American. Jr. of Mathematics and Sciences
Vol. 1, No.1 ,(January 2012)

www.journalshub.com

NEW GENERALIZED PARAMETRIC MEASURES OF ENTROPY AND CROSS ENTROPY

Om Parkash
Department of Mathematics
Guru Nanak Dev University
Amritsar- 143005(India)
omparkash777@yahoo.co.in

Mukesh
Department of Mathematics
Guru Nanak Dev University
Amritsar-143005(India)
sarangal.mukesh@yahoo.co.in

ABSTRACT

The measure of entropy introduced by Shannon [12] is the key concept in the literature of information theory and has found tremendous applications in different disciplines of science and technology. The various researchers have generalized this entropy with different approaches. The object of the present manuscript is to develop a generalized measure of entropy by using the property of concavity. The proposed measure satisfies all essential and some desirable properties of the *original Shannon's [1*2] entropy. Moreover, we have developed a new generalized measure of cross entropy which corresponds to the newly introduced measure of entropy.

Keywords: Entropy, Concavity, Additivity, Cross entropy, Expansibility, Monotonicity.

INTRODUCTION

In communication theory, it was Shannon [12] who first introduced the concept of entropy and it was then realized that entropy is a property of any stochastic system and the concept is now widely prevalent in different disciplines. The tendency of the system to become more disordered over time is described by the second law of thermodynamics, which states that the entropy of the system cannot spontaneously decrease. Today, information theory is still principally concerned with communications systems, but there are widespread applications in Mathematical Sciences. Shannon [12] called the probabilistic uncertainty as entropy and developed his measure given by

$$H(P) = -\sum_{i=1}^{n} p_i \log p_i \tag{1.1}$$

Hu [7] proposed two new broad classes for measures of uncertainty as the survival exponential and the generalized survival exponential entropies and thus improving Shannon's [12] entropy. Honda and Okazaki [6] generalized Shannon's [12] entropy and proved that their entropy has applicability to the capacity on set systems. Immediately, after Shannon [12] gave his measure, Renyi [11] was the first to derive entropy of order α as follows:

$${}_{\alpha}H(P) = \frac{1}{1-\alpha} \log\left(\sum_{i=1}^{n} p_i^{\alpha} \Big/ \sum_{i=1}^{n} p_i \right), \alpha \neq 1, \alpha > 0 \tag{1.2}$$

The generalized measure of entropy (1.2) includes Shannon's [12] entropy as a limiting case as $\alpha \to 1$. Zyczkowski [14] explored the relationships between the Shannon's [12] and Renyi's [11] entropies of integer order. Havrada and Charvat [4] introduced non-additive entropy, given by:

$$H^{\alpha}(P) = \frac{\left[\sum_{i=1}^{n} p_i^{\alpha}\right] - 1}{2^{1-\alpha} - 1}, \alpha \neq 1, \alpha > 0 \tag{1.3}$$

Many other probabilistic measures of entropy have been discussed and derived by Brissaud [1], Chen [2], Garbaczewski [3], Herremoes [5], Lavenda [9], Nanda and Paul [10], Sharma and Taneja [13] etc. The applications of the results on probabilistic information measures obtained by various authors have been provided to different fields of Linguistics, Biological Sciences, Economics, Social Sciences and Engineering Sciences for developing new entropic models.

A measure $D(P;Q)$ of divergence or cross entropy or directed divergence is found to be very important in various disciplines of Mathematical and Engineering Sciences. This measure is probabilistic in nature and measures the distance of a probability distribution P from Q. The most important and useful measure of divergence is due to Kullback and Leibler [8] and is given by

$$D(P;Q)=\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \tag{1.4}$$

In this communication, we have introduced a new parametric generalized probabilistic measure of entropy and consequently, developed a new generalized measure of cross entropy.

2. A NEW GENERALIZED PARAMETRIC MEASURE OF ENTROPY

In this section, we have proposed a new generalized measure of entropy for a probability distribution $P=\left\{(p_1, p_2, ..., p_n), p_i \ge 0, \sum_{i=1}^{n} p_i = 1\right\}$ and studied its essential and desirable properties. This new entropy measure of order α is given by the following mathematical expression:

$$H_\alpha(P)=\frac{1}{\alpha}\sum_{i=1}^{n}\left(1-p_i^{\alpha p_i}\right), \quad \alpha \ne 0, \alpha > 0 \tag{2.1}$$

Under convention, we take $(0)^{\alpha.0}=1$. Obviously, we have $\lim_{\alpha\to 0} H_\alpha(P) = -\sum_{i=1}^{n} p_i \log p_i$

Thus, $H_\alpha(P)$ can be taken as a generalization of well known Shannon's [12] measure of entropy.

Next, to prove that the measure (2.1) is a valid measure of entropy, we study its essential properties as follows:

(i) Clearly $H_\alpha(P) \ge 0$

(ii) $H_\alpha(P)$ is permutationally symmetric as it does not change if $p_1, p_2, ..., p_n$ are re-ordered among themselves.

(iii) $H_\alpha(P)$ is a continuous function of p_i for all p_i's.

(iv) **Concavity:** To prove concavity property, we proceed as follows:

Let $\phi(p)=\frac{1}{\alpha}\left(1-p^{\alpha p}\right)$

Then $\phi''(p) = -p^{\alpha p}\left[\alpha(1+\log p)^2+\frac{1}{p}\right] < 0$ for all $\alpha > 0$.

Thus, $\phi(p)$ is a concave function of p. Since the sum of concave functions is also a concave function, $H_\alpha(P)$ is a concave function of $p_1, p_2, ..., p_n$. Moreover, with the help of numerical data shown in the following Table-2.1, we have presented $H_\alpha(P)$ as shown in Fig.-2.1.

Table-2.1: $H_\alpha(P)$ against p for $n=2$ and $\alpha=2$

p	$H_\alpha(P)$
0	0
0.1	0.271
0.2	0.388
0.3	0.454
0.4	0.489
0.5	0.5
0.6	0.489
0.7	0.454
0.8	0.388
0.9	0.271

1	0

The Fig.-2.1 clearly shows that the measure (2.1) is a concave function.

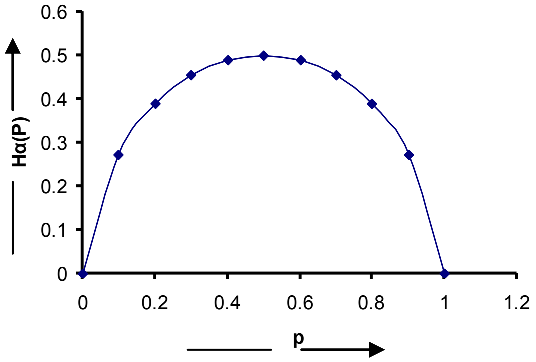

Fig.-2.1

Hence, under the above conditions, the function $H_\alpha(P)$ is a correct measure of entropy.

Next, we study some most desirable properties of $H_\alpha(P)$.

(i) Expansibility: $H_\alpha(p_1, p_2, ..., p_n, 0) = H_\alpha(p_1, p_2, ..., p_n)$

That is, the entropy does not change by the inclusion of an impossible event.

(ii) For degenerate distributions, $H_\alpha(P) = 0$.

This indicates that for certain outcomes, the uncertainty should be zero.

(iii) Maximization: We use Lagrange's method to maximize the entropy measure (2.1) subject to the natural constraint $\sum_{i=1}^{n} p_i = 1$. In this case, the corresponding Lagrangian is

$$L \equiv \frac{1}{\alpha}\sum_{i=1}^{n}\left(1 - p_i^{\alpha p_i}\right) - \lambda\left(\sum_{i=1}^{n} p_i - 1\right) \quad (2.2)$$

Differentiating equation (2.2) with respect to $p_1, p_2, ..., p_n$ and equating the derivatives to zero, we get $p_1 = p_2 = ... = p_n$. This further gives $p_i = \frac{1}{n}\ \forall i$. Thus, we observe that the maximum value of $H_\alpha(P)$ arises for the uniform distribution and this result is most desirable.

(iv) The maximum value $\psi(n)$ of the entropy is given by

$$\psi(n) = \frac{n}{\alpha}\left[1 - \left(\frac{1}{n}\right)^{\frac{\alpha}{n}}\right]$$

or $f(x) = \frac{1}{\alpha}\frac{1}{x}\left(1 - e^{\alpha x \log x}\right)$ where $x = \frac{1}{n},\ x \in (0,1]$

Thus $f'(x) = -\frac{1}{\alpha x}\left[\alpha x^{\alpha x}(1 + \log x) + \frac{1}{x}\left(1 - x^{\alpha x}\right)\right] < 0$

which shows that $\psi(n)$ is an increasing function of n, which is again a desirable result as the maximum value of an entropy should always increase.

(v) Monotonicity: From equation (2.1), we have

$\frac{d}{d\alpha}\left(H_\alpha(P)\right) = -\frac{1}{\alpha^2}\sum_{i=1}^{n}\left[p_i^{\alpha p_i}\left(\log p_i^{\alpha p_i} - 1\right) + 1\right] \le 0$ which is possible iff $p_i^{\alpha p_i}\left(\log p_i^{\alpha p_i}-1\right)+1\geq 0 \quad \forall\, i=1,2,\ldots,n$

or $\log p_i^{-\alpha p_i}\leq p_i^{-\alpha p_i}-1$ which is always true.

Hence $H_\alpha(P)$ is a monotonic decreasing function of α. For the probability distribution $(p,1-p)$,

$$[H_\alpha(P)]_{max}=\frac{1}{\alpha}\left(1-\left(\frac{1}{2}\right)^{\alpha/2}\right)+\frac{1}{\alpha}\left(1-\left(\frac{1}{2}\right)^{\alpha/2}\right)=\frac{2}{\alpha}\left(1-\left(\frac{1}{2}\right)^{\alpha/2}\right)$$

The various values of the maximum entropy have been displayed in the following Table-2.2:

Table-2.2: $\max H_\alpha(P)$ against α

α	0	0.5	1	1.5	2	5	10	20	∞
$[H_\alpha(P)]_{max}$	∞	0.6364	0.5857	0.5405	0.5	0.3292	0.1937	0.0999	0

The graph of $H_\alpha(p,1-p)$ for different values of α is shown in Fig.-2.2.

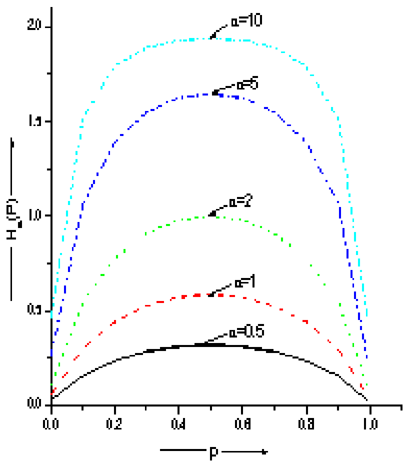

Fig.-2.2

3. A NEW PARAMETRIC MEASURE OF CROSS ENTROPY

Consider the following set of all complete finite discrete probability distributions:

$$\Omega_n=\left\{P=(p_1,p_2,\ldots,p_n):\ p_i>0,\sum_{i=1}^{n}p_i=1\right\},n\geq 2, \tag{3.1}$$

Let $P,Q\in\Omega_n$ be any two probability distributions. Then, the measure of cross entropy corresponding to the new entropy measure introduced in the above section is given by

$$D_\alpha(P;Q)=-\frac{1}{\alpha}\sum_{i=1}^{n}q_i\left(1-\left(\frac{p_i}{q_i}\right)^{\frac{\alpha p_i}{q_i}}\right),\alpha\neq 0,\alpha>0. \tag{3.2}$$

Note: We have

$$\lim_{\alpha\to 0}D_\alpha(P;Q)=\lim_{\alpha\to 0}\left[-\frac{1}{\alpha}\sum_{i=1}^{n}q_i\left(1-\left(\frac{p_i}{q_i}\right)^{\frac{\alpha p_i}{q_i}}\right)\right]=\sum_{i=1}^{n}p_i\log\frac{p_i}{q_i}.$$

Thus, $D_\alpha(P;Q)$ is a generalization of Kullback-Leibler's [8] measure of cross-entropy.

Some of the important properties of this cross-entropy measure are:

(1) $D_\alpha(P;Q)$ is a continuous function of $p_1, p_2, \ldots, p_n$ and of $q_1, q_2, \ldots, q_n$.

(2) $D_\alpha(P;Q) \geq 0$ and vanishes if and only if $P = Q$.

(3) We can deduce from condition (2) that the minimum value of $D_\alpha(P;Q)$ is zero.

(4) We shall now prove that $D_\alpha(P;Q)$ is a convex function of both P and Q. This result is important in establishing the property of global minimum.

Let $D_\alpha(P;Q) = f(p_1, p_2, \ldots, p_n; q_1, q_2, \ldots, q_n)$

$$= -\frac{1}{\alpha}\sum_{i=1}^{n} q_i\left(1-\left(\frac{p_i}{q_i}\right)^{\frac{\alpha p_i}{q_i}}\right),$$

Then we have $\dfrac{\partial f}{\partial p_i} = \left(\dfrac{p_i}{q_i}\right)^{\frac{\alpha p_i}{q_i}}\left(1+\log\dfrac{p_i}{q_i}\right)$.

Also, $\dfrac{\partial^2 f}{\partial p_i^2} = \left(\dfrac{p_i}{q_i}\right)^{\frac{\alpha p_i}{q_i}}\left[\dfrac{\alpha}{q_i}\left(1+\log\dfrac{p_i}{q_i}\right)^2 + \dfrac{1}{p_i}\right] > 0$ for $\alpha > 0$.

And $\dfrac{\partial^2 f}{\partial p_i \partial p_j} = 0$ for $i, j = 1, 2, \ldots, n; i \neq j$.

Similarly, $\dfrac{\partial f}{\partial q_i} = -\dfrac{1}{\alpha}\left[1-\left(\dfrac{p_i}{q_i}\right)^{\frac{\alpha p_i}{q_i}} + \dfrac{\alpha p_i}{q_i}\left(\dfrac{p_i}{q_i}\right)^{\frac{\alpha p_i}{q_i}}\left(1+\log\dfrac{p_i}{q_i}\right)\right]$,

$$\frac{\partial^2 f}{\partial q_i^2} = \frac{p_i}{q_i^2}\left(\frac{p_i}{q_i}\right)^{\frac{\alpha p_i}{q_i}}\left[\frac{\alpha p_i}{q_i}\left(1+\log\frac{p_i}{q_i}\right)^2 + 1\right] > 0 \text{ for } \alpha > 0,$$

and $\dfrac{\partial^2 f}{\partial q_i \partial q_j} = 0$ for $i, j = 1, 2, \ldots, n; i \neq j$.

The Hessian matrix of second order partial derivatives of f with respect to $p_1, p_2, \ldots, p_n$ is

$$\begin{bmatrix} \left(\frac{p_1}{q_1}\right)^{\frac{\alpha p_1}{q_1}}\left[\frac{\alpha}{q_1}\left(1+\log\frac{p_1}{q_1}\right)^2+\frac{1}{p_1}\right] & 0 & \cdots & 0 \\ 0 & \left(\frac{p_2}{q_2}\right)^{\frac{\alpha p_2}{q_2}}\left[\frac{\alpha}{q_2}\left(1+\log\frac{p_2}{q_2}\right)^2+\frac{1}{p_2}\right] & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \left(\frac{p_n}{q_n}\right)^{\frac{\alpha p_n}{q_n}}\left[\frac{\alpha}{q_n}\left(1+\log\frac{p_n}{q_n}\right)^2+\frac{1}{p_n}\right] \end{bmatrix}$$

which is positive definite. A similar result is also true with respect to $q_1, q_2, \ldots, q_n$. Thus, $D_\alpha(P;Q)$ is a convex function of both $p_1, p_2, \ldots, p_n$ and $q_1, q_2, \ldots, q_n$. Moreover, with the help of numerical data, we have presented $D_\alpha(P;Q)$ as shown in the following Fig.-3.1.

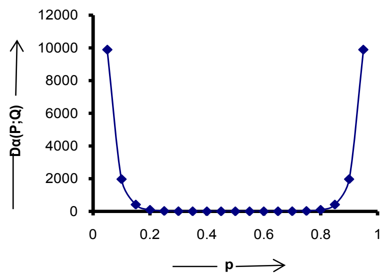

Fig.-3.1

Under the above conditions, the function $D_\alpha(P;Q)$ is a correct measure of cross entropy.

ACKNOWLEDGEMENTS: The authors are thankful to the University Grants Commission, New Delhi, for providing financial assistance for the preparation of the manuscript.

REFERENCES

[1] Brissaud, J. B. (2005). The meaning of entropy. Entropy **7**: 68-96.

[2] Chen, Y. (2006). Properties of quasi-entropy and their applications. Journal of Southeast University Natural Sciences **36**: 222-225.

[3] Garbaczewski, P. (2006). Differential entropy and dynamics of uncertainty. Journal of Statistical Physics **123**: 315-355.

[4] Havrada, J. H. and Charvat, F. (1967). Quantification methods of classification process: Concept of structural α-entropy. Kybernetika **3**: 30-35.

[5] Herremoes, P. (2006). Interpretations of Renyi entropies and divergences. Journal of Physics A **365**: 57-62.

[6] Honda, A. and Okazaki, Y. (2007). Generalization of Shannon entropy of capacities on set systems. Applied Functional Analysis: 89-98.

[7] Hu, H. (2007). Survival exponential entropies and their analytic properties. Journal of Huazhong Normal University, Natural Sciences **41**: 493-496.

[8] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. Annals of Mathematical Statistics **22**: 79-86.

[9] Lavenda, B. H. (2005). Mean Entropies. Open Systems & Information Dynamics **12**: 289-302.

[10] Nanda, A. K. and Paul, P. (2006). Some results on generalized residual entropy. Information Sciences **176**: 27-47.

[11] Renyi, A. (1961). On measures of entropy and information. Proceedings 4th Berkeley Symposium on Mathematical Statistics and Probability **1**: 547-561.

[12] Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal **27**: 379-423, 623-659.

[13] Sharma, B. D. and Taneja, I. J. (1975). Entropies of type (α,β) and other generalized measures of information theory. Metrica **22**: 202-215.

[14] Zyczkowski, K. (2003). Renyi extrapolation of Shannon entropy. Open Systems and Information Dynamics **10**: 297-310.