International Journal of Electronic and Electrical Engineering.
ISSN 0974-2174 Volume 12, Number 1 (2019), pp. 1-9

http://www.irphouse.com

New Generalized Parametric M-Measure of Information and Its Application in Coding Theory

A. Chandravanshi

Dept. of Mathematics, Govt. J.Y. Chhattisgarh College, Raipur (C.G.), India.

Abstract

New generalized parametric M- measures of information are defined and studied some of their properties. Application of these measures is provided to the discipline of coding theory.

Keyword: Entropy, directed divergence, mean codeword length, uniquely decipherable code.

INTRODUCTION

Shannon led to the discovery of information theory. As there is similarity between Shannon's uncertainty measure and entropy function, this subject is closely related to thermodynamics and Physics. It then realization that entropy is a property of any stochastic system and this concept is widely used now. The system disorders over time which can be described by the second law of thermodynamics, which states that the entropy of the system cannot decrease spontaneously. In the present day, information theory is chiefly concerned with communication system but it has application in statistics, information processing and computing.

Shannon (1948)[13] entropy, also known as measure of uncertainty for a probability distribution $P=(p_1, p_2, \ldots, p_n)$ is given by

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

with the convention that $0\log 0 = 0$. It is to be noted that the base of logarithm is assumed to be 2, unless until specified.

Besides entropy, directed divergence is another basic and fundamental concept usually applied in information theory. The most important and desirable measure of directed divergence associated with the probability distributions $P=(p_1, p_2,..., p_n)$ and $Q=(q_1, q_2,..., q_n)$ is due to Kullback and Leibler(1951) [9] and is given by

$$D(P:Q)=\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \qquad (2)$$

The above information measures can be used in various field like genetics, finance, economics, political science, biology, analysis of contingency of tables, statistics, signal processing and pattern recognition. It should be noted that the above non-parametric measures are insufficient while use in a number of disciplines. For example, Shannon's measure of entropy always leads to exponential families of distributions but in practice, many families and distributions are not exponential in nature. Therefore restriction of Shannon's entropy means restriction of only exponential family thus the system is leaved to be least flexible. The use of generalized parametric measures of information can be an alternative. Here the word 'generalized' means to be more flexible and not superior or more useful.

Csiszar(1977)[2] investigated Shannon's measure and also made a summary of importance of this measure and it generalization with their scope in the coding theory. Renyi(1961)[12], Havrda and Charvat(1967)[5] etc. investigate and studied some other parametric generalization of Shannon's entropy. Garrido(2011)[4] studies a variety of information measure and their mutual relationships. Dahl and Osteras(2010)[3] used Shannon's entropy as a measure of information content in survey data and information efficiency was also defined by them as the empirical entropy divided by the maximum entropy that can be attained. Generalized entropies were introduced by Mathai and Haubold(2007)[10]. They studied their properties and examined those situations where generalized entropy of order α and type β can be used in variety of mathematical models.

Present paper objective is introduction of new measures of information and to extend their use in coding theory. The organization of this paper is as follows: In section 2, new measures of entropy have been described along with necessary and desirable property. Section 3 deals with the use of proposed measures in the discipline of coding theory.

2. NEW GENERALIZED PARAMETRIC MEASURE OF ENTROPY

In this section, we propose a new generalized measure of entropy to be called parametric M-entropy for a probability distribution $P=\left\{(p_1, p_2,..., p_n), p_i \geq 0, \sum_{i-1}^{n} p_i = 1\right\}$ and study its essential and desirable properties.

This new generalized measure of entropy is given by the following mathematical expression:

$$M_\alpha(P)=\frac{\alpha}{\alpha-1}\left[1-\prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i\frac{(1-\alpha)}{\alpha}}\right],\quad \alpha>0,\ \alpha\neq 1 \tag{3}$$

with the convention $0^0=1$.

We observe that for $\alpha\to 1$, measure (3) reduce to Shannon's (1948) [13] entropy as shown below:

$$\lim_{\alpha\to 1} M_\alpha(P) = \lim_{\alpha\to 1}\frac{\alpha}{\alpha-1}\left[1-\prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i\frac{(1-\alpha)}{\alpha}}\right]$$

$$=-\sum_{i=1}^{n} p_i \log p_i$$

Hence, the measure (3) is generalization of Shannon's measure.

Essential properties of measure:

1. $M_\alpha(P)$ is non-negative, that is, $M_\alpha(P)\geq 0$.

Proof: Case-I: When $0<\alpha<1$

$$\frac{\alpha}{\alpha-1}\left[1-\prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i\frac{(1-\alpha)}{\alpha}}\right]\geq 0$$

that is, iff $$\log\left[\prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i\frac{(1-\alpha)}{\alpha}}\right]\geq 0$$

that is, iff $-\sum_{i=1}^{n} p_i \log p_i \geq 0$ which is true.

Case-II: When $\alpha>1$

$$\frac{\alpha}{\alpha-1}\left[1-\prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i\frac{(1-\alpha)}{\alpha}}\right]\geq 0$$

that is, iff $$\log\left[\prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i\frac{(1-\alpha)}{\alpha}}\right]\leq 0$$

that is, iff $-\sum_{i=1}^{n} p_i \log p_i \geq 0$ which is true.

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

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

4. Concavity: $M_\alpha(P)$ is a concave function of p_i for all p_i's.

To prove concavity property, we proceed as follows: We have

$$\frac{\partial^2 M_\alpha(P)}{\partial p_i^2} = \left[\left(\frac{1-\alpha}{\alpha}\right)(1+\log p_i)^2 - \frac{1}{p_i}\right] \prod_{i-1}^{n} \left(\frac{1}{p_i}\right)^{p_i \frac{(1-\alpha)}{\alpha}} \tag{4}$$

Now, we know that for all $i = 1,2,...,n$, we have $0 \leq p_i \leq 1$,

that is, $(1+\log p_i)^2 - \frac{1}{p_i} \leq 0$

and for $\alpha > 0$, we have

$$\left(\frac{1-\alpha}{\alpha}\right)(1+\log p_i)^2 - \frac{1}{p_i} \leq 0, \quad i = 1,2,...,n, \tag{5}$$

that is $\frac{\partial^2 M_\alpha(P)}{\partial p_i^2} \leq 0$.

So $M_\alpha(P)$ is a concave function of p_i. Similarly, it can be proved that $M_\alpha(P)$ is a concave function of all p_i's. Hence, under the above condition $M_\alpha(P)$ is a correct measure of entropy.

Desirable Properties of measure:

1. Expansibility:

   We have $M_\alpha(p_1, p_2, ..., p_n, 0) = M_\alpha(p_1, p_2, ..., p_n)$. That is, the entropy does not change by the inclusion of an impossible event.

2. For n degenerate distribution, we have $M_\alpha(P) = 0$. This indicates that for certain outcomes, the uncertainty should be zero.

3. Maximization of entropy: Maximize the entropy function $M_\alpha(P)$ subject to the natural constraint $\sum_{i=1}^{n} p_i = 1$, by using Lagrange's method, we have

$$J = \frac{\alpha}{\alpha - 1}\left[1 - \prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i \frac{(1-\alpha)}{\alpha}}\right] - \lambda\left[\sum_{i=1}^{n} p_i - 1\right] \tag{6}$$

Differentiating equation (6) 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 $M_\alpha(P)$ arises for the uniform distribution.

4. Maximum value: Maximum value of the entropy is given by $M_\alpha\left(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n}\right) = \frac{\alpha}{\alpha - 1}\left[1 - n^{\frac{(1-\alpha)}{\alpha}}\right]$ which is an increasing function of n and is again a desirable result as the maximum value of entropy should always increase.

5. Non-additive:

   The entropy of the joint probability distribution denoted by

$$\begin{aligned} M_\alpha(P * Q) &= \frac{\alpha}{\alpha - 1}\left[1 - \prod_{i=1}^{n}\prod_{j=1}^{n}(p_i q_j)^{-p_i q_j \frac{(1-\alpha)}{\alpha}}\right] \\ &= \frac{\alpha}{\alpha - 1}\left[1 - \prod_{i=1}^{n} p_i^{-p\left(\frac{1-\alpha}{\alpha}\right)_i} \prod_{j=1}^{n} q_j^{-q_j\left(\frac{1-\alpha}{\alpha}\right)}\right] \end{aligned} \tag{7}$$

Also, we have

$$M_\alpha(P) + M_\alpha(Q) + \frac{\alpha}{\alpha - 1} M_\alpha(P) M_\alpha(Q) = \frac{\alpha}{\alpha - 1}\left[1 - \prod_{i=1}^{n} p_i^{-p_i\left(\frac{1-\alpha}{\alpha}\right)} \prod_{j=1}^{n} q_j^{-q_j\left(\frac{1-\alpha}{\alpha}\right)}\right] \tag{8}$$

From equation (7) and (8), we have

$$M_\alpha(P) + M_\alpha(Q) + \frac{\alpha}{\alpha - 1} M_\alpha(P) M_\alpha(Q) = M_\alpha(P * Q)$$

Thus, we claim that the new measure of entropy $M_\alpha(P)$ introduced in (3) satisfies all the essential and as well as desirable properties of being entropy measure, it is a valid measure of entropy.

3. NEW GENERALIZED MEASURE OF DIRECTED DIVERGENCE

We propose a new generalized parametric measure of directed divergence of probability distribution $P=(p_1, p_2,...., p_n)$ from another probability distribution $Q=(q_1, q_2,..., q_n)$, given by

$$D_\alpha(P:Q)=\frac{\alpha}{1-\alpha}\left(1-\prod_{i=1}^{n}\left(\frac{p_i}{q_i}\right)^{-p_i\frac{(1-\alpha)}{\alpha}}\right) \tag{9}$$

It is observe that for $\alpha \to 1$ in (9), we get Kullback - Leibler's(1956)[9] measure of directed divergence.

Thus, we claim that the measure (9) is a correct measure of directed divergence as it satisfies all the requisite properties.

4. SOME NEW SOURCE CODING THEOREM

4.1 Source coding theorem with generalized measure of entropy

Let $x_1, x_2,.......x_n$ be n inputs which have to be encoded in terms of an alphabet of size $D(D \geq 2)$. If $l_1, l_2......., l_n$ be the n codeword lengths and $p_1, p_2,...., p_n$ be the probabilities, then the arithmetic mean L of the codeword length is

$$L=\sum_{i=1}^{n} p_i l_i \tag{10}$$

Shannon [13] minimize (10) subject to Kraft's [8] inequality

$$\sum_{i=1}^{n} D^{-l_i} \leq 1 \tag{11}$$

for uniquely decipherable codes and proved that the minimum value of L lies between $H(P)$ and $H(P)+1$, where the logarithm in the definition of Shannon entropy is taken in base D. This result indicates that the Shannon entropy $H(P)$ is the fundamental limit on the minimum average length of any code constructed for the source. The lengths of the individual code words, are given by

$$l_i = -\log_D p_i \tag{12}$$

Also, we have the following relation between the Shannon's entropy and the generalized entropy (3)

$$M_\alpha(P)=\log\left(D^{H(P)}\right) \tag{13}$$

where $\log_\alpha(.)$ is the $\alpha-$ deformed logarithm defined as $\log_\alpha x = \frac{\alpha}{\alpha-1}\left(1-x^{\frac{(1-\alpha)}{\alpha}}\right)$.

Now, we consider the following cases:

Case-I When $0 < \alpha < 1$, we have

$$L \geq H(P)$$

$$\Rightarrow \quad \frac{\alpha}{\alpha - 1}\left[1 - D^{L\left(\frac{1-\alpha}{\alpha}\right)}\right] \geq \frac{\alpha}{\alpha - 1}\left[1 - D^{H(P)\left(\frac{1-\alpha}{\alpha}\right)}\right]$$

$$\Rightarrow \quad L_\alpha = \log_\alpha\left(D^L\right) \geq \log_\alpha\left(D^{H(P)}\right) = M_\alpha(P) \tag{14}$$

Case-II When $\alpha > 1$, we have

$$L \geq H(P)$$

$$\Rightarrow \quad L_\alpha = \log_\alpha\left(D^L\right) \geq \log_\alpha\left(D^{H(P)}\right) = M_\alpha(P) \tag{15}$$

Here comes out the new generalized length L_α from (14) and (15) to which the generalized entropy $M_\alpha(P)$ forms a lower bound. It is a monotonic increasing function of mean codeword length L and it reduces to L when $\alpha \to 1$. The optimal codeword length are given by $l_i = -\log_D p_i$, which is similar to as in the case of Shannon's source coding theorem. L_α is not an average of the type $\phi^{-1}\left(\sum_{i=1}^{n} p_i \phi(l_i)\right)$ as introduced by Kolmogorov(1930)[7] and Nagumo(1930) [11] but is a simple expression of the $\alpha-$ deformed logarithm.

Note: When $l_1 = l_2 = ... = l_n = l$, then $L_\alpha \neq l$. Instead, it reduces to $\frac{\alpha}{\alpha - 1}\left[1 - D^{l\left(\frac{1-\alpha}{\alpha}\right)}\right]$ which further reduces to l when $\alpha \to 1$. The above result (14) and (15) can also be stated in the form of following theorem:

Theorem: If $l_1, l_2, ..., l_n$ denote the lengths of the uniquely decipherable code for the random variable X, then $L_\alpha \geq M_\alpha(P)$ with equality if and only if $l_i = -\log_D p_i$.

Proof: We have to minimize the following length

$$L_\alpha = \frac{\alpha}{\alpha - 1}\left[1 - D^{\frac{(1-\alpha)}{\alpha}\sum_{i=1}^{n} p_i l_i}\right] \tag{16}$$

subject to the Kraft's (1949)[8] inequality

$$\sum_{i=1}^{n} D^{-l_i} \leq 1$$

The corresponding Lagrangian is given by

$$J_1 = \frac{\alpha}{\alpha - 1}\left[1 - D^{\frac{(1-\alpha)}{\alpha}\sum_{i=1}^{n} p_i l_i}\right] + \lambda\left[\sum_{i=1}^{n} D^{-l_i} - 1\right] \tag{17}$$

Differentiating (17) with respect to l_i , $\forall i = 1,2,...,n$ and equating to zero, we get

$$p_i = \lambda D^{-l_i} \tag{18}$$

Using $\sum_{i=1}^{n} D^{-l_i} = 1$ and $\sum_{i=1}^{n} p_i = 1$, equation (18) gives $\lambda = 1$ and hence $p_i = D^{-l_i}$, that is $l_i = -\log_D p_i$. Substituting l_i in (16), we get the minimum value of L_α as

$$[L_\alpha]_{\min} = \frac{\alpha}{\alpha - 1}\left[1 - \prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i\frac{(1-\alpha)}{\alpha}}\right] = M_\alpha(P)$$

4.2 Source coding theorem via new measure of directed divergence

We know that the measure of directed divergence as given by (9) is non-negative, that is,

$$D_\alpha(P:Q) = \frac{\alpha}{1-\alpha}\left(1 - \prod_{i=1}^{n}\left(\frac{p_i}{q_i}\right)^{-p_i\frac{(1-\alpha)}{\alpha}}\right) \geq 0,\ 0 < \alpha < 1, \alpha \neq 1 \tag{19}$$

Substituting $q_i = \frac{D^{-l_i}}{\sum_{i=1}^{n} D^{-l_i}}$, $i = 1,2,...,n$ in (9), we get

$$\Rightarrow \qquad \prod_{i=1}^{n}\left(\frac{p_i}{D^{-l_i}}\sum_{i=1}^{n} D^{-l_i}\right)^{-p_i\left(\frac{1-\alpha}{\alpha}\right)} \leq 1$$

$$\Rightarrow \qquad -\sum_{i=1}^{n} p_i \log_D p_i - \log_D \sum_{i=1}^{n} D^{-l_i} \leq \sum_{i=1}^{n} p_i l_i = L$$

Now, since $\frac{1}{D} \leq \sum_{i=1}^{n} D^{-l_i} \leq 1$, therefore the lower bound for L lies between $H(P)$ and $H(P)+1$. Hence, we obtain the following result $H(P) \leq L \leq H(P)+1$ which is the Shannon's source coding theorem for uniquely decipherable codes.

REFERENCES

[1] Campbell L.L. "A coding theorem and Renyi's entropy". *Information and control*, Vol. 8, 1965, 423 - 429.

[2] Csiszar I. "Information measure: a critical survey " *In: Transaction of the 7th conference on Information Theory, Statistical Decision Functions, Random Processes. Eds. Kozesnik, J. Academia, Prague.* 1977, B: 73-86.

[3] Dahl F.A. and Osteras N. "Quantifying information content in survey data by entropy " , *Entropy*, 2010, 12: 161-163.

[4] Garrido A. "Classifying entropy measures " , *Symmetry,* 2011, 3: 487-502.

[5] Havrda J.H. and Charvat F. C. "Quantification methods of classificatory Processes. The Concept of structural α entropy". *Kybinetika*, 3, 1967, 30-35.

[6] Kerridge D.F. "Inaccuracy and Inference". *J. Royal Statist. Soc. Ser*, B23, 1961, P. 184-194.

[7] Kolmogorov A. "Sur la notion de la moyenne " *Atti Acad. Naz. Lincei*, 1930, 12: 388-391.

[8] Kraft L.G. *A device for quantizing grouping and coding amplitude modulated pulses.*MIT: MS. Thesis, electrical engineering Deptt.1949.

[9] Kullback S. and Leibler R.A. "On information and sufficiency". *Ann. Maths. Stat.* 22, 1951, pp. 76 - 86.

[10] Mathai A.M. and Haubold H.J. "On generalized entropy measures and pathways " , *Physica A.*, 2007, 385: 493-500.

[11] Nagumo M. "Uber eine Klass der Mittlwerte " , *Japan J. Math.*, 1930, 7: 71-79.

[12] Renyi A. "On measure of entropy and information". *Proc. 4th Berk symp. Statistics Probl.*, Vol.1, 1961, p. 541-561.

[13] Shannon C.E. "The mathematical theory of communication". *Bell System Tech.*, Vol. 27, 1948, 379-425, 623-656.