A CONVERSE INEQUALITY FOR THE CSISZÁR $\Phi-$DIVERGENCE

S.S. DRAGOMIR

Abstract. A converse inequality for the Csiszár $\Phi-$divergence and applications for some particular distance functions such as: Kullback-Leibler distance, Renyi distance, χ^2-distance, Bhattacharyya distance, J-distance, etc. 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 [1], [2] 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 [2], we interpret undefined expressions by

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

The following results were essentially given by Csiszár and Körner [3].

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 $\Phi-$divergence functional also holds.

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

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

Date: September 7, 1999.

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

Key words and phrases. The Csiszár $\Phi-$Divergence, Kullback-Leibler distance, Rényi $\alpha-$entropy.

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, the 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 $\Phi : \mathbb{R}_+ \to \mathbb{R}$,

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

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 [4]-[5], information theory [6]-[8] and signal processing [9]-[10].

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** [11]-[12].*

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** [13]*

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

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 [14].*

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. *(**Variational distance**). Let $\Phi(t) = |t-1|, \ t > 0$. The corresponding divergence, called the **variational distance**, is symmetric,*

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

For other examples of divergence measures, see the paper [22] by J.N. Kapur, where further references are given.

2. Some Converse Inequalities for the Csiszár $\Phi-$Divergence

We start with the following converse of Jensen's discrete inequality established in 1994 in [15] by Dragomir and Ionescu.

Lemma 1. *Let $f : I \subseteq \mathbb{R} \to \mathbb{R}$ be a differentiable convex mapping on I. If $x_i \in \mathring{I}$, $t_i \geq 0$ with $T_n := \sum_{i=1}^{n} t_i > 0 \ \ (i = 1, ..., n)$, then we have the inequality*

$$\begin{aligned} 0 &\leq \frac{1}{T_n}\sum_{i=1}^{n} t_i f(x_i) - f\left(\frac{1}{T_n}\sum_{i=1}^{n} t_i x_i\right) \\ &\leq \frac{1}{T_n}\sum_{i=1}^{n} t_i x_i f'(x_i) - \frac{1}{T_n}\sum_{i=1}^{n} t_i x_i \cdot \frac{1}{T_n}\sum_{i=1}^{n} t_i f'(x_i); \end{aligned} \tag{2.1}$$

where $f' : \mathring{I} \to \mathbb{R}$ is the derivative of f on $\mathring{I}$.
If $t_i > 0 \ (i = 1, ..., n)$ and the mapping f is strictly convex on $\mathring{I}$, then the case of equality holds in (2.1) iff $x_1 = x_2 = ... = x_n$.

Proof. As $f : I \to \mathbb{R}$ is convex on $\mathring{I}$, then we have the following inequality,

$$f(x) - f(y) \geq f'(y)(x - y) \text{ for all } x, y \in \mathring{I}. \tag{2.2}$$

Choose in (2.2) $x = \frac{1}{T_n}\sum_{i=1}^{n} t_i x_i \in \mathring{I}$, $y = x_j, \ j = 1, ..., n$ to obtain

$$f\left(\frac{1}{T_n}\sum_{i=1}^{n} t_i x_i\right) - f(x_j) \geq f'(x_j)\left(\frac{1}{T_n}\sum_{i=1}^{n} t_i x_i - x_j\right) \tag{2.3}$$

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

If we multiply (2.3) by $t_j \geq 0$ and sum over j from 0 to n, we deduce

$$\sum_{j=1}^{n} t_j \left[f\left(\frac{1}{T_n} \sum_{i=1}^{n} t_i x_i \right) - f(x_j) \right] \geq \sum_{j=1}^{n} t_j f'(x_j) \left(\frac{1}{T_n} \sum_{i=1}^{n} t_i x_i - x_j \right). \tag{2.4}$$

However, a simple calculation shows that

$$\sum_{j=1}^{n} t_j \left[f\left(\frac{1}{T_n} \sum_{i=1}^{n} t_i x_i \right) - f(x_j) \right] = T_n f\left(\frac{1}{T_n} \sum_{i=1}^{n} t_i x_i \right) - \sum_{i=1}^{n} t_i f(x_i)$$

and

$$\sum_{j=1}^{n} t_j f'(x_j) \left(\frac{1}{T_n} \sum_{i=1}^{n} t_i x_i - x_j \right) = \frac{1}{T_n} \sum_{i=1}^{n} t_i x_i \sum_{i=1}^{n} t_i f'(x_i) - \sum_{i=1}^{n} t_i x_i f'(x_i)$$

and then, by (2.4), we can state

$$T_n f\left(\frac{1}{T_n} \sum_{i=1}^{n} t_i x_i \right) - \sum_{i=1}^{n} t_i f(x_i) \geq \frac{1}{T_n} \sum_{i=1}^{n} t_i x_i \sum_{i=1}^{n} t_i f'(x_i) - \sum_{i=1}^{n} t_i x_i f'(x_i).$$

Dividing by $T_n > 0$, we obtain the second inequality in (2.1).

If f is strictly convex, then the equality holds in (2.2) iff $x = y$. Using this and an obvious argument, we obtain the fact that the equality holds in (2.1) iff $x_1 = ... = x_n$.

We omit the details. ∎

Remark 1. *If, in the above lemma we drop the differentiability condition, then can choose instead of $f'(x_i)$ any number $l = l(x_i) \in [f'_-(x_i), f'_+(x_i)]$ and the inequality still remains valid. This follows by the fact that for a convex mapping $f : \mathbb{R}_+ \to \mathbb{R}_+$, we have*

$$f(x) - f(y) \geq l(y)(x - y) \text{ for } x, y \in \mathbb{R}_+,$$

where $l(y) \in [f'_-(y), f'_+(y)]$ and $f'_-(y)$ is the left derivative of f in y and $f'_+(y)$ is the right derivative of f in y. We omit the details.

Remark 2. *For an extension of this theorem for convex mappings of several variables see the paper* [16] *by Dragomir and Goh where further applications in Information Theory for Shannon's Entropy, Mutual Information, Conditional Entropy, etc. are given. An integral version of this result can be found in the paper* [19] *by Dragomir and Goh where further applications for the continuous case of the Shannon Entropy have been given. Extensions of the above results for convex mappings defined on convex sets in linear spaces and particularly in normed spaces can be found in the recent Ph.D. Dissertation* [20] *of M. Matić where other applications in Information Theory for Shannon's Entropy have been considered.*

In the recent paper [23], we proved the following theorem.

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*

$$\Phi'(1)(P_n - Q_n) \leq I_\Phi(p,q) - Q_n \Phi(1) \leq I_{\Phi'}\left(\frac{p^2}{q}, q\right) - I_{\Phi'}(p,q), \tag{2.5}$$

where $P_n := \sum_{i=1}^{n} p_i > 0$, $Q_n := \sum_{i=1}^{n} q_i > 0$ and Φ' is the derivative of Φ. If Φ is strictly convex and p_i, $q_i > 0$ $(i = 1, ..., n)$, then the equality holds in (2.5) iff $p = q$.

The following converse inequality for the Csiszár $\Phi-$divergence also holds.

Theorem 4. *Let Φ, p and q be as in Theorem 2. Then we have the inequality:*

$$\begin{aligned} 0 &\leq I_\Phi(p,q) - Q_n \Phi\left(\frac{P_n}{Q_n}\right) \\ &\leq I_{\Phi'}\left(\frac{p^2}{q}, p\right) - \frac{P_n}{Q_n} I_{\Phi'}(p,q), \end{aligned} \tag{2.6}$$

where $Q_n := \sum_{i=1}^{n} q_i > 0$, $P_n := \sum_{i=1}^{n} p_i > 0$ and Φ' is the derivative of Φ. If Φ is strictly convex, and p_i, $q_i > 0$, $(i = 1, ..., n)$, then the equality holds in (2.6) iff $\frac{p_1}{q_1} = ... = \frac{p_n}{q_n}$.

Proof. Choose in Lemma 1 $f = \Phi$, $t_i := q_i$ and $x_i = \frac{p_i}{q_i}$ to obtain

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

which is equivalent to (2.6). The case of equality is obvious. ∎

The following corollary is natural.

Corollary 2. *Let $\Phi : \mathbb{R}_+ \to \mathbb{R}$ be differentiable convex and normalized. Then, for any $p, q \in \mathbb{R}_+^n$ with $P_n = Q_n$, we have the converse of the positivity inequality (1.6)*

$$0 \leq I_\Phi(p,q) \leq I_{\Phi'}\left(\frac{p^2}{q}, p\right) - I_{\Phi'}(p,q). \tag{2.7}$$

The equality holds in (2.7) for a strictly convex mapping Φ iff $p = q$.

3. Applications for some Particular $\Phi-$Divergences

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

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

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

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

The following inequality holds.

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

$$0 \le KL(q,p) - Q_n \log\left(\frac{Q_n}{P_n}\right) \le \frac{1}{Q_n}\left[P_n \sum_{i=1}^{n} \frac{q_i^2}{p_i} - Q_n^2\right] = \frac{1}{2Q_n}\sum_{i,j=1}^{n} p_i p_j \left(\frac{q_i}{p_i}-\frac{q_j}{p_j}\right)^2. \tag{3.2}$$

The case of equality holds iff $\frac{p_1}{q_1}=\dots=\frac{p_n}{q_n}$.

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

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

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

and then from (2.6), we can state that

$$0 \le KL(q,p) + Q_n\log\left(\frac{P_n}{Q_n}\right) \le -Q_n + \frac{P_n}{Q_n}\sum_{i=1}^{n}\frac{q_i^2}{p_i} = \frac{1}{Q_n}\left[P_n\sum_{i=1}^{n}\frac{q_i^2}{p_i}-Q_n^2\right],$$

i.e.,

$$0 \le KL(q,p) - Q_n\log\left(\frac{Q_n}{P_n}\right) \le \frac{1}{Q_n}\left[P_n\sum_{i=1}^{n}\frac{q_i^2}{p_i}-Q_n^2\right].$$

On the other hand, we observe that

$$\begin{aligned}\frac{1}{2}\sum_{i,j=1}^{n} p_i p_j\left(\frac{q_i}{p_i}-\frac{q_j}{p_j}\right)^2 &= \frac{1}{2}\sum_{i,j=1}^{n} p_i p_j\left(\frac{q_i^2}{p_i^2}-2\frac{q_i q_j}{p_i p_j}+\frac{q_j^2}{p_j^2}\right)\\ &= \frac{1}{2}\left[\sum_{i,j=1}^{n} p_j\frac{q_i^2}{p_i} - 2\sum_{i,j=1}^{n} q_i q_j + \sum_{i,j=1}^{n} p_i\frac{q_j^2}{p_j}\right]\\ &= P_n\sum_{i=1}^{n}\frac{q_i^2}{p_i}-Q_n^2\end{aligned}$$

and the inequality (3.2) is proved.

The case of equality follows by the strict convexity of the mapping $-\log$. ∎

Corollary 3. *If we assume that $P_n=Q_n$ in (3.2), we have*

$$0\le KL(q,p)\le\sum_{i=1}^{n}\left[\frac{q_i^2-p_i^2}{p_i}\right]=\frac{1}{2Q_n}\sum_{i,j=1}^{n}p_i p_j\left(\frac{q_i}{p_i}-\frac{q_j}{p_j}\right)^2 \tag{3.3}$$

with equality iff $p=q$.

Remark 3. *We know that the* χ^2-***distance*** *between* p *and* q *is (see (1.15))*

$$D_{\chi^2}(p,q) := \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i} = \left(\sum_{i=1}^{n} \frac{p_i^2 - q_i^2}{q_i} \text{ if } P_n = Q_n \right).$$

As

$$D_{\chi^2}(q,p) = \sum_{i=1}^{n} \frac{(q_i - p_i)^2}{p_i} = \left(\sum_{i=1}^{n} \frac{q_i^2 - p_i^2}{p_i} \text{ if } P_n = Q_n \right),$$

then the inequality (3.3) can be rewritten in the following manner

$$0 \le KL(q,p) \le D_{\chi^2}(q,p). \tag{3.4}$$

Corollary 4. *Let* p *and* q *be two probability distributions. Then*

$$0 \le KL(q,p) \le D_{\chi^2}(q,p) \tag{3.5}$$

with equality iff $p = q$.

Remark 4. *For a direct proof of the inequality (3.5) see* [21] *where further bounds are also given.*

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

$$D_\alpha(p,q) := \sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha},\ \alpha > 1 \tag{3.6}$$

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

We know that Renyi'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 2. *Let* $p, q \in \mathbb{R}_+^n$. *Then we have the inequality*

$$\begin{aligned} 0 &\le D_\alpha(p,q) - P_n^\alpha Q_n^{1-\alpha} \\ &\le \frac{\alpha}{Q_n}\left[Q_n D_\alpha(p,q) - P_n D_\alpha\left(q^{\frac{2-\alpha}{\alpha}}, \frac{1}{p} \right) \right]. \end{aligned} \tag{3.7}$$

The case of equality holds iff $\frac{p_1}{q_1} = \ldots = \frac{p_n}{q_n}$.

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

We have

$$\begin{aligned} I_{\Phi'}\left(\frac{p^2}{q}, p \right) &= \sum_{i=1}^{n} p_i \cdot \left[\alpha \left(\frac{p_i^2}{q_i p_i} \right)^{\alpha-1} \right] \\ &= \alpha \sum_{i=1}^{n} p_i \left(\frac{p_i}{q_i} \right)^{\alpha-1} = \alpha \sum_{i=1}^{n} q_i^{1-\alpha} p_i^\alpha \\ &= \alpha D_\alpha(p,q) \end{aligned} \tag{3.8}$$

and

$$I_{\Phi'}(p,q) = \sum_{i=1}^{n} q_i \left[\alpha \left(\frac{p_i}{q_i} \right)^{\alpha-1} \right] = \sum_{i=1}^{n} p_i^{\alpha-1} q_i^{2-\alpha} = \alpha D_\alpha\left(q^{\frac{2-\alpha}{\alpha}}, \frac{1}{p} \right). \tag{3.9}$$

Using the inequality (2.6), we can state that

$$\begin{aligned} 0 &\leq D_\alpha(p,q) - Q_n\left(\frac{P_n}{Q_n}\right)^\alpha \\ &\leq \alpha D_\alpha(p,q) - \alpha\frac{P_n}{Q_n}D_\alpha\left(q^{\frac{2-\alpha}{\alpha}}, \frac{1}{p}\right) \\ &= \frac{\alpha}{Q_n}\left[Q_n D_\alpha(p,q) - P_n D_\alpha\left(q^{\frac{2-\alpha}{\alpha}}, \frac{1}{p}\right)\right] \end{aligned}$$

and the inequality (3.7) is obtained.

The case if equality follows by the fact that the mapping $\Phi(t) = t^\alpha$ $(\alpha > 1,\ t > 0)$ is strictly convex on $(0,\infty)$. ∎

Corollary 5. *If we assume that $P_n = Q_n$ in (3.7), we obtain*

$$0 \leq D_\alpha(p,q) - P_n \leq \alpha\left[D_\alpha(p,q) - D_\alpha\left(q^{\frac{2-\alpha}{\alpha}}, \frac{1}{p}\right)\right] \tag{3.10}$$

with equality iff $p = q$.
In particular, if p, q are probability distributions, then

$$0 \leq D_\alpha(p,q) - 1 \leq \alpha\left[D_\alpha(p,q) - D_\alpha\left(q^{\frac{2-\alpha}{\alpha}}, \frac{1}{p}\right)\right] \tag{3.11}$$

with equality iff $p = q$.

Now, let us consider the Bhattacharyya distance given by (see [20])

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

If we consider the convex mapping $\Phi : (0,\infty) \to \mathbb{R}$, $\Phi(t) = -\sqrt{t}$, then

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

The following inequality holds.

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

$$0 \leq \sqrt{Q_n P_n} - B(p,q) \leq \frac{1}{2}\left[\frac{P_n}{Q_n}\sum_{i=1}^n q_i\sqrt{\frac{q_i}{p_i}} - B(p,q)\right]. \tag{3.12}$$

The case of equality holds iff $\frac{p_1}{q_1} = \dots = \frac{p_n}{q_n}$.

Proof. If $\Phi(t) = -\sqrt{t}$, then $\Phi'(t) = -\frac{1}{2}\cdot\frac{1}{\sqrt{t}}$, $t > 0$. We have

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

and then, from (2.6), we can state that

$$0 \leq -B(p,q) + Q_n \sqrt{\frac{P_n}{Q_n}} \leq \frac{1}{2} B(p,q) + \frac{1}{2} \frac{P_n}{Q_n} \sum_{i=1}^{n} q_i \sqrt{\frac{q_i}{p_i}},$$

which is equivalent to (3.12).

The case of equality follows by the strict convexity of the mapping Φ. ∎

Remark 5. *The second inequality in (3.12) is equivalent to*

$$\sqrt{P_n Q_n} \leq \frac{1}{2} \left[\frac{P_n}{Q_n} \sum_{i=1}^{n} q_i \sqrt{\frac{q_i}{p_i}} + B(p,q) \right] \tag{3.13}$$

with equality iff $\frac{p_1}{q_1} = ... = \frac{p_n}{q_n}$.

Corollary 6. *If we assume that* $P_n = Q_n$ *in (3.12), then we obtain*

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

with equality iff $p = q$.

Another important divergence measure in Information Theory is the *J-divergence* defined by (see for example [2])

$$J(p,q) = \sum_{i=1}^{n} (p_i - q_i) \log \left(\frac{p_i}{q_i} \right).$$

Let us observe that the mapping $\Phi(t) := (t-1)\ln t$, $t \in (0,\infty)$, has the derivatives

$$\begin{aligned} \Phi'(t) &= \ln t - \frac{1}{t} + 1, \ t > 0 \\ \Phi''(t) &= \frac{t+1}{t^2} > 0 \ \text{ for } \ t > 0, \end{aligned}$$

which shows that Φ is convex on $(0,1)$ and

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

We can state the following proposition.

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

$$\begin{aligned} 0 &\leq J(p,q) - (P_n - Q_n) \ln \left(\frac{P_n}{Q_n} \right) \\ &\leq KL(p,q) + \frac{P_n}{Q_n} KL(q,p) + \frac{P_n}{Q_n} \sum_{i=1}^{n} \frac{q_i^2}{p_i} - Q_n. \end{aligned} \tag{3.14}$$

The case of equality holds iff $\frac{p_1}{q_1} = ... = \frac{p_n}{q_n}$.

Proof. We have

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

and then, from (2.6), we can state that

$$\begin{aligned} 0 &\le J(p,q)-Q_n\left(\frac{P_n}{Q_n}-1\right)\ln\left(\frac{P_n}{Q_n}\right) \\ &\le KL(p,q)-Q_n+P_n-\frac{P_n}{Q_n}\left[-KL(q,p)-\sum_{i=1}^{n}\frac{q_i^2}{p_i}+Q_n\right] \\ &= KL(p,q)+P_n+\frac{P_n}{Q_n}\sum_{i=1}^{n}\frac{q_i^2}{p_i}-P_n-Q_n \\ &= KL(p,q)+\frac{P_n}{Q_n}KL(q,p)+\frac{P_n}{Q_n}\sum_{i=1}^{n}\frac{q_i^2}{p_i}-Q_n \end{aligned}$$

and the inequality (3.14) is obtained. ∎

4. Further Bounds via the Discrete Grüss Inequality

The following inequality is well known in the literature as the weighted discrete Grüss inequality.

Lemma 2. *Let $a_i, b_i, p_i \in \mathbb{R}$ $(i=1,...,n)$ be such that $p_i \ge 0$ with $P_n := \sum_{i=1}^{n} p_i > 0$ and*

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

Then we have the inequality

$$\left|\frac{1}{P_n}\sum_{i=1}^{n}p_ia_ib_i-\frac{1}{P_n}\sum_{i=1}^{n}p_ia_i\cdot\frac{1}{P_n}\sum_{i=1}^{n}p_ib_i\right| \le \frac{1}{4}(A-a)(B-b) \tag{4.2}$$

and the constant $\frac{1}{4}$ is the best possible one in the sense that it cannot be replaced by a smaller constant independent of n.

Using this result, we can point out the following counterpart of Jensen's inequality.

Lemma 3. *Let $f,\ x_i,\ t_i$ be as in Lemma 1. If there exist the real constants $m, M \in \mathring{I}$ such that $m \leq x_i \leq M$ for all $i \in \{1, ..., n\}$, then we have the inequality:*

$$0 \leq \frac{1}{T_n} \sum_{i=1}^{n} t_i f(x_i) - f\left(\frac{1}{T_n} \sum_{i=1}^{n} t_i x_i\right) \leq \frac{1}{4}(M-m)(f'(M) - f'(m)), \tag{4.3}$$

where $f' : \mathring{I} \to \mathbb{R}$ is the derivative of f.

Proof. Define in Grüss' inequality $a_i := x_i$ and $b_i := f'(x_i)$. As f is convex, f' is monotonic nondecreasing and therefore $f'(m) \leq b_i \leq f'(M)$. Applying the Grüss inequality, we can state that

$$\begin{aligned} & \frac{1}{T_n} \sum_{i=1}^{n} t_i x_i f'(x_i) - \frac{1}{T_n} \sum_{i=1}^{n} t_i x_i \cdot \frac{1}{T_n} \sum_{i=1}^{n} t_i f'(x_i) \\ \leq & \frac{1}{4}(M-m)(f'(M) - f'(m)) \end{aligned}$$

and by the inequality (2.1) we deduce (4.3). ∎

Remark 6. *Similar results can be obtained by using other Grüss type inequalities. However, we omit the details.*

The following converse inequality for the Csiszár $\Phi-$divergence holds.

Theorem 5. *Let Φ, p and q be as in Theorem 2. If there exist the real numbers r, R such that $0 < r \leq \frac{p_i}{q_i} \leq R < \infty$ for all $i \in \{1, ..., n\}$, then we have the inequality*

$$0 \leq I_\Phi(p, q) - Q_n \Phi\left(\frac{P_n}{Q_n}\right) \leq \frac{1}{4}(R-r)(\Phi'(R) - \Phi'(r)) Q_n. \tag{4.4}$$

Proof. Apply Lemma 3 for $f = \Phi,\ t_i = q_i$ and $x_i = \frac{p_i}{q_i}$ $(i = 1, ..., n)$. ∎

The following particular inequalities can be noted as well:

$$\begin{aligned} 0 & \leq KL(p, q) - P_n \log\left(\frac{P_n}{Q_n}\right) \leq \frac{P_n}{4}(R-r)[\log(R) - \log(r)] \\ & \leq \frac{P_n}{4} \cdot \frac{(R-r)^2}{\sqrt{Rr}}. \end{aligned} \tag{4.5}$$

Indeed, if we choose $\Phi(t) = t \log t$ in (4.4), we obtain the Kullback-Leibler divergence. The last inequality in (4.5) follows by the well known inequality between the *geometric mean* $G(a, b) = \sqrt{ab}$ and the *logarithmic mean* $L(a, b) := \frac{b-a}{\log b - \log a}$ $(a, b > 0,\ a \neq b)$ which says that

$$L(a, b) \geq G(a, b) \quad \text{for all } a, b > 0,\ a \neq b. \tag{4.6}$$

In addition, if in (4.4) we put $\Phi(t) = -\log t$, we deduce

$$0 \leq KL(q, p) - Q_n \log\left(\frac{Q_n}{P_n}\right) \leq \frac{Q_n}{4} \cdot \frac{(R-r)^2}{Rr}, \tag{4.7}$$

provided that $0 < r \leq \frac{p_i}{q_i} \leq R < \infty$ for all $i \in \{1, ..., n\}$.

Now, if in (4.4) we choose $\Phi(t) = t^\alpha$ $(\alpha > 1)$, $t > 0$, then we get the following inequality for the $\alpha-$order Renyi entropy

$$0 \leq D_\alpha(p, q) - P_n^\alpha Q_n^{1-\alpha} \leq \frac{\alpha}{4} \cdot Q_n (R-r)\left(R^{\alpha-1} - r^{\alpha-1}\right), \tag{4.8}$$

provided that $0 < r \leq \frac{p_i}{q_i} \leq R < \infty$ for all $i \in \{1, ..., n\}$.

If we apply Theorem 5 for the Bhattacharyya distance, we get the inequality

$$0 \leq \sqrt{P_n Q_n} - B(p,q) \leq \frac{1}{8} \cdot \frac{(R-r)\left(\sqrt{R}-\sqrt{r}\right)}{\sqrt{rR}} \cdot Q_n, \tag{4.9}$$

provided that $0 < r \leq \frac{p_i}{q_i} \leq R < \infty$ for all $i \in \{1, ..., n\}$.

Finally, if we apply Theorem 5 for the J-divergence, we may obtain the inequality

$$0 \leq J(p,q) - (P_n - Q_n) \ln \frac{P_n}{Q_n} \leq \frac{1}{4}(R-r)\left(\ln \frac{R}{r} + \frac{R-r}{rR}\right) Q_n, \tag{4.10}$$

provided that $0 < r \leq \frac{p_i}{q_i} \leq R < \infty$ for all $i \in \{1, ..., n\}$.

References

[1] 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.

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

[3] I. CSISZÁR and J. KÖRNER, *Information Theory: Coding Theorem for Discrete Memoryless Systems,* Academic Press, New York, 1981.

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

[5] 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.

[6] 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.

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

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

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

[10] 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.

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

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

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

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

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

[16] 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.

[17] 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.

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

[19] 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`

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

[21] S.S. DRAGOMIR, J. ŠUNDE and M. SCHOLZ, Some upper bounds for relative entropy and applications, *Comp. and Math. with Appl. (* in press).

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

[23] S.S. DRAGOMIR, Some inequalities for the Csiszár Φ-divergence, *submitted.*

School of Commuincations and Informatics, Victoria University of Technology, PO Box 14428, Melbourne City MC 8001, Victoria, Australia

E-mail address: sever@matilda.vu.edu.au

URL: http://rgmia.vu.edu.au/SSDragomirWeb.html