Int. J. of Mathematical Sciences and Applications,
Vol. 1, No. 3, September 2011

www.journalshub.com

SOME NEW DIVERGENCE MEASURES & THEIR PROPERTIES

[1]P.K.Bhatia [2]Surender Singh* and [3]Vinod Kumar

Department of Mathematics DCRUST , Murthal (Sonepat) ,India , bhatiapk@gmail.com

Department of Mathematics SDITM , Israna (panipat),India, surender1976@gmail.com

Department of Mathematics RPIIT, Bastara (Karnal) ,India , vkmajra@gmail.com

Abstract

Apart from classical means, contra-harmonic mean is also known. In this paper, we have constructed divergence measures based on nonnegative differences among these means, and established an interesting inequality by use of properties of Csisz´ar's f-divergence. Metric properties of these divergence measures also presented. Comparison of new mean divergence measures with classical divergence measures such as J-divergence, Kullback Leibler is also established.Three new measures of image registration have been defined.

Keywords: **Contra-harmonic means, Csisz´ar's f-divergence, classical divergence measures, J-divergence, Kullback Leibler, Image registration.**

AMS subject classification: 94A17

1. Introduction

The measure of information was defined Claude .E.Shannon in his treatise paper [7] in 1948.

$$H(P) = \sum_{i=1}^{n} p_i \log p_i \text{ ,} P\epsilon\Gamma n \tag{1}$$

where $\Gamma n = \{P=(p_1,p_2,p_3,\ldots\ldots p_n) \,/\, p_i \geq 0 \;, \sum_{i=1}^{n} p_i = 1 \quad ; n\geq 2\}$

is the set of all complete finite discrete probability distributions.

The relative entropy is a measure of the distance between two probability distributions. In statistics, it arises as the expected logarithm of the likelihood ratio. The relative entropy K(P||Q) is the measure of inefficiency of assuming that the distribution is q when the true distribution is p . For example, if we knew the true distribution of the random variable, then we could construct a code with average description length H(P). If, instead, we used the code for a distribution q, we would need H(P) + K(P||Q) bits on the average to describe the random variable[3,4].

The *relative entropy* or *Kullback Leibler distance* [3] between two probability distributions is defined as

$$K(P||Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}. \tag{2}$$

Taneja [8] defined some new divergence measures using the classical means. In the present communication some new divergence measures have been defined using contra harmonic mean and classical means. Comparison of these measures have been done with some already defined measures and new measures of image registration have been defined. In this communication inductive approach is used to define divergence measure with improved results.

2. New Generalized Mean of Order p

Let us consider the following well known mean of order p [1]

$$L_p(a,b) = \frac{a^p+b^p}{a^{p-1}+b^{p-1}} \text{ , } p\varepsilon R$$

$$L_p(a,b)=\begin{cases} \frac{a^2+b^2}{a+b}, & p=2 \\ \frac{a+b}{2}, & p=1 \\ \sqrt{ab}, & p=1/2 \\ \frac{2ab}{a+b}, & p=0 \end{cases} \tag{3}$$

Clearly,

$$L_0(a,b)=H(a,b)=\frac{2ab}{a+b} \quad \text{--- Harmonic Mean} \tag{4}$$

$$L_1(a,b)=A(a,b)=\frac{a+b}{2} \quad \text{---Arithmetic Mean} \tag{5}$$

$$L_{1/2}(a,b)=G(a,b)=\sqrt{ab} \quad \text{---Geometric Mean} \tag{6}$$

$$L_2(a,b)=C(a,b)=\frac{a^2+b^2}{a+b} \quad \text{--Contra-Harmonic Mean} \tag{7}$$

We can establish that $L_p(a,b)$ is monotonically non decreasing function in relation to p[1] . This allow us to conclude the following inequality

$$L_0(a,b)\le L_{1/2}(a,b) \le L_1(a,b) \le L_2(a,b) \tag{8}$$

In view of (8) , we write

$$H(a,b) \le G(a,b) \le A(a,b) \le C(a,b) \tag{9}$$

Let us now consider the following non-negative differences arising due to inequality (9)

$$M_{CA}= C(a,b) - A(a,b) = \frac{a^2+b^2}{a+b} - \frac{a+b}{2} \tag{10}$$

$$M_{CG}= C(a,b) - G(a,b) = \frac{a^2+b^2}{a+b} - \sqrt{ab} \tag{11}$$

$$M_{CH}= C(a,b) - H(a,b) = \frac{a^2+b^2}{a+b} - \frac{2ab}{a+b} \tag{12}$$

In view of (8) , we have the following inequality among the mean divergences measures:

$$0 \le M_{CA}(a,b) \le M_{CG}(a,b) \le M_{CH}(a,b) \tag{13}$$

3. Mean Difference Divergence Measures

Let $\Gamma_n= \{P=(p_1,p_2,p_3,\ldots\ldots p_n) \,/\, p_i\ge 0 \;, \sum_{i=1}^{n} p_i = 1 \quad ; n\ge 2\}$
be the set of all complete finite discrete probability distributions.
Let us take a = p_i and b = q_i in the differences given above and sum over all i = 1, 2, .., n, then for all P,Q $\epsilon\Gamma_n$, we have the following *mean divergence measures*:

- **Contra-harmonic – Arithmetic Mean Divergence Measures**

$$M_{CA}\,(P||Q) = \sum_{i=1}^{n} \frac{p_i^2 + q_i^2}{p_i + q_i} - 1$$

- **Contra-harmonic – Geometric Mean Divergence Measures**

$$M_{CG}\,(P||Q) = \sum_{i=1}^{n} \left(\frac{p_i^2 + q_i^2}{p_i + q_i} - \sqrt{p_i q_i} \right)$$

- **Contra-harmonic harmonic Mean Divergence Measures**

$$M_{CH}\,(P||Q) = \sum_{i=1}^{n} \left(\frac{p_i^2 + q_i^2}{p_i + q_i} - \frac{2p_i q_i}{p_i + q_i} \right)$$

4. Symmetric Measures of Information

Here we shall give some symmetric measures of information.

• Hellinger Discrimination [9]

$h(P||Q) = 1 - B(P||Q) = \frac{1}{2}\sum_{i=1}^{n}(\sqrt{p_i} - \sqrt{q_i})^2$

where $B(P||Q) = \sum_{i=1}^{n} \sqrt{p_i q_i}$ **is** well known Bhattacharya distance.

• Triangular Discrimination [9]

$\Delta(P||Q) = 2\,[1 - W(P||Q)] = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{p_i + q_i}$

where $W(P||Q) = \sum_{i=1}^{n} \frac{2p_i q_i}{p_i + q_i}$

is well known harmonic mean divergence .

• **J-divergence** [9]

$J(P||Q) = K(P||Q) + K(Q||P)$

$= \sum_{i=1}^{n} (p_i - q_i) \log \frac{p_i}{q_i}$

we have,

$M_{CA}\,(P||Q) = \frac{1}{2}\Delta(P||Q)$ (14)

Also, $M_{CG}\,(P||Q) = h(P||Q) + \frac{1}{2}\Delta(P||Q)$ (15)

Again,

$M_{CH}\,(P||Q) = \Delta(P||Q)$ (16)

$M_{CA}\,(P||Q), M_{CG}\,(P||Q), M_{CH}\,(P||Q)$ are also symmetric divergence measures and

$M_{CA}\,(P||Q) = \frac{1}{2}\Delta(P||Q)$

$M_{CG}\,(P||Q) = h(P||Q) + \frac{1}{2}\Delta(P||Q)$

In view of (8) and measure defined in section (4), we have

$0 \le M_{CA}\,(P||Q) \le M_{CG}\,(P||Q) \le M_{CH}\,(P||Q)$ and

$0 \leq h(P||Q) \leq \frac{1}{2}\Delta(P||Q)$ (17)

5. Csiszar's f−Divergence and Mean Divergence Measures

Given a convex function f : (0,∞) → R, the f−*divergence* measure introduced by Csisz´ar [10] is given by

$C_f\,(P||Q) = \sum_{i=1}^{n} q_i\, f(\frac{p_i}{q_i})$, for all P, Q Γ_n

The following theorem is well known in the literature [1]:

Property 5.1. Let the function f : [0,∞) → R be differentiable convex and normalized ,i.e., f(1) = 0, then the Csisz´ar f−divergence, $C_f\,(P||Q)$ is nonnegative and convex in the pair of probability distribution (P,Q) ϵ Γn × Γn.

The *mean divergence measures* given in Section 3 can be written as examples of (5.1) and applying property 5.1 we can check the *nonnegativity* and *convexity* of these measures. Here below we shall give these as examples.

Example 5.1. *Let us consider*

$f_{CA}(x) = \sqrt{\frac{x^2+1}{x+1}} - \frac{x+1}{2}$, $x \in (0, \infty)$

in (5.1) , then we have

$C_f(P||Q) = M_{CA}\,(P||Q)$

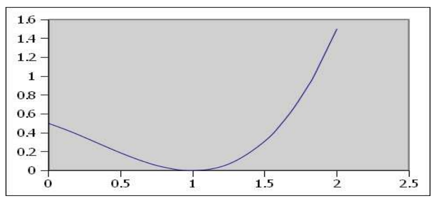
Moreover, $f''_{CA}(x) > 0 \quad , \ \forall x \in (0, \infty)$
Also we have $f_{CA}(1) = 0$.
In view of this we can say that Contra-harmonic – Arithmetic Mean Divergence Measure is non-negative and convex in pairs of probability distributions (P,Q) $\epsilon \Gamma n \times \Gamma n$

The graphical representation of symmetric contra harmonic –Arithmetic divergence function is given by Figure 1 with respect to Table 1

Table 1

x	**0**	**0.125**	**0.25**	**0.375**	**0.5**	**0.625**	**0.75**	**0.875**	**1**	**1.25**	**1.5**	**1.6**	**1.7**	**1.8**	**1.85**	**2**
Y=f_{CA}(x)	**0.5**	**0.430**	**0.351**	**0.268**	**0.187**	**0.114**	**0.054**	**0.014**	**0**	**0.070**	**0.312**	**0.468**	**0.661**	**0.896**	**1.029**	**1.5**

Figure 1



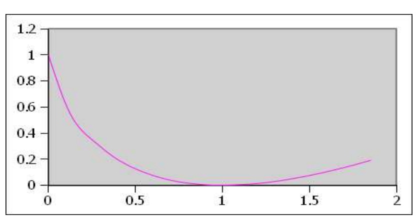
Example 5.2. *Let us consider*
$f_{CG}(x) = \frac{x^2 + 1}{x+1} - \sqrt{x} \quad , x\varepsilon(0, \infty)$
in (5.1) , then we have
$C_f(P||Q) = M_{CG}$ (P||Q)
Moreover, $f''_{CG}(x) > 0 \quad , \ \forall x \in (0, \infty)$
Also we have $f_{CG}(1) = 0$.
In view of this we can say that Contra-harmonic – Geometric Mean Divergence Measure is non-negative and convex in pairs of probability distributions (P,Q) $\in \Gamma n \times \Gamma n$

The graphical representation of symmetric contra harmonic –Geometric divergence function is given by Figure 2 with respect to Table 2.

Table 2

x	0	0.125	0.25	0.375	0.5	0.625	0.75	0.875	1	1.25	1.5	1.6	1.7	1.8	1.85
Y=f_{CG}(x)	1	0.549	0.35	0.217	0.126	0.065	0.026	0.006	0	0.020	0.075	0.104	0.136	0.172	0.191

Figure 2



Example 5.3. *Let us consider*

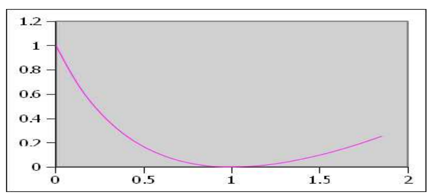
$f_{CH}(x) = \frac{x^2+1}{x+1} - \frac{2x}{x+1}$, $x\varepsilon(0,\infty)$
in (5.1) , then we have
$C_f(P||Q) = M_{CH}(P||Q)$
Moreover, $f''_{CH}(x) > 0$, $\forall x \in (0,\infty)$
Also we have $f_{CH}(1) = 0$.
In view of this we can say that Contra-harmonic – Harmonic Mean Divergence Measure is non-negative and convex in pairs of probability distributions $(P, Q) \in \Gamma n \times \Gamma n$

The graphical representation of symmetric contra harmonic –Arithmetic divergence function is given by Figure 3 with respect to Table 3

Table 3

x	0	0.125	0.25	0.375	0.5	0.625	0.75	0.875	1	1.25	1.5	1.6	1.7	1.8	1.85
$Y=f_{CH}(x)$	1	0.681	0.450	0.284	0.167	0.087	0.036	0.008	0.000	0.028	0.100	0.138	0.181	0.229	0.254

Figure 3



6. Majorization and Schur-Convex Functions
In this section, we recall the basic notions of majorization theory which are relevant to our context [6]
Definition 1: Given two probability distributions $P=(p_1,\ldots,p_n)$ and
$Q=(q_1,\ldots,q_n)$ with $p_1 \geq \ldots\ldots \geq p_n$ and $q_1 \geq \ldots\ldots \geq q_n$, we say that P is majorized by Q in symbols $P \preccurlyeq Q$, if and only if $\sum_{i=1}^{n} p_i \leq \sum_{i=1}^{n} q_i$, k= 1; . . . n
The main link between the concept of majorization and the theory of inequalities is established by the notion of *Schur-convex functions*.
Definition 2: A real-valued function $\emptyset$ defined on the set of n-dimensional probability vectors is said to be Schur-convex (resp., concave) if is order preserving (resp., inverse-order preserving) with respect to the partial order $\preccurlyeq$, that is , if

$$P \preccurlyeq Q \Rightarrow \emptyset(P) \leq \emptyset(Q)$$

A wide class of schur-convex (concave) function is given by the following results.
Lemma 1 [6] If $\emptyset$ is invariant with respect to any permutation of its inputs and convex (concave) then $\emptyset$ is schur – convex (respectively; concave)
By this lemma, we get that entropy function
$H(P) = -\sum_{i=1}^{n} p_i (\log p_i)$ is schur concave that is

$$P \preccurlyeq Q \Rightarrow H(P) \geq H(Q)$$

7. The Majorization Lattice
We first recall [6] that a lattice is a quadruple $(\mathcal{L} / \leq / \vee / \wedge)$ where $\mathcal{L}$ is a set, '$\leq$' is partial ordering on $\mathcal{L}$, and for all $a,b \in \mathcal{L}$ there is a unique greatest lower bound (g.l.b) $a \wedge b$ and a unique least upper bound (l.u.b) $a \vee b$,
More precisely, $a \wedge b \leq b$ and $a \wedge b \leq b$
and $a \leq a \vee b$, $b \leq a \vee b$
Now for $n \geq 2$, let

$\Gamma_n = \{P=(p_1; \ldots; p_n) ; p_i \epsilon [0,1] , \sum_{i=1}^{n} p_i = 1 , p_i \geq p_{i+1} \}$

be the set of all n-dimensional probability distributions , with components in non decreasing order . Then,

(Γ_n /≤/V /Λ) is a partially ordered set such that $\left(\frac{1}{n}, \frac{1}{n}, \frac{1}{n}, \ldots \frac{1}{n}\right) \preccurlyeq \left(\frac{1}{n-1}, \frac{1}{n-1}, \frac{1}{n-1}, \ldots .0\right) \preccurlyeq \cdots \left(\frac{1}{2}, \frac{1}{2}, ..0\right) \preccurlyeq (1,0, \ldots ..0)$

that is $\left(\frac{1}{n}, \frac{1}{n}, \frac{1}{n}, \ldots ..\frac{1}{n}\right) \preccurlyeq P \preccurlyeq (1,0, \ldots ..0)$ for all $P \epsilon \Gamma_n$

we first turn our attention to appropriate definition of greatest lower bound p∧q of any two arbitrary elements of Γ_n.

Definition: For any P,Q ϵ Γ_n let α(P,Q)=(a_1,a_2, ……,a_n)

with a_i= min.$\{\sum_{j=1}^{i} p_j , \sum_{j=1}^{i} q_j \} - \sum_{j=1}^{i-1} a_j$,

Lemma [1]For all P,Q ϵ Γ_n let α(p,q)= $p \wedge q$,we also define β(P,Q) =(b_1,b_2,……b_n) with

b_i= max.$\{\sum_{j=1}^{i} p_j , \sum_{j=1}^{i} q_j \} - \sum_{j=1}^{i-1} b_j$,

Note that components of β(P,Q) might not be in non increasing order . That is, it is not true in general that β(P,Q) ϵ Γ_n

Let β'(P,Q) = $(b_1', b_2', b_3', \ldots \ldots .. b_n')$ be n-tuple of reals obtained by rearranging the components of β(P,Q) in non increasing order .

Therefore, β'(P,Q) ϵ Γ_n

And it is immediate that

$\sum_{j=1}^{k} b_i' \geq$Max.$\{\sum_{j=1}^{k} p_j , \sum_{j=1}^{k} q_j \}$ for k=1,2,………n.

Hence, β'(P,Q) is an upper bound for P, Q in Γ_n .

8. Metric Properties

Definition: For a set X, a function d : X × X→ R
is called a distance if for every x, y ϵ X
1. d(x; y) ≥0 with equality if x = y.
2. d is symmetric: d(x, y) = d(y, x).
The pair (X; d) is then called a distance space. If
in addition to 1 and 2, for every triple x, y , z ϵ X
the function d satisfies
3. d(x, y) + d(x, z) ≥d(y, z) (the triangle inequality),
then d is called a pseudo metric and (X,d) a pseudo metric space
If also, d(x, y) = 0 holds if and only if x = y; then we speak of a metric and a metric space.

Define d: $\Gamma_n \times \Gamma_n \rightarrow R$ by

d(P,Q) =M_{CA}(P||Q) then we have

1. d(P,Q) ≥ 0 for all P,Q ϵ Γ_n
2. If P=Q if and only if p_i=q_i for all i and for all P,Q ϵ Γ_n
   then P=Q implies d(P,Q) = 0
3. Let P,Q ϵ Γ_n and β (P,Q)=(b_1,b_2,……b_n)
   β is as defined in section (7)
   Then we have
   d(P,Q)= $\sum_{i=1}^{n} \frac{p_i^2+q_i^2}{p_i+q_i} - 1$
   d(P, β)= $\sum_{i=1}^{n} \frac{p_i^2+q_i^2}{p_i+q_i} - 1$
   d(β,Q)= $\sum_{i=1}^{n} \frac{p_i^2+q_i^2}{p_i+q_i} - 1$
   Here β (P,Q) is upper bound of p and q as defined in section (7).

   Therefore, P≼ β (P, Q) and Q≼ β (P,Q)

   $\Rightarrow \sum_{i=1}^{n} p_i \leq \sum_{i=1}^{n} b_i$ and $\sum_{i=1}^{n} q_i \leq \sum_{i=1}^{n} b_i$, k= 1,2, . . . n
   We have,

$d(P,\beta) + d(\beta,Q) - d(P,Q) \geq 0$

Therefore, $d(P,Q) \leq d(P,\beta) + d(\beta,Q)$

Hence d is a pseudo metric on Γ_n.

Similarly, if we define

$d_1(P,Q) = M_{CG}(P||Q)$ and $d_2(P,Q) = M_{CH}(P||Q)$

d_1 and d_2 are also found to be pseudo metric on Γ_n.

9. Comparison with Some Classical Divergence measures

The *relative entropy* or *Kullback Leibler distance* [3,4] between two probability distributions is defined as

$$K(P||Q) = \sum_{i=1}^{n} p_i \log\frac{p_i}{q_i}. \tag{18}$$

Also, *J-divergence* [3] is defined as

$$J(P||Q) = K(P||Q) + K(P||Q) = \sum_{i=1}^{n}(p_i - q_i)\log\frac{p_i}{q_i} \tag{19}$$

In view of (14) to (17) & (18)-(19), it can be concluded that

$$0 \leq M_{CA}(P||Q) \leq M_{CG}(P||Q) \leq M_{CH}(P||Q) \leq K(P||Q) \leq J(P||Q) \tag{20}$$

Numerical Verifications:

Consider the probability distributions p=(0.2,0.3,0.1,0.25,0.15) and q=(0.12,0.18,0.3,0.2,0.2). Using this distribution we compare the values of following divergence measures in the table:

Table 4

D(P||Q)	**0.228213**
J P||Q)	**0.501234**
M_{CA} (P||Q)	**0.081349**
M_{CG} (P||Q)	**0.124034**
M_{CH} (P||Q)	**0.162698**

From the table 4 and equation (20) the divergence measures defined in the section (3) shows lesser inefficiency i.e. greater efficiency while choosing distribution q instead of true distribution p.

10. New Measure for Image Registration

Some measures of image registration based on mean divergences have already been defined [5], here we define some new measures . For two images *A* and *B* , let p_{ab} and $p_a\, p_b$ denote $p(a,b)$ and $p(a)\, p(b)$ for simplicity. By substituting p_{ab} , $p_a\, p_b$ for p_i , q_i respectively in the divergence measures defined in section 3, we obtain the following form of CA, CG, and CH measures.

$$M_{CA}(A,B) = \sum_{a,b}\frac{p_{ab}^2 + (p_a p_b)^2}{p_{ab} + p_a p_b} - 1$$

$$M_{CG}(A,B) = \sum_{a,b}\left(\frac{p_{ab}^2 + (p_a p_b)^2}{p_{ab} + p_a p_b} - \sqrt{p_{ab} p_a p_b}\right)$$

$$M_{CH}(A,B) = \sum_{a,b}\left(\frac{p_{ab}^2 + (p_a p_b)^2}{p_{ab} + p_a p_b} - \frac{2 p_{ab} p_a p_b}{p_{ab} + p_a p_b}\right)$$

These measures can determine the distance of $p(a,b)$ and $p(a)\, p(b)$ because of their convexity. The more high value denotes more dependence between the two images. When the two images are aligned, maximal dependence of the two images is assumed to occur, then the value $p(a,b) / p(a)\, p(b)$ is far from 1.

REFERENCES

[1] Bagdasar ,Ovidiu, ‘’*Inequalities and Applications*’’Monograph *http://rgmia.vu.edu.au/monographs.asp*

[2] Cicalese, Ferdinando , Vaccaro ,Ugo , ‘’*Supermodularity and subadditivity properties of the entropy on the majorization lattice*’’, *IEEE Transactions on Information Theory* ,vol 48 ,No. 4,April 2002.

[3] Cover ,T.M and Thomas ,Joy A, *Elements of Information Theory,* Wiley India Pvt. Ltd. New Delhi,2009

[4] Kullback, S. and Leibler ,R.A., *On Information and Sufficiency*, *Ann. Math. Statist.*, 22(1951), 79-86.

[5] Liu ,Changchun, Yang ,Jinbao, Peng Shao, Jason J. Gu , and Mengsun Yu ,*''Mean Divergence Measures for Multimodality Medical Image Registration'' ,IEEE Transactions On Information Theory,2007,(*pp. 1171-1174)

[6] Marshall ,A.W and Olkin, L. ,*Inequalities*, *Theory of Majorization and its Applications*: New York :Academic Press,1979.

[7] Shannon, C.E*''A Mathematical theory of communications*'' Bell System Technical Journal ,Vol 27(1948) 379

[8] Taneja ,I.J.,'' *On mean divergence Measures*" , Available on-line at: arXiv:math/0501298v2 [math.ST] 13,June 2005

[9] 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.

[10] Taneja ,I.J., ''*Bounds on Non-Symmetric Divergence Measures in Terms of Symmetric Divergence Measures''*, Available on-line at: arXiv:math.PR/0506256 v1 13 Jun 2005