BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **13**, No 2

Sofia • 2013 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2013-0013

On a New Csiszar's *f*-Divergence Measure

P. K. Bhatia[1], *Surender Singh*[2]

[1] *DCR University of Science and Technology, Murthal-131039 (HARYANA), INDIA*
[2] *Shri Mata Vaishno Devi University, sub post office, Katra-182320 (J & K), INDIA*
Emails: bhatiapk@rediffmail.com surender1976@gmail.com

Abstract: *A new parametric probabilistic measure of information and a corresponding symmetric divergence (distance) measure is proposed. The information measure is useful to study the uncertainty and the corresponding divergence measure is useful for comparing two probability distributions. The proposed parametric divergence measure belongs to the family of Csiszar's f-divergence measures. The bounds of this divergence measures are obtained in terms of some well known divergence measures. Some properties of the proposed information and divergence measures are studied.*

Keywords: *Csiszar's f-Divergence, distance measure information, information bounds.*

2000 Mathematics Subject Classification: 94A17.

1. Introduction

The information and divergence (or distance) measures are of key importance in a number of theoretical and applied statistical inference and data processing problems. M a j i [13] presented generalized *f*-information measures as evaluation criteria for gene selection problem.

The literature on the development and applications of information and divergence measures has expanded considerably in recent years. T a n e j a [18], B e s s e v i l l e [1], E s t e b a n and M o r a l e s [10] are good references to review the development of generalized information and divergence measures. Depending on the nature of the problem, the different information and divergence measures are suitable. So it is always desirable to develop a new information or divergence measure.

In this paper we present a new parametric information measure and a corresponding parametric divergence measure which belongs to the class of

Csiszar's *f*-divergences. In Section 2 some preliminaries are presented, in Section 3 some approaches to develop new information and divergence measures are presented. A new parametric information measure and its axiomatic characterization is presented in Section 4. The main advantage of this new parametric measure of information is that its maximum value depends on parameter α and the disadvantage is that it is non-additive. Shannon's measure [17] is the limiting case of the proposed measure of information, therefore it has more flexibility of application than Shannon's measure. In order to increase the flexibility of application of this measure of information, two parametric generalizations may be explored. A new parametric divergence measure and its characterization are presented in Section 5. The advantage of the new parametric divergence measure is that it is a distance measure and the value of divergence can be adjusted by adjusting the value of parameter α. Further, it is approximated in terms of Pearson [14] divergence measure. In order to increase the flexibility of application of this divergence measure, two parametric generalizations may be explored. In Section 6 a new symmetric divergence (distance) measure is proposed and in Section 7 some bounds in terms of some well known divergence measures are presented. In Section 8 the approximation in terms of a well known divergence measure is obtained.

2. Preliminaries

2.1. Information measure

The measure of information was defined by Claude E. Shannon in his treatise paper [17] in 1948:

$$H(P)=\sum_{i=1}^{n} p_i \log p_i\,, \qquad P\in\Gamma_n, \tag{2.1}$$

where $\Gamma_n=\{P=(p_1,p_2,...,p_n)/\,p_i\geq 0, \sum_{i=1}^{n} p_i=1\,; n\geq 2\}$, is the set of all complete finite discrete probability distributions. To improve the weakness of Shannon's measure in certain situations Renyi [15] took the first step and proposed a parametric measure of information

$$H_\alpha(P)=\frac{1}{1-\alpha}\log\,(\sum_{i=1}^{n} p_i^{\alpha}),\ \alpha\neq 1\,,\alpha>0. \tag{2.2}$$

After Renyi many generalized information and divergence measures have been developed. Taneja [18], Basseville [1], Esteban and Morales [10] and Wang [21] can be consulted for survey of generalized information and divergence measures.

2.2. Divergence measure

The relative entropy or the directed divergence 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 $D(P, Q)$ is the measure of inefficiency 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 $H(P) + D(P, Q)$ bits on the average to describe the random variable, Cover and Thomas [3]. The *relative entropy* or *Kullback Leibler distance* Kullback and Leibler [12] between two probability distributions is defined as

(2.3) $$D(P,Q)=\sum_{i=1}^{n} p_i \log\frac{p_i}{q_i}.$$

A correct measure of directed divergence must satisfy the following postulates:

(a) $D(P, Q) \geq 0$;

(b) $D(P, Q) = 0$ iff $P = Q$;

(c) $D(P,Q)$ is a convex function of both $P = (p_1, p_2, p_3, \dots, p_n)$ and $Q = (q_1, q_2, q_3, \dots, q_n)$.

If in addition symmetry and triangle inequality is also satisfied by $D(P, Q)$, then it is called a distance measure. Properties (a)-(c) are essential to define a new measure of directed divergence. A parametric measure of directed divergence can also be characterized in terms of its parameter(s).

3. Approaches to develop information and divergence measures

3.1. From an entropy functional

Esteban and Morales [10] proposed a mathematical expression. Most of the information measures cited in literature can be obtained as a particular or limiting case of this entropy functional $H_{h,v}^{\phi_1,\phi_2}(P)$. Let (X, β_X, P), $P\in\Gamma_n$, be a statistical space, where $X=\{x_1, x_2, x_3, \dots, x_n\}$, $\Gamma_n=\{P=(p_1,p_2,\dots,p_n)/\,p_i\geq 0,\ \sum_{i=1}^{n} p_i=1;\ n\geq 2\}$ and β_X is the σ-field of all subsets of X. Then the entropy functional is defined as

(3.1) $$H_{h,v}^{\phi_1,\phi_2}(P)=h\left(\frac{\sum_{i=1}^{n} v_i\,\phi_1(p_i)}{\sum_{i=1}^{n} v_i\,\phi_2(p_i)}\right),$$

where v_i is the weight associated with the $x_i\in X$ and $\phi_1:[0,1)\to R$, $\phi_2:[0,1)\to R$ and $h:[0,1)\to R$ are any suitable functions.

Let $H(P)$ be the entropy function and $P,Q\in\Gamma_n$. Then the divergence measure of Jenson-Shannon type is given by

(3.2) $$D_{\text{J-S}}(P,Q)=H(\lambda_1 P+\lambda_2 Q)-\lambda_1 H(P)-\lambda_2 H(Q),$$

where $\lambda=(\lambda_1,\lambda_2)$, $\lambda_1\geq 0$, $\lambda_2\geq 0$ and $\lambda_1+\lambda_2=1$.

3.2. Csiszar's *f*-divergences

Csiszar's *f*-divergence between two probability distributions introduced by C s i s z a r [5] and are defined as

(3.3)
$$C_f(P,Q)=\sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right),$$

where *f* is a convex function satisfying $f(1)=0,\ f'(1)=0,\ f''(1)=1$.

A list of *f*-divergence measures is provided in S a l i c r u [16] and T a n e j a [19].

3.3. Bregman's divergences

Bregman divergences, introduced by B r e g m a n [2] and are defined for vectors, matrices, functions and probability distributions. The Bregman divergence between vectors is defined as

(3.4)
$$D_\phi(x,y)=\phi(x)-\phi(y)-(x-y)^{\mathrm{T}}\nabla\phi(y).$$

With ϕ a differentiable strictly convex function $\Re^d\to\Re$. The symmetrised Bregman divergence is

(3.5)
$$\overline{D_\phi(x,y)}=(\nabla\phi(x)-\nabla\phi(y))^{\mathrm{T}}(x-y).$$

The Bregman matrix divergence is defined as

(3.6)
$$D_\phi(x,y)=\phi(X)-\phi(Y)-\mathrm{Tr}(\nabla\phi(Y))^{\mathrm{T}}(X-Y).$$

For *X*, *Y* real symmetric *d*×*d* matrices, and ϕ a differentiable strictly convex function $S^d\to\Re$.

3.4. Mean divergences

The divergence measure can be formed by taking the difference of some classical means and Greek means, suggested by T a n e j a [20]. Further T a n e j a [19] proposed their generalized version. These divergence measures also belong to the class of Csiszar's *f*-divergences.

4. A new parametric measure of information

We propose a new parametric information measure using entropy functional approach. In equation (3.1), taking

$$v_i=1,\ \phi_1(x)=-x\sinh(\alpha\log x),\ \phi_2(x)=x,\ h(x)=x\,(\sinh(\alpha))^{-1}$$

we have the following parametric measure of information:

(4.1)
$$H_\alpha(P)=-\frac{1}{\sinh(\alpha)}\big(p_i\sinh(\alpha\log p_i)\big),\ \alpha>0,$$

(4.2)
$$\lim_{\alpha\to 0}H_\alpha(P)=H(P).$$

Thus the Shannon measure (2.1) is the limiting case of the measure proposed in (4.1).

Theorem 4.1. For $P \in \Gamma_n$, the measure of information (4.1) satisfies the following properties:

Symmetry. From (4.1) it is clear that $H_\alpha(P)$ is a permutationally symmetric function of p_i.

Continuity. $H_\alpha(p, 1-p)$ is a continuous function.

Normality. $H_\alpha(1/2, 1/2)=1$.

Non-additive. $H_\alpha(P*Q) \neq H_\alpha(P) + H_\alpha(Q)$, P, $Q \in \Gamma_n$.

Monotonicity w.r.t. p_i. $H_\alpha(p_1, p_2, ... p_n)$ is a monotonic decreasing function of p_i for all i=1, 2, 3, …, n.

Monotonicity w.r.t. α. We have

$$\frac{dH_\alpha(P)}{d\alpha} = -\frac{1}{\sinh(\alpha)}\left(\sum_{i=1}^{n} p_i \log p_i \sinh(\alpha)\cosh(\alpha \log p_i) + \right.$$

$$\left. + p_i \cosh(\alpha)\sinh(\alpha \log p_i)\right), \alpha > 0.$$

That is, $\frac{dH_\alpha(P)}{d\alpha} \geq 0$. Therefore, $H_\alpha(P)$ is a monotonic increasing function of α.

Concavity. The determining function for measure (9) is

$$f(x) = -\frac{x\sinh(\alpha \log x)}{\sinh(\alpha)}, \ \alpha > 0,$$

which gives $f''(x) \leq 0$, for all $\alpha > 0$.

Therefore, $H_\alpha(P)$ is a concave function.

5. A new parametric divergence measure

We propose a new parametric divergence measure using Csiszar's *f*-divergence approach. Consider the function

(5.1) $$f(x) = \frac{x\sinh(\alpha \log x)}{\sinh(\alpha)}, \ \alpha > 0,$$

we have

(5.2) $$f'(x) = \frac{\alpha\cosh(\alpha \log x) + \sinh(\alpha \log x)}{\sinh(\alpha)}, \ \alpha > 0,$$

(5.3) $$f''(x) = \frac{\alpha\cosh(\alpha \log x) + \alpha^2 \sinh(\alpha \log x)}{x\sinh(\alpha)}, \ \alpha > 0.$$

From (5.1), (5.2) and (5.3) we have $f''(x) \geq 0$ for all $x > 0$, therefore $f(x)$ is convex for $x > 0$ and $a > 0$.

Also, $f(1)$=0, $f'(1) = 0$, Here the condition $f''(1)$=1 is to be relaxed.

Thus using (3.3) and (5.1) the new parametric non symmetric divergence measure is given by

(5.4) $$D_\alpha(P,Q)=\sum_{i=1}^{n}\frac{p_i\sinh\left(\alpha\log\frac{p_i}{q_i}\right)}{\sinh(\alpha)},\ \alpha>0.$$

Here we observe that the axioms (a) and (b) in Section 2 are obviously satisfied.

The following theorem is well known in literature [5], related to Csiszar's f-divergences.

Theorem 5.1. Let the function f: $[0, \infty) \rightarrow \Re$ be differentiable convex and normalized, i.e., if(1)=0, then the Csiszar's f-divergence, $C_f(P, Q)$ is nonnegative and convex in the pair of probability distribution $(P, Q) \in \Gamma_n\times\Gamma_n$.

All the requirements in this theorem have already been satisfied, so $C_f(P, Q) = D_\alpha(P, Q)$ is convex in pairs probability distribution $(P, Q) \in \Gamma_n\times\Gamma_n$. This proves axiom (c) of Section 2. Hence, $D_\alpha(P, Q)$ is a correct measure of divergence.

The determining function of the new parametric divergence measure (5.4) is

$$f(x)=\frac{x\sinh(\alpha\log x)}{\sinh(\alpha)},\ \alpha>0.$$

The behaviour of the divergence measure (5.4) with increasing value of α is shown in Fig. 1.

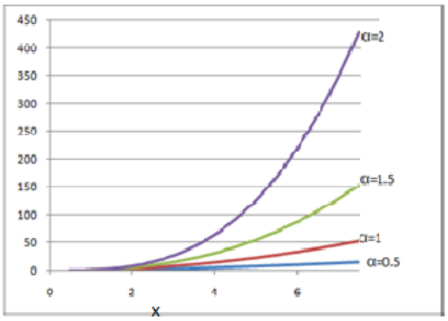

Fig. 1

From Fig. 1 it is clear that $D_\alpha(P, Q)$ has a steeper slope for increasing values of α.

6. New parametric symmetric divergence (distance) measure

The divergence measure proposed in (5.4) is non symmetric. Now we propose a symmetric divergence measure using (5.4) as follows:

(6.1) $$D_{J\alpha}(P, Q) = D_\alpha(P, Q)+ D_\alpha(Q, P).$$

From the discussion carried in Section 5 and from (6.1) we conclude that:

- $D_\alpha(P,Q)\geq 0$;
- $D_{J\alpha}(P, Q) = 0$ iff $P = Q$;

- $D_{J\alpha}(P, Q) = D_{J\alpha}(P, Q)$.

Now if the triangle inequality is satisfied, then $D_{J\alpha}(P, Q)$ becomes a distance measure. For this we prove the following.

Theorem 6.1. Let

(6.2) $$d_\alpha(p,q) = \frac{p\sinh\left(\alpha\log\frac{p}{q}\right) + q\sinh\left(\alpha\log\frac{q}{p}\right)}{\sinh(\alpha)}, \quad \alpha > 0.$$

Then $d_\alpha(p, q) \le d_\alpha(p, r) + d_\alpha(r, q)$ for $p,\ q,\ r \in \Re^+$.

Proof. (i) First we prove the result for $a \neq 0$.

To prove the desired result we prove:

(6.3) $$\sqrt{d_\alpha(p,q)} \le \sqrt{d_\alpha(p,r)} + \sqrt{d_\alpha(r,q)}.$$

Let

(6.4) $$K_{pq} = \sqrt{d_\alpha(p,r)} + \sqrt{d_\alpha(r,q)}.$$

We have

(6.5) $$K'_{pq} = \frac{d}{dr}(K_{pq}) = \frac{d'_\alpha(p,r)}{2\sqrt{d_\alpha(p,r)}} + \frac{d'_\alpha(r,q)}{2\sqrt{d_\alpha(r,q)}}.$$

Now

(6.6) $$d_\alpha(p,r) = \frac{p\sinh\left(\alpha\log\frac{p}{r}\right) + r\sinh\left(\alpha\log\frac{r}{p}\right)}{\sinh(\alpha)}, \quad \alpha > 0.$$

Therefore,

(6.7) $$d_\alpha(t,1) = \frac{t\sinh\left(\alpha\log t\right) + \sinh\left(\alpha\log\frac{1}{t}\right)}{\sinh(\alpha)}, \quad \alpha > 0.$$

From (6.6) and (6.7) we have

(6.8) $$\sqrt{d_\alpha(p,r)} = \sqrt{r}\sqrt{d_\alpha(t,1)}.$$

Now

(6.9) $$d'_\alpha(p,r) = \frac{d}{dr}(d_\alpha(p,r)) =$$

$$= \frac{1}{\sinh(\alpha)}\left[-\alpha t\cosh(\alpha\log t) + \sinh\left(\alpha\log\frac{1}{t}\right) + \alpha\cosh\left(\alpha\log\frac{1}{t}\right)\right], \text{ where } p/r = t.$$

Let

$$n(t) = d'_\alpha(p,r)\big|_{\frac{p}{r}=t} = \frac{1}{\sinh(\alpha)} -$$

$$-\left[\alpha t\cosh(\alpha\log t) + \sinh\left(\alpha\log\frac{1}{t}\right) + \alpha\cosh\left(\alpha\log\frac{1}{t}\right)\right], \text{ and } \ d(t) = \sqrt{d_\alpha(t,1)}.$$

Define a function

(6.10) $$h(t) = \frac{n(t)}{d(t)}.$$

Clearly $d(t)$ is positive for all $t > 0$ and $\alpha > 0$.

Therefore the sign of $h(t)$ depends on $n(t)$. We have

$$n'(t) = -\frac{1}{\sinh(\alpha)}\Big[-\alpha^2 \sinh(\alpha \log t) + \alpha \cosh\left(\alpha \log t\right) +$$
$$+\alpha\, t \cosh\left(\alpha \log \frac{1}{t}\right) + \frac{\alpha^2}{t}\sinh\left(\alpha \log \frac{1}{t}\right)\Big].$$

Thus, $n'(t) \le 0$, for all $t > 0$ and $\alpha > 0$; $n(t)$ is a monotonic decreasing function with $n(1) = 0$ therefore $h(t)$ is a monotonic decreasing function with $h(1) = 0$, hence $h(t)$ changes its sign at $t=1$ and it is observed that:

$$h(t) = \begin{cases} > 0, & t < 1, \\ < 0, & t > 1, \end{cases}$$

as $\frac{p}{r} = t$, then $\frac{q}{r} = \frac{q}{p}\frac{p}{r} = \beta t$, where $\beta = \frac{q}{p}$.

Now (6.5) can be written as

(6.11) $$2\sqrt{r}K'_{pq}(r) = h(t) + h(\beta t).$$

Now suppose $\beta > 1$, $q > p$, this gives:

- for $t < \frac{1}{\beta} < 1$: $h(t)$ and $h(\beta t)$ both are positive;
- for $t > 1$: $h(t)$ and $h(\beta t)$ both are negative;
- for $t \in \left(\frac{1}{\beta},\ 1\right)$: $h(t) > 0$ and $h(\beta t) < 0$.

Since $h(t)$ is a monotonic decreasing function, therefore for

- $t \in \left(\frac{1}{\beta}, 1\right)$ we have $h(t) > h(\beta t)$, this implies $h(t) + h(\beta t) > 0$;
- for $t > 1$, $h(t) < 0$ and $h(\beta t) < 0$ that is, $h(t) + h(\beta t) < 0$.

Therefore $K'_{pq} = \frac{d}{dr}(K_{pq})$, indeed changes the sign at $t = 1$, $r = p$. Thus there is a minimum at $t = 1$.

Since $h(t)$ is monotonically decreasing, this implies that $h'(t) \le 0$, for all $t > 0$ and we know that $h(t)$ changes the sign only once. This gives

$$\frac{d}{dt}\big(h(t) + h(\beta t)\big) = h'(t) + \beta\, h'(\beta t) < 0.$$

The case $\beta < 1$, $q < p$, can be investigated similarly.

Now by symmetry $K'_{pq} = \frac{d}{dr}(K_{pq})$ changes the sign at $r = q$.

Hence the proof of (6.3) follows.

Next we prove the result for $\alpha = 0$.

When $\alpha \to 0$ then (6.2) reduces to

(6.12) $$T(p,q) = (p-q)\log\frac{p}{q}.$$

And we shall show that

(6.13) $$\sqrt{T(p,q)} \le \sqrt{T(p,q)} + \sqrt{T(p,r)}, \text{ for } p,q,r \in \Re^{+}.$$

Let

$$T_{pq}(r) = \sqrt{T(p,r)} + \sqrt{T(r,p)},$$

then

(6.14) $$T'_{pq}(r) = \frac{d}{dr}(T_{pq}) = \frac{T'(p,q)}{2\sqrt{T(p,q)}} + \frac{T'(r,q)}{2\sqrt{T(r,q)}}.$$

Now we have

(6.15) $$T(p,r) = (p-r)\log\frac{p}{r} = r\left(\frac{p}{r}-1\right)\log\frac{p}{r},$$

(6.16) $$T'(p,r) = \frac{d}{dr}(T(p,r)) = 1 - \frac{p}{r} - \log\frac{p}{r}.$$

This gives

(6.17) $$\frac{T'(p,r)}{2\sqrt{T(p,r)}} = \frac{h_1(t)}{2\sqrt{r}}, \text{ where } t = \frac{p}{r},$$

(6.18) $$h_1(t) = \frac{(1-t)-\log t}{\sqrt{(t-1)\log t}} = \frac{n_1(t)}{d_1(t)}.$$

Here $d_1(t) = \sqrt{(t-1)}\log t$ is positive for all $t > 0$, the sign of $h_1(t)$ is determined by $n_1(t)$.

Now $n_1'(t) = -1 - \frac{1}{t} < 0$, for all $t > 0$.

Since $n_1(t) = 0$, then $n_1(t)$ changes the sign at $t = 1$. This results in the fact that $h_1(t)$ changes the sign at $t = 1$ and it is observed that

$$h_1(t) = \begin{cases} > 0, & t < 1, \\ < 0, & t > 1. \end{cases}$$

As $\frac{p}{r} = t$, then $\frac{q}{r} = \frac{q}{p}\frac{p}{r} = \beta t$, where $\beta = \frac{q}{p}$.

Now (6.14) can be written as

(6.19) $$2\sqrt{r}T'_{pq}(r) = h_1(t) + h_1(\beta t),\ t = \frac{p}{r}.$$

Now suppose $\beta > 1$, $q > p$, this gives:

- for $t < \dfrac{1}{\beta} < 1$: $h_1(t)$ and $h_1(\beta t)$ both are positive;

- for $t > 1$: $h_1(t)$ and $h_1(\beta t)$ both are negative;
- for $t \in (\frac{1}{\beta}, 1)$: $h_1(t) > 0$ and $h_1(\beta t) < 0$.

Since $h(t)$ is a monotonic decreasing function, therefore for $t \in (\frac{1}{\beta}, 1)$, $h(t) > h(\beta t)$ this implies $h_1(t) + h_1(\beta t) > 0$.

- for $t > 1$, $h_1(t) < 0$ and $h_1(\beta t) < 0$, that is, $h_1(t) + h_1(\beta t) < 0$.

Therefore $T'_{pq}(r) = \frac{d}{dr}(T_{pq})$ indeed changes the sign at $t = 1$, $r = p$. Thus there is a minimum at $t = 1$.

Since $h_1(t)$ is monotonically decreasing, this implies that $h'_1(t) \le 0$. For all $t > 0$ and we know that $h(t)$ changes the sign only once. This gives

$$\frac{d}{dt}(h_1(t) + h_1(\beta t)) = h'_1(t) + h'_1(\beta t) < 0.$$

The case $\beta < 1$, $q < p$, can be investigated similarly.

Now by symmetry $T'_{pq}(r) = \frac{d}{dr}(T_{pq})$ changes the sign at $r = q$.

Hence the proof of (6.13) follows.

In view of this result we conclude that the new parametric symmetric divergence measure

$D_{J\alpha}(P, Q) = D_\alpha(P, Q) + D_\alpha(Q, P)$ is a *distance measure.*

7. Information bounds of the new parametric divergence measure

Many of the divergence measures used in statistics are of the *f*-divergence type. Often one is interested in the inequalities for one *f*-divergence in terms of another *f*-divergence. Such inequalities are for instance needed in order to calculate the relative efficiency of two *f*-divergences when used for testing goodness of fit but there are many other applications.

First we cite a divergence measure in terms of which bounds are obtained in this section:

• **Hellinger Discrimination (**H e l l i n g e r [11]**)**

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

Following the theorems provides bounds on Csiszár's *f*-divergence measure.

Theorem 7.1 (D r a g o m i r [6]). Let $f : \Re^+ \to \Re$ be differentiable convex. If $p, q \in \Re_+^n$ are such that $P_n = Q_n$, where $P_n = \sum_{i=1}^{n} p_i$ and $Q_n = \sum_{i=1}^{n} p_i$ and

$$m \le p_i - q_i \le M,\ i = 1, 2, ..., n,$$

$$0 \le r \le \frac{p_i}{q_i} \le R < \infty,\ i = 1, 2, ..., n,$$

then we have the inequality

$$0 \le I_f(p,q) - Q_n\, f(1) \le \left[\frac{n}{2}\right]\left(1 - \frac{1}{n}\left[\frac{n}{2}\right]\right)(m - M)(f'(R) - f'(r)). \tag{7.1}$$

Here we have the case in which $P_n = Q_n = 1$, $f(1) = 0$, $p = P$, $q = Q$ and $I_f(p, q) = D_\alpha(P,Q)$, therefore the inequality (7.1) is reduced to

$$0 \le D_\alpha(P,Q) \le \left[\frac{n}{2}\right]\left(1 - \frac{1}{n}\left[\frac{n}{2}\right]\right)(m - M)(f'(R) - f'(r)). \tag{7.2}$$

Theorem 7.2 (Dragomir [6]).

Let $f : \Re^+ \to \Re$ be differentiable convex. If $p, q \in \Re_+^n$ are such that $P_n = Q_n$, where $P_n = \sum_{i=1}^{n} p_i$, and $Q_n = \sum_{i=1}^{n} p_i$, and $0 \le r \le \frac{p_i}{q_i} \le R < \infty,\ i = 1, 2, ..., n,$

then we have the inequality

$$0 \le I_f(p,q) - Q_n\, f(1) \le \frac{1}{4}(R - r)(f'(R) - f'(r)). \tag{7.3}$$

Here we have the case in which $P_n = Q_n = 1$, $f(1) = 0$, $p = P$, $q = Q$ and $I_f(p, q) = D_\alpha(P, Q)$, therefore the inequality (7.3) is reduced to

$$0 \le D_\alpha(P,Q) \le \frac{1}{4}(R - r)(f'(R) - f'(r)). \tag{7.4}$$

Theorem 7.3 (Dragomir [7]). Assume that the generating mapping $f : (0,1) \to \Re$ is normalized, i.e., $f(1) = 0$ and it satisfies the assumptions:

(i) f is twice differentiable on (r, R), where $0 < r \le R < \infty$,

(ii) there exist real constants m, M, such that $m \le x\, f''(x) \le M$ for $x \in (r, R)$.

If P, Q are discrete probability distributions satisfying the assumption

$$0 \le r = r_i \le \frac{p_i}{q_i} \le R,\ i = 1, 2, ..., n,$$

then we have the inequality

$$m\, D(P,Q) \le I_f(P,Q) \le M\, D(P,Q), \tag{7.5}$$

where $D(P, Q)$ is Kullback Liebler's divergence measure.

Here we have $I_f(P, Q) = D_\alpha(P, Q)$,

$$f(x) = \frac{x \sinh(\alpha \log x)}{\sinh(\alpha)},\ \alpha > 0,$$

$$x\, f''(x) = \frac{\alpha \cosh(\alpha \log x) + \alpha^2 \sinh(\alpha \log x)}{\sinh(\alpha)},\ \alpha > 0. \tag{7.6}$$

It can be observed that the minimum value of $x\,f''(x)$ is obtained at $x=\left(\dfrac{\alpha+1}{\alpha-1}\right)^{\frac{1}{2\alpha}}$ provided that $\alpha>1$ and the maximum value of $x\,f''(x)$ cannot be determined here. We can find m by putting this value of x in (7.6). Consequently, we find a lower bound for $D_\alpha(P, Q)$ from (7.5) as follows:

(7.7) $$m\,D(P,Q)\le D_\alpha(P,Q).$$

Theorem 7.4 (Dragomir [8]). Assume that the generating mapping $f:(0,1)\to\Re$ is normalized, i.e., $f(1) = 0$ and satisfies the assumptions:

(i) f is twice differentiable on (r, R), where $0<r\le 1\le R<\infty$,

(ii) there exist real constants m, M such that $m\le x^{3/2}f''(x)\le M$, for $x\in(r,R)$.

If P, Q are discrete probability distributions satisfying the assumption

$$0\le r=r_i\le\frac{p_i}{q_i}\le R,\ i=1,2,...,n,$$

then we have the inequality

(7.8) $$4m\,h^2(P,Q)\le I_f(P,Q)\le 4M\,h^2(P,\ Q),$$

where $h^2(P,Q)$ is Hellinger discrimination.

Here we have $I_f(p, q) = D_\alpha(P, Q)$,

$$f(x)=\frac{x\sinh(\alpha\log x)}{\sinh(\alpha)},\ \alpha>0,$$

(7.9) $$x^{3/2}f''(x)=x^{1/2}\left[\frac{\alpha\cosh(\alpha\log x)+\alpha^2\sinh(\alpha\log x)}{\sinh(\alpha)}\right],\ \alpha>0.$$

It can be observed that the minimum value of $x^{1/2}f''(x)$ is obtained at $x=\left(\dfrac{(\alpha+1)(2\alpha+1)}{(\alpha-1)(2\alpha-1)}\right)^{\frac{1}{2\alpha}}$ provided that $\alpha>1$ and the maximum value of $x^{3/2}f''(x)$ cannot be determined here. We can find m by putting this value of x in (7.9). Consequently, we find a lower bound for $D_\alpha(P, Q)$ from (7.8) as follows:

(7.10) $$4m\,h^2(P,Q)\le D_\alpha(P,Q).$$

The upper bound for $D_\alpha(P, Q)$ is given by the following

Theorem 7.5 (Dragmir [6, 9]). Let the function $f:\Re_+\to\Re$ be differentiable convex and normalized, i.e., $f(1) = 0$. Then

(7.11) $$0\le C_f(P,Q)\le E_{C_f}(P,Q),$$

where

(7.12) $$E_{C_f}(P,Q)=\sum_{i=1}^{n}(p_i-q_i)\,f'\!\left(\frac{p_i}{q_i}\right)\ \text{ for all }\ P,\ Q\in\Gamma_n.$$

Let $P,\ Q \in \Gamma_n$ be such that $0 < r = r_i \le \frac{p_i}{q_i} \le R < \infty, \quad i = 1, 2, ..., n,$ then

(7.13) $$0 \le C_f(P,Q) \le A_{C_f}(P,Q),$$

where

(7.14) $$A_{C_f}(r,R) = \frac{1}{4}(R-r)[f'(R) - f'(r)].$$

Further, if we suppose that $0 < r \le 1 \le R < \infty,\ r \ne R,$ then

(7.15) $$0 \le C_f(P,Q) \le B_{C_f}(P,Q),$$

where

(7.16) $$B_{C_f}(r,\ R) = \frac{(R-r)f(r) + (1-r)f(r)}{R-r}.$$

Theorem (7.5) provides the upper bounds for $D_\alpha(P, Q)$.

8. Approximation of the new divergence measure in terms of χ^2 divergence measure

χ^2 divergence proposed by P e a r s o n s [14]

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

In this section we shall bring the asymptotic approximation of the divergence measure given by (14) in terms of χ^2 divergence.

Lemma 8.1. If f is twice differentiable at $x = 1$ and $f''(x) \ge 0$, $f(1) = 0$, then

(8.1) $$C_f(P,Q) \approx \frac{f''(1)}{2}\chi^2(P,Q).$$

P r o o f. From Taylor's expansion, we have

$$f(x) = f'(1)(x-1) + \frac{f''(1)(x-1)^2}{2} + K(x)(x-1)^2,$$

where $f(1) = 0$ and $K(x) \to 0$ as $x \to 1$.

Hence, $q_i f\left(\frac{p_i}{q_i}\right) = f'(1)(p_i - q_i) + \frac{f''(1)(p_i - q_i)^2}{2q_i} + K\left(\frac{p_i}{q_i}\right)\frac{(p_i - q_i)^2}{q_i}.$

Approximating $p_i \to q_i$ and summing over $i = 1, 2, 3, \ldots, n$ we have

$$C_f(P,Q) \approx \frac{f''(1)}{2}\chi^2(P,Q).$$

Theorem 8.1. For $D_\alpha(P, Q)$ as defined in (14) the following result holds:

(8.2) $$D_\alpha(P,\ Q) \approx \frac{\alpha}{2\sinh(\alpha)}\chi^2(P,\ Q).$$

P r o o f. We have

$$f(x) = \frac{x \sinh(\alpha \log x)}{\sinh(\alpha)}, \ \alpha > 0.$$

Clearly, $f(x)$ satisfies the conditions of Lemma 8.1, thus we have

$$C_f(P,Q) \approx \frac{f''(1)}{2}\chi^2(P,Q), \text{ and } f''(1) = \frac{\alpha}{\sinh(\alpha)}.$$

From Section 3 we also know that

$$D_\alpha(P,Q) = C_f(P,Q).$$

This gives
$$D_\alpha(P,Q) \approx \frac{\alpha}{2\sinh(\alpha)}\chi^2(P,Q).$$

Also, when $\alpha \to 0$, we have $\frac{\alpha}{\sinh(\alpha)} \to 1.$

Consequently,

$$D_\alpha(P,Q) \approx \frac{1}{2}\chi^2(P,Q).$$

Thus the desired result follows.

References

1. B a s s e v i l l e, M. Divergence Measures for Statistical Data Processing. Publications Internes de l'IRISA, November 2010.
2. B r e g m a n, L. M. The Relaxation Method of Finding the Common Point of Convex Sets and Its Application to the Solution of Problems in Convex Programming. – USSR Computational Mathematics and Mathematical Physics, Vol. **7**, 1967, No 3, 200-217.
3. C o v e r, T. M., A. J. T h o m a s. Elements of Information Theory. Wiley India Pvt. Ltd. New Delhi, 2009.
4. C s i s z a r, I. Eine Informations Theoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizitat von marko_schen ketten. Magyar. – Tud. Akad. Mat. Kutato Int. Kozl, Vol. **8**, 1963, 85-108.
5. C s i s z a r, I. On Topological Properties of *f*-Divergences. – Studia Math. Hungarica, Vol. **2**, 1967, 329-339.
6. D r a g o m i r, S. S. Some Inequalities for the Csiszar's ∅-Divergence. Inequalities for the Csizar's *f*-Divergence in Information Theory, 2000.
   http://rgmia.vu.edu.au/monographs/ciszar.htm
7. D r a g o m i r, S. S. Upper and Lower Bounds for Csiszar's *f*-Divergence in Terms of the Kullback-Leibder Distance and Applications. Inequalities for the Csiszar's *f*-divergence in Information Theory, 2000.
   http://rgmia.vu.edu.au/monographs/ciszar.htm
8. D r a g o m i r, S. S. Upper and Lower Bounds for Csiszar's *f*-Divergence in Terms of the Hellinger Discrimination and Applications. Inequalities for the Csiszar's *f*-Divergence in Information Theory, 2000.
   http://rgmia.vu.edu.au/monographs/csiszar.htm
9. D r a g o m i r, S. S. Other Inequalities for the Csiszar's Divergence and Applications. Inequalities for the Csizar's *f*-Divergence in Information Theory, 2000.
   http://rgmia.vu.edu.au/monographs/ciszar.htm
10. E s t e b a n, M. D., D. M o r a l e s. A Summary on Entropy Statistic. – Kybernetika, Vol. **31**, 1995, No 4, 337-346.

11. H e l l i n g e r, E. Neue Berunduring der Theorie der Quadratischen Formen von Un endlichen Vieden Veran derliehen. – J. Reine Aug. Math., Vol. **136**, 1909, 210-271.
12. K u l l b a c k, S., R. A. L e i b l e r. On Information and Sufficiency. – Ann. Math. Statist., Vol. **22**, 1951, 79-86.
13. M a j i, P. *f*-Information Measures for Efficient Selection of Discriminative Genes From Microarray Data. – IEEE Transactions on Biomedical Engineering, Vol. **56**, 2009, No 4, 1063-1069.
14. P e a r s o n, 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., Vol. **50**, 1900, 157-172.
15. R e n y i, A. On Measures of Entropy and Information. – In: Proc. of 4th Berkeley Symp. Math. Stat. Probab., Vol. **1**, 1961, 547-561.
16. S a l i c r u, M. Measures of Information Associated with Csiszar's Divergences. – Kybernetika, Vol. **50**, 1994, No 5, 563-573.
17. S h a n n o n, C. E. The Mathematical Theory of Communications. – Bell Syst. Tech. Journal, Vol. **27**, 1948, 423-467.
18. T a n e j a, I. J. Generalized Information Measures and their Applications. – On-Line Book**,** 2001. **http://www.mtm.ufsc.br/~taneja/book/book.html**
19. T a n e j a, I. J. Generalized Arithmetic and Geometric Mean Divergence Measure and Their Statistical Aspects. Available at: arXiv:math/0501297v1[math.ST] 19 Jan 2005.
20. T a n e j a, I. J. On Mean Divergence Measure. Available at: arXiv:math/0501298v2[math.ST] 13 June 2005.
21. W a n g,Y. Generelized Information Theory: A Review and Outlook. – Information Technology Journal, Vol. **10**, 2011, No 3, 461-469.