Canadian Journal of Pure and Applied Sciences
Vol. 10, No. 2, pp. 3921-3925, June 2016
Online ISSN: 1920-3853; Print ISSN: 1715-9997
Available online at www.cjpas.net

NEW PARAMETRIC AND NON –PARAMETRIC MEASURES OF CROSS ENTROPY

*Om Parkash and Rakesh Kumar
Department of Mathematics, Guru Nanak Dev University, Amritsar- 143005, India

ABSTRACT

It has been observed that in realistic situations, a single mathematical model measuring distance between two probability distributions may not be appropriate one. This motivation encourages to introduce new measures of divergence or cross-entropy so as to induce flexibility in different disciplines. The present communication is a step in this direction and deals with the development of new parametric and non-parametric measures of divergence in the probability spaces.

Keywords: Entropy, cross entropy, convex function, uncertainty.

INTRODUCTION

Divergence measures play a crucial role in measuring the distance between two probability distributions and in the theory of probability, one of the core issue is to find such an appropriate measures of distance. In the literature of information theory, such a measure of distance also known as measures of cross entropy or measures of divergence provides the difference between two probability distributions P and Q. Such a divergence measure is a useful instrument in solving many optimization problems dealing with a variety of disciplines in physical, biological, engineering and mathematical sciences. The mathematical model for such a divergence measure is probabilistic in nature and was first investigated by Kullback and Leibler (1951). The mathematical expression of this model is given by

$$D(P,Q)=\sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} \tag{1.1}$$

Dragomir (2005) has introduced several new classes of general divergence measures, presented their main properties and developed the connection with known information measure like Csiszar's $f-$divergence or the Jeffreys divergence (1946). Amari (2009) has also contributed towards the development of some divergence measures. Wada (2009) has developed a relation giving two parametric generalization of the well-known relation between log-likelihood and Kullback and Leibler's (1951) divergence measures.

Keeping in view, the fundamental properties and applications areas of such divergence measures in probability spaces, many new measures of divergence have been investigated by various researchers. Some of these are:

$$^{\alpha}D(P;Q)=\frac{1}{\alpha-1}\left[\sum_{i=1}^{n} p_i^{\alpha} q_i^{1-a}-1\right],$$

$$\alpha\neq 1,\ \alpha>0 \tag{1.2}$$

which is Havrda and Charvat (1967) probabilistic measure of divergence.

$$_{\alpha}D(P;Q)=\frac{1}{\alpha-1}\log\sum_{i=1}^{n} p_i^{\alpha} q_i^{1-a},$$

$$\alpha\neq 1,\ \alpha>0 \tag{1.3}$$

which is Renyi's (1961) probabilistic measure of divergence.

Parkash and Mukesh (2011, 2014) developed the following divergence measures and applied these measures to the fields of Statistics and Operations Research:

$$D(P;Q)=\sum_{i=1}^{n}\left(\frac{p_i^2}{q_i}+\frac{q_i^2}{p_i}-2p_i\right) \tag{1.4}$$

$$S(P;Q)=\sum_{i=1}^{n}\left[\left(\frac{1}{\sqrt{p_i}}+\frac{1}{\sqrt{q_i}}\right)\left(\frac{p_i+q_i}{2}\right)^{3/2}-2p_i\right] \tag{1.5}$$

---

*Corresponding author e-mail: omparkash777@yahoo.co.in

$$D^{\alpha,\beta}(P;Q)=\frac{\sum_{i=1}^{n}p_i\left(\frac{\alpha}{2\beta}+\frac{1}{2}\right)^{\log\frac{p_i}{q_i}}-1}{\alpha-\beta},$$

$$\alpha>0,\beta>0,\alpha\neq\beta \qquad (1.6)$$

where α,β are real parameters.

Furthermore, Parkash and Priyanka (2013, 2014) developed the following measures of divergence and applied their findings to the field of Coding Theory:

$$^{\alpha}D(P;Q)=\frac{1}{\alpha-1}\left(\prod_{i=1}^{n}\left(\frac{p_i}{q_i}\right)^{-p_i(1-\alpha)}-1\right),\alpha>1,\alpha\neq 1 \qquad (1.7)$$

$$^{\alpha}D(P;Q)=\frac{1}{1-\alpha}\left(\prod_{i=1}^{n}q_i^{-p_i(1-\alpha)}-\prod_{i=1}^{n}p_i^{-p_i(1-\alpha)}\right),$$

$$0\leq\alpha<1,\alpha\neq 1 \qquad (1.8)$$

$$_{\alpha}D(P;Q)=\frac{\sum_{i=1}^{n}p_i\alpha^{\ln\frac{p_i}{q_i}}-1}{1-\alpha},\alpha>1 \qquad (1.9)$$

Huang *et al*. (2016) presented a generalization of Kullback and Leibler (1951) divergence measure in the form of Tsallis statistics and studied several important properties of the new generalization such as pseudo-additivity, positivity and monotonicity in addition to the essential properties of a divergence measure. Some other pioneers who have made the study of divergence measures more rigorous include Chen *et al*. (2012), Sankaran *et al*. (2016), Di Crescenzo and Longobardi (2015), Liu *et al*. (2014), Taneja (2004) and Zhao and Blahut (2007) etc.

In section 2, we propose two new measures of divergence and study their detailed properties.

2. TWO NEW GENERALIZED MEASURES OF CROSS ENTROPY

(i) A New Non Parametric Measure of Cross-Entropy

For any two discrete probability distributions P and Q, we propose the following measure of divergence:

$$D(P;Q)=\sum_{i=1}^{n}p_i\ln\frac{p_i}{\left(\frac{\sqrt{p_i}+\sqrt{q_i}}{2}\right)^2} \qquad (2.1)$$

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

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

(ii) $D(P;Q)=0$ iff $P=Q$

(iii) $D(P;Q)$ is convex function of both

$p_1,p_2,...,p_n$ and $q_1,q_2,...,q_n$.

(i) To prove $D(P;Q)\geq 0$, we use the inequality,

$$\ln x\leq x-1 \qquad (2.2)$$

The result (2.1) can be written as

$$-\sum_{i=1}^{n}p_i\ln\frac{\left(\frac{\sqrt{p_i}+\sqrt{q_i}}{2}\right)^2}{p_i}\geq-\sum_{i=1}^{n}p_i\left[\frac{\left(\frac{\sqrt{p_i}+\sqrt{q_i}}{2}\right)^2}{p_i}-1\right]$$

$$=-\sum_{i=1}^{n}\left(\frac{\sqrt{p_i}+\sqrt{q_i}}{2}\right)^2+\sum_{i=1}^{n}p_i$$

$$=\frac{1}{2}\left(1-\sum_{i=1}^{n}\sqrt{p_iq_i}\right)\geq 0$$

(ii) $D(P;Q)=0$ if and only if $P=Q$.

(iii) $D(P;Q)$ is convex.

Let $D(P,Q)=f(p_1,p_2,...p_n;q_1,q_2,...,q_n)$

$$=\sum_{i=1}^{n}p_i\ln\frac{p_i}{\left(\frac{\sqrt{p_i}+\sqrt{q_i}}{2}\right)^2}$$

Then, $\frac{\partial f}{\partial p_i}=\frac{\sqrt{q_i}}{\sqrt{p_i}+\sqrt{q_i}}+\ln\frac{4p_i}{\left(\sqrt{p_i}+\sqrt{q_i}\right)^2}$

Also, $\frac{\partial^2 f}{\partial p_i^2}=\frac{\sqrt{q_i}\left(p_i+2\sqrt{p_iq_i}\right)}{2p_i^{\frac{3}{2}}\left(\sqrt{p_i}+\sqrt{q_i}\right)^2}>0$

and $\frac{\partial^2 f}{\partial p_i\partial p_j}=0$ 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

$$\begin{bmatrix}\frac{\sqrt{q_1}\left(p_1+2\sqrt{p_1q_1}\right)}{2p_1^{\frac{3}{2}}\left(\sqrt{p_1}+\sqrt{q_1}\right)^2} & 0 & \cdots & 0\\ 0 & \frac{\sqrt{q_2}\left(p_2+2\sqrt{p_2q_2}\right)}{2p_2^{\frac{3}{2}}\left(\sqrt{p_2}+\sqrt{q_2}\right)^2} & \cdots & 0\\ \vdots & & & \\ 0 & 0 & \cdots & \frac{\sqrt{q_n}\left(p_n+2\sqrt{p_nq_n}\right)}{2p_n^{\frac{3}{2}}\left(\sqrt{p_n}+\sqrt{q_n}\right)^2}\end{bmatrix}$$

which is positive definite matrix. Thus, $D(P;Q)$ is a convex function of $p_1,p_2,\ldots,p_n$.

Proceeding on similar lines, we have proved that $D(P;Q)$ is a convex function of $q_1, q_2, \ldots, q_n$.

For this purpose, we have made the following observations:

$$\frac{\partial f}{\partial q_i} = -\frac{p_i}{\sqrt{p_i q_i} + q_i}$$

Also, $$\frac{\partial^2 f}{\partial q_i^2} = \frac{p_i\left(p_i + 2\sqrt{p_i q_i}\right)}{2\sqrt{p_i q_i}\left(\sqrt{p_i q_i} + q_i\right)^2} > 0$$

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

The Hessian matrix of second order partial derivatives of f with respect to $q_1, q_2, \ldots, q_n$ is

$$\begin{bmatrix} \frac{p_1\left(p_1+2\sqrt{p_1q_1}\right)}{2\sqrt{p_1q_1}\left(\sqrt{p_1q_1}+\sqrt{q_1}\right)^2} & 0 & \ldots & 0 \\ 0 & \frac{p_2\left(p_2+2\sqrt{p_2q_2}\right)}{2\sqrt{p_2q_2}\left(\sqrt{p_2q_2}+\sqrt{q_2}\right)^2} & \ldots & 0 \\ . & & & \\ . & & & \\ . & & & \\ 0 & 0 & \ldots & \frac{p_n\left(p_n+2\sqrt{p_nq_n}\right)}{2\sqrt{p_nq_n}\left(\sqrt{p_nq_n}+\sqrt{q_n}\right)^2} \end{bmatrix}$$

which is positive definite matrix. Thus, $D(P;Q)$ 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 shown in the Table 1, we have presented $D(P;Q)$ as shown in the Figure 1.

Table 1. $D(P,Q)$ against p and $q = 0.5$

p	q	$D(P,Q)$
0.1	0.5	0.531004
0.2	0.5	0.278072
0.3	0.5	0.118709
0.4	0.5	0.029049
0.5	0.5	0
0.6	0.5	0.029049
0.7	0.5	0.118709
0.8	0.5	0.278072
0.9	0.5	0.531004

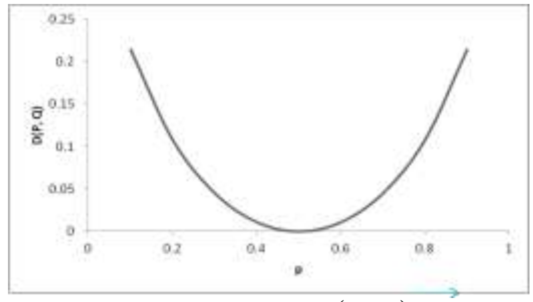

Fig. 1. Convexiety of $D(P,Q)$

Hence, under the above conditions, the function $D(P,Q)$ is a correct measure of cross-entropy.

(ii) A New Generalized Parametric Measure of Cross-Entropy

Now, we propose a new generalized parametric measure of cross entropy is given by

$$D_\alpha(P;Q) = \frac{\ln\left(\prod_{i=1}^{n}\left(\frac{p_i}{q_i}\right)^{(\alpha-1)p_i}\right)}{(\alpha-1)},$$

$$\alpha > 0,\ \alpha \neq 1 \qquad (2.3)$$

We have,

$$\underset{\alpha\to1}{lt}\left[D_\alpha(P;Q)\right] = \underset{\alpha\to1}{lt}\left(\frac{\ln\left(\prod_{i=1}^{n}\left(\frac{p_i}{q_i}\right)^{(\alpha-1)p_i}\right)}{\alpha-1}\right)$$

$$= \sum_{i=1}^{n} p_i \ln\frac{p_i}{q_i}$$

Thus $D_\alpha(P;Q)$ is a generalization of Kullback and Leibler (1951) measure of cross entropy. To prove this is valid measures of cross-entropy, we prove the following properties.

(i) We have,

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

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

$$= \sum_{i=1}^{n} p_i \ln\frac{p_i}{q_i} \geq 0.$$

Thus, $D_\alpha(P;Q) \geq 0$.

(ii) $D_\alpha(P;Q) = 0$ iff $P = Q$.

(iii) To prove $D_\alpha(P;Q)$ is convex function of both P and Q, we have

$$\frac{\partial(D_\alpha(P,Q))}{\partial p_i} = 1 + \ln\frac{p_i}{q_i}$$

Also, $\frac{\partial^2(D_\alpha(P,Q))}{\partial p_i^2} = \frac{1}{pi} > 0$

and $\frac{\partial^2(D_\alpha(P,Q))}{\partial p_i \partial p_j} = 0$ for

$i, j = 1, 2, \ldots, n; i \neq j$.

Therefore, the Hessian matrix is given by

$$\begin{bmatrix} \frac{1}{p_1} & 0 & \ldots & 0 \\ 0 & \frac{1}{p_2} & \ldots & 0 \\ . & & & \\ . & & & \\ . & & & \\ 0 & 0 & \ldots & \frac{1}{p_n} \end{bmatrix}$$

which is positive definite matrix. So, it is convex function of $p_1, p_2, ..., p_n$.

Similarly, $\frac{\partial(D_\alpha(P,Q))}{\partial q_i} = -\frac{p_i}{q_i}$

Also, $\frac{\partial^2(D_\alpha(P,Q))}{\partial q_i^2} = \frac{p_i}{q_i^2} > 0$

and $\frac{\partial(D_\alpha(P,Q))}{\partial q_i \partial q_j} = 0$ for

$i, j = 1, 2, \ldots, n; i \neq j$.

Therefore, Hessian matrix is given by

$$\begin{bmatrix} \frac{p_1}{q_1^2} & 0 & \ldots & 0 \\ 0 & \frac{p_2}{q_2^2} & \ldots & 0 \\ . & & & \\ . & & & \\ . & & & \\ 0 & 0 & \ldots & \frac{p_n}{q_n^2} \end{bmatrix}$$

which is also positive definite matrix. So, it is convex function of $q_1, q_2, ..., q_n$. Moreover, with the help of numerical data shown in Table 2, we have presented $D_\alpha(P;Q)$ as shown in the following Fig.ure 2.

Table 2. $D_\alpha(P,Q)$ against p for $n = 2$, and $\alpha = 2$

p	q	$D_\alpha(P,Q)$
0.1	0.5	0.214813
0.2	0.5	0.108127
0.3	0.5	0.045178
0.4	0.5	0.010932
0.5	0.5	0
0.6	0.5	0.010932
0.7	0.5	0.045178
0.8	0.5	0.108127
0.9	0.5	0.214813

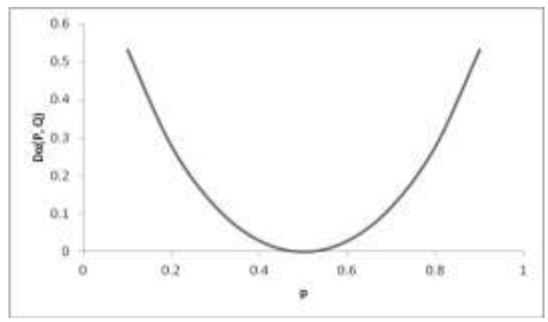

Fig. 2. Convexity of $D_\alpha(P;Q)$

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

CONCLUDING REMARKS

To induce flexibility in many real life situations, we need a variety of parametric and non-parametric measures of divergence so as to extend their applications in different areas of research. Our study is a step in this direction and such a study can be made for other distributions including continuous and fuzzy distributions.

ACKNOWLEDGEMENTS

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

REFERENCES

Amari, SI. 2009. $\alpha-$ divergence is unique, belonging to both $f-$divergence and Bregman divergence classes. IEEE Trans. Inform. Theory. 55(11):4925-4931.

Chen, X., Kar, S. and Ralescu, DA. 2012. Cross-entropy measure of uncertain variables. Inform. Sci. 201:53-60.

Di Crescenzo, A. and Longobardi, M. 2015. Some properties and applications of commulative Kullback-Leibler information. Appl. Stoch. Models Bus. Ind. 31:875-891.

Dragomir, SS. 2005. Some general divergence measures for probability distribution. Acta Mathematica. Hungarica. 109:331-345.

Havrda, JH. and Charvat, F. 1967. Quantification methods of classification process: Concept of structural α-entropy. Kybernetika. 3:30-35.

Huang, J., Yong, WA. and Hong, L. 2016. Generalization of the Kullback-Leibler divergence in the Tsallis Statistics. J. Math. Anal. Appl. 436(1):501-512.

Jeffreys, H. 1946. An Invariant Form for the Prior Probability in Estimation Problems. Proc. of the Royal Society of London. Series A .186:453-461.

Kullback, S. and Leibler, RA. 1951. On information and sufficiency. Annals of Mathematical Statistics. 22:79-86.

Liu, R., Chakrabarti, A., Samanta, TK. and Ghosh, JK. 2014. On divergence measures leading to Jeffreys and other reference priors. Bayesian Anal. 9(2):331.

Parkash, O. and Kakkar, P. 2013. New measures of information and their applications in coding theory. Canadian Journal of Pure and Applied Sciences. 8(2):2905-2912.

Parkash, O. and Kakkar, P. 2014. New information theoretic models, their detailed properties and new inequalities. Canadian Journal of Pure and Applied Sciences. 8(3): 3115-3123.

Parkash, O. and Mukesh. 2011. Two new symmetric divergence measures and information inequalities. International Journal of Mathematics and Application. 4:165-179.

Parkash, O. and Mukesh. 2014. Portfolio optimization using measures of cross-entropy. Canadian Journal of Pure and Applied Sciences. 8(2):2963-2967.

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

Sankaran, PG., Sunoj, SM. and Nair, NU. 2016. Kullback-Leibler divergence: A quantile approach. Statistics and Probability Letters. 111:72-79.

Taneja, IJ. 2004. Relative Divergence Measures and Information Inequality-Inequality Theory and Applications. Nova Science Publishers Inc. Huntington, New York, USA.

Wada, T. 2009. Generalized log-likelihood functions and Bregman divergences. J. Math. Phys. 50(11):7.

Zhao, Z. and Blahut, RE. 2007. On the asymptotic consistency of minimum divergence and least square principles. IEEE Trans. Inform. Theory. 53:3283-3287.

Received: March 9, 2016; Accepted: April 13, 2016