 \le p_i \le 1; \qquad \sum_{i=1}^{n} p_i = 1$$

is a concave function.

Proof. Let

$$f(p) = pe^{1-p}; \tag{5}$$

then

$$f'(p) = (1-p)e^{1-p} \tag{6}$$

and

$$f''(p) = (p-2)e^{1-p}, \tag{7}$$

i.e., $f''(p) < 0$ for $0 \le p \le 1$.

We know that a function $f(x)$ is concave in the interval (a,b), if at all points in the interval (a,b) the second derivative of $f(x)$ is negative, i.e., $f''(x) < 0$, hence $f(p)$ (just defined) is a concave function. It is also known that a sum of concave functions is also a concave function. Hence H is a concave function.

THEOREM 2. *Any change towards equalization of the probabilities $p_1, p_2, \ldots, p_n$ increases H. Thus if $p_1 < p_2$ and we increase p_1, decreasing p_2 by an equal amount so that p_1 and p_2 become more nearly equal, then H increases.*

Proof. Suppose $p_1 < p_2$. Now p_1 is increased by an amount δ ($\delta > 0$) and p_2 is reduced by the same amount, such that $p_1 + \delta < p_2 - \delta$. With this change p_1 and p_2 has become more equal. To prove the theorem, we need to show that

$$H(p_1 + \delta, p_2 - \delta, p_3, \ldots, p_n) - H(p_1, p_2, p_3, \ldots, p_n) > 0 \tag{8}$$

or

$$(p_1 + \delta)e^{1-p_1-\delta} + (p_2 - \delta)e^{1-p_2+\delta} + \sum_{i=3}^{n} p_i e^{1-p_i} - \sum_{i=1}^{n} p_i e^{1-p_i} > 0 \tag{9}$$

or

$$(p_1 + \delta)e^{1-p_1-\delta} + (p_2 - \delta)e^{1-p_2+\delta} - p_1 e^{1-p_1} - p_2 e^{1-p_2} > 0$$

or

$$p_1 e^{1-p_1}(e^{-\delta} - 1) + p_2 e^{1-p_2}(e^{\delta} - 1) + \delta(e^{1-p_1-\delta} - e^{1-p_2+\delta}) > 0. \tag{10}$$

Let $\delta(e^{1-p_1-\delta} - e^{1-p_2+\delta}) = M$. (11)
Now

$$p_1 + \delta < p_2 - \delta \Rightarrow 1 - p_1 - \delta > 1 - p_2 + \delta$$

$$\Rightarrow \delta(e^{1-p_1-\delta} - e^{1-p_2+\delta}) > 0,$$

i.e., $M > 0$.

Let

$$p_1 e^{1-p_1}(e^{-\delta} - 1) + p_2 e^{1-p_2}(e^{\delta} - 1) = K \tag{12}$$

or

$$K = p_2 e^{1-p_2}(e^{\delta} - 1) - p_1 e^{1-p_1}(1 - e^{-\delta}) \tag{13}$$

or

$$K = p_2 e^{1-p_2}(e^{\delta} - 1) - p_1 e^{1-p_1}\left(\frac{e^{\delta} - 1}{e^{\delta}}\right). \tag{14}$$

Since pe^{1-p} is a monotonically increasing function of p, for $p \in [0, 1]$, $p_1 e^{1-p_1} < p_2 e^{1-p_2}$. In addition to this $\{(e^{\delta} - 1)/e^{\delta}\} < (e^{\delta} - 1)$ as $e^{\delta} > 1$ for $\delta > 0$. Hence $K > 0$.

Thus

$$H(p_1 + \delta, p_2 - \delta, p_3, \ldots, p_n) - H(p_1, p_2, \ldots, p_n) = M + K > 0. \tag{15}$$

Hence the proof.

THEOREM 3. *If any zero probability is changed to a nonzero probability by reducing some other probability, then the entropy increases, i.e.,*

$$H(\delta, p_2 - \delta, p_3, \ldots, p_n) - H(0, p_2, \ldots, p_n) > 0.$$

Proof.

$$H(\delta, p_2 - \delta, p_3, \ldots, p_n) - H(0, p_2, \ldots, p_n)$$

$$= \delta e^{1-\delta} + (p_2 - \delta)e^{1-p_2+\delta} + \sum_{i=3}^{n} p_i e^{1-p_i} - \sum_{i=2}^{n} p_i e^{1-p_i} \tag{16}$$

$$= \delta e^{1-\delta} + (p_2 - \delta)e^{1-p_2+\delta} - p_2 e^{1-p_2} \tag{17}$$

$$= \delta e^{1-\delta} + (p_2 - \delta)e^{1-p_2+\delta} - \delta e^{1-p_2} - (p_2 - \delta)e^{1-p_2} \tag{18}$$

$$= \delta(e^{1-\delta} - e^{1-p_2}) + (p_2 - \delta)e^{1-p_2}(e^{\delta} - 1). \tag{19}$$

Now

$$\begin{aligned} &\delta < p_2 \\ &\Rightarrow 1-\delta > 1-p_2 \\ &\Rightarrow e^{1-\delta} > e^{1-p_2} \\ &\Rightarrow \delta(e^{1-\delta} - e^{1-p_2}) > 0. \end{aligned}$$

Again

$$\begin{aligned} &\delta > 0 \\ &\Rightarrow e^{\delta} > 1 \\ &\Rightarrow (e^{\delta} - 1) > 0 \\ &\Rightarrow (p_2 - \delta)e^{1-p_2}(e^{\delta} - 1) > 0. \end{aligned}$$

Hence the proof.

THEOREM 4. *H is minimum, if and only if all the p_i but one are zero, this one having the value of unity.*

Proof. Suppose $p_i = 0$ for all i except p_k, which is 1; then H must be minimum. If not so, the minimum of H occurs when at least there are two nonzero probabilities, say p_i and p_j. In this case if we merge the two probabilities and generate two states, one with probability of 0 and the other with $(p_i + p_j)$, then the entropy will reduce (by Theorem 3). This contradicts our assumption that entropy is minimum when there are at least two states with $p_i \neq 0$.

Again if H is minimum then p_i will be equal to zero for all i except some k for which $p_k = 1$. Now suppose this is not true, i.e., H is minimum and there are at least two states with nonzero probabilities, say with p_i and p_k. In this case if we change the probabilities of the two states as 0 and $(p_i + p_k)$, then in the new system entropy will be reduced (by Theorem 3).

In addition to this, H is a symmetric function (i.e., permutation of p_i does not change the entropy). Thus, H will be minimum when all p_i's are zero except one, which takes a value of unity.

Hence,

$$H(1,0,0,\dots,0) = H(0,1,0,\dots,0) = \cdots = H(0,0,\dots,1) = H_{\min}.$$

4. PROPERTIES OF COMPOUND PROBABILISTIC EXPERIMENT

Consider two probabilistic experiments A and B with possible outcomes $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_m)$, respectively. Define $(A * B)$ as a compound probabilistic experiment with outcomes of the form (a_k, b_l); $k = 1, 2, \ldots, n$; $l = 1, 2, \ldots, m$.

Let p_{kl} be the probability of the event (a_k, b_l). Now, if p_k denotes the probability of the outcome a_k in the first experiment regardless of the second, then,

$$p_k = \sum_{l=1}^{m} p_{kl}. \tag{20}$$

Similarly, q_l, the probability of b_l in the second experiment regardless of the first, is

$$q_l = \sum_{k=1}^{n} p_{kl}. \tag{21}$$

Now, $p(a_k / b_l)$, the conditional probability of a_k given that b_l has occurred, is given by

$$p(a_k / b_l) = \frac{p_{kl}}{q_l} = p_{k/l} \quad \text{(say)}. \tag{22}$$

Similarly,

$$p(b_l / a_k) = \frac{p_{kl}}{p_k} = q_{l/k}. \tag{23}$$

Therefore, the entropy of the system A, given that b_l has occurred in the experiment B, is given by

$$H_n(A / b_l) = \sum_{k=1}^{n} p_{k/l} e^{1 - p_{k/l}} \tag{24}$$

and

$$H_n(A/B) = \sum_{l=1}^{m} \sum_{k=1}^{n} q_l p_{k/l} e^{1-p_{k/l}}. \tag{25}$$

Similarly,

$$H_m(B/A) = \sum_{k=1}^{n} \sum_{l=1}^{m} p_k q_{l/k} e^{1-q_{l/k}}. \tag{26}$$

Again, the entropy of the compound experiment $(A*B)$ is defined as

$$H_{nm}(A*B) = \sum_{k=1}^{n} \sum_{l=1}^{m} p_{kl} e^{1-p_{kl}}. \tag{27}$$

Under the preceding framework, the following theorems can be stated.

THEOREM 5.

$$H_m(B/A) \le H_m(B)$$

and

$$H_n(A/B) \le H_n(A).$$

This theorem says that the average amount of gain in information from the realization of the experiment A (or B) can decrease if another experiment B (or A) has already been realized.

Proof. We know that for any real valued continuous concave function $f(x)$ over $[0,1]$ and for any set of $(x_1, x_2, \ldots, x_n)$ over $[0,1]$, the following inequality holds good.

$$\sum_{k=1}^{n} \lambda_k f(x_k) \le f\left(\sum_{k=1}^{n} \lambda_k x_k\right) \quad \text{(Jensen inequality)}. \tag{28}$$

where λ_i's are non-negative real numbers such that $\sum_{i=1}^{n} \lambda_i = 1$.

It has been already proved that $f(x) = xe^{1-x}$ is a concave function over $[0,1]$. Now taking $\lambda_k = p_k$ and $x_k = q_{l/k}$ ($\sum \lambda_k = 1$ and $0 \le x_k \le 1$), we have

$$\sum_{k=1}^{n} p_k q_{l/k} e^{1-q_{l/k}} \le \left(\sum_{k=1}^{n} p_k q_{l/k}\right) e^{1-\sum p_k q_{l/k}} \tag{29}$$

or

$$\sum_{k=1}^{n} p_k / q_{l/k} e^{1-q_{l/k}} \leq q_l e^{1-q_l} \tag{30}$$

or

$$H_m(B/A) \leq H_m(B). \tag{31}$$

Similarly, it can also be proved that

$$H_n(A/B) \leq H_n(A). \tag{32}$$

THEOREM 6.

$$H_{nm}(A * B) \leq H_n(A) + H_m(B/A).$$

Proof.

$$H_{nm}(A * B) = \sum_{k=1}^{n} \sum_{l=1}^{m} p_{kl} e^{1-p_{kl}} \tag{33}$$

$$= \sum_{k=1}^{n} \sum_{l=1}^{m} p_k q_{l/k} e^{1-p_k q_{l/k}} \tag{34}$$

$$< \sum_{k=1}^{n} \sum_{l=1}^{m} p_k q_{l/k} (e^{1-p_k} + e^{1-q_{l/k}}) \tag{35}$$

(as $e^{1-x} + e^{1-y} > e^{1-xy}$; $0 \leq x, y \leq 1$; this will be shown later)

$$\doteq \sum_{k=1}^{n} p_k e^{1-p_k} + \sum_{k=1}^{n} \sum_{l=1}^{m} p_k q_{l/k} e^{1-q_{l/k}} \tag{38}$$

$$= H_n(A) + H_m(B/A). \tag{39}$$

Hence,

$$H_{nm}(A * B) \leq H_n(A) + H_m(B/A). \tag{40}$$

Equality is included considering the case when B is a null experiment.

Proof of $e^{1-x} + e^{1-y} > e^{1-xy}; \quad 0 \leq x, y \leq 1.$
We need to show

$$e^{1-x} + e^{1-y} > e^{1-xy}; \qquad 0 \leq x, y \leq 1$$

or

$$e^{-x(1-y)} + e^{-y(1-x)} > 1.$$

We know that

$$\log(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots$$

or

$$\log(1-x) \leq -x$$

or

$$1 - x \leq e^{-x}$$

Using this result, we can write that

$$e^{-x(1-y)} + \bar{e}^{y(1-x)} \geq 1 - x(1-y) + 1 - y(1-x)$$

or

$$e^{-x(1-y)} + \bar{e}^{y(1-x)} \geq 2 - x - y + 2xy$$

or

$$e^{-x(1-y)} + \bar{e}^{y(1-x)} \geq 2 - x - y + xy$$

or

$$e^{-x(1-y)} + \bar{e}^{y(1-x)} \geq 1 + (1-y)(1-x)$$

or

$$e^{-x(1-y)} + \bar{e}^{y(1-x)} > 1 \quad \text{as } 0 \leq x, y \leq 1.$$

or

$$e^{1-x} + e^{1-y} > e^{1-xy}.$$

Hence the proof.

THEOREM 7.

$$H_{nm}(A * B) \leq H_n(A) + H_m(B).$$

This result follows from Theorems 5 and 6.

The next theorem involves the entropy of an information source, so before proceeding further we shall define the entropy of an information source. Let $p(s_i)$ be the probability of a sequence of symbols s_i of length q, then $H^{(q)}$, the entropy of order q of the system, is defined as

$$H^{(q)} = \sum_i p(s_i) e^{1-p(s_i)}. \tag{41}$$

where the summation is taken over all possible sequences of length q. The entropy of the source H is defined as

$$H = \lim_{q \to \infty} \frac{1}{q} H^{(q)}. \tag{42}$$

if the limit exists. This H can be viewed as the limiting value of the average amount of information conveyed per symbol.

Suppose the source can generate symbols

$$(g_1, g_2, \ldots, g_L) = G_L \quad \text{(say)}. \tag{43}$$

Let us now consider a sequence of symbols $(x_1, x_2, \ldots, x_n)$, $x_i \in G_L$, such that x_i occurs at time t_i. Let us denote such a sequence by $(x_1/t_1, \ldots, x_n/t_n)$. Suppose $p(x_1/t_1, \ldots, x_n/t_n)$ represents the probability of the sequence $(x_1/t_1, \ldots, x_n/t_n)$. The information source is said to be stationary, if for any integer r the following is true.

$$p(x_1/t_1 + r, \ldots, x_n/t_n + r) = p(x_1/t_1, \ldots, x_n/t_n). \tag{44}$$

Let $s_q^t = (x_1/t, x_2/t+1, \ldots, x_q/t+q-1)$ and S_q^t be the finite probability space whose elementery events are sequences of the form s_t. It is to be noted that there are L^q such sequences. Therefore, the entropy of such a finite probability space S_q^t can be written as

$$H(S_q^t) = \sum_{S_q^t} p(x_1/t, \ldots, x_q/t+q-1) e^{1-p(x_1/t, \ldots, x_q/t+q-1)}. \tag{45}$$

Since the source is stationary, we can write that $s_q^t = s_q^0$. Hence

$$H(S_q^t) = H(S_q^0) = H(S_q) = H^{(q)}. \tag{46}$$

i.e., the entropy $H(S_q^t)$ does not depend on t but on the probabilistic structure of the source and q.

THEOREM 8. *For a stationary information source $H^{(k)} > H^{(n)}$ where $k > n$.*

Proof. To prove Theorem 8 it is enough to show that $H^{(q+1)} > H^{(q)}$, $q \geq 1$. Since the source can generate L distinct symbols, the probability space S_q will have L^q sequences (events) of length q. In other words, $H^{(q)}$ is computed on the basis of L^q sequences. On the other hand, $H^{(q+1)}$ will have $L^{(q+1)}$ sequences of length $(q+1)$. Consider a sequence $s_{q+1} = \{x_1, \ldots, x_q, x_{q+1}\} = \{s_q, x_{q+1}\}$. Now there are exactly L sequence of length $(q+1)$ for which s_q is same and x_{q+1} takes one of the L possible symbols. Obviously, $\sum_{x_{q+1}} p_{q+1}(s_q, x_{q+1}) = p_q(x_1, \ldots x_q)$; where the summation is taken over the L possible distinct values of x_{q+1}. Note that p_q and p_{q+1} are the probabilities of sequences of lengths q and $(q+1)$ in S_q and S_{q+1}, respectively. Let us now add $(L^{q+1} - L^q)$ number of dummy events, each with probability of zero, to the probability space S_q and generate a probability space S_{q+1}^{new}; i.e., S_{q+1}^{new} has all the events of S_q and $(L^{q+1} - L^q)$ dummy events, each with zero probability. Thus we can write

$$S_{q+1}^{new} = \left(s_q^1, d_1^1, \ldots, d_1^{L-1}, s_q^2; d_2^1, \ldots, d_2^{L-1}, \ldots s_q^{L^q}, d_{L^q}^1, \ldots, d_{L^q}^{L-1}\right)$$

where d_i^j's are dummy events each with zero probability.

Under this situation,

$$H(S_{q+1}^{new}) = H(S_q). \tag{47}$$

Now distribute the probability of s_q^i; i.e., $p_q(s_q^i)$ among s_q^i and $d_i^1, \ldots, d_i^{L-1}$ in such a manner that the following conditions hold good.

$$p_{q+1}^{new}(s_q^i) = p_{q+1}(s_q^i, x_{q+1}), \tag{48}$$

$$p_{q+1}^{new}(d_i^j) = p_{q+1}(s_q^i, x_{q+1}^j). \tag{49}$$

for $j = 1, \ldots, L-1$ and x_{q+1}^j is one of the L symbols excluding the symbol represented by x_{q+1}. Let us call this revised probability space S_{q+1}^{rev}. The probability space S_{q+1}^{rev} is now exactly identical to S_{q+1}. It may be noted that

$$p_{q+1}^{rev}(s_q') + \sum_{j=1}^{L-1} p_{q+1}^{rev}(d_i^j) = \sum_{x_{q+1} \in G_L} p_{q+1}(s_q', x_{q+1}) = p_q(s_q')$$

Hence,

$$H(S_{q+1}^{rev}) = H(S_{q+1}). \tag{50}$$

Again,

$$H(S_{q+1}^{rev}) > H(S_{q+1}^{new}) \quad \text{(by Theorem 3)}. \tag{51}$$

From Equations 47, 50, and 51 one gets $H(S_{q+1}) > H(S_q)$. Hence the theorem.

THEOREM 9. *For a stationary information source,*

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

exists.

Proof. Let us consider a sequence of length $(q+k)$ as

$$s_{q+k}^0 = (x_1/0, x_2/1, \ldots, x_q/q-1, x_{q+1}/q, \ldots, x_{q+k}/q+k-1) \tag{52}$$

and the associated finite probability space S_{q+k}^0. Now,

$$s_{q+k}^0 = (x_1/0, x_2/1, \ldots, x_q/q-1) \cap (x_{q+1}/q, \ldots, x_{q+k}/q+k-1) \tag{53}$$

$$= s_q^0 \cap s_k^q;$$

as this implies, a sequence of length $(q+k)$ in which s_k^q follows s_q^0. Thus we have

$$S_{q+k}^0 = S_q^0 * S_k^q, \tag{54}$$

i.e., the probability space S^0_{q+k} is equal to the compound (product) probabilistic space $S^0_q * S^q_k$. We have already seen that (Theorem 7) $H(A * B) \le H(A) + H(B)$. So

$$H\left(S^0_q * S^q_k\right) \le H\left(S^0_q\right) + H\left(S^q_k / S^0_q\right) \tag{55}$$

$$\le H\left(S^0_q\right) + H\left(S^q_k\right) \tag{56}$$

(by Theorem 6). Since the source is stationary, $S^q_k = S^0_k$

$$H\left(S^0_q * S^q_k\right) \le H\left(S^0_q\right) + H\left(S^0_k\right) \tag{57}$$

or

$$H\left(S^0_{q+k}\right) \le H\left(S^0_q\right) + H\left(S^0_k\right). \tag{58}$$

For the sake of notational simplicity, we can write,

$$H^{(q+k)} \le H^{(q)} + H^{(k)}. \tag{59}$$

From Equation 59, we can write

$$H^{(nk)} \le nH^{(k)}. \tag{60}$$

Taking $k = 1$,

$$H^{(n)} \le nH^{(1)} \tag{61}$$

or

$$\frac{H^{(n)}}{n} \le H^{(1)} \tag{62}$$

where $H^{(1)}$ is the entropy of the finite probability space S^0_1 (i.e., only sequence of length one). Now when the source can generate L distinct symbols, $H^{(1)} \le e^{1-1/L}$. (It has already been proved in [1] that $H^{(1)}$ attains the maximum value when $p_i = 1/L$ for all $i = ,\ldots L$.) Now since $H^{(n)}/n \le H^{(1)}$ we can write,

$$\operatorname*{Inf}_n \frac{H^{(n)}}{n} \le H^{(1)}. \tag{63}$$

Let us write

$$H = \operatorname*{Inf}_{n} \frac{H^{(n)}}{n}. \tag{64}$$

Thus for every $\epsilon > 0$, there exists a positive integer m, such that

$$\frac{H^{(m)}}{m} - H < \epsilon \tag{65}$$

or

$$\frac{H^{(m)}}{m} < H + \epsilon. \tag{66}$$

Now, we can always find a pair of integers n and r such that $n > m$ and $(r-1)m < n \le rm$. From Theorem 8 and Equations 59 and 60, we can write

$$H^{(n)} \le H^{(rm)} \le rH^{(m)} \tag{67}$$

or

$$H^{(n)} \le rH^{(m)} \tag{68}$$

or

$$\frac{H^{(n)}}{n} \le \frac{r}{n} H^{(m)} \tag{69}$$

or

$$\frac{H^{(n)}}{n} \le \frac{rm}{n} \frac{H^{(m)}}{m} \tag{70}$$

or

$$\frac{H^{(n)}}{n} < \frac{r}{r-1} \frac{H^{(m)}}{m} \tag{71}$$

(as $r/(r-1) > (rm)/n$ will be shown later) or

$$\frac{H^{(n)}}{n} < \frac{r}{r-1}(H + \epsilon) \tag{72}$$

(using Equation 66). Therefore,

$$\mathrm{Lim}_n \mathrm{Sup} \frac{H^{(n)}}{n} \le H + \epsilon \tag{73}$$

(as $n \to \infty$, $r \to \infty$ and $\lim_{r \to \infty}(r/(r-1)) \to 1$). Since ϵ is a positive number, we can write

$$\mathrm{Lim}_n \mathrm{Sup} \frac{H^{(n)}}{n} \le H. \tag{74}$$

Again, we have seen that

$$\frac{H^{(n)}}{n} \ge H.$$

Therefore,

$$\mathrm{Lim}_n \mathrm{Inf} \frac{H^{(n)}}{n} \ge H. \tag{75}$$

Now, combining Equations 74 and 75, we get

$$\lim_{n \to \infty} \frac{H^{(n)}}{n} = H;$$

hence the proof.

Proof of $(r/(r-1) > (rm)/n)$

$$n/(rm) = (rm - x)/(rm) \quad \text{where } 0 \le x \le m-1$$
$$= 1 - x/(rm)$$

since $0 \le x \le m-1$, $x/(rm) < 1/r$.
Now

$$(r-1)/r = 1 - 1/r < 1 - x/(rm)$$
$$= (rm - x)/(rm)$$
$$= n/(rm).$$

Thus

$$(r-1)/r < n/(rm) \Rightarrow r/(r-1) > (rm)/n.$$

then

$$H'(A)=\frac{1}{n}\sum_{x_i\in X} f(\mu_A(x_i)) \tag{83}$$

and

$$H'(B)=\frac{1}{n}\sum_{x_i\in X} f(\mu_B(x_i)) \tag{84}$$

$$H'(A\cup B)=\frac{1}{n}\sum_{x\in X} f(\mu_{A\cup B}(x_i)) \tag{85}$$

and

$$H'(A\cap B)=\frac{1}{n}\sum_{x_i\in X} f(\mu_{A\cap B}(x_i)). \tag{86}$$

We know that

$$\mu_{A\cup B}(x_i)=\max\{\mu_A(x_i),\mu_B(x_i)\},\qquad \forall x_i\in X \tag{87}$$

and

$$\mu_{A\cap B}(x_i)=\min\{\mu_A(x_i),\mu_B(x_i)\},\qquad \forall x_i\in X. \tag{88}$$

Therefore, we can write

$$\sum_{x_i\in X} f(\mu_{A\cup B}(x_i))=\sum_{x_i\in D} f(\mu_A(x_i)+\sum_{x_i\in D'} f(\mu_B(x_i)). \tag{89}$$

Similarly,

$$\sum_{x_i\in X} f(\mu_{A\cap B}(x_i))=\sum_{x_i\in D'} f(\mu_A(x_i))+\sum_{x_i\in D} f(\mu_B(x_i)). \tag{90}$$

Now, taking the sum of Equations 89 and 90, we obtain the following.

$$\sum_{x_i\in X} f(\mu_{A\cup B}(x_i))+\sum_{x_i\in X} f(\mu_{A\cap B}(x_i))$$
$$=\sum_{x_i\in X} f(\mu_A(x_i))+\sum_{x_i\in X} f(\mu_B(x_i)) \tag{91}$$

as $D\cup D'=X$ and $D\cap D'=\varnothing$.

then

$$H'(A) = \frac{1}{n} \sum_{x_i \in X} f(\mu_A(x_i)) \tag{83}$$

and

$$H'(B) = \frac{1}{n} \sum_{x_i \in X} f(\mu_B(x_i)) \tag{84}$$

$$H'(A \cup B) = \frac{1}{n} \sum_{x \in X} f(\mu_{A \cup B}(x_i)) \tag{85}$$

and

$$H'(A \cap B) = \frac{1}{n} \sum_{x_i \in X} f(\mu_{A \cap B}(x_i)). \tag{86}$$

We know that

$$\mu_{A \cup B}(x_i) = \max\{\mu_A(x_i), \mu_B(x_i)\}, \qquad \forall x_i \in X \tag{87}$$

and

$$\mu_{A \cap B}(x_i) = \min\{\mu_A(x_i), \mu_B(x_i)\}, \qquad \forall x_i \in X. \tag{88}$$

Therefore, we can write

$$\sum_{x_i \in X} f(\mu_{A \cup B}(x_i)) = \sum_{x_i \in D} f(\mu_A(x_i) + \sum_{x_i \in D'} f(\mu_B(x_i)). \tag{89}$$

Similarly,

$$\sum_{x_i \in X} f(\mu_{A \cap B}(x_i)) = \sum_{x_i \in D'} f(\mu_A(x_i)) + \sum_{x_i \in D} f(\mu_B(x_i)). \tag{90}$$

Now, taking the sum of Equations 89 and 90, we obtain the following.

$$\sum_{x_i \in X} f(\mu_{A \cup B}(x_i)) + \sum_{x_i \in X} f(\mu_{A \cap B}(x_i))$$
$$= \sum_{x_i \in X} f(\mu_A(x_i)) + \sum_{x_i \in X} f(\mu_B(x_i)) \tag{91}$$

as $D \cup D' = X$ and $D \cap D' = \varnothing$.

Now dividing Equation 91 by n, we get

$$H'(A \cup B) + H'(A \cap B) = H'(A) + H'(B) \tag{92}$$

and this completes the proof.

6. CONCLUSION

A new definition of probabilistic entropy that can also be extended to define the nonprobabilistic entropy of a fuzzy set has recently been introduced by the authors [1, 2]. The present work is an attempt to investigate some desirable properties of the new definition. In this regard several theorems have been stated and proved. Some of the theorems are for a compound probabilistic experiment. These properties establish that the mathematical structure of the new exponential entropic function is close to that of Shannon's logarithmic entropy, thereby enhancing the range of its applications. However, many other properties still need to be investigated.

The authors gratefully acknowledge useful discussions with Professors D. Duttamajumdar and B. K. Roy.

REFERENCES

1. N. R. Pal and S. K. Pal, Object background segmentation using new definitions of entropy, *IEE Proc.*, 136(4) (Pt. E):284–295 (1989).
2. N. R. Pal and S. K. Pal, Entropy: A new definition and its applications, *IEEE Trans. Syst. Man Cybern.*, 21(5):1260–1270 (1991).
3. N. R. Pal and S. K. Pal, Higher order fuzzy entropy and hybrid entropy of a set, *Inform. Sci.*, 61(3):211–231 (1992).
4. N. R. Pal and S. K. Pal, Image model, Poisson distribution and object extraction, *Int. J. Patt. Recogn. Artif. Intell.*, 5(3):459–483 (1991).
5. N. R. Pal and S. K. Pal, Some information measures on fuzzy sets and their applications to image processing, NACONECS-1989, Tata McGraw-Hill, New Delhi:94–96 (1989).
6. C. E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.* 27:379–423 (July, 1948).

Received 10 January 1991; revised 20 June 1991