

ON A NEW NON- SYMMETRIC DIVERGENCE MEASURE

Dr. Praphull Chhabra

Communicated by P. K. Banerji

MSC 2010 Classifications: Primary 94A17; Secondary 26D15.

Keywords and phrases: Csiszar's divergence measure, New divergence measure, Bounds, Metric space, Comparison graph.

Abstract. New divergence measure of Csiszar's class has been introduced. Bounds of this divergence are evaluated in terms of other well known symmetric and non- symmetric divergences. Metric nature has been discussed as well.

1 Introduction

Divergence measures are basically measures of distance between two probability distributions or compare two probability distributions. Divergence measure must increase as probability distributions move apart.

Divergence measures have been demonstrated very useful in a variety of disciplines such as Bayesian model validation (1996) [45], quantum information theory (2008, 2000) [29, 31], model validation (1987) [3], robust detection (1980) [35], economics and political science (1972, 1967) [43, 44], biology (1975) [34], analysis of contingency tables (1978) [14], approximation of probability distributions (1968, 1980) [8, 26], signal processing (1967, 1967) [24, 25], pattern recognition (1978, 1979, 1973, 1990) [2, 5, 7, 23], color image segmentation (2010) [30], 3D image segmentation and word alignment (2006) [42], cost- sensitive classification for medical diagnosis (2009) [37], magnetic resonance image analysis (2010) [46] etc.

Also we can use divergence measures in fuzzy mathematics as fuzzy directed divergences and fuzzy entropies (2010, 2004, 2012) [1, 16, 22], which are very useful to find the amount of average ambiguity or difficulty in making a decision whether an element belongs to a set or not. Fuzzy information measures have recently found applications to fuzzy aircraft control, fuzzy traffic control, engineering, medicines, computer science, management and decision making etc. Divergence measures are also very useful to find the utility of an event (2010, 1986) [4, 39], i.e., an event is how much useful compare to other event.

Let $\Gamma_n = \{P = (p_1, p_2, p_3, \dots, p_n) : p_i > 0, \sum_{i=1}^n p_i = 1\}$, $n \geq 2$ be the set of all complete finite discrete probability distributions. The restriction here to discrete distributions is only for convenience, similar results hold for continuous distributions as well. If we take $p_i \geq 0$ for some $i = 1, 2, 3, \dots, n$, then we have to suppose that $0f(0) = 0f\left(\frac{0}{0}\right) = 0$.

Csiszar's f - divergence (1974, 1967)[9, 10] is widely used due to its compact nature, which is given by

$$C_f(P, Q) = \sum_{i=1}^n q_i f\left(\frac{p_i}{q_i}\right), \quad (1.1)$$

where $f : (0, \infty) \rightarrow R$ (set of real no.) is real, continuous, and convex function and $P = (p_1, p_2, \dots, p_n)$, $Q = (q_1, q_2, \dots, q_n) \in \Gamma_n$, where p_i and q_i are probabilities.

$C_f(P, Q)$ is a natural distance measure from a true probability distribution P to an arbitrary probability distribution Q . Typically P represents observations or a precise calculated probability distribution, whereas Q represents a model, a description or an approximation of P .

Definition 1.1. Convex function: A function $f(t)$ is said to be convex over an interval (a, b) if for every $t_1, t_2 \in (a, b)$ and $0 \leq \lambda \leq 1$, we have

$$f[\lambda t_1 + (1 - \lambda)t_2] \leq \lambda f(t_1) + (1 - \lambda)f(t_2),$$

and said to be strictly convex if equality does not hold only if $\lambda \neq 0$ or $\lambda \neq 1$.

Geometrically, it means that if A, B, C are three distinct points on the graph of convex function f with B between A and C , then B is on or below chord AC .

Definition 1.2. Jensen Inequality: Let $f : I \subset \mathbb{R} \rightarrow \mathbb{R}$ be differentiable convex on I^0 (I^0 is the interior of the interval I), $t_i \in I^0$, $\lambda_i > 0 \forall i = 1, 2, \dots, n$ and $\sum_{i=1}^n \lambda_i = 1$, then we have the following inequality.

$$f\left(\sum_{i=1}^n \lambda_i t_i\right) \leq \sum_{i=1}^n \lambda_i f(t_i). \quad (1.2)$$

If function is concave, then Jensen's inequality will be reversed.

Corollary 1.3. After replacing λ_i with q_i as $\sum_{i=1}^n q_i = 1$ and t_i with $\frac{p_i}{q_i}$ in (1.2) for each $i = 1, \dots, n$ by assuming that the function is normalized, i.e., $f(1) = 0$, we get

$$f(1) \leq \sum_{i=1}^n q_i f\left(\frac{p_i}{q_i}\right), \text{ i.e., } C_f(P, Q) \geq 0. \quad (1.3)$$

The following theorem is well known in literature (1967) [10].

Theorem 1.4. If the function f is convex and normalized, i.e., $f''(t) \geq 0 \forall t > 0$ and $f(1) = 0$ respectively, then $C_f(P, Q)$ and its adjoint $C_f(Q, P)$ are both non-negative and convex in the pair of probability distribution $(P, Q) \in \Gamma_n \times \Gamma_n$.

The following theorem (2005) [41] is being used in this article for further calculation of bounds. This theorem relates two f -divergence measures.

Theorem 1.5. Let $f_1, f_2 : I \subset \mathbb{R}^+ \rightarrow \mathbb{R}$ be two convex differentiable and normalized functions, i.e., $f_1''(t), f_2''(t) \geq 0 \forall t > 0$ and $f_1(1) = f_2(1) = 0$ respectively and suppose the following assumptions.

- (i) f_1 and f_2 are twice differentiable on (α, β) , $0 < \alpha \leq 1 \leq \beta < \infty$ with $\alpha \neq \beta$.
- (ii) There exists the real constants m, M such that $m < M$ and

$$m \leq \frac{f_1''(t)}{f_2''(t)} \leq M, f_2''(t) \neq 0 \forall t \in (\alpha, \beta). \quad (1.4)$$

If $P, Q \in \Gamma_n$ is such that $0 < \alpha \leq \frac{p_i}{q_i} \leq \beta < \infty \forall i = 1, 2, 3, \dots, n$, then we have the following inequalities

$$mC_{f_2}(P, Q) \leq C_{f_1}(P, Q) \leq MC_{f_2}(P, Q), \quad (1.5)$$

where $C_f(P, Q)$ is given by (1.1).

2 New Divergence Measure

In this section, we introduce a new divergence measure of Csiszar's class and define the properties.

Let $f : (0, \infty) \rightarrow \mathbb{R}$ be a real differentiable mapping, which is defined as

$$f(t) = f_1(t) = e^t(t^2 - 1), \forall t \in (0, \infty), \quad (2.1)$$

$$f_1'(t) = e^t(t^2 + 2t - 1)$$

and

$$f_1''(t) = e^t(t^2 + 4t + 1). \quad (2.2)$$

We can check that the function $f_1(t)$ strictly convex and normalized because $f_1''(t) > 0 \forall t \in (0, \infty)$ and $f_1(1) = 0$ respectively.

After putting this convex function in (1.1), we obtain

$$C_{f_1}(P, Q) = C_*(P, Q) = \sum_{i=1}^n \frac{e^{\frac{p_i}{q_i}}(p_i^2 - q_i^2)}{q_i}. \quad (2.3)$$

In view of corollary (1.3) and theorem (1.4), we see that $C_*(P, Q)$ is positive and convex for the pair of probability distribution $(P, Q) \in \Gamma_n \times \Gamma_n$ and equal to zero (Non- degeneracy) or attains its minimum value when $p_i = q_i$, i.e., when probability distributions are parallel to each other. It will be maximum when probability distributions are perpendicular to each other. We can also see that $C_*(P, Q)$ is non- symmetric divergence w.r.t. P and Q because $C_*(P, Q) \neq C_*(Q, P)$.

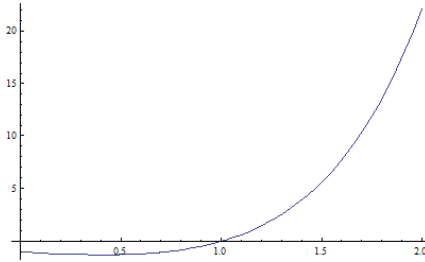


Figure 1. Convex function $f_1(t)$

3 Upper and Lower Bounds

To estimate the new exponential divergence $C_*(P, Q)$, it would be very interesting to establish some upper and lower bounds. So in this section, we obtain bounds of the divergence measure (2.3) in terms of other well known divergence measures.

I. With symmetric divergence measures:

Proposition 3.1. *Let $P, Q \in \Gamma_n$ and $0 < \alpha \leq \frac{p_i}{q_i} \leq \beta < \infty$, then we have*

$$(3.1) \quad \frac{e^\alpha (\alpha + 1)^3 (\alpha^2 + 4\alpha + 1)}{8} \Delta(P, Q) \leq C_*(P, Q) \leq \frac{e^\beta (\beta + 1)^3 (\beta^2 + 4\beta + 1)}{8} \Delta(P, Q),$$

where $\Delta(P, Q)$ is given by (3.4).

Proof. Let us consider

$$f_2(t) = \frac{(t - 1)^2}{t + 1}, t \in (0, \infty) \tag{3.2}$$

and

$$f_2'(t) = \frac{(t - 1)(t + 3)}{(t + 1)^2},$$

$$f_2''(t) = \frac{8}{(t + 1)^3}. \tag{3.3}$$

Since $f_2''(t) > 0 \forall t > 0$ and $f_2(1) = 0$, so $f_2(t)$ is strictly convex and normalized function respectively. By putting $f_2(t)$ in (1.1), we get

$$C_{f_2}(P, Q) = \sum_{i=1}^n \frac{(p_i - q_i)^2}{p_i + q_i} = \Delta(P, Q), \tag{3.4}$$

where $\Delta(P, Q)$ is called the Triangular discrimination (1978) [11].

Now, let

$$g(t) = \frac{f_1''(t)}{f_2''(t)} = \frac{e^t (t^2 + 4t + 1) (t + 1)^3}{8},$$

where $f_1''(t)$ and $f_2''(t)$ are given by (2.2) and (3.3) respectively and

$$g'(t) = \frac{e^t (t + 1)^2 (t^3 + 10t^2 + 23t + 8)}{8}.$$

It is clear that $g'(t) > 0$ for $t > 0$, therefore $g(t)$ is strictly increasing function in interval $(0, \infty)$. So

$$m = \inf_{t \in (\alpha, \beta)} g(t) = g(\alpha) = \frac{e^\alpha (\alpha + 1)^3 (\alpha^2 + 4\alpha + 1)}{8} \quad (3.5)$$

and

$$M = \sup_{t \in (\alpha, \beta)} g(t) = g(\beta) = \frac{e^\beta (\beta + 1)^3 (\beta^2 + 4\beta + 1)}{8}. \quad (3.6)$$

The result (3.1) is obtained by using (2.3), (3.4), (3.5), and (3.6) in inequalities (1.5).

Proposition 3.2. Let $P, Q \in \Gamma_n$ and $0 < \alpha \leq \frac{p_i}{q_i} \leq \beta < \infty$, then we have

$$\frac{e^\alpha \alpha^2 (\alpha^2 + 4\alpha + 1)}{1 + \alpha} J(P, Q) \leq C_*(P, Q) \leq \frac{e^\beta \beta^2 (\beta^2 + 4\beta + 1)}{1 + \beta} J(P, Q), \quad (3.7)$$

where $J(P, Q)$ is given by (3.10).

Proof. Let us consider

$$f_2(t) = (t - 1) \log t, t \in (0, \infty) \quad (3.8)$$

and

$$f_2'(t) = \frac{t - 1}{t} + \log t, \quad (3.9)$$

$$f_2''(t) = \frac{1 + t}{t^2}.$$

Since $f_2''(t) > 0 \forall t > 0$ and $f_2(1) = 0$, so $f_2(t)$ is strictly convex and normalized function respectively. By putting $f_2(t)$ in (1.1), we get

$$C_{f_2}(P, Q) = \sum_{i=1}^n (p_i - q_i) \log \frac{p_i}{q_i} = J(P, Q), \quad (3.10)$$

where $J(P, Q)$ is called the J- divergence or Jeffrey- Kullback divergence (1946, 1951) [21, 27]. Now, let

$$g(t) = \frac{f_1''(t)}{f_2''(t)} = \frac{e^{tt^2} (t^2 + 4t + 1)}{1 + t},$$

where $f_1''(t)$ and $f_2''(t)$ are given by (2.2) and (3.9) respectively and

$$g'(t) = \frac{e^{tt} (t^4 + 8t^3 + 17t^2 + 14t + 2)}{(1 + t)^2}.$$

It is clear that $g'(t) > 0$ for $t > 0$, therefore $g(t)$ is strictly increasing function in interval $(0, \infty)$. So

$$m = \inf_{t \in (\alpha, \beta)} g(t) = g(\alpha) = \frac{e^\alpha \alpha^2 (\alpha^2 + 4\alpha + 1)}{1 + \alpha} \quad (3.11)$$

and

$$M = \sup_{t \in (\alpha, \beta)} g(t) = g(\beta) = \frac{e^\beta \beta^2 (\beta^2 + 4\beta + 1)}{1 + \beta}. \quad (3.12)$$

The result (3.7) is obtained by using (2.3), (3.10), (3.11), and (3.12) in inequalities (1.5).

Proposition 3.3. Let $P, Q \in \Gamma_n$ and $0 < \alpha \leq \frac{p_i}{q_i} \leq \beta < \infty$, then we have

$$\frac{e^\alpha \alpha^3 (\alpha^2 + 4\alpha + 1)}{2(\alpha^3 + 1)} \psi(P, Q) \leq C_*(P, Q) \leq \frac{e^\beta \beta^3 (\beta^2 + 4\beta + 1)}{2(\beta^3 + 1)} \psi(P, Q), \quad (3.13)$$

where $\psi(P, Q)$ is given by (3.16).

Proof. Let us consider

$$f_2(t) = \frac{(t-1)^2(t+1)}{t}, t \in (0, \infty) \tag{3.14}$$

and

$$\begin{aligned} f_2'(t) &= 2t - \frac{1}{t^2} - 1, \\ f_2''(t) &= \frac{2(1+t^3)}{t^3}. \end{aligned} \tag{3.15}$$

Since $f_2''(t) > 0 \forall t > 0$ and $f_2(1) = 0$, so $f_2(t)$ is strictly convex and normalized function respectively. By putting $f_2(t)$ in (1.1), we get

$$C_{f_2}(P, Q) = \sum_{i=1}^n \frac{(p_i - q_i)^2 (p_i + q_i)}{p_i q_i} = \psi(P, Q), \tag{3.16}$$

where $\psi(P, Q)$ is called the Symmetric chi- square divergence (2000) [13]. Now, let

$$g(t) = \frac{f_1''(t)}{f_2''(t)} = \frac{e^t t^3 (t^2 + 4t + 1)}{2(t^3 + 1)},$$

where $f_1''(t)$ and $f_2''(t)$ are given by (2.2) and (3.15) respectively and

$$g'(t) = \frac{e^t t^2 (t^6 + 6t^5 + 5t^4 + t^3 + 9t^2 + 17t + 3)}{2(t^3 + 1)^2}.$$

It is clear that $g'(t) > 0$ for $t > 0$, therefore $g(t)$ is strictly increasing function in interval $(0, \infty)$. So

$$m = \inf_{t \in (\alpha, \beta)} g(t) = g(\alpha) = \frac{e^\alpha \alpha^3 (\alpha^2 + 4\alpha + 1)}{2(\alpha^3 + 1)} \tag{3.17}$$

and

$$M = \sup_{t \in (\alpha, \beta)} g(t) = g(\beta) = \frac{e^\beta \beta^3 (\beta^2 + 4\beta + 1)}{2(\beta^3 + 1)}. \tag{3.18}$$

The result (3.13) is obtained by using (2.3), (3.16), (3.17), and (3.18) in inequalities (1.5). In a similar procedure, we obtain the bounds of $C_*(P, Q)$ with the other well known symmetric divergence measures. The results are as follows.

Proposition 3.4. If $f_2(t) = \frac{t}{2} \log t + (\frac{t+1}{2}) \log \frac{2}{t+1}$, then we have

$$2e^\alpha \alpha (1 + \alpha) (\alpha^2 + 4\alpha + 1) I(P, Q) \leq C_*(P, Q) \leq 2e^\beta \beta (1 + \beta) (\beta^2 + 4\beta + 1) I(P, Q), \tag{3.19}$$

where

$$I(P, Q) = \frac{1}{2} \left[\sum_{i=1}^n p_i \log \frac{2p_i}{p_i + q_i} + \sum_{i=1}^n q_i \log \frac{2q_i}{p_i + q_i} \right] \tag{3.20}$$

is the Jensen- Shannon divergence or Information radius (1982, 1969) [6, 38].

Proposition 3.5. If $f_2(t) = \frac{(1-\sqrt{t})^2}{2}$, then we have

$$4e^\alpha (\alpha^2 + 4\alpha + 1) \alpha^{\frac{3}{2}} h(P, Q) \leq C_*(P, Q) \leq 4e^\beta (\beta^2 + 4\beta + 1) \beta^{\frac{3}{2}} h(P, Q), \tag{3.21}$$

where

$$h(P, Q) = \sum_{i=1}^n \frac{(\sqrt{p_i} - \sqrt{q_i})^2}{2} \tag{3.22}$$

is the Hellinger discrimination or Kolmogorov's divergence (1909) [15].

Proposition 3.6. If $f_2(t) = \frac{(t-1)^2}{\sqrt{t}}$, then we have

$$\frac{4e^\alpha (\alpha^2 + 4\alpha + 1) \alpha^{\frac{5}{2}}}{3\alpha^2 + 2\alpha + 3} E(P, Q) \leq C_*(P, Q) \leq \frac{4e^\beta (\beta^2 + 4\beta + 1) \beta^{\frac{5}{2}}}{3\beta^2 + 2\beta + 3} E(P, Q), \quad (3.23)$$

where

$$E(P, Q) = \sum_{i=1}^n \frac{(p_i - q_i)^2}{\sqrt{p_i q_i}} \quad (3.24)$$

is the Jain- Srivastava divergence (2007)[20].

Proposition 3.7. If $f_2(t) = \left(\frac{t+1}{2}\right) \log \frac{t+1}{2\sqrt{t}}$, then we have

$$\frac{4e^\alpha (\alpha^2 + 4\alpha + 1) \alpha^2 (1 + \alpha)}{\alpha^2 + 1} T(P, Q) \leq C_*(P, Q) \leq \frac{4e^\beta (\beta^2 + 4\beta + 1) \beta^2 (1 + \beta)}{\beta^2 + 1} T(P, Q), \quad (3.25)$$

where

$$T(P, Q) = \sum_{i=1}^n \left(\frac{p_i + q_i}{2} \right) \log \frac{p_i + q_i}{2\sqrt{p_i q_i}} \quad (3.26)$$

is the Arithmetic- Geometric mean divergence (1995) [40].

Proposition 3.8. If $f_2(t) = \frac{(t^2-1)^2}{2t^{\frac{3}{2}}}$, then we have

$$\frac{8e^\alpha \alpha^{\frac{7}{2}} (\alpha^2 + 4\alpha + 1)}{15\alpha^4 + 2\alpha^2 + 15} \psi_M(P, Q) \leq C_*(P, Q) \leq \frac{8e^\beta \beta^{\frac{7}{2}} (\beta^2 + 4\beta + 1)}{15\beta^4 + 2\beta^2 + 15} \psi_M(P, Q), \quad (3.27)$$

where

$$\psi_M(P, Q) = \sum_{i=1}^n \frac{(p_i^2 - q_i^2)^2}{2(p_i q_i)^{\frac{3}{2}}} \quad (3.28)$$

is the Kumar- Johnson divergence (2005) [28].

II. With Non- symmetric divergence measures:

Proposition 3.9. Let $P, Q \in \Gamma_n$ and $0 < \alpha \leq \frac{p_i}{q_i} \leq \beta < \infty$, then we have

$$\frac{e^\alpha (\alpha^2 + 4\alpha + 1)}{2} \chi^2(P, Q) \leq C_*(P, Q) \leq \frac{e^\beta (\beta^2 + 4\beta + 1)}{2} \chi^2(P, Q), \quad (3.29)$$

where $\chi^2(P, Q)$ is given by (3.32).

Proof. Let us consider

$$f_2(t) = (t-1)^2, t \in (0, \infty) \quad (3.30)$$

and

$$f_2'(t) = 2(t-1),$$

$$f_2''(t) = 2. \quad (3.31)$$

Since $f_2''(t) > 0 \forall t > 0$ and $f_2(1) = 0$, so $f_2(t)$ is strictly convex and normalized function respectively. By putting $f_2(t)$ in (1.1), we get

$$C_{f_2}(P, Q) = \sum_{i=1}^n \frac{(p_i - q_i)^2}{q_i} = \chi^2(P, Q), \quad (3.32)$$

where $\chi^2(P, Q)$ is called the Chi- square divergence or Pearson divergence (1900) [33].

Now, let

$$g(t) = \frac{f_1''(t)}{f_2''(t)} = \frac{e^t (t^2 + 4t + 1)}{2},$$

where $f_1''(t)$ and $f_2''(t)$ are given by (2.2) and (3.31) respectively and

$$g'(t) = \frac{e^t(t^2 + 6t + 5)}{2}.$$

It is clear that $g'(t) > 0$ for $t > 0$, therefore $g(t)$ is strictly increasing function in interval $(0, \infty)$. So

$$m = \inf_{t \in (\alpha, \beta)} g(t) = g(\alpha) = \frac{e^\alpha(\alpha^2 + 4\alpha + 1)}{2} \tag{3.33}$$

and

$$M = \sup_{t \in (\alpha, \beta)} g(t) = g(\beta) = \frac{e^\beta(\beta^2 + 4\beta + 1)}{2}. \tag{3.34}$$

The result (3.29) is obtained by using (2.3), (3.32), (3.33), and (3.34) in inequalities (1.5).

Proposition 3.10. *Let $P, Q \in \Gamma_n$ and $0 < \alpha \leq \frac{p_i}{q_i} \leq \beta < \infty$, then we have*

$$e^\alpha \alpha (\alpha^2 + 4\alpha + 1) K(P, Q) \leq C_*(P, Q) \leq e^\beta \beta (\beta^2 + 4\beta + 1) K(P, Q), \tag{3.35}$$

where $K(P, Q)$ is given by (3.38).

Proof. Let us consider

$$f_2(t) = t \log t, t \in (0, \infty) \tag{3.36}$$

and

$$f_2'(t) = 1 + \log t,$$

$$f_2''(t) = \frac{1}{t}. \tag{3.37}$$

Since $f_2''(t) > 0 \forall t > 0$ and $f_2(1) = 0$, so $f_2(t)$ is strictly convex and normalized function respectively. By putting $f_2(t)$ in (1.1), we get

$$C_{f_2}(P, Q) = \sum_{i=1}^n p_i \log \frac{p_i}{q_i} = K(P, Q), \tag{3.38}$$

where $K(P, Q)$ is called the Kullback- Leibler divergence or Relative entropy or Directed divergence or Information gain (1951) [27].

Now, let

$$g(t) = \frac{f_1''(t)}{f_2''(t)} = e^t t (t^2 + 4t + 1),$$

where $f_1''(t)$ and $f_2''(t)$ are given by (2.2) and (3.37) respectively and

$$g'(t) = e^t (t^3 + 7t^2 + 9t + 1).$$

It is clear that $g'(t) > 0$ for $t > 0$, therefore $g(t)$ is strictly increasing function in interval $(0, \infty)$. So

$$m = \inf_{t \in (\alpha, \beta)} g(t) = g(\alpha) = e^\alpha \alpha (\alpha^2 + 4\alpha + 1) \tag{3.39}$$

and

$$M = \sup_{t \in (\alpha, \beta)} g(t) = g(\beta) = e^\beta \beta (\beta^2 + 4\beta + 1). \tag{3.40}$$

The result (3.35) is obtained by using (2.3), (3.38), (3.39), and (3.40) in inequalities (1.5).

Proposition 3.11. *Let $P, Q \in \Gamma_n$ and $0 < \alpha \leq \frac{p_i}{q_i} \leq \beta < \infty$, then we have*

$$2\alpha^2 (1 + \alpha) e^\alpha (\alpha^2 + 4\alpha + 1) G(P, Q) \leq C_*(P, Q) \leq 2\beta^2 (1 + \beta) e^\beta (\beta^2 + 4\beta + 1) G(P, Q), \tag{3.41}$$

where $G(P, Q)$ is given by (3.44).

Proof. Let us consider

$$f_2(t) = \left(\frac{t+1}{2}\right) \log \frac{t+1}{2t}, t \in (0, \infty) \quad (3.42)$$

and

$$f_2'(t) = \frac{1}{2} \left[\log \frac{t+1}{2t} - \frac{1}{t} \right],$$

$$f_2''(t) = \frac{1}{2t^2(t+1)}. \quad (3.43)$$

Since $f_2''(t) > 0 \forall t > 0$ and $f_2(1) = 0$, so $f_2(t)$ is strictly convex and normalized function respectively. By putting $f_2(t)$ in (1.1), we get

$$C_{f_2}(P, Q) = \sum_{i=1}^n \left(\frac{p_i + q_i}{2}\right) \log \frac{p_i + q_i}{2p_i} = G(P, Q), \quad (3.44)$$

where $G(P, Q)$ is called the Relative Arithmetic- Geometric divergence (1995) [40]. Now, let

$$g(t) = \frac{f_1''(t)}{f_2''(t)} = 2e^t t^2 (t+1) (t^2 + 4t + 1),$$

where $f_1''(t)$ and $f_2''(t)$ are given by (2.2) and (3.43) respectively and

$$g'(t) = 2te^t (t^4 + 10t^3 + 25t^2 + 16t + 2).$$

It is clear that $g'(t) > 0$ for $t > 0$, therefore $g(t)$ is strictly increasing function in interval $(0, \infty)$. So

$$m = \inf_{t \in (\alpha, \beta)} g(t) = g(\alpha) = 2\alpha^2 (1 + \alpha) e^\alpha (\alpha^2 + 4\alpha + 1) \quad (3.45)$$

and

$$M = \sup_{t \in (\alpha, \beta)} g(t) = g(\beta) = 2\beta^2 (1 + \beta) e^\beta (\beta^2 + 4\beta + 1). \quad (3.46)$$

The result (3.41) is obtained by using (2.3), (3.44), (3.45), and (3.46) in inequalities (1.5). In a similar procedure, we obtain the bounds of $C_*(P, Q)$ with the other well known non-symmetric divergence measures. The results are as follows.

Proposition 3.12. If $f_2(t) = (t-1) \log \frac{t+1}{2}$, then we have

$$\frac{e^\alpha (1 + \alpha)^2 (\alpha^2 + 4\alpha + 1)}{\alpha + 3} J_R(P, Q) \leq C_*(P, Q) \leq \frac{e^\beta (1 + \beta)^2 (\beta^2 + 4\beta + 1)}{\beta + 3} J_R(P, Q), \quad (3.47)$$

where

$$J_R(P, Q) = \sum_{i=1}^n (p_i - q_i) \log \left(\frac{p_i + q_i}{2q_i}\right) \quad (3.48)$$

is the Relative J -divergence (2001) [12].

Proposition 3.13. If $f_2(t) = t \log \frac{2t}{t+1}$, then we have

$$e^\alpha \alpha (1 + \alpha)^2 (\alpha^2 + 4\alpha + 1) F(P, Q) \leq C_*(P, Q) \leq e^\beta \beta (1 + \beta)^2 (\beta^2 + 4\beta + 1) F(P, Q), \quad (3.49)$$

where

$$F(P, Q) = \sum_{i=1}^n p_i \log \frac{2p_i}{p_i + q_i} \quad (3.50)$$

is the Relative Jensen- Shannon divergence (1969) [38].

Proposition 3.14. If $f_2(t) = \frac{(t^2-1)^2}{t}$, then we have

$$\frac{e^\alpha \alpha^3 (\alpha^2 + 4\alpha + 1)}{2(3\alpha^4 + 1)} \gamma_1(P, Q) \leq C_*(P, Q) \leq \frac{e^\beta \beta^3 (\beta^2 + 4\beta + 1)}{2(3\beta^4 + 1)} \gamma_1(P, Q), \tag{3.51}$$

where

$$\gamma_1(P, Q) = \sum_{i=1}^n \frac{(p_i^2 - q_i^2)^2}{p_i q_i^2} \tag{3.52}$$

is the Jain Chhabra divergence (2014) [17].

Proposition 3.15. If $f_2(t) = e^t(t - 1)$, then we have

$$\frac{\alpha^2 + 4\alpha + 1}{\alpha + 1} G_{\text{exp}}(P, Q) \leq C_*(P, Q) \leq \frac{\beta^2 + 4\beta + 1}{\beta + 1} G_{\text{exp}}(P, Q), \tag{3.53}$$

where

$$G_{\text{exp}}(P, Q) = \sum_{i=1}^n e^{\frac{p_i}{q_i}} (p_i - q_i) \tag{3.54}$$

is the Jain Chhabra Exponential divergence (2016) [18].

Proposition 3.16. If $f_2(t) = \frac{(t^2-1)^2}{\sqrt{t}}$, then we have

$$\frac{4e^\alpha \alpha^{\frac{5}{2}} (\alpha^2 + 4\alpha + 1)}{35\alpha^4 - 6\alpha^2 + 3} \xi_1(P, Q) \leq C_*(P, Q) \leq \frac{4e^\beta \beta^{\frac{5}{2}} (\beta^2 + 4\beta + 1)}{35\beta^4 - 6\beta^2 + 3} \xi_1(P, Q), \tag{3.55}$$

where

$$\xi_1(P, Q) = \sum_{i=1}^n \frac{(p_i^2 - q_i^2)^2}{(p_i q_i)^{\frac{1}{2}} q_i^2} \tag{3.56}$$

is the Jain Chhabra divergence (2014) [19].

4 Verification of Bounds

In this section, we take example for calculating the divergences $\Delta(P, Q)$, $h(P, Q)$, $G(P, Q)$, $\gamma_1(P, Q)$ and $C_*(P, Q)$ and verify numerically the inequalities (3.1), (3.21), (3.41) and (3.51).

Example 4.1. Let P be the binomial probability distribution with parameters $(n = 10, p = 0.7)$ and Q its approximated Poisson probability distribution with parameter $(\lambda = np = 7)$ for the random variable X , then we have

x_i	0	1	2	3	4	5	6	7	8	9	10
$p_i \approx$.0000059	.000137	.00144	.009	.036	.102	.200	.266	.233	.121	.0282
$q_i \approx$.000911	.00638	.022	.052	.091	.177	.199	.149	.130	.101	.0709
$\frac{p_i}{q_i} \approx$.00647	.0214	.0654	.173	.395	.871	1.005	1.785	1.792	1.198	.397

Table 1. Evaluation of discrete probability distributions for $n = 10, p = 0.7, q = 0.3$

By using Table 1, we get the followings:

$$\alpha (= .00647) \leq \frac{p_i}{q_i} \leq \beta (= 1.792). \tag{4.1}$$

$$\Delta(P, Q) = \sum_{i=1}^{11} \frac{(p_i - q_i)^2}{p_i + q_i} \approx .1812. \tag{4.2}$$

$$h(P, Q) = \sum_{i=1}^{11} \frac{(\sqrt{p_i} - \sqrt{q_i})^2}{2} \approx .0502. \quad (4.3)$$

$$G(P, Q) = \sum_{i=1}^{11} \frac{p_i + q_i}{2} \log \left(\frac{p_i + q_i}{2p_i} \right) \approx .0746. \quad (4.4)$$

$$\gamma_1(P, Q) = \sum_{i=1}^{11} \frac{(p_i^2 - q_i^2)^2}{p_i q_i^2} \approx 2.25065. \quad (4.5)$$

$$C_*(P, Q) = \sum_{i=1}^{11} \frac{e^{\frac{p_i}{q_i}} (p_i^2 - q_i^2)}{q_i} \approx 3.818206. \quad (4.6)$$

Put the approximated values from (4.1) to (4.6) in inequalities (3.1), (3.21), (3.41) and (3.51) respectively and get the following results

$$.02384 \leq [C_*(P, Q) = 3.818206] \leq 33.66544,$$

$$1.07905 \times 10^{-4} \leq [C_*(P, Q) = 3.818206] \leq 32.8958,$$

$$6.49086 \times 10^{-6} \leq [C_*(P, Q) = 3.818206] \leq 91.3545 \text{ and}$$

$$3.14708 \times 10^{-7} \leq [C_*(P, Q) = 3.818206] \leq 13.8475 \text{ respectively.}$$

Hence verified the bounds of $C_*(P, Q)$ in terms of the $\Delta(P, Q)$, $h(P, Q)$, $G(P, Q)$ and $\gamma_1(P, Q)$ for $p = 0.7$, where $C_*(P, Q)$, $\Delta(P, Q)$, $h(P, Q)$, $G(P, Q)$ and $\gamma_1(P, Q)$ are given by (2.3), (3.4), (3.22), (3.44) and (3.52) respectively.

Similarly, we can verify the bounds of $C_*(P, Q)$ in terms of the other divergences or can verify the other inequalities for different values of p and q and for other discrete probability distributions as well, like; Negative binomial, Geometric, uniform etc.

5 Metric Space Nature

We know that $C_*(P, Q)$ is non- symmetric but

$$\begin{aligned} C_*(P, Q) + C_*(Q, P) &= \sum_{i=1}^n \frac{(p_i^2 - q_i^2) e^{\frac{p_i}{q_i}}}{q_i} + \sum_{i=1}^n \frac{(q_i^2 - p_i^2) e^{\frac{q_i}{p_i}}}{p_i} \\ &= \sum_{i=1}^n (p_i^2 - q_i^2) \left(\frac{e^{\frac{p_i}{q_i}}}{q_i} - \frac{e^{\frac{q_i}{p_i}}}{p_i} \right) = C_*^*(P, Q). \end{aligned} \quad (5.1)$$

is symmetric with respect to probability distributions $P, Q \in \Gamma_n$, as $C_*^*(P, Q) = C_*^*(Q, P)$.

We can see that $\sqrt{C_*^*(P, Q)} > 0$ and $= 0$ if and only if $P = Q$ or $p_i = q_i \forall i = 1, 2, 3, \dots, n$. The $\sqrt{C_*^*(P, Q)}$ is symmetric because $C_*^*(P, Q)$ is symmetric or $\sqrt{C_*^*(P, Q)} = \sqrt{C_*^*(Q, P)}$.

In this section we prove that $\sqrt{C_*^*(P, Q)}$ satisfies triangle inequality and then obtain a new metric space over an interval $(0, \infty)$. For this, we prove the following theorem.

Theorem 5.1. Let $x(p, q) : (0, \infty) \times (0, \infty) \rightarrow (0, \infty)$ be defined as

$$x(p, q) = (p^2 - q^2) \left(\frac{e^{\frac{p}{q}}}{q} - \frac{e^{\frac{q}{p}}}{p} \right), \quad (5.2)$$

i.e., we can write

$$C_*^*(P, Q) = \sum_{i=1}^n x(p_i, q_i). \quad (5.3)$$

Then triangle inequality will be

$$\sqrt{x(p, q)} \leq \sqrt{x(p, r)} + \sqrt{x(r, q)}, \quad (5.4)$$

where $p, q, r \in (0, \infty)$.

Proof. To prove the inequality (5.4), first let us consider

$$X_{pq}(r) = \sqrt{x(p,r)} + \sqrt{x(r,q)}, \tag{5.5}$$

then

$$\frac{d}{dr} X_{pq}(r) = X'_{pq}(r) = \frac{x'(p,r)}{2\sqrt{x(p,r)}} + \frac{x'(r,q)}{2\sqrt{x(r,q)}}. \tag{5.6}$$

Now from (5.2), we can write

$$x(p,r) = (p^2 - r^2) \left(\frac{e^{\frac{p}{r}}}{r} - \frac{e^{\frac{r}{p}}}{p} \right) \tag{5.7}$$

and after differentiating (5.7) w.r.t r , we obtain

$$x'(p,r) = -\frac{e^{\frac{p}{r}}}{r^3} [p^3 + r^3 + rp(p-r)] - \frac{e^{\frac{r}{p}}}{p^2} (p^2 - r^2 - 2rp). \tag{5.8}$$

Put $p = rt$, i.e., $t = \frac{p}{r} \in (0, \infty)$ in (5.8), we get

$$[x'(p,r)]_{p=rt} = k(t) = -e^t (t^3 + t^2 - t + 1) - \frac{e^{\frac{1}{t}}}{t^2} (t^2 - 2t - 1). \tag{5.9}$$

Now from (5.7), we can write

$$x(t,1) = \frac{(t^2 - 1) (te^t - e^{\frac{1}{t}})}{t}. \tag{5.10}$$

From (5.7) and (5.10), we have the following relation for $p = rt$

$$\sqrt{x(p,r)} = \sqrt{r} \sqrt{x(t,1)} = \sqrt{r} l(t), \tag{5.11}$$

where we are assuming

$$\sqrt{x(t,1)} = l(t). \tag{5.12}$$

Now, differentiate (5.9) w.r.t. t , we obtain

$$k'(t) = -\frac{(t^2 + 4t + 1) (e^t t^5 + e^{\frac{1}{t}})}{t^4}. \tag{5.13}$$

Now, let we define a function

$$s(t) = \frac{k(t)}{l(t)}, \forall t \in (0, \infty). \tag{5.14}$$

From (5.10) and (5.13), we can see that $l(t) = \sqrt{x(t,1)} \geq 0$ and $k'(t) < 0 \forall t \in (0, \infty)$, i.e., $k(t)$ is monotonically decreasing function and $k(1) = 0$, so $s(t)$ will be decreasing as well in $(0, \infty)$ with $\lim_{t \rightarrow 1} s(t) = 0$ or the nature of $s(t)$ depends on the nature of $k(t)$ only as $l(t)$ is fix and positive. Therefore, we conclude that $s(t)$ changes the sign at $t = 1$, so

$$s(t) = \begin{cases} > 0 & \text{if } t < 1 \\ < 0 & \text{if } t > 1 \\ = 0 & \text{if } t \rightarrow 1 \end{cases}. \tag{5.15}$$

Now suppose $u = \frac{q}{p} \in (0, \infty) \Rightarrow \frac{q}{r} = \frac{qp}{pr} = ut \in (0, \infty)$, so (5.6) can be written as

$$2\sqrt{r} X'_{pq}(r) = s(t) + s(ut). \tag{5.16}$$

Now we have two cases on u , as follows.

Case I: If we are taking $u > 1$ or $q > p$, then (by considering that $s(t)$ is decreasing function)

- (a) For $t > 1 \Rightarrow s(t) < 0$ and $s(ut) < 0 \Rightarrow s(t) + s(ut) < 0$.
- (b) For $\frac{1}{u} < t < 1 \Rightarrow s(t) > 0$ and $s(ut) < 0 \Rightarrow s(t) > s(ut) \Rightarrow s(t) + s(ut) > 0$.
- (c) For $t < \frac{1}{u} < 1 \Rightarrow s(t) > 0$ and $s(ut) > 0$.

It means $X'_{pq}(r) = \frac{s(t)+s(ut)}{2\sqrt{r}}$ changes the sign at $t = 1$ or $r = p$, so $X_{pq}(r)$ attains its minimum value at $t = 1$ or $r = p$.

Case II: This case is for $u < 1$ or $q < p$, can be done in a similar manner.

Similarly, repeating the above procedure by considering $t = \frac{q}{r} \in (0, \infty)$ and $u = \frac{p}{q} \in (0, \infty) \Rightarrow \frac{p}{r} = \frac{p}{q} \frac{q}{r} = ut \in (0, \infty)$, then we get that $X'_{pq}(r)$ changes the sign at $t = 1$ or $r = q$, so $X_{pq}(r)$ attains its minimum value at $t = 1$ or $r = q$. Therefore, right side of (5.4) has its minimum value at $p = q = r \forall p, q, r \in (0, \infty)$.

Hence proof the result (5.4) or theorem 5.1.

In view of this proof, we conclude that the new divergence measure $\sqrt{C^*(P, Q)}$ is a metric or we obtain a new metric space $\sqrt{(C^*, (0, \infty))}$ over $(0, \infty)$.

Comparison Graph

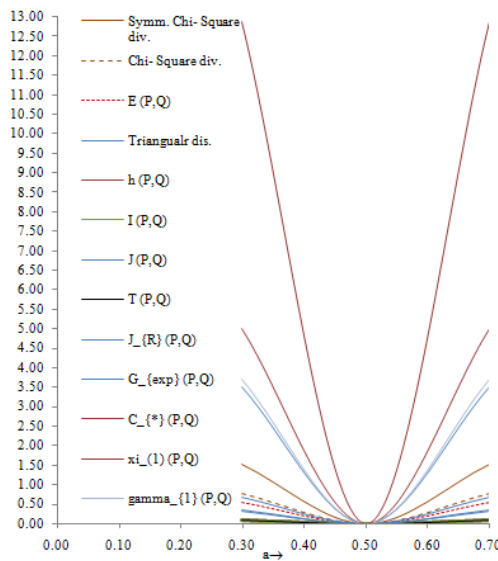


Figure 2. Comparison of Divergence Measures

In Figure 2, we have considered $p_i = (a, 1 - a), q_i = (1 - a, a)$, where $a \in (0, 1)$. It is clear from the Figure that the new divergence $C^*(P, Q)$ has a steeper slope than the other well known divergences.

References

- [1] R. K. Bajaj and D.S. Hooda, Generalized measures of fuzzy directed divergence, total ambiguity and information improvement, Journal of Applied Mathematics, Statistics and Informatics, vol. 6 (2010), no. 2, pp: 31- 44.
- [2] M.B. Bassat, f- Entropies, probability of error and feature selection, Inform. Control, vol. 39, 1978, pp: 227-242.
- [3] A. Benveniste, M. Basseville, G. Moustakides, The asymptotic local approach to change detection and model validation, IEEE Trans. Automatic Control, vol. AC-32, no. 7, pp: 583- 592, 1987.
- [4] J.S. Bhullar, O.P. Vinocha, and M. Gupta, Generalized measure for two utility distributions, Proceedings of the World Congress on Engineering, vol. 3, June 30- July 2, 2010.
- [5] D.E. Boekee, J.C.A. Van Der Lubbe, Some aspects of error bounds in feature selection, Pattern Recognition, vol. 11, pp: 353- 360, 1979.
- [6] J. Burbea and C.R. Rao, On the convexity of some divergence measures based on entropy functions, IEEE Trans. on Inform. Theory, IT-28 (1982), pp: 489-495.
- [7] H.C. Chen, Statistical pattern recognition, Hoyderc Book Co., Rocelle Park, New York, 1973.

-
- [8] C.K. Chow and C.N. Lin, Approximating discrete probability distributions with dependence trees, *IEEE Trans. Inform. Theory*, vol. 14, 1968, no 3, pp: 462-467.
- [9] I. Csiszar, Information measures: A Critical survey, in *Trans. In: Seventh Prague Conf. on Information Theory*, Academia, Prague, 1974, pp: 73-86.
- [10] I. Csiszar, Information type measures of differences of probability distribution and indirect observations, *Studia Math. Hungarica*, vol. 2, pp: 299- 318, 1967.
- [11] D. Dacunha- Castelle, *Ecole d'Ete de Probabilites de, Saint-Flour VII-1977*, Berlin, Heidelberg, New York: Springer, 1978.
- [12] S.S. Dragomir, V. Gluscevic, and C.E.M. Pearce, Approximation for the Csiszar's 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, Vol. 1, 2001, pp: 139-154.
- [13] S.S. Dragomir, J. Sunde and C. Buse, New inequalities for Jeffreys divergence measure, *Tamusi Oxford Journal of Mathematical Sciences*, 16(2) (2000), 295-309.
- [14] D.V. Gokhale and S. Kullback, *Information in contingency Tables*, New York, Marcel Dekker, 1978.
- [15] E. Hellinger, Neue begrundung der theorie der quadratischen formen von unendlichen vielen veranderlichen, *J. Rein.Aug. Math.*, 136 (1909), pp: 210-271.
- [16] D.S. Hooda, On generalized measures of fuzzy entropy, *Mathematica Slovaca*, vol. 54 (2004), no. 3, pp: 315- 325.
- [17] K.C. Jain and P. Chhabra, New series of information divergence measures and their properties, *Applied Mathematics and Information Sciences*, vol. 10, no.4 (2016), pp: 1433- 1446.
- [18] K.C. Jain and P. Chhabra, Applications of new exponential information divergence measure, *Palestine Journal of Mathematics*, vol. 5, 02, pp: 249- 257, 2016.
- [19] K.C. Jain and P. Chhabra, Series of new information divergences, properties and corresponding series of metric spaces, *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 3, no.5 (2014), pp: 12124- 12132.
- [20] K.C. Jain and A. Srivastava, On symmetric information divergence measures of Csiszar's f- divergence class, *Journal of Applied Mathematics, Statistics and Informatics (JAMSI)*, 3 (2007), no.1, pp- 85- 102.
- [21] Jeffreys, An invariant form for the prior probability in estimation problem, *Proc. Roy. Soc. Lon. Ser. A*, 186(1946), pp: 453-461.
- [22] P. Jha and V. K. Mishra, Some new trigonometric, hyperbolic and exponential measures of fuzzy entropy and fuzzy directed divergence, *International Journal of Scientific and Engineering Research*, vol. 3 (2012), no. 4, pp: 1- 5.
- [23] L. Jones and C. Byrne, General entropy criteria for inverse problems with applications to data compression, pattern classification and cluster analysis, *IEEE Trans. Inform. Theory*, vol. 36, 1990, pp: 23- 30.
- [24] T.T. Kadota and L.A. Shepp, On the best finite set of linear observables for discriminating two Gaussian signals, *IEEE Trans. Inform. Theory*, vol. 13, 1967, pp: 288-294.
- [25] T. Kailath, The divergence and Bhattacharyya distance measures in signal selection, *IEEE Trans. Comm. Technology*, vol. COM-15, 1967, pp: 52- 60.
- [26] D. Kazakos and T. Cotsidas, A decision theory approach to the approximation of discrete probability densities, *IEEE Trans. Perform. Anal. Machine Intell.*, vol. 1, 1980, pp: 61- 67.
- [27] S. Kullback and R.A. Leibler, On information and sufficiency, *Ann. Math. Statist.*, 22 (1951), pp: 79-86.
- [28] P. Kumar and A. Johnson, On a symmetric divergence measure and information inequalities, *Journal of Inequalities in Pure and Applied Mathematics*, 6(3) (2005), Article 65, 1-13.
- [29] P.W. Lamberti, A.P. Majtey, A. Borrás, M. Casas, A. Plastino, Metric character of the quantum Jensen-Shannon divergence, *Physical Review A*, 77:052311, 2008.
- [30] F. Nielsen and S. Boltz, The Burbea- Rao and Bhattacharyya centroids, Apr. 2010, Arxiv.
- [31] M.A. Nielsen, I.L. Chuang, *Quantum computation and information*, Cambridge University Press, Cambridge, UK, 3 (8):9, 2000.
- [32] F. Osterreicher, Csiszar's f- divergence basic properties, Homepage: <http://www.sbg.ac.at/mat/home.html>, November 22, 2002.
- [33] 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), pp: 157-172.
- [34] E.C. Pielou, *Ecological diversity*, New York, Wiley, 1975.
- [35] H.V. Poor, Robust decision design using a distance criterion, *IEEE Trans. Inf. Th.*, vol.- IT 26, no. 5, pp: 575- 587, 1980.

- [36] M. Salicru, Measures of information associated with Csiszar's divergences, *Kybernetika*, vol. 30 (1994), no. 5, pp: 563- 573.
- [37] R. Santos-Rodriguez, D. Garcia-Garcia, and J. Cid-Sueiro, Cost-sensitive classification based on Bregman divergences for medical diagnosis, In M.A. Wani, editor, *Proceedings of the 8th International Conference on Machine Learning and Applications (ICMLA'09)*, Miami Beach, Fl., USA, December 13-15, 2009, pp: 551- 556, 2009.
- [38] R. Sibson, Information radius, *Z. Wahrs. Undverw. Geb.*, (14) (1969), pp: 149-160.
- [39] H.C. Taneja and R.K. Tuteja, Characterization of a quantitative- qualitative measure of inaccuracy, *Kybernetika*, vol. 22 (1986), no. 5, pp: 393- 402.
- [40] I.J. Taneja, New developments in generalized information measures, Chapter in: *Advances in Imaging and Electron Physics*, Ed. P.W. Hawkes, 91(1995), pp: 37-135.
- [41] I.J. Taneja, Generalized symmetric divergence measures and inequalities, *RGMA Research Report Collection*, <http://rgmia.vu.edu.au>, 7(4) (2004), Art. 9. Available on-line at: arXiv:math.ST/0501301 v1 19 Jan 2005.
- [42] B. Taskar, S. Lacoste-Julien, and M.I. Jordan, Structured prediction, dual extra gradient and Bregman projections, *Journal of Machine Learning Research*, 7, pp: 1627- 1653, July 2006.
- [43] H. Theil, *Statistical decomposition analysis*, Amsterdam, North-Holland, 1972.
- [44] H. Theil, *Economics and information theory*, Amsterdam, North-Holland, 1967.
- [45] K. Tumer, J. Ghosh, Estimating the Bayes error rate through classifier combining, *Proceedings of 13th International Conference on Pattern Recognition*, pp: 695- 699, 1996.
- [46] B. Vemuri, M. Liu, S. Amari, and F. Nielsen, Total Bregman divergence and its applications to DTI analysis, *IEEE Transactions on Medical Imaging*, 2010.

Author information

Dr. Praphull Chhabra, Department of Mathematics, University of Engineering and Management, Jaipur-303807 (Rajasthan), India.
E-mail: prf124@gmail.com

Received: September 22, 2017.

Accepted: December 6, 2017