ACTA MATHEMATICA VIETNAMICA
Volume 28, Number 2, 2003, pp. 123-134

NEW INEQUALITIES FOR CSISZÁR DIVERGENCE AND APPLICATIONS

S. S. DRAGOMIR

ABSTRACT. In this paper we point out some new inequalities for Csiszár f-divergence and apply them for particular instances of distances between two probability distributions.

1. INTRODUCTION

One of the important issues in many applications of Probability Theory is finding an appropriate measure of *distance* (or *difference* or *discrimination*) between two probability distributions. A number of divergence measures for this purpose have been proposed and extensively studied by Jeffreys [1], Kullback and Leibler [2], Rényi [3], Havrda and Charvat [4], Kapur [5], Sharma and Mittal [6], Burbea and Rao [7], Rao [8], Lin [9], Csiszár [10], Ali and Silvey [12], Vajda [13], Shioya and Da-te [40] and others (see for example [5] and the references therein).

These measures have been applied in a variety of fields such as: anthropology [8], genetics [14], finance, economics, and political science [15], [16], [17], biology [18], the analysis of contingency tables [19], approximation of probability distributions [20], [21], signal processing [22], [23] and pattern recognition [24], [25]. A number of these measures of distance are specific cases of Csiszár f-divergence and so further exploration of this concept will have a flow on effect to other measures of distance and to areas in which they are applied.

Assume that a set χ and the σ-finite measure μ are given. Consider the set of all probability densities on μ to be

$$\Omega := \left\{ p | p : \chi \to \mathbb{R},\, p(x) \geq 0,\ \int_\chi p(x) d\mu(x) = 1 \right\}.$$

The Kullback-Leibler divergence [2] is well known among the information divergences. It is defined as:

$$D_{KL}(p, q) := \int_\chi p(x) \log \left[\frac{p(x)}{q(x)} \right] d\mu(x), \quad p, q \in \Omega, \tag{1.1}$$

where log is to base 2.

Received September 28, 2001.
1991 *Mathematics Subject Classification.* Primary 26D15; Secondary 94A99.
Key words and phrases. Csiszár Divergence.

In Information Theory and Statistics, various divergences are applied in addition to the Kullback-Leibler divergence. These are the: *variation distance* D_v, *Hellinger distance* D_H [1], χ^2-*divergence* D_{χ^2}, α-*divergence* D_α, *Bhattacharyya distance* D_B [2], *Harmonic distance* D_{Ha}, *Jeffreys distance* D_J [1], *triangular discrimination* D_Δ [35], etc... They are defined as follows:

$$D_v(p,q) := \int_\chi |p(x) - q(x)|\, d\mu(x), \quad p,q \in \Omega; \tag{1.2}$$

$$D_H(p,q) := \int_\chi \left[\sqrt{p(x)} - \sqrt{q(x)}\right]^2 d\mu(x), \quad p,q \in \Omega; \tag{1.3}$$

$$D_{\chi^2}(p,q) := \int_\chi p(x)\left[\left(\frac{q(x)}{p(x)}\right)^2 - 1\right] d\mu(x), \quad p,q \in \Omega; \tag{1.4}$$

$$D_\alpha(p,q) := \frac{4}{1-\alpha^2}\left[1 - \int_\chi [p(x)]^{\frac{1-\alpha}{2}} [q(x)]^{\frac{1+\alpha}{2}} d\mu(x)\right], \quad p,q \in \Omega; \tag{1.5}$$

$$D_B(p,q) := \int_\chi \sqrt{p(x)\, q(x)}\, d\mu(x), \quad p,q \in \Omega; \tag{1.6}$$

$$D_{Ha}(p,q) := \int_\chi \frac{2p(x)\, q(x)}{p(x) + q(x)} d\mu(x), \quad p,q \in \Omega; \tag{1.7}$$

$$D_J(p,q) := \int_\chi [p(x) - q(x)] \ln\left[\frac{p(x)}{q(x)}\right] d\mu(x), \quad p,q \in \Omega; \tag{1.8}$$

$$D_\Delta(p,q) := \int_\chi \frac{[p(x) - q(x)]^2}{p(x) + q(x)} d\mu(x), \quad p,q \in \Omega. \tag{1.9}$$

For other divergence measures, see the paper [5] by Kapur or the book on line [6] by Taneja. For a comprehensive collection of preprints available on line, see the RGMIA web site `http://rgmia.vu.edu.au/papersinfth.html`

Csiszár $f-$divergence is defined as follows [10]

$$I_f(p,q) := \int_\chi p(x)\, f\left[\frac{q(x)}{p(x)}\right] d\mu(x), \quad p,q \in \Omega, \tag{1.10}$$

where f is convex on $(0,\infty)$. It is assumed that $f(u)$ is zero and strictly convex at $u = 1$. By appropriately defining this convex function, various divergences are derived. All the above distances $(1.1) - (1.9)$, are particular instances of Csiszár $f-$divergence. There are also many others which are not in this class (see for example [5] or [6]). For the basic properties of Csiszár f-divergence see [7]-[10].

2. The results

We start with the following result.

Theorem 1. *Let $\phi : [0,\infty) \to \mathbb{R}$ be a convex mapping on the interval $[r,R] \subset [0,\infty)$ with $r \le 1 \le R$. If $p,q \in \Omega$ and $r \le \frac{p(y)}{q(y)} \le R$ for all $y \in \chi$, then we have* *the inequality*

$$I_{\phi}(p,q) \leq \frac{R-1}{R-r} \cdot \phi(r) + \frac{1-r}{R-r} \cdot \phi(R). \tag{2.1}$$

Proof. As ϕ is convex on $[r,R]$, we may write that

$$\phi(tr + (1-t)R) \leq t\phi(r) + (1-t)\phi(R) \tag{2.2}$$

for all $t \in [0,1]$.

Choose $t = \dfrac{R-x}{R-r}$, $x \in [r,R]$. Then $1-t = \dfrac{x-r}{R-r}$ and from (2.2) we deduce (see also [46, p. 98])

$$\phi(x) \leq \frac{R-x}{R-r} \cdot \phi(r) + \frac{x-r}{R-r} \cdot \phi(R) \tag{2.3}$$

for all $x \in [r,R]$, as a simple calculation shows that $\dfrac{R-x}{R-r} \cdot r + \dfrac{x-r}{R-r} \cdot R = x$. Put in (2.3) $x = \dfrac{p(y)}{q(y)}$, $y \in \chi$, to get

$$\phi\left(\frac{p(y)}{q(y)}\right) \leq \frac{R - \frac{p(y)}{q(y)}}{R-r} \cdot \phi(r) + \frac{\frac{p(y)}{q(y)} - r}{R-r} \cdot \phi(R) \tag{2.4}$$

for all $y \in \chi$.

If we multiply (2.4) by $q(y) \geq 0$, integrate on χ and take into account that

$$\int_{\chi} p(y)\, d\mu(y) = \int_{\chi} q(y)\, d\mu(y) = 1$$

then by (2.4) we obtain (2.1). □

The following result also holds.

Theorem 2. *Let $\phi : [0,\infty) \to \mathbb{R}$ be differentiable convex on $[r,R]$ and p,q be as in Theorem 1. Then we have the inequality:*

$$\begin{aligned} 0 &\leq \frac{R-1}{R-r} \cdot \phi(r) + \frac{1-r}{R-r} \cdot \phi(R) - I_{\phi}(p,q) \\ &\leq \frac{\phi'(R) - \phi'(r)}{R-r} \cdot \left[(R-1)(1-r) - D_{\chi^2}(p,q)\right] \\ &\leq \frac{1}{4}(R-r)\left[\phi'(R) - \phi'(r)\right], \end{aligned} \tag{2.5}$$

where $D_{\chi^2}(\cdot,\cdot)$ is the chi-square divergence.

Proof. Since the mapping ϕ is differentiable convex, we can write

$$\phi(u) - \phi(v) \geq \phi'(v)(u-v) \tag{2.6}$$

for all $u,v \in (r,R)$.

Now, assume that $\alpha, \beta \geq 0$ and $\alpha + \beta > 0$. Then, by (2.6), we have

$$\phi\left(\frac{\alpha a+\beta b}{\alpha+\beta}\right)-\phi(a) \geq \phi'(a)\left(\frac{\alpha a+\beta b}{\alpha+\beta}-a\right) = \frac{\beta}{\alpha+\beta}\cdot\phi'(a)(b-a) \tag{2.7}$$

and

$$\phi\left(\frac{\alpha a+\beta b}{\alpha+\beta}\right)-\phi(b) \geq \phi'(b)\left(\frac{\alpha a+\beta b}{\alpha+\beta}-b\right) = -\frac{\alpha}{\alpha+\beta}\cdot\phi'(b)(b-a). \tag{2.8}$$

Now, if we multiply (2.7) by α and (2.8) by β and add the obtained results, we get

$$(\alpha+\beta)\phi\left(\frac{\alpha a+\beta b}{\alpha+\beta}\right)-\alpha\phi(a)-\beta\phi(b) \geq \frac{\alpha\beta}{\alpha+\beta}(b-a)\left(\phi'(a)-\phi'(b)\right)$$

which is equivalent to:

$$0 \leq \frac{\alpha\phi(a)+\beta\phi(b)}{\alpha+\beta}-\phi\left(\frac{\alpha a+\beta b}{\alpha+\beta}\right) \leq \frac{\alpha\beta}{(\alpha+\beta)^2}\left(\phi'(b)-\phi'(a)\right)(b-a). \tag{2.9}$$

Now, if in (2.9) we choose $\alpha = R - x$, $\beta = x - r$, $a = r$, $b = R$, then we obtain

$$0 \leq \frac{(R-x)\phi(r)+(x-r)\phi(R)}{R-r}-\phi(x) \leq \frac{(R-x)(x-r)}{R-r}\left(\phi'(R)-\phi'(r)\right). \tag{2.10}$$

If in (2.10), we choose $x = \frac{p(y)}{q(y)}$ and then multiply with $q(y)$ we get

$$\frac{(Rq(y)-p(y))\phi(r)+(p(y)-rq(y))\phi(R)}{R-r}-q(y)\phi\left(\frac{p(y)}{q(y)}\right) \leq \frac{(Rq(y)-p(y))(p(y)-rq(y))}{(R-r)q(y)}\left(\phi'(R)-\phi'(r)\right) \tag{2.11}$$

for all $y \in \chi$.

If we integrate (2.11) on χ and take into consideration that

$$\int_\chi p(y)\,d\mu(y) = \int_\chi q(y)\,d\mu(y) = 1,$$

we get

$$\frac{(R-1)\phi(r)+(1-r)\phi(R)}{R-r}-I_\phi(p,q) \leq \frac{(\phi'(R)-\phi'(r))}{R-r}\int_\chi \frac{(Rq(y)-p(y))(p(y)-rq(y))}{q(y)}d\mu(y). \tag{2.12}$$

However,

$$\begin{aligned} 0 &\le \int_{\chi} \frac{(Rq(y) - p(y))(p(y) - rq(y))}{q(y)} d\mu(y) \\ &= R - \int_{\chi} \frac{p^2(y)}{q(y)} d\mu(y) - rR + r = R + r - rR - 1 - D_{\chi^2}(p,q) \\ &= (R-1)(1-r) - D_{\chi^2}(p,q). \end{aligned}$$

As

$$(R-1)(1-r) \le \frac{1}{4}(R-r)^2 \quad \text{and} \quad D_{\chi^2}(p,q) \ge 0,$$

the last inequality is obvious. □

The following results also holds.

Theorem 3. *Assume that the function $\Phi : [0,\infty) \to \mathbb{R}$ is twice differentiable on $[r,R]$ and*

$$m \le \Phi''(t) \le M \quad \text{for all } t \in [r,R]. \tag{2.13}$$

If the probability distributions $p, q \in \Omega$ satisfy the conditions of Theorem 1, then we have the inequality:

$$\begin{aligned} \frac{1}{2} m \left[(R-1)(1-r) - D_{\chi^2}(p,q)\right] &\le \frac{R-1}{R-r} \cdot \Phi(r) + \frac{1-r}{R-r} \cdot \Phi(R) - I_{\Phi}(p,q) \\ &\le \frac{1}{2} M \left[(R-1)(1-r) - D_{\chi^2}(p,q)\right]. \end{aligned} \tag{2.14}$$

Proof. Define the function $\Phi_m : [0,\infty) \to \mathbb{R}$, $\Phi_m(t) = \Phi(t) - \frac{1}{2} m t^2$. Then Φ_m is twice differentiable and $\Phi_m''(t) = \Phi''(t) - m \ge 0$, $t \in [r,R]$, which shows that Φ_m is convex on $[r,R]$.

If we write the inequality (2.1) for the convex mapping Φ_m, we obtain

$$I_{\Phi - \frac{1}{2}m(\cdot)^2}(p,q) \le \frac{R-1}{R-r}\left[\Phi(r) - \frac{1}{2} m r^2\right] + \frac{1-r}{R-r}\left[\Phi(R) - \frac{1}{2} m R^2\right]. \tag{2.15}$$

However,

$$\begin{aligned} I_{\Phi - \frac{1}{2}m(\cdot)^2}(p,q) &= I_{\Phi}(p,q) - \frac{1}{2} m \left[\int_{\chi} \frac{p^2(y)}{q(y)} d\mu(y) - 1 + 1\right] \\ &= I_{\Phi}(p,q) - \frac{1}{2} m D_{\chi^2}(p,q) - \frac{1}{2} m \end{aligned}$$

and then, by (2.15), we can get

$$\begin{aligned} &\frac{R-1}{R-r} \cdot \Phi(r) + \frac{1-r}{R-r} \cdot \Phi(R) - I_{\Phi}(p,q) \\ &\ge \frac{1}{2} m R^2 \cdot \frac{(1-r)}{R-r} + \frac{1}{2} m r^2 \cdot \frac{(R-1)}{R-r} - \frac{1}{2} m D_{\chi^2}(p,q) - \frac{1}{2} m \end{aligned} \tag{2.16}$$

Nonetheless, the right hand side of (2.16) is

$$\frac{1}{2}m\left[(R-1)(1-r)-D_{\chi^2}(p,q)\right]$$

and the first inequality in (2.14) is obtained.

The second inequality follows by a similar argument applied for the mapping $\Phi_m(t) := \frac{1}{2}Mt^2 - \Phi(t)$. We omit the details. □

Corollary 1. *With the assumptions in Theorem 3, and if $m \geq 0$, then*

$$\begin{aligned} 0 &\leq \frac{1}{2}m\left[(R-1)(1-r)-D_{\chi^2}(p,q)\right] \\ &\leq \frac{R-1}{R-r}\cdot\Phi(r)+\frac{1-r}{R-r}\cdot\Phi(R)-I_\phi(p,q). \end{aligned} \tag{2.17}$$

Proof. We only have to prove the fact that

$$D_{\chi^2}(p,q) \leq (R-1)(1-r), \tag{2.18}$$

which follows by the fact that (see the proof of Theorem 2)

$$\begin{aligned} 0 &\leq \int_\chi \frac{(Rq(y)-p(y))(p(y)-rq(y))}{q(y)}d\mu(y) \\ &= (R-1)(1-r)-D_{\chi^2}(p,q). \end{aligned}$$

□

3. Applications for particular divergences

Before we point out some applications of the above results, we would like to recall the following special means:

$$L(\alpha,\beta) := \begin{cases} \beta & \text{if } \alpha=\beta; \\ \dfrac{\beta-\alpha}{\ln\beta-\ln\alpha} & \text{if } \beta\neq\alpha,\ \alpha,\beta>0 \ (\textit{logarithmic mean}) \end{cases}$$

and

$$I(\alpha,\beta) := \begin{cases} \beta & \text{if } \alpha=\beta; \\ \dfrac{1}{e}\left(\dfrac{\beta^\beta}{\alpha^\alpha}\right)^{\frac{1}{\beta-\alpha}} & \text{if } \beta\neq\alpha,\ (\textit{identric mean}). \end{cases}$$

1. *Kullback-Leibler Divergence.* Consider the convex mapping $\phi:(0,\infty)\to\mathbb{R}$, $\phi(t)=t\ln t$. Then

$$I_\phi(p,q)=\int_\chi p(x)\ln\left[\frac{p(x)}{q(x)}\right]d\mu(x)=D(p,q),$$

where $D(p,q)$ is the *Kullback-Leibler distance.*

Proposition 1. *Let $p, q \in \Omega$ with the property that:*

$$r \leq \frac{p(y)}{q(y)} \leq R \quad \text{for all} \quad y \in \chi. \tag{3.1}$$

Then we have the inequality

$$D(p,q) \leq \ln I(r,R) - \frac{G^2(r,R)}{L(r,R)} + 1, \tag{3.2}$$

where $I(\cdot,\cdot)$ is the identric mean, $L(\cdot,\cdot)$ is the logarithmic mean and $G(\cdot,\cdot)$ is the usual geometric mean.

Proof. We apply Theorem 1 for $\phi(t) = t \ln t$ to get

$$\begin{aligned} D(p,q) &\leq \frac{R-1}{R-r} r \ln r + \frac{1-r}{R-r} R \ln R \\ &= \frac{R \ln R - r \ln r}{R-r} - rR \cdot \frac{\ln R - \ln r}{R-r} \\ &= \ln I(r,R) + 1 - \frac{G^2(r,R)}{L(r,R)} \end{aligned}$$

and the inequality (3.2) is proved. □

Proposition 2. *With the assumptions of Proposition 1, we have*

$$\begin{aligned} 0 &\leq \ln I(r,R) - \frac{G^2(r,R)}{L(r,R)} + 1 - D(p,q) \\ &\leq \frac{(R-1)(1-r) - D_{\chi^2}(p,q)}{L(r,R)}. \end{aligned} \tag{3.3}$$

The proof follows by Theorem 2 applied for $\phi(t) = t \ln t$, and taking into account that

$$\frac{\phi'(R) - \phi'(r)}{R-r} = \frac{1}{L(r,R)}.$$

Using Theorem 3, we may be able to improve the inequality (3.3) as follows.

Proposition 3. *Let $p, q \in \Omega$ satisfy the condition* (3.1)*. Then we have the inequality:*

$$\begin{aligned} \frac{1}{2R}\left[(R-1)(1-r) - D_{\chi^2}(p,q)\right] &\leq \ln I(r,R) - \frac{G^2(r,R)}{L(r,R)} + 1 - D(p,q) \\ &\leq \frac{1}{2r}\left[(R-1)(1-r) - D_{\chi^2}(p,q)\right]. \end{aligned} \tag{3.4}$$

Proof. We have $\phi''(t) = \frac{1}{t}$, $t \in [r,R]$ and then

$$\frac{1}{R} \leq \phi''(t) \leq \frac{1}{r}, \quad t \in [r,R].$$

Applying Theorem 3 for $\phi(t) = t \ln t$, we obtain (3.4). □

Now, assume that $\phi(t) = -\ln t$, which is a convex mapping as well.

We have

$$I_{\phi}(p,q) = -\int_{\chi} q(y)\ln\left[\frac{p(y)}{q(y)}\right] d\mu(y) = \int_{\chi} q(y)\ln\left[\frac{q(y)}{p(y)}\right] d\mu(y) = D(q,p).$$

Using Theorem 1, we may state the following proposition.

Proposition 4. *Let $p,q \in \Omega$ with the property that* (3.1) *holds. Then we have the inequality:*

$$D(q,p) \le \ln I\left(\frac{1}{r},\frac{1}{R}\right) - \frac{1}{L(r,R)} + 1. \tag{3.5}$$

Proof. Applying the inequality (2.1) for $\phi(t) = -\ln t$, we may write that

$$\begin{aligned} D(q,p) &\le \frac{(R-1)(-\ln r)+(1-r)(-\ln R)}{R-r} \\ &= \frac{r\ln R - R\ln r}{R-r} - \frac{\ln R - \ln r}{R-r} = \frac{rR\left(\frac{1}{R}\ln R - \frac{1}{r}\ln r\right)}{R-r} - \frac{1}{L(r,R)} \\ &= \frac{\frac{1}{r}\ln\frac{1}{r} - \frac{1}{R}\ln\frac{1}{R}}{\frac{1}{r}-\frac{1}{R}} - \frac{1}{L(r,R)} = \ln I\left(\frac{1}{r},\frac{1}{R}\right) + 1 - \frac{1}{L(r,R)} \end{aligned}$$

and the inequality (3.5) is proved. □

Proposition 5. *Let p,q be as in Proposition 1. Then*

$$\begin{aligned} 0 &\le \ln I\left(\frac{1}{r},\frac{1}{R}\right) - \frac{1}{L(r,R)} + 1 - D(q,p) \\ &\le \frac{1}{G^2(r,R)}\left[(R-1)(1-r) - D_{\chi^2}(p,q)\right]. \end{aligned} \tag{3.6}$$

The proof follows by Theorem 2 applied for the function $\phi(t) = -\ln t$, and taking into account that

$$\frac{\phi'(R)-\phi'(r)}{R-r} = \frac{1}{rR} = \frac{1}{G^2(r,R)}.$$

The inequality (3.6) can be improved as follows.

Proposition 6. *Let p,q be as in Proposition 1. Then*

$$\begin{aligned} \frac{1}{2R^2}\left[(R-1)(1-r) - D_{\chi^2}(p,q)\right] &\le \ln I\left(\frac{1}{r},\frac{1}{R}\right) - \frac{1}{L(r,R)} + 1 - D(q,p) \\ &\le \frac{1}{2r^2}\left[(R-1)(1-r) - D_{\chi^2}(p,q)\right]. \end{aligned} \tag{3.7}$$

The proof is obvious by Theorem 3, taking into account that $\phi''(t) = \frac{1}{t^2}$ and $\frac{1}{R^2} \le \phi''(t) \le \frac{1}{r^2}$ for all $t \in [r, R]$.

2. *Hellinger discrimination.* Consider the convex mapping $\phi : [0, \infty) \to \mathbb{R}$, $\phi(t) = \frac{1}{2}\left(\sqrt{t} - 1\right)^2$. Then

$$\begin{aligned} I_\phi(p, q) &= \frac{1}{2}\int_\chi q(x)\left(\sqrt{\frac{p(x)}{q(x)}} - 1\right)^2 d\mu(x) \\ &= \frac{1}{2}\int_\chi \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 d\mu(x) = h^2(p, q), \end{aligned}$$

where $h^2(p, q)$ is the *Hellinger discrimination.*

Proposition 7. *With the assumptions of Proposition 1, we have*

$$h^2(p, q) \le \frac{\left(\sqrt{R} - 1\right)(1 - \sqrt{r})}{\sqrt{R} + \sqrt{r}}. \tag{3.8}$$

Proof. We apply Theorem 1 for $\phi(t) = \frac{1}{2}\left(\sqrt{t} - 1\right)^2$ to get

$$\begin{aligned} h^2(p, q) &\le \frac{(R-1)\frac{1}{2}(\sqrt{r} - 1)^2 + (1 - r)\frac{1}{2}\left(\sqrt{R} - 1\right)^2}{R - r} \\ &= \frac{\frac{1}{2}\left(\sqrt{R} - 1\right)(\sqrt{r} - 1)}{R - r}\left[\left(\sqrt{R} + 1\right)(1 - \sqrt{r}) + (1 + \sqrt{r})\left(\sqrt{R} - 1\right)\right] \\ &= \frac{\left(\sqrt{R} - 1\right)(\sqrt{r} - 1)\left(\sqrt{R} - \sqrt{r}\right)}{R - r} \\ &= \frac{\left(\sqrt{R} - 1\right)(1 - \sqrt{r})}{\sqrt{R} + \sqrt{r}}, \end{aligned}$$

and the inequality (3.8) is proved. □

Using Theorem 2, we may state the following proposition as well.

Proposition 8. *With the assumptions of Proposition 1, we have*

$$\begin{aligned} 0 &\le \frac{\left(\sqrt{R} - 1\right)(1 - \sqrt{r})}{\sqrt{R} + \sqrt{r}} - h^2(p, q) \\ &\le \frac{1}{4(r - R)A\left(\sqrt{r}, \sqrt{R}\right)}\left[(R - 1)(1 - r) - D_{\chi^2}(p, q)\right], \end{aligned} \tag{3.9}$$

where $A(\cdot, \cdot)$ is the arithmetic mean.

The proof is obvious by Theorem 2 applied for $\phi(t) = \frac{1}{2}\left(\sqrt{t}-1\right)^2$, taking into account that $\phi'(t) = \frac{1}{2} - \frac{1}{2\sqrt{t}}$, and

$$\frac{\phi'(R) - \phi'(r)}{R-r} = \frac{\sqrt{R}-\sqrt{r}}{2\sqrt{rR}\,(R-r)} = \frac{1}{2\sqrt{rR}\left(\sqrt{R}+\sqrt{r}\right)}.$$

Finally, by the use of Theorem 3, we may state:

Proposition 9. *Assume that $p, q \in \Omega$ are as in Proposition 1. Then*

(3.10)

$$\begin{aligned}\frac{1}{8\sqrt{R^3}}\left[(R-1)(1-r) - D_{\chi^2}(p,q)\right] &\le \frac{\left(\sqrt{R}-1\right)\left(1-\sqrt{r}\right)}{\sqrt{R}+\sqrt{r}} - h^2(p,q) \\ &\le \frac{1}{8\sqrt{r^3}}\left[(R-1)(1-r) - D_{\chi^2}(p,q)\right].\end{aligned}$$

The proof follows by Theorem 3 applied for the mapping $\phi(t) = \frac{1}{2}\left(\sqrt{t}-1\right)^2$ for which $\phi''(t) = \frac{1}{4\sqrt{t^3}}$ and, obviously,

$$\frac{1}{4\sqrt{R^3}} \le \phi''(t) \le \frac{1}{4\sqrt{r^3}} \quad \text{for all } t \in [r, R].$$

References

[1] H. Jeffreys, *An invariant form for the prior probability in estimating problems*, Proc. Roy. Soc. London **186** A (1946), 453-461.

[2] S. Kullback and R. A. Leibler, *On information and sufficiency*, Ann. Math. Stat. **22** (1951), 79-86.

[3] A. Rényi, *On measures of entropy and information*, Proc. Fourth Berkeley Symp. Math. Stat. and Prob., University of California Press **1** (1961), 547-561.

[4] J. H. Havrda and F. Charvat, *Quantification method classification process: concept of structural α-entropy*, Kybernetika **3** (1967), 30-35.

[5] J. N. Kapur, *A comparative assessment of various measures of directed divergence*, Advances in Management Studies **3** (1984), 1-16.

[6] B. D. Sharma and D. P. Mittal, *New non-additive measures of relative information*, Journ. Comb. Inf. Sys. Sci. **2** (4) (1977), 122-132.

[7] I. Burbea and C.R. Rao, *On the convexity of some divergence measures based on entropy function*, IEEE Trans. Inf. Th. **28** (3) (1982), 489-495.

[8] C. R. Rao, *Diversity and dissimilarity coefficients: a unified approach*, Theoretic Population Biology **21** (1982), 24-43.

[9] J. Lin, *Divergence measures based on the Shannon entropy*, IEEE Trans. Inf. Th. **37** (1) (1991), 145-151.

[10] I. Csiszár, *Information-type measures of difference of probability distributions and indirect observations*, Studia Math. Hungarica **2** (1967), 299-318.

[11] I. Csiszár, *On topological properties of $f-$divergences*, Studia Math. Hungarica **2** (1967), 329-339.

[12] S. M. Ali and S. D. Silvey, *A general class of coefficients of divergence of one distribution from another*, J. Roy. Statist. Soc. Sec B **28** (1966), 131-142.

[13] I. Vajda, *Theory of Statistical Inference and Information*, Dordrecht-Boston, Kluwer Academic Publishers, 1989.

[14] M. Mei, *The theory of genetic distance and evaluation of human races*, Japan J. Human Genetics **23** (1978), 341-369.

[15] A. Sen, *On Economic Inequality*, Oxford University Press, London 1973.

[16] H. Theil, *Economics and Information Theory*, North-Holland, Amsterdam, 1967.

[17] H. Theil, *Statistical Decomposition Analysis*, North-Holland, Amsterdam, 1972.

[18] E. C. Pielou, *Ecological Diversity,* Wiley, New York, 1975.

[19] D. V. Gokhale and S. Kullback, *Information in Contingency Tables*, New York, Marcel Dekker, 1978.

[20] C. K. Chow and C. N. Lin, *Approximating discrete probability distributions with dependence trees*, IEEE Trans. Inf. Th. **14** (3) (1968), 462-467.

[21] D. Kazakos and T. Cotsidas, *A decision theory approach to the approximation of discrete probability densities*, IEEE Trans. Perform. Anal. Machine Intell. **1** (1980), 61-67.

[22] T. T. Kadota and L. A. Shepp, *On the best finite set of linear observables for discriminating two Gaussian signals*, IEEE Trans. Inf. Th. **13** (1967), 288-294.

[23] T. Kailath, *The divergence and Bhattacharyya distance measures in signal selection*, IEEE Trans. Comm. Technology. Vol COM-15 (1967), 52-60.

[24] M. Beth Bassat, *f-entropies, probability of error and feature selection*, Inform. Control **39** (1978), 227-242.

[25] C. H. Chen, *Statistical Pattern Recognition*, Rocelle Park, New York, Hoyderc Book Co., 1973.

[26] V. A. Volkonski and J. A. Rozanov, *Some limit theorems for random function -I*, (English Trans.), Theory Prob. Appl. (USSR) **4** (1959), 178-197.

[27] M. S. Pinsker, *Information and information stability of random variables and processes*, (in Russian), Moscow: Izv. Akad. Nouk, 1960.

[28] I. Csiszár, *A note on Jensen's inequality*, Studia Sci. Math. Hung. **1** (1966), 185-188.

[29] H. P. McKean, JR., Speed of approach to equilibrium for Koc's caricature of a Maximilian gas, Arch. Ration. Mech. Anal. **21** (1966), 343-367.

[30] J. H. B. Kemperman, *On the optimum note of transmitting information*, Ann. Math. Statist. **40** (1969), 2158-2177.

[31] S. Kullback, *A lower bound for discrimination information in terms of variation*, IEEE Trans. Inf. Th. **13** (1967), 126-127.

[32] S. Kullback, *Correction to a lower bound for discrimination information in terms of variation*, IEEE Trans. Inf. Th. **16** (1970), 771-773.

[33] I. Vajda, *Note on discrimination information and variation*, IEEE Trans. Inf. Th. **16** (1970), 771-773.

[34] G. T. Toussaint, *Sharper lower bounds for discrimination in terms of variation*, IEEE Trans. Inf. Th. **21** (1975), 99-100.

[35] F. Topsoe, *Some inequalities for information divergence and related measures of discrimination*, Res. Rep. Coll., RGMIA **2** (1) (1999), 85-98.

[36] L. Lecam, *Asymptotic Methods in Statistical Decision Theory*, Springer, New York, 1986.

[37] D. Dacunha-Castelle, *Ecole d'ete de Probability de Saint-Flour*, III-1977, Springer, Berlin–Heidelberg, 1978.

[38] C. Kraft, *Some conditions for consistency and uniform consistency of statistical procedures*, Univ. of California Pub. in Statistics **1** (1955), 125-142.

[39] S. Kullback and R. A. Leibler, *On information and sufficiency*, Annals Math. Statist. **22** (1951), 79-86.

[40] E. Hellinger, *Neue Bergrüirdung du Theorie quadratisher Formerus von uneudlichvieleu Veränderlicher*, J. für reine and Augeur. Math. **36** (1909), 210-271.

[41] A. Bhattacharyya, *On a measure of divergence between two statistical populations defined by their probability distributions*, Bull. Calcutta Math. Soc. **35** (1943), 99-109.

[42] I. J. Taneja, *Generalised Information Measures and their Applications* (http://www.mtm.ufsc.br/~taneja/bhtml/bhtml.html).
[43] I. Csiszár and J. Körner, *Information Theory: Coding Theorem for Discrete Memoryless Systems*, Academic Press, New York, 1981.
[44] J. Lin and S. K. M. Wong, *A new directed divergence measure and its characterization*, Int. J. General Systems **17** (1990), 73-81.
[45] H. Shioya and T. Da-Te, *A generalisation of Lin divergence and the derivative of a new information divergence*, Elec. and Comm. in Japan **78** (7) (1995), 37-40.
[46] J. Pečarić, F. Proschan and Y. L. Tong, *Convex Functions, Partial Orderings and Statistical Applications*, Academic Press Inc., 1992.

School of Computer Science & Mathematics
Victoria University of Technology
Melbourne City MC
Victoria 8001, Australia.

E-mail address: sever@matilda.vu.edu.au
http://rgmia.vu.edu.au/SSDragomirWeb.html