Jñānābha, Special Issue, 2018, 72-80
Proceedings : MSSCID-2017, 20^{th} Annual Conference of VPI, Jaipur, India

AN INFORMATION MEASURE OF f-DIVERGENCE FAMILY AND ITS PROPERTIES

By
K.C.Jain[1] and Devendra Singh[2]
[1]Department of Mathematics, IIIT Kota, JLN Marg, MNIT Campus, Jaipur-302017, Rajasthan, India
[2]Department of Mathematics, Malaviya National Institute of Technology, Jaipur-302017, Rajasthan, India
Email: devendra.mnit2013@gmail.com , 2013rma9061@mnit.ac.in

(*Received : November* 24, 2017 ; *Revised: September* 03, 2018)

Abstract

A non-parametric symmetric information divergence measure is proposed. This information measure belongs to the family of Csiszrs f-divergences. Its properties are studied and discussed. Further, we have derived some new bounds for this information divergence measure in terms of some recognized divergence measures based on two distinct probability distributions.

2010 Mathematics Subject Classification: 94A17, 26D15.
Keywords and phrases: Non-parametric measure, Kullback-Leibler Measure, Csiszàr's f-divergence, Information measure, Triangular Discremination, Symmetric chi-square divergence.

1 Introduction

The information divergence measures are used to find out distance or difference or affinity between two probability distributions. Non parametric divergence measures give the amount of information supplied by the data for discriminating in favor of a probability distribution P against another Q. The construction of information divergence measure for two distinct probability distributions is not an easy task.

Let $\Gamma_n = \left\{ P = (p_1, p_2, p_3, ..., p_n); p_i > 0, \sum\limits_{i=1}^{n} p_i = 1 \right\}, n \geq 0$ be the set of all discrete probability distributions. For convex function $f : (0, \infty) \to \mathbb{R}$ (set of real numbers) and for probability distributions $P, Q \in \Gamma_n$, the f-divergence measure, by Csiszàr's, [3,4] and Ali & Silvey, [1], is defined as

$$C_f(P, Q) \sum_{i=1}^{n} q_i f(p_i/q_i) \tag{1.1}$$

An important property of this divergence measure is that many known divergence measures can be obtained from this information measure by appropriately defining the convex function f.

In this paper, we present a new non-parametric symmetric information divergence measure which belongs to the family of Csiszàr's f-divergences [3, 4].In section 2; we discuss information inequalities of Csiszàr's f-divergences with other known divergence measures. New symmetric information divergence measure is obtained in section 3. In section 4, we have derived some information inequalities for the new information divergence measure in terms of some well-known divergence measures. Section 5, concludes the paper.

For brevity, we will denote $C_f(P,Q), p_i, q$ and $\sum_{i=1}^{n}$ by $C(P,Q), p, q$ and $\sum$, respectively.

Some well-known information divergence measures are as follows.

Chi-square Divergence [18]

$$\xi^2(P,Q) = \sum \frac{(p-1)^2}{q}. \tag{1.2}$$

Symmetric Chi-square Divergence

$$\psi\left(P,Q\right) = \chi^2\left(P,Q\right) + \chi^2\left(Q,P\right) = \sum \frac{(p+q)\ \left(p-q\right)^2}{pq}. \tag{1.3}$$

Kullback and Leibler Divergence [14]

$$K\left(P,Q\right) = \sum p \ln(p/q). \tag{1.4}$$

Kullback and Leibler Symmetric Divergence

$$J\left(P,Q\right) = K\left(P,Q\right) + K\left(Q,P\right) = \sum \left(p-q\right)\ \ln(p/q). \tag{1.5}$$

Triangular Discrimination [5, 24]

$$\Delta\left(P,Q\right) = \sum \frac{\left(p-q\right)^2}{p+q}. \tag{1.6}$$

Bhattacharyya Distance [2]

$$B\left(P,Q\right) = \sum \sqrt{pq}. \tag{1.7}$$

Harmonic Mean Divergence [17]

$$W\left(P,Q\right) = \sum \frac{2p\ q}{p+q}. \tag{1.8}$$

Kumar and Johnson [15]

$$\Psi M\left(P,Q\right) = \sum \frac{\left(p^2-q^2\right)^2}{2\left(pq\right)^{3/2}} \tag{1.9}$$

2 Information Inequalities

Different kinds of bounds on the information divergence measures have been studied during the recent past [6-11,25]. In [25] Kumar and Taneja unified and generalized information bounds for $C(P,Q)$ studied by Dragomir [6-11]. The main results in [25] given in the following theorem.

Theorem 2.1. *Let $f\ : I \subset R_+ \to \mathbb{R}$ be a mapping which is normalized and suppose that*

(i) f is twice differentiable on $(r,R), 0 \le r \le 1 \le R < \infty$

(ii) There exists real constants $m, M,$ such that $m < M$ and $m \leq t^{2-s} f''(t) \leq M, \ \forall\ t \in (r, R)\,,\ s \in \mathbb{R}$. If $P, Q\ \in \Omega_n$ are discrete probability distributions with $0 < r \leq p/q \leq R < \infty$,

$$m\ \Phi_s(P,Q) \quad \leq \quad C(P,Q) \quad \leq\ M\ \Phi_s(P,Q) \tag{2.1}$$

and

$$m\ \{\ \eta_s(P,Q) - \Phi_S(P,Q)\ \} \quad \leq\ C_\rho(P,Q) - C(P,Q)\ \leq M\ \{\ \eta_s(P,Q) - \Phi_s(P,Q)\ \}\,, \tag{2.2}$$

where

$$\Phi_s(P,Q) = \begin{cases} {}^2K_s(P,Q)\,, & s \neq 0, 1, \\ K(Q,P)\,, & s = 0, \\ K(P,Q)\,, & s = 1, \end{cases} \tag{2.3}$$

$${}^2K_s(P,Q) = [s(s-1)]^{-1}\left[\sum p^s q^{1-s} - 1\right]\ , \quad s \neq 0, 1, \tag{2.4}$$

$$K(P,Q) = \sum p \ln\left(\frac{p}{q}\right)\ ,$$

$$C_\rho(P,Q) = \sum (p-q)\ \ f'\ \left(\frac{p}{q}\right) \tag{2.5}$$

$$\eta_s(P,Q) = C_{\varphi'_s}\left(\frac{P^2}{Q}, P\right) - C_{\varphi'_s}(P,Q)$$

$$= \begin{cases} (s-1)^{-1} \sum (p-q)\left(\frac{p}{q}\right)^{s-1}\ , & s \neq 1, \\ \sum (p-q)\ \ln\ \left(\frac{p}{q}\right)\ , & s = 1. \end{cases} \tag{2.6}$$

As a consequence of this theorem, following information inequalities which are interesting from the information-theoretic point of view are also obtained in [11].

(i) The case $s = 2$ provides the information bounds in terms of the Chi-square divergence, $\chi^2(P,Q)$,

$$\frac{m}{2}\chi^2(P,Q)\ \leq\ \ C\ \ (P,Q)\ \ \leq \frac{M}{2}\chi^2(P,Q) \tag{2.7}$$

and

$$\frac{m}{2}\chi^2(P,Q)\ \ \leq\ C_\rho(P,Q)\ -C\ \ (P,Q)\ \leq\ \frac{M}{2}\chi^2(P,Q)\,. \tag{2.8}$$

(ii) For $s = 1$, the information bounds in terms of the Kullback-Leibler divergence,$K(P,Q)$,

$$m\ K\ \ (P,Q)\ \leq\ C\ \ (P,Q)\ \ \leq\ M\ K\ \ (P,Q) \tag{2.9}$$

and

$$m\ K\ \ (P,Q)\ \leq\ C\rho\ \ (P,Q) - C\ \ (P,Q)\ \leq\ M\ K\ \ (P,Q)\,. \tag{2.10}$$

(iii) For $s = 1/2$ yields the information bounds in terms of the Hellingers discrimination, $h(P,Q)$,

$$4m\ h\ \ (P,Q)\ \leq\ C\ \ (P,Q)\ \leq\ 4M\ h\ \ (P,Q) \tag{2.11}$$

and

$$4\,m\,\left(\frac{1}{4}\eta_{1/2}\left(P,Q\right)-h\left(P,Q\right)\right)\;\leq\;C_{\rho}\left(P,Q\right)-C\left(P,Q\right)\;\leq\;4\,M\left(\frac{1}{4}\eta_{1/2}\left(P,Q\right)-h\left(P,Q\right)\right). \tag{2.12}$$

(iv) For $s=0$, the information bounds in terms of the Kullback-Leibler and Chi-square divergence

$$mK\;\left(P,Q\right)\leq C\left(P,Q\right)\;\leq\;MK\;\left(P,Q\right) \tag{2.13}$$

and

$$m\;\left\{\;\chi^{2}\left(Q,P\right)-K\left(Q,P\right)\;\right\}\;\leq\;C\rho\left(P,Q\right)-C\;\left(P,Q\right)\;\leq\;M\;\left\{\;\chi^{2}\left(Q,P\right)-K\;\left(Q,P\right)\;\right\} \tag{2.14}$$

3 New Symmetric Information Divergence Measure

Now, we consider the function $f:(0,\infty)\to\mathbb{R}$ given by

$$f\left(t\right)=\frac{\left(t-1\right)^{4}\;\left(t^{2}+1\right)\;\left(t+1\right)\;\left(t^{2}+3t+1\right)}{t^{4}}. \tag{3.1}$$

Then

$$f'\left(t\right)=\frac{\left(t-1\right)^{3}\left(5t^{6}+15t^{5}+15t^{4}+15t^{3}+14t^{2}+12t+4\right)}{t^{5}} \tag{3.2}$$

and

$$f''\left(t\right)=\frac{10(t+1)^{2}(t-1)^{2}\left(2t^{5}+t^{3}+t^{2}+2\right)}{t^{6}}. \tag{3.3}$$

The function $f(t)$ is convex since $f''(t)>0$ for all $t>0$ and normalized also since $f(1)=0$.

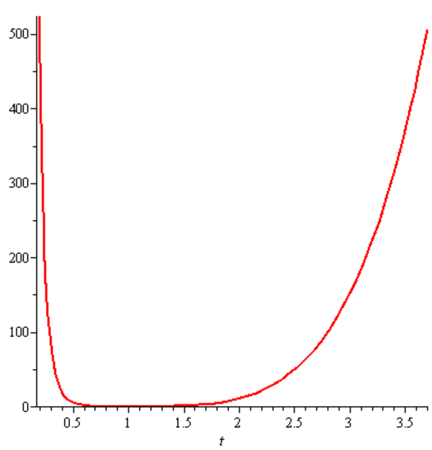

Figure - 1

Figure 1: Shows the behavior of the function $f(t)$ and which is always convex.

Now, we have the following new information divergence measure belonging to the Csiszàr's f-divergence family,

$$Z\left(P,Q\right)=\sum\frac{\left(p-q\right)^{4}\left(p^{2}+q^{2}\right)\;\left(p+q\right)\;\left(p^{2}+3pq+q^{2}\right)}{p^{4}q^{4}}, \tag{3.4}$$

where

(a) Divergence measure $Z(P,Q)$ is symmetric with respect to probability distributions.

(b) $Z\left(P,Q\right) \geq 0$ *and* $Z\left(P,Q\right) = 0,\ iff\ P = Q$

(c) The function $f(t)$ is *-self conjugate since $f*(t) \equiv t\ f(1/t) = f(t)$.

4 Bounds For $Z(P,Q)$

We now derive information inequalities providing bounds for $Z(P,Q)$ in terms of the well-known divergence measures in the following propositions.

Proposition 4.1. *Let* $(P,Q) \in \Gamma_n \times \Gamma_n$*, then we have the following new inequality*

$$Z\left(P,Q\right) \leq \left[\chi^2\left(P,Q\right) + \chi^2\left(Q,P\right)\right]^2 \left[\left(\frac{2}{W\left(P,Q\right)}\right)^2 + \left(\frac{1}{B\left(P,Q\right)}\right)^2\right], \tag{4.1}$$

where $\chi^2(P,Q) + \chi^2(Q,P)$*,* $W(P,Q)$ *and* $B(P,Q)$ *are given by* (1.3)*,* (1.8) *and* (1.7) *respectively.*

Proof. From (3.4), we have,

$$\begin{aligned}
Z(P,Q) &= \sum \frac{(p-q)^4\,(p+q)\,(p^2+q^2)\,(p^2+3pq+q^2)}{p^4q^4} \\
&\leq \sum \frac{(p-q)^4\,(p+q)\,(p+q)\,(p^2+3pq+q^2)}{p^4q^4}; \qquad \text{Since } p^2+q^2 \leq p+q \\
&= \sum \frac{(p-q)^4(p+q)^2}{p^2q^2} \sum \frac{(p^2+3pq+q^2)}{p^2q^2} \\
&= \left[\sum \frac{(p-q)^2\,(p+q)}{pq}\right]^2 \left[4\left(\sum \frac{p+q}{2pq}\right)^2 + \left(\sum \frac{1}{\sqrt{pq}}\right)^2\right].
\end{aligned}$$

Therefore,

$$Z(P,Q) \leq \left[\chi^2\left(P,Q\right) + \chi^2\left(Q,P\right)\right]^2 \left[\left(\frac{2}{W\left(P,Q\right)}\right)^2 + \left(\frac{1}{B\left(P,Q\right)}\right)^2\right].$$

Hence the result. □

Proposition 4.2. *Let* $(P,Q) \in \Gamma_n \times \Gamma_n$*, then we have the following new inequality*

$$Z(P,Q) \leq 8\ \Psi M(P,Q)\ \Delta(P,Q)\ \left[\frac{1}{W(P,Q)}\right]^2 \frac{1}{B(P,Q)}. \tag{4.2}$$

where $Z(P,Q), \Psi M(P,Q), \Delta(P,Q), W(P,Q)$ *and* $B(P,Q)$ *are given by* (3.4)*,* (1.9)*,* (1.6)*,* (1.8) *and* (1.7) *respectively.*

Proof. From (3.4), we have,

$$Z(P,Q) = \sum \frac{(p-q)^4(p+q)(p^2+q^2)(p^2+3pq+q^2)}{p^4q^4}$$

$$\leq \sum \frac{(p-q)^4(p+q)(p+q)(p^2+3pq+q^2)}{p^4q^4}$$
$$= 2\sum \frac{(p^2-q^2)^2}{2(pq)^{3/2}} \sum \frac{1}{\sqrt{pq}} \sum \frac{(p-q)^2}{p+q} \sum \frac{(p+q)(p^2+3pq+q^2}{p^2q^2}$$
$$\leq 2\sum \frac{(p^2-q^2)^2}{2(pq)^{3/2}} \sum \frac{1}{\sqrt{pq}} \sum \frac{(p-q)^2}{p+q} 4 \left(\sum \frac{p+q}{2pq}\right)^2$$

Since $p^2+3pq+q^2 \leq p+q$.

Therefore,

$$Z(P,Q) \leq 8\ \Psi M(P,Q) \quad \Delta(P,Q) \quad \left[\frac{1}{W(P,Q)}\right]^2 \ \frac{1}{B(P,Q)}.$$

Hence the result. □

Proposition 4.3. *Let* $\chi^2(P,Q)$ *and* $Z(P,Q)$ *be defined as in* (1.2) *and* (3.4) *respectively. For* $P,Q \in \Gamma_n \times \Gamma_n$ *and* $0 < r \leq p/q \leq R < \infty$, *we have*

$$\frac{5(R+1)^2(R-1)^2(2R^5+R^3+R^2+2)}{R^6}\chi^2(P,Q) \leq Z(P,Q)$$
$$\leq \frac{5(r+1)^2(r-1)^2(2r^5+r^3+r^2+2)}{r^6}\chi^2(P,Q), \quad (4.3)$$

and

$$\frac{5(R+1)^2(R-1)^2(2R^5+R^3+R^2+2)}{R^6}\chi^2(P,Q) \leq Z_\rho(P,Q) - Z(P,Q)$$
$$\leq \frac{5(r+1)^2(r-1)^2(2r^5+r^3+r^2+2)}{r^6}\chi^2(P,Q). \quad (4.4)$$

where

$$Z_\rho(P,Q) = \sum \frac{(p-q)^4\left(5p^6+15p^5q+15p^4q^2+15p^3q^3+14p^2q^4+12p\ q^5+4q^6\right)}{p^5q^4}. \quad (4.5)$$

Proof. From (3.2), we have

$$f'(t) = \frac{(t-1)^3\left(5t^6+15t^5+15t^4+15t^3+14t^2+12t+4\right)}{t^5}, \quad (4.6)$$

so that

$$Z_\rho(P,Q) = \sum (p-q)\, f'\ (p/q)$$
$$= \sum \frac{(p-q)^4\left(5p^6+15p^5q+15p^4q^2+15p^3q^3+14p^2q^4+12p\ q^5+4q^6\right)}{p^5q^4}. \quad (4.7)$$

Further, from (3.3), we have

$$f''(t) = \frac{10(t+1)^2(t-1)^2\left(2t^5+t^3+t^2+2\right)}{t^6}. \tag{4.8}$$

If $t \in [a,b] \subset (0,\infty)$, then

$$\frac{10(b+1)^2(b-1)^2\left(2b^5+b^3+b^2+2\right)}{b^6} \le f''(t) \le \frac{10(a+1)^2(b-1)^2\left(2a^5+a^3+a^2+2\right)}{a^6}$$

Or, consequently

$$\frac{10(R+1)^2(R-1)^2\left(2R^5+R^3+R^2+2\right)}{R^6} \le f''(t) \le \frac{10(r+1)^2(r-1)^2\left(2r^5+r^3+r^2+2\right)}{r^6}. \tag{4.9}$$

Hence, from (2.7) and (2.8), we get inequalities (4.3) and (4.4), respectively. □

Proposition 4.4. *Let* $K(P,Q)$, $Z(P,Q)$ *and* $Z_\rho(P,Q)$ *be defined as in* (1.4), (3.4) *and* (4.5) *respectively. For* $P,\ Q \in \Gamma_n \times \Gamma_n$ *and* $0 < r \le p/q \le R < \infty$, *we have*

$$\frac{10(R+1)^2(R-1)^2\left(2R^5+R^3+R^2+2\right)}{R^5}K(P,Q) \le Z(P,Q)$$
$$\le \frac{10(r+1)^2(r-1)^2\left(2r^5+r^3+r^2+2\right)}{r^5}K(P,Q) \tag{4.10}$$

and

$$\frac{10(R+1)^2(R-1)^2\left(2R^5+R^3+R^2+2\right)}{R^5}K(P,Q) \le Z_\rho(P,Q) - Z(P,Q)$$
$$\le \frac{10(r+1)^2(r-1)^2\left(2r^5+r^3+r^2+2\right)}{r^5}K(P,Q). \tag{4.11}$$

Proof. From (3.3), we have

$$f''(t) = \frac{10(t+1)^2(t-1)^2\left(2t^5+t^3+t^2+2\right)}{t^6}. \tag{4.12}$$

Let the function $h : [r,R] \to \mathbb{R}$, such that

$$h(t) = t\ f''(t) = \frac{10(t+1)^2(t-1)^2\left(2t^5+t^3+t^2+2\right)}{t^5}. \tag{4.13}$$

Then

$$\inf_{t \in [r,R]} h(t) = \frac{10(R+1)^2(R-1)^2\left(2R^5+R^3+R^2+2\right)}{R^5} \tag{4.14}$$

and

$$\sup_{t \in [r,R]} h(t) = \frac{10(r+1)^2(r-1)^2\left(2r^5+r^3+r^2+2\right)}{r^5}. \tag{4.15}$$

Hence, from (2.9) and (2.10), we get inequalities (4.10) and (4.11), using (4.14) and (4.15), respectively. □

5 Conclusion

In this new work, we have obtained the new information divergence measure using properties of Csiszàr's f-Divergence category. This work is very interesting in the field of the information theory. We have also described different bounds for new derived f-divergence measure in terms of other recognized information divergence measures. Work on generalization of these bounds is in progress and will be reported timely.

References

[1] S.M. Ali and S.D. Silvey, A general class of coefficients of divergence of one distribution from another, *Jour. Roy. Statist. Soc.,* B **28** (1966), 131-142.

[2] A. Bhattacharyya, On some analogues to the amount of information and their uses in statistical estimation, **8** (1946), 1-14, .

[3] I. Csiszár's, Information-type measure of difference of probability distributions and indirect observations, *Studia Sci. Math. Hungar.,* **2** (1967), 299-318.

[4] I. Csiszár's, Information measures: A critical survey, *In Trans. Seventh Prague Conf. On Information Theory*, A, Academia, Prague (1974), 73-86.

[5] L. Le Cam, *Asymptotic Methods in Statistical Decision Theory*, Springer-Verlag, New York, (1986).

[6] S.S. Dragomir, *Some inequalities for the Csiszár's ϕ-divergence, Inequalities for the Csiszár's f-divergence in Information Theory*, (Edited by S.S. Dragomir), `http://rgmia.vu.edu.au/monographs/csiszar.html`, (2000).

[7] S.S. Dragomir, *A converse inequality for the Csiszar's - divergence, Inequalities for the Csiszár's f-divergence in Information Theory*, (Edited by S.S. Dragomir), `http://rgmia.vu.edu.au/monographs/csiszar.html`, (2000).

[8] S.S. Dragomir, *Some inequalities for (m, M)-convex mappings and applications for the Csiszár's ϕ-divergence in information theory, Inequalities for the Csiszár's f-divergence in Information Theory*, (Edited by S.S. Dragomir), `http://rgmia.vu.edu.au/monographs/csiszar.html`, (2000).

[9] S.S. Dragomir, *Upper and lower bounds for Csiszár's f-divergence in terms of the Kullback-Leibler distance and applications, Inequalities for the Csiszár's f-divergence in Information Theory*, (Editedby S.S. Dragomir), `http://rgmia.vu.edu.au/monographs/csiszar.html`, (2000).

[10] S.S. Dragomir, *Upper and lower bounds for Csiszár's f-divergence in terms of the Hellinger discrimination and applications, Inequalities for the Csiszar's f-divergence in Information Theory*, (Edited by S.S. Dragomir), `http://rgmia.vu.edu.au/monographs/csiszar.html`, (2000).

[11] S.S.Dragomir, V.Gluscevic, and C.E.M.Pearce, *Inequality theory and Applications,* (Edited by Y.J.Cho, J.K. Kim and S.S. Dragomir), Nova Science Publishers, Huntington, New York, (2001).

[12] E. Hellinger, Neue Berudung der Theorie der quadratischen Formen von Un endlichen Vielen Veran derlichen, *J.Reine Aug.Math.,* **136** (1909), 210-271 .

[13] H. Jeffreys, An invariant form for the prior probability in estimation problems, *Proc. Roy. Soc. London Ser.,* A **186** (1946), 453-461.

[14] S. Kullback and A.S. Leibler, *On information and sufficiency*, **22** (1951), 79-86.

[15] P. Kumar and A. Johnson, On a symmetric divergence measure and information inequalities, *J. Inequalities in Pure Appl.Math.,* **6(3)** (2005), 1-13.
[16] P. Kafka, F. Osterreicher and I.Vincze, On powers of f-divergence defining a distance,*Studia Sci.Math.Hungar.,* **26** (1991), 415-422 .
[17] F. Liese and I. Vajda, *Teubner-Texte zur Mathematik,* **95**, Leipzig, (1987).
[18] K. Pearson ,On the Criterion that a given system of deviations from the probable in the case of correlated system of variables is such that it can be reasonable supposed to have arisen from random sampling, *Phil. Mag.,* **50** (1900), 157-172.
[19] C.E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. Jour.* **27**, (1958), 623-659.
[20] R. Sibson, Information radius,*Z. Wahrs. undverw Geb.,* **14** (1969), 149-160.
[21] I.J. Taneja,Inequalities among logarithmic mean measures,2011.Available online: `http://arxiv.org/abs/1103.2580v1`.
[22] I.J. Taneja, New developments in generalized information measures, Chapter in: *Advances in Imaging and Electron Physics* (Edited by P.W. Hawkes) **91** (1995), 37-135.
[23] I.J.Taneja, Generalized symmetric divergence measures and Inequalities, *RGMIA Research Report Collection*, `http://rgmia.vu.edu.au`, **7(4)** (2004), Art. 9. Available online: `arxiv:math.ST/0501301v1` 19 Jan 2005.
[24] F. Topsϕe, Res. Rep. Collection: *RGMIA,* **2(1)** (1999), 85-98.
[25] I.J. Taneja and P. Kumar, Relative information of Csiszàr's f-divergence and information inequalities, *Information Sciences,* (2003).