J. Appl. Math. & Informatics Vol. **34**(2016), No. 3 - 4, pp. 295 - 308
http://dx.doi.org/10.14317/jami.2016.295

A NEW EXPONENTIAL DIRECTED DIVERGENCE INFORMATION MEASURE

K.C. JAIN*, PRAPHULL CHHABRA

Abstract. Depending upon the nature of the problem, different divergence measures are suitable. So it is always desirable to develop a new divergence measure. In the present work, new information divergence measure, which is exponential in nature, is introduced and characterized. Bounds of this new measure are obtained in terms of various symmetric and non- symmetric measures together with numerical verification by using two discrete distributions: Binomial and Poisson. Fuzzy information measure and Useful information measure corresponding to new exponential divergence measure are also introduced.

AMS Mathematics Subject Classification : Primary 94A17, Secondary 26D15.
Key words and phrases : New exponential divergence measure, Bounds, Numerical verification, Comparison of divergence measures.

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 [50], quantum information theory [33, 35], model validation [4], robust detection [39], economics and political science [48, 49], biology [38], analysis of contingency tables [18], approximation of probability distributions [11, 29], signal processing [27, 28], pattern recognition [2, 7, 10, 26], color image segmentation [34], 3D image segmentation and word alignment [47], cost- sensitive classification for medical diagnosis [42], magnetic resonance image analysis [51] etc.

Received September 18, 2015. Revised January 18, 2016. Accepted January 24, 2016.
*Corresponding author.

Also we can use divergence measures in fuzzy mathematics as fuzzy directed divergences and fuzzy entropies [1, 20, 25], 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 [6, 44], i.e., an event is how much useful compare to other event.

Without essential loss of insight, we have restricted ourselves to discrete probability distributions, so let $\Gamma_n = \{P = (p_1, p_2, p_3, ..., p_n) : p_i > 0, \sum_{i=1}^{n} p_i = 1\}$, $n \geq 2$ 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 $i = 1, 2, 3..., n$, then we have to suppose that $0f(0) = 0f\left(\frac{0}{0}\right) = 0$.

Some generalized f- information divergence measures had been introduced, characterized and applied in variety of fields, such as: Csiszar's f- divergence [12, 13], Bregman's f- divergence [8], Burbea- Rao's f- divergence [9], Renyi's like f- divergence [40], M- divergence [41], Jain- Saraswat f- divergence [22] etc.

Besides these, The f- divergence measure [3] with respect to two functions (f, g) is also introduced, which is

$$d(P, Q) = g\left[\sum_{i=1}^{n} p_i f\left(\frac{q_i}{p_i}\right)\right],$$

where g is an increasing function on R and f is real, continuous, and convex function on R^+. We obtain many standard divergence measures by suitably defining the function f and g, such as: for $f(t) = -t^{1-r}, g(t) = -\log(-t), 0 \leq r \leq 1$, we get $d(P, Q) = -\log\left(\sum_{i=1}^{n} p_i^r q_i^{1-r}\right)$ called Chernoff Coefficient and at $r = \frac{1}{2}$, we obtain $-\log\left(\sum_{i=1}^{n} \sqrt{p_i q_i}\right)$ well known the Bhattacharyya distance [5]. Similarly for $f(t) = \left|1 - t^{\frac{1}{r}}\right|^r, r \geq 1, g(t) = t^{\frac{1}{r}}$, we obtain $d(P, Q) = \left(\sum_{i=1}^{n} \left|p_i^{\frac{1}{r}} - q_i^{\frac{1}{r}}\right|^r\right)^{\frac{1}{r}}$ so called Generalized Matusita distance and at $r = 1$, we obtain $\sum_{i=1}^{n} |p_i - q_i|$ the well known Variational distance or l_1 distance [30].

Csiszar's f- divergence is widely used due to its compact nature, which is given by

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

where $f : (0, \infty) \to R$ (set of real no.) is real, continuous, and convex function and $P = (p_1, p_2, ..., p_n), Q = (q_1, q_2, ..., q_n) \in \Gamma_n$, 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. Fundamental properties of $C_f(P,Q)$ can be seen in literature [36], in detail.

Remark 1.1. For comparing multiple number of discrete probability distributions, following will be the Csiszar's generalized f- divergence [15]

$$C_f^n(P_1,...,P_n,Q_1,...,Q_n)=\sum_{i=1}^m...\sum_{i=1}^m q_{i1}...q_{in} f\left(\frac{\frac{p_{i1}}{q_{i1}}+...+\frac{p_{in}}{q_{in}}}{n}\right)$$

and following relation can be seen as well in the same literature

$$C_f^1(P_1,Q_1)\geq C_f^2(P_1,P_2,Q_1,Q_2)\geq ...\geq C_f^n(P_1,...,P_n,Q_1,...,Q_n)$$
$$\geq C_f^{n+1}(P_1,...,P_{n+1},Q_1,...,Q_{n+1})\geq f(1).$$

Divergences between more than two probability distributions are useful for discrimination and taxonomy.

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

$$f\left(\sum_{i=1}^n\lambda_i t_i\right)\leq\sum_{i=1}^n\lambda_i f(t_i). \tag{2}$$

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 $\frac{p_i}{q_i}$ for each $i=1,...,n$ by assuming that the function is normalized, i.e., $f(1)=0$, we get*

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

The following theorem is well known in literature [13].

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 is given by Taneja (2005), which relates two generalized f- divergence measures.

Theorem 1.5. *Let $f_1, f_2 : I \subset R^+ \to R$ 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), $0 < \alpha \leq 1 < \beta < \infty$.
(ii) There exists the real constants m, M such that $m < M$ and

$$m \leq \frac{f_1'' (t)}{f_2'' (t)} \leq M, f_2'' (t) \neq 0 \; \forall \; t \in (\alpha, \beta) . \tag{3}$$

If $P, Q \in \Gamma_n$, then we have the following inequalities

$$m C_{f_2} (P, Q) \leq C_{f_1} (P, Q) \leq M C_{f_2} (P, Q) , \tag{4}$$

where $C_f (P, Q)$ is given by (1).

2. New exponential divergence measure and properties

In this section, we introduce a new exponential divergence measure of Csiszar's class and define the properties.
Let $f : (0, \infty) \to R$ be a real differentiable mapping, which is defined as

$$f (t) = f_1 (t) = e^t (t - 1) , \forall \; t \in (0, \infty) , \tag{5}$$

$$f_1' (t) = t e^t$$

and

$$f_1'' (t) = e^t (t + 1) . \tag{6}$$

We can check that the function $f_1 (t)$ is exponential in nature and convex normalized because $f_1'' (t) > 0 \; \forall \; t \in (0, \infty)$ and $f_1 (1) = 0$ respectively. Further $f_1 (t)$ is monotonically increasing in $(0, \infty)$ as $f_1' (t) > 0$ in $(0, \infty)$.
After putting this exponential function in (1), we obtain

$$C_{f_1} (P, Q) = G_{\exp} (P, Q) = \sum_{i=1}^{n} e^{\frac{p_i}{q_i}} (p_i - q_i) . \tag{7}$$

In view of corollary (1.3) and theorem (1.4), we see that $G_{\exp} (P, Q)$ 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 $p_i = q_i$. We can also see that $G_{\exp} (P, Q)$ is non- symmetric divergence w.r.t. P and Q because $G_{\exp} (P, Q) \neq G_{\exp} (Q, P)$.

Remark 2.1. If function $f (t)$ is convex in interval $(0, \infty)$, then $f_* (t) = t f \left(\frac{1}{t} \right) = e^{\frac{1}{t}} (1 - t)$ will be a convex function as well because $f_*'' (t) = \frac{e^{\frac{1}{t}} \left(4t^2 - 3t + 1 \right)}{t^2} > 0 \; \forall \; t > 0$, called conjugate of $f (t)$. By putting this conjugate convex function in (1), we get $C_{f_*} (P, Q) = G_{\exp}^* (P, Q) = \sum_{i=1}^n e^{\frac{q_i}{p_i}} (q_i - p_i)$, and we can see

$$G_{\exp} (P, Q) + G_{\exp}^* (P, Q) = \sum_{i=1}^{n} (p_i - q_i) \left(e^{\frac{p_i}{q_i}} - e^{\frac{q_i}{p_i}} \right) \tag{8}$$

is a symmetric exponential divergence.

Consequently, we obtain the following intra relations among new exponential divergences by applying remark (1.1), for comparing multiple number of discrete probability distributions taking normalized function

$$G_{\exp}^{1}(P_1,Q_1) \geq G_{\exp}^{2}(P_1,P_2,Q_1,Q_2) \geq ... \geq G_{\exp}^{n}(P_1,...,P_n,Q_1,...,Q_n)$$
$$\geq G_{\exp}^{n+1}(P_1,...,P_{n+1},Q_1,...,Q_{n+1}) \geq 0. \tag{9}$$

Remark 2.2. Bajaj and Hooda (2010) have defined 'useful' fuzzy directed divergence of fuzzy set A from fuzzy set B. We can also define a new exponential measure of 'useful' fuzzy directed divergence on the same lines, for this let A and B be two standard fuzzy sets with same supporting points $x_1, x_2, ..., x_n$ and with fuzzy vectors $\mu_A(x_1), ..., \mu_A(x_n)$ and $\mu_B(x_1), ..., \mu_B(x_n)$, then fuzzy information measure corresponding to new exponential measure (7), will be

$$G_{\exp}^{\wedge}(A,B) = \sum_{i=1}^{n} [\mu_A(x_i) - \mu_B(x_i)]\, e^{\frac{\mu_A(x_i)}{\mu_B(x_i)}} + \sum_{i=1}^{n} [\mu_B(x_i) - \mu_A(x_i)]\, e^{\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}},$$

$$\text{i.e., } G_{\exp}^{\wedge}(A,B) = \sum_{i=1}^{n} [\mu_A(x_i) - \mu_B(x_i)] \left[e^{\frac{\mu_A(x_i)}{\mu_B(x_i)}} - e^{\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}} \right]. \tag{10}$$

Consequently, let $u_i > 0$ be the utilities of the events E_i with probabilities p_i and revised probabilities q_i respectively, for all $i = 1, 2, ..., n$. Then Useful information measure corresponding to new exponential divergence measure (7), will be

$$G_{\exp}(P,Q;u) = \frac{\sum_{i=1}^{n} u_i e^{\frac{p_i}{q_i}} (p_i - q_i)}{\sum_{i=1}^{n} u_i p_i}. \tag{11}$$

If utilities are ignored, i.e., $u_i = 1$ for each i, then we obtain the as usual $G_{\exp}(P,Q)$. Fuzzy information measures 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, whereas Useful information measure are very useful to find utility of an event, i.e., an event is how much useful compare to other event.

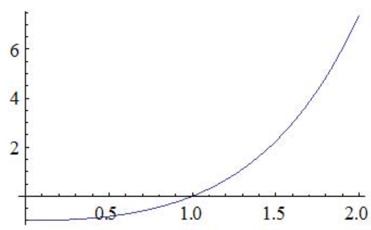

FIGURE 1. Convex function $f_1(t)$

3. Bounds of new divergence measure

To estimate the new exponential divergence $G_{\exp}(P,Q)$, it would be very interesting to establish some upper and lower bounds. So in this section, we obtain bounds of the exponential divergence measure (7) in terms of other symmetric and non- symmetric divergence measures.

Proposition 3.1. *Let $P, Q \in \Gamma_n$ and $0 < \alpha \leq 1 < \beta < \infty$, then we have*

$$\frac{e^{\alpha}(1+\alpha)^4}{8}\Delta(P,Q) \leq G_{\exp}(P,Q) \leq \frac{e^{\beta}(1+\beta)^4}{8}\Delta(P,Q), \tag{12}$$

where $G_{\exp}(P,Q)$ and $\Delta(P,Q)$ are given by (7) and (15) respectively.

Proof. Let us consider

$$f_2(t) = \frac{(t-1)^2}{t+1}, t \in (0,\infty) \tag{13}$$

and

$$f_2'(t) = \frac{(t-1)(t+3)}{(t+1)^2},$$
$$f_2''(t) = \frac{8}{(t+1)^3}. \tag{14}$$

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), we get

$$C_{f_2}(P,Q) = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{p_i + q_i} = \Delta(P,Q), \tag{15}$$

where $\Delta(P,Q)$ is called the Triangular discrimination [14].
Now, let

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

where $f_1''(t)$ and $f_2''(t)$ are given by (6) and (14) respectively and

$$g'(t) = \frac{e^t(1+t)^3(5+t)}{8}.$$

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

$$m = \inf_{t \in (\alpha,\beta)} g(t) = g(\alpha) = \frac{e^{\alpha}(1+\alpha)^4}{8} \tag{16}$$

and

$$M = \sup_{t \in (\alpha,\beta)} g(t) = g(\beta) = \frac{e^{\beta}(1+\beta)^4}{8}. \tag{17}$$

The result (12) is obtained by using (7), (15), (16), and (17) in inequality (4). □

Proposition 3.2. *Let $P, Q \in \Gamma_n$ and $0 < \alpha \leq 1 < \beta < \infty$, then we have*

$$4e^{\alpha}(1+\alpha)\alpha^{\frac{3}{2}}h(P,Q) \leq G_{\exp}(P,Q) \leq 4e^{\beta}(1+\beta)\beta^{\frac{3}{2}}h(P,Q), \tag{18}$$

where $G_{\exp}(P,Q)$ and $h(P,Q)$ are given by (7) and (21) respectively.

Proof. Let us consider

$$f_2(t) = \frac{\left(1-\sqrt{t}\right)^2}{2}, t \in (0,\infty) \tag{19}$$

and

$$f_2'(t) = -\frac{\left(1-\sqrt{t}\right)}{2\sqrt{t}},$$
$$f_2''(t) = \frac{1}{4t^{\frac{3}{2}}}. \tag{20}$$

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), we get

$$C_{f_2}(P,Q) = \sum_{i=1}^{n} \frac{\left(\sqrt{p_i}-\sqrt{q_i}\right)^2}{2} = h(P,Q), \tag{21}$$

where $h(P,Q)$ is called the Hellinger discrimination or Kolmogorov's divergence [19].

Now, let

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

where $f_1''(t)$ and $f_2''(t)$ are given by (6) and (20) respectively and

$$g'(t) = 2e^{t}\sqrt{t}(3+t)(2t+1).$$

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

$$m = \inf_{t\in(\alpha,\beta)} g(t) = g(\alpha) = 4e^{\alpha}(1+\alpha)\alpha^{\frac{3}{2}} \tag{22}$$

and

$$M = \sup_{t\in(\alpha,\beta)} g(t) = g(\beta) = 4e^{\beta}(1+\beta)\beta^{\frac{3}{2}}. \tag{23}$$

The result (18) is obtained by using (7), (21), (22), and (23) in inequality (4). □

In a similar procedure, we obtain the bounds of $G_{\exp}(P,Q)$ in terms of the other well known divergence measures. The results are as follows.

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

$$2e^{\alpha}\alpha(1+\alpha)^2 I(P,Q) \leq G_{\exp}(P,Q) \leq 2e^{\beta}\beta(1+\beta)^2 I(P,Q), \tag{24}$$

where

$$I(P,Q) = \frac{1}{2}\left[\sum_{i=1}^{n} p_i \log\frac{2p_i}{p_i+q_i} + \sum_{i=1}^{n} q_i \log\frac{2q_i}{p_i+q_i}\right] \tag{25}$$

is the Jensen- Shannon divergence or Information radius [9, 43].
(b) If $f_2(t) = (t-1)\log t$, then we have

$$e^{\alpha}\alpha^2 J(P,Q) \le G_{\exp}(P,Q) \le e^{\beta}\beta^2 J(P,Q), \tag{26}$$

where

$$J(P,Q) = \sum_{i=1}^{n} (p_i - q_i)\log\frac{p_i}{q_i} \tag{27}$$

is the J- divergence or Jeffrey- Kullback divergence [24, 31].
(c) If $f_2(t) = \frac{(t-1)^2}{\sqrt{t}}$, then we have

$$\frac{4e^{\alpha}(1+\alpha)\alpha^{\frac{5}{2}}}{3\alpha^2+2\alpha+3}E(P,Q) \le G_{\exp}(P,Q) \le \frac{4e^{\beta}(1+\beta)\beta^{\frac{5}{2}}}{3\beta^2+2\beta+3}E(P,Q), \tag{28}$$

where

$$E(P,Q) = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{\sqrt{p_i q_i}} \tag{29}$$

is the Jain- Srivastava divergence [23].
(d) If $f_2(t) = \frac{(t-1)^2(t+1)}{t}$, then we have

$$\frac{e^{\alpha}\alpha^3}{2(\alpha^2-\alpha+1)}\psi(P,Q) \le G_{\exp}(P,Q) \le \frac{e^{\beta}\beta^3}{2(\beta^2-\beta+1)}\psi(P,Q), \tag{30}$$

where

$$\psi(P,Q) = \sum_{i=1}^{n} \frac{(p_i - q_i)^2(p_i+q_i)}{p_i q_i} \tag{31}$$

is the Symmetric chi- square divergence [17].
(e) If $f_2(t) = \left(\frac{t+1}{2}\right)\log\frac{t+1}{2\sqrt{t}}$, then we have

$$\frac{4e^{\alpha}\alpha^2(1+\alpha)^2}{1+\alpha^2}T(P,Q) \le G_{\exp}(P,Q) \le \frac{4e^{\beta}\beta^2(1+\beta)^2}{1+\beta^2}T(P,Q), \tag{32}$$

where

$$T(P,Q) = \sum_{i=1}^{n} \left(\frac{p_i+q_i}{2}\right)\log\frac{p_i+q_i}{2\sqrt{p_i q_i}} \tag{33}$$

is the Arithmetic- Geometric mean divergence [45].
(f) If $f_2(t) = \frac{\left(t^2-1\right)^2}{2t^{\frac{3}{2}}}$, then we have

$$\frac{8e^{\alpha}\alpha^{\frac{7}{2}}(1+\alpha)}{15\alpha^4+2\alpha^2+15}\psi_M(P,Q) \le G_{\exp}(P,Q) \le \frac{8e^{\beta}\beta^{\frac{7}{2}}(1+\beta)}{15\beta^4+2\beta^2+15}\psi_M(P,Q), \tag{34}$$

where

$$\psi_M(P,Q) = \sum_{i=1}^{n} \frac{\left(p_i^2-q_i^2\right)^2}{2(p_i q_i)^{\frac{3}{2}}} \tag{35}$$

is the Kumar- Johnson divergence [32].
(g) If $f_2(t) = (t-1)\log\frac{t+1}{2}$, then we have

$$\frac{e^{\alpha}(1+\alpha)^3}{\alpha+3} J_R(P,Q) \le G_{\exp}(P,Q) \le \frac{e^{\beta}(1+\beta)^3}{\beta+3} J_R(P,Q), \tag{36}$$

where

$$J_R(P,Q) = \sum_{i=1}^{n} (p_i - q_i)\log\left(\frac{p_i+q_i}{2q_i}\right) \tag{37}$$

is the Relative J- divergence [16].
(h) If $f_2(t) = t\log t$, then we have

$$\alpha(1+\alpha)e^{\alpha}K(P,Q) \le G_{\exp}(P,Q) \le \beta(1+\beta)e^{\beta}K(P,Q), \tag{38}$$

where

$$K(P,Q) = \sum_{i=1}^{n} p_i \log\frac{p_i}{q_i} \tag{39}$$

is the Kullback- Leibler divergence or Relative entropy or Directed divergence or Information gain [31].
(i) If $f_2(t) = \left(\frac{t+1}{2}\right)\log\frac{t+1}{2t}$, then we have

$$2\alpha^2(1+\alpha)^2 e^{\alpha}G(P,Q) \le G_{\exp}(P,Q) \le 2\beta^2(1+\beta)^2 e^{\beta}G(P,Q), \tag{40}$$

where

$$G(P,Q) = \sum_{i=1}^{n}\left(\frac{p_i+q_i}{2}\right)\log\frac{p_i+q_i}{2p_i} \tag{41}$$

is the Relative Arithmetic- Geometric divergence [45].
(j) If $f_2(t) = (t-1)^2$, then we have

$$\frac{e^{\alpha}(1+\alpha)}{2}\chi^2(P,Q) \le G_{\exp}(P,Q) \le \frac{e^{\beta}(1+\beta)}{2}\chi^2(P,Q), \tag{42}$$

where

$$\chi^2(P,Q) = \sum_{i=1}^{n}\frac{(p_i-q_i)^2}{q_i} \tag{43}$$

is the Chi- square divergence or Pearson divergence [37].
(k) If $f_2(t) = t\log\frac{2t}{t+1}$, then we have

$$\alpha e^{\alpha}(1+\alpha)^3 F(P,Q) \le G_{\exp}(P,Q) \le \alpha e^{\beta}(1+\beta)^3 F(P,Q), \tag{44}$$

where

$$F(P,Q) = \sum_{i=1}^{n} p_i\log\frac{2p_i}{p_i+q_i} \tag{45}$$

is the Relative Jensen- Shannon divergence [43].
(l) If $f_2(t) = \frac{(t^2-1)^2}{t}$, then we have

$$\frac{e^{\alpha}(\alpha+1)\alpha^3}{2(3\alpha^4+1)}\gamma(P,Q) \le G_{\exp}(P,Q) \le \frac{e^{\beta}(\beta+1)\beta^3}{2(3\beta^4+1)}\gamma(P,Q), \tag{46}$$

where

$$\gamma\left(P,Q\right)=\sum_{i=1}^{n}\frac{\left(p_i^2-q_i^2\right)^2}{p_iq_i^2} \tag{47}$$

is the Jain and Chhabra divergence [21].

Remark 3.1. Divergences (15), (21), (25), (27), (29), (31), (33), (35) are symmetric and divergences (37), (39), (41), (43), (45), (47) are non- symmetric with respect to probability distributions $P,Q\in\Gamma_n$.

4. Numerical verification of bounds

In this section, we take an example for calculating the divergences $\Delta\left(P,Q\right)$, $h\left(P,Q\right), G\left(P,Q\right)$ and $G_{exp}\left(P,Q\right)$ and verify numerically the inequalities (12), (18), and (40) or verify the bounds of $G_{exp}\left(P,Q\right)$.

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

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

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

By using Table 1, we get the followings.

$$\alpha\left(=.00647\right)\leq\frac{p_i}{q_i}\leq\beta\left(=1.792\right). \tag{48}$$

$$\Delta\left(P,Q\right)=\sum_{i=1}^{11}\frac{\left(p_i-q_i\right)^2}{p_i+q_i}\approx .1812. \tag{49}$$

$$h\left(P,Q\right)=\sum_{i=1}^{11}\frac{\left(\sqrt{p_i}-\sqrt{q_i}\right)^2}{2}\approx .0502. \tag{50}$$

$$G\left(P,Q\right)=\sum_{i=1}^{11}\frac{p_i+q_i}{2}\log\left(\frac{p_i+q_i}{2p_i}\right)\approx .0746. \tag{51}$$

$$G_{\text{exp}}\left(P,Q\right)=\sum_{i=1}^{11}e^{\frac{p_i}{q_i}}\left(p_i-q_i\right)\approx .97971. \tag{52}$$

Put the approximated values from (48) to (52) in inequalities (12), (18), and (40) respectively and get the following results

$$\begin{aligned}.0233\leq .97971\leq 8.260, 1.508\times10^{-4}\leq .97971\leq 8.071, 6.367\times10^{-6}\\ \leq .97971\leq 22.414\end{aligned} \tag{53}$$

respectively. Hence verified the bounds of $G_{\exp}(P,Q)$ in terms of $\Delta(P,Q), h(P,Q)$ and $G(P,Q)$ for $p = 0.7$.

Similarly, we can verify the bounds of $G_{\exp}(P,Q)$ in terms of 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.

In Figure 2, we have considered $p_i = (a, 1-a), q_i = (1-a, a)$, where $a \in (0,1)$.

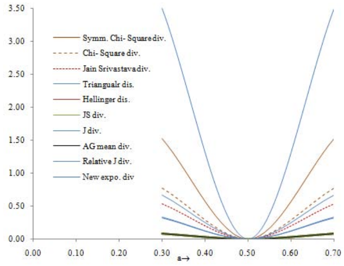

FIGURE 2. Comparison of divergence measures with new exponential divergence measure

It is clear from Figure 2 that the new exponential divergence $G_{\exp}(P,Q)$ has a steeper slope than $\psi(P,Q), \chi^2(P,Q), E(P,Q), \Delta(P,Q), h(P,Q), I(P,Q), J(P,Q)$, $T(P,Q)$, and $J_R(P,Q)$.

5. Conclusion and discussion

To design a communication system with a specific message handling capability, we need a measure of information content to be transmitted. Divergence measures are for quantifying the dissimilarity among probability distributions. In this work we introduced a new exponential divergence measure and obtained the bounds by using Csiszar's information inequality and verified the bounds numerically as well in the interval (α,β), $0 < \alpha \leq 1 < \beta < \infty$. Fuzzy exponential information measure and Useful exponential information measure also introduced. Work on further generalizations of this new divergence measure is in progress and will be reported elsewhere, like: Application to the mutual information, other relations by using standard algebraic and exponential inequalities, square root of this new measure is a metric space etc.

We hope that this work will motivate the reader to consider the extensions of divergence measures in information theory, other problems of functional analysis and fuzzy mathematics.

Acknowledgements

The authors are greatly thankful to the reviewer/reviewers for his/her precious comments and suggestions for the enhancement of this paper.

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, **6** (2010), 31- 44.
2. M.B. Bassat, *f- Entropies, probability of error and feature selection*, Inform. Control, **39** (1978), 227-242.
3. M. Basseville, *Distance measures for signal processing and pattern recognition*, Signal Processing, **18** (1989), 349- 369.
4. A. Benveniste, M. Basseville and G. Moustakides, *The asymptotic local approach to change detection and model validation*, IEEE Trans. Automatic Control, **AC-32** (1987), 583- 592.
5. A. Bhattacharyya, *On a measure of divergence between two multinomial populations*, Sankhaya: The Indian Journal of Statistics (1933- 1960), **7** (1946), 401- 406.
6. J.S. Bhullar, O.P. Vinocha and M. Gupta, *Generalized measure for two utility distributions*, Proceedings of the World Congress on Engineering, **3** (2010).
7. D.E. Boekee and J.C.A. Van Der Lubbe, *Some aspects of error bounds in feature selection*, Pattern Recognition, **11** (1979), 353- 360.
8. L.M. Bregman, *The relaxation method to find the common point of convex sets and its applications to the solution of problems in convex programming*, USSR Comput. Math. Phys., **7** (1967), 200-217.
9. 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), 489-495.
10. H.C. Chen, *Statistical pattern recognition*, Hoyderc Book Co., Rocelle Park, New York, (1973).
11. C.K. Chow and C.N. Lin, *Approximating discrete probability distributions with dependence trees*, IEEE Trans. Inform. Theory, **14** (1968), 462-467.
12. I. Csiszar, *Information measures: A Critical survey*, in Trans. In: Seventh Prague Conf. on Information Theory, Academia, Prague, (1974), 73-86.
13. I. Csiszar, *Information type measures of differences of probability distribution and indirect observations*, Studia Math. Hungarica, **2** (1967), 299- 318.
14. D. Dacunha- Castelle, *Ecole dEte de Probabilites de*, Saint-Flour VII-1977, Berlin, Heidelberg, New York: Springer, (1978).
15. S.S. Dragomir, *A generalized f- divergence for probability vectors and applications*, Research Report Collection **5** (2000).
16. S.S. Dragomir, V. Gluscevic and C.E.M. Pearce, *Approximation for the Csiszars f- divergence via midpoint inequalities*, in Inequality Theory and Applications - Y.J. Cho, J.K. Kim, and S.S. Dragomir (Eds.), Nova Science Publishers, Inc., Huntington, New York, **1** (2001), 139-154.
17. S.S. Dragomir, J. Sunde and C. Buse, *New inequalities for Jeffreys divergence measure*, Tamusi Oxford Journal of Mathematical Sciences, **16** (2000), 295-309.
18. D.V. Gokhale and S. Kullback, *Information in contingency Tables*, New York, Marcel Dekker, (1978).
19. E. Hellinger, *Neue begrundung der theorie der quadratischen formen von unendlichen vielen veranderlichen*, J. Rein.Aug. Math., **136** (1909), 210-271.
20. D.S. Hooda, *On generalized measures of fuzzy entropy*, Mathematica Slovaca, **54** (2004), 315- 325.
21. K.C. Jain and P. Chhabra, *New series of information divergence measures and their properties*, Accepted in Applied Mathematics and Information Sciences.
22. K.C. Jain and R.N. Saraswat, *Some new information inequalities and its applications in information theory*, International Journal of Mathematics Research, **4** (2012), 295- 307.
23. K.C. Jain and A. Srivastava, *On symmetric information divergence measures of Csiszar's f- divergence class*, Journal of Applied Mathematics, Statistics and Informatics, **3** (2007), 85-102.

24. H. Jeffreys, *An invariant form for the prior probability in estimation problem*, Proc. Roy. Soc. Lon. Ser. A, **186** (1946), 453-461.
25. 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, **3** (2012), 1-5.
26. 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, **36** (1990), 23-30.
27. T.T. Kadota and L.A. Shepp, *On the best finite set of linear observables for discriminating two Gaussian signals*, IEEE Trans. Inform. Theory, **13** (1967), 288-294.
28. T. Kailath, *The divergence and Bhattacharyya distance measures in signal selection*, IEEE Trans. Comm. Technology, **COM-15** (1967), 52- 60.
29. D. Kazakos and T. Cotsidas, *A decision theory approach to the approximation of discrete probability densities*, IEEE Trans. Perform. Anal. Machine Intell, **1** (1980), 61- 67.
30. A.N. Kolmogorov, *On the approximation of distributions of sums of independent summands by infinitely divisible distributions*, Sankhya, **25**, 159-174.
31. S. Kullback and R.A. Leibler, *On information and sufficiency*, Ann. Math. Statist., **22** (1951), 79-86.
32. P. Kumar and A. Johnson, *On a symmetric divergence measure and information inequalities*, Journal of Inequalities in Pure and Applied Mathematics, **6** (2005), 1-13.
33. P.W. Lamberti, A.P. Majtey, A. Borras, M. Casas and A. Plastino, *Metric character of the quantum Jensen- Shannon divergence*, Physical Review A, **77** (2008), 052311.
34. F. Nielsen and S. Boltz, *The Burbea-Rao and Bhattacharyya centroids*, Apr. (2010), Arxiv.
35. M.A. Nielsen and I.L. Chuang, *Quantum computation and information*, Cambridge University Press, Cambridge, UK, **3** (2000), 9.
36. F. Osterreicher, *Csiszar's f- divergence basic properties*, Homepage: http://www.sbg.ac.at/mat/home.html, November 22, (2002).
37. 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.
38. E.C. Pielou, *Ecological diversity*, New York, Wiley, (1975).
39. H.V. Poor, *Robust decision design using a distance criterion*, IEEE Trans. Inf. Th., **IT 26** (1980), 575- 587.
40. A. Renyi, *On measures of entropy and information*, Proc. 4th Berkeley Symposium on Math. Statist. and Prob., **1** (1961), 547-561.
41. M. Salicru, *Measures of information associated with Csiszar's divergences*, Kybernetika, **30** (1994), 563- 573.
42. 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), 551- 556.
43. R. Sibson, *Information radius*, Z. Wahrs. Undverw. Geb., **1** (1969), 149-160.
44. H.C. Taneja and R.K. Tuteja, *Characterization of a quantitative- qualitative measure of inaccuracy*, Kybernetika, **22** (1986), 393- 402.
45. I.J. Taneja, *New developments in generalized information measures*, Chapter in: Advances in Imaging and Electron Physics, Ed. P.W. Hawkes, **91** (1995), 37-135.
46. I.J. Taneja, *Generalized symmetric divergence measures and inequalities*, RGMIA Research Report Collection, http://rgmia.vu.edu.au, **7**(2004), Art. 9. Available on-line at: arXiv:math.ST/0501301 v1 19 Jan 2005.
47. B. Taskar, S. Lacoste-Julien and M.I. Jordan, *Structured prediction, dual extra gradient and Bregman projections*, Journal of Machine Learning Research, **7** (2006), 1627- 1653.
48. H. Theil *Statistical decomposition analysis*, Amsterdam, North-Holland, 1972.

49. H. Theil, *Economics and information theory*, Amsterdam, North-Holland, 1967.
50. K. Tumer and J. Ghosh, *Estimating the Bayes error rate through classifier combining*, Proceedings of 13th International Conference on Pattern Recognition, (1996), 695- 699.
51. B. Vemuri, M. Liu, S. Amari and F. Nielsen, *Total Bregman divergence and its applications to DTI analysis*, IEEE Transactions on Medical Imaging, (2010).

K.C. Jain received the M.Sc and Ph.D degrees from the University of Rajasthan, India in 1973 and 1979 respectively. He joined the Malaviya Regional Engineering College, Jaipur (Rajasthan), India (Now Malaviya National Institute of Technology, Jaipur) as a lecturer in 1977. He was appointed professor in 2006, Department of Mathematics, MNIT, Jaipur. His research area is Information Divergence Measures. He delivered invited lectures at more than 30 places on the topics: Stochastic process and applications, Information Theory and Coding, Applications of Statistical tools, Testing of Hypothesis, Error-Correcting Codes, Better understanding of Grading System, and Continuous analogue of Shannons Entropy etc.

Department of Mathematics, Malaviya National Institute of Technology, Jaipur (Rajasthan) 302017, India.
e-mail: jainkc_2003@yahoo.com

Praphull Chhabra received his B.Sc, M.Sc and B.Ed degrees from the University of Rajasthan, India in 2008, 2010, and 2013 respectively. He has qualified the National Eligibility Test (NET)- June 2011 and Graduate Aptitude Test for Engineering (GATE)- Feb. 2011, in Mathematical sciences. Currently, he is doing Ph.D in Information Theory under the supervision of Prof. K.C. Jain, from the Department of Mathematics, Malaviya National Institute of Technology, Jaipur (Rajasthan), India. He is the Co- author of 20 research papers with Prof. K.C. Jain, out of them 16 have been published in Reputed International journals. He has presented his research work in more than 10 national and international conferences. He has attended more than 15 workshops, short term courses and seminars.

Department of Mathematics, Malaviya National Institute of Technology, Jaipur (Rajasthan) 302017, India.
e-mail: prfl24@gmail.com