American Journal of Mathematics and Sciences
Vol. 4, No.1- 2, (January-December, 2015)

ISSN No: 2250-3102
www.journalshub.com

ON INFORMATION RADIUS BASED UPON ENTROPY MEASURE

Jatinder Kumar
Associate Professor
Guru Gibind Singh Khalsa College,
Sarhali, Tarn Taran (INDIA)
jk.anita@gmail.com

ABSTRACT

Measures of entropy and divergence play a significant role towards application areas and also for the manipulation of new results. One such a manipulation has been made resulting in the development of a measure usually known as R-divergence. The present manuscript deals with the introduction of R-divergence and the study of its essential properties so as to prove its validity.

Keywords: Uncertainty, Probability distribution, Entropy, Divergence, R-divergence, Shannon-Gibbs inequality.

INTRODUCTION

It is well known in the literature of information theory that uncertainty associated with probability of outcomes is called entropy and it was Shannon [11] who for the first time introduced the concept of information theoretic entropy by associating uncertainty with every probability distribution $P=\left(p_1,\, p_2,\, ...,\, p_n\right)$ and found that the only function that can measure uncertainty, is given by

$$H(P) = -\sum_{i=1}^{n} p_i \ \text{In}\ p_i \qquad (1.1)$$

The probabilistic measure of entropy (1.1) possesses a number of interesting properties. Immediately after Shannon gave his measure, research workers in many fields saw the potential of the application of this expression. Psychologists used this for measuring (i) subjects response to stimuli (ii) information in absolute judgments (iii) information in visual perception (iv) uncertainty associated with conflicts (v) uncertainty associated with learning (vi) information in dot and matrix patterns (vii) information in auditory displays and so on. After Shannon's [11] measure of entropy, a large number of other measures of information theoretic entropies have been derived. Renyi [10] defined entropy of order α as

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

which includes Shannon's [11] entropy as a limiting case as $\alpha \to 1$.

Havrada and Charvat [6] introduced first non-additive entropy, given by:

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

Behara and Chawla [1] defined the non-additive τ-entropy, given by

$$H_\tau(P)=\frac{1-\left(\sum_{i=1}^{n} p_i^{1/\tau}\right)^{\tau}}{1-2^{\tau-1}},\ \tau \neq 1,\ \tau > 0. \tag{1.4}$$

Let $P=(p_1, p_2, ..., p_n)$ and $Q=(q_1, q_2, ..., q_n)$ be any two discrete probability distributions. A measure D (P: Q) of divergence or cross entropy or directed divergence is defined as the discrepancy of the probability distribution P from another probability distribution Q. In some sense, it measures the distance of P from Q and the most important and useful measure of directed divergence is due to Kullback and Leibler [9] and is given by

$$D(P:Q)=\sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} \tag{1.5}$$

A general measure of divergence of a discrete probability distribution P from another probability distribution Q has been developed by Csiszer's [4] and is given by

$$D(P:Q)=\sum_{i=1}^{n} q_i\, \phi\left(\frac{p_i}{q_i}\right) \tag{1.6}$$

where $\phi(\cdot)$ is a twice differentiable convex function for which $\phi(1)=0$

It is to be observed that Kullback-Leibler's [9] divergence is a fundamental information measure, special cases of which are mutual information and entropy but the problem of divergence estimation of sources whose distributions are unknown has received relatively little attention.

Some parametric measures of directed divergence are:

$$D_\alpha(P:Q)=\frac{1}{\alpha-1}\ln\sum_{i=1}^{n} p_i^{\alpha} q_i^{1-\alpha},\ \alpha \neq 1,\ \alpha > 0 \tag{1.7}$$

which is Renyi's [10] probabilistic measure of directed divergence.

$$D^{\alpha}(P:Q)=\frac{1}{\alpha-1}\left[\sum_{i=1}^{n} p_i^{\alpha}\, q_i^{1-\alpha} -1\right], \alpha\neq 1,\ \alpha>0 \tag{1.8}$$

which is Havrada and Charvat's [6] probabilistic measure of divergence.

$$D_{\lambda}(P:Q)=\frac{1}{\lambda}\sum_{i=1}^{n}(1+\lambda\, p_i)\ln\frac{1+\lambda\, p_i}{1+\lambda\, q_i},\ \lambda>0 \tag{1.9}$$

which is Ferreri's [5] probabilistic measure of divergence.

Kapur [7] introduced the following measures of directed divergence:

$$D_1(P:Q)=\sum_{i=1}^{n} p_i\ln\frac{p_i}{q_i}-\frac{1}{a}\sum_{i=1}^{n}(q_i+ap_i)\ln\frac{(q_i+ap_i)}{q_i\left(1+a\right)} \tag{1.10}$$

$$D_2(P:Q)=\sum_{i=1}^{n} p_i\ln\frac{p_i}{q_i}-\frac{1}{a}\sum_{i=1}^{n}(1+ap_i)\ln\frac{(1+ap_i)}{1+aq_i};\ a\geq -1 \tag{1.11}$$

In the next section, we have developed a new measure of divergence usually known as R-divergence and studied its important properties.

2. INFORMATION RADIUS BASED ON BURG'S [3] MEASURE OF ENTROPY

It is well-known in the literature of information theory that Shannon's [11] measure of entropy satisfies the following inequalities:

$$H\left(P\right)\leq H\left(P\square Q\right) \tag{2.1}$$

and

$$\left(\frac{H\left(P\right)+H\left(Q\right)}{2}\right)\leq H\left(\frac{P+Q}{2}\right) \tag{2.2}$$

where $H\left(P\right)$ is Shannon's [11] entropy and $H\left(P\square Q\right)$ is a measure of inaccuracy due to Kerridge's [8].

The inequality (2.1) is known as Shannon-Gibbs inequality and the inequality (2.2) arises due to concavity property of Shannon entropy.

The difference in two inequalities (2.1) and (2.2) given by

$$D\left(P;Q\right)=H\left(P\square Q\right)-H\left(P\right)$$

$$=\sum_{i=1}^{n} p_i\log\frac{p_i}{q_i} \tag{2.3}$$

is known as Kullback and Leibler's [9] directed divergence.

And the difference

$$R(P;Q)=H\left(\frac{P+Q}{2}\right)-\left(\frac{H(P)+H(Q)}{2}\right) \qquad (2.4)$$

is known as information radius or Jensen difference divergence measure introduced by Burbea and Rao [2]. By simple calculations, we can write

$$R(P;Q)=\frac{1}{2}\left\{D\left(P;\frac{P+Q}{2}\right)+D\left(Q;\frac{P+Q}{2}\right)\right\} \qquad (2.5)$$

The expression (2.5) is known as R-divergence.

Now, we introduce Burg's [3] entropy instead of Shannon's [11] entropy in (2.4) to obtain a new measure of divergence.

By Burg's [3] entropy, we have

$$B(P)=\sum_{i=1}^{n}\log p_i\ , \qquad B(Q)=\sum_{i=1}^{n}\log q_i \qquad (2.6)$$

Thus, equation (2.4) becomes

$$\begin{aligned}R(P;Q)&=B\left(\frac{P+Q}{2}\right)-\left(\frac{B(P)+B(Q)}{2}\right)\\&=\sum_{i=1}^{n}\log\left(\frac{p_i+q_i}{2}\right)-\frac{1}{2}\sum_{i=1}^{n}\log p_i-\frac{1}{2}\sum_{i=1}^{n}\log q_i\\&=\frac{1}{2}\left\{\sum_{i=1}^{n}\log\left(\frac{p_i+q_i}{2p_i}\right)+\sum_{i=1}^{n}\log\left(\frac{p_i+q_i}{2q_i}\right)\right\}\\&=\frac{1}{2}\left\{D\left(P;\frac{P+Q}{2}\right)+D\left(Q;\frac{P+Q}{2}\right)\right\} \qquad (2.7)\end{aligned}$$

where

$$R(P;Q)\neq D(P;Q)+D(Q;P)=\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{p_iq_i} \qquad (2.8)$$

is directed divergence between P and Q corresponding to Burg's [3] measure of entropy with following properties:

(1) $D(P;Q)\geq 0$

(2) $D(P;Q)=0 \; iff \; P=Q$

(3) $D(P;Q)$ is a convex function of $p_1,p_2,\ldots,p_n$ and $q_1,q_2,\ldots,q_n$

Under these three properties (2.8) is a valid measure of directed divergence.

Next, we discuss properties and generalizations of $R(P;Q)$ **:**

Properties:

It is obvious that

(1) $R(P;Q) \geq 0$

(2) $R(P;Q) = 0 \Rightarrow P = Q = \dfrac{P+Q}{2}$

(3) $R(P;Q)$ being sum of directed divergences is a convex function of $p_1, p_2, \ldots, p_n$ and $q_1, q_2, \ldots, q_n$.

(4) $R(P;Q)$ is symmetric in P and Q that is $R(P;Q) = R(Q;P)$

Of course $R(P;Q) \neq D(P;Q) + D(Q;P) = \sum_{i=1}^{n} \dfrac{(p_i - q_i)^2}{p_i q_i}$

Thus, we see that $R(P;Q)$ defined above satisfies all the essential properties of a measure of directed divergence. Hence, it is a valid measure of R-divergence.

Next, we generalize the R-divergence $R(P;Q)$ by introducing non-negative λ such that 0 ≤ λ ≤ 1 as follows:

$$R_\lambda(P;Q) = B\left(\lambda P + \overline{1-\lambda}Q\right) - \lambda B(P) - (1-\lambda) B(Q) \tag{2.9}$$

$$= \sum_{i=1}^{n} \log\left(\lambda p_i + \overline{1-\lambda} q_i\right) - \lambda \sum_{i=1}^{n} \log p_i - (1-\lambda) \sum_{i=1}^{n} \log q_i$$

$$= \lambda D\left(P; \lambda P + \overline{1-\lambda}Q\right) + (1-\lambda) D\left(Q; \lambda P + \overline{1-\lambda}Q\right) \tag{2.10}$$

Now, we study properties of $R_\lambda(P;Q)$:

(1) $R_\lambda(P;Q) \geq 0$ being sum of directed divergence with +ve constants

(2) $R_0(P;Q) = R_1(P;Q) = 0$

(3) $R_{\frac{1}{2}-\lambda}(P;Q) = R_\lambda(P;Q)$

(4) $R_{1-\lambda}(P;Q) = R_\lambda(Q;P)$

(5) $R_{\frac{1}{2}+\lambda}(P;Q) = R_{\frac{1}{2}-\lambda}(Q;P)$

This indicates that about λ = ½, $R_\lambda(P;Q)$ and $R_{1-\lambda}(Q;P)$ are symmetrical as shown in Fig.-2.1.

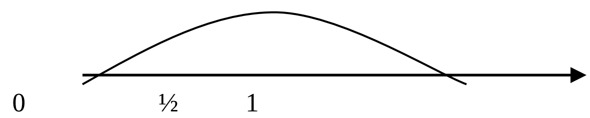

Fig.-2.1

Again, we have the following expression:

$$\frac{dR}{d\lambda}=\sum_{i=1}^{n}\frac{p_i-q_i}{\lambda p_i+(1-\lambda)q_i}+\ln\sum_{i=1}^{n}\frac{q_i}{p_i}$$

$$\frac{d^2R}{d\lambda^2}=-\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{(\lambda p_i+(1-\lambda)q_i)^2}\le 0$$

So $R_\lambda(P;Q)$ is concave function of λ.

We can further generalize $R_\lambda(P;Q)$ as follows:

Let $\lambda_j \ge 0$ and $\sum_{j=1}^{m}\lambda_j = 1$

Then, we have $R(P_1,P_2,.....,P_m)=\sum_{j=1}^{m}\lambda_j.D\left(P_j;\sum_{j=1}^{m}\lambda_j P_j\right)$

is another generalization of information radius given by $R_\lambda(P;Q)$. $R(P_1,P_2,.....,P_m)$ is called information radius and $\sum_{j=1}^{m}\lambda_j P_j$ is called centre where

$0\le \lambda_j \le 1$ and $\sum_{j=1}^{m}\lambda_j = 1$.

REFERENCES

[1] Behara, M. and Chawla, J. S. (1974). Generalized γ-entropy. *Selecta Statistica Canadiana* **2:**15-38.

[2] Burbea, J. and Rao, C.R. (1982). On the convexity of some divergence measures based on entropy functions, IEEE Trans. Inf. The., 28, 489-495.

[3] Burg, J.P. (1972). The relationship between maximum entropy spectra and maximum likelihood spectra. In: Childrers, D.G.(ed). *Modern Spectral Analysis*. pp. 130-131.

[4] Csiszer, I.(1967). Information type measures of differences of probabilistic distributions and indirect observations. Stud. Science Math. Hindi, 2, 299-308.

[5] Ferreri, C. (1980): "Hypoentropy and related heterogeneity divergence measures", *Statistica*, 40(2), 155-168.

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

[7] Kapur, J.N. (1995). *Measures of Information and Their Applications*. Wiley Eastern, New York.

[8] Kerridge, D.F. (1961). Inaccuracy and Inference. *Jr. Royal Statist. Soc. Ser. B*, **23:** 184–194.

[9] Kullback, S. and Leibler, R.A. (1951). On information and sufficiency. *Ann. Maths. Stat.* **22**: 79-86.

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

[11] Shannon, C.E. (1948): "A mathematical theory of communication", *Bell. Sys. Tech. Jr.*, **27**, 379-423, 623-659.