J. Appl. Math. & Informatics Vol. **35***(2017), No. 3 - 4, pp.371 - 385*
https://doi.org/10.14317/jami.2017.371

NEW INFORMATION INEQUALITIES ON ABSOLUTE VALUE OF THE FUNCTIONS AND ITS APPLICATION

PRAPHULL CHHABRA

Abstract. Jain and Saraswat (2012) introduced new generalized f- information divergence measure, by which we obtained many well known and new information divergences.
In this work, we introduce new information inequalities in absolute form on this new generalized divergence by considering convex normalized functions. Further, we apply these inequalities for getting new relations among well known divergences, together with numerical verification. Application to the Mutual information is also presented. Asymptotic approximation in terms of Chi- square divergence is done as well.

AMS Mathematics Subject Classification: 94A17, 26D15.
Key words and phrases : New generalized f- divergence, New information inequalities, Convex and normalized function, Asymptotic approximation, Mutual information.

1. Introduction

Divergence measures have been demonstrated very useful in a variety of disciplines such as economics and political science [30, 31], biology [23], analysis of contingency tables [10], approximation of probability distributions [5, 18], signal processing [16, 17], pattern recognition [1, 4, 15], color image segmentation [21], 3D image segmentation and word alignment [29], cost- sensitive classification for medical diagnosis [25], magnetic resonance image analysis [32] etc.
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. 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 information divergence measures had been introduced, characterized and applied in variety of fields. Such as: Csiszar's f- divergence [6, 7], Bregman's f- divergence [2], Burbea- Rao's f- divergence [3], Renyi's like f-

Received December 22, 2016. Revised February 26, 2017. Accepted March 25, 2017.

divergence [24]. Similarly, Jain and Saraswat [14] defined new generalized f-divergence measure, which is given by

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

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.

Many divergence measures can be obtained from these generalized f- measures by suitably defining the function f. Specially $C_f(P,Q)$ and $S_f(P,Q)$ are widely used due to its compact nature. Some resultant divergences by $S_f(P,Q)$, are as follows.

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

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

where $F(Q,P)$ is called adjoint of the Relative JS divergence $F(P,Q)$ [27].

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

$$S_f(P,Q)=\frac{1}{2}\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{p_i+q_i}=\frac{1}{2}\Delta(P,Q), \tag{3}$$

where $\Delta(P,Q)$ is called the Triangular discrimination [8].

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

$$S_f(P,Q)=\sum_{i=1}^{n}\frac{p_i+q_i}{2}\log\left(\frac{p_i+q_i}{2q_i}\right)=G(Q,P), \tag{4}$$

where $G(Q,P)$ is called adjoint of the Relative AG divergence $G(P,Q)$ [28].

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

$$S_f(P,Q)=\frac{1}{2}\sum_{i=1}^{n}(p_i-q_i)\log\left(\frac{p_i+q_i}{2q_i}\right)=\frac{1}{2}J_R(P,Q), \tag{5}$$

where $J_R(P,Q)$ is called the Relative J- divergence [9].

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

$$S_f(P,Q)=\frac{1}{4}\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{q_i}=\frac{1}{4}\chi^2(P,Q), \tag{6}$$

where $\chi^2(P,Q)$ is called the Chi- square divergence or Pearson divergence measure [22].

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

$$S_f(P,Q)=\frac{1}{2}\sum_{i=1}^{n}|p_i-q_i|=\frac{1}{2}V(P,Q), \tag{7}$$

where $V(P,Q)$ is called the Variational distance (l_1 distance) [19].
(g). Particularly, by taking $f(t) = (2t-1)\log(2t-1), t \in \left(\frac{1}{2}, \infty\right)$ in (1), we obtain

$$S_f(P,Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} = K(P,Q), \tag{8}$$

where $K(P,Q)$ is called the Relative entropy (Kullback- Leibler distance) [20]. Similarly, we can obtain many divergences by using linear convex functions. Since these divergences are not worthful in practice, therefore we can skip them. Now, for a differentiable function $f : \left(\frac{1}{2}, \infty\right) \to R$, consider the associated function $g : \left(\frac{1}{2}, \infty\right) \to R$, which is given by

$$g(t) = (t-1) f'\left(\frac{t+1}{2}\right). \tag{9}$$

Put (9) in (1), we get

$$E^*_{S_f}(P,Q) = \sum_{i=1}^{n} \left(\frac{p_i - q_i}{2}\right) f'\left(\frac{p_i + 3q_i}{4q_i}\right). \tag{10}$$

2. New Information Inequalities

In this section, we introduce new information inequalities on absolute value of the functions for $S_f(P,Q)$. Such inequalities are for instance needed in order to calculate the relative efficiency of two divergences.

Theorem 2.1. *Let f_1 and f_2 be two real, convex and normalized differentiable functions of $(\alpha,\beta) \subset (0,\infty)$, where $0 < \alpha \le 1 \le \beta < \infty, \alpha \ne \beta$. If there exists the real constants m, M such that $m < M$, then*

$$m \le \frac{|f_1(t_1) - f_1(t_2)|}{|f_2(t_1) - f_2(t_2)|} \le M. \tag{11}$$

By using Cauchy's theorem equation (11) can be written as

$$m \le \frac{|f_1'(t)|}{|f_2'(t)|} = \left|\frac{f_1'(t)}{f_2'(t)}\right| \le M, \tag{12}$$

for all $t_1, t_2 \in (\alpha,\beta) \subset (0,\infty)$.
If $P, Q \in \Gamma_n$ is such that $0 < \alpha \le \frac{p_i+q_i}{2q_i} \le \beta < \infty \; \forall \; i = 1,2,3...,n$, then we have the following inequalities

$$mS_{|f_2|}(P,Q) \le S_{|f_1|}(P,Q) \le MS_{|f_2|}(P,Q), \tag{13}$$

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

Proof. Firstly, we can see that (12) is obtained from (11) by using Cauchy's theorem.
Now put $t_1 = \frac{p_i+q_i}{2q_i}$ and $t_2 = 1$ in (12), multiply with q_i and then sum over all $i = 1,2,3...,n$, we get desire result (13). □

3. Application of New Information Inequalities

In this section, we will obtain bounds of different divergences in terms of the Variational distance by using new inequalities (13). For this, let $f_2 : (0,\infty) \to R$ be a function defined as

$$f_2(t) = |t-1|, f_2(1) = 0, f_2'(t) = \begin{cases} -1 & \text{if } 0 < t < 1 \\ 1 & \text{if } 1 < t < \infty \end{cases}, f_2''(t) = 0 \; \forall\, t \in (0,\infty)$$

but $f(t)$ is not differentiable at $t = 1$ and

$$|f_2'(t)| = 1. \tag{14}$$

Since $f_2''(t) \geq 0 \; \forall\, t > 0 - [1]$ and $f_2(1) = 0$, so $f_2(t)$ is convex and normalized function respectively. Now put $f_2(t)$ in (1), we get

$$S_{|f_2|}(P,Q) = \frac{1}{2}\sum_{i=1}^{n} |p_i - q_i| = \frac{1}{2} V(P,Q). \tag{15}$$

Proposition 3.1. *Let $V(P,Q)$ and $F(P,Q)$ be defined as in (7) and (2) respectively. For $P,Q \in \Gamma_n$, we have*

$$\frac{1}{2\beta} V(P,Q) \leq |F|(P,Q) \leq \frac{1}{2\alpha} V(P,Q). \tag{16}$$

Proof. Let us consider

$$f_1(t) = -\log t, t \in (0,\infty), f_1(1) = 0, f_1'(t) = -\frac{1}{t} \text{ and } f_1''(t) = \frac{1}{t^2}.$$

Since $f_1''(t) > 0 \; \forall\, t > 0$ and $f_1(1) = 0$, so $f_1(t)$ is convex and normalized function respectively.
Now put $f_1(t)$ in (1), we get

$$S_{|f_1|}(P,Q) = \sum_{i=1}^{n} q_i \left|\log\left(\frac{2q_i}{p_i + q_i}\right)\right| = |F|(Q,P). \tag{17}$$

Now, let $g(t) = \left|\frac{f_1'(t)}{f_2'(t)}\right| = \left|-\frac{1}{t}\right| = \frac{1}{t}$, where $|f_2'(t)| = 1$ and $g'(t) = -\frac{1}{t^2} < 0$.
It is clear that $g(t)$ is always decreasing in $(0,\infty)$, so

$$m = \inf_{t\in(\alpha,\beta)} g(t) = g(\beta) = |f_1'(\beta)| = \frac{1}{\beta}. \tag{18}$$

$$M = \sup_{t\in(\alpha,\beta)} g(t) = g(\alpha) = |f_1'(\alpha)| = \frac{1}{\alpha}. \tag{19}$$

The result (16) is obtained by using (15), (17), (18), and (19) in (13), after interchanging P and Q. □

Proposition 3.2. *Let $V(P,Q)$ and $\Delta(P,Q)$ be defined as in (7) and (3) respectively. For $P,Q \in \Gamma_n$, we have*
(a). If $0<\alpha<1$, then

$$0 \leq \Delta(P,Q) \leq \frac{1}{2}\left[\frac{\beta^2-\alpha^2}{\alpha^2\beta^2} + \left|\frac{\beta^2+\alpha^2}{\alpha^2\beta^2} - 2\right|\right] V(P,Q). \tag{20}$$

(b). If $\alpha = 1$, then

$$0 \leq \Delta(P,Q) \leq \frac{\beta^2-1}{\beta^2} V(P,Q). \tag{21}$$

Proof. Let us consider

$$f_1(t) = \frac{(t-1)^2}{t}, t \in (0,\infty), f_1(1) = 0, f_1'(t) = \frac{t^2-1}{t^2} \text{ and } f_1''(t) = \frac{2}{t^3}.$$

Since $f_1''(t) > 0\ \forall\ t > 0$ and $f_1(1) = 0$, so $f_1(t)$ is convex and normalized function respectively.
Now put $f_1(t)$ in (1), we get

$$S_{|f_1|}(P,Q) = \frac{1}{2}\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{p_i+q_i} = \frac{1}{2}\Delta(P,Q). \tag{22}$$

Now, let

$$g(t) = \left|\frac{f_1'(t)}{f_2'(t)}\right| = \left|\frac{t^2-1}{t^2}\right| = \begin{cases} -\left(\frac{t^2-1}{t^2}\right) & \text{if } 0<t<1 \\ \frac{t^2-1}{t^2} & \text{if } 1 \leq t < \infty \end{cases},$$

where $|f_2'(t)| = 1$ and $g'(t) = \begin{cases} -\frac{2}{t^3} < 0 & \text{if } 0<t<1 \\ \frac{2}{t^3} > 0 & \text{if } 1 \leq t < \infty \end{cases}$.
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 $g(t)$ has a minimum value at $t=1$, therefore

$$m = \inf_{t\in(0,\infty)} g(t) = g(1) = |f_1'(1)| = 0. \tag{23}$$

$$\begin{aligned} M &= \sup_{t\in(\alpha,\beta)} g(t) \\ &= \begin{cases} max\left(|f_1'(\alpha)|, |f_1'(\beta)|\right) = \frac{|f_1'(\alpha)|+|f_1'(\beta)|+\left||f_1'(\alpha)|-|f_1'(\beta)|\right|}{2} & \text{if } 0<\alpha<1 \\ |f_1'(\beta)| & \text{if } \alpha = 1 \end{cases}, \end{aligned}$$

i.e.,

$$M = \begin{cases} \frac{1}{2}\left[\frac{\beta^2-\alpha^2}{\alpha^2\beta^2} + \left|\frac{\beta^2+\alpha^2}{\alpha^2\beta^2} - 2\right|\right] & \text{if } 0<\alpha<1 \\ \frac{\beta^2-1}{\beta^2} & \text{if } \alpha = 1 \end{cases}. \tag{24}$$

The results (20) and (21) are obtained by using (15), (22), (23), and (24) in (13). □

Proposition 3.3. *Let $V(P,Q)$ and $G(P,Q)$ be defined as in (7) and (4) respectively. For $P,Q \in \Gamma_n$, we have*
(a). If $0 < \alpha \leq \frac{1}{e}$, then

$$0 \leq |G|(P,Q) \leq \frac{1}{2}\left[\log\sqrt{\frac{\beta}{\alpha}} + \left|\log e\sqrt{\alpha\beta}\right|\right] V(P,Q). \tag{25}$$

(b). If $\frac{1}{e} < \alpha \leq 1$, then

$$\frac{\log e\alpha}{2} V(P,Q) \leq |G|(P,Q) \leq \frac{\log e\beta}{2} V(P,Q). \tag{26}$$

Proof. Let us consider

$$f_1(t) = t\log t, t \in (0,\infty), f_1(1) = 0, f_1'(t) = 1 + \log t \text{ and } f_1''(t) = \frac{1}{t}.$$

Since $f_1''(t) > 0 \ \forall\ t > 0$ and $f_1(1) = 0$, so $f_1(t)$ is convex and normalized function respectively.
Now put $f_1(t)$ in (1), we get

$$S_{|f_1|}(P,Q) = \sum_{i=1}^{n}\left(\frac{p_i + q_i}{2}\right)\left|\log\left(\frac{p_i + q_i}{2q_i}\right)\right| = |G|(Q,P). \tag{27}$$

Now, let

$$g(t) = \left|\frac{f_1'(t)}{f_2'(t)}\right| = |1 + \log t| = \begin{cases} -(1 + \log t) & \text{if } 0 < t \leq \frac{1}{e} \\ 1 + \log t & \text{if } \frac{1}{e} < t < \infty \end{cases},$$

where $|f_2'(t)| = 1$ and $g'(t) = \begin{cases} -\frac{1}{t} < 0 & \text{if } 0 < t \leq \frac{1}{e} \\ \frac{1}{t} > 0 & \text{if } \frac{1}{e} < t < \infty \end{cases}$.
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 $g(t)$ has a minimum value at $t = \frac{1}{e}$, therefore

$$m = \inf_{t\in(\alpha,\beta)} g(t) = \begin{cases} \left|f_1'\left(\frac{1}{e}\right)\right| = 0 & \text{if } 0 < \alpha \leq \frac{1}{e} \\ |f_1'(\alpha)| = 1 + \log\alpha & \text{if } \frac{1}{e} < \alpha \leq 1 \end{cases}. \tag{28}$$

$$\begin{aligned} M &= \sup_{t\in(\alpha,\beta)} g(t) \\ &= \begin{cases} max\left(|f_1'(\alpha)|, |f_1'(\beta)|\right) = \left[\log\sqrt{\frac{\beta}{\alpha}} + \left|\log e\sqrt{\alpha\beta}\right|\right] & \text{if } 0 < \alpha \leq \frac{1}{e} \\ |f_1'(\beta)| = 1 + \log\beta & \text{if } \frac{1}{e} < \alpha \leq 1 \end{cases}. \end{aligned} \tag{29}$$

The results (25) and (26) are obtained by using (15), (27), (28), and (29) in (13), after interchanging P and Q. □

Proposition 3.4. *Let $V(P,Q)$ and $J_R(P,Q)$ be defined as in (7) and (5) respectively. For $P,Q \in \Gamma_n$, we have* *(a). If $0 < \alpha < 1$, then*

$$0 \leq |J_R|(P,Q) \leq \left[\log\sqrt{\frac{\beta}{\alpha}} + \frac{\beta-\alpha}{2\alpha\beta} + \left|\frac{\beta+\alpha}{2\alpha\beta} - \log e\sqrt{\alpha\beta}\right|\right] V(P,Q). \quad (30)$$

(b). If $\alpha = 1$, then

$$0 \leq |J_R|(P,Q) \leq \left(\log e\beta - \frac{1}{\beta}\right) V(P,Q). \quad (31)$$

Proof. Let us consider

$$f_1(t) = (t-1)\log t, t \in (0,\infty), f_1(1) = 0,$$
$$f_1'(t) = \frac{t-1}{t} + \log t \text{ and } f_1''(t) = \frac{t+1}{t^2}.$$

Since $f_1''(t) > 0 \ \forall \ t > 0$ and $f_1(1) = 0$, so $f_1(t)$ is convex and normalized function respectively.
Now put $f_1(t)$ in (1), we get

$$S_{|f_1|}(P,Q) = \frac{1}{2}\sum_{i=1}^{n}\left|(p_i - q_i)\log\left(\frac{p_i+q_i}{2q_i}\right)\right| = \frac{1}{2}|J_R|(P,Q). \quad (32)$$

Now, let

$$g(t) = \left|\frac{f_1'(t)}{f_2'(t)}\right| = \left|\frac{t-1}{t} + \log t\right| = \begin{cases} -\left(\frac{t-1}{t} + \log t\right) & \text{if } 0 < t < 1 \\ \frac{t-1}{t} + \log t & \text{if } 1 \leq t < \infty \end{cases},$$

where $|f_2'(t)| = 1$ and $g'(t) = \begin{cases} -\left(\frac{t+1}{t^2}\right) < 0 & \text{if } 0 < t < 1 \\ \frac{t+1}{t^2} > 0 & \text{if } 1 \leq t < \infty \end{cases}$.
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 $g(t)$ has a minimum value at $t = 1$, therefore

$$m = \inf_{t\in(0,\infty)} g(t) = g(1) = |f_1'(1)| = 0. \quad (33)$$

$$\begin{aligned} M &= \sup_{t\in(\alpha,\beta)} g(t) \\ &= \begin{cases} max\left(|f_1'(\alpha)|, |f_1'(\beta)|\right) = \left[\log\sqrt{\frac{\beta}{\alpha}} + \frac{\beta-\alpha}{2\alpha\beta} + \left|\frac{\beta+\alpha}{2\alpha\beta} - \log e\sqrt{\alpha\beta}\right|\right] & \text{if } 0 < \alpha < 1 \\ \left(\log e\beta - \frac{1}{\beta}\right) & \text{if } \alpha = 1 \end{cases}. \end{aligned} \quad (34)$$

The results (30) and (31) are obtained by using (15), (32), (33), and (34) in (13). □

Proposition 3.5. *Let $V(P,Q)$ and $\chi^2(P,Q)$ be defined as in (7) and (6) respectively. For $P, Q \in \Gamma_n$, we have*
(a). If $0 < \alpha < 1$, then

$$0 \leq \chi^2(P,Q) \leq 2\left[\beta - \alpha + |2 - (\alpha+\beta)|\right] V(P,Q). \quad (35)$$

(b). If $\alpha = 1$, then

$$0 \leq \chi^2(P,Q) \leq 4(\beta - 1) V(P,Q). \tag{36}$$

Proof. Let us consider

$$f_1(t) = (t-1)^2, t \in (0,\infty), f_1(1) = 0, f_1'(t) = 2(t-1) \text{ and } f_1''(t) = 2.$$

Since $f_1''(t) > 0 \ \forall\ t > 0$ and $f_1(1) = 0$, so $f_1(t)$ is convex and normalized function respectively.
Now put $f_1(t)$ in (1), we get

$$S_{|f_1|}(P,Q) = \frac{1}{4}\sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i} = \frac{1}{4}\chi^2(P,Q). \tag{37}$$

Now, let

$$g(t) = \left|\frac{f_1'(t)}{f_2'(t)}\right| = |2(t-1)| = \begin{cases} -2(t-1) & \text{if } 0 < t < 1 \\ 2(t-1) & \text{if } 1 \leq t < \infty \end{cases},$$

where $|f_2'(t)| = 1$ and $g'(t) = \begin{cases} -2 < 0 & \text{if } 0 < t < 1 \\ 2 > 0 & \text{if } 1 \leq t < \infty \end{cases}$.

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 $g(t)$ has a minimum value at $t = 1$, therefore

$$m = \inf_{t\in(0,\infty)} g(t) = g(1) = |f_1'(1)| = 0. \tag{38}$$

$$\begin{aligned} M &= \sup_{t\in(\alpha,\beta)} g(t) \\ &= \begin{cases} max(|f_1'(\alpha)|, |f_1'(\beta)|) = [\beta - \alpha + |2 - (\alpha+\beta)|] & \text{if } 0 < \alpha < 1 \\ 2(\beta - 1) & \text{if } \alpha = 1 \end{cases}. \end{aligned} \tag{39}$$

The results (35) and (36) are obtained by using (15), (37), (38), and (39) in (13). □

4. Numerical Verification

In this section, we give two examples for calculating the divergences $|F|(P,Q), \Delta(P,Q), |J_R|(P,Q)$, and $V(P,Q)$ and then verify the inequalities (16), (20), and (30), numerically.

Example 4.1. 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

x_i	0	1	2	3	4	5	6	7	8	9	10
$p_i \approx$	.000976	.00976	.043	.117	.205	.246	.205	.117	.043	.00976	.000976
$q_i \approx$	.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

TABLE 1. Evaluation of probability distributions for $(n = 10, p = 0.5, q = 0.5)$

By using Table 1, we get the followings.

$$\alpha\,(=.527) \leq \frac{p_i+q_i}{2q_i} \leq \beta\,(=1.203)\,. \tag{40}$$

$$|F|\,(P,Q) = \sum_{i=1}^{11} p_i \left|\log\left(\frac{2p_i}{p_i+q_i}\right)\right| \approx .1495. \tag{41}$$

$$\Delta\,(P,Q) = \sum_{i=1}^{11} \frac{(p_i-q_i)^2}{p_i+q_i} \approx .0917. \tag{42}$$

$$|J_R|\,(P,Q) = \sum_{i=1}^{11} \left|(p_i-q_i)\log\left(\frac{p_i+q_i}{2q_i}\right)\right| \approx .0808. \tag{43}$$

$$V\,(P,Q) = \sum_{i=1}^{11} |p_i-q_i| \approx .3312. \tag{44}$$

Put the approximated numerical values from (40) to (44) in (16), (20), and (30), we obtain the followings respectively.

$$.1376 \leq .1495\,(|F|\,(P,Q)) \leq .3142,$$

$$0 \leq .0917\,(\Delta\,(P,Q)) \leq .8613,$$

and

$$0 \leq .0808\,(|J_R|\,(P,Q)) \leq .51042.$$

Hence verified the inequalities (16), (20), and (30) for $p = 0.5$.

Example 4.2. 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
$\frac{p_i+q_i}{2q_i} \approx$	.503	.510	.532	.586	.697	.788	1.002	1.392	1.396	1.099	.698

TABLE 2. Evaluation of probability distributions for $(n=10, p=0.7, q=0.3)$

By using Table 2, we get the followings.

$$\alpha\,(=.503) \leq \frac{p_i+q_i}{2q_i} \leq \beta\,(=1.396)\,. \tag{45}$$

$$|F|\,(P,Q) = \sum_{i=1}^{11} p_i \left|\log\left(\frac{2p_i}{p_i+q_i}\right)\right| \approx .21792. \tag{46}$$

$$\Delta\,(P,Q) = \sum_{i=1}^{11} \frac{(p_i-q_i)^2}{p_i+q_i} \approx .1812. \tag{47}$$

$$|J_R|(P,Q)=\sum_{i=1}^{11}\left|(p_i-q_i)\log\left(\frac{p_i+q_i}{2q_i}\right)\right|\approx .1686. \tag{48}$$

$$V(P,Q)=\sum_{i=1}^{11}|p_i-q_i|\approx .4844. \tag{49}$$

Put the approximated numerical values from (45) to (49) in (16), (20), and (30), we obtain the followings respectively.

$$.1734\leq .2179\left(|F|(P,Q)\right)\leq .4815,$$

$$0\leq .1812\left(\Delta(P,Q)\right)\leq 1.4301,$$

and

$$0\leq .1686\left(|J_R|(P,Q)\right)\leq .8129.$$

Hence verified the inequalities (16), (20), and (30) for $p=0.7$.

Remark 4.1. Similarly we can verify the inequalities for different values of p and q and for different discrete probability distributions, like: Geometric, Negative Binomial, Uniform etc.

5. Application to the Mutual Information

Mutual information [26] 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

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

By viewing $K(P,Q)$ (Relative entropy (8)), 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 (8). So $I(X,Y)$ can also be written as

$$I(X,Y)=K\left(p(x,y),p(x)p(y)\right)=\sum_{(x,y)\in(X,Y)}p(x,y)\log\frac{p(x,y)}{p(x)p(y)}. \tag{50}$$

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

In $|F|(P,Q)$ manner:

$$I_{|F|}(X,Y)=\sum_{(x,y)\in(X,Y)}p(x,y)\left|\log\frac{2p(x,y)}{p(x,y)+p(x)p(y)}\right|, \tag{51}$$

In $\chi^2(P,Q)$ manner:

$$I_{\chi^2}(X,Y) = \sum_{(x,y)\in(X,Y)} \frac{\left[p(x,y) - p(x)\,p(y)\right]^2}{p(x)\,p(y)}, \tag{52}$$

In $J_R(P,Q)$ manner:

$$I_{J_R}(X,Y) = \sum_{(x,y)\in(X,Y)} \left(p(x,y) - p(x)\,p(y)\right)\log\frac{p(x,y) + p(x)\,p(y)}{2p(x)\,p(y)}, \tag{53}$$

In $V(P,Q)$ manner:

$$I_V(X,Y) = \sum_{(x,y)\in(X,Y)} \left|p(x,y) - p(x)\,p(y)\right|, \tag{54}$$

and In $|G|(P,Q)$ manner:

$$I_{|G|}(X,Y) = \sum_{(x,y)\in(X,Y)} \left(\frac{p(x,y) + p(x)\,p(y)}{2}\right)\left|\log\frac{p(x,y) + p(x)\,p(y)}{2p(x)\,p(y)}\right|, \tag{55}$$

where $F(P,Q), J_R(P,Q), \chi^2(P,Q), V(P,Q)$, and $G(P,Q)$ are given by (2), (5), (6), (7), and (4) respectively.
So (50) to (55) tell us that how far the joint distribution is from its independency or

$$I(X,Y) = 0 = I_{|F|}(X,Y) = I_{\chi^2}(X,Y) = I_{J_R}(X,Y) = I_V(X,Y) = I_{|G|}(X,Y)$$

if distributions are independent to each other.
For applications of mutual information, the papers [11] and [12] due to Jain and Chhabra are referred.
Now we introduce the following proposition to obtain results in mutual information sense.

Proposition 5.1. *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*

$$\begin{aligned} &\left|I(X,Y) - I_{J_R}(X,Y)\right| \le \frac{(\beta-\alpha)}{4(2\alpha-1)(2\beta-1)} I_{\chi^2}(X,Y) \\ &\le \frac{(\beta-\alpha)}{2(2\alpha-1)(2\beta-1)}\left(\beta-\alpha+|2-(\alpha+\beta)|\right) I_V(X,Y) \\ &\le \frac{\beta(\beta-\alpha)}{(2\alpha-1)(2\beta-1)}\left(\beta-\alpha+|2-(\alpha+\beta)|\right) I_{|F|}(X,Y) \end{aligned} \tag{56}$$

and

$$\begin{aligned} &\left|I(X,Y) - I_{J_R}(X,Y)\right| \le \frac{1}{2}\log\left(\frac{2\beta-1}{2\alpha-1}\right) I_V(X,Y) \\ &\le \frac{1}{1+\log\alpha}\log\left(\frac{2\beta-1}{2\alpha-1}\right) I_{|G|}(X,Y). \end{aligned} \tag{57}$$

Proof. Let us consider

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

$$\text{and } f''(t) = \frac{4}{2t-1}. \tag{58}$$

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) and put $f'(t)$ in (10) then replace $p_i, q_i \; \forall i = 1, 2, ..., n$ by $p(x,y), p(x)p(y) \; \forall (x,y) \in (X,Y)$, we get

$$S_f(P,Q) = \sum_{(x,y)\in(X,Y)} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} = I(X,Y) \tag{59}$$

and

$$E^*_{S_f}(P,Q) = \sum_{(x,y)\in(X,Y)} \left(p(x,y) - p(x)p(y)\right)\log\frac{p(x,y) + p(x)p(y)}{2p(x)p(y)} = I_{J_R}(X,Y) \tag{60}$$

respectively, and

$$A^\beta_\alpha(f') = \int_\alpha^\beta |f''(t)|\, dt = \int_\alpha^\beta \left|\frac{4}{2t-1}\right| dt = 2\log\left(\frac{2\beta-1}{2\alpha-1}\right). \tag{61}$$

Now, let $g(t) = f''(t) = \frac{4}{2t-1}$, where $f''(t)$ is given by (58) and $g'(t) = -\frac{2}{\left(t-\frac{1}{2}\right)^2} < 0$.
It is clear that $g(t)$ is always decreasing in $\left(\frac{1}{2}, \infty\right)$, so

$$m = \inf_{t\in(\alpha,\beta)} g(t) = g(\beta) = \frac{4}{2\beta-1}. \tag{62}$$

$$M = \sup_{t\in(\alpha,\beta)} g(t) = g(\alpha) = \frac{4}{2\alpha-1}. \tag{63}$$

The results (56) and (57) are obtained by using (50) to (55), (59) to (63) together with first inequality of (16) and second inequality of (35) in [11] and first inequality of (25) in [12] after replacing p_i, q_i by $p(x,y), p(x)p(y)$ respectively. □

6. Asymptotic Approximation on $S_f(P,Q)$

In this section, we introduce asymptotic approximation of the generalized f-divergence measure (1) in terms of well known Chi- square divergence (6).

Theorem 6.1. *If $f : (0,\infty) \to R$ is twice differentiable, convex, and normalized function, i.e., $f''(t) > 0$ and $f(1) = 0$ respectively, then we have*

$$S_f(P,Q) \approx \frac{f''(1)}{8}\chi^2(P,Q). \tag{64}$$

Equivalently

$$\left|\frac{S_f(P,Q)}{\chi^2(P,Q)} - \frac{f''(1)}{8}\right| < \epsilon \text{ ,}when\, |P-Q| < \delta,$$

where $\epsilon, \delta \to 0$*, i.e.,* ϵ, δ *are very small.*

Proof. We know by Taylors series expansion of function $f(t)$ at $t=1$, that

$$f(t) = f(1) + (t-1) f'(1) + \frac{(t-1)^2}{2!} f''(1) + (t-1)^2 g(t), \tag{65}$$

where $g(t) = \frac{(t-1)}{3!} f'''(1) + \frac{(t-1)^2}{4!} f''''(1) + ...$ and we can see that $g(t) \to 0$ as $t \to 1$, $f(1) = 0$ because $f(t)$ is normalized, therefore from (65) we get

$$f(t) \approx (t-1) f'(1) + \frac{(t-1)^2}{2!} f''(1). \tag{66}$$

Now Put $t = \frac{p_i+q_i}{2q_i}$ in (66), multiply with q_i and then sum over all $i = 1, 2, 3..., n$, we get the desire result (64). □

7. Conclusion and discussion

In this work, we presented new information inequalities on absolute functions for $S_f(P,Q)$. Further, bounds of various well known divergences have been obtained in terms of the Variational distance in an interval (α, β), $0 < \alpha \leq 1 \leq \beta < \infty$ with $\alpha \neq \beta$ as an application of new inequalities. These bounds have been verified numerically by taking two discrete distributions: Binomial and Poisson. An approximation on $S_f(P,Q)$ has been done, which relates $S_f(P,Q)$ to $\chi^2(P,Q)$ approximately. Lastly, a very important application to the Mutual information has been discussed, which tells us how far the joint distribution is from its independency.

We found in our previous article [13] that square root of some particular divergences of Csiszar's class 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, 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

1. M.B. Bassat, *f- Entropies, probability of error and feature selection*, Inform. Control **39** (1978), 227-242.
2. 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.

3. 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.
4. H.C. Chen, *Statistical pattern recognition*, Hoyderc Book Co., Rocelle Park, New York, (1973).
5. C.K. Chow and C.N. Lin, *Approximating discrete probability distributions with dependence trees*, IEEE Trans. Inform. Theory, **14** (1968), no 3, 462-467.
6. I. Csiszar, *Information measures: A Critical survey*, in Trans. In: Seventh Prague Conf. on Information Theory, Academia, Prague, (1974), 73-86.
7. I. Csiszar, *Information type measures of differences of probability distribution and indirect observations*, Studia Math. Hungarica, **2** (1967), 299- 318.
8. D. Dacunha- Castelle, *Ecole dEte de Probabilites de*, Saint-Flour VII-1977, Berlin, Heidelberg, New York: Springer, (1978).
9. 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.
10. D.V. Gokhale and S. Kullback, *Information in contingency Tables*, New York, Marcel Dekker, (1978).
11. K.C. Jain and P. Chhabra, *New information inequalities in terms of Chi- square divergence and its application*, International Bulletin of Mathematical Research **1**, no. 1, pp: 37- 48.
12. K.C. Jain and P. Chhabra, *New information inequalities in terms of Variational distance and its application*, Journal of New Results in Science **11** (2016), 30- 40.
13. K.C. Jain and P. Chhabra, *Series of new information divergences, properties and corresponding series of metric spaces*, International Journal of Innovative Research in Science, Engineering and Technology **3** (2014), 12124- 12132.
14. K.C. Jain and R.N. Saraswat, *Some new information inequalities and its applications in information theory*, International Journal of Mathematics Research, **4**, no.3 (2012), 295-307.
15. 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.
16. 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.
17. T. Kailath, *The divergence and Bhattacharyya distance measures in signal selection*, IEEE Trans. Comm. Technology **COM-15** (1967), 52- 60.
18. 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.
19. A.N. Kolmogorov, *On the approximation of distributions of sums of independent summands by infinitely divisible distributions*, Sankhya **25**, 159-174.
20. S. Kullback and R.A. Leibler, *On information and sufficiency*, Ann. Math. Statist. **22** (1951), 79-86.
21. F. Nielsen and S. Boltz, *The Burbea-Rao and Bhattacharyya centroids*, Apr. (2010), Arxiv.
22. 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.
23. E.C. Pielou, *Ecological diversity*, New York, Wiley, (1975).
24. A. Renyi, *On measures of entropy and information*, Proc. 4th Berkeley Symposium on Math. Statist. and Prob. **1** (1961), 547-561.
25. 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.

26. C.E. Shannon, *A mathematical theory of communication*, Bull. Sept. Tech. J. **27** (1948), 370-423 and 623-656.
27. R. Sibson, *Information radius*, Z. Wahrs. Undverw. Geb. **1** (1969), 149-160.
28. 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.
29. 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.
30. H. Theil *Statistical decomposition analysis*, Amsterdam, North-Holland, 1972.
31. H. Theil, *Economics and information theory*, Amsterdam, North-Holland, 1967.
32. B. Vemuri, M. Liu, S. Amari, and F. Nielsen, *Total Bregman divergence and its applications to DTI analysis*, IEEE Transactions on Medical Imaging, 2010.

Praphull Chhabra received M.Sc. from University of Rajasthan, Jaipur (Rajasthan), India and Ph.D at MNIT, Jaipur (Rajasthan), India. Currently working in University of Engineering and Management, Jaipur (Rajasthan), India. His research interests include information measures and fuzzy divergence measures.

University of Engineering and Management, Jaipur- 303807 (Rajasthan), India.
e-mail: prfl24@gmail.com