SUP-ENTROPY OF DEGREE α

A. Awad*

Yarmouk University, Department of Statistics
Irbid, Jordan

and

M. Azzam and D. Hussein

Department of Mathematics
University of Jordan
Amman, Jordan

الخلاصـــة :

يوضح هذا البحث عدم امكانية تعميم مقياس الانتظام من الدرجة (أ) ، من حالة التوزيعات المنفصلة الى حالة التوزيعات المتصلة بطريقة مباشرة . كما يوضح فشل اسلوب التحليل غير المالوف في اجراء مثل هذا التعميم . لذا تم تعريف نوع جديد من مقياس الانتظام اسميناه مقياس الانتظام الاعظم من الدرجة (أ) ، وتم ايضاح هذا المقياس الجديد الذي يمكن تعميمه من حالة التوزيعات المنفصلة الى حالة التوزيعات المتصلة بطريقة مباشرة . إضافة إلى دراسة خواص هذا المقياس الجديد .

ABSTRACT

In this paper, we show that the entropy of degree α cannot be generalized in a natural way from the discrete to the continuous case. Since the non-standard analysis fails to do this extension, we define what we call the sup-entropy of degree α which can be extended naturally from the discrete case to the continuous case. Moreover, we study the properties of the new suggested entropy. A brief review of the application areas of the new entropy is given.

*Address for correspondence :
Department of Mathematics
University of Jordan
Amman, Jordan

SUP-ENTROPY OF DEGREE α

1. INTRODUCTION

There are four types of information measures:

(*i*) Parametric measures such as the Fisher information.

(*ii*) Non-parametric measures such as the Kullback-Leibler measure.

(*iii*) Entropy measures such as the Shannon entropy, and

(*iv*) Statistical information measures such as the likelihood function.

For more details the reader is referred to Ferentinos and Papaioannon [1] and Basu [2].

One of the oldest and most widely used measures of entropy is Shannon's entropy [3]. This measure proved to be successful in the analysis of communication systems. It was also the basis for several extensions, see for example Csiszar [4]. Several characterizations of this measure have been considered in the literature. These characterizations depend on one of the following approaches:

(*i*) Maximum probability approach (see for example Kapur [5]).

(*ii*) Probabilistic axiomatic approach (see for example Frote and Ng [6]).

(*iii*) Non-probabilistic axiomatic approach (see for example Cerny and Brunovsky [7]).

(*iv*) Parent function approach (see for example Behara and Nath [8]).

We will now review some of the extensions of the Shannon entropy.

Let X be a discrete random variable taking values $x_1,\dots,x_n$ with probabilities $p_1,\dots,p_n$. One of the well-known entropies is the Shannon entropy which is defined by:

$$H_n(X) = H_n(p_1,\dots,p_n) = -\sum_{i=1}^{n} p_i \log p_i \,. \qquad (1)$$

If X is a continuous random variable with density $f(x)$ then its Shannon entropy is defined by:

$$H(X) = -\int f(x) \log f(x)\, dx \,.$$

One way to generalize the Shannon entropy from the discrete case to the continuous case is to partition the range of X into n intervals $A_1,\dots,A_n$ of equal lengths Δx, then

$$p_i = P(X \in A_i) = f(x_i)\,\Delta x$$

where x_i is some point in A_i. So one expects

$$H(X) = \lim_{\Delta x \to 0} H_n(p_1,\dots,p_n)$$

$$= \lim_{\Delta x \to 0}\left[-\sum_{i=1}^{n}\{f(x_i)\,\Delta x \log\{f(x_i)\,\Delta x\}\}\right].$$

It is well known that the entropy of a continuous distribution defined by Shannon [3] is not a natural extension of the entropy of a discrete distribution despite their analogous forms. Ingels (reference [9], pp. 91–92) showed that if X is a continuous random variable with density $f(x)$, then the average Shannon entropy of X is

$$S(X) = \lim_{\Delta x \to 0}\left[-\sum_{i=-\infty}^{\infty}(f(x_i)\,\Delta x)\log_2(f(x_i)\,\Delta x)\right]$$

$$= -\int_{-\infty}^{\infty} f(x)\log_2 f(x)\,dx - \lim_{\Delta x \to 0}(\log_2 \Delta x).$$

The divergent term, $\lim_{\Delta x \to 0}(\log_2 \Delta x)$, will not allow us to define the Shannon entropy of a continuous random variable X as

$$H(X) = -\int_{-\infty}^{\infty} f(x)\log_2 f(x)\,dx.$$

Ozeki [10] used non-standard analysis to overcome this difficulty. He has shown that if X is a continuous random variable then for any positive infinitesimal δx the non-standard Shannon's entropy of X is

$$^*H(X, \delta x) = {}^*\phi(\delta x) - {}^*\log_2(\delta x)$$

where $^*a$ denotes the hyperreal number a. He has noted that for any positive infinitesimal δx, the standard part of $^*\phi(\delta x)$,

$$\mathrm{st}(^*\phi(\delta x)) = -\int f(x)\log_2 f(x)\,dx$$

coincides with $H(X)$. The second term $-{}^*\log_2(\delta x)$ is a positive infinite hyperreal number and it is independent of the density of X. This justifies the use of $H(X)$ as the entropy of X. For the details of the terminology of non-standard analysis, the reader is referred to Keisler [11].

Awad [12] suggested another approach to overcome this difficulty. He suggested an extension of the Shannon entropy, namely,

$$A_n(X) = -\sum_{i=1}^{n} p_i \log(p_i/s) \tag{2}$$

where $s = \sup\{p_1,\dots,p_n\}$ and $p_1,\dots,p_n$ are the probabilities assumed by the discrete random variable X. He has shown that this definition can be extended naturally to the continuous case as

$$A(X) = -\int \mathrm{f}(x) \log(\mathrm{f}(x)/s)\,\mathrm{d}x$$

where $s = \sup_x \mathrm{f}(x)$.

The Shannon entropy has been generalized in two directions in the literature, namely, the entropy of order α (Renyi's entropy) and the entropy of degree α. If X is a discrete random variable with probabilities, $p_1,\dots,p_n$ then the entropy of order α is

$${}_\alpha H_n(X) = R_\alpha(X) = \frac{1}{1-\alpha} \log \sum_{i=1}^{n} p_i^\alpha,\ \alpha \neq 1,\ \alpha > 0 \tag{3}$$

and the entropy of degree α is

$$H_n^\alpha(X) = \frac{1}{2^{1-\alpha}-1}\left(\sum_{i=1}^{n} p_i^\alpha - 1\right), \quad \alpha \neq 1,\ \alpha > 0. \tag{4}$$

These are extensions of the Shannon entropy Equation (1), since

$$\lim_{\alpha\to 1} {}_\alpha H_n(X) = \lim_{\alpha\to 1} H_n^\alpha(X) = H_n(X).$$

In Section 2 we will show that ${}_\alpha H_n(X)$ and $H_n^\alpha(X)$ cannot be generalized naturally to the continuous case. The non-standard analysis will fail to extend $H_n^\alpha(X)$ to the continuous case even though it is successful in both $H_n(X)$ and ${}_\alpha H_n(X)$. However, Awad's approach will be successful in both the cases $H_n^\alpha(X)$ and ${}_\alpha H_n(X)$. Section 3 gives some properties of the new generalizations of ${}_\alpha H_n(X)$ and $H_n^\alpha(X)$. Section 4 gives a brief survey of statistical applications which support the use of the new entropies instead of the entropies (1), (3), and (4).

2. SUP-α-ENTROPY

Let X be a continuous random variable with density $\mathrm{f}(x)$; then it can be shown that:

$$\lim_{\Delta x\to 0} {}_\alpha H_n(X) = \frac{1}{1-\alpha} \lim_{\Delta x\to 0} \log \sum_{i=1}^{n} (\mathrm{f}(x_i)\Delta x)^\alpha$$

$$= {}_\alpha H(X) - \lim_{\Delta x\to 0} (\log \Delta x)$$

where

$${}_\alpha H(X) = \frac{1}{1-\alpha} \log \int \mathrm{f}^\alpha(x)\,\mathrm{d}x.$$

Using Ozeki's approach it can be shown that

$${}^*_\alpha H(X,\ \delta x)\ {}^*R_\alpha(\delta x)\ -\ {}^*\log(\delta x)$$

where $\mathrm{st}({}^*R_\alpha(\delta x)) = {}_\alpha H(X)$.

On the other hand it can be shown that

$$\lim_{\Delta x\to 0} H_n^\alpha(X) = H^\alpha(X).\ \lim_{\Delta x\to 0} (\Delta x)^{\alpha-1}$$

$$+ \frac{1}{2^{1-\alpha}-1}(\lim(\Delta x)^{\alpha-1} - 1),$$

where

$$H^\alpha(X) = \frac{1}{2^{1-\alpha}-1}\left[\int \mathrm{f}^\alpha(x)\mathrm{d}x - 1\right].$$

It is clear that if $\alpha > 1$ then $\lim_{\Delta x\to 0} H_n^\alpha(X)$ is free of $H^\alpha(X)$ and hence the non-standard analysis will not help in generalizing $H_n^\alpha(X)$ to the continuous case.

Now, we suggest the following extended α-entropy measures.

Definition 1: Let X be a discrete random variable assuming probabilities $p_1,\dots,p_n$ where $p_i \geq 0$, $i = 1,\dots,n$ and $\sum_{i=1}^{n} p_i = 1$. The generalized entropy of order α of X is

$${}_\alpha A_n(X) = \frac{1}{1-\alpha} \log \sum_{i=1}^{n} (p_i/s)^{\alpha-1} p_i,$$

$$\alpha \neq 1 \text{ and } \alpha > 0, \tag{5}$$

and the generalized entropy of degree α of X is

$$A_n^\alpha(X) = \frac{1}{2^{1-\alpha}-1}\left(\sum_{i=1}^{n} (p_i/s)^{\alpha-1} p_i - 1\right),$$

$$\alpha \neq 1 \text{ and } \alpha > 0 \tag{6}$$

where $s = \sup\{p_1,\dots,p_n\}$.

Using L'Hopital's rule, it can be shown that

$$\lim_{\alpha\to 1} {}_\alpha A_n(X) = \lim_{\alpha\to 1} A_n^\alpha(X) = A_n(X),$$

so ${}_\alpha A_n(X)$ and $A_n^\alpha(X)$ are extensions of $A_n(X)$ defined in Equation (2).

If X is a continuous random variable with density $\mathrm{f}(x)$ then it can be shown that

$$\lim_{\Delta x\to 0} {}_\alpha A_n(X) = {}_\alpha A(X) = \frac{1}{1-\alpha} \log \int (f(x)/s)^{\alpha-1} f(x)dx$$

and

$$\lim_{\Delta x\to 0} A_n^\alpha(X) = A^\alpha(X)$$
$$= \frac{1}{2^{1-\alpha}-1}\left[\int (f(x)/s)^{\alpha-1} f(x)\,dx - 1\right],$$

where $s = \sup_x f(x)$. So Definition 1 has been generalized naturally to the continuous case. Therefore the use of the sup-method is more appropriate than the non-standard analysis method to generalize the definitions of entropies from the discrete to the continuous case.

3. PROPERTIES OF $A_n^\alpha(X)$

Definition 2: (see for example Aczel and Darozy [13], pp. 51–53).

Let

$$\Gamma_n = \{(p_1,\ldots,p_n);\ 0\le p_i\le 1,\ i=1,\ldots,n \text{ and } \sum_{i=1}^n p_i = 1\},$$

$$\Delta_n = \{(p_1,\ldots,p_n);\ 0\le p_i \le 1 \text{ and } \sum_{i=1}^n p_i \le 1\}$$

and

$I_n : \Gamma_n \to R$ be a sequence of real valued functions on Γ_n. I_n is said to be

(*i*) Decisive if $I_2(1,0) = I_2(0,1) = 0$.

(*ii*) Bounded from above if $I_2(1-p,p) \le K$ for some constant K.

(*iii*) Normalized if $I_2(½,½) = 1$.

(*iv*) Monotonic if the function $p \to I_2(1-p,p)$ is non-decreasing on $[0,½]$.

(*v*) Measurable if $p\to I_2(1-p,p)$ is Lebesgue measurable on $]0,1[$ (or on $[0,1]$).

(*vi*) Small for small probabilities if $\lim_{p\to 0^+} I_2(1-p,p) = 0$.

(*vii*) Stable at p_0 if $\lim_{q\to 0^+} I_2(p_0,q) = I_1(p_0)$ provided that $p_0 \in]0,1]$ and $p_0+q\le 1$.

Theorem 1: The entropies $A_n^\alpha : \Gamma_n \to R (n=2,3,\ldots)$ of degree α are decisive, bounded, measurable, small for small probabilities, and stable. They are not normalized and not monotone.

Proof: It is clear that

$$A_2^\alpha(1-x,x) = \begin{cases} \dfrac{1}{2^{1-\alpha}-1}[(1-x)\{1+(x/(1-x))^\alpha\}-1] & \text{if } 0\le x\le ½ \\ \dfrac{1}{2^{1-\alpha}-1}[x\{1+((1-x)/x)^\alpha\}-1] & \text{if } ½ < x \le 1. \end{cases}$$

Hence

(*i*) $A_2^\alpha(1,0) = A_2^\alpha(0,1) = 0$, so A_n^α is decisive.

(*ii*) $A_2^\alpha((1-x),x) \le 1/(2^{1-\alpha}-1)$, so A_n^α is bounded from above.

(*iii*) $A_2^\alpha(½,½) = 0$, so A_n^α is not normalized.

(*iv*) Note that if $x\in[0,½]$ then

$$\frac{\partial A_2^\alpha}{\partial x} = \frac{1}{2^{1-\alpha}-1}\left[\left(\frac{x}{1-x}\right)^{\alpha-1}\frac{\alpha-x}{1-x} - 1\right]$$

may be positive and may be negative when $\alpha > 1$ or $0<\alpha<1$. Therefore A_2^α is not monotone.

(*v*) It is clear that the function $p\to A_2^\alpha(1-p,p)$ is Lebesgue measurable on $]0,1[$.

(*vi*) $\lim_{x\to 0^+} A_2^\alpha(1-x,x) = 0$, so A_n^α is small for small probabilities.

(*vii*) Consider p_0 fixed such that $0<q+p_0\le 1$. In $A_2^\alpha(p_0,q)$, we have

$$\lim_{q\to 0^+} A_2^\alpha(p_0,q) = \frac{1}{2^{1-\alpha}-1}[p_0-1]\ A_1^\alpha(p_0).$$

Hence it is stable at p_0.

Definition 3: The sequence of functions $I_n : \Delta_n \to R(n=1,2,\ldots)$ or $\Gamma_n \to R(n=2,3,\ldots)$ is:

(*i*) Symmetric if for all n

$$I_n(p_1,\ldots,p_n) = I_n(p_{k(1)},\ldots,p_{k(n)})$$

for all $(p_1,\ldots,p_n) \in \Gamma_n$, where k is an arbitrary permutation on $\{1,\ldots,n\}$;

(*ii*) Expansible (that is, null events discarded), if for all n

$$I_n(p_1,\ldots,p_n) = I_{n+1}(0,p_1,\ldots,p_n)$$
$$= I_{n+1}(p_1,0,p_2,\ldots,p_n) = \ldots$$
$$= I_{n+1}(p_1,\ldots,p_n,0)\ ;$$

(*iii*) Nonnegative if for all n

$$I_n(p_1,\ldots,p_n) \ge 0\ ;$$

(*iv*) Maximal if for all n

$$I_n(p_1,\ldots,p_n) \le I_n(1/n,\ldots,1/n),$$

and it is minimal if for all n

$I_n(p_1,\dots,p_n) \geq I_n(1/n,\dots,1/n)$,

for all $(p_1,\dots,p_n) \in \Gamma_n$;

(v) Continuous if for all n, I_n is continuous on Γ_n.

Note that all parts of this definition are given in Aczel and Daroczy (reference [13], pp. 51–53), except the minimal property, which reflects the statistical fact that a uniform distribution is non-informative from a Bayesian point of view.

Theorem 3: The entropies $A_n^\alpha : \Gamma_n \to R$ of degree α are symmetric, expansible, minimal, nonnegative, and continuous. They are not maximal.

Proof: It is clear from the definition of

$$A_n^\alpha(p_1,\dots,p_n) = \frac{1}{2^{1-\alpha}-1}\left[\sum_{i=1}^{n}(p_i/s)^{\alpha-1}p_i - 1\right],$$

that:

(i) The cumulative property of the summation operator implies that A_n^α is symmetric;

(ii) Since $\sup\{p_1,\dots,p_n\} = \sup\{p_1,\dots,p_n,0\}$ and the arguments of A_n^α are added to each other through $(p_i/s)^{\alpha-1}p_i$. $A_n^\alpha(p_1,\dots,p_n) = A_n^\alpha(0,p_1,\dots,p_n)$. This together with the symmetry property implies that A_n^α is expansible.

(iii) Note that for all i, $p_i/s \leq 1$ and hence:

$$\sum\left(\frac{p_i}{s}\right)^{\alpha-1} p_i \begin{cases} \leq 1, & \text{if} \quad \alpha \geq 1 \\ \geq 1, & \text{if} \quad \alpha < 1 . \end{cases}$$

This, together with the fact that:

$$2^{1-\alpha}-1 \begin{cases} \leq 0, & \text{if} \quad \alpha \geq 1 \\ \geq 0, & \text{if} \quad \alpha < 1 , \end{cases}$$

implies that $A_n^\alpha(p_1,\dots,p_n)$ is nonnegative.

(iv) Since $A_n^\alpha(p_1,\dots,p_n)$ is nonnegative and $A_n^\alpha(1/n,\dots,1/n) = 0$, $A_n^\alpha(p_1,\dots,p_n)$ is minimal. And hence it is not maximal.

(v) The continuity property is obvious.

Consider a discrete bivariate random vector with joint probabilities

$$(p_{11},\dots p_{1n};p_{21},\dots,p_{2n};\dots;p_{m1},\dots,p_{mn}) \in \Gamma_{mn}.$$

Let $p_i = \sum_{j=1}^{n} p_{ij}$ for $i = 1,\dots,m$ and $q_j = \sum_{i=1}^{m} p_{ij}$ for $i = 1,\dots,n$ be the corresponding marginal probabilities. For $i = 1,\dots,m$ and $j = 1,\dots,n$ let $q_{ij} = p_{ij}/p_i$ be the corresponding conditional probabilities.

For a given non-negative real number $\alpha \neq 1$, let $I_{mn}^\alpha : \Gamma_{mn} \to R$, $I_m^\alpha : \Gamma_m \to R$ and $I_n^\alpha : \Gamma_n \to R$ be sequences of real valued functions.

Now we will use the definition of an additive function as given by Aczel and Daroczy (reference [13], p. 52) and generalize the definitions of strong additive and the sum property, to state:

Definition 4:

(i) Additive of degree α if

$$I_{mn}^\alpha(p_1q_1,\ p_1q_2,\dots,p_1q_n;\ p_2q_1,\dots,p_2q_n;\dots;\ p_mq_1,\dots,p_mq_n)$$
$$= I_m^\alpha(p_1,\dots,p_m) + I_n^\alpha(q_1,\dots,q_n) + (2^{1-\alpha}-1)\ I_m^\alpha(p_1,\dots,p_m).\ I_n^\alpha(q_1,\dots,q_n)$$

for all $(p_1,\dots,p_m) \in \Gamma_m$, $(q_1,\dots,q_n) \in \Gamma_n$, $n = 2,3,\dots$ $m = 2,3,\dots$

(ii) Strong additive of degree α if

$$I_{mn}^\alpha(p_1q_{11},\dots,p_1q_{1n};\dots;\ p_mq_{m1},\dots,p_mq_{mn}) = I_m^\alpha(p_1,\dots,p_m) + \sum_{j=1}^{m} I_n^\alpha\ (q_{j1},\dots,q_{jn})\left(\frac{p_j}{s_p}\right)^{\alpha-1} p_j.$$

for all $(p_1,\dots,p_m) \in \Gamma_m$, $(q_{j1},\dots,q_{jn}) \in \Gamma_n$, $j=1,\dots,m$, $m=2,3,\dots$; $n=2,3,\dots$, where $s_p = \sup\{p_1,\dots,p_n\}$.

(iii) I_n^α satisfies the sum property of degree α if there exists a function $g_{s_p,\alpha}$, measurable in $]0,1[$ such that $I_n^\alpha(p_1,\dots,p_n) = \sum_{k=1}^{n} g_{s_p,\alpha}(p_k)$ for all $(p_1,\dots,p_n) \in \Gamma_n$ $(n = 2,3,\dots)$.

Theorem 4: The entropies $A_n^\alpha : \Gamma_n \to R$ of degree α satisfy the sum property, and they are additive of degree α.

Proof (i): Set $P = (p_1,\dots,p_m)$, $Q = (q_1,\dots,q_n)$ and use the notation

$$I_{mn}^\alpha(p_1Q, p_2Q,\dots,p_mQ) = I_m^\alpha(P) + I_n^\alpha(Q) + (2^{1-\alpha}-1)\, I_m^\alpha(P).\, I_m^\alpha(Q)$$

for additivity of degree α.

Let $s_q = \sup\{q_1,\dots,q_n\}$ and $s = \sup\{p_1Q,\dots,p_mQ\}$.

It is clear that $s = s_p\,.\,s_q$ and

$$(2^{1-\alpha}-1)\,[A^\alpha_{mn}(p_1Q,\dots,p_mQ)-A^\alpha_m(P)-A^\alpha_n(Q)]$$
$$=\sum_{j=1}^{n}\sum_{i=1}^{m}\left(\frac{p_iq_j}{s_ps_q}\right)^{\alpha-1}p_iq_j-\sum_{i=1}^{m}\left(\frac{p_i}{s_p}\right)^{\alpha-1}p_i-\sum_{j=1}^{n}\left(\frac{q_j}{s_p}\right)^{\alpha-1}q_j+1$$
$$=\left[\sum_{i=1}^{m}\left(\frac{p_i}{s}\right)^{\alpha-1}p_i-1\right]\left[\sum_{j=1}^{n}\left(\frac{q_j}{s_q}\right)^{\alpha-1}q_j-1\right]$$
$$=(2^{1-\alpha}-1)^2\,A^\alpha_m(P).\,A^\alpha_n(Q).$$

Hence A^α_n is additive of degree α.

(*ii*) A^α_n satisfies the sum property with

$$g_{s_p,\alpha}(p_k)=\frac{1}{2^{1-\alpha}-1}\left[\left(\frac{p_k}{s_p}\right)^{\alpha-1}-1\right]p_k.$$

Theorem 5: A^α_n is strongly additive with degree α if $s_{q_j}=s_{q^*}$ for all j, where $s_{q_j}=\sup\{q_{j1},\dots,q_{jn}\}$ $i=1,\dots,m$

$$s_{q^*}=\sup\{q_{11},q_{12},\dots,q_{mn}\}$$

Proof: Note that

$$A^\alpha_m(p_1,\dots,p_m)+\sum_{j=1}^{m}\left(\frac{p_j}{s_p}\right)^{\alpha-1}p_j\,A^\alpha_n(q_{j1},\dots,q_{jn})$$
$$=\frac{1}{2^{1-\alpha}-1}\left[\sum_{j=1}^{m}\left(\frac{p_j}{s_p}\right)^{\alpha-1}p_j-1\right.$$
$$\left.+\sum_{j=1}^{m}\left(\frac{p_j}{s_p}\right)^{\alpha-1}p_j\left\{\sum_{r=1}^{n}\left(\frac{q_{jr}}{s_{q_j}}\right)^{\alpha-1}q_{jr}-1\right\}\right]$$
$$=\frac{1}{2^{1-\alpha}-1}\left[\sum_{j=1}^{m}\sum_{r=1}^{n}\left(\frac{q_{jr}\,p_j}{s_{q_j}\,s_p}\right)^{\alpha-1}q_{jr}\,p_j-1\right]$$
$$=A_{mn}(p_1q_{11},\dots,p_1q_{1n};\dots;p_mq_{m1},\dots,p_mq_{mn})$$

if $s_{q_j}=s_{q^*}$ for all j.

Definition 5.

(*i*) Let $(\pi_{11},\dots,\pi_{1n};\pi_{21},\dots,\pi_{2n};\dots;\pi_{m1},\dots,\pi_{mn})\in\Gamma_{mn}$.

Set $p_i=\sum_{j=1}^{n}\pi_{ij}$, $q_j=\sum_{i=1}^{m}\pi_{ij}$, $i=1,\dots,m, j=1,\dots,n$
$s_m=\sup\{p_1,\dots,p_m\}$, $s_n=\sup\{q_1,\dots,q_n\}$, and
$s=\sup\{\pi_{11},\dots,\pi_{1n};\dots;\pi_{m1},\dots,\pi_{mn}\}$. The discriminant function of Γ_{mn} of degree α is

$$G(n,m,s_n,s_m,s,\alpha)=\left[\left(\frac{ns}{s_m}\right)^{\alpha-1}+\left(\frac{ms}{s_n}\right)^{\alpha-1}\right]^{-1}.$$

(*ii*) I^α_n is subadditive if

$$I^\alpha_{mn}(\pi_{11},\dots,\pi_{1n};\pi_{21},\dots,\pi_{2n};\dots;\pi_{m1},\dots,\pi_{mn})$$
$$\le I^\alpha_m(p_1,\dots,p_m)+I_n(q_1,\dots,q_n)$$

for all $(\pi_{11},\dots,\pi_{mn})\in\Gamma_{mn}$ and all m and n where p_i and q_j as in (*i*) above.

Note that (*ii*) is given in Aczel and Daroczy (Reference [13], p. 52).

Theorem 6: A^α_n is subadditive if $0<\alpha<1$ and the discriminant function of degree α is less than or equal to 1.

Proof: Note that

$$(2^{1-\alpha}-1)\left[A^\alpha_m\left(\sum_{i=1}^{n}\pi_{i1},\dots,\sum_{i=1}^{n}\pi_{im}\right)+A^\alpha_n\left(\sum_{j=1}^{m}\pi_{ij},\dots,\sum_{j=1}^{m}\pi_{nj}\right)\right]$$
$$=s_m^{1-\alpha}\sum_{j=1}^{m}\left(\sum_{i=1}^{n}\pi_{ij}\right)^\alpha+s_n^{1-\alpha}\sum_{i=1}^{n}\left(\sum_{j=1}^{m}\pi_{ij}\right)^\alpha-2=J,\ \text{say}.$$

Using Holder's inequality we obtain that for a given j,

$$\left(\sum_{i=1}^{n}\pi_{ij}\right)^\alpha\ge n^{\alpha-1}\sum_{i=1}^{n}\pi^\alpha_{ij};\ 0<\alpha<1.$$

Hence

$$j\ge n^{\alpha-1}\,s_m^{1-\alpha}\sum_{j=1}^{m}\sum_{i=1}^{n}\pi^\alpha_{ij}+m^{\alpha-1}\,s_n^{1-\alpha}\sum_{i=1}^{n}\sum_{j=1}^{m}\pi^\alpha_{ij}-2,$$
$$\frac{1}{2^{1-\alpha}-1}J\ge\left[\left(\frac{ns}{s_m}\right)^{\alpha-1}+\left(\frac{ms}{s_n}\right)^{\alpha-1}\right]A^\alpha_{nm}(\pi_{11},\dots,\pi_{1m};\dots;\pi_{n1},\dots,\pi_{nm}).$$

So, $A^\alpha_{nm}(\pi_{11},\dots,\pi_{1m};\dots;\pi_{n1},\dots,\pi_{nm})$

$$\le G(n,m,s_n,s_m,s,\alpha)\left[A^\alpha_m\left(\sum_{i=1}^{n}\pi_{i1},\dots,\sum_{i=1}^{n}\pi_{im}\right)+A^\alpha_n\left(\sum_{j=1}^{m}\pi_{1j},\dots,\sum_{j=1}^{m}\pi_{nj}\right)\right].$$

Therefore A^α_n is subadditive.

Definition 6 [14].

A sequence of function $I_n:\Gamma_n\to R\,(n=2,3,\dots)$ is said to satisfy the independence inequality if $I_{nm}(\pi_{11},\dots,\pi_{1m};\pi_{21},\dots,\pi_{2m};\dots;\pi_{n1},\dots,\pi_{nm})$
$\le I_{nm}(p_1q_1,\dots,p_1q_m;p_2q_1,\dots,p_2q_m;\dots;p_nq_1,\dots,p_nq_m)$
for all n,m and all $(\pi_{11},\dots,\pi_{nm})\in\Gamma_{nm}$ with π_{ij}, p_i and q_j as given in Definition 5.

Theorem 7: A^α_n satisfies the independence inequality if $0<\alpha<1$ and the discriminant function of degree α of A^α_n is less than or equal to 1.

Proof: The proof depends on a theorem given by El-Sayed [14]. Since $0<\alpha<1$,

$$A^{\alpha}_{mn}(p_1q_1,\ldots,p_mq_n) \geq A^{\alpha}_{m}(p_1,\ldots,p_m)+A^{\alpha}_{n}(q_1,\ldots,q_n),$$

from the additivity of degree α.

Since the discriminant function is less than or equal to one, A^{α}_{n} is subadditive, i.e.

$$A^{\alpha}_{mn}(\pi_{11},\ldots,\pi_{mn}) \leq A^{\alpha}_{m}(p_1,\ldots,p_m)+A^{\alpha}_{n}(q_1,\ldots,q_n).$$

Therefore A^{α}_{n} satisfies the independence inequality.

Definition 7: I^{α}_{n} is said to be k-recursive of degree α if there exist three functions $g_1(s_p, s, \alpha)$, $f_k(s_q, s, \alpha, p_1)$ and $g_2(\delta_q, \delta, \alpha, p_1)$ such that

$$I^{\alpha}_{n+k-1}(p_1q_1,\ldots,p_1q_k;p_2,\ldots,p_n)=g_1(s_p,s,\alpha)I^{\alpha}_{n}(p_1,\ldots,p_n)$$
$$+f_k(s_q,s,\alpha,p_1)I^{\alpha}_{k}(q_1,\ldots,q_k)+g_2(s_q,s,\alpha,p_1)$$

where $s_p=\sup\{p_1,\ldots,p_n\}$, $s_q=\sup\{q_1,\ldots,q_k\}$ and $s=\sup\{p_1q_1,\ldots,p_1q_k,p_2,\ldots p_n\}$.

It is clear that:

(*i*) if $\alpha=1$, $g_1(s_q,s,\alpha,p_1)=1$, $g_2(s_q,s,\alpha,p_1)=0$ and $f_k(s_q,s,\alpha,p_1)=G_k(p_1)$ then this definition reduces to the k-generalized recursive property given in Ebanks [16].

(*ii*) if $k=2$, $g_1(s_p,s,\alpha)=1$, $g_2(s_q,s,\alpha,p_1)=0$ and $f_k(s_q,s,\alpha,p_1)=p_1^{\alpha}$ then this definition reduces to the recursive property of degree α given in Aczel and Daroczy ([13]; p. 186). Moreover if $\alpha=1$ then this definition reduces to the recursivity property given in Aczel and Daroczy ([13]; p. 51).

Theorem 8: A^{α}_{n} is the k-recursive of degree α.

Proof:

$$A^{\alpha}_{n+k-1}(p_1q_1,\ p_1q_2,\ldots,p_1q_k;\ p_2,\ldots,p_n)$$
$$=\frac{1}{2^{1-\alpha}-1}\left[\sum_{i=1}^{k}(p_1q_i/s)^{\alpha-1}p_1q_i+\sum_{j=2}^{n}(p_j/s)^{\alpha-1}p_j-1\right]$$
$$=(s_p/s)^{\alpha-1}A^{\alpha}_{n}(p_1,\ldots,p_n)+p_1^{\alpha}(s_q/s)^{\alpha-1}A^{\alpha}_{k}(q_1,\ldots,q_k)$$
$$+\frac{p_1^{\alpha}s^{1-\alpha}}{2^{1-\alpha}-1}(s_q^{\alpha-1}-1).$$

Hence A^{α}_{n} is k-generalized recursive with

$$g_1(s_p,s,\alpha)=(s_p/s)^{\alpha-1},$$
$$f_k(s_q,s,\alpha,p_1)=p_1^{\alpha}(s_q/s)^{\alpha-1},$$

and

$$g_2(s_q,s,\alpha,p_1)=p_1^{\alpha}s^{1-\alpha}(s_q^{\alpha-1}-1)/(2^{1-\alpha}-1).$$
$$=p_1^{\alpha}((s_q/s)^{\alpha-1}-1)/(2^{1-\alpha}-1).$$

Definition 8:

If there is a function $\psi_{mn}: R^2\times[0,1]\to R$ such that

$$I^{\alpha}_{m+n}[(1-q)p_1,\ldots,(1-q)p_m;q\,q_1,\ldots,q\,q_n]$$
$$=\psi_{mn}[I^{\alpha}_{m}(p_1,\ldots,p_m),\ I^{\alpha}_{n}(q_1,\ldots,q_n),\ q,\alpha,s_p,s_q,s]$$
$$\forall(p_1,\ldots,p_m)\in\Gamma_m,\ (q_1,\ldots,q_n)\in\Gamma_n,\ q\in[0,1].$$

then I^{α} is said to satisfy the (m,n)-compositivity property of degree α. It is compositive if it satisfies the (m,n)-compositivity property for all m and n.

Theorem 9: A^{α}_{n} is compositive of degree α.

Proof:

$$A^{\alpha}_{m+n}[(1-q)p_1,\ldots,(1-q)p_m;\ q\,q_1,\ldots,q\,q_n]$$
$$=\frac{1}{2^{1-\alpha}-1}\left[\sum_{i=1}^{m}\left\{(1-q)\frac{p_i}{s}\right\}^{\alpha-1}(1-q)p_i\right.$$
$$\left.+\sum_{j=1}^{n}\left(\frac{q\,q_j}{s}\right)^{\alpha-1}q\,q_j-1\right]$$
$$=\frac{1}{2^{1-\alpha}-1}\left[\frac{(1-q)^{\alpha}}{(s/s_p)^{\alpha-1}}\left\{\sum_{i=1}^{m}\left(\frac{p_i}{s_p}\right)^{\alpha-1}p_i-1+1\right\}\right.$$
$$\left.+\frac{q^{\alpha}}{(s/s_q)^{\alpha-1}}\left\{\sum_{j=1}^{n}\left(\frac{q_j}{s_q}\right)^{\alpha-1}q_j-1+1\right\}-1\right]$$
$$=\left(\frac{s_p}{s}\right)^{\alpha-1}(1-q)^{\alpha}\ A^{\alpha}_{m}(p_1,\ldots,p_m)+\frac{(1-q)^{\alpha}}{2^{1-\alpha}-1}\left(\frac{s_p}{s}\right)^{\alpha-1}$$
$$+\left(\frac{s_q}{s}\right)^{\alpha-1}q^{\alpha}\ A^{\alpha}_{n}(q_1,\ldots,q_n)$$
$$+\frac{q^{\alpha}}{2^{1-\alpha}-1}\left(\frac{s_q}{s}\right)^{\alpha-1}-\frac{1}{2^{1-\alpha}-1}$$
$$=\left(\frac{s_p}{s}\right)^{\alpha-1}(1-q)^{\alpha}\ A^{\alpha}_{m}(p_1,\ldots,p_m)$$
$$+\left(\frac{s_q}{s}\right)^{\alpha-1}q^{\alpha}\ A^{\alpha}_{n}(q_1,\ldots,q_n)$$
$$+\frac{1}{2^{1-\alpha}-1}\left[\left(\frac{s_p}{s}\right)^{\alpha-1}(1-q)^{\alpha}+\left(\frac{s_p}{s}\right)^{\alpha-1}q^{\alpha}-1\right]$$
$$=\psi_{mn}[A^{\alpha}_{m}(p_1,\ldots,p_m),\ A^{\alpha}_{n}(q_1,\ldots,q_n);\ q,\alpha,s_p,s_q,s].$$

Therefore A^{α}_{n} is compositive of degree α.

4. CONCLUSIONS

It is clear from Section 2 that the use of the sup-method is more appropriate than the non-standard analysis method to generalize the definitions of entropies from the discrete to the continuous case.

In Section 3 it was shown that the new entropy has almost the same properties of the entropies given in the literature. It is interesting to note that the new entropy is not normalized and it is not maximal. More specifically $A_n^\alpha(1/n,\ldots,1/n) = 0$ for all $n \geq 2$. This property is usually used in Bayes methods to define what is called non-informative distribution, *i.e.* the uniform distribution is non-informative.

Awad [12] gave eight critical comments on the Shannon entropy. These comments motivated the definition of his entropy (2) which does not suffer from the drawbacks of the Shannon entropy. Some of these comments may be used to motivate the definition of the entropies (5) and (6). Since the arguments are the same as those given by Awad [12] we have not mentioned them here.

Awad [12] applied the entropies (1) and (2) to evaluate the information stability coefficient when the model is Bernoulli, uniform or normal. He noted that entropy (2) gives more meaningful results on this problem than those given by using entropy (1).

The results of this paper were applied by Abu-Taleb [15] to define several normed information rates as informational correlation and association measures. Some of these measures depend on the H-entropies (1), (3), and (4). Others depend on the A-entropies (2), (5), and (6). Comparing the behavior of these normed information rates, Abu-Taleb concluded that the rates derived from A-entropies are more meaningful than the rates derived from H-entropies.

Alawneh [16] also applied the results of this paper to find a truncation point t_0 such that the relative loss of information in using an exponential model truncated at t_0 instead of an exponential model is less than a given constant ε. Using the six entropies (1)–(6), he concluded that the A-entropies give more meaningful results than the H-entropies.

These three applications support the use of the A-entropies instead of the H-entropies.

ACKNOWLEDGEMENT

The authors would like to thank the referees for their useful comments, which improved the presentation of the paper. Thanks are also due to one of the referees who pointed out a mistake in the proof of the monotonicity property.

REFERENCES

[1] K. Ferentinos and T. Papaioannou, "New Parametric Measures of Information", *Information and Control,* **51** (1981), p. 193.

[2] D. Basu, "Statistical Information and Likelihood", *Sankhya,* **37A1** (1975), p. 1.

[3] E. E. Shannon, "A Mathematical Theory of Communication", *Bulletin of Systems Technology Journal,* **27** (1948), p. 379 and p. 623.

[4] I. Csiszar, "Information Measures: A Critical Survey", in *Transactions of the 7th Prague Conference on Information Theory; Statistical Decision Functions, Random Processes. Vol. B.*, Prague: Academia, 1978, p. 73.

[5] J. N. Kapur, "A Note on the Relationships Between Some Probability Distributions, Measures of Entropy, Measures of Directed Divergence and Distributions of Statistical Mechanics", *Indian Journal of Pure and Applied Mathematics,* **14(12)** (1983), p. 1433.

[6] B. Frote and C. T. Ng, "Derivation of a Class of Entropies Including Those of Degree β", *Information and Control,* **28** (1975), p. 335.

[7] J. Cerny and P. Brunovsky, "A Note on Information Without Probability", *Information and Control,* **25** (1974), p. 134.

[8] M. Behara and P. Nath, "Additive and Non-Additive Entropies of Finite Measurable Partitions", in *Lecture Notes in Mathematics, 296: Probability and Information Theory II,* 1973, p. 102.

[9] F. M. Ingels, *Information and Coding Theory.* Intext Educational Publishers, 1971.

[10] K. Ozeki, "A New Definition of the Entropy of General Probability Distributions Based on the Non-Standard Analysis", *Information and Control,* **74** (1980), p. 94.

[11] H. J. Keisler, *Foundations of Infinitesimal Calculus.* Boston: Prindle, Weber and Schmidt, 1976.

[12] A. Awad, "A Statistical Information Measure", submitted for publication.

[13] J. Aczel and Z. Davoczy, *On Measures of Information and Their Characterizations.* New York: Academic Press, 1975.

[14] Abu-Bakr El-Sayed, "The Independence Inequality and Its Application to Information Theory", *Information and Control,* **35** (1977), p. 229.

[15] A. Abu-Taleb, "Normed Information Rates and Their Applications", *Master's Thesis, Department of Statistics, Yarmouk University, Jordan, 1986.*

[16] B. Ebanks, "On a Generalized Recursivity and Entropies of Degree α", *Information and Control,* **32** (1976), p. 33.

[17] A. Alawneh, "Extension of Shannon's Entropies with Applications", *Master's Thesis, Department of Statistics, Yarmouk University, Jordan, 1986.*

[18] G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities.* Cambridge University Press, 1967.

Paper Received 8 October 1985; Revised 3 May 1986.