Available online at http://www.journalcra.com

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

International Journal of Current Research
Vol. 7, Issue, 03, pp.13836-13853, March, 2015

ISSN: 0975-833X

RESEARCH ARTICLE

NEW INFORMATION INEQUALITIES ON NEW F- DIVERGENCE BY USING OSTROWSKI'S INEQUALITIES AND ITS APPLICATION

***Jain, K. C. and Praphull Chhabra**

Department of Mathematics, Malaviya National Institute of Technology, Jaipur- 302017 (Rajasthan), India

ARTICLE INFO

Article History:
Received 24th December, 2014
Received in revised form
06th January, 2015
Accepted 28th February, 2015
Published online 31st March, 2015

Key words:
New information inequalities;
Logarithmic power mean;
Identric mean;
Binomial and Poisson distributions;
Mutual information;
Numerical approximation.

ABSTRACT

Many relations have been obtained among several divergences by using several information inequalities. In this work, we also relate the Relative Jensen- Shannon divergence, Triangular discrimination, Relative Arithmetic- Geometric divergence, Relative J- divergence, and Hellinger discrimination to the Chi- square divergence and Varitional distance independently in a specific interval by using two different new information inequalities on new generalized f- divergence, together with numerical verification by taking two discrete probability distributions: Binomial and Poisson. These new information inequalities are derived by using Ostrowski's inequalities. Application to the Mutual information and numerical approximation are done as well.

INTRODUCTION

Divergence measures are basically measures of distance between two probability distributions or compare two probability distributions, i.e., divergence measures are directly propositional to the distance between two probability distributions. It means that any divergence measure must take its minimum value zero when probability distributions are equal and maximum when probability distributions are perpendicular to each other. So, any divergence measure must increase as probability distributions move apart.

Divergence measures have been demonstrated very useful in a variety of disciplines such as economics and political science (Theil, 1972 and Theil, 1967), biology (Pielou, 1975), analysis of contingency tables (Gokhale and Kullback, 1978), approximation of probability distributions (Chow and Lin, 1968 and Kazakos and Cotsidas, 1980), signal processing (Kadota and Shepp, 1967 and Kailath, 1967), pattern recognition (Bassat, 1978; Chen, 1973 and Jones and Byrne, 1990), color image segmentation (Nielsen and Boltz, 2010), 3D image segmentation and word alignment (Taskar *et al*., 2006), cost- sensitive classification for medical diagnosis (Santos-Rodriguez *et al*., 2009), magnetic resonance image analysis (Vemuri *et al*., 2010) etc.

Also we can use divergences in fuzzy mathematics as fuzzy directed divergences and fuzzy entropies (Bajaj and Hooda, 2010; Hooda, 2004 and Jha and Mishra, 2012), 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.

Without essential loss of insight, we have restricted ourselves to discrete probability distributions, so let $\Gamma_n = \left\{ P = \left(p_1, p_2, p_3 ..., p_n\right) : p_i > 0, \sum_{i=1}^{n} p_i = 1 \right\}, n \geq 2$ be the set of all complete finite discrete probability distributions.

***Corresponding author: Jain, K. C.**
Department of Mathematics, Malaviya National Institute of Technology, Jaipur- 302017 (Rajasthan), India.

The restriction here to discrete distributions is only for convenience, similar results hold for continuous distributions. If we take $p_i \geq 0$ for some $i = 1, 2, 3, ..., n$, then we have to suppose that $0 f(0) = 0 f\left(\frac{0}{0}\right) = 0$.

Some generalized functional information divergence measures had been introduced, characterized and applied in variety of fields, such as: Csiszar's f - divergence (Csiszar, 1974 and Csiszar, 1967), Bregman's f - divergence (Bregman, 1967), Burbea- Rao's f - divergence (Burbea and Rao, 1982), Renyi's like f - divergence (Renyi, 1961) etc. Similarly, Jain and Saraswat (Jain and Saraswat, 2012) defined new generalized f - divergence measure, which is given by

$$S_f(P,Q) = \sum_{i=1}^{n} q_i f\left(\frac{p_i + q_i}{2q_i}\right), \tag{1.1}$$

where $f : (0, \infty) \to R$ (set of real no.) is real, continuous, and convex function and $P = (p_1, p_2, p_3..., p_n), Q = (q_1, q_2, q_3..., q_n) \in \Gamma_n$, where p_i and q_i are probability mass functions. The advantage of these generalized divergences is that many divergence measures can be obtained from these generalized f - divergences by suitably defining the function f. Some resultant divergences by $S_f(P,Q)$, are as follows (Properties of $S_f(P,Q)$ can be seen in literature (Jain and Saraswat, 2012).

a. If we take $f(t) = -\log t$ in (1.1), we obtain

$$S_f(P,Q) = \sum_{i=1}^{n} q_i \log\left(\frac{2q_i}{p_i + q_i}\right) = F(Q,P), \tag{1.2}$$

where $F(Q,P)$ is called adjoint of the **Relative JS divergence** $F(P,Q)$ (Sibson, 1969**).**

b. If we take $f(t) = \frac{(t-1)^2}{t}$ in (1.1), we obtain

$$S_f(P,Q) = \frac{1}{2}\sum_{i=1}^{n} \frac{(p_i - q_i)^2}{p_i + q_i} = \frac{1}{2}\Delta(P,Q), \tag{1.3}$$

where $\Delta(P,Q)$ is called the **Triangular discrimination** (Dacunha- Castelle, 1978).

c. If we take $f(t) = t \log t$ in (1.1), we obtain

$$S_f(P,Q) = \sum_{i=1}^{n} \left(\frac{p_i + q_i}{2}\right) \log\left(\frac{p_i + q_i}{2q_i}\right) = G(Q,P), \tag{1.4}$$

where $G(Q,P)$ is called adjoint of the **Relative AG divergence** $G(P,Q)$ (Taneja, 1995**).**

d. If we take $f(t) = (t-1)\log t$ in (1.1), we obtain

$$S_f(P,Q) = \frac{1}{2}\sum_{i=1}^{n} (p_i - q_i) \log\left(\frac{p_i + q_i}{2q_i}\right) = \frac{1}{2} J_R(P,Q), \tag{1.5}$$

where $J_R(P,Q)$ is called the **Relative J- divergence (**Dragomir *et al*., 2001**)**.

e. If we take $f(t)=1-\sqrt{t}$ in (1.1), we obtain

$$S_f(P,Q)=\frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{\frac{p_i+q_i}{2}}-\sqrt{q_i}\right)^2=1-G^*\left(\frac{P+Q}{2},Q\right)=h\left(\frac{P+Q}{2},Q\right), \tag{1.6}$$

where $G^*(P,Q)=\sum_{i=1}^{n}\sqrt{p_iq_i}$ and $h(P,Q)=\sum_{i=1}^{n}\frac{\left(\sqrt{p_i}-\sqrt{q_i}\right)^2}{2}$ are called the Geometric divergence and Hellinger discrimination (Kolmogorov's divergence) (Hellinger, 1909), respectively.

f. If we take $f(t)=(t-1)^2$ in (1.1), we obtain

$$S_f(P,Q)=\frac{1}{4}\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{q_i}=\frac{1}{4}\chi^2(P,Q), \tag{1.7}$$

where $\chi^2(P,Q)$ is called the **Chi- Square divergence (Pearson div. measure) (**Pearson, 1900**).**

g. If we take $f(t)=|t-1|$ in (1.1), we obtain

$$S_f(P,Q)=\frac{1}{2}\sum_{i=1}^{n}|p_i-q_i|=\frac{1}{2}V(P,Q), \tag{1.8}$$

where $V(P,Q)$ is called the **Variational distance (l_1 distance) (Kolmogorov, 1963).**

h. Particularly, by taking $f(t)=(2t-1)\log(2t-1),\ t\in\left(\frac{1}{2},\infty\right)$ in (1.1), we obtain

$$S_f(P,Q)=\sum_{i=1}^{n}p_i\log\left(\frac{p_i}{q_i}\right)=K(P,Q)\ , \tag{1.9}$$

where $K(P,Q)$ is called the **Relative entropy (Kullback- Leibler distance) (**Kullback and Leibler, 1951**).**

Similarly, we can obtain many divergences by using linear convex functions. Since these divergences are not worthful in practice, therefore we can skip them. We can see that, divergence (1.3) and (1.8) are symmetric while (1.2), (1.4), (1.5) to (1.7), and (1.9) are non- symmetric with respect to probability distributions. Now, there are two generalized means which are being used in this paper for calculations only. These are as follows.

$$L_p(a,b)=\begin{cases}\frac{b^{p+1}-a^{p+1}}{(p+1)(b-a)}, & p\neq -1,0\\ \frac{\log b-\log a}{b-a}, & p=-1\\ 1, & p=0\end{cases}\quad a,b>0,\ a\neq b. \tag{1.10}$$

$$I(a,b)=\begin{cases}\frac{1}{e}\left(\frac{b^b}{a^a}\right)^{\frac{1}{b-a}}, & a\neq b \\ b, & a=b\end{cases} \quad a,b>0. \tag{1.11}$$

Means (1.10) and (1.11) are called p - Logarithmic power mean and Identric mean respectively.

New information inequalities

In this section, we introduce two new information inequalities (Theorems 2.1 and 2.2) on $S_f(P,Q)$ in terms of $\chi^2(P,Q)$ and $V(P,Q)$. Such inequalities are for instance needed in order to calculate the relative efficiency of two divergences.

Firstly, following lemmas 2.1 (Dragomir *et al*., 2001) and 2.2 (Dragomir, 1999) are very important to introduce new information inequalities. These are as follows.

Lemma 2.1 Let $f:(a,b)\subseteq R\to R$ be an absolutely continuous function in (a,b) with $a<b$ and $f':(a,b)\to R$ is essentially bounded or $f'\in L_\infty(a,b)$, i.e.,

$\|f'\|_\infty = ess \sup_{t\in(a,b)}|f'(t)|<\infty$, then we have

$$\left|f(x)-\frac{1}{(b-a)}\int_a^b f(t)\,dt\right|\le\left[\frac{1}{4}+\left(\frac{x-\frac{(a+b)}{2}}{(b-a)}\right)^2\right](b-a)\|f'\|_\infty, \tag{2.1}$$

for all $x,t\in(a,b)$.

Lemma 2.2 Let $f:(a,b)\subseteq R\to R$ be a differentiable function and is of bounded variation on (a,b), i.e.,

$\bigvee_a^b(f)=\int_a^b|f'(t)|\,dt<\infty$, then we have

$$\left|\int_a^b f(t)\,dt-f(x)(b-a)\right|\le\left[\frac{b-a}{2}+\left|x-\frac{a+b}{2}\right|\right]\bigvee_a^b(f), \tag{2.2}$$

for all $x,t\in(a,b)$.

Now, we will obtain two new information inequalities by using above two Ostrowski's inequalities (2.1) and (2.2).

Theorem 2.1 Let $f:(\alpha,\beta)\subseteq R_+\to R$ be an absolutely continuous function in (α,β) with $0<\alpha\le 1\le\beta<\infty,\ \alpha\neq\beta$ and $f':(\alpha,\beta)\to R$ is essentially bounded or $f'\in L_\infty(\alpha,\beta)$, i.e.,

$$\|f'\|_\infty = ess \sup_{t\in(\alpha,\beta)}|f'(t)|<\infty, \tag{2.3}$$

for all $t\in(\alpha,\beta)$.

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

$$\left|S_f\left(P,Q\right)-\frac{1}{\left(\beta-\alpha\right)}\int_{\alpha}^{\beta}f\left(t\right)dt\right|\le\frac{\left(\beta-\alpha\right)\left\|f'\right\|_{\infty}}{4}\left[2+\frac{1}{\left(\beta-\alpha\right)^2}\chi^2\left(P,Q\right)\right], \quad (2.4)$$

where $S_f\left(P,Q\right)$ and $\chi^2\left(P,Q\right)$ are given by (1.1) and (1.7) respectively.

Proof: Put $a=\alpha, b=\beta$, and $x=\frac{p_i+q_i}{2q_i}$ in inequality (2.1), multiply by q_i and then sum over all $i=1,2,...n$, we get

$$\left|\sum_{i=1}^{n}q_i f\left(\frac{p_i+q_i}{2q_i}\right)-\frac{1}{\left(\beta-\alpha\right)}\int_{\alpha}^{\beta}f\left(t\right)dt\sum_{i=1}^{n}q_i\right|\le\left[\frac{1}{4}\sum_{i=1}^{n}q_i+\frac{1}{\left(\beta-\alpha\right)^2}\sum_{i=1}^{n}q_i\left(\frac{p_i+q_i}{2q_i}-\frac{\left(\alpha+\beta\right)}{2}\right)^2\right]\left(\beta-\alpha\right)\left\|f'\right\|_{\infty}, \text{ i.e.,}$$

$$\left|S_f\left(P,Q\right)-\frac{1}{\left(\beta-\alpha\right)}\int_{\alpha}^{\beta}f\left(t\right)dt\right|\le\left[\frac{1}{4}+\frac{1}{4}\frac{1}{\left(\beta-\alpha\right)^2}\sum_{i=1}^{n}q_i\left(\frac{p_i+q_i}{q_i}-\left(\alpha+\beta\right)\right)^2\right]\left(\beta-\alpha\right)\left\|f'\right\|_{\infty}, \text{ i.e.,}$$

$$\le\frac{\left(\beta-\alpha\right)\left\|f'\right\|_{\infty}}{4}\left[1+\frac{1}{\left(\beta-\alpha\right)^2}\left\{\sum_{i=1}^{n}\frac{\left(p_i+q_i\right)^2}{q_i}+\left(\alpha+\beta\right)^2\sum_{i=1}^{n}q_i-2\left(\alpha+\beta\right)\sum_{i=1}^{n}\left(p_i+q_i\right)\right\}\right], \text{ i.e.,}$$

$$\le\frac{\left(\beta-\alpha\right)\left\|f'\right\|_{\infty}}{4}\left[1+\frac{1}{\left(\beta-\alpha\right)^2}\left\{\left(\alpha+\beta\right)^2-4\left(\alpha+\beta\right)+4+\sum_{i=1}^{n}\frac{\left(p_i-q_i\right)^2}{q_i}\right\}\right], \text{ i.e.,}$$

$$\le\frac{\left(\beta-\alpha\right)\left\|f'\right\|_{\infty}}{4}\left[1+\frac{1}{\left(\beta-\alpha\right)^2}\left\{\left(\alpha+\beta-2\right)^2+\chi^2\left(P,Q\right)\right\}\right], \text{ i.e.,}$$

$$\le\frac{\left(\beta-\alpha\right)\left\|f'\right\|_{\infty}}{4}\left[2+\frac{1}{\left(\beta-\alpha\right)^2}\chi^2\left(P,Q\right)\right], \left[\because\left(\alpha+\beta-2\right)^2\le\left(\beta-\alpha\right)^2\right].$$

Hence prove the inequality (2.4).

Theorem 2.2 Let $f:\left(\alpha,\beta\right)\subseteq R_+\to R$ be a differentiable function and is of bounded variation on $\left(\alpha,\beta\right)$ with $0<\alpha\le1\le\beta<\infty, \alpha\ne\beta$, i.e.,

$$\underset{\alpha}{\overset{\beta}{A}}\left(f\right)=\int_{\alpha}^{\beta}\left|f'\left(t\right)\right|dt<\infty, \quad (2.5)$$

for all $t\in\left(\alpha,\beta\right)$.

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

$$\left|S_f\left(P,Q\right)-\frac{1}{\left(\beta-\alpha\right)}\int_{\alpha}^{\beta}f\left(t\right)dt\right|\le\frac{\underset{\alpha}{\overset{\beta}{A}}\left(f\right)}{2}\left[2+\frac{1}{\left(\beta-\alpha\right)}V\left(P,Q\right)\right], \quad (2.6)$$

where $S_f\left(P,Q\right)$ and $V\left(P,Q\right)$ are given by (1.1) and (1.8) respectively.

Proof: Put $a=\alpha, b=\beta$, and $x=\dfrac{p_i+q_i}{2q_i}$ in inequality (2.2), multiply by q_i and then sum over all $i=1,2,\ldots n$, we get

$$\left|\sum_{i=1}^{n} q_i f\left(\frac{p_i+q_i}{2q_i}\right)-\frac{1}{(\beta-\alpha)}\int_{\alpha}^{\beta} f(t)\,dt\sum_{i=1}^{n} q_i\right| \le \left[\frac{1}{2}+\frac{1}{(\beta-\alpha)}\sum_{i=1}^{n} q_i\left|\frac{p_i+q_i}{2q_i}-\frac{\alpha+\beta}{2}\right|\right]\overset{\beta}{\underset{\alpha}{A}}(f), \text{ i.e.,}$$

$$\left|S_f(P,Q)-\frac{1}{(\beta-\alpha)}\int_{\alpha}^{\beta} f(t)\,dt\right| \le \left[\frac{1}{2}+\frac{1}{(\beta-\alpha)}\sum_{i=1}^{n} q_i\left|\frac{p_i+q_i}{2q_i}-1-\left(\frac{\alpha+\beta}{2}-1\right)\right|\right]\overset{\beta}{\underset{\alpha}{A}}(f), \text{ i.e.,}$$

$$\le \left[\frac{1}{2}+\frac{1}{(\beta-\alpha)}\left\{\sum_{i=1}^{n} q_i\left|\frac{p_i+q_i}{2q_i}-1\right|+\sum_{i=1}^{n} q_i\left|\frac{\alpha+\beta}{2}-1\right|\right\}\right]\overset{\beta}{\underset{\alpha}{A}}(f), \text{ i.e.,}$$

$$\le \left[\frac{1}{2}+\frac{1}{(\beta-\alpha)}\left\{\frac{1}{2}V(P,Q)+\left|\frac{\alpha+\beta}{2}-1\right|\right\}\right]\overset{\beta}{\underset{\alpha}{A}}(f), \text{ i.e.,}$$

$$\le \frac{1}{2}\left[2+\frac{1}{(\beta-\alpha)}V(P,Q)\right]\overset{\beta}{\underset{\alpha}{A}}(f), \left[\because \left|\frac{\alpha+\beta}{2}-1\right| \le \frac{\beta-\alpha}{2}\right].$$

Hence prove the inequality (2.6).

Application of obtained new information inequalities

In this section, we obtain relations of different divergences in terms of the Chi- square divergence and Variational distance independently by using new inequalities (2.4) and (2.6). We are considering only convex functions, the inequalities hold good for concave functions as well. Means (1.10) and (1.11) are taking for simplification and calculations.

Proposition 3.1 Let $F(P,Q)$, $\chi^2(P,Q)$, and $V(P,Q)$ be defined as in (1.2), (1.7), and (1.8) respectively. For $P,Q \in \Gamma_n$, we have

$$\left|F(Q,P)+\log I(\alpha,\beta)-2\right| \le \frac{(\beta-\alpha)}{4\alpha}\left[2+\frac{1}{(\beta-\alpha)^2}\chi^2(P,Q)\right]. \tag{3.1}$$

$$\left|F(P,Q)+\log I(\alpha,\beta)-2\right| \le \left[\beta-\alpha+V(P,Q)\right]L_{-1}(\alpha,\beta). \tag{3.2}$$

Proof: Let us consider

$$f(t)=-\log t, t\in R_+, f(1)=0, f'(t)=-\frac{1}{t} \text{ and } f''(t)=\frac{1}{t^2}.$$

Since $f''(t)>0\,\forall t>0$ and $f(1)=0$, so $f(t)$ is convex and normalized function respectively.

Now put $f(t)$ in (1.1) and put $f'(t)$ in (2.3) and (2.5), we get the followings

$$S_f(P,Q)=\sum_{i=1}^{n} q_i \log\left(\frac{2q_i}{p_i+q_i}\right)=F(Q,P). \tag{3.3}$$

$$\overset{\beta}{\underset{\alpha}{A}}(f)=\int_{\alpha}^{\beta}\left|f'(t)\right|dt=\int_{\alpha}^{\beta}\left|-\frac{1}{t}\right|dt=\int_{\alpha}^{\beta}\frac{1}{t}\,dt=\log\beta-\log\alpha. \tag{3.4}$$

Now, let $g(t)=\left|f'(t)\right|=\left|-\frac{1}{t}\right|=\frac{1}{t}$, and $g'(t)=-\frac{1}{t^2}<0$.

It is clear that $g(t)$ is always decreasing in (0, ∞), so

$$\|f'\|_\infty=\sup_{t\in(\alpha,\beta)}\left|f'(t)\right|=\sup_{t\in(\alpha,\beta)}g(t)=g(\alpha)=\frac{1}{\alpha}. \tag{3.5}$$

The results (3.1) and (3.2) are obtained by using (3.3), (3.4), and (3.5) in (2.4) and (2.6) respectively.

Proposition 3.2 Let $\Delta(P,Q)$, $\chi^2(P,Q)$, and $V(P,Q)$ be defined as in (1.3), (1.7), and (1.8) respectively. For $P,Q\in\Gamma_n$, we have

a. If $0<\alpha<1$, then

$$\left|\Delta(P,Q)-2\left\{L_1(\alpha,\beta)+L_{-1}(\alpha,\beta)+2\right\}\right|$$
$$\leq\frac{(\beta-\alpha)}{4}\left[2+\frac{1}{(\beta-\alpha)^2}\chi^2(P,Q)\right]\left[\frac{\beta^2-\alpha^2}{\alpha^2\beta^2}+\left|\frac{\beta^2+\alpha^2}{\alpha^2\beta^2}-2\right|\right], \tag{3.6}$$

b. If $\alpha=1$, then

$$\left|\Delta(P,Q)-2\left\{L_1(\alpha,\beta)+L_{-1}(\alpha,\beta)+2\right\}\right|$$

$$\leq\left[2+\frac{1}{(\beta-\alpha)^2}\chi^2(P,Q)\right]\frac{(\beta-\alpha)(\beta+1)(\beta-1)}{2\beta^2}, \tag{3.7}$$

and

$$\left|\Delta(P,Q)-2\left\{L_1(\alpha,\beta)+L_{-1}(\alpha,\beta)+2\right\}\right|$$

$$\leq\left[2+\frac{1}{(\beta-\alpha)}V(P,Q)\right]\left(\frac{\alpha+\beta}{\alpha\beta}+2L_1(\alpha,\beta)-4\right). \tag{3.8}$$

Proof: Let us consider

$$f(t)=\frac{(t-1)^2}{t},t\in R_+,f(1)=0,f'(t)=\frac{t^2-1}{t^2}\text{ and }f''(t)=\frac{2}{t^3}.$$

Since $f''(t)>0\,\forall t>0$ and $f(1)=0$, so $f(t)$ is convex and normalized function respectively.
Now put $f(t)$ in (1.1) and put $f'(t)$ in (2.3) and (2.5), we get the followings

$$S_f(P,Q)=\frac{1}{2}\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{p_i+q_i}=\frac{1}{2}\Delta(P,Q). \tag{3.9}$$

$$\overset{\beta}{\underset{\alpha}{A}}(f)=\int_{\alpha}^{\beta}|f'(t)|dt=\int_{\alpha}^{1}\frac{1-t^2}{t^2}+\int_{1}^{\beta}\frac{t^2-1}{t^2}=\frac{\alpha+\beta}{\alpha\beta}+\alpha+\beta-4. \qquad (3.10)$$

Now, let $g(t)=|f'(t)|=\left|\frac{(t+1)(t-1)}{t^2}\right|=\begin{cases}\frac{(t+1)(1-t)}{t^2}, & 0<t<1\\ \frac{(t+1)(t-1)}{t^2}, & 1\le t<\infty\end{cases}$, and

$$g'(t)=\begin{cases}-\frac{2}{t^3}<0, & 0<t<1\\ \frac{2}{t^3}>0, & 1\le t<\infty\end{cases}.$$

It is clear that $g'(t)<0$ in $(0,1)$ and >0 in $(1,\infty)$, i.e., $g(t)$ is decreasing in $(0,1)$ and increasing in $(1,\infty)$, so

$$\|f'\|_{\infty}=\sup_{t\in(\alpha,\beta)}|f'(t)|=\sup_{t\in(\alpha,\beta)}g(t)$$

$$=\begin{cases}\max\left[|f'(\alpha)|,|f'(\beta)|\right]=\frac{|f'(\alpha)|+|f'(\beta)|+\left||f'(\alpha)|-|f'(\beta)|\right|}{2}, & 0<\alpha<1\\ |f'(\beta)|, & \alpha=1\end{cases}, \text{ i.e.,}$$

$$=\begin{cases}\frac{1}{2}\left[\frac{\beta^2-\alpha^2}{\alpha^2\beta^2}+\left|\frac{\beta^2+\alpha^2}{\alpha^2\beta^2}-2\right|\right], & 0<\alpha<1\\ \frac{(\beta+1)(\beta-1)}{\beta^2}, & \alpha=1\end{cases}. \qquad (3.11)$$

The results (3.6), (3.7), and (3.8) are obtained by using (3.9), (3.10), and (3.11) in (2.4) and (2.6) respectively.

Proposition 3.3 Let $G(P,Q)$, $\chi^2(P,Q)$, and $V(P,Q)$ be defined as in (1.4), (1.7), and (1.8) respectively. For $P,Q\in\Gamma_n$, we have

a. If $0<\alpha\le\frac{1}{e}$, then

$$\left|2G(Q,P)+L_1(\alpha,\beta)-\frac{(\beta^2\log\beta-\alpha^2\log\alpha)}{(\beta-\alpha)}\right|$$

$$\le\frac{(\beta-\alpha)}{2}\left[2+\frac{1}{(\beta-\alpha)^2}\chi^2(P,Q)\right]\left[\log\sqrt{\frac{\beta}{\alpha}}+\left|1+\log\sqrt{\alpha\beta}\right|\right], \qquad (3.12)$$

b. If $\frac{1}{e}<\alpha\le1$, then

$$\left|2G(Q,P)+L_1(\alpha,\beta)-\frac{(\beta^2\log\beta-\alpha^2\log\alpha)}{(\beta-\alpha)}\right|$$

$$\leq \frac{(\beta-\alpha)\log(e\alpha)}{2}\left[2+\frac{1}{(\beta-\alpha)^2}\chi^2(P,Q)\right], \tag{3.13}$$

and

$$\left|2G(P,Q)+L_1(\alpha,\beta)-\frac{\left(\beta^2\log\beta-\alpha^2\log\alpha\right)}{(\beta-\alpha)}\right|$$
$$\leq\left[2+\frac{1}{(\beta-\alpha)}V(P,Q)\right]\left(\alpha\log\alpha+\beta\log\beta+\frac{2}{e}\right). \tag{3.14}$$

Proof: Let us consider

$f(t)=t\log t, t\in R_+, f(1)=0, f'(t)=1+\log t$ and $f''(t)=\frac{1}{t}$.

Since $f''(t)>0\,\forall t>0$ and $f(1)=0$, so $f(t)$ is convex and normalized function respectively.

Now put $f(t)$ in (1.1) and put $f'(t)$ in (2.3) and (2.5), we get the followings

$$S_f(P,Q)=\sum_{i=1}^{n}\left(\frac{p_i+q_i}{2}\right)\log\left(\frac{p_i+q_i}{2q_i}\right)=G(Q,P)\ . \tag{3.15}$$

$$\overset{\beta}{\underset{\alpha}{A}}(f)=\int_{\alpha}^{\beta}\left|f'(t)\right|dt=\int_{\alpha}^{1/e}(-1-\log t)\,dt+\int_{1/e}^{\beta}(1+\log t)\,dt=\alpha\log\alpha+\beta\log\beta+\frac{2}{e}. \tag{316}$$

Now, let $g(t)=\left|f'(t)\right|=\left|1+\log t\right|=\begin{cases}-1-\log t, & 0<t\leq\frac{1}{e}\\ 1+\log t, & \frac{1}{e}<t<\infty\end{cases}$,

and $$g'(t)=\begin{cases}-\frac{1}{t}<0, & 0<t\leq\frac{1}{e}\\ \frac{1}{t}>0, & \frac{1}{e}<t<\infty\end{cases}.$$

It is clear that $g'(t)<0$ in $\left(0,\frac{1}{e}\right)$ and >0 in $\left(\frac{1}{e},\infty\right)$, i.e., $g(t)$ is decreasing in $\left(0,\frac{1}{e}\right)$ and increasing in $\left(\frac{1}{e},\infty\right)$, so

$$\left\|f'\right\|_\infty=\sup_{t\in(\alpha,\beta)}\left|f'(t)\right|=\sup_{t\in(\alpha,\beta)}g(t)$$
$$=\begin{cases}\max\left[\left|f'(\alpha)\right|,\left|f'(\beta)\right|\right]=\left[\log\sqrt{\frac{\beta}{\alpha}}+\left|1+\log\sqrt{\alpha\beta}\right|\right], & 0<\alpha\leq\frac{1}{e}\\ \left|f'(\beta)\right|=1+\log\alpha, & \frac{1}{e}<\alpha\leq 1\end{cases}. \tag{3.17}$$

The results (3.12), (3.13), and (3.14) are obtained by using (3.15), (3.16), and (3.17) in (2.4) and (2.6) respectively.

Proposition 3.4 Let $J_R(P,Q)$, $\chi^2(P,Q)$, and $V(P,Q)$ be defined as in (1.5), (1.7), and (1.8) respectively. For $P,Q \in \Gamma_n$, we have

a. If $0<\alpha<1$, then

$$\left| J_R(P,Q)+2\log I(\alpha,\beta)+L_1(\alpha,\beta)-\frac{\left(\beta^2\log\beta-\alpha^2\log\alpha\right)}{(\beta-\alpha)}-4 \right|$$
$$\le \frac{(\beta-\alpha)}{2}\left[2+\frac{1}{(\beta-\alpha)^2}\chi^2(P,Q)\right]\left[\log\sqrt{\frac{\beta}{\alpha}}+\frac{\beta-\alpha}{2\alpha\beta}+\left|\frac{\beta+\alpha}{2\alpha\beta}-\log e\sqrt{\alpha\beta}\right|\right], \quad (3.18)$$

b. If $\alpha=1$, then

$$\left| J_R(P,Q)+2\log I(\alpha,\beta)+L_1(\alpha,\beta)-\frac{\left(\beta^2\log\beta-\alpha^2\log\alpha\right)}{(\beta-\alpha)}-4 \right|$$
$$\le \frac{(\beta-\alpha)}{2}\left[2+\frac{1}{(\beta-\alpha)^2}\chi^2(P,Q)\right]\left(\log e\beta-\frac{1}{\beta}\right), \quad (3.19)$$

and

$$\left| J_R(P,Q)+2\log I(\alpha,\beta)+L_1(\alpha,\beta)-\frac{\left(\beta^2\log\beta-\alpha^2\log\alpha\right)}{(\beta-\alpha)}-4 \right|$$
$$\le \left[2+\frac{1}{(\beta-\alpha)}V(P,Q)\right]\left(\alpha\log\alpha+\beta\log\beta-\log\alpha-\log\beta\right). \quad (3.20)$$

Proof: Let us consider

$f(t)=(t-1)\log t, t\in R_+, f(1)=0, f'(t)=\frac{t-1}{t}+\log t$ and $f''(t)=\frac{t+1}{t^2}$.

Since $f''(t)>0\,\forall t>0$ and $f(1)=0$, so $f(t)$ is convex and normalized function respectively.

Now put $f(t)$ in (1.1) and put $f'(t)$ in (2.3) and (2.5), we get the followings

$$S_f(P,Q)=\frac{1}{2}\sum_{i=1}^{n}(p_i-q_i)\log\left(\frac{p_i+q_i}{2q_i}\right)=\frac{1}{2}J_R(P,Q). \quad (3.21)$$

$$\underset{\alpha}{\overset{\beta}{A}}(f)=\int_\alpha^\beta |f'(t)|dt=\int_\alpha^1\left(-1-\log t+\frac{1}{t}\right)dt+\int_1^\beta\left(1+\log t-\frac{1}{t}\right)dt$$
$$=\alpha\log\alpha+\beta\log\beta-\log\alpha-\log\beta. \quad (3.22)$$

Now, let $g(t)=|f'(t)|=\left|\frac{t-1}{t}+\log t\right|=\begin{cases}-1-\log t+\frac{1}{t}, & 0<t<1\\ 1+\log t-\frac{1}{t}, & 1\le t<\infty\end{cases}$, and $g'(t)=\begin{cases} -\left(\dfrac{t+1}{t^2}\right)<0, & 0<t<1 \\ \left(\dfrac{t+1}{t^2}\right)>0, & 1\le t<\infty \end{cases}$.

It is clear that $g'(t)<0$ in $(0,1)$ and >0 in $(1,\infty)$, i.e., $g(t)$ is decreasing in $(0,1)$ and increasing in $(1,\infty)$, so

$$\|f'\|_\infty = \sup_{t\in(\alpha,\beta)}|f'(t)| = \sup_{t\in(\alpha,\beta)} g(t)$$
$$=\begin{cases} \max\left[|f'(\alpha)|,|f'(\beta)|\right]=\left[\log\sqrt{\dfrac{\beta}{\alpha}}+\dfrac{\beta-\alpha}{2\alpha\beta}+\left|\dfrac{\beta+\alpha}{2\alpha\beta}-\log e\sqrt{\alpha\beta}\right|\right], & 0<\alpha<1 \\ |f'(\beta)|=\left(\log e\beta-\dfrac{1}{\beta}\right), & \alpha=1 \end{cases}. \quad (3.23)$$

The results (3.18), (3.19), and (3.20) are obtained by using (3.21), (3.22), and (3.23) in (2.4) and (2.6) respectively.

Proposition 3.5 Let $h(P,Q)$, $\chi^2(P,Q)$, and $V(P,Q)$ be defined as in (1.6), (1.7), and (1.8) respectively. For $P,Q\in\Gamma_n$, we have

$$\left|h\left(\frac{P+Q}{2},Q\right)+L_{1/2}(\alpha,\beta)-1\right|\le\frac{(\beta-\alpha)}{8\sqrt{\alpha}}\left[2+\frac{1}{(\beta-\alpha)^2}\chi^2(P,Q)\right]. \quad (3.24)$$

$$\left|h\left(\frac{P+Q}{2},Q\right)+L_{1/2}(\alpha,\beta)-1\right|\le\frac{1}{2}\left[2\left(\sqrt{\beta}-\sqrt{\alpha}\right)+\frac{1}{2}L_{-1/2}(\alpha,\beta)V(P,Q)\right]. \quad (3.25)$$

Proof: Let us consider

$$f(t)=1-\sqrt{t},\ t\in R_+,\ f(1)=0,\ f'(t)=-\frac{1}{2\sqrt{t}}\ \text{and}\ f''(t)=\frac{1}{4t^{3/2}}.$$

Since $f''(t)>0\,\forall t>0$ and $f(1)=0$, so $f(t)$ is convex and normalized function respectively.
Now put $f(t)$ in (1.1) and put $f'(t)$ in (2.3) and (2.5), we get the followings

$$S_f(P,Q)=\sum_{i=1}^n q_i\left(1-\sqrt{\frac{p_i+q_i}{2q_i}}\right)=1-\sum_{i=1}^n\sqrt{\frac{q_i(p_i+q_i)}{2}}=\frac{1}{2}\left[2-2\sum_{i=1}^n\sqrt{\frac{q_i(p_i+q_i)}{2}}\right]$$
$$=\frac{1}{2}\left[\sum_{i=1}^n\frac{p_i+q_i}{2}+\sum_{i=1}^n q_i-\sqrt{2q_i(p_i+q_i)}\right]=\frac{1}{2}\sum_{i=1}^n\left(\sqrt{\frac{p_i+q_i}{2}}-\sqrt{q_i}\right)^2=h\left(\frac{P+Q}{2},Q\right). \quad (3.26)$$

$$\overset{\beta}{\underset{\alpha}{A}}(f)=\int_\alpha^\beta|f'(t)|\,dt=\frac{1}{2}\int_\alpha^\beta\left|-\frac{1}{\sqrt{t}}\right|dt=\frac{1}{2}\int_\alpha^\beta\frac{1}{\sqrt{t}}\,dt=\sqrt{\beta}-\sqrt{\alpha}. \quad (3.27)$$

Now, let $g(t)=|f'(t)|=\dfrac{1}{2}\left|-\dfrac{1}{\sqrt{t}}\right|=\dfrac{1}{2\sqrt{t}}$, and $g'(t)=-\dfrac{1}{4t^{3/2}}<0$.

It is clear that $g(t)$ is always decreasing in (0, ∞), so

$$\|f'\|_\infty = \sup_{t\in(\alpha,\beta)} |f'(t)| = \sup_{t\in(\alpha,\beta)} g(t) = g(\alpha) = \frac{1}{2\sqrt{\alpha}}. \quad (3.28)$$

The results (3.24) and (3.25) are obtained by using (3.26), (3.27), and (3.28) in (2.4) and (2.6) respectively.

Mutual information

Mutual information (Shannon, 1948) 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

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

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

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

where

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

is known as Marginal entropy (Shannon, 1948) and

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

is known as Conditional entropy (Shannon, 1948).

By viewing $K(P,Q)$ (Relative entropy (1.9)), 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 (1.9). So $I(X,Y)$ can also be written as

$$I(X,Y) = K\left(p(x,y), p(x)\,p(y)\right) = \sum_{(x,y)\in(X,Y)} p(x,y)\log\left[\frac{p(x,y)}{p(x)\,p(y)}\right]. \quad (4.5)$$

Similarly, we can define the Mutual information in following manners as well.

In $\chi^2(P,Q)$ manner: $I_{\chi^2}(X,Y) = \sum_{(x,y)\in(X,Y)} \frac{\left[p(x,y)-p(x)p(y)\right]^2}{p(x)p(y)}$, (4.6)

In $V(P,Q)$ manner: $I_V(X,Y) = \sum_{(x,y)\in(X,Y)} \left|p(x,y)-p(x)p(y)\right|$, (4.7)

where $\chi^2(P,Q)$ and $V(P,Q)$ are given by (1.7) and (1.8) respectively.

Proposition 4.1 For $\frac{1}{2} < \alpha \le \frac{p(x,y)+p(x)p(y)}{2p(x)p(y)} \le \beta < \infty \ \forall (x,y)\in(X,Y)$, we get the following new information inequalities in Mutual information sense

$$\begin{cases} \left|I(X,Y)-A(\alpha,\beta)\right| \le B(\alpha,\beta)\left[2+\frac{1}{(\beta-\alpha)^2} I_{\chi^2}(X,Y)\right], \frac{1}{2}<\alpha\le\frac{1}{2}\left(1+\frac{1}{e}\right) \\ \left|I(X,Y)-A(\alpha,\beta)\right| \le C(\alpha,\beta)\left[2+\frac{1}{(\beta-\alpha)^2} I_{\chi^2}(X,Y)\right], \frac{1}{2}\left(1+\frac{1}{e}\right)<\alpha\le 1 \end{cases} \tag{4.8}$$

and

$$\left|I(X,Y)-A(\alpha,\beta)\right| \le \frac{(2\beta-1)\log(2\beta-1)}{2}\left[2+\frac{1}{(\beta-\alpha)} I_V(X,Y)\right], \tag{4.9}$$

where

$$A(\alpha,\beta) = \frac{1}{2}\left[\frac{1}{2(\beta-\alpha)}\log\left\{\frac{(2\beta-1)^{(2\beta-1)^2}}{(2\alpha-1)^{(2\alpha-1)^2}}\right\} - 2L_1(\alpha,\beta)+1\right], \tag{4.10}$$

$$B(\alpha,\beta) = \frac{(\beta-\alpha)}{2}\left[\log\left(\frac{2\beta-1}{2\alpha-1}\right)+\left|\log(2\alpha-1)(2\beta-1)+2\right|\right], \tag{4.11}$$

$$C(\alpha,\beta) = \frac{(\beta-\alpha)}{2}\left[1+\log(2\beta-1)\right], \tag{4.12}$$

and $I(X,Y), I_{\chi^2}(X,Y), I_V(X,Y)$ are given by (4.5), (4.6), and (4.7) respectively.

Proof: Let

$f(t) = (2t-1)\log(2t-1), t\in\left(\frac{1}{2},\infty\right), f(1)=0, f'(t) = 2\left[1+\log(2t-1)\right]$ and

$f''(t) = \frac{4}{2t-1} > 0$.

Since $f''(t)>0 \forall t>\frac{1}{2}$ and $f(1)=0$, so $f(t)$ is convex and normalized function respectively. Now put $f(t)$ in (1.1) and put $f'(t)$ in (2.3) and (2.5) then after replacing $p_i, q_i \ \forall i=1,2,3,...n$ by $p(x,y)$, $p(x)p(y) \ \forall (x,y)\in(X,Y)$, we get the followings

$$S_f(P,Q)=\sum_{(x,y)\in(X,Y)} p(x,y)\log\left[\frac{p(x,y)}{p(x)p(y)}\right]=I(X,Y). \quad (4.13)$$

$$\mathop{A}_{\alpha}^{\beta}(f)=\int_{\alpha}^{\beta}\left|f'(t)\right|dt=2\int_{1/2}^{\beta}\left[1+\log(2t-1)\right]dt=(2\beta-1)\log(2\beta-1). \quad (4.14)$$

Now, let $g(t)=\left|f'(t)\right|=2\left|1+\log(2t-1)\right|=\begin{cases}-2\left[1+\log(2t-1)\right], & \frac{1}{2}<t\le\frac{1}{2}\left(1+\frac{1}{e}\right)\\ 2\left[1+\log(2t-1)\right], & \frac{1}{2}\left(1+\frac{1}{e}\right)<t<\infty\end{cases}$,

and $g'(t)=\begin{cases}-\frac{4}{(2t-1)}\le 0, & \frac{1}{2}<t\le\frac{1}{2}\left(1+\frac{1}{e}\right)\\ \frac{4}{(2t-1)}>0, & \frac{1}{2}\left(1+\frac{1}{e}\right)<t<\infty\end{cases}$.

It is clear that $g'(t)<0$ in $\left(\frac{1}{2},\frac{1}{2}\left(1+\frac{1}{e}\right)\right)$ and >0 in $\left(\frac{1}{2}\left(1+\frac{1}{e}\right),\infty\right)$, i.e., $g(t)$ is decreasing in $\left(\frac{1}{2},\frac{1}{2}\left(1+\frac{1}{e}\right)\right)$ and increasing in $\left(\frac{1}{2}\left(1+\frac{1}{e}\right),\infty\right)$, so

$$\left\|f'\right\|_{\infty}=\sup_{t\in(\alpha,\beta)}\left|f'(t)\right|=\sup_{t\in(\alpha,\beta)}g(t)=\begin{cases}\max\left[\left|f'(\alpha)\right|,\left|f'(\beta)\right|\right], & \frac{1}{2}<\alpha\le\frac{1}{2}\left(1+\frac{1}{e}\right)\\ \left|f'(\beta)\right|, & \frac{1}{2}\left(1+\frac{1}{e}\right)<\alpha\le 1\end{cases}$$

$$=\begin{cases}2\left[\log\left(\frac{2\beta-1}{2\alpha-1}\right)+\left|\log(2\alpha-1)(2\beta-1)+2\right|\right], & \frac{1}{2}<\alpha\le\frac{1}{2}\left(1+\frac{1}{e}\right)\\ 2\left[1+\log(2\beta-1)\right], & \frac{1}{2}\left(1+\frac{1}{e}\right)<\alpha\le 1\end{cases}. \quad (4.15)$$

The results (4.8) and (4.9) are obtained by using (4.6), (4.7), (4.13), (4.14), (4.15) in (2.4) and (2.6) after replacing p_i, q_i by $p(x,y), p(x)p(y)$ respectively.

Numerical approximation

If p_i and q_i are very close to each other and $p_i=p_i(\delta)$, $q_i=q_i(\delta)$, i.e.,

$\left|\frac{p_i(\delta)+q_i(\delta)}{2q_i(\delta)}-1\right|\le\delta, \delta\in(0,1) \ \forall \ i=1,2,3...,n$, with

$\beta = 1+\delta$ and $\alpha = 1-\delta$, then from (3.1) and (3.2) we obtain the followings

$$\left|F\left[Q(\delta),P(\delta)\right]-\left\{2-\log I(1-\delta,1+\delta)\right\}\right| \le \frac{\delta}{2(1-\delta)}\left[2+\frac{1}{4\delta^2}\chi^2\left[P(\delta),Q(\delta)\right]\right]. \quad (5.1)$$

$$\left|F\left[P(\delta),Q(\delta)\right]-\left\{2-\log I(1-\delta,1+\delta)\right\}\right| \le \left[2\delta+V\left[P(\delta),Q(\delta)\right]\right]L_{-1}(1-\delta,1+\delta). \quad (5.2)$$

We conclude from (5.1) and (5.2) that adjoint of the Relative JS divergence and Relative JS divergence $F(P,Q)$ respectively can be approximated by $\left[2-\log I(1-\delta,1+\delta)\right]$ and the error of the approximation is less than and equal to

$$\frac{\delta}{2(1-\delta)}\left[2+\frac{1}{4\delta^2}\chi^2\left[P(\delta),Q(\delta)\right]\right] \text{ and}$$

$$\left[2\delta+V\left[P(\delta),Q(\delta)\right]\right]L_{-1}(1-\delta,1+\delta)$$

respectively, for all $\delta \in (0,1)$.

In a similar manner, from (4.8) and (4.9), we can say that the Mutual information is approximated by $A(1-\delta,1+\delta)$ and error of approximation is less than and equal to

$$B(1-\delta,1+\delta)\left[2+\frac{1}{4\delta^2}I_{\chi^2}\left[X(\delta),Y(\delta)\right]\right] \text{ for } 0<\delta<1,$$

$$C(1-\delta,1+\delta)\left[2+\frac{1}{4\delta^2}I_{\chi^2}\left[X(\delta),Y(\delta)\right]\right] \text{ for } 0<\delta<\frac{e-1}{2e},$$

and

$$\frac{(2\delta+1)\log(2\delta+1)}{2}\left[2+\frac{1}{2\delta}I_V\left[X(\delta),Y(\delta)\right]\right] \text{ for } 0<\delta<1,$$

respectively.

Similarly, we can approximate and can find the error of the approximation for other results as well. We leave to the readers to verify these.

Numerical verification

In this section, we give two examples for calculating the divergences $\Delta(P,Q)$, $\chi^2(P,Q)$, $V(P,Q)$ and verify the inequalities (3.6) and (3.8).

Example 6.1 Let P be the binomial probability distribution with parameters $(n=10, p=0.5)$ and Q its approximated Poisson probability distribution with parameter $(\lambda = np = 5)$ for the random variable X, then we have

Table 1. $(n=10, p=0.5, q=0.5)$

x_i	0	1	2	3	4	5	6	7	8	9	10
$p(x_i)=p_i \approx$	.000976	.00976	.043	.117	.205	.246	.205	.117	.043	.00976	.000976
$q(x_i)=q_i \approx$	.00673	.033	.084	.140	.175	.175	.146	.104	.065	.036	.018
$\frac{p_i+q_i}{2q_i} \approx$	.573	.648	.757	.918	1.086	1.203	1.202	1.063	.831	.636	.527

By using Table 1, we get the followings.

$$\alpha\left(=0.527\right) \leq \frac{p_i + q_i}{2q_i} \leq \beta\left(=1.203\right). \tag{6.1}$$

$$\Delta\left(P,Q\right) = \sum_{i=1}^{11} \frac{\left(p_i - q_i\right)^2}{p_i + q_i} \approx 0.0917. \tag{6.2}$$

$$\chi^2\left(P,Q\right) = \sum_{i=1}^{11} \frac{\left(p_i - q_i\right)^2}{q_i} \approx 0.1471. \tag{6.3}$$

$$V\left(P,Q\right) = \sum_{i=1}^{11} \left|p_i - q_i\right| \approx 0.3312. \tag{6.4}$$

Put the approximated numerical values from (6.1) to (6.4) in (3.6) and (3.8) and verify them for $p = 0.5$. We omit the details.

Example 6.2 Let P be the binomial probability distribution with parameters $\left(n = 10, p = 0.7\right)$ and Q its approximated Poisson probability distribution with parameter $\left(\lambda = np = 7\right)$ for the random variable X, then we have

Table 2. $\left(n = 10, p = 0.7, q = 0.3\right)$

x_i	0	1	2	3	4	5	6	7	8	9	10
$p\left(x_i\right) = p_i \approx$	.0000059	.000137	.00144	.009	.036	.102	.20	.266	.233	.121	.0282
$q\left(x_i\right) = q_i \approx$	.000911	.00638	.022	.052	.091	.177	.199	.149	.130	.101	.0709
$\frac{p_i + q_i}{2q_i} \approx$	.503	.510	.532	.586	.697	.788	1.002	1.392	1.396	1.099	.698

By using Table 2, we get the followings.

$$\alpha\left(=0.503\right) \leq \frac{p_i + q_i}{2q_i} \leq \beta\left(=1.396\right). \tag{6.5}$$

$$\Delta\left(P,Q\right) = \sum_{i=1}^{11} \frac{\left(p_i - q_i\right)^2}{p_i + q_i} \approx 0.1812. \tag{6.6}$$

$$\chi^2\left(P,Q\right) = \sum_{i=1}^{11} \frac{\left(p_i - q_i\right)^2}{q_i} \approx 0.3298. \tag{6.7}$$

$$V\left(P,Q\right) = \sum_{i=1}^{11} \left|p_i - q_i\right| \approx \approx 0.4844. \tag{6.8}$$

Put the approximated numerical values from (6.5) to (6.8) in (3.6) and (3.8) and verify them for $p = 0.7$. We omit the details. Similarly, we can verify the other inequalities (3.1), (3.2), (3.13), (3.14), (3.18), (3.20), (3.24), and (3.25).

CONCLUSION AND DISCUSSION

In this work, we presented new information inequalities on $S_f\left(P,Q\right)$. Further, we relate various well known divergences to the Chi- square divergence and Variational distance separately in an interval $\left(\alpha,\beta\right), 0<\alpha\leq 1\leq\beta<\infty$ with $\alpha\neq\beta$ as an application of new inequalities. These relations have been verified numerically by taking two discrete distributions: Binomial and Poisson. A numerical approximation has been done as well, which shows that we can approximate a particular divergence and can find the error of approximation. Lastly, a very important application to the Mutual information has been discussed, which tells us how far the joint distribution is from its independency, i.e., $I\left(X,Y\right)=0=I_{\chi^2}\left(X,Y\right)=I_V\left(X,Y\right)$ if distributions are independent to each other.

We found in articles (Bhatia and Singh, 2013 and Jain and Chhabra, 2014) that square root of some particular divergences is a metric space but not each because of violation of triangle inequality, so we strongly believe that divergence measures can be extended to other significant problems of functional analysis and its applications, such investigations are actually in progress because this is also an area worth being investigated. Such types of divergences are also very useful to find the utility of an event (Bhullar *et al*., 2010 and Taneja and Tuteja, 1986)[, i.e., an event is how much useful compare to other event.

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.

REFERENCES

Bajaj R. K. and Hooda D.S., "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.

Bassat M.B., "f- Entropies, probability of error and feature selection", Inform. Control, vol. 39, 1978, pp: 227-242.

Bhatia P.K. and Singh S., "On a new Csiszar's f- divergence measure", Cybernetics and Information Technologies, vol. 13 (2013), no. 2, pp: 43- 56.

Bhullar J.S., Vinocha O.P., and Gupta M., "Generalized measure for two utility distributions", Proceedings of the World Congress on Engineering, vol. 3, June 30- July 2, 2010.

Bregman L.M., "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.

Burbea J. and Rao C.R, "On the convexity of some divergence measures based on entropy functions", IEEE Trans. on Inform. Theory, IT-28 (1982), pp: 489-495.

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

Chow C.K. and Lin C.N., "Approximating discrete probability distributions with dependence trees", IEEE Trans. Inform. Theory, vol. 14, 1968, no 3, pp: 462-467.

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

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

Dacunha- Castelle D., "Ecole d'Ete de Probabilites de", Saint-Flour VII-1977, Berlin, Heidelberg, New York: Springer, 1978.

Dragomir S.S., "On the Ostrowski's integral inequality for mappings of bounded variation and applications", RGMIA Res. Rep. Coll. 2 (1999) no.1, pp: 73–80.

Dragomir S.S., Gluscevic V., and Pearce C.E.M., "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.

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

Hellinger E., "Neue begrundung der theorie der quadratischen formen von unendlichen vielen veranderlichen", J. Rein.Aug. Math., 136 (1909), pp: 210-271.

Hooda D.S., "On generalized measures of fuzzy entropy", Mathematica Slovaca, vol. 54 (2004), no. 3, pp: 315- 325.

Jain K.C. and Chhabra P., "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.

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

Jha P. and Mishra V. K., "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.

Jones L. and Byrne C., "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.

Kadota T.T. and Shepp L.A., "On the best finite set of linear observables for discriminating two Gaussian signals", IEEE Trans. Inform. Theory, vol. 13, 1967, pp: 288-294.
Kailath T, "The divergence and Bhattacharyya distance measures in signal selection", IEEE Trans. Comm. Technology, vol. COM-15, 1967, pp: 52- 60.
Kazakos D. and Cotsidas T, "A decision theory approach to the approximation of discrete probability densities", IEEE Trans. Perform. Anal. Machine Intell, vol. 1, 1980, pp: 61- 67.
Kolmogorov A.N., "On the approximation of distributions of sums of independent summands by infinitely divisible distributions", Sankhya, 25, pp: 159-174.
Kullback S. and Leibler R.A., "On information and sufficiency", Ann. Math. Statist., 22 (1951), pp: 79-86.
Nielsen F. and Boltz S., "The Burbea-Rao and Bhattacharyya centroids", Apr. 2010, Arxiv.
Pearson K., "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.
Pielou E.C, "Ecological diversity", New York, Wiley, 1975.
Renyi A., "On measures of entropy and information", Proc. 4th Berkeley Symposium on Math. Statist. and Prob., 1 (1961), pp: 547-561.
Santos-Rodriguez R., Garcia-Garcia D., and Cid-Sueiro J., "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.
Shannon C.E., "A mathematical theory of communication", Bull. Sept. Tech. J., 27 (1948), 370-423 and 623-656.
Sibson R., "Information radius", Z. Wahrs. Undverw. Geb., (14) (1969), pp: 149-160.
Taneja H.C. and Tuteja R.K., "Characterization of a quantitative- qualitative measure of inaccuracy", Kybernetika, vol. 22 (1986), no. 5, pp: 393- 402.
Taneja I.J., "New developments in generalized information measures", Chapter in: Advances in Imaging and Electron Physics, Ed. P.W. Hawkes, 91(1995), pp: 37-135.
Taskar B., Lacoste-Julien S., and Jordan M.I., "Structured prediction, dual extra gradient and Bregman projections", Journal of Machine Learning Research, 7, pp: 1627- 1653, July 2006.
Theil H "Statistical decomposition analysis", Amsterdam, North-Holland, 1972.
Theil H. "Economics and information theory", Amsterdam, North-Holland, 1967.
Vemuri B., Liu M., Amari S., and Nielsen F., "Total Bregman divergence and its applications to DTI analysis", IEEE Transactions on Medical Imaging, 2010.

****