Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/100422

New Series of Information Divergence Measures and their Properties

K. C. Jain and Praphull Chhabra*

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

Received: 4 Apr. 2016, Revised: 21 May 2016, Accepted: 22 May 2016
Published online: 1 Jul. 2016

Abstract: In this work, we introduce new series of divergence measures as a family of Csiszar's functional divergence, characterize the properties of convex functions and divergences, compare several divergences, and derive various important and interesting relations among divergences of these new series and other well known divergence measures. Also get the bounds of a particular member of that series together with numerical verification. Application to the mutual information is presented as well.

Keywords: New series of information divergences, various relations among divergences, comparison of divergences, bounds, mutual information, numerical verification

1 Introduction

Divergence measures are basically measures of distance between two probability distributions or compare two probability distributions. It means that any divergence measure must take its minimum value zero when probability distributions are equal and maximum value when probability distributions are perpendicular to each other. Depending on the nature of the problem, different divergence measures are suitable. So it is always desirable to develop a new divergence measure.
In recently years, lot of work had been done on information divergence measures by Dragomir [9,10,11,12], Jain [15,16,19,20,21,23], Taneja [38,39,42,43,44] and others, who gave the idea of divergence measures, their properties, their bounds and relations with other measures.
Divergence measures have been demonstrated very useful in a variety of disciplines such as economics and political science [46,47], biology [33], analysis of contingency tables [13], approximation of probability distributions [5,29], signal processing [26,28], pattern recognition [1,4,25], color image segmentation [31], 3D image segmentation and word alignment [45], cost- sensitive classification for medical diagnosis [35], magnetic resonance image analysis [49] etc.
Also we can use divergences in fuzzy mathematics as fuzzy directed divergences and fuzzy entropies 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 $\Gamma_n = \{P = (p_1, p_2, p_3, ..., p_n) : p_i > 0, \sum_{i=1}^{n} p_i = 1\}$, $n \geq 2$ 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. 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 [6,7], Bregman's f- divergence [2], Burbea- Rao's f- divergence [3], Renyi's like f- divergence [34], and Jain-Saraswat f- divergence [22].
Many divergence measures can be obtained from these generalized f- measures by suitably defining the function f. Especially Csiszar's f- divergence is widely used due to its compact nature, which is given by

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

* Corresponding author e-mail: jainkc_2003@yahoo.com

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. Some resultant divergences by $C_f(P,Q)$, are as follows.

$$E_m^*(P,Q)=\sum_{i=1}^{n}\frac{(p_i-q_i)^{2m}}{(p_iq_i)^{\frac{2m-1}{2}}}, m=1,2,3,...[23]. \quad (2)$$

$$J_m^*(P,Q)=\sum_{i=1}^{n}\frac{(p_i-q_i)^{2m}}{(p_iq_i)^{\frac{2m-1}{2}}}\exp\frac{(p_i-q_i)^2}{p_iq_i}, m=1,2,3,...[23]. \quad (3)$$

$$N_m^*(P,Q)=\sum_{i=1}^{n}\frac{(p_i-q_i)^{2m}}{(p_i+q_i)^{2m-1}}\exp\frac{(p_i-q_i)^2}{(p_i+q_i)^2}, m=1,2,3,...[21]. \quad (4)$$

$$P^*(P,Q)=\sum_{i=1}^{n}\frac{(p_i-q_i)^4(p_i+q_i)\left(p_i^2+q_i^2\right)}{p_i^3q_i^3}[20]. \quad (5)$$

$$\Delta_m(P,Q)=\sum_{i=1}^{n}\frac{(p_i-q_i)^{2m}}{(p_i+q_i)^{2m-1}}, m=1,2,3... \quad (6)$$
$$=\text{Puri and Vineze Divergences } [27].$$

$$\chi^{2m}(P,Q)=\sum_{i=1}^{n}\frac{(p_i-q_i)^{2m}}{q_i^{2m-1}}, m=1,2,3... \quad (7)$$
$$=\text{Chi- } m \text{ divergences } [48],$$

where

$$E_1^*(P,Q)=E^*(P,Q)=\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{\sqrt{p_iq_i}}, \quad (8)$$

$$\Delta_1(P,Q)=\Delta(P,Q)=\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{p_i+q_i} \quad (9)$$
$$=\text{Triangular discrimination } [8],$$

and

$$\chi^2(P,Q)=\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{q_i}=\text{Chi- square divergence } [32]. \quad (10)$$

(8), (9), and (10) are the particular cases of (2), (6), and (7) respectively at $m=1$.

$$K(P,Q)=\sum_{i=1}^{n}p_i\log\frac{p_i}{q_i}=\text{Relative information } [30]. \quad (11)$$

$$G(P,Q)=\sum_{i=1}^{n}\frac{p_i+q_i}{2}\log\left(\frac{p_i+q_i}{2p_i}\right)$$
$$=\text{Relative Arithmetic- Geometric Divergence } [42]. \quad (12)$$

$$F(P,Q)=\sum_{i=1}^{n}p_i\log\frac{2p_i}{p_i+q_i}$$
$$=\text{Relative Jensen- Shannon divergence } [37]. \quad (13)$$

Some means can be seen in literature [41], these are as follows [(14)- (20)].

$$H^*(P,Q)=\sum_{i=1}^{n}\frac{2p_iq_i}{p_i+q_i}=\text{Harmonic mean.} \quad (14)$$

$$A(P,Q)=\sum_{i=1}^{n}\frac{p_i+q_i}{2}=\text{Arithmetic mean.} \quad (15)$$

$$N_1(P,Q)=\sum_{i=1}^{n}\left(\frac{\sqrt{p_i}+\sqrt{q_i}}{2}\right)^2=\text{Square root mean.} \quad (16)$$

$$N_3(P,Q)=\sum_{i=1}^{n}\frac{p_i+\sqrt{p_iq_i}+q_i}{3}=\text{Heronian mean.} \quad (17)$$

$$L^*(P,Q)=\sum_{i=1}^{n}\frac{p_i-q_i}{\log p_i-\log q_i}, p_i\neq q_i\ \forall\, i=\text{Logarithmic mean.} \quad (18)$$

$$G^*(P,Q)=\sum_{i=1}^{n}\sqrt{p_iq_i}=\text{Geometric mean.} \quad (19)$$

$$N_2(P,Q)=\sum_{i=1}^{n}\left(\frac{\sqrt{p_i}+\sqrt{q_i}}{2}\right)\sqrt{\frac{p_i+q_i}{2}}=N_2\text{ mean.} \quad (20)$$

$$J_R(P,Q)=2[F(Q,P)+G(Q,P)]=\sum_{i=1}^{n}(p_i-q_i)\log\left(\frac{p_i+q_i}{2q_i}\right)$$
$$=\text{Relative J- Divergence } [11], \quad (21)$$

where $F(P,Q)$ and $G(P,Q)$ are given by (13) and (12) respectively.

$$h(P,Q)=1-G^*(P,Q)=\sum_{i=1}^{n}\frac{\left(\sqrt{p_i}-\sqrt{q_i}\right)^2}{2} \quad (22)$$
$$=\text{Hellinger discrimination } [14],$$

where $G^*(P,Q)$ is given by (19).

$$I(P,Q)=\frac{1}{2}[F(P,Q)+F(Q,P)]$$
$$=\frac{1}{2}\left[\sum_{i=1}^{n}p_i\log\frac{2p_i}{p_i+q_i}+\sum_{i=1}^{n}q_i\log\frac{2q_i}{p_i+q_i}\right] \quad (23)$$
$$=\text{JS divergence } [3,37],$$

where $F(P,Q)$ is given by (13).

$$J(P,Q)=K(P,Q)+K(Q,P)=J_R(P,Q)+J_R(Q,P)$$
$$=\sum_{i=1}^{n}(p_i-q_i)\log\frac{p_i}{q_i}=\text{J- divergence } [24,30], \quad (24)$$

where $J_R(P,Q)$ and $K(P,Q)$ are given by (21) and (11) respectively.

$$T(P,Q) = \frac{1}{2}[G(P,Q)+G(Q,P)] = \sum_{i=1}^{n}\frac{p_i+q_i}{2}\log\frac{p_i+q_i}{2\sqrt{p_iq_i}} = \text{AG Mean Divergence[42]}, \quad (25)$$

where $G(P,Q)$ is given by (12).

$$\psi(P,Q) = \chi^2(P,Q)+\chi^2(Q,P) = \sum_{i=1}^{n}\frac{(p_i-q_i)^2(p_i+q_i)}{p_iq_i} = \text{Symmetric Chi- square Divergence [12]}, \quad (26)$$

where $\chi^2(P,Q)$ is given by (10).
Divergences (2) to (4), (6), and (7) are series of divergence measures corresponding to series of convex functions. Out of them, divergences (2) to (4) are introduced by Jain and others. Divergences (2) to (6), Means (14) to (20), and (22) to (26) are symmetric while (7), (11) to (13), and (21) are non- symmetric with respect to probability distributions $P,Q \in \Gamma_n$.
Now, for a differentiable function $f:(0,\infty)\to R$, consider the associated function $g:(0,\infty)\to R$, is given by

$$g(t) = (t-1)f'\left(\frac{t+1}{2}\right). \quad (27)$$

After putting (27) in (1), we get

$$E^*_{C_f}(P,Q) = \sum_{i=1}^{n}(p_i-q_i)f'\left(\frac{p_i+q_i}{2q_i}\right). \quad (28)$$

2 New series of convex functions and properties

In this section, we develop some new series of convex functions and study their properties. For this, firstly let $f:(0,\infty)\to R$(set of real no.) be a mapping defined as

$$f_m(t) = \frac{(t^2-1)^{2m}}{t^{2m-1}}, m=1,2,3... \quad (29)$$

and

$$f'_m(t) = \frac{(t^2-1)^{2m-1}\left[t^2(2m+1)+2m-1\right]}{t^{2m}}, \quad (30)$$

$$f''_m(t) = \frac{2m(t^2-1)^{2m-2}}{t^{2m+1}}\left[t^4(2m+1)+4t^2(m-1)+2m-1\right]. \quad (31)$$

From (29), we get the following new convex functions at $m=1,2,3...$ respectively.

$$f_1(t) = \frac{(t^2-1)^2}{t}, f_2(t) = \frac{(t^2-1)^4}{t^3}, f_3(t) = \frac{(t^2-1)^6}{t^5}... \quad (32)$$

Since, we know that the linear combination of convex functions is also a convex function, i.e., $a_1f_1(t)+a_2f_2(t)+a_3f_3(t)+...$ is a convex function as well, where $a_1,a_2,a_3,...$ are positive constants. Therefore, we have following two cases to obtain new series of convex functions.
(i) If we take $a_1=a_2=1, a_3=a_4=a_5=...=0$, then we have

$$f_{1,2}(t) = f_1(t)+f_2(t) = \frac{(t^2-1)^2}{t}+\frac{(t^2-1)^4}{t^3} = \frac{(t^2-1)^2(t^4-t^2+1)}{t^3}. \quad (33)$$

Similarly, if we take $a_2=a_3=1, a_1=a_4=a_5=...=0$, then we have

$$f_{2,3}(t) = f_2(t)+f_3(t) = \frac{(t^2-1)^4}{t^3}+\frac{(t^2-1)^6}{t^5} = \frac{(t^2-1)^4(t^4-t^2+1)}{t^5}. \quad (34)$$

In this way, we can write for $m=1,2,3...$

$$f_{m,m+1}(t) = f_m(t)+f_{m+1}(t) = \frac{(t^2-1)^{2m}}{t^{2m-1}}+\frac{(t^2-1)^{2m+2}}{t^{2m+1}} = \frac{(t^2-1)^{2m}(t^4-t^2+1)}{t^{2m+1}}. \quad (35)$$

(ii) If we take $a_1=1, a_2=\log_e b, a_3=\frac{(\log_e b)^2}{2!}, a_4=\frac{(\log_e b)^3}{3!},..., b>1$, then we have

$$g_1(t) = f_1(t)+(\log_e b)f_2(t)+\frac{(\log_e b)^2}{2!}f_3(t)+... = \frac{(t^2-1)^2}{t}+(\log_e b)\frac{(t^2-1)^4}{t^3}+... = \frac{(t^2-1)^2}{t}\left[1+(\log_e b)\frac{(t^2-1)^2}{t^2}+...\right] = \frac{(t^2-1)^2}{t}b^{\frac{(t^2-1)^2}{t^2}}, b>1. \quad (36)$$

Similarly, if we take $a_1=0, a_2=1, a_3=\log_e b, a_4=\frac{(\log_e b)^2}{2!}, a_5=\frac{(\log_e b)^3}{3!},..., b>1$, then we have

$$g_2(t) = \frac{(t^2-1)^4}{t^3}+(\log_e b)\frac{(t^2-1)^6}{t^5}+..., b>1 = \frac{(t^2-1)^4}{t^3}\left[1+(\log_e b)\frac{(t^2-1)^2}{t^2}+...\right] = \frac{(t^2-1)^4}{t^3}b^{\frac{(t^2-1)^2}{t^2}}, b>1. \quad (37)$$

In this way, we can write

$$g_m(t) = \frac{(t^2-1)^{2m}}{t^{2m-1}} b^{\frac{(t^2-1)^2}{t^2}}, b>1, m=1,2,3,.... \quad (38)$$

Remark: If we take $b = e \approx 2.71828$ then from (38), we obtain the following series.

$$g_m(t) = \frac{(t^2-1)^{2m}}{t^{2m-1}} e^{\frac{(t^2-1)^2}{t^2}} = \frac{(t^2-1)^{2m}}{t^{2m-1}} \exp\frac{(t^2-1)^2}{t^2}, m=1,2,3,.... \quad (39)$$

Properties of functions defined by (29), (35) and (39), are as follows.

- Since $f_m(1) = 0 = f_{m,m+1}(1) = g_m(1) \Rightarrow f_m(t), f_{m,m+1}(t)$ and $g_m(t)$ are normalized functions for each m.
- Since $f_m''(t) \geq 0 \; \forall \; t \in (0,\infty), m = 1,2,3... \Rightarrow f_m(t)$ are convex functions and so $f_{m,m+1}(t), g_m(t)$ are as well.
- Since $f_m'(t) < 0$ at $(0,1)$ and > 0 at $(1,\infty) \Rightarrow f_m(t)$ are monotonically decreasing in $(0,1)$ and monotonically increasing in $(1,\infty)$, for each value of m and $f_m'(1) = 0$.

3 New series of information divergence measures and properties

In this section, we obtain new series of divergence measures corresponding to series of convex functions defined in section 2 and study their properties. For this, firstly the following theorem is well known in literature [7].

Theorem 3.1 If the function f is convex and normalized, i.e., $f''(t) \geq 0 \; \forall \; t > 0$ and $f(1) = 0$ respectively, then $C_f(P,Q)$ and its adjoint $C_f(Q,P)$ are both non-negative and convex in the pair of probability distribution $(P,Q) \in \Gamma_n \times \Gamma_n$.

Now put (29) in (1), we get the following new series of divergences.

$$C_f(P,Q) = \gamma_m(P,Q) = \sum_{i=1}^n \frac{(p_i^2-q_i^2)^{2m}}{p_i^{2m-1} q_i^{2m}}, m=1,2,3... \quad (40)$$

$$\gamma_1(P,Q) = \sum_{i=1}^n \frac{(p_i^2-q_i^2)^2}{p_i q_i^2}, \gamma_2(P,Q) = \sum_{i=1}^n \frac{(p_i^2-q_i^2)^4}{p_i^3 q_i^4}, ... \quad (41)$$

Similarly put (35) in (1), we get the following new series of divergences.

$$C_f(P,Q) = \eta_m(P,Q) = \sum_{i=1}^n \frac{(p_i^2-q_i^2)^{2m}(p_i^4-p_i^2q_i^2+q_i^4)}{p_i^{2m+1} q_i^{2m+2}}, m=1,2... \quad (42)$$

$$\eta_1(P,Q) = \sum_{i=1}^n \frac{(p_i^2-q_i^2)^2 (p_i^4-p_i^2q_i^2+q_i^4)}{p_i^3 q_i^4}, \quad (43)$$

$$\eta_2(P,Q) = \sum_{i=1}^n \frac{(p_i^2-q_i^2)^4 (p_i^4-p_i^2q_i^2+q_i^4)}{p_i^5 q_i^6}, ... \quad (44)$$

Similarly put (39) in (1), we get the following new series of divergences.

$$C_f(P,Q) = \rho_m(P,Q) = \sum_{i=1}^n \frac{(p_i^2-q_i^2)^{2m}}{p_i^{2m-1} q_i^{2m}} \exp\frac{(p_i^2-q_i^2)^2}{(p_i q_i)^2}, m=1,2... \quad (45)$$

$$\rho_1(P,Q) = \sum_{i=1}^n \frac{(p_i^2-q_i^2)^2}{p_i q_i^2} \exp\frac{(p_i^2-q_i^2)^2}{(p_i q_i)^2}, \quad (46)$$

$$\rho_2(P,Q) = \sum_{i=1}^n \frac{(p_i^2-q_i^2)^4}{p_i^3 q_i^4} \exp\frac{(p_i^2-q_i^2)^2}{(p_i q_i)^2}, ... \quad (47)$$

Properties of divergences defined by (40), (42) and (45), are as follows.

- In view of theorem 3.1, we can say that $\gamma_m(P,Q), \eta_m(P,Q), \rho_m(P,Q) > 0$ and are convex in the pair of probability distribution $P,Q \in \Gamma_n$.
- $\gamma_m(P,Q) = 0 = \eta_m(P,Q) = \rho_m(P,Q)$ if $P = Q$ or $p_i = q_i$ (attains its minimum value).
- Since $\gamma_m(P,Q) \neq \gamma_m(Q,P), \eta_m(P,Q) \neq \eta_m(Q,P), \rho_m(P,Q) \neq \rho_m(Q,P) \Rightarrow \gamma_m(P,Q), \eta_m(P,Q), \rho_m(P,Q)$ are non- symmetric divergence measures.

4 Csiszar's information inequality and its application

In this section, we are taking well known information inequalities on $C_f(P,Q)$; such inequalities are for instance needed in order to calculate the relative efficiency of two divergences. By using these inequalities, we will obtain the bounds of $\gamma_1(P,Q)$ in terms of the other well known divergence measures. The following theorem is due to literature [40], which relates two generalized f-divergence measures.

Theorem 4.1 Let $f_1, f_2 : I \subset (0,\infty) \to R$ be two convex and normalized functions, i.e., $f_1''(t), f_2''(t) \geq 0 \; \forall \; t > 0$ and $f_1(1) = f_2(1) = 0$ respectively and suppose the following assumptions.

(i) f_1 and f_2 are twice differentiable on (α,β), $0 < \alpha \leq 1 \leq \beta < \infty$ with $\alpha \neq \beta$.

(ii) There exists the real constants m, M such that $m < M$ and

$$m \leq \frac{f_1''(t)}{f_2''(t)} \leq M, f_2''(t) \neq 0 \; \forall \; t \in (\alpha,\beta). \quad (48)$$

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

$$mC_{f_2}(P,Q) \leq C_{f_1}(P,Q) \leq MC_{f_2}(P,Q), \quad (49)$$

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

Now by using theorem 4.1 or inequalities (49), we will get the bounds of $\gamma_1(P,Q)$ in terms of other well known standard divergences. Firstly, let us consider

$$f_1(t)=\frac{(t^2-1)^2}{t}, t>0,\ f_1(1)=0,\ f_1'(t)=\frac{(t^2-1)(3t^2+1)}{t^2}$$

and

$$f_1''(t)=\frac{2(3t^4+1)}{t^3}. \tag{50}$$

Put $f_1(t)$ in (1), we get

$$C_{f_1}(P,Q)=\sum_{i=1}^{n}\frac{(p_i^2-q_i^2)^2}{p_iq_i^2}=\gamma_1(P,Q). \tag{51}$$

Now, we will obtain bounds of $\gamma_1(P,Q)$ in terms of other well known divergences, by the following propositions.

Proposition 4.1 Let $\gamma_1(P,Q)$ and $h(P,Q)$ be defined as in (51) and (22) respectively. For $P,Q\in\Gamma_n$, we have

(i) If $0<\alpha\le .67$, then

$$23.4h(P,Q)\le\gamma_1(P,Q)\le 8max\left[\frac{3\alpha^4+1}{\alpha^{\frac{3}{2}}},\frac{3\beta^4+1}{\beta^{\frac{3}{2}}}\right]h(P,Q). \tag{52}$$

(ii) If $.67<\alpha\le 1$, then

$$\frac{8(3\alpha^4+1)}{\alpha^{\frac{3}{2}}}h(P,Q)\le\gamma_1(P,Q)\le\frac{8(3\beta^4+1)}{\beta^{\frac{3}{2}}}h(P,Q). \tag{53}$$

Proof: Let us consider $f_2(t)=\frac{1}{2}(1-\sqrt{t})^2, t\in(0,\infty), f_2(1)=0, f_2'(t)=\frac{1}{2}\left(1-\frac{1}{\sqrt{t}}\right)$ and

$$f_2''(t)=\frac{1}{4t^{\frac{3}{2}}}. \tag{54}$$

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

$$C_{f_2}(P,Q)=\sum_{i=1}^{n}\frac{(\sqrt{p_i}-\sqrt{q_i})^2}{2}=h(P,Q). \tag{55}$$

Now, let $g(t)=\frac{f_1''(t)}{f_2''(t)}=\frac{8(3t^4+1)}{t^{\frac{3}{2}}}$ and $g'(t)=\frac{4(15t^4-3)}{t^{\frac{5}{2}}}$, $g''(t)=30\left(3\sqrt{t}+\frac{1}{t^{\frac{7}{2}}}\right)$, where $f_1''(t)$ and $f_2''(t)$ are given by (50) and (54) respectively.

If $g'(t)=0\Rightarrow t=.6687403\approx .67$.

It is clear by Figure 1 of $g'(t)$ that $g'(t)<0$ in $(0,.67)$ and >0 in $(.67,\infty)$, i.e., $g(t)$ is decreasing in $(0,.67)$ and increasing in $(.67,\infty)$. So $g(t)$ has a minimum value at $t=.67$ because $g''(.67)=195.5276\approx 195.5>0$. So

(i) If $0<\alpha\le .67$, then

$$m=\inf_{t\in(\alpha,\beta)}g(t)=g(.67)=23.405968\approx 23.4. \tag{56}$$

$$M=\sup_{t\in(\alpha,\beta)}g(t)=max[g(\alpha),g(\beta)]=max\left[\frac{8(3\alpha^4+1)}{\alpha^{\frac{3}{2}}},\frac{8(3\beta^4+1)}{\beta^{\frac{3}{2}}}\right]. \tag{57}$$

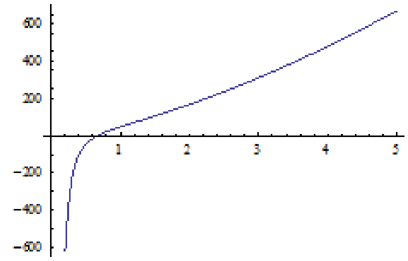

Fig. 1: Graph of $g'(t)$

(ii) If $.67<\alpha\le 1$, then

$$m=\inf_{t\in(\alpha,\beta)}g(t)=g(\alpha)=\frac{8(3\alpha^4+1)}{\alpha^{\frac{3}{2}}}. \tag{58}$$

$$M=\sup_{t\in(\alpha,\beta)}g(t)=g(\beta)=\frac{8(3\beta^4+1)}{\beta^{\frac{3}{2}}}. \tag{59}$$

The results (52) and (53) are obtained by using (51), (55), (56), (57), (58), and (59) in (49).

Proposition 4.2 Let $\gamma_1(P,Q)$ and $G(P,Q)$ be defined as in (51) and (12) respectively. For $P,Q\in\Gamma_n$, we have

(i) If $0<\alpha\le .51$, then

$$14.24G(P,Q)\le\gamma_1(P,Q)$$
$$\le 4max\left[\frac{(\alpha+1)(3\alpha^4+1)}{\alpha},\frac{(\beta+1)(3\beta^4+1)}{\beta}\right]G(P,Q). \tag{60}$$

(ii) If $.51<\alpha\le 1$, then

$$\frac{4(\alpha+1)(3\alpha^4+1)}{\alpha}G(P,Q)\le\gamma_1(P,Q)\le\frac{4(\beta+1)(3\beta^4+1)}{\beta}G(P,Q). \tag{61}$$

Proof: Let us consider

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

$$f_2(1)=0,\ f_2'(t)=\frac{1}{2}\left[\log\frac{t+1}{2t}-\frac{1}{t}\right] \text{ and}$$

$$f_2''(t)=\frac{1}{2t^2(t+1)}. \tag{62}$$

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

$$C_{f_2}(P,Q)=\sum_{i=1}^{n}\left(\frac{p_i+q_i}{2}\right)\log\frac{p_i+q_i}{2p_i}=G(P,Q). \tag{63}$$

Now, let $g(t)=\frac{f_1''(t)}{f_2''(t)}=\frac{4(t+1)(3t^4+1)}{t}$ and $g'(t)=\frac{4(12t^5+9t^4-1)}{t^2}$, $g''(t)=8\left(18t^2+9t+\frac{1}{t^3}\right)$, where $f_1''(t)$ and $f_2''(t)$ are given by (50) and (62) respectively.
If $g'(t)=0\Rightarrow t=.507385\approx .51$.

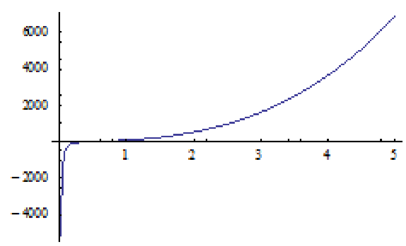

Fig. 2: Graph of $g'(t)$

It is clear by Figure 2 of $g'(t)$ that $g'(t)<0$ in $(0,.51)$ and >0 in $(.51,\infty)$, i.e., $g(t)$ is decreasing in $(0,.51)$ and increasing in $(.51,\infty)$. So $g(t)$ has a minimum value at $t=.51$ because $g''(.51)=134.4830294\approx 134.45>0$. So
(i) If $0<\alpha\le .51$, then

$$m=\inf_{t\in(\alpha,\beta)} g(t)=g(.51)=14.24677337\approx 14.24. \quad (64)$$

$$M=\sup_{t\in(\alpha,\beta)} g(t)=max[g(\alpha),g(\beta)] =max\left[\frac{4(\alpha+1)(3\alpha^4+1)}{\alpha},\frac{4(\beta+1)(3\beta^4+1)}{\beta}\right]. \quad (65)$$

(ii) If $.51<\alpha\le 1$, then

$$m=\inf_{t\in(\alpha,\beta)} g(t)=g(\alpha)=\frac{4(\alpha+1)(3\alpha^4+1)}{\alpha}. \quad (66)$$

$$M=\sup_{t\in(\alpha,\beta)} g(t)=g(\beta)=\frac{4(\beta+1)(3\beta^4+1)}{\beta}. \quad (67)$$

The results (60) and (61) are obtained by using (51), (63), (64), (65), (66), and (67) in (49).
Proposition 4.3 Let $\gamma_1(P,Q)$ and $\chi^2(P,Q)$ be defined as in (51) and (10) respectively. For $P,Q\in\Gamma_n$, we have
(i) If $0<\alpha<1$, then

$$4\chi^2(P,Q)\le\gamma_1(P,Q)\le max\left[\frac{3\alpha^4+1}{\alpha^3},\frac{3\beta^4+1}{\beta^3}\right]\chi^2(P,Q). \quad (68)$$

(ii) If $\alpha=1$, then

$$4\chi^2(P,Q)\le\gamma_1(P,Q)\le\frac{3\beta^4+1}{\beta^3}\chi^2(P,Q). \quad (69)$$

Proof: Let us consider

$f_2(t)=(t-1)^2, t\in(0,\infty), f_2(1)=0, f_2'(t)=2(t-1)$ and

$$f_2''(t)=2. \quad (70)$$

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

$$C_{f_2}(P,Q)=\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{q_i}=\chi^2(P,Q). \quad (71)$$

Now, let $g(t)=\frac{f_1''(t)}{f_2''(t)}=\frac{3t^4+1}{t^3}$ and $g'(t)=\frac{3(t^4-1)}{t^4}$, $g''(t)=\frac{12}{t^5}$, where $f_1''(t)$ and $f_2''(t)$ are given by (50) and (70) respectively.
If $g'(t)=0\Rightarrow t=1$.

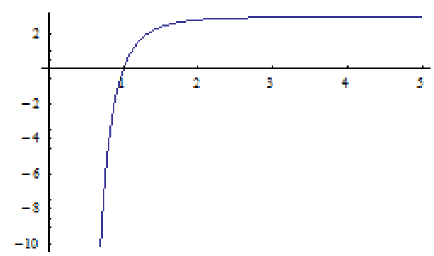

Fig. 3: Graph of $g'(t)$

It is clear by Figure 3 of $g'(t)$ 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 $t=1$ because $g''(1)=12>0$. So

$$m=\inf_{t\in(0,\infty)} g(t)=g(1)=4. \quad (72)$$

(i) If $0<\alpha<1$, then

$$M=\sup_{t\in(\alpha,\beta)} g(t)=max[g(\alpha),g(\beta)] =max\left[\frac{3\alpha^4+1}{\alpha^3},\frac{3\beta^4+1}{\beta^3}\right]. \quad (73)$$

(ii) If $\alpha=1$, then

$$M=\sup_{t\in(1,\beta)} g(t)=g(\beta)=\frac{3\beta^4+1}{\beta^3}. \quad (74)$$

The results (68) and (69) are obtained by using (51), (71), (72), (73), and (74) in (49).

By using the similar approach, we obtain the bounds of $\gamma_1(P,Q)$ in terms of other standard divergences; these inequalities are as follows (we leave to the readers to prove the followings, omitting the details).

Proposition 4.4 If we take $f_2(t) = t\log t$, then we have

$$6.9K(P,Q) \le \gamma_1(P,Q) \le 2max\left[\frac{3\alpha^4+1}{\alpha^2}, \frac{3\beta^4+1}{\beta^2}\right] K(P,Q) \text{ if } 0<\alpha\le .76,$$

$$\frac{2(3\alpha^4+1)}{\alpha^2} K(P,Q) \le \gamma_1(P,Q) \le \frac{2(3\beta^4+1)}{\beta^2} K(P,Q) \text{ if } .76<\alpha\le 1.$$

Proposition 4.5 If we take $f_2(t) = t\log\frac{2t}{t+1}$, then we have

$$19.7F(P,Q) \le \gamma_1(P,Q) \le 2max\left[\frac{(\alpha+1)^2(3\alpha^4+1)}{\alpha^2}, \frac{(\beta+1)^2(3\beta^4+1)}{\beta^2}\right] F(P,Q)$$
if $0<\alpha\le .62$,

$$\frac{2(\alpha+1)^2(3\alpha^4+1)}{\alpha^2} F(P,Q) \le \gamma_1(P,Q) \le \frac{2(\beta+1)^2(3\beta^4+1)}{\beta^2} F(P,Q) \text{ if } .62<\alpha\le 1.$$

Proposition 4.6 If we take $f_2(t) = (t-1)\log t$, then we have

$$2.87J(P,Q) \le \gamma_1(P,Q) \le 2max\left[\frac{3\alpha^4+1}{\alpha(\alpha+1)}, \frac{3\beta^4+1}{\beta(\beta+1)}\right] J(P,Q) \text{ if } 0<\alpha\le .65,$$

$$\frac{2(3\alpha^4+1)}{\alpha(\alpha+1)} J(P,Q) \le \gamma_1(P,Q) \le \frac{2(3\beta^4+1)}{\beta(\beta+1)} J(P,Q) \text{ if } .65<\alpha\le 1.$$

Proposition 4.7 If we take $f_2(t) = \frac{t+1}{2}\log\frac{t+1}{2\sqrt{t}}$, then we have

$$21.8T(P,Q) \le \gamma_1(P,Q) \le 8max\left[\frac{(3\alpha^4+1)(\alpha+1)}{\alpha(\alpha^2+1)}, \frac{(3\beta^4+1)(\beta+1)}{\beta(\beta^2+1)}\right] T(P,Q)$$
if $0<\alpha\le .62$,

$$\frac{8(3\alpha^4+1)(\alpha+1)}{\alpha(\alpha^2+1)} T(P,Q) \le \gamma_1(P,Q) \le \frac{8(3\beta^4+1)(\beta+1)}{\beta(\beta^2+1)} T(P,Q) \text{ if } .62<\alpha\le 1.$$

Proposition 4.8 If we take $f_2(t) = \frac{(t-1)^2(t+1)}{t}$, then we have

$$\psi(P,Q) \le \gamma_1(P,Q) \le max\left[\frac{3\alpha^4+1}{\alpha^3+1}, \frac{3\beta^4+1}{\beta^3+1}\right] \psi(P,Q)$$
if $0<\alpha\le .25$,

$$\frac{3\alpha^4+1}{\alpha^3+1}\psi(P,Q) \le \gamma_1(P,Q) \le \frac{3\beta^4+1}{\beta^3+1}\psi(P,Q)$$
if $.25<\alpha\le 1$.

Proposition 4.9 If we take $f_2(t) = \frac{t}{2}\log t + \frac{t+1}{2}\log\frac{2}{t+1}$, then we have

$$23.86I(P,Q) \le \gamma_1(P,Q) \le 4max\left[\frac{(\alpha+1)(3\alpha^4+1)}{\alpha^2}, \frac{(\beta+1)(3\beta^4+1)}{\beta^2}\right] I(P,Q)$$
if $0<\alpha\le .69$,

$$\frac{4(\alpha+1)(3\alpha^4+1)}{\alpha^2} I(P,Q) \le \gamma_1(P,Q) \le \frac{4(\beta+1)(3\beta^4+1)}{\beta^2} I(P,Q) \text{ if } .69<\alpha\le 1.$$

5 Some new relations among divergences

In this section, we obtain various new important and interesting relations on new divergence measures (40), (42), and (45) with other standard divergence measures.

Proposition 5.1 Let $P,Q \in \Gamma_n$, then we have the following new intra relation.

$$\gamma_m(P,Q) \le \eta_m(P,Q) \le \rho_m(P,Q), \quad (75)$$

where $m = 1,2,3...$ and $\gamma_m(P,Q), \eta_m(P,Q)$, and $\rho_m(P,Q)$ are given by (40), (42), and (45) respectively.

Proof: Since

$$\frac{(t^2-1)^{2m}(t^4-t^2+1)}{t^{2m+1}} = \frac{(t^2-1)^{2m}}{t^{2m-1}} + \frac{(t^2-1)^{2m+2}}{t^{2m+1}}$$

and

$$\frac{(t^2-1)^{2m}}{t^{2m-1}} \exp\frac{(t^2-1)^2}{t^2} = \frac{(t^2-1)^{2m}}{t^{2m-1}}\left[1+\frac{(t^2-1)^2}{t^2}+\frac{(t^2-1)^4}{2!t^4}+...\right].$$

Therefore, for $m = 1,2,3...$ and $t>0$, we have the following inequalities.

$$\frac{(t^2-1)^{2m}}{t^{2m-1}} \le \frac{(t^2-1)^{2m}}{t^{2m-1}} + \frac{(t^2-1)^{2m+2}}{t^{2m+1}} \le \frac{(t^2-1)^{2m}}{t^{2m-1}}\left[1+\frac{(t^2-1)^2}{t^2}+\frac{(t^2-1)^4}{2!t^4}+...\right]. \quad (76)$$

Now put $t = \frac{p_i}{q_i}, i = 1,2,3...,n$ in (76), multiply by q_i and then sum over all $i = 1,2,3...,n$, we get the relation (75). Particularly from (75), we will obtain the followings.

$$\gamma_1(P,Q) \le \eta_1(P,Q) \le \rho_1(P,Q), \quad \gamma_2(P,Q) \le \eta_2(P,Q) \le \rho_2(P,Q), ... \quad (77)$$

Now there are some new algebraic and exponential inequalities, which are important tool to derive some interesting and important new relations in this paper. These inequalities are as follows.

Proposition 5.2 Let $t \in (0,\infty)$ and $m = 1,2,3...$ then we have the following new inequalities.

$$\frac{\left(t^2-1\right)^{2m}}{t^{2m-1}} > \frac{(t-1)^{2m}}{t^{\frac{2m-1}{2}}}, \tag{78}$$

$$\frac{\left(t^2-1\right)^{2m}}{t^{2m-1}} > \frac{(t-1)^{2m}}{(t+1)^{2m-1}}, \tag{79}$$

$$\frac{\left(t^2-1\right)^{2m}}{t^{2m-1}} > (t-1)^{2m}, \tag{80}$$

and

$$\frac{\left(t^2-1\right)^{2m}}{t^{2m-1}} \exp\frac{\left(t^2-1\right)^2}{t^2} > \frac{(t-1)^{2m}}{t^{\frac{2m-1}{2}}} \exp\frac{(t-1)^2}{t}. \tag{81}$$

All functions involve in (78) to (81) are convex and normalized, since $f''(t) \geq 0 \ \forall\ t > 0$ and $f(1) = 0$ respectively.

Proof:From (78), we have to prove that

$$\frac{\left(t^2-1\right)^{2m}}{t^{2m-1}} > \frac{(t-1)^{2m}}{t^{\frac{2m-1}{2}}} \Rightarrow (t+1)^{2m} > t^{m-\frac{1}{2}}$$
$$\Rightarrow \sqrt{t}\,(t+1)^{2m} - t^m > 0,$$

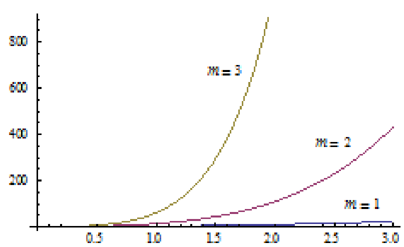

Fig. 4: Graph of $\sqrt{t}\,(t+1)^{2m} - t^m$

which is true (Figure 4) for $t > 0, m = 1,2,3....$ Hence proved the result (78).

Now from (79), we have to prove that

$$\frac{\left(t^2-1\right)^{2m}}{t^{2m-1}} > \frac{(t-1)^{2m}}{(t+1)^{2m-1}} \Rightarrow (t+1)^{4m-1} > t^{2m-1}$$
$$\Rightarrow (t+1)^{4m-1} - t^{2m-1} > 0,$$

which is true (Figure 5) for $t > 0, m = 1,2,3....$ Hence

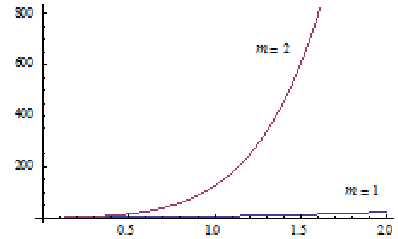

Fig. 5: Graph of $(t+1)^{4m-1} - t^{2m-1}$

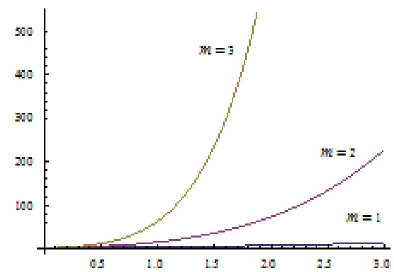

Fig. 6: Graph of $(t+1)^{2m} - t^{2m-1}$

proved the result (79).

Similarly from (80), we have to prove that

$$\frac{\left(t^2-1\right)^{2m}}{t^{2m-1}} > (t-1)^{2m} \Rightarrow (t+1)^{2m} - t^{2m-1} > 0,$$

which is true (Figure 6) for $t > 0, m = 1,2,3....$ Hence proved the result (80).

Similarly from (81), we have to prove that

$$\frac{\left(t^2-1\right)^{2m}}{t^{2m-1}} \exp\frac{\left(t^2-1\right)^2}{t^2} > \frac{(t-1)^{2m}}{t^{\frac{2m-1}{2}}} \exp\frac{(t-1)^2}{t}$$
$$\Rightarrow \frac{(t+1)^{2m} e^{\frac{(t-1)^2\left(t^2+t+1\right)}{t^2}}}{t^{m-\frac{1}{2}}} > 1$$
$$\Rightarrow (t+1)^{2m} e^{\frac{(t-1)^2\left(t^2+t+1\right)}{t^2}} - t^{m-\frac{1}{2}} > 0,$$

which is true (Figure 7) for $t > 0, m = 1,2,3....$ Hence proved the result (81).

Proposition 5.3 Let $P, Q \in \Gamma_n$, then we have the followings new inter relations.

$$\gamma_m(P,Q) > E_m^*(P,Q), \tag{82}$$

$$\gamma_m(P,Q) > \Delta_m(P,Q), \tag{83}$$

$$\gamma_m(P,Q) > \chi^{2m}(P,Q), \tag{84}$$

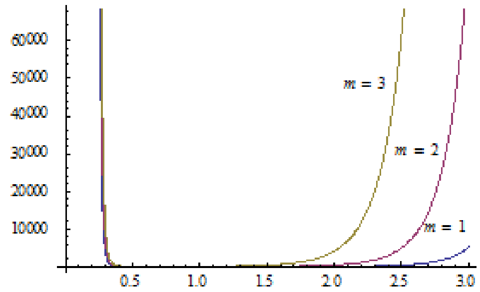

Fig. 7: Graph of $(t+1)^{2m} e^{\frac{(t-1)^2(t^2+t+1)}{t^2}} - t^{m-\frac{1}{2}}$

and

$$\rho_m(P,Q) > J_m^*(P,Q), \tag{85}$$

where $\gamma_m(P,Q), E_m^*(P,Q), \Delta_m(P,Q), \chi^{2m}(P,Q), \rho_m(P,Q)$, and $J_m^*(P,Q)$ are given by (40), (2), (6), (7), (45), and (3) respectively.

Proof: If we put $t = \frac{p_i}{q_i}, i = 1,2,3...,n$ in (78) to (81), multiply by q_i and then sum over all $i = 1,2,3...,n$, we get the desired relations (82) to (85) respectively.

Now we can easily say from (82) to (85) that

$$\gamma_1(P,Q) > E_1^*(P,Q) = E^*(P,Q), \gamma_2(P,Q) > E_2^*(P,Q), ..., \tag{86}$$

$$\gamma_1(P,Q) > \Delta_1(P,Q) = \Delta(P,Q), \gamma_2(P,Q) > \Delta_2(P,Q), ..., \tag{87}$$

$$\gamma_1(P,Q) > \chi^2(P,Q), \gamma_2(P,Q) > \chi^4(P,Q), ..., \tag{88}$$

and

$$\rho_1(P,Q) > J_1^*(P,Q), \rho_2(P,Q) > J_2^*(P,Q), ..., \tag{89}$$

respectively.

Proposition 5.4 Let $P,Q \in \Gamma_n$, then we have the followings new inter relations.

$$\rho_m(P,Q) > J_m^*(P,Q) \geq E_m^*(P,Q), \tag{90}$$

$$\rho_1(P,Q) > 2\Delta(P,Q) \geq 2[N_1^*(P,Q) - N_2^*(P,Q)], \tag{91}$$

$$\rho_1(P,Q) > 8T(P,Q) \geq J(P,Q) \geq 8h(P,Q) \geq 8I(P,Q), \tag{92}$$

and

$$\rho_1(P,Q) > 8A(P,Q) \geq 8N_2(P,Q) \geq 8N_3(P,Q) \geq 8N_1(P,Q) \geq 8L^*(P,Q) \geq 8G^*(P,Q) \geq 8H^*(P,Q), \tag{93}$$

where $\rho_m(P,Q), J_m^*(P,Q), E_m^*(P,Q), N_m^*(P,Q), \Delta(P,Q), T(P,Q), J(P,Q)$, $h(P,Q)$, $I(P,Q)$ and means $H^*(P,Q), A(P,Q), N_1(P,Q), N_3(P,Q), L^*(P,Q), G^*(P,Q), N_2(P,Q)$ are given by (45), (3), (2), (4), (9), (25), (24), (22), (23), (14), (15), (16), (17), (18), (19), and (20) respectively.

Proof: Since we know the followings.

$$J_m^*(P,Q) \geq E_m^*(P,Q) \text{ [17]}, \tag{94}$$

$$\frac{1}{2}E^*(P,Q) \geq \Delta(P,Q) \geq [N_1^*(P,Q) - N_2^*(P,Q)] \text{ [17]}, \tag{95}$$

$$\frac{1}{2}E^*(P,Q) \geq T(P,Q) \geq \frac{1}{8}J(P,Q) \geq h(P,Q) \geq I(P,Q) \text{ [23]}, \tag{96}$$

$$T(P,Q) \geq A(P,Q) \text{ [17]}, \tag{97}$$

and

$$A(P,Q) \geq N_2(P,Q) \geq N_3(P,Q) \geq N_1(P,Q) \geq L^*(P,Q) \geq G^*(P,Q) \geq H^*(P,Q) \text{ [41]}. \tag{98}$$

By taking (85) and (94) together, we get the relation (90).

By taking first and third part of the proved relation (90) at $m = 1$ together with (95), we get the relation (91).

By taking first and third part of the proved relation (90) at $m = 1$ together with (96), we get the relation (92).

By taking first and second part of the proved relation (92) together with (97) and (98), we get the relation (93).

6 Application to the Mutual information

Mutual information [36] 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 $p(x_i) = \sum_{j=1}^{n}, i = 1,2,...,m$ 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

$$\begin{aligned} I(X,Y) &= \sum_{i=1}^{m}\sum_{j=1}^{n} p(x_i,y_j)\log\frac{p(x_i,y_j)}{p(x_i)\,p(y_j)} \\ &= \sum_{(x,y)\in(X,Y)} p(x,y)\log\frac{p(x,y)}{p(x)\,p(y)}. \end{aligned} \tag{99}$$

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

$$\begin{aligned} I(X,Y) = I(Y,X) &= H(X) - H\left(\frac{X}{Y}\right) \\ &= H(Y) - H\left(\frac{Y}{X}\right), \end{aligned} \tag{100}$$

where

$$H(X) = -\sum_{i=1}^{m} p(x_i)\log p(x_i) = -\sum_{i=1}^{m}\sum_{j=1}^{n} p(x_i,y_j)\log\left(\sum_{j=1}^{n} p(x_i,y_j)\right) \quad (101)$$

is known as Marginal entropy [36] and

$$H\left(\frac{X}{Y}\right) = -\sum_{i=1}^{m}\sum_{j=1}^{n} p(x_i,y_j)\log p\left(\frac{x_i}{y_j}\right) \quad (102)$$

is known as Conditional entropy [36].
By viewing $K(P,Q)$ (Relative entropy (11)), 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 (11). So $I(X,Y)$ can also be written as

$$I(X,Y) = K(p(x,y), p(x)p(y)) = \sum_{(x,y)\in(X,Y)} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}. \quad (103)$$

Similarly, we can define the Mutual information in following manners as well.
In $\gamma_1(P,Q)$ manner:

$$I_{\gamma_1}(X,Y) = \sum_{(x,y)\in(X,Y)} \frac{\left[p^2(x,y) - p^2(x)p^2(y)\right]^2}{p(x,y)p^2(x)p^2(y)}, \quad (104)$$

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

$$I_{\chi^2}(X,Y) = \sum_{(x,y)\in(X,Y)} \frac{[p(x,y) - p(x)p(y)]^2}{p(x)p(y)}, \quad (105)$$

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

$$I_{J_R}(X,Y) = \sum_{(x,y)\in(X,Y)} [p(x,y) - p(x)p(y)]\log\frac{p(x,y)+p(x)p(y)}{2p(x)p(y)}, \quad (106)$$

where $\chi^2(P,Q), J_R(P,Q)$ and $\gamma_1(P,Q)$ are given by (10), (21) and (51) respectively.
So (103), (104), (105), and (106) tell us that how far the joint distribution is from its independency or $I(X,Y) = 0 = I_{\gamma_1}(X,Y) = I_{\chi^2}(X,Y) = I_{J_R}(X,Y)$ if distributions are independent to each other.
Now, the following theorem can be seen in literature [10].
Theorem 6.1 Let $f:(\alpha,\beta)\subset(0,\infty)\to R$ be a mapping which is normalized, i.e., $f(1)=0$ and f' is locally absolutely continuous on (α,β) then there exist the constants $m,M\in R$ with $m<M$, such that

$$m \le f''(t) \le M \;\forall\, t\in(\alpha,\beta).$$

If $P,Q\in\Gamma_n$ such that $0<\alpha\le\frac{p_i}{q_i}\le\beta<\infty\;\forall\, i=1,2,3\ldots,n$ for some α and β with $0<\alpha\le 1\le\beta<\infty, \alpha\ne\beta$, then we have the following inequalities

$$\left|C_f(P,Q) - E^*_{C_f}(P,Q)\right| \le \frac{1}{8}(M-m)\chi^2(P,Q), \quad (107)$$

where $C_f(P,Q), \chi^2(P,Q)$ and $E^*_{C_f}(P,Q)$ are given by (1), (10), and (28) respectively.

Now by using theorem 6.1, we introduce a new information inequalities which relates $I(X,Y)$ and new divergence measure $\gamma_1(P,Q)$.
Proposition 6.1
For $0<\alpha\le\frac{p(x,y)}{p(x)p(y)}\le\beta<\infty\;\forall(x,y)\in(X,Y)$, we get the following new information inequalities in Mutual information sense

$$|I(X,Y) - I_{J_R}(X,Y)| \le \frac{1}{8}\left(\frac{\beta-\alpha}{\alpha\beta}\right) I_{\chi^2}(X,Y) \le \frac{1}{32}\left(\frac{\beta-\alpha}{\alpha\beta}\right) I_{\gamma_1}(X,Y), \quad (108)$$

where $I(X,Y), I_{\gamma_1}(X,Y), I_{\chi^2}(X,Y)$, and $I_{J_R}(X,Y)$ are given by (103), (104), (105) and (106) respectively.
Proof: Let us consider

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

$$f''(t) = \frac{1}{t}. \quad (109)$$

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) and $f'(t)$ in (28) then after replacing $p_i,q_i\;\forall i=1,2,\ldots,n$ by $p(x,y), p(x)p(y)\;\forall(x,y)\in(X,Y)$, we get

$$C_f(P,Q) = \sum_{(x,y)\in(X,Y)} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} = I(X,Y) \quad (110)$$

and

$$E^*_{C_f}(P,Q) = \sum_{(x,y)\in(X,Y)} [p(x,y)-p(x)p(y)]\log\frac{p(x,y)+p(x)p(y)}{2p(x)p(y)} = I_{J_R}(X,Y), \quad (111)$$

respectively.
Now, let $g(t) = f''(t) = \frac{1}{t}$ and $g'(t) = -\frac{1}{t^2}$, where $f''(t)$ is given by (109).
It is clear that $g(t)$ is always decreasing in $(0,\infty)$, so

$$m = \inf_{t\in(\alpha,\beta)} g(t) = g(\beta) = \frac{1}{\beta}. \quad (112)$$

$$M = \sup_{t\in(\alpha,\beta)} g(t) = g(\alpha) = \frac{1}{\alpha}. \quad (113)$$

The result (108) is obtained by using (104), (105), (110), (111), (112), (113) together with first inequality of (68) or (69) in (107), after replacing p_i, q_i by $p(x,y), p(x)p(y)$ respectively.

7 Numerical verification of the obtained bounds

In this section, we give two examples for calculating the divergences $h(P,Q), G(P,Q)$ and $\gamma_1(P,Q)$ and verify the inequalities (52) and (60) or verify the bounds of $\gamma_1(P,Q)$ numerically.

Example 7.1 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

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

x_i	0	1	2	3	4
$p_i \approx$	.000976	.00976	.043	.117	.205
$q_i \approx$	.00673	.033	.084	.140	.175
$\frac{p_i}{q_i} \approx$	.1450	.2957	.5119	.8357	1.171

5	6	7	8	9	10
.246	.205	.117	.043	.00976	.000976
.175	.146	.104	.065	.036	.018
1.405	1.404	1.125	.6615	.2711	.0542

by using Table 1, we get the followings.

$$\alpha(=.0542) \le \frac{p_i}{q_i} \le \beta(=1.405). \quad (114)$$

$$h(P,Q) = \sum_{i=1}^{11} \frac{\left(\sqrt{p_i} - \sqrt{q_i}\right)^2}{2} \approx .02549. \quad (115)$$

$$G(P,Q) = \sum_{i=1}^{11} \frac{p_i + q_i}{2} \log\left(\frac{p_i + q_i}{2p_i}\right) \approx .031. \quad (116)$$

$$\gamma_1(P,Q) = \sum_{i=1}^{11} \frac{\left(p_i^2 - q_i^2\right)^2}{p_i q_i^2} \approx .9610. \quad (117)$$

Put the approximated numerical values from (114) to (117) in (52) and (60), we get the followings respectively

$$.5964 \le .9610\,(= \gamma_1(P,Q)) \le 16.161 \text{ and}$$
$$.44144 \le .9610\,(= \gamma_1(P,Q)) \le 2.6936.$$

Hence verify the inequalities (52) and (60) for $p = 0.5$.

Example 7.2 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 by using Table 2, we get the followings.

Table 2: Evaluation of probability distributions $(n=10, p=0.7, q=0.3)$

x_i	0	1	2	3	4
$p_i \approx$	.0000059	.000137	.00144	.009	.036
$q_i \approx$	.000911	.00638	.022	.052	.091
$\frac{p_i}{q_i} \approx$	.00647	.0214	.0654	.173	.395

5	6	7	8	9	10
.102	.20	.266	.233	.121	.0282
.177	.199	.149	.130	.101	.0709
.871	1.005	1.785	1.792	1.198	.397

$$\alpha(=.00647) \le \frac{p_i}{q_i} \le \beta(=1.792). \quad (118)$$

$$h(P,Q) = \sum_{i=1}^{11} \frac{\left(\sqrt{p_i} - \sqrt{q_i}\right)^2}{2} \approx .0502. \quad (119)$$

$$G(P,Q) = \sum_{i=1}^{11} \frac{p_i + q_i}{2} \log\left(\frac{p_i + q_i}{2p_i}\right) \approx .0746. \quad (120)$$

$$\gamma_1(P,Q) = \sum_{i=1}^{11} \frac{\left(p_i^2 - q_i^2\right)^2}{p_i q_i^2} \approx 2.25065. \quad (121)$$

Put the approximated numerical values from (118) to (121) in (52) and (60), we get the followings respectively

$$1.17468 \le 2.25065\,(= \gamma_1(P,Q)) \le 771.68 \text{ and}$$
$$1.062304 \le 2.25065\,(= \gamma_1(P,Q)) \le 46.4161.$$

Hence verify the inequalities (52) and (60) for $p = 0.7$.
Similarly, we can verify the other obtained inequalities numerically for different values of p and q by taking other discrete probability distributions, like: Geometric, Negative Binomial, Uniform etc.

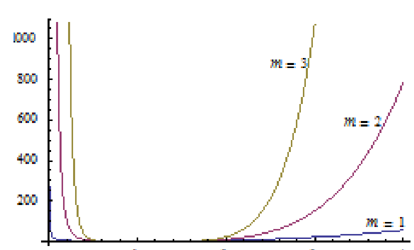

Fig. 8: Convex functions $f_m(t)$

Figure 8, 9, and 10 shows the behavior of convex functions and shows that $f_m(t)$, $f_{m,m+1}(t)$, and $g_m(t)$ has

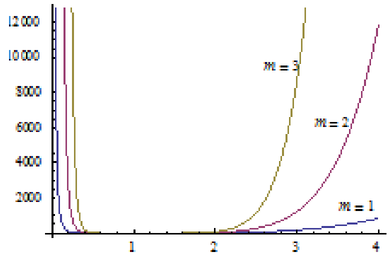

Fig. 9: Convex functions $f_{m,m+1}(t)$

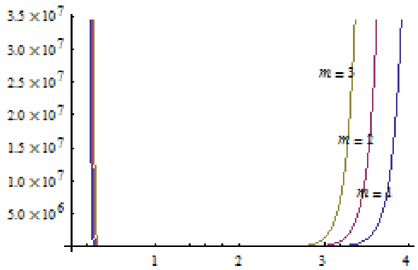

Fig. 10: Convex functions $g_m(t)$

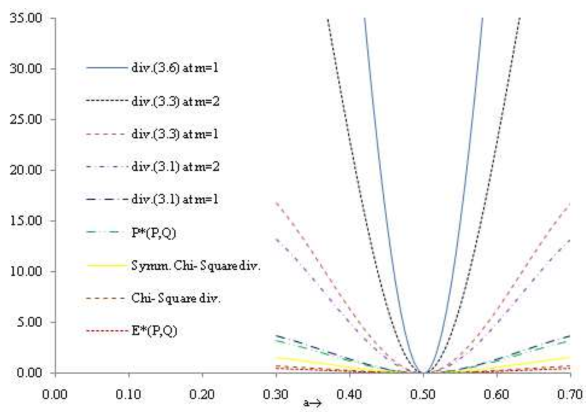

Fig. 11: Comparison of divergence measures

a stepper slope for increasing values of m respectively, while 11 shows the behavior of $\gamma_1(P,Q)$, $\gamma_2(P,Q)$, $\eta_1(P,Q)$, $\eta_2(P,Q)$, $\rho_1(P,Q)$, $P^*(P,Q)$, $\psi(P,Q)$, $\chi^2(P,Q)$, and $E^*(P,Q)$. We have considered $p_i = (a, 1-a), q_i = (1-a, a)$, where $a \in (0,1)$. It is clear from figure 11 that the new divergences $\gamma_1(P,Q), \gamma_2(P,Q), \eta_1(P,Q), \eta_2(P,Q)$, and $\rho_1(P,Q)$ has a steeper slope than $P^*(P,Q), \psi(P,Q), \chi^2(P,Q)$, and $E^*(P,Q)$.

8 Conclusion and discussion

In this paper, we introduced new series of information divergences $\gamma_m(P,Q), \eta_m(P,Q)$, and $\rho_m(P,Q)$ together with characterized their properties. Various important and interesting relations have been obtained among these new divergences and other well known divergences with comparison by using the standard algebraic and exponential inequalities. The upper and lower bounds of a member of new divergence series have been obtained in terms of the other well known divergences in an interval (α,β), $0 < \alpha \leq 1 \leq \beta < \infty$ with $\alpha \neq \beta$ by using Csiszar's inequalities and have been verified numerically by taking two discrete distributions: Binomial and Poisson. Lastly, a very important application to the Mutual information has been discussed, which tells us how far the joint distribution is from its independency and relates new divergence and mutual information.

We found in our previous article [18] that square root of some particular divergences of Csiszars class is a metric space but $C_f(P,Q)$ itself, is not a metric 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 and such investigations are actually in progress because this is also an area worth being investigated.

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. Such types of divergences are also very useful to find utility of an event i.e. an event is how much useful compare to other event.

References

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

[2] 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.

[3] 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.

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

[5] 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.

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

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

[8] Dacunha- Castelle D., Heyer H., and Roynette B., Ecole dEte de probabilities de, Saint-Flour VII-1977, Berlin, Heidelberg, New York: Springer, 1978.

[9] Dragomir S.S., A converse result for Jensen's discrete inequality via Gruss's inequality and applications in information theory, available on line:http://rgmia.vu.edu.au/authors/SSDragomir.htm, 1999.
[10] Dragomir S.S., Gluscevic V., and Pearce C.E.M., New approximation for f- divergence via trapezoid and midpoint inequalities, RGMIA research report collection, 5 (4), 2002.
[11] 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.
[12] Dragomir S.S., Sunde J. and Buse C., New inequalities for Jeffreys divergence measure, Tamusi Oxford Journal of Mathematical Sciences, 16(2) (2000), 295-309.
[13] Gokhale D.V. and Kullback S., Information in contingency Tables, New York, Marcel Dekker, 1978.
[14] Hellinger E., Neue begrundung der theorie der quadratischen formen von unendlichen vielen veranderlichen, J. Rein.Aug. Math., 136 (1909), pp: 210-271.
[15] Jain K.C. and Chhabra P., A new non- symmetric information divergence of Csiszar?s class, properties and its bounds, International Journal of Research in Engineering and Technology, vol. 3, no.5 (2014), pp: 665- 671.
[16] Jain K.C. and Chhabra P., Bounds of non- symmetric divergence measure in terms of other symmetric and non-symmetric divergence measures, International Scholarly Research Notices, Volume 2014, Article ID 820375, 9 pages.
[17] Jain K.C. and Chhabra P., Establishing relations among various measures by using well known inequalities, International Journal of Modern Engineering Research, vol. 4, no.1 (2014), pp: 238- 246.
[18] 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.
[19] Jain K.C. and Chhabra P., Various relations on new information divergence measures, International Journal on Information Theory, vol. 3, no. 4 (2014), pp: 1- 18.
[20] Jain K.C. and Mathur R., A symmetric divergence measure and its bounds, Tamkang Journal of Mathematics, vol. 42, no. 4, 2011, pp: 493-503.
[21] Jain K.C. and Saraswat R. N., Series of information divergence measures using new f- divergences, convex properties and inequalities, International Journal of Modern Engineering Research, vol. 2 (2012), pp: 3226- 3231.
[22] 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.
[23] Jain K.C. and Srivastava A., On symmetric information divergence measures of Csiszar?s f- divergence class, Journal of Applied Mathematics, Statistics and Informatics (JAMSI), 3 (2007), no.1, pp- 85- 102.
[24] Jeffreys, An invariant form for the prior probability in estimation problem, Proc. Roy. Soc. Lon. Ser. A, 186(1946), pp: 453-461.
[25] 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.
[26] 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.
[27] Kafka P., Osterreicher F. and Vincze I., On powers of f? divergence defining a distance, Studia Sci. Math. Hungar., 26 (1991), 415-422.
[28] Kailath T, The divergence and Bhattacharyya distance measures in signal selection, IEEE Trans. Comm. Technology, vol. COM-15, 1967, pp: 52- 60.
[29] 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.
[30] Kullback S. and Leibler R.A., On information and sufficiency, Ann. Math. Statist., 22 (1951), pp: 79-86.
[31] Nielsen F. and Boltz S., The Burbea-Rao and Bhattacharyya centroids, Apr. 2010, Arxiv.
[32] 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.
[33] Pielou E.C, Ecological diversity, New York, Wiley, 1975.
[34] Renyi A., On measures of entropy and information, Proc. 4th Berkeley Symposium on Math. Statist. and Prob., 1 (1961), pp: 547-561.
[35] 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.
[36] Shannon C.E., A mathematical theory of communication, Bull. Sept. Tech. J., 27 (1948), 370-423 and 623-656.
[37] Sibson R., Information radius, Z. Wahrs. Undverw. Geb., (14) (1969), pp: 149-160.
[38] Taneja I.J., A sequence of inequalities among difference of symmetric divergence measures, 2011, Available online: http://arxiv.org/abs/1104.5700v1.
[39] Taneja I.J., Bounds on non symmetric divergence measures in terms of symmetric divergence measures, Journal of Combinatorics, Information, and System Sciences, 29, 14 (2005), pp: 115-134.
[40] Taneja I.J., Generalized symmetric divergence measures and inequalities, RGMIA Research Report Collection, http://rgmia.vu.edu.au, 7(4) (2004), Art. 9. Available on-line at: arXiv:math.ST/0501301 v1 19 Jan 2005.
[41] Taneja I.J., Inequalities among logarithmic mean measures, 2011, Available online: http://arxiv.org/abs/1103.2580v1.
[42] 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.
[43] Taneja I.J., On symmetric and non symmetric divergence measures and their generalizations, Chapter in: Advances in Imaging and Electron Physics, 138(2005), pp: 177-250.
[44] Taneja I.J., Seven means, generalized triangular discrimination and generating divergence measures, Information 4 (2013), pp: 198-239.

[45] 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.
[46] Theil H. Statistical decomposition analysis, Amsterdam, North-Holland, 1972.
[47] Theil H. Economics and information theory, Amsterdam, North-Holland, 1967.
[48] Vajda I., On the f- divergence and singularity of probability measures, Periodica Math. Hunger, 2 (1972), pp: 223- 234.
[49] Vemuri B., Liu M., Amari S., and Nielsen F., Total Bregman divergence and its applications to DTI analysis, IEEE Transactions on Medical Imaging, 2010.

K. C. Jain He received the M.Sc and Ph.D degrees from the University of Rajasthan, India in 1973 and 1979 respectively. He joined the Malaviya Regional Engineering College, Jaipur (Rajasthan), India (Now Malaviya National Institute of Technology, Jaipur) as a lecturer in 1977. Then senior lecturer in 1986, and then Reader in Mathematics from 1992- 1994 and 1997- 2006, in the same institute. He was appointed professor in 2006, Department of Mathematics, MNIT, Jaipur. He was the head of the department in 1994- 1997 and 2010- 2013, as well. Presently, he is a professor in Mathematics department, MNIT, Jaipur. He has supervised 5 Ph.D students and currently supervising 2 Ph.D students. He is the author and Co- author of more than 60 research papers and more than 10 books. He has presented his research work in more than 25 national and international conferences. He has attended more than 30 workshops, short term courses and seminars. He is a life member of Indian Society for Technical Education, Rajasthan Ganita Parishad, International Journal of Mathematical Sciences and Engineering Applications, and International Bulletin of Mathematical Research. He is also a referee of the Sarajevo Journal of Mathematics, Ganita sandesh, and Journal of the Rajasthan Academy of physical Sciences. He also developed computer programs in FORTRAN, C and FOXPRO and developed computer software as well, named: STAFF EXPENDITURE. He delivered invited lectures at more than 30 places on the topics: Stochastic process and applications, Information Theory and Coding, Applications of Statistical tools, Testing of Hypothesis, Error-Correcting Codes, Better understanding of Grading System, and Continuous analogue of Shannons Entropy etc. His research area is Information Divergence Measures.

Praphull Chhabra He received his B.Sc, M.Sc and B.Ed degrees from the University of Rajasthan, India in 2008, 2010, and 2013 respectively. He has qualified the National Eligibility Test (NET)- June 2011 and Graduate Aptitude Test for Engineering (GATE)- Feb. 2011, in Mathematical sciences. He worked 2 years in private engineering college as an Assistant professor after post graduation. Currently, he is doing Ph.D in Information Theory under the supervision of Prof. K.C. Jain, from the Department of Mathematics, Malaviya National Institute of Technology, Jaipur (Rajasthan), India. He is the Co- author of 20 research papers with Prof. K.C. Jain, out of them 16 have been published in Reputed International journals. He has presented his research work in more than 10 national and international conferences. He has attended more than 15 workshops, short term courses and seminars. He is an editorial member of the journal International Bulletin of Mathematical Research. He has technical knowledge of Latex, Matlab and C Language.