ISSN No.: 2454- 2024 (online)

International Journal of Technical Research & Science

NEW DEVELOPMENTS OF INFORMATION INEQUALITIES

RAM NARESH SARASWAT
Department of Mathematics and Statistics, Manipal University Jaipur
Jaipur (Rajasthan),India, E-mail: sarswatrn@gmail.com

Abstract-Information and divergence measure are very useful and play an important role in many areas like as Sensor Networks [11], Testing the order in a Markov chain [12], Risk for binary experiments[13], Region segmentation and estimation [14] etc. In this paper we establish an upper and lower bounds of Relative arithmetic-geometric divergence, Relative J-divergence, Jenson-Shannon's divergence, Triangular discrimination, Unified Relative Jensen-Shannon and arithmetic-geometric divergence of types etc. in terms of Chi-square divergence using a new f-divergence measure and inequalities.

Key Word- Chi-square divergence, Jenson-Shannon's divergence, Triangular discrimination etc.

AMS Classification 62B-10, 94A-17,26D15

1. INTRODUCTION

Let

$$\Gamma_n = \left\{ P = (p_1, p_2, ...p_n) \middle| p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}, n \geq 2 \qquad (1.1)$$

be the set of all complete finite discrete probability distributions. There are many information and divergence measures exists in the literature of information theory and statistics. Csiszar [2] & [3] introduced a generalized measure of information using f-divergence measure given by

$$I_f(P,Q) = \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right) \qquad (1.2)$$

where $f : \mathbf{R}_+ \to \mathbf{R}_+$ is a convex function and $P,Q \in \Gamma_n$.

As in Csiszar [3], The Csiszar's f-divergence is a general class of divergence measures that includes several divergences used in measuring the distance or affinity between two probability distributions. This class is introduced by using a convex function f, defined on (0, ∞). An important property of this divergence is that many known divergences can be obtained from this measure by appropriately defining the convex function f.

There are some examples of divergence measures in the category of Csiszar's f-divergence measure. Bhattacharya divergence [1], Triangular discrimination [5], Relative J-divergence [7], Hellinger discrimination [8], Chi-square divergence [15], Relative Jensen-shannon divergence [16], Relative arithmetic-geometric divergence [17], Unified relative Jensen-Shannon and arithmetic-geometric divergence of type s [18]

2. NEW f-DIVERGENCE MEASURE AND ITS PARTICULAR CASES

In this section we shall consider some properties of a new f-divergence measure [Jain and Saraswat, 10] and its particular cases which are may be interesting in areas of information theory is given by

$$S_f(P,Q) = \sum_{i=1}^{n} q_i f\left(\frac{p_i + q_i}{2q_i}\right) \qquad (2.1)$$

Where $f : \mathbf{R}_+ \to \mathbf{R}_+$ is a convex function and $P,Q \in \Gamma_n$.

It is shown that using new f-divergence measure we derive some well-known divergence measures such as Chi-square divergence, Relative J-divergence, Jenson-Shannon's divergence, Triangular discrimination, Hellinger discrimination, Bhattacharya divergence, Unified relative Jensen-Shannon and arithmetic-geometric divergence of type's etc. in this section. An inequality of f-divergence in terms of Chi-square divergence measure is established in section 3. Using the inequality of section 3, bounds of various particular measures are found in terms of Chi-square divergence measure in section 4.

The following results are on similar lines the result presented by Csisz´ar & K¨orner [4] and Dragomir [6].

2.1 PROPOSITION

Let $f:[0,\infty)\to\mathbf{R}$ be convex and $P,Q\in\Gamma_n$ then we have the following inequality

$$S_f(P,Q)\ge f(1)$$

(2.2)

Equality holds in (2.2) iff

$$p_i=q_i \ \forall\, i=1,2,..,n \tag{2.3}$$

Proof: If f is a convex function then using well known Jensen's discrete inequality, we get

$$\sum_{i=1}^{n} q_i\, f(x_i)\ge f\left(\sum_{i=1}^{n} q_i\, x_i\right)$$

$$\text{Put } x_i=\frac{p_i+q_i}{2q_i}$$

$$\sum_{i=1}^{n} q_i\, f\left(\frac{p_i+q_i}{2q_i}\right)\ge f\left(\sum_{i=1}^{n} q_i\left(\frac{p_i+q_i}{2q_i}\right)\right)$$

$$\sum_{i=1}^{n} q_i\, f\left(\frac{p_i+q_i}{2q_i}\right)\ge f\left(\frac{\sum_{i=1}^{n} p_i+\sum_{i=1}^{n} q_i}{2\sum_{i=1}^{n} q_i}\right)\ge 0$$

$$\sum_{i=1}^{n} q_i\, f\left(\frac{p_i+q_i}{2q_i}\right)\ge f(1)$$

$$\sum_{i=1}^{n} q_i\, f\left(\frac{p_i+q_i}{2q_i}\right)=S_f(P,Q)\ge f(1)$$

$$S_f(P,Q)\ge f(1)$$

Equality holds if

$$p_i=q_i \quad ,\forall\ i=1,2,..,n$$

2.1.1 COROLLARY (Non-negativity of new f-divergence measure)

Let $f:[0,\infty)\to\mathbf{R}$ be convex and normalized, i.e.

$$f(1)=0 \tag{2.4}$$

Then for any $P,Q\in\Gamma_n$ from (2.2) of proposition 2.1 and (2.4), we have the inequality

$$S_f(P,Q)\ge 0 \tag{2.5}$$

If f is strictly convex, equality holds in (2.5) iff

$$p_i=q_i \quad \forall\ i\in[i,2,...............n] \tag{2.6}$$

and

$$S_f(P,Q)\ge 0 \ \text{ and } \ S_f(P,Q)=0 \ \text{ iff } \ P=Q \tag{2.7}$$

2.2 PROPOSITION

Let $f_1\,\&\,f_2$ are two convex functions and $g=a\,f_1+b\,f_2$ then $S_g(P,Q)=a\,S_{f_1}(P,Q)+b\,S_{f_2}(P,Q)$, where $a\,\&\,b$ are constants and $P,Q\in\Gamma_n$

Proof: Let $f:[0,\infty)\to\mathbf{R}$ be convex mapping. Let $f_1\,\&\,f_2$ are two convex functions and $g=a\,f_1+b\,f_2$ then we get

$$S_g(P,Q)=\sum_{i=1}^{n} q_i\, g\left(\frac{p_i+q_i}{2q_i}\right)=a\sum_{i=1}^{n} q_i\ f_1\left(\frac{p_i+q_i}{2q_i}\right)+b\sum_{i=1}^{n} q_i\ \ f_2\left(\frac{p_i+q_i}{2q_i}\right)$$

$$S_g(P,Q)=\sum_{i=1}^{n} q_i\, g\left(\frac{p_i+q_i}{2q_i}\right)=a\,S_{f_1}+b\,S_{f_2}$$

We now give some examples of well- known information divergence measures which are obtained from new f-divergence measure.

- If $f(t)=(t-1)^2$ then Chi-square divergence measure is given by

$$S_f(P,Q)=\frac{1}{4}\left[\sum_{i=1}^{n}\frac{p_i^2}{q_i}-1\right]=\frac{1}{4}\chi^2(P,Q) \tag{2.8}$$

- If $f(t)=-\log t$ then relative Jensen-Shannon divergence measure is given by

$$S_f(P,Q)=\sum_{i=1}^{n} q_i \log\left(\frac{2q_i}{p_i+q_i}\right)=F(Q,P) \tag{2.9}$$

- If $f(t)=t\log t$ then relative arithmetic-geometric divergence measure is given by

$$S_f(P,Q)=\sum_{i=1}^{n}\left(\frac{p_i+q_i}{2}\right)\log\left(\frac{p_i+q_i}{2q_i}\right)=G(Q,P) \tag{2.10}$$

- If $f(t)=\frac{(t-1)^2}{t},\ \forall t>0$ then Triangular discrimination is given by

$$S_f(P,Q)=\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{2(p_i+q_i)}=\frac{1}{2}\Delta(P,Q) \tag{2.11}$$

- If $f(t)=(t-1)\log t$ then Relative J-divergence measure is given by

$$S_f(P,Q)=\sum_{i=1}^{n}\left(\frac{p_i-q_i}{2}\right)\log\left(\frac{p_i+q_i}{2q_i}\right)=\frac{1}{2}J_R(P,Q) \tag{2.12}$$

- If $f(t)=(1-\sqrt{t})$ then Hellinger discrimination is given by

$$S_f(P,Q)=\left[1-B\left(\frac{P+Q}{2},Q\right)\right]=h\left(\frac{P+Q}{2},Q\right) \tag{2.13}$$

- If $$f(t)=\begin{cases}[\alpha(\alpha-1)]^{-1}\left[t^{\alpha}-1\right], & \alpha\neq 0,1\\ -\log t & if\,\alpha=0\\ t\log t & if\,\alpha=1\end{cases} \tag{2.14}$$

Then Unified relative Jensen-Shannon and Arithmetic-Geometric divergence measure of type α is given by

$$S_f(P,Q)=\Omega_\alpha(Q,P)=\begin{cases}FG_\alpha(Q,P)=[\alpha(\alpha-1)]^{-1}\left[\sum_{i=1}^{n} q_i\left(\frac{p_i+q_i}{2q_i}\right)^{\alpha}-1\right], & \alpha\neq 0,1\\ F(Q,P)=\sum_{i=1}^{n} q_i\log\left(\frac{2q_i}{p_i+q_i}\right), & \alpha=0\\ G(Q,P)=\sum_{i=1}^{n}\left(\frac{p_i+q_i}{2}\right)\log\left(\frac{p_i+q_i}{2q_i}\right), & \alpha=1\end{cases} \tag{2.15}$$

3. SOME INEQUALITIES BETWEEN NEW F-DIVERGENCE AND THE CHI-SQUARE DIVERGENCE MEASURE

The following theorem concerning inequalities among new f-divergence measure and Chi-square divergence measure holds. Its particular cases are given in Section 4.

The results are on similar lines to the result presented by Dragomir [6] and Jain & Saraswat [9].

3.1 THEOREM

Let $f:(0,\infty)\to\mathbf{R}$ is normalized mapping i.e. $f(1)=0$ and satisfy the assumptions.

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

(ii) There exist constants m, M such that

$$2m\le f''(t)\le 2M \tag{3.1}$$

If P, Q are discrete probability distributions satisfying the assumptions

$$r<\frac{1}{2}\le r_i=\frac{p_i+q_i}{2q_i}\le R,\ \forall i\in\{1,2,..........n\} \tag{3.2}$$

Then we have the following inequality

$$\frac{m}{4}\chi^2(P,Q)\le S_f(P,Q)\le\frac{M}{4}\chi^2(P,Q) \tag{3.3}$$

Proof: - Define a mapping $F_m:(0,\infty)\to\mathbf{R}, F_m(t)=f(t)-m(t-1)^2$. Then $F_m(.)$ is normalized, twice differentiable and since

$$F_m''(t)=f''(t)-2m=\left[f''(t)-2m\right]\ge 0 \tag{3.4}$$

For all $t\in(r,R)$, it follows that $F_m(.)$ is convex on (r,R). Applying non-negativity property of new f-divergence measure for $F_m(.)$ and by proposition2.2, we may state that

$$0\le S_{F_m}(P,Q)=S_f(P,Q)-mS_{(t-1)^2}(P,Q)=S_f(P,Q)-\frac{m}{4}\chi^2(P,Q)$$

$$\Rightarrow\qquad 0\le S_f(P,Q)-\frac{m}{4}\chi^2(P,Q) \tag{3.5}$$

from where the first inequality in (3.3).

Now we again Define a mapping $F_M:(0,\infty)\to\mathbf{R}, F_M(t)=M(t-1)^2-f(t)$, which is obviously normalized, twice differentiable and by (3.1), convex on (r,R). Applying non-negativity property of new f-divergence measure for $F_M(.)$ and the linearity property, we obtain the second part of (3.3) i.e.

$$0\le\frac{M}{4}\chi^2(P,Q)-S_f(P,Q) \tag{3.6}$$

from results (3.5) and (3.6) give result (3.3).

REMARK.1 If we have strict inequality ">" in (3.3) for any $t\in(r,R)$ then the mapping $F_m(.)$ and $F_M(.)$ are strictly convex and equality holds in (3.3) iff $P=Q$

REMARK.2 It is important note that f is twice differentiable on $(0,\infty)$ and $m\le f''(t)\le M<\infty,\forall t\in(0,\infty)$, then inequality (3.1) holds for any probability distributions P, Q.

4. SOME PARTICULAR CASES

Using equation (3.3) of Theorem 3.1 we able to point out the following particular cases which may be interested in Information Theory and Statistics.

The results are on similar lines to the result presented by Dragomir [6] and Jain & Saraswat [9].

4.1 PROPOSITION

Let $P,Q\in\Gamma_n$ be two probability distributions with the property that

$$r<\frac{1}{2}\le r_i=\frac{p_i+q_i}{2q_i}\le R,\ \forall i\in\{1,2,..........n\}$$

Then we have the following inequality

$$\frac{1}{4R^2}\chi^2(P,Q)\le F(Q,P)\le\frac{1}{4r^2}\chi^2(P,Q) \tag{4.1}$$

Proof:-Consider the mapping $f:(r,R)\to\mathbf{R}$.

$$f(t) = -\log t,\ f'(t) = -\frac{1}{t},\ f''(t) = \frac{1}{t^2} > 0,\ \forall t > 0$$

$f''(t) \geq 0$ and $f(1) = 0$, So function f is convex and normalized.

Define $g(t) = f''(t) = \frac{1}{t^2}$

Then obviously

$$M = \inf_{t\in[r,R]} g(t) = \frac{1}{r^2},\ m = \sup_{t\in[r,R]} g(t) = \frac{1}{R^2} \quad (4.2)$$

Also $S_f(P,Q) = F(Q,P)$ from equation (2.9)

From equation (2.9), (3.3) & (4.2) give the result (4.1).

4.2 PROPOSITION

Let $P,Q \in \Gamma_n$ be two probability distributions satisfying (3.2) then we have the following inequality

$$\frac{1}{4R}\chi^2(P,Q) \leq G(Q,P) \leq \frac{1}{4r}\chi^2(P,Q) \quad (4.3)$$

Proof:-Consider the mapping $f:(r,R) \to \mathbf{R}$.

$$f(t) = t\log t,\ f'(t) = 1+\log t,\ f''(t) = \frac{1}{t} > 0,\ \forall t > 0$$

$f''(t) \geq 0$ and $f(1) = 0$, So function f is convex and normalized.

Define $g(t) = f''(t) = \frac{1}{t}$

Then obviously

$$M = \inf_{t\in[r,R]} g(t) = \frac{1}{R},\ m = \sup_{t\in[r,R]} g(t) = \frac{1}{r} \quad (4.4)$$

Since $S_f(P,Q) = G(Q,P)$ from equation (2.10)

From equation (2.10), (3.3), & (4.4) give the result (4.3).

4.3 PROPOSITION

Let $P,Q \in \Gamma_n$ be two probability distributions satisfying (3.2) then we have the following inequality

$$\frac{1}{R^3}\chi^2(P,Q) \leq \Delta(P,Q) \leq \frac{1}{r^3}\chi^2(P,Q) \quad (4.5)$$

Proof:-Consider the mapping $f:(r,R) \to \mathbf{R}$.

$$f(t) = \frac{(t-1)^2}{t} = \left(t + \frac{1}{t} - 2\right),\ f'(t) = \left(1 - \frac{1}{t^2}\right),\ f''(t) = \frac{2}{t^3}$$

$f''(t) \geq 0$ and $f(1) = 0$, So function f is convex and normalized.

Define $g(t) = f''(t) = \frac{2}{t^3}$

Then obviously

$$M = \sup_{t\in[r,R]} g(t) = \frac{2}{r^3},\ m = \inf_{t\in[r,R]} g(t) = \frac{2}{R^3} \quad (4.6)$$

Since $S_f(P,Q) = \frac{1}{2}\Delta(P,Q)$ from equation (2.11)

From equation (2.11), (3.3) & (4.6) give the result (4.5).

4.4 PROPOSITION

Let $P,Q\in\Gamma_n$ be two probability distributions satisfying (3.2) then we have the following inequality

$$\frac{1+R}{2R^2}\chi^2(P,Q)\le J_R(P,Q)\le\frac{1+r}{2r^2}\chi^2(P,Q) \qquad (4.7)$$

Proof:-Consider the mapping $f:(r,R)\to\mathbf{R}$.

$$f(t)=(t-1)\log t,\ f'(t)=\left(1-\frac{1}{t}\right)+\log t,\ f''(t)=\frac{(1+t)}{t^2}>0,\ \forall t>0$$

$f''(t)\ge 0$ and $f(1)=0$, So function f is convex and normalized.

Define $g(t)=f''(t)=\dfrac{(1+t)}{t^2}$

$g(t)$ is monotonic deceasing $\forall t\ge 0$

Then obviously

$$M=\sup_{t\in[r,R]} g(t)=\frac{(1+r)}{r^2},\ m=\inf_{t\in[r,R]} g(t)=\frac{(1+R)}{R^2} \qquad (4.8)$$

Since $S_f(P,Q)=\dfrac{1}{2}J_R(P,Q)$ from equation (2.12)

From equation (2.12), (3.3) & (4.8) give the result (4.7).

4.5 PROPOSITION

Let $P,Q\in\Gamma_n$ be two probability distributions satisfying (3.2) then we have the following inequality

$$\frac{r^{\alpha-2}}{4}\chi^2(P,Q)\le\Omega_\alpha(Q,P)\le\frac{R^{\alpha-2}}{4}\chi^2(P,Q),\quad \alpha\ge 2 \qquad (4.9)$$

$$\frac{R^{\alpha-2}}{4}\chi^2(P,Q)\le\Omega_\alpha(Q,P)\le\frac{r^{\alpha-2}}{4}\chi^2(P,Q),\quad \alpha<2 \qquad (4.10)$$

Proof:-Consider the mapping $f:(r,R)\to\mathbf{R}$.

$$f(t)=\left[\alpha(\alpha-1)\right]^{-1}\left[t^\alpha-1\right],\ \alpha\ne 0,1$$

$$f(t)=\left[\alpha(\alpha-1)\right]^{-1}\left[t^\alpha-1\right],\ f'(t)=\left[\alpha-1\right]^{-1}t^{\alpha-1},\ f''(t)=t^{\alpha-2}>0,\ \forall t>0$$

$f''(t)\ge 0$ and $f(1)=0$, So function f is convex and normalized.

Define $g(t)=f''(t)=t^{\alpha-2}$

Then obviously

$$M=\sup_{t\in[r,R]} g(t)=R^{\alpha-2},\ m=\inf_{t\in[r,R]} g(t)=r^{\alpha-2} \qquad (4.11)$$

Since $S_f(P,Q)=\Omega_\alpha(Q,P)$ from equation (2.14)

From equation (2.14), (3.3) & (4.11) give the result (4.9) & (4.10)

4.5.1 COROLLARY

$\alpha=\dfrac{1}{2}$ for result (4.10) of proposition 4.5 and Let $P,Q\in\Gamma_n$ be two probability distributions satisfying (3.2) then we have the following inequality

$$\frac{1}{16R^{3/2}}\chi^2(P,Q)\le\left[1-B\left(\frac{P+Q}{2},Q\right)\right]\le\frac{1}{16r^{3/2}}\chi^2(P,Q) \qquad (4.12)$$

and

$$\frac{1}{16R^{3/2}}\chi^2(P,Q) \le h\left(\frac{P+Q}{2},Q\right) \le \frac{1}{16r^{3/2}}\chi^2(P,Q) \quad (4.13)$$

Proof: - Consider the mapping $f:(r,R)\to \mathbf{R}$. If $\alpha = \frac{1}{2}$ then function

$$f(t) = 4(1-t^{\frac{1}{2}}),\ f'(t) = -2t^{-\frac{1}{2}},\ f''(t) = t^{-\frac{3}{2}} > 0,\ \forall t > 0$$

$f''(t) \ge 0$ and $f(1) = 0$, So function f is convex and normalized.

$$g(t) = f''(t) = t^{-\frac{3}{2}}$$

$$M = \sup_{t\in[r,R]} g(t) = r^{-\frac{3}{2}},\ m = \inf_{t\in[r,R]} g(t) = R^{-\frac{3}{2}} \quad (4.14)$$

Since $S_f(P,Q) = 4\left[1 - B\left(\frac{P+Q}{2},Q\right)\right] = 4h\left(\frac{P+Q}{2},Q\right)$ from (2.13) (4.15)

From equation (2.13), (3.3), (4.14) & (4.15) give the result (4.12) & (4.13)

4.5.2 Corollary

For $\alpha = 0 \& 1$ of result (4.9) are already proved in propositions (4.1) and (4.2).

REFERENCES

[1]. Bhattacharya A., "Some analogues to amount of information and their uses in statistical estimation", Sankhya8 (1946) 1-14.

[2]. Csiszar I. "Information measure: A critical survey. Trans.7th prague conf. on info. Th. Statist. Decius. Funct, Random Processes" and 8th European meeting of statist Volume B. Acadmia Prague, 1978, PP-73-86.

[3]. Csiszar I., "Information-type measures of difference of probability functions and indirect observations" studia Sci.Math.hunger.2 (1961).299-318.

[4]. Csiszar I. and Korner,J. "Information Theory: Coding Theorem for Discrete Memory-Less Systems", Academic Press, New York, 1981.

[5]. Dacunha-Castella D., "Ecole d'Ete de probabilities de Saint-Flour VII-1977 Berline, Heidelberg", New-York:Springer 1978.

[6]. Dragomir S.S., "Some inequalities for (m,M)-convex mappings and applications for the Csiszar's Φ-divergence in information theory". Math. J. Ibaraki Univ. (Japan) 33 (2001), 35-50.

[7]. Dragomir S.S., V. Gluscevic and C.E.M. Pearce, "Approximation for the Csiszar f –divergence via mid-point inequalities, in inequality theory and applications- Y. J. Cho, J. K. Kim and S.S. dragomir (Eds), nova science publishers, inc., Huntington, new York, vol1, 2001, pp.139-154.

[8]. Hellinger E., "Neue Begrundung der Theorie der quadratischen Formen Von unendlichen vielen veranderlichen", J.Rein.Aug.Math,136(1909),210-271.

[9]. Jain K. C. and R. N. Saraswat , "Some bounds of Csiszar's f-divergence measure in terms of the well-known divergence measures of Information Theory" International Journal of Mathematical Sciences and Engineering Applications, 5(5), 2011.

[10]. Jain K. C. and R. N. Saraswat "A New Information Inequality and its Application in Establishing Relation among various f-Divergence Measures", Journal of Applied Mathematics, Statistics and Informatics, 8(1) (2012), 17-32.

[11]. Leandro Pardo, "Statistical Inference Based on Divergence measures", Chapman & Hall/CRC Taylor & Francis Group 2006.

[12]. Lyndon S. Hibbard, "Region segmentation using information divergence measures Medical Image Analysis" 8 (2004) 233–244.

[13]. Mark D. Reid, Robert C. Williamson, "Information, Divergence and Risk for Binary Experiments" Journal of Machine Learning Research 12 (2011) 731-817 Submitted 8/09; Revised 11/10; Published 3/11.

[14]. Men4ndez M. L. , J. A. Pardo, and L. Pardo, "Csiszar's q -divergences for testing the order in a Markov chain" Statistical Papers 42, 313--328 (2001).

[15]. Pearson K., "On the criterion that a give system of deviations from the probable in the case of correlated system of variables in such that it can be reasonable supposed to have arisen from random sampling", Phil. Mag., 50(1900),157-172.

[16]. Sibson R., Information Radius, Z, Wahrs. undverw.geb. (14) (1969), 149-160.

[17]. Taneja I.J., "New Developments in generalized information measures", Chapter in: Advances in imaging and Electron Physics, Ed. P. W. Hawkes 91 (1995),37-135.

[18]. Taneja I. J. and Pranesh Kumar, "Generalized non-symmetric divergence measures and inequalities" (2000) The Natural Science and Engineering Research Council's Discovery grant to Pranesh Kumar.