Journal of the Applied Mathematics, Statistics and Informatics (JAMSI), **7** (2011), No. 1

A SYMMETRIC INFORMATION DIVERGENCE MEASURE OF CSISZAR'S *F* DIVERGENCE CLASS

K.C. JAIN AND R. MATHUR

Abstract

Information divergence measures are often useful for comparing two probability distributions. A new divergence measure is introduced which is symmetric and belongs to the well known class of Csiszar's *f* divergence. Its properties are studied and bounds in terms of some well known divergence measures are obtained. A numerical illustration to compare this measure with some known divergence measures is carried out.

Mathematics Subject Classification 2000: 94A17, 26D15
Additional Key Words and Phrases: Csiszar's *f* divergence, symmetric divergence, information inequalities .

1. INTRODUCTION

A general class of divergence measure that includes several divergences used in measuring the distance or affinity between two probability distributions is the Csiszar's *f* divergence class.

This class is introduced by using a convex function *f* defined on $(0,\infty)$. An important property of this divergence is that many known divergences can be obtained from this measure by appropriately defining the convex function *f*. These measures have been applied in a variety of fields such as anthropology, genetics, finance, economics, analysis of contingency tables approximation of probability distribution, signal processing and pattern recognition. In this paper we have introduced a new symmetric information divergence measure which belongs to the Csiszar's *f* divergence [2,3] In section 2 we discuss Csiszar's *f* Divergence and inequalities. A New symmetric measure is obtained in section 3. In section 4 we have derived information divergence inequalities providing bounds for the new measure in terms of some well known divergence measures. In section 5 the bounds obtained in previous section are checked numerically and the suggested measure is compared with other measures. Section 6 concludes the paper.

2. CSISZAR'S *f* DIVERGENCE

Let $\Omega = \{x_1, x_2,...\}$ be a set with at least two elements and $\mathbb{P}$ the set of all probability distribution $P = (p(x) : x\in\Omega)$ For a convex function $f : (0,\infty) \to \mathbb{R}$, the *f- divergence* of the probability distributions P and Q by Csiszar, [2,3] and Ali & Silvey, [1] is defined as

$$C_f(P, Q) = \sum_{x\in\Omega} q(x) f\left(\frac{p(x)}{q(x)}\right)$$

Henceforth, for brevity we will denote $C_f(P,Q), p(x)$ and $\sum_{x\in\Omega}$ by $C(P,Q), p, q$ and Σ respectively.

Osterreicher [7] has discussed basic general properties of *f* - divergences including their axiomatic properties and some important classes. During the recent past, there has been a considerable amount of work providing different kind of bounds on the distance, information and divergence measures [9, 4, 5 and 6]. Taneja and Kumar [8] unified and generalized three theorems studied by Dragomir [4, 5, 6] which provide bounds on C (P, Q). The main result in [8] is the following theorem:

Theorem 2.1 Let $f: I \subset \mathbb{R}_+ \to \mathbb{R}$ be a mapping which is normalized, i.e., $f(1) = 0$ and suppose that

(i) *f is twice differentiable on (r,R),* $0 \le r \le 1 \le R < \infty$, (f' and f'' denote the first and second derivatives of *f*),

(ii) there exist real constants m, M such that $m < M$ *and* $m \le x^{2-s} f''(x) \le M$, $\forall x \in (r,R)$, $s \in \mathbb{R}$.

if $P,Q \in \mathbb{P}^2$ are discrete probability distributions with $0 < r \le \frac{p}{q} \le R \le \infty$, then

$$m\Phi_s(P,Q) \le C(P,Q) \le M\Phi_s(P,Q),$$

and

$$m(\eta_s(P,Q) - \Phi_s(P,Q)) \le C_\rho(P,Q) - C(P,Q) \le M(\eta_s(P,Q) - \Phi_s(P,Q)),$$

where

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

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

$$K(P,Q) = \sum p \ln\left(\frac{p}{q}\right), \tag{2.1}$$

$$C\rho(P,Q) = C_{f'}\left(\frac{p^2}{q}, P\right) - C_{f'}(P,Q) = \sum (p-q) f'\left(\frac{p}{q}\right), \tag{2.2}$$

and

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

$$= \begin{cases} (s-1)_s^{-1}\sum(p-q)\left(\frac{p}{q}\right)^{s-1}, & s \ne 1 \\ \sum(p-q)\ln\left(\frac{p}{q}\right), & s = 1 \end{cases}$$

The following information inequalities which are interesting from the information-theoretic point of view, are obtained from Theorem 2.1 and discussed in [5]

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

$$\frac{m}{2}\chi^2(P,Q) \le C(P,Q) \le \frac{M}{2}\chi^2(P,Q) \tag{2.3}$$

and

$$\frac{m}{2}\chi^2(P,Q) \leq [C_\rho(P,Q) - C(P,Q)] \leq \frac{M}{2}\chi^2(P,Q) \tag{2.4}$$

where

$$\chi^2(P,Q) = \sum \frac{(p-q)^2}{q}. \tag{2.5}$$

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

$$mK(P,Q) \leq C(P,Q) \leq MK(P,Q), \tag{2.6}$$

and

$$mK(Q,P) \leq [C_\rho(P,Q) - C(P,Q)] \leq MK(Q,P). \tag{2.7}$$

(iii) The case $s = \frac{1}{2}$ provides the information bounds in terms of the Hellinger's discrimination, h (P,Q):

$$4mh(P,Q) \leq C(P,Q) \leq 4Mh(P,Q). \tag{2.8}$$

and

$$4m\left(\frac{1}{4}\eta_{1/2}(P,Q) - h(P,Q)\right) \leq [C_\rho(P,Q) - C(P,Q)] \tag{2.9}$$

$$\leq 4M\left(\frac{1}{4}\eta_{1/2}(P,Q) - h(P,Q)\right)$$

where

$$h(P,Q) = \sum \frac{(\sqrt{p}-\sqrt{q})^2}{2} \tag{2.10}$$

(iv) For s = 0, the information bounds in terms of the Kullback- Leibler and χ^2 – divergences:

$$mK(Q,P) \leq C(P,Q) \leq MK(Q,P), \tag{2.11}$$

and

$$m[\chi^2(Q,P) - K(Q,P)] \leq [C_\rho(P,Q) - C(P,Q)] \leq M[\chi^2(Q,P) - K(Q,P)] \tag{2.12}$$

3. NEW INFORMATION DIVERGENCE MEASURE

In this section we have found a new divergence measure which belongs to the Csiszar’s family and is symmetric with respect to the two probability distributions P and Q.We have derived the new measure from the convex function f (u).We have also shown that second derivative of this function is positive for all u > 0, hence f (u) is convex. Further f (1)=0 .Thus we can say that the new measure is nonnegative and convex in the pair of probability distributions (P,Q)∈ Ω.

We consider the function $f: (0,\infty) \to \mathbb{R}$

$$f(u) = \frac{(u+1)^2(u-1)^4(u^3+1)}{u^4} \tag{3.1}$$

and thus we have the new divergence measure.

$$R^*(P,Q) = \Sigma \frac{(p+q)^2(p-q)^4(p^3+q^3)}{p^4q^4} \tag{3.2}$$

$$f'(u) = \frac{(u+1)(u-1)^3[5u^5+2u^4+u^3+2u^2+2u+4]}{u^5} \tag{3.3}$$

$$f''(u) = \frac{(u-1)^2}{u^6}[20u^7 + 16u^6 + 6u^5 + 6u^4 + 6u^3 + 6u^2 + 16u + 20]$$

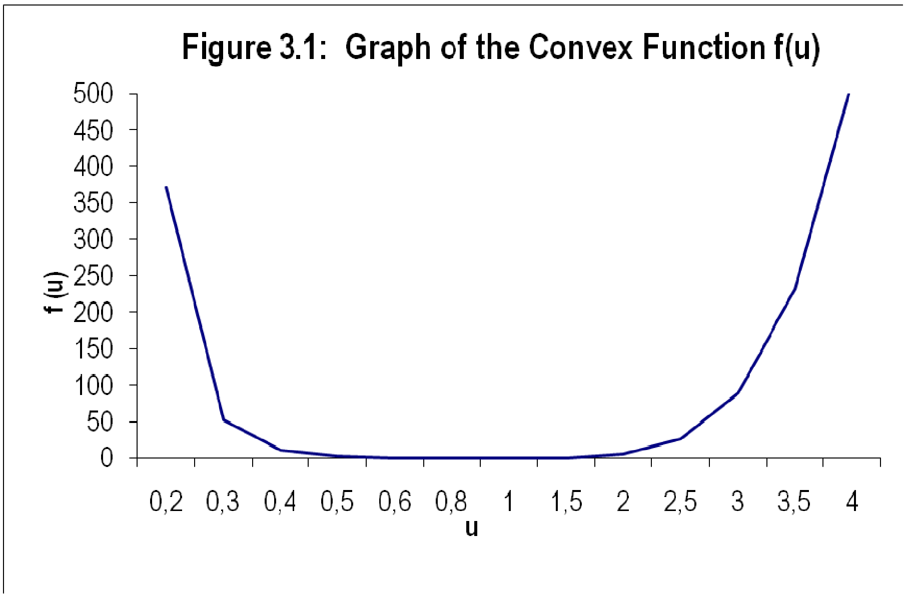
$$= 2\frac{(u-1)^2}{u^6}[10u^7 + 8u^6 + 3u^5 + 3u^4 + 3u^3 + 3u^2 + 8u + 10] \tag{3.4}$$

$$= 2\frac{(u-1)^2}{u^6}T(u)$$

Where $T(u) = [10u^7 + 8u^6 + 3u^5 + 3u^4 + 3u^3 + 3u^2 + 8u + 10]$

It follows that $f''(u) > 0$ for all $u > 0$

Hence $f(u)$ is convex for all $u > 0$.



Further $f(1) = 0$ thus we can say that the measure is non negative and convex in the pair of probability distribution $(P, Q) \in \Omega$

Its corresponding non-symmetric function is given by

$$f(u) = \frac{(u+1)^2(u-1)^4}{u}$$

$$f''(u) = \frac{(u-1)^2}{u^3}[20u^4 + 3u^3 + 2u^2 + 2u + 2]$$

which is convex for $u > 0$

4. BOUNDS FOR R*(P, Q)

In this section we will derive information divergence inequalities providing bounds for R* (P,Q) in terms of some well known divergence measures, using the inequalities discussed in section 2.

Proposition (4.1)

Let χ^2 (P,Q,) and R* (P,Q) be defined as in (2.5) and (3.2) respectively. For P, Q $\in \mathbb{P}^2$ *and* $0< r \leq \frac{p}{q} \leq R < \infty$ we have

$$\frac{(R-1)^2}{R^6} T(R)\chi^2(P,Q) \leq R^*(P,Q) \leq \frac{(r-1)^2}{r^6} T(r)\chi^2(P,Q) \tag{4.1}$$

and

$$\frac{(R-1)^2}{R^6} T(R)\chi^2(P,Q) \leq [R_p^*(P,Q) - R^*(P,Q)] \leq \frac{(r-1)^2}{r^6} T(r)\chi^2(P,Q) \tag{4.2}$$

where $R_p^*(P,Q) = (p-q)f'\left(\frac{p}{q}\right)$

$$= \Sigma \frac{(p-q)^4 (p+q)(5p^5+2p^4q+p^3q^2+2p^2q^3+2pq^4+4q^5}{p^5q^4} \tag{4.3}$$

and

T(R)= $[10R^7 + 8R^6 + 3R^5 + 3R^4 + 3R^3 + 3R^2 + 8R + 10]$

T(r)= $[10r^7 + 3r^6 + 3r^5 + 3r^4 + 3r^3 + 3r^2 + 8r + 10]$

Proof. From the function f(u) in (3.1) we have

$$f'(u) = \frac{(u+1)(u-1)^3[5u^5 + 2u^4 + u^3 + 2u^2 + 2u + 4]}{u^5}$$

And thus the divergence measure:

$$R_p^*(P,Q) = \Sigma(p-q)f'\left(\frac{p}{q}\right)$$

$$= \Sigma \frac{(p-q)^4 (p+q)(5p^5 + 2p^4q + p^3q^2 + 2p^2q^3 + 2pq^4 + 4q^5)}{p^5q^4}$$

Further $f''(u) = \frac{2(u-1)^2}{u^6} T(u)$

Now if u $\in$ [a,b] $\subset$ (0,∞) then

$$\frac{2(b-1)^2}{b^6}[10b^7+8b^6+3b^5+3b^4+3b^3+3b^2+8b+10] \leq f''(u) \leq \frac{2(a-1)^2}{a^6}[10a^7+8a^6+3a^5+3a^4+3a^3+3a^2+8b+10]$$

and accordingly

$$\frac{2(R-1)^2}{R^6}T(R) \leq f''(u) \leq \frac{2(r-1)^2}{r^6}T(r) \tag{4.4}$$

where r and R are defined above.

Thus in view of (2.3) , (2.4) and (4.4) we get inequalities (4.1) and (4.2) respectively

The information bounds in terms of Kullback - Leibler divergence K (P, Q) defined in (2.1) are as follows

Proposition (4.2)

Let K (P, Q), R* (P, Q) and R_p*(P, Q) be defined as (2.1), (3.2) and (4.3) respectively.

For P, Q, $\in \mathbb{P}^2$ and $0 < r < \frac{p}{q} \leq R < \infty$ we have

$$\frac{2(R-1)^2}{R^5}T(R)\,K(P,Q) \leq R^*(P,Q) \leq \frac{2(r-1)^2}{r^5}T(r)\,K(P,Q)$$

and (4.5)

$$\frac{2(R-1)^2}{R^5}T(R)K(P,Q) \leq [R_p^*(P,Q) - [R^*(P,Q)] \leq \frac{2(r-1)^2}{r^5}T(r)K(P,Q) \tag{4.6}$$

Proof. We have $f''(u) = \frac{2(u-1)^2}{u^6}T(u)$

Let the function g= [r, R] $\to \mathbb{R}$ be such that

$$g(u) = u\,f''(u) = \frac{2(u-1)^2}{u^5}T(u)$$

Then

$$\inf_{u\in(r,R)} g(u) = \frac{2(R-1)^2}{R^5}T(R) \tag{4.7}$$

and

$$\sup g(u) = \frac{2(r-1)^2}{r^5}T(r) \tag{4.8}$$

The inequalities (4.5) and (4.6) follow from (2.6) and (2.7) using (4.7) and (4.8)

The following proposition provides the information bounds in terms of the Hellinger's discrimination h (P,Q) and $\eta_{½}$ (P,Q).

Proposition 4.3

Let $\eta_{1/2}$ (P,Q), h (P,Q), R* (P,Q) and R_ρ* (P,Q) be defined as in (2.2), (2.9)

(3.2) and (4.3) respectively. For P, Q ∈ $\mathbb{P}^2$ and and $0 < r < \frac{p}{q} \leq R < \infty$

$$\frac{8(r-1)^2}{r^{9/2}}\, T(r)\, h(P,Q) \leq R^*(P,Q) \leq \frac{8(R-1)^2}{R^{\frac{9}{2}}}\, T(R)\, h(P,Q) \tag{4.9}$$

And

$$\frac{8(r-1)^2 T(r)}{r^{\frac{9}{2}}}\left[\frac{1}{4}\eta_{1/2}(P,Q) - h(P,Q)\right] \leq \left[R_\rho^*(P,Q) - R^*(P,Q)\right] \leq \frac{8(R-1)^2}{R^{\frac{9}{2}}}\, T(R)\left[\frac{1}{4}\eta_{1/2}(P,Q) - h(P,Q)\right] \tag{4.10}$$

Let the function g = [r, R] $\rightarrow \mathbb{R}$ be such that

$$g(u) = u^{\frac{3}{2}} f''(u) = \frac{2(u-1)^2}{u^{\frac{9}{2}}}\, T(u)$$

$$\inf_{u \in (r,R)} g(u) = \frac{2(r-1)^2}{r^{\frac{9}{2}}}\, T(r) \tag{4.11}$$

$$\sup_{u \in (r,R)} g(u) = \frac{2(R-1)^2}{R^{\frac{9}{2}}}\, T(R) \tag{4.12}$$

Thus the inequalities (4.9) and (4.10) are established using (2.8), (2.9), (4.11) and (4.12)
Next follows the information bounds in terms of Kullback – Leibler and χ^2 divergences.

Proposition 4.4

Let K (P, Q), χ^2 (P, Q), R* (P, Q) and $R_\rho^*(P,Q)$ be defined as in (2.1), (2.4), (3.4) and (4.3) respectively.

For P, Q ∈ $\mathbb{P}^2$ and $0 < r < \frac{p}{q} \leq R < \infty$ we have

$$\frac{2(r-1)^2}{r^4}\, T(r)\, K(Q,P) \leq R^*(P,Q) \leq \frac{2(R-1)^2}{R^4}\, T(R)\, K(Q,P) \tag{4.13}$$

And

$$\frac{2(r-1)^2}{r^4}T(r)\,[\chi^2(Q,P) - K(Q,P)] \leq [R_p^*(P,Q) - R^*(P,Q)]$$
$$\leq \frac{2(R-1)^2}{R^4}T(R)[\chi^2(Q,P) - K(Q,P)] \qquad (4.14)$$

Proof. We have $f''(u) = \frac{2(u-1)^2}{u^6}\,T(u)$

$$g(u) = u^2 f''(u) = \frac{2(u-1)^2}{u^4}\,T(u)$$

then $\inf_{u\in(r,R)} g(u) = \frac{2(r-1)^2}{r^4}\,T(r) \qquad (4.15)$

and

$$\sup_{u\in(r,R)} g(u) = \frac{2(R-1)^2}{R^4}\,T(R) \qquad (4.16)$$

Thus (4.13) and (4.14) follows from (2.11) and (2.12) using (4.15) and (4.16)

5. NUMERICAL ILLUSTRATION

In this section we will numerically verify the bounds obtained in the previous section and also compare the new measure with some known divergence measures. Here $J(P,Q) = K(P,Q) + K(Q,P) = \Sigma(p-q)\ln\frac{p}{q}$ is the Kullback-Leibler divergence

and $\Psi(P,Q) = \chi^2(P,Q) + \chi^2(Q,P) = \Sigma\frac{(p-q)^2(p+q)}{pq}$ is the Symmetric Chi-square Divergence.

Ex. 5.1 (Symmetrical):

Let P be the binomial probability distribution for the random valuable X with parameter (n=8 p=0.5) and Q its approximated normal probability distribution.

Table 5.1 Binomial Probability Distribution (n=8 p=0.5)

X	0	1	2	3	4	5	6	7	8
p (x)	.004	.031	.109	.219	.274	.219	.109	.031	.004
q (x)	.005	.030	.104	.220	.282	.220	.104	.030	.005

$\frac{p(x)}{q(x)}$	.774	1.042	1.0503	.997	.968	.997	1.0503	1.042	.774

The measure R* (P, Q), Ψ (P,Q), χ^2 (P,Q), J(P,Q) are

R* (P,Q) = 0.00020185, χ^2 (P,Q) = 0.00145837, Ψ (P,Q) = 0.00305063, J (P,Q) = 0.0015184

It is noted that

$$r(0.77419933) \le \frac{p}{q} \le R(1.050330018)$$

The lower and upper bounds for R* (P, Q)

Lower bound =

$$\frac{(R-1)^2}{R^6}[10R^7+8R^6+3R^5+3R^4+3R^3+3R^2+8R+10]\,\chi^2(P,Q) = 0.00015764$$

Upper bound =

$$\frac{(r-1)^2}{r^6}[10r^7+8r^6+3r^5+3r^4+3r^3+3r^2+8r+10]\,\chi^2(P,Q) = 0.00853097$$

thus 0.00015764 < R* (P, Q) = 0.000201854 < 0.00853097

width of the interval is 0.00837335.

Example 5.2 (Asymmetrical)

Let P be the binomial probability distribution for the random variable X with parameter n = 8 p = 0.4 and Q its approximated normal probability distribution. Then

Table 2 Binomial Probability Distribution (n=8, p=0.4)

X	0	1	2	3	4	5	6	7	8
p (x)	.017	.090	.209	.279	.232	.124	.041	.008	.001
q (x)	.020	.082	.198	.285	.244	.124	.037	.007	.0007
$\frac{p(x)}{q(x)}$	.850	1.102	1.056	.979	.952	1.001	1.097	1.194	1.401

From the above data measures R* (P, Q) ,χ^2 (P,Q), Ψ (P,Q) and J (P,Q) are calculated

$R^*(P, Q) = 0.04883318$, $\Psi(P,Q) = 0.00657063$

$\chi^2(P,Q) = 0.00333883$, $J(P,Q) = 0.0032778$

$$r(0.849782156) \leq \frac{p}{q} \leq R\,(1.40121965)$$

the lower and upper bounds for R* (P, Q) is

Lower bound =

$$\frac{(R-1)^2}{R^6}\left[10R^7 + 8R^6 + 3R^5 + 3R^4 + 3R^3 + 3R^2 + 8R + 10\right]\chi^2(P,Q)$$
$$= 0.01590959$$

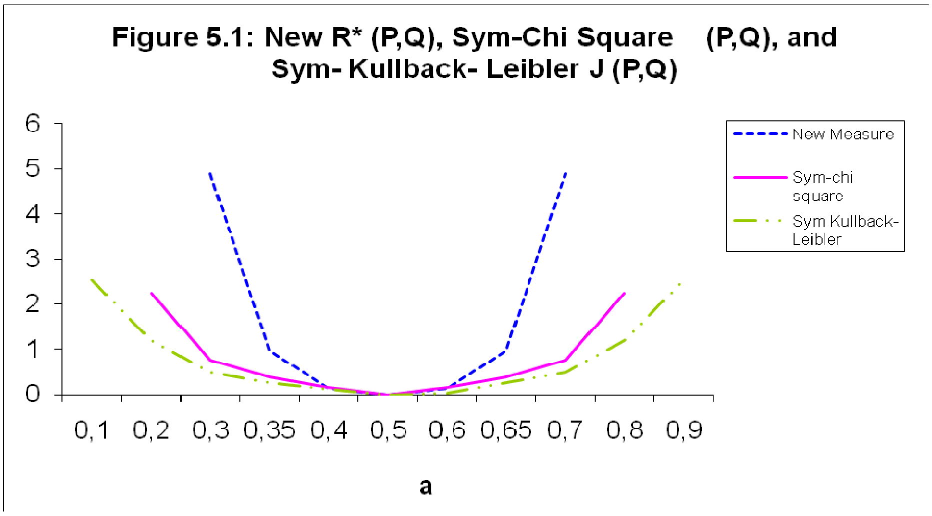
Upper bound =

$$\frac{(r-1)^2}{r^6}\left[10r^7 + 8r^6 + 3r^5 + 3r^4 + 3r^3 + 3r^2 + 8r + 10\right]\chi^2(P,Q)$$
$$= 0.09986605$$

Thus $0.01590959 < R^*(P, Q) = 0.04883318 < 0.09986605$

The width of the interval is 0.08395646.

The following figure shows the behavior of R* (P, Q) [New measure], Ψ (P, Q) [Symmetric Chi-square divergence] and J (P, Q) [Symmetric Kullback-Leibler divergence] we have considered p = (a, 1-a) and q = (1-a, a) a ∈ (0, 1). It is clear from the figure 5.1 that measure R* (P, Q) and Ψ (P, Q) have a steeper slope than J (P, Q).



6. CONCLUDING REMARKS

The Csiszar's *f*-divergence is a general class of divergence measures, which includes several divergences used in measuring the distance or affinity between two probability distribution. This class is introduced by using a convex function *f*, defined on $(0,\infty)$. An important property of this divergence is that many known divergence can be obtained by appropriately defining the convex function *f*. Non-parametric measures for the Csiszar's *f*-divergences are also available. We have introduced a new symmetric divergence measure by considering a convex function and have investigated its properties. Further we have established its bounds in terms of known divergence measures.We have also compared the new measure with some well known divergence measures.

7. REFERENCES

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

[2] Csiszar I., "Information measures: A critical survey", *In Trans. Seventh Prague Conf. On Information Theory*, A, Academia, Prague (1967), pp. 73-86.

[3] Csiszar I., "Information-type measure of difference of probability distributions and indirect observations", *Studia Sci. Math. Hungar*. **2**(1974), 299-318.

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

[5] Dragomir S.S., "Some inequalities for the Csiszar's ϕ divergence", *Inequalities for the Csizar's f-divergence in Information Theory"*, (Edited - S.S. Dragomir),

http://rgmia.vu.edu.au/monographs/ciszar.htm' (2000).

[6] Dragomir S.S., "Upper and lower bounds for Csiszar's *f*-divergence in terms of the Kullback-Leibler distance and applications", *Inequalities for the Csiszar's f-divergence in Information Theory*, (Edited) by S.S. Dragomir), http://rgmia.vu.edu.au/monographs/ciszar.htm' (2000)

[7] Österreircher F., "Csiszar's *f*-divergences-Basic properties", *Res. Report Collection http://rgmia.vu.edu.au/monographs/csiszar.htm'* (2002).

[8] Taneja I.J. and Kumar P., "Relative Information of type-s, Csiszar's *f*-divergence and information inequalities", *Information Sciences*, (2003).

[9] Topsoe F., "Res. Rep. Collection", RGMIA **2 (1**) (1999) 85-98.

K.C. JAIN
Deparment of Mathematics,
Malaviya National Institute of Technology,
Jaipur-302017, INDIA
E-mail : jainkc_2003@yahoo.co.in
Email : kantesh_g@indiatimes.com

R. MATHUR (Corresponding Author)
Deparment of Mathematics,
Malaviya National Institute of Technology,
Jaipur-302017, INDIA
E-mail : ruchimathur99@yahoo.in

Received September 2009