Palestine Journal of Mathematics

Vol. 5(2) (2016) , 249–257

APPLICATIONS OF NEW EXPONENTIAL INFORMATION DIVERGENCE MEASURE

K.C. Jain and Praphull Chhabra

Communicated by Ayman Badawi

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

Keywords and phrases: Metric, Mutual information, Variational distance, New exponential divergence measure, Numerical verification.

Abstract Jain and Chhabra (2015) introduced and characterized a new exponential divergence measure. Also in the same literature, they introduced new fuzzy and useful information divergence measure corresponding to same exponential divergence. They also evaluated the upper and lower bounds of this new divergence in terms of various well known symmetric and non-symmetric divergence measures together with numerical verification.
In this work, we extends the previous work, i.e., we apply the new exponential divergence to the Mutual information and as a metric to the metric space. Bounds in terms of Variational distance and numerical verification are evaluated as well.

1 Introduction

Divergence measures are basically measures of distance between two probability distributions or compare two probability distributions. Divergence measures have been demonstrated very useful in a variety of disciplines such as Bayesian model validation (1996) [40], quantum information theory (2008, 2000) [26, 28], model validation (1987) [3], robust detection (1980) [31], economics and political science (1972, 1967) [38, 39], biology (1975) [30], analysis of contingency tables (1978) [15], approximation of probability distributions (1968, 1980) [9, 23], signal processing (1967, 1967) [21, 22], pattern recognition (1978, 1979, 1973, 1990) [2, 8, 20, 5], color image segmentation (2010) [27], 3D image segmentation and word alignment (2006) [37], cost- sensitive classification for medical diagnosis (2009) [34], magnetic resonance image analysis (2010) [41] etc.
Also we can use divergence measures in fuzzy mathematics as fuzzy directed divergences and fuzzy entropies (2010, 2004, 2012) [1, 16, 19], 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, 36], 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 (1974, 1967) [10, 11], Bregman's f- divergence (1967) [6], Burbea- Rao's f- divergence (1982) [7], Renyi's like f- divergence (1961) [32], M- divergence (1994) [33], Jain- Saraswat f- divergence (2012) [18] etc.
Csiszar's f- divergence is widely used due to its compact nature, which is given by

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

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 (2002) [29], in detail.

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.

Jain and Chhabra (2015) [17] introduced the following divergence measure which is exponential in nature and did a quality work on it in the same literature

$$G_{\text{exp}}(P, Q) = \sum_{i=1}^{n} e^{\frac{p_i}{q_i}} (p_i - q_i). \tag{1.2}$$

We see that $G_{\text{exp}}(P, Q)$ 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 $G_{\text{exp}}(P, Q)$ is non- symmetric divergence w.r.t. P and Q because $G_{\text{exp}}(P, Q) \neq G_{\text{exp}}(Q, P)$.
In this paper, we introduce two important applications of this exponential divergence to the information and statistical theory, one is to the Mutual information (section 4) and second is as a Metric space (section 2). In section 3, we obtain bounds of this new divergence in terms of Variational distance and in section 5, verification of these bounds is done numerically.

2 Application as a metric space

We know that $G_{\text{exp}}(P, Q)$ is non- symmetric but

$$\begin{aligned} G_{\text{exp}}(P, Q) + G_{\text{exp}}(Q, P) &= \sum_{i=1}^{n} (p_i - q_i) e^{\frac{p_i}{q_i}} + \sum_{i=1}^{n} (q_i - p_i) e^{\frac{q_i}{p_i}} \\ &= \sum_{i=1}^{n} (p_i - q_i) \left(e^{\frac{p_i}{q_i}} - e^{\frac{q_i}{p_i}} \right) = G^*_{\text{exp}}(P, Q). \end{aligned} \tag{2.1}$$

is symmetric with respect to probability distributions $P, Q \in \Gamma_n$, as $G^*_{\text{exp}}(P, Q) = G^*_{\text{exp}}(Q, P)$.
We can see that $\sqrt{G^*_{\text{exp}}(P, Q)} > 0$ and $= 0$ if and only if $P = Q$ or $p_i = q_i \ \forall \ i = 1, 2, 3..., n$. The $\sqrt{G^*_{\text{exp}}(P, Q)}$ is symmetric because $G^*_{\text{exp}}(P, Q)$ is symmetric or $\sqrt{G^*_{\text{exp}}(P, Q)} = \sqrt{G^*_{\text{exp}}(Q, P)}$.
In this section we prove that $\sqrt{G^*_{\text{exp}}(P, Q)}$ satisfies triangle inequality and then obtain a new exponential metric space over an interval $(0, \infty)$. For this, we prove the following theorem, which is stated as

Theorem 2.1. *Let $x(p, q) : (0, \infty) \times (0, \infty) \to (0, \infty)$ be defined as*

$$x(p, q) = (p - q) \left(e^{\frac{p}{q}} - e^{\frac{q}{p}} \right), \tag{2.2}$$

i.e., we can write

$$G^*_{\text{exp}}(P, Q) = \sum_{i=1}^{n} x(p_i, q_i). \tag{2.3}$$

Then triangle inequality will be

$$\sqrt{x\left(p,q\right)} \leq \sqrt{x\left(p,r\right)} + \sqrt{x\left(r,q\right)}, \tag{2.4}$$

where $p, q, r \in (0, \infty)$ *and* $G^{*}_{\text{exp}}\left(P,Q\right)$ *is given by (2.1).*

Proof: To prove the inequality (2.4), first let us consider

$$X_{pq}\left(r\right) = \sqrt{x\left(p,r\right)} + \sqrt{x\left(r,q\right)}, \tag{2.5}$$

then

$$\frac{d}{dr}X_{pq}\left(r\right) = X'_{pq}\left(r\right) = \frac{x'\left(p,r\right)}{2\sqrt{x\left(p,r\right)}} + \frac{x'\left(r,q\right)}{2\sqrt{x\left(r,q\right)}}. \tag{2.6}$$

Now from (2.2), we can write

$$x\left(p,r\right) = \left(p-r\right)\left(e^{\frac{p}{r}} - e^{\frac{r}{p}}\right) \tag{2.7}$$

and after differentiating (2.7) w.r.t r, we obtain

$$x'\left(p,r\right) = -\left[e^{\frac{p}{r}}\left(\frac{p^2-pr+r^2}{r^2}\right) - \frac{re^{\frac{r}{p}}}{p}\right]. \tag{2.8}$$

Put $p = rt$, i.e., $t = \frac{p}{r} \in (0,\infty)$ in (2.8), we get

$$\left[x'\left(p,r\right)\right]_{p=rt} = k\left(t\right) = \frac{e^{\frac{1}{t}} - te^{t}\left(t^2-t+1\right)}{t}. \tag{2.9}$$

Now from (2.7), we can write

$$x\left(t,1\right) = \left(t-1\right)\left(e^{t} - e^{\frac{1}{t}}\right). \tag{2.10}$$

From (2.7) and (2.10), we have the following relation for $p = rt$

$$\sqrt{x\left(p,r\right)} = \sqrt{r}\sqrt{x\left(t,1\right)} = \sqrt{r}\, l\left(t\right), \tag{2.11}$$

where we are assuming

$$\sqrt{x\left(t,1\right)} = l\left(t\right). \tag{2.12}$$

Now, differentiate (2.9) w.r.t t, we obtain

$$k'\left(t\right) = -\left[\frac{\left(t+1\right)\left(t^4e^{t} + e^{\frac{1}{t}}\right)}{t^3}\right]. \tag{2.13}$$

Now, let we define a function

$$s\left(t\right) = \frac{k\left(t\right)}{l\left(t\right)},\ \forall\, t \in (0,\infty). \tag{2.14}$$

From (2.10) and (2.13), we can see that $l\left(t\right) = \sqrt{x\left(t,1\right)} \geq 0$ and $k'\left(t\right) < 0\ \forall\ t \in (0,\infty)$, i.e., $k\left(t\right)$ is monotonically decreasing function and $k\left(1\right) = 0$, so $s\left(t\right)$ will be decreasing as well in $(0,\infty)$ with $\lim_{t\to 1} s\left(t\right) = 0$ or the nature of $s\left(t\right)$ depends on the nature of $k\left(t\right)$ only as $l\left(t\right)$ is fix and positive. Therefore, we conclude that $s\left(t\right)$ changes the sign at $t = 1$, so

$$s\left(t\right) = \begin{cases} > 0 & \text{if } t < 1 \\ < 0 & \text{if } t > 1 \\ = 0 & \text{if } t \to 1 \end{cases}. \tag{2.15}$$

Now suppose $u = \frac{q}{p} \in (0, \infty) \Rightarrow \frac{q}{r} = \frac{q}{p}\frac{p}{r} = ut \in (0, \infty)$, so (2.6) can be written as

$$2\sqrt{r}X'_{pq}(r) = s(t) + s(ut). \tag{2.16}$$

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 (2.4) has its minimum value at $p = q = r\ \forall\, p, q, r \in (0, \infty)$.
Hence proof the result (2.4) or theorem 2.1.
In view of this proof, we conclude that the new exponential symmetric divergence measure $\sqrt{G^*_{\exp}(P, Q)}$ is a metric or we obtain a new metric space $\sqrt{\left(G^*_{\exp}, (0, \infty)\right)}$ over $(0, \infty)$.

3 Bounds in terms of Variational distance

We had obtained bounds of new exponential divergence in terms of the various well known symmetric and non- symmetric divergences in last literature (2015) [17]. Now in this section, we obtain bounds in terms of Variational distance. For getting the bounds, we consider the following theorem from literature (2001) [12].

Theorem 3.1. *Let $f_1, f_2 : (\alpha, \beta) \subset (0, \infty) \to R$ be two real, convex and normalized differentiable functions, i.e., $f_1''(t), f_2''(t) \geq 0\ \forall\ t > 0$ and $f_1(1) = f_2(1) = 0$ respectively with $0 < \alpha \leq 1 \leq \beta < \infty, \alpha \neq \beta$. If there exists the real constants m, M such that $m < M$ and*

$$m \leq \frac{|f_1(t_1) - f_1(t_2)|}{|f_2(t_1) - f_2(t_2)|} \leq M,$$

i.e.,

$$m \leq \frac{|f_1'(t)|}{|f_2'(t)|} = \left|\frac{f_1'(t)}{f_2'(t)}\right| \leq M, \tag{3.1}$$

for all $t_1, t_2 \in (\alpha, \beta) \subset (0, \infty)$.
If $P, Q \in \Gamma_n$, then we have the following inequalities

$$mC_{|f_2|}(P, Q) \leq C_{|f_1|}(P, Q) \leq MC_{|f_2|}(P, Q), \tag{3.2}$$

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

Now by using the above theorem, we will obtain the bounds of $G_{\exp}(P, Q)$ in terms of $V(P, Q)$.

Proposition 3.2. *Let $G_{\exp}(P, Q)$ and $V(P, Q)$ be defined as in (1.2) and (3.5) respectively. For $P, Q \in \Gamma_n$, we have*

$$\alpha e^{\alpha} V(P, Q) \leq \left|G_{\exp}\right|(P, Q) \leq \beta e^{\beta} V(P, Q). \tag{3.3}$$

Proof: Let us consider

$$f_1(t) = e^t(t-1), f_2(t) = |t-1|\ \forall\, t \in (0, \infty),$$

$$f_1'(t) = te^t, f_2'(t) = \begin{cases} -1 & \text{if } 0 < t < 1 \\ 1 & \text{if } 1 \leq t < \infty \end{cases}$$

and

$$f_1''(t) = e^t(t+1), f_2''(t) = 0.$$

We can see that both functions $f_1(t), f_2(t)$ are convex and normalized because $f_1''(t), f_2''(t) \geq 0 \; \forall \, t > 0$ and $f_1(1) = 0 = f_2(1)$ respectively.
Now put $f_1(t), f_2(t)$ in (1.1), we obtain the followings

$$C_{|f_1|}(P,Q) = \sum_{i=1}^{n} e^{\frac{p_i}{q_i}} |p_i - q_i| = |G_{\exp}|(P,Q) \tag{3.4}$$

and

$$C_{|f_2|}(P,Q) = \sum_{i=1}^{n} |p_i - q_i| = V(P,Q) \tag{3.5}$$

respectively. Where $V(P,Q)$ is well known Variational distance or l_1 distance (1963) [24].
Now, let $g(t) = \frac{|f_1'(t)|}{|f_2'(t)|} = \left|\frac{f_1'(t)}{f_2'(t)}\right| = |te^t| = te^t$, where $|f_2'(t)| = 1$ and $g'(t) = e^t(t+1) > 0$.
It is clear that $g(t)$ is strictly increasing in $(0,\infty)$, so

$$m = \inf_{t\in(\alpha,\beta)} g(t) = g(\alpha) = \alpha e^{\alpha}. \tag{3.6}$$

$$M = \sup_{t\in(\alpha,\beta)} g(t) = g(\beta) = \beta e^{\beta}. \tag{3.7}$$

The result (3.3) is obtained by using (3.4), (3.5), (3.6), and (3.7) in (3.2).

4 Application to the Mutual information

Mutual information (1948) [35] is a measure of amount of information that one random variable contains about another or amount of information conveyed about one random variable by another.
Let X and Y be two discrete random variables with a joint probability mass function $p(x_i, y_j) = p_{ij}$ with $i = 1,2,...,m; j = 1,2,...,n$ and marginal probability mass functions $p(x_i) = \sum_{j=1}^{n} p(x_i, y_j), i = 1,2,...,m$ and $p(y_j) = \sum_{i=1}^{m} p(x_i, y_j), j = 1,2,...,n$, where $x_i \in X, y_j \in Y$, then Mutual information $I(X,Y)$ is defined by

$$I(X,Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} p(x_i,y_j)\log\frac{p(x_i,y_j)}{p(x_i)\,p(y_j)} = \sum_{(x,y)\in(X,Y)} p(x,y)\log\frac{p(x,y)}{p(x)\,p(y)}. \tag{4.1}$$

Since $I(X,Y)$ is symmetric in X, Y therefore it can also be written as

$$I(X,Y) = I(Y,X) = H(X) - H\left(\frac{X}{Y}\right) = H(Y) - H\left(\frac{Y}{X}\right), \tag{4.2}$$

where

$$H(X) = -\sum_{i=1}^{m} p(x_i)\log p(x_i) = -\sum_{i=1}^{m}\sum_{j=1}^{n} p(x_i,y_j)\log\left(\sum_{j=1}^{n} p(x_i,y_j)\right) \tag{4.3}$$

is known as Marginal entropy (1948) [35] and

$$H\left(\frac{X}{Y}\right) = -\sum_{i=1}^{m}\sum_{j=1}^{n} p(x_i,y_j)\log p\left(\frac{x_i}{y_j}\right) \tag{4.4}$$

is known as Conditional entropy (1948) [35].
By viewing

$$\begin{aligned} K(P,Q) &= \sum_{i=1}^{n} p_i \log\frac{p_i}{q_i} \\ &= \text{Kullback- Leibler divergence or Relative entropy}(1951)[25], \end{aligned} \tag{4.5}$$

we can say that the Mutual information is nothing but a Relative entropy between joint distribution $p(x,y)$ and product of marginal distributions $p(x)$ and $p(y)$ after replacing $p(x)$ and $q(x)$ by $p(x,y)$ and $p(x)p(y)$ respectively in (4.5). So $I(X,Y)$ can also be written as

$$I(X,Y) = K(p(x,y), p(x)p(y)) = \sum_{(x,y)\in(X,Y)} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}. \tag{4.6}$$

Similarly, we can define the Mutual information in following manners as well.
In $\left|G_{\exp}\right|(P,Q)$ manner:

$$I_{|G_{\exp}|}(X,Y) = \sum_{(x,y)\in(X,Y)} |p(x,y) - p(x)p(y)|\, e^{\frac{p(x,y)}{p(x)p(y)}}, \tag{4.7}$$

In $V(P,Q)$ manner:

$$I_V(X,Y) = \sum_{(x,y)\in(X,Y)} |p(x,y) - p(x)p(y)| \tag{4.8}$$

and
In $J_R(P,Q)$ manner:

$$I_{J_R}(X,Y) = \sum_{(x,y)\in(X,Y)} [p(x,y) - p(x)p(y)] \log \frac{p(x,y) + p(x)p(y)}{2p(x)p(y)}, \tag{4.9}$$

where $\left|G_{\exp}\right|(P,Q), V(P,Q)$ are given by (3.4) and (3.5) respectively and

$$J_R(P,Q) = \sum_{i=1}^{n} (p_i - q_i) \log\left(\frac{p_i + q_i}{2q_i}\right) = \text{Relative J- Divergence}(2001)[14]. \tag{4.10}$$

Equations (4.6) to (4.9) tell us that how far the joint distribution is from its independency or $I(X,Y) = 0 = I_{|G_{\exp}|}(X,Y) = I_V(X,Y) = I_{J_R}(X,Y)$ if distributions are independent to each other.
Now, the following theorem can be seen in literature (1999) [13].

Theorem 4.1. *Let $f : [\alpha,\beta] \subset (0,\infty) \to R$ be a convex twice differentiable function which is normalized, i.e., $f(1) = 0$ and f' is of bounded variation on $[\alpha,\beta]$, i.e., $A_\alpha^\beta(f') = \int_\alpha^\beta |f''(t)|\,dt < \infty$.*
If $P,Q \in \Gamma_n$ for some α and β with $0 < \alpha \le 1 \le \beta < \infty, \alpha \ne \beta$, then we have the following inequality

$$\left|C_f(P,Q) - E^*_{C_f}(P,Q)\right| \le A_\alpha^\beta(f')\, V(P,Q), \tag{4.11}$$

where $C_f(P,Q), V(P,Q)$ are given by (1.1) and (3.5) respectively and

$$E^*_{C_f}(P,Q) = \sum_{i=1}^{n} (p_i - q_i) f'\left(\frac{p_i + q_i}{2q_i}\right). \tag{4.12}$$

Now by using the above theorem, we introduce a new information inequalities which relates $I(X,Y)$ and $I_{|G_{\exp}|}(X,Y)$.

Proposition 4.2. *For $0 < \alpha \le \frac{p(x,y)}{p(x)p(y)} \le \beta < \infty\ \forall (x,y) \in (X,Y)$, we obtain the following new information inequalities in mutual information sense*

$$|I(X,Y) - I_{J_R}(X,Y)| \le \log\left(\frac{\beta}{\alpha}\right) I_V(X,Y) \le \frac{\log\frac{\beta}{\alpha}}{\alpha e^\alpha} I_{|G_{\exp}|}(X,Y), \tag{4.13}$$

where $I(X,Y), I_{|G_{\exp}|}(X,Y), I_V(X,Y)$, and $I_{J_R}(X,Y)$ are given by (4.6) to (4.9) respectively.

Proof: Let us consider

$$f(t) = t\log t, t \in (0,\infty), f(1) = 0, f'(t) = 1 + \log t$$

and

$$f''(t) = \frac{1}{t}. \tag{4.14}$$

Since $f''(t) > 0\ \forall\ t > 0$ and $f(1) = 0$, so $f(t)$ is strictly convex and normalized function respectively. Now put $f(t)$ in (1.1) and $f'(t)$ in (4.12) then after replacing $p_i, q_i\ \forall i = 1, 2, ..., n$ by $p(x,y), p(x)p(y)\ \forall (x,y) \in (X,Y)$, we get

$$C_f(P,Q) = I(X,Y), \tag{4.15}$$

$$E^*_{C_f}(P,Q) = I_{J_R}(X,Y), \tag{4.16}$$

respectively. Also we obtain

$$A^\beta_\alpha(f') = \int_\alpha^\beta |f''(t)|\,dt = \int_\alpha^\beta \frac{1}{t}dt = \log\beta - \log\alpha = \log\frac{\beta}{\alpha}. \tag{4.17}$$

The result (4.13) is obtained by using (4.15), (4.16), (4.17) together with first inequality of (3.3) in (4.11), after replacing p_i, q_i by $p(x,y), p(x)p(y)$ respectively.

5 Numerical verification of bounds

In this section, we take an example for calculating the divergences $|G_{exp}|(P,Q)$ and $V(P,Q)$ and verify numerically the inequalities (3.3) or verify the bounds of $|G_{exp}|(P,Q)$.

Example 5.1. 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.

$$\alpha(= .00647) \le \frac{p_i}{q_i} \le \beta(= 1.792). \tag{5.1}$$

$$\left|G_{\text{exp}}\right|(P,Q) = \sum_{i=1}^{11} e^{\frac{p_i}{q_i}}|p_i - q_i| \approx 1.78872. \tag{5.2}$$

$$V(P,Q) = \sum_{i=1}^{11} |p_i - q_i| \approx 0.4844. \tag{5.3}$$

Put the approximated values from (5.1) to (5.3) in inequalities (3.3) and get the following result

$$3.154 \times 10^{-3} \le 1.78872\left(= \left|G_{\text{exp}}\right|(P,Q)\right) \le 5.2095. \tag{5.4}$$

Hence verified the bounds of $\left|G_{\text{exp}}\right|(P,Q)$ in terms of $V(P,Q)$ for $p = 0.7$.

Similarly, we can verify the bounds of $\left|G_{\text{exp}}\right|(P,Q)$ for different values of p and q and for other discrete probability distributions as well, like; Negative binomial, Geometric, uniform etc.

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] 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., vol. 7, pp: 200- 217, 1967.

[7] 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.

[8] H.C. Chen, Statistical pattern recognition, Hoyderc Book Co., Rocelle Park, New York, 1973.

[9] 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.

[10] I. Csiszar, Information measures: A Critical survey, in Trans.- In: Seventh Prague Conf. on Information Theory, Academia, Prague, 1974, pp: 73-86.

[11] I. Csiszar, Information type measures of differences of probability distribution and indirect observations, Studia Math. Hungarica, vol. 2, pp: 299- 318, 1967.

[12] S.S. Dragomir, Some inequalities for two Csiszar's divergences and applications, Mat. Bilten (Macedonia), 25 (2001), pp: 73- 90.

[13] S.S. Dragomir, On the Ostrowski's integral inequality for mappings of bounded variation and applications, RGMIA Res. Rep. Coll. 2 (1999) No.1, 73- 80.

[14] S.S. Dragomir, V. Gluscevic, and C.E.M. Pearce, Approximation for the Csiszar's 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, Vol. 1, 2001, pp: 139-154.

[15] D.V. Gokhale and S. Kullback, Information in contingency Tables, New York, Marcel Dekker, 1978.

[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, On a new exponential information divergence measure, Communicated (2015).

[18] K.C. Jain and R.N. Saraswat, Some new information inequalities and its applications in information theory, International Journal of Mathematics Research, vol. 4, no.3 (2012), pp: 295- 307.

[19] 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.

[20] 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.

[21] 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.

[22] T. Kailath, The divergence and Bhattacharyya distance measures in signal selection, IEEE Trans. Comm. Technology, vol. COM-15, 1967, pp: 52- 60.

[23] 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.

[24] A.N. Kolmogorov, On the approximation of distributions of sums of independent summands by infinitely divisible distributions, Sankhya, 25, pp: 159-174.

[25] S. Kullback and R.A. Leibler, On information and sufficiency, Ann. Math. Statist., 22 (1951), pp: 79-86.

[26] P.W. Lamberti, A.P. Majtey, A. Borras, M. Casas, A. Plastino , Metric character of the quantum Jensen-Shannon divergence, Physical Review A, 77:052311, 2008.

[27] F. Nielsen and S. Boltz, The Burbea-Rao and Bhattacharyya centroids, Apr. 2010, Arxiv.

[28] M.A. Nielsen, I.L. Chuang, Quantum computation and information, Cambridge University Press, Cambridge, UK, 3 (8):9, 2000.

[29] F. Osterreicher, Csiszar's f- divergence basic properties, Homepage: http//www.sbg.ac.at/mat/home.html, November 22, 2002.

[30] E.C. Pielou, Ecological diversity, New York, Wiley, 1975.

[31] H.V. Poor, Robust decision design using a distance criterion, IEEE Trans. Inf. Th., vol.- IT 26, no. 5, pp: 575- 587, 1980.

[32] A. Renyi, On measures of entropy and information, Proc. 4th Berkeley Symposium on Math. Statist. and Prob., 1 (1961), pp: 547-561.

[33] M. Salicru, Measures of information associated with Csiszar's divergences, Kybernetika, vol. 30 (1994), no. 5, pp: 563- 573.

[34] 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.

[35] C.E. Shannon, A mathematical theory of communication, Bull. Sept. Tech. J., 27 (1948), 370-423 and 623-656.

[36] H.C. Taneja and R.K. Tuteja, Characterization of a quantitative- qualitative measure of inaccuracy, Kybernetika, vol. 22 (1986), no. 5, pp: 393- 402.

[37] 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.

[38] H. Theil, Statistical decomposition analysis, Amsterdam, North-Holland, 1972.

[39] H. Theil, Economics and information theory, Amsterdam, North-Holland, 1967.

[40] K. Tumer, J. Ghosh, Estimating the Bayes error rate through classifier combining, Proceedings of 13th International Conference on Pattern Recognition, pp: 695- 699, 1996.

[41] 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

K.C. Jain and Praphull Chhabra, Department of Mathematics, Malaviya National Institute of Technology, Jaipur- 302017 (Rajasthan), India.
E-mail: jainkc_2003@yahoo.com, prfl24@gmail.com

Received: May 28, 2015.

Accepted: September 26, 2015.