J. KSIAM **Vol.7**, No.1, 63-77, 2003

SOME INEQUALITIES FOR THE CSISZÁR Φ-DIVERGENCE

S.S. DRAGOMIR

ABSTRACT. Some inequalities for the Csiszár Φ-divergence and applications for the Kullback-Leibler, Rényi, Hellinger and Bhattacharyya distances in Information Theory are given.

1. INTRODUCTION

Given a convex function $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$, the Φ-*divergence functional*

$$I_\Phi(p,q) := \sum_{i=1}^{n} q_i \Phi\left(\frac{p_i}{q_i}\right) \tag{1.1}$$

was introduced in Csiszár [3], [4] as a generalized measure of information, a "distance function" on the set of probability distributions $\mathbb{P}^n$. The restriction here to discrete distribution is only for convenience, similar results hold for general distributions.

As in Csiszár [4], we interpret undefined expressions by

$$\Phi(0) = \lim_{t \to 0+} \Phi(t), \quad 0\Phi\left(\frac{0}{0}\right) = 0,$$
$$0\Phi\left(\frac{a}{0}\right) = \lim_{\varepsilon \to 0+} \Phi\left(\frac{a}{\varepsilon}\right) = a \lim_{t \to \infty} \frac{\Phi(t)}{t}, \quad a > 0.$$

The following results were essentially given by Csiszár and Körner [5].

Theorem 1. *If $\Phi : \mathbb{R}_+ \to \mathbb{R}$ is convex, then $I_\Phi(p,q)$ is jointly convex in p and q.*

The following lower bound for the Φ−divergence functional also holds.

1991 *Mathematics Subject Classification.* Primary 94Xxx; Secondary 26D15.

Key words and phrases. The Csiszár Φ-Divergence, Kullback-Leibler distance, R ényi α−entropy, Hellinger discrimination, Bhattacharyya distance.

Theorem 2. *Let $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ be convex. Then for every $p, q \in \mathbb{R}_+^n$, we have the inequality:*

$$I_\Phi(p,q) \geq \sum_{i=1}^{n} q_i \Phi\left(\frac{\sum_{i=1}^{n} p_i}{\sum_{i=1}^{n} q_i}\right). \tag{1.2}$$

If Φ is strictly convex, equality holds in (1.2) iff

$$\frac{p_1}{q_1} = \frac{p_2}{q_2} = \dots = \frac{p_n}{q_n}. \tag{1.3}$$

Corollary 1. *Let $\Phi : \mathbb{R}_+ \to \mathbb{R}$ be convex and normalized, i.e.,*

$$\Phi(1) = 0. \tag{1.4}$$

Then for any $p, q \in \mathbb{R}_+^n$ with

$$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i, \tag{1.5}$$

we have the inequality

$$I_\Phi(p,q) \geq 0. \tag{1.6}$$

If Φ is strictly convex, equality holds in (1.6) iff $p_i = q_i$ for all $i \in \{1, ..., n\}$.

In particular, if p, q are probability vectors, then (1.5) is assured. Corollary 1 then shows, for strictly convex and normalized $\Phi : \mathbb{R}_+ \to \mathbb{R}$, that

$$I_\Phi(p,q) \geq 0 \text{ for all } p, q \in \mathbb{P}^n. \tag{1.7}$$

The equality holds in (1.7) iff $p = q$.

These are "distance properties". However, I_Φ is not a metric: It violates the triangle inequality, and is **asymmetric**, i.e, for general $p, q \in \mathbb{R}_+^n$, $I_\Phi(p,q) \neq I_\Phi(q,p)$.

In the examples below we obtain, for suitable choices of the kernel Φ, some of the best known distance functions I_Φ used in mathematical statistics [15]-[17], information theory [2]-[22] and signal processing [13], [20].

Example 1. ***(Kullback-Leibler)*** *For*

$$\Phi(t) := t \log t, \; t > 0; \tag{1.8}$$

the $\Phi-$divergence is

$$I_\Phi(p,q) = \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i}\right), \tag{1.9}$$

*the **Kullback-Leibler distance*** [18]-[19].

Example 2. ***(Hellinger)*** *Let*

$$\Phi(t) = \frac{1}{2}\left(1-\sqrt{t}\right)^2, \; t>0. \tag{1.10}$$

Then I_Φ *gives the* ***Hellinger distance*** [1]

$$I_\Phi(p,q) = \frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{p_i}-\sqrt{q_i}\right)^2, \tag{1.11}$$

which is symmetric.

Example 3. ***(Renyi)*** *For* $\alpha > 1$, *let*

$$\Phi(t) = t^\alpha, \; t>0. \tag{1.12}$$

Then

$$I_\Phi(p,q) = \sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha}, \tag{1.13}$$

which is the $\alpha-$***order*** *entropy* [21].

Example 4. *(*χ^2-***distance****)* *Let*

$$\Phi(t) = (t-1)^2, \; t>0. \tag{1.14}$$

Then

$$I_\Phi(p,q) = \sum_{i=1}^{n}\frac{(p_i-q_i)^2}{q_i} \tag{1.15}$$

is the χ^2-***distance*** *between* p *and* q.

Finally, we have

Example 5. ***(Variation distance)****. Let* $\Phi(t) = |t-1|, \; t>0$. *The corresponding divergence, called the* ***variation distance****, is symmetric,*

$$I_\Phi(p,q) = \sum_{i=1}^{n}|p_i-q_i|.$$

For other examples of divergence measures, see the paper [16] by J.N. Kapur, where further references are given.

2. Other Inequalities for the Csiszár Φ-Divergence

We start with the following result.

Theorem 3. *Let $\Phi : \mathbb{R}_+ \to \mathbb{R}$ be differentiable convex. Then for all $p, q \in \mathbb{R}_+^n$ we have the inequality*

$$\Phi'(1)(P_n - Q_n) \le I_\Phi(p,q) - Q_n \Phi(1) \le I_{\Phi'}\left(\frac{p^2}{q}, p\right) - I_{\Phi'}(p,q), \tag{2.1}$$

where $P_n := \sum_{i=1}^{n} p_i > 0$, $Q_n := \sum_{i=1}^{n} q_i > 0$ and $\Phi' : (0,\infty) \to \mathbb{R}$ is the derivative of Φ. If Φ is strictly convex and $p_i, q_i > 0$ $(i = 1, ..., n)$, then the equality holds in (2.1) iff $p = q$.

Proof. As Φ is differentiable convex on $\mathbb{R}_+$, then we have the inequality

$$\Phi'(y)(y-x) \ge \Phi(y) - \Phi(x) \ge \Phi'(x)(y-x) \tag{2.2}$$

for all $x, y \in \mathbb{R}_+$.

Choose in (2.2) $y = \frac{p_i}{q_i}$ and $x = 1$, to obtain

$$\Phi'\left(\frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i} - 1\right) \ge \Phi\left(\frac{p_i}{q_i}\right) - \Phi(1) \ge \Phi'(1)\left(\frac{p_i}{q_i} - 1\right) \tag{2.3}$$

for all $i \in \{1, ..., n\}$.

Now, if we multiply (2.3) by $q_i \ge 0$ $(i = 1, ..., n)$ and sum over i from 1 to n, we can deduce

$$\sum_{i=1}^{n} (p_i - q_i)\Phi'\left(\frac{p_i}{q_i}\right) \ge I_\Phi(p,q) - Q_n\Phi(1) \ge \Phi'(1)(P_n - Q_n)$$

and as

$$\sum_{i=1}^{n} (p_i - q_i)\Phi'\left(\frac{p_i}{q_i}\right) = I_{\Phi'}\left(\frac{p^2}{q}, p\right) - I_{\Phi'}(p,q),$$

the inequality in (2.1) is thus obtained.

The case of equality holds in (2.2) for a strictly convex mapping iff $x = y$ and so the equality holds in (2.1) iff $\frac{p_i}{q_i} = 1$ for all $i \in \{1, ..., n\}$, and the theorem is proved. ∎

Remark 1. *In the above theorem, if we would like to drop the differentiability condition, we can choose instead of $\Phi'(x)$ any number $l = l(x) \in [\Phi'_-(x), \Phi'_+(x)]$ and the inequality will still be valid. This follows by the fact that for the convex mapping $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ we have*

$$l_2(x)(x-y) \ge \Phi(x) - \Phi(y) \ge l_1(y)(x-y), \quad x, y \in (0,\infty);$$

where $l_1(y) \in [\Phi'_-(y), \Phi'_+(y)]$ and $l_2(x) \in [\Phi'_-(x), \Phi'_+(x)]$, where Φ'_- is the left and Φ'_+ is the right derivative of Φ respectively. We omit the details.

The following corollary is a natural consequence of the above theorem.

Corollary 2. *Let $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ be convex and normalized. If $\Phi'(1)(P_n - Q_n) \geq 0$, then we have the positivity inequality*

$$0 \leq I_\Phi(p,q) \leq I_{\Phi'}\left(\frac{p^2}{q}, p\right) - I_{\Phi'}(p,q). \tag{2.4}$$

The equality holds in (2.4) for a strictly convex mapping Φ iff $p = q$.

Remark 2. *Corollary 2 shows that the positivity inequality (1.6) holds for a larger class of $(p,q) \in \mathbb{R}^n_+$ than that one considered in Corollary 1, namely, for $(p,q) \in \left\{\mathbb{R}^n_+ \times \mathbb{R}^n_+ : P_n = Q_n\right\}$.*

We have the following theorem as well.

Theorem 4. *Assume that Φ is differentiable convex on $(0,\infty)$. If $p^{(j)}$, $q^{(j)}$ $(j = 1,2)$ are probability distributions, then for all $\lambda \in [0,1]$ we have the inequality*

$$\begin{aligned} 0 &\leq \lambda I_\Phi\left(p^{(1)}, q^{(1)}\right) + (1-\lambda) I_\Phi\left(p^{(2)}, q^{(2)}\right) \\ &\quad - I_\Phi\left(\lambda p^{(1)} + (1-\lambda) q^{(1)}, \lambda p^{(2)} + (1-\lambda) q^{(2)}\right) \\ &\leq \lambda(1-\lambda) \sum_{i=1}^{n} \frac{\begin{vmatrix} p_i^{(1)} & p_i^{(2)} \\ q_i^{(1)} & q_i^{(2)} \end{vmatrix}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}} \left[\Phi'\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right) - \Phi'\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right)\right], \end{aligned} \tag{2.5}$$

where Φ' is the derivative of Φ.

Proof. Using the inequality (2.2), we may state

$$\begin{aligned} &\Phi'\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}}\right)\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}} - \frac{p_i^{(1)}}{q_i^{(1)}}\right) \\ \geq\ &\Phi\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}}\right) - \Phi\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right) \\ \geq\ &\Phi'\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right)\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}} - \frac{p_i^{(1)}}{q_i^{(1)}}\right) \end{aligned} \tag{2.6}$$

and

$$\begin{aligned} &\Phi'\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}}\right)\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}} - \frac{p_i^{(2)}}{q_i^{(2)}}\right) \\ \geq\ &\Phi\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}}\right) - \Phi\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right) \\ \geq\ &\Phi'\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right)\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}} - \frac{p_i^{(2)}}{q_i^{(2)}}\right). \end{aligned} \tag{2.7}$$

Multiply (2.6) by $\lambda q_i^{(1)}$ and (2.7) by $(1-\lambda)\, q_i^{(2)}$ and add the obtained inequalities to get

$$\begin{aligned}
(2.8)\qquad & \sum_{i=1}^{n} \Phi'\left(\frac{\lambda p_i^{(1)} + (1-\lambda)\, p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}}\right) \left[\lambda q_i^{(1)} \left(\frac{\lambda p_i^{(1)} + (1-\lambda)\, p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} - \frac{p_i^{(1)}}{q_i^{(1)}}\right)\right. \\
& \left. + (1-\lambda)\, q_i^{(2)} \left(\frac{\lambda p_i^{(1)} + (1-\lambda)\, p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} - \frac{p_i^{(2)}}{q_i^{(2)}}\right)\right] \\
\geq\ & I_\Phi\left(\lambda p^{(1)} + (1-\lambda)\, p^{(2)}, \lambda q^{(1)} + (1-\lambda)\, q^{(2)}\right) \\
& - \lambda I_\Phi\left(p^{(1)}, q^{(1)}\right) - (1-\lambda)\, I_\Phi\left(p^{(2)}, q^{(2)}\right) \\
\geq\ & \sum_{i=1}^{n} \left[\lambda q_i^{(1)} \Phi'\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right) \left(\frac{\lambda p_i^{(1)} + (1-\lambda)\, p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} - \frac{p_i^{(1)}}{q_i^{(1)}}\right)\right. \\
& \left. + (1-\lambda)\, q_i^{(2)} \Phi'\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right) \left(\frac{\lambda p_i^{(1)} + (1-\lambda)\, p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} - \frac{p_i^{(2)}}{q_i^{(2)}}\right)\right].
\end{aligned}$$

However,

$$\begin{aligned}
& \lambda q_i^{(1)} \left(\frac{\lambda p_i^{(1)} + (1-\lambda)\, p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} - \frac{p_i^{(1)}}{q_i^{(1)}}\right) \\
& + (1-\lambda)\, q_i^{(2)} \left(\frac{\lambda p_i^{(1)} + (1-\lambda)\, p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} - \frac{p_i^{(2)}}{q_i^{(2)}}\right) \\
=\ & -\frac{\lambda(1-\lambda)\begin{vmatrix} p_i^{(1)} & p_i^{(2)} \\ q_i^{(1)} & q_i^{(2)} \end{vmatrix}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} + \frac{\lambda(1-\lambda)\begin{vmatrix} p_i^{(1)} & p_i^{(2)} \\ q_i^{(1)} & q_i^{(2)} \end{vmatrix}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} = 0,
\end{aligned}$$

which shows that the first part in (2.8) is zero.

In addition,

$$\lambda q_i^{(1)} \left(\frac{\lambda p_i^{(1)} + (1-\lambda)\, p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} - \frac{p_i^{(1)}}{q_i^{(1)}}\right) = -\frac{\lambda(1-\lambda)\begin{vmatrix} p_i^{(1)} & p_i^{(2)} \\ q_i^{(1)} & q_i^{(2)} \end{vmatrix}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}},$$

and

$$(1-\lambda)\, q_i^{(2)} \left(\frac{\lambda p_i^{(1)} + (1-\lambda)\, p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} - \frac{p_i^{(2)}}{q_i^{(2)}}\right) = -\frac{\lambda(1-\lambda)\begin{vmatrix} p_i^{(1)} & p_i^{(2)} \\ q_i^{(1)} & q_i^{(2)} \end{vmatrix}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}},$$

and then, the second part in (2.4) is

$$-\lambda(1-\lambda)\sum_{i=1}^{n}\frac{\begin{vmatrix} p_i^{(1)} & p_i^{(2)} \\ q_i^{(1)} & q_i^{(2)} \end{vmatrix}}{\lambda q_i^{(1)}+(1-\lambda)q_i^{(2)}}\left[\Phi'\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right)-\Phi'\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right)\right],$$

which proves the theorem. ∎

Remark 3. *The first inequality in (2.5) is actually the joint convexity property of* $I_\Phi(\cdot,\cdot)$ *which has been proven here in a different manner than in* [5].

3. Applications for Some Particular Φ−Divergences

Let us consider the Kullback-Leibler distance given by (1.9)

$$KL(p,q):=\sum_{i=1}^{n}p_i\log\left(\frac{p_i}{q_i}\right). \tag{3.1}$$

Consider the convex mapping $\Phi(t)=-\log t$, $t>0$. For this mapping we have the Csiszár Φ−divergence

$$\begin{aligned} I_\Phi(p,q) &= \sum_{i=1}^{n}q_i\left[-\log\left(\frac{p_i}{q_i}\right)\right] \\ &= \sum_{i=1}^{n}q_i\log\left(\frac{q_i}{p_i}\right)=KL(q,p). \end{aligned} \tag{3.2}$$

The following inequality holds.

Proposition 1. *Let* $p,q\in\mathbb{R}^n$*. Then we have the inequality*

$$Q_n-P_n\le KL(q,p)\le\sum_{i=1}^{n}\frac{q_i^2}{p_i}-Q_n. \tag{3.3}$$

The case of equality holds iff $p=q$.

Proof. Since $\Phi(t)=-\log t$, then $\Phi'(t)=-\frac{1}{t}$, $t>0$. We have

$$\begin{aligned} I_{\Phi'}\left(\frac{p^2}{q},p\right) &= \sum_{i=1}^{n}p_i\cdot\left[-\frac{1}{\left(\frac{p_i^2}{q_i}\right)\cdot\frac{1}{p_i}}\right]=-Q_n, \\ I_{\Phi'}(p,q) &= \sum_{i=1}^{n}q_i\cdot\left[-\frac{1}{\frac{p_i}{q_i}}\right]=-\sum_{i=1}^{n}\frac{q_i^2}{p_i}, \end{aligned}$$

and then, from (2.1), we get

$$-(P_n-Q_n)\le KL(q,p)\le -Q_n+\sum_{i=1}^{n}\frac{q_i^2}{p_i},$$

which is the desired inequality (3.3).

The case of equality is obvious taking into account that $-\log$ is a strictly convex mapping on $(0,\infty)$. ∎

The following result for the Kullback-Leibler distance also holds.

Proposition 2. *Let $p,q\in\mathbb{R}^n$. Then we have the inequality*

$$P_n - Q_n \le KL(p,q) \le P_n - Q_n + KL(q,p) - KL\left(p,\frac{p^2}{q}\right). \tag{3.4}$$

The case of equality holds iff $p=q$.

Proof. As $\Phi(t) = t\log(t)$, then $\Phi'(t) = \log t + 1$. We have

$$\begin{aligned} I_\Phi(p,q) &= KL(p,q), \\ I_{\Phi'}\left(\frac{p^2}{q},p\right) &= I_{\log(\cdot)+1}\left(\frac{p^2}{q},p\right) = P_n + I_{\log(\cdot)}\left(\frac{p^2}{q},p\right). \end{aligned}$$

As we know that $I_{-\log(\cdot)}(p,q) = KL(q,p)$ (see (3.2)), then we have that

$$I_{\log(\cdot)}\left(\frac{p^2}{q},p\right) = -KL\left(p,\frac{p^2}{q}\right).$$

In addition, we have

$$\begin{aligned} I_{\Phi'}(p,q) &= I_{\log(\cdot)+1}(p,q) = Q_n + I_{\log(\cdot)}(p,q) \\ &= Q_n - KL(q,p) \end{aligned}$$

and then, by (2.1), we can state that

$$P_n - Q_n \le KL(p,q) \le P_n - Q_n - KL\left(p,\frac{p^2}{q}\right)Q_n + KL(q,p)$$

and the inequality (3.4) is obtained.

The case of equality holds from the fact that the mapping $\Phi(t) = t\log t$ is strictly convex on $(0,\infty)$. ∎

Now, let us consider the *$\alpha-$order entropy* of Rényi (see (1.13))

$$D_\alpha(p,q) := \sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha}, \quad \alpha > 1 \tag{3.5}$$

and $p,q\in\mathbb{R}_+^n$.

We know that Rényi's entropy is actually the Csiszár $\Phi-$divergence for the convex mapping $\Phi(t) = t^\alpha$, $\alpha>1$, $t>0$ (see Example 3).

The following proposition holds.

Proposition 3. *Let $p,q \in \mathbb{R}_+^n$. Then we have the inequality*

$$\alpha\left(P_n - Q_n\right) \leq D_\alpha\left(p,q\right) - Q_n \leq \alpha\left[D_\alpha\left(p,q\right) - D_\alpha\left(q^{\frac{2-\alpha}{\alpha}}, p^{-1}\right)\right]. \tag{3.6}$$

The case of equality holds iff $p = q$.

Proof. Since $\Phi\left(t\right) = t^\alpha$, then $\Phi'\left(t\right) = \alpha t^{\alpha-1}$.

We have

$$\begin{aligned} I_{\Phi'}\left(\frac{p^2}{q}, p\right) &= \sum_{i=1}^{n} p_i\left[\alpha \cdot \left(\frac{p_i^2}{q_i p_i}\right)^{\alpha-1}\right] \\ &= \alpha\sum_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1} = \alpha\sum_{i=1}^{n} q_i^{1-\alpha}p_i^\alpha = \alpha D_\alpha\left(p,q\right) \end{aligned}$$

and

$$\begin{aligned} I_{\Phi'}\left(p,q\right) &= \sum_{i=1}^{n} q_i\left[\alpha \cdot \left(\frac{p_i}{q_i}\right)^{\alpha-1}\right] \\ &= \alpha\sum_{i=1}^{n} p_i^{\alpha-1}q_i^{2-\alpha} = \alpha D_\alpha\left(q^{\frac{2-\alpha}{\alpha}}, \frac{1}{p}\right). \end{aligned}$$

Using the inequality (2.1), we have

$$\alpha\left(P_n - Q_n\right) \leq D_\alpha\left(p,q\right) - Q_n \leq \alpha\left[D_\alpha\left(p,q\right) - D_\alpha\left(q^{\frac{2-\alpha}{\alpha}}, \frac{1}{p}\right)\right]$$

and the inequality (3.6) is proved.

The case of equality holds since the mapping $\Phi\left(t\right) = t^\alpha$ is strictly convex on $\left(0,\infty\right)$ for $\alpha > 1$. ∎

Consider now the *Hellinger discrimination (see for example* [16])

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

where $p,q \in \mathbb{R}_+^n$.

We know that Hellinger discrimination is actually the Csiszár Φ−divergence for the convex mapping $\Phi\left(t\right) = \frac{1}{2}\left(\sqrt{t} - 1\right)^2$.

We may state the following proposition.

Proposition 4. *Let $p,q \in \mathbb{R}_+^n$. Then we have the inequality*

$$0 \leq h^2\left(p,q\right) \leq \frac{1}{2}\left[P_n - Q_n\right] + \frac{1}{2}\left[\sum_{i=1}^{n} q_i\left(\sqrt{\frac{q_i}{p_i}} - \sqrt{\frac{p_i}{q_i}}\right)\right]. \tag{3.7}$$

The equality holds iff $p = q$.

Proof. As $\Phi(t) = \frac{1}{2}\left(\sqrt{t}-1\right)^2$, we have $\Phi'(t) = \frac{1}{2} - \frac{1}{2\sqrt{t}}$ and $\Phi''(t) = \frac{1}{4}\cdot\frac{1}{\sqrt{t^3}} > 0$ $(t \in (0,\infty))$ which shows that Φ is indeed strictly convex on $(0,\infty)$.

We also have:

$$\begin{aligned} I_{\Phi}(p,q) &= h^2(p,q), \\ I_{\Phi'}\left(\frac{p^2}{q},p\right) &= \sum_{i=1}^{n} p_i \left[\frac{1}{2} - \frac{1}{2\sqrt{\frac{p_i^2}{q_i p_i}}}\right] \\ &= \frac{1}{2}P_n - \frac{1}{2}\sum_{i=1}^{n}\sqrt{p_i q_i} = \frac{1}{2}\left[P_n - \sum_{i=1}^{n}\sqrt{p_i q_i}\right] \\ I_{\Phi'}(p,q) &= \sum_{i=1}^{n} q_i \left[\frac{1}{2} - \frac{1}{2\sqrt{\frac{p_i}{q_i}}}\right] = \frac{1}{2}\left[Q_n - \sum_{i=1}^{n} q_i\sqrt{\frac{q_i}{p_i}}\right] \end{aligned}$$

and as $\Phi'(1) = 0$ and $\Phi(1) = 0$, then, by (2.1) applied for Φ as above, we deduce (3.7). The case of equality is obvious by the strict convexity of Φ. ∎

Consider now the *Bhattacharyya distance* (see for example [16])

$$B(p,q) = \sum_{i=1}^{n}\sqrt{p_i q_i},$$

where $p,q \in \mathbb{R}_+^n$.

We know that for the convex mapping $f(t) = -\sqrt{t}$, we have

$$I_{\Phi}(p,q) = -\sum_{i=1}^{n} q_i\sqrt{\frac{p_i}{q_i}} = -B(p,q).$$

We may state the following proposition.

Proposition 5. *Let $p,q \in \mathbb{R}_+^n$. Then we have the inequality*

$$\frac{1}{2}(Q_n - P_n) \le Q_n - B(p,q) \le \frac{1}{2}\sum_{i=1}^{n} q_i\left(\sqrt{\frac{q_i}{p_i}} - \sqrt{\frac{p_i}{q_i}}\right). \tag{3.8}$$

The case of equality holds iff $p = q$.

Proof. As $\Phi(1) = -\sqrt{t}$, $t > 0$, then $\Phi'(t) = -\frac{1}{2\sqrt{t}}$ and $\Phi''(t) = \frac{1}{4\sqrt{t^3}}$, $t > 0$, which also shows that $\Phi(\cdot)$ is strictly convex on $(0, \infty)$. We also have

$$I_{\Phi'}\left(\frac{p^2}{q}, p\right) = \sum_{i=1}^{n} p_i \left[-\frac{1}{2\sqrt{\frac{p_i^2}{q_i p_i}}} \right] = -\frac{1}{2}\sum_{i=1}^{n} \sqrt{p_i q_i} = -\frac{1}{2} B(p, q),$$

$$I_{\Phi'}(p, q) = -\frac{1}{2}\sum_{i=1}^{n} q_i \frac{1}{\sqrt{\frac{p_i}{q_i}}} = -\frac{1}{2}\sum_{i=1}^{n} q_i \sqrt{\frac{q_i}{p_i}}$$

and as $\Phi'(1) = -\frac{1}{2}$, $\Phi(1) = -1$, then by (2.1) applied for the mapping Φ as defined above, we deduce (3.8).

The case of equality is obvious by the strict convexity of Φ. ∎

4. Further Bounds for the Case when $P_n = Q_n$

The following inequality of Grüss type is well known in the literature as the Biernacki, Pidek and Ryll-Nardzewski inequality (see for example [24]).

Lemma 1. *Let a_i, b_i $(i = 1, ..., n)$ be real numbers such that*

$$a \le a_i \le A, \ b \le b_i \le B \ \text{ for all } i \in \{1, ..., n\}. \tag{4.1}$$

Then we have the inequality:

$$\left| \frac{1}{n}\sum_{i=1}^{n} a_i b_i - \frac{1}{n^2}\sum_{i=1}^{n} a_i \sum_{i=1}^{n} b_i \right| \le \frac{1}{n}\left[\frac{n}{2}\right]\left(1 - \frac{1}{n}\left[\frac{n}{2}\right]\right)(A - a)(B - b), \tag{4.2}$$

where $[x]$ denotes the integer part of x.

The following inequality holds.

Theorem 5. *Let $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ be differentiable convex. If $p, q \in \mathbb{R}_+^n$ are such that $P_n = Q_n$ and*

$$m \le p_i - q_i \le M, \ \ i = 1, ..., n \tag{4.3}$$

$$0 < r \le \frac{p_i}{q_i} \le R < \infty, \ \ i = 1, ..., n, \tag{4.4}$$

then we have the inequality

$$0 \le I_\Phi(p, q) - Q_n \Phi(1) \le \left[\frac{n}{2}\right]\left(1 - \frac{1}{n}\left[\frac{n}{2}\right]\right)(M - m)\left(\Phi'(R) - \Phi'(r)\right). \tag{4.5}$$

Proof. From (2.1) we have

$$0 \le I_\Phi(p, q) - Q_n \Phi(1) \le \sum_{i=1}^{n} (p_i - q_i)\, \Phi'\left(\frac{p_i}{q_i}\right). \tag{4.6}$$

Applying (4.2) we have

$$\left|\frac{1}{n}\sum_{i=1}^{n}(p_i-q_i)\,\Phi'\left(\frac{p_i}{q_i}\right)-\frac{1}{n^2}\sum_{i=1}^{n}(p_i-q_i)\sum_{i=1}^{n}\Phi'\left(\frac{p_i}{q_i}\right)\right| \leq \frac{1}{n}\left[\frac{n}{2}\right]\left(1-\frac{1}{n}\left[\frac{n}{2}\right]\right)(M-m)\left(\Phi'(R)-\Phi'(r)\right) \tag{4.7}$$

as the mapping Φ' is monotonic nondecreasing, and then

$$\Phi'(r)\leq\Phi'\left(\frac{p_i}{q_i}\right)\leq\Phi'(R)\quad\text{for all } i\in\{1,...,n\}.$$

As $\sum_{i=1}^{n}(p_i-q_i)=0$, we deduce by (4.6) and (4.7) the desired result (4.5). ∎

The following inequalities for particular distances are valid.

(1) If $p,q\in\mathbb{R}_n^+$ are such that the conditions (4.3) and (4.4) hold, then we have the inequalities

$$0\leq KL(q,p)\leq\left[\frac{n}{2}\right]\left(1-\frac{1}{n}\left[\frac{n}{2}\right]\right)(M-m)\frac{R-r}{rR}, \tag{4.8}$$

and

$$0\leq KL(p,q)\leq\left[\frac{n}{2}\right]\left(1-\frac{1}{n}\left[\frac{n}{2}\right]\right)(M-m)\left[\log\left(\frac{R}{r}\right)\right]. \tag{4.9}$$

(2) If p,q are as in (4.3) and (4.4), we have the inequality ($\alpha\geq 1$)

$$0\leq D_\alpha(p,q)-Q_n\leq\alpha\left[\frac{n}{2}\right]\left(1-\frac{1}{n}\left[\frac{n}{2}\right]\right)(M-m)\left(R^{\alpha-1}-r^{\alpha-1}\right). \tag{4.10}$$

(3) If p,q are as in (4.3) and (4.4), we have the inequality

$$0\leq h^2(p,q)\leq\frac{1}{2}\left[\frac{n}{2}\right]\left(1-\frac{1}{n}\cdot\left[\frac{n}{2}\right]\right)(M-m)\frac{\sqrt{R}-\sqrt{r}}{\sqrt{rR}}. \tag{4.11}$$

(4) Under the above assumptions for p and q, we have

$$0\leq Q_n-B(p,q)\leq\frac{1}{2}\left[\frac{n}{2}\right]\left(1-\frac{1}{n}\left[\frac{n}{2}\right]\right)(M-m)\frac{\sqrt{R}-\sqrt{r}}{\sqrt{rR}}. \tag{4.12}$$

Using the following Grüss' weighted inequality.

Lemma 2. *Assume that a_i, b_i $(i=1,...,n)$ are as in Lemma 1. If $q_i\geq 0$, $\sum_{i=1}^{n}q_i=1$, then we have the inequality*

$$\left|\sum_{i=1}^{n}q_ia_ib_i-\sum_{i=1}^{n}q_ia_i\sum_{i=1}^{n}q_ib_i\right|\leq\frac{1}{4}(A-a)(B-b). \tag{4.13}$$

We may prove the following converse inequality as well.

Theorem 6. *Let $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ be differentiable convex. If $p, q \in \mathbb{R}_+^n$ are such that $P_n = Q_n$ and*

$$0 < r \leq \frac{p_i}{q_i} \leq R < \infty, \quad i = 1, ..., n, \tag{4.14}$$

then we have the inequality

$$0 \leq I_\Phi(p,q) - Q_n \Phi(1) \leq \frac{1}{4}(R-r)\left[\Phi'(R) - \Phi'(r)\right]. \tag{4.15}$$

Proof. From (2.1) we have

$$\begin{aligned} 0 \ &\leq \ I_\Phi(p,q) - Q_n\Phi(1) \leq \sum_{i=1}^n (p_i - q_i)\Phi'\left(\frac{p_i}{q_i}\right) \\ &= \ \sum_{i=1}^n q_i\left(\frac{p_i}{q_i} - 1\right)\Phi'\left(\frac{p_i}{q_i}\right). \end{aligned} \tag{4.16}$$

As $\Phi'(\cdot)$ is monotonic nondecreasing, then

$$\Phi'(r) \leq \Phi'\left(\frac{p_i}{q_i}\right) \leq \Phi'(R) \quad \text{for all } i \in \{1, ..., n\}.$$

Applying (4.13) for $a_i = \frac{p_i}{q_i} - 1$, $b_i = \Phi'\left(\frac{p_i}{q_i}\right)$, we obtain

$$\begin{aligned} &\left|\sum_{i=1}^n q_i\left(\frac{p_i}{q_i} - 1\right)\Phi'\left(\frac{p_i}{q_i}\right) - \sum_{i=1}^n q_i\left(\frac{p_i}{q_i} - 1\right)\sum_{i=1}^n q_i\Phi'\left(\frac{p_i}{q_i}\right)\right| \\ &\leq \ \frac{1}{4}(R-r)\left[\Phi'(R) - \Phi'(r)\right] \end{aligned} \tag{4.17}$$

and as

$$\sum_{i=1}^n q_i\left(\frac{p_i}{q_i} - 1\right) = 0,$$

then, by (4.16) and (4.17) we deduce (4.15). ∎

The following inequalities for particular distances are valid.

(1) If p, q are such that $P_n = Q_n$ and (4.14) holds, then

$$0 \leq KL(q,p) \leq \frac{(R-r)^2}{4rR} \tag{4.18}$$

and

$$0 \leq KL(q,p) \leq \frac{1}{4}(R-r)^2 \ln\left(\frac{R}{r}\right). \tag{4.19}$$

(2) With the same assumptions for p, q, we have

$$0 \leq D_{\alpha}(p,q) - Q_n \leq \frac{\alpha}{4}(R-r)\left(R^{\alpha-1} - r^{\alpha-1}\right) \quad (\alpha \geq 1); \tag{4.20}$$

$$0 \leq h^2(p,q) \leq \frac{1}{8}(R-r)\frac{\sqrt{R}-\sqrt{r}}{\sqrt{Rr}} \tag{4.21}$$

and

$$0 \leq Q_n - B(p,q) \leq \frac{1}{8}(R-r)\frac{\sqrt{R}-\sqrt{r}}{\sqrt{Rr}}. \tag{4.22}$$

Remark 4. *Any other Grüss type inequality can be used to provide different bounds for the difference*

$$\Delta := \sum_{i=1}^{n}(p_i - q_i)\Phi'\left(\frac{p_i}{q_i}\right).$$

We omit the details.

References

[1] R. BERAN, Minimum Hellinger distance estimates for parametric models, *Ann. Statist.*, 5 (1977), 445-463.

[2] I. BURBEA and C.R. RAO, On the convexity of some divergence measures based on entropy functions, *IEEE Transactions on Information Theory,* **28** (1982), 489-495.

[3] I. CSISZÁR, Information measures: A critical survey, *Trans. 7th Prague Conf. on Info. Th., Statist. Decis. Funct., Random Processes and 8th European Meeting of Statist.*, Volume B, Academia Prague, 1978, 73-86.

[4] I. CSISZÁR, Information-type measures of difference of probability distributions and indirect observations, *Studia Sci. Math. Hungar,* **2** (1967), 299-318.

[5] I. CSISZÁR and J. KÖRNER, *Information Theory: Coding Theorems for Discrete Memory-less Systems,* Academic Press, New York, 1981.

[6] S.S. DRAGOMIR, Some inequalities for the Csiszár Φ-divergence, Preprint, *RGMIA Monographs, Victoria University, 2001,* ON LINE: http://rgmia.vu.edu.au/monographs/csiszar_list.html

[7] S.S. DRAGOMIR and C.J. GOH, A counterpart of Jensen's discrete inequality for differentiable convex mappings and applications in information theory, *Math. Comput. Modelling,* **24** (2) (1996), 1-11.

[8] S.S. DRAGOMIR and C.J. GOH, Some counterpart inequalities in for a functional associated with Jensen's inequality, *J. of Ineq. & Appl.,* **1** (1997), 311-325.

[9] S.S. DRAGOMIR and C.J. GOH, Some bounds on entropy measures in information theory, *Appl. Math. Lett.,* **10** (1997), 23-28.

[10] S.S. DRAGOMIR and C.J. GOH, A counterpart of Jensen's continuous inequality and applications in information theory, *RGMIA Preprint,* `http://matilda.vu.edu.au/~rgmia/InfTheory/Continuse.dvi`

[11] S.S. DRAGOMIR and N.M. IONESCU, Some converse of Jensen's inequality and applications, *Anal. Num. Theor. Approx.,* **23** (1994), 71-78.

[12] S.S. DRAGOMIR, J. ŠUNDE and M. SCHOLZ, Some upper bounds for relative entropy and applications, *Comp. & Math. with Applications,* **39** (2000), 91-100.

[13] B.R. FRIEDEN, Image enhancement and restoration, *Picture Processing and Digital Filtering* (T.S. Huang, Editor), Springer-Verlag, Berlin, 1975.

[14] R.G. GALLAGER, *Information Theory and Reliable Communications,* J. Wiley, New York, 1968.

[15] J.H. JUSTICE (editor), *Maximum Entropy and Bayssian Methods in Applied Statistics,* Cambridge University Press, Cambridge, 1986.

[16] J.N. KAPUR, A comparative assessment of various measures of directed divergence, *Advances in Management Studies,* **3** (1984), No. 1, 1-16.

[17] J.N. KAPUR, On the roles of maximum entropy and minimum discrimination information principles in Statistics, *Technical Address of the 38th Annual Conference of the Indian Society of Agricultural Statistics,* 1984, 1-44.

[18] S. KULLBACK, *Information Theory and Statistics,* J. Wiley, New York, 1959.

[19] S. KULLBACK and R.A. LEIBLER, On information and sufficiency, *Annals Math. Statist.,* 22 (1951), 79-86.

[20] R.M. LEAHY and C.E. GOUTIS, An optimal technique for constraint-based image restoration and mensuration, *IEEE Trans. on Acoustics, Speech and Signal Processing,* **34** (1986), 1692-1642.

[21] A. RENYI, On measures of entropy and information, *Proc. Fourth Berkeley Symp. Math. Statist. Prob., Vol. 1, University of California Press, Berkeley, 1961.*

[22] C.E. SHANNON, A mathematical theory of communication, *Bull. Sept. Tech. J.,* **27** *(1948), 370-423 and 623-656.*

[23] M. MATIĆ, Jensen's inequality and applications in Information Theory (in Croatian), Ph.D. Thesis, Univ. of Zagreb, 1999.

[24] D.S. MITRINOVIĆ, J.E. PEČARIĆ and A.M. FINK, *Classical and New Inequalities in Analysis,* Kluwer Academic Publishers, 1993.

School of Computer Science & Mathematics
Victoria University of Technology
PO Box 14428
Melbourne City MC 8001
Victoria, Australia
email : sever.dragomir@vu.edu.au
urladdrhttp://rgmia.vu.edu.au/SSDragomirWeb.html