Received 04/01/14

Some Inequalities for f-Divergences Via Slater's Inequality for Convex Functions

S. S. Dragomir[1,2]

Abstract. Some inequalities for f-divergence measures by the use of Slater's inequality for convex functions of a real varaible are established.

1. Introduction

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

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

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

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

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

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

The following lower bound for the f-divergence functional also holds.

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

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

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

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

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

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

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

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

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

If f 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 $f : \mathbb{R}_+ \to \mathbb{R}$,

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

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

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

In the examples below we obtain, for suitable choices of the kernel f, some of the best known distance functions D_f used in mathematical statistics [**15**], information theory [**2**]-[**24**] and signal processing [**13**], [**19**].

Example 1. *(Kullback-Leibler) For*

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

the f-divergence is

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

the Kullback-Leibler distance [**17**]-[**18**].

Example 2. *(Hellinger) Let*

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

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

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

which is symmetric.

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

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

Then

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

which is the α-order entropy [**23**].

EXAMPLE 4. *(χ^2-distance) Let*

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

Then

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

is the χ^2-distance between p and q.

Finally, we have:

EXAMPLE 5. *(Variational distance). Let $f(t) = |t-1|, \ t > 0$. The corresponding divergence, called the variational distance, is symmetric,*

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

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

2. Slater Type Inequalities

Suppose that I is an interval of real numbers with interior $\mathring{I}$ and $f : I \to \mathbb{R}$ is a convex function on I. Then f is continuous on $\mathring{I}$ and has finite left and right derivatives at each point of $\mathring{I}$. Moreover, if $x, y \in \mathring{I}$ and $x < y$, then

$$D^- f(x) \le D^+ f(x) \le D^- f(y) \le D^+(y),$$

which shows that both $D^- f$ and $D^+ f$ are nondecreasing functions on $\mathring{I}$. It is also known that a convex function must be differentiable except for at most countably many points.

For a convex function $f : I \to \mathbb{R}$, the subdifferential of f denoted by ∂f is the set of all functions $\varphi : I \to [-\infty, \infty]$ such that $\varphi(\mathring{I}) \subseteq \mathbb{R}$ and

$$f(x) \ge f(a) + (x-a)\varphi(a) \quad \text{for any } x, a \in I. \tag{2.1}$$

It is also well known that if f is convex on I, then ∂f is nonempty, $D^+ f, \ D^- f \in \partial f$ and if $\varphi \in \partial f$, then

$$D^- f(x) \le \varphi(x) \le D^+ f(x) \tag{2.2}$$

for every $x \in \mathring{I}$. In particular, φ is a nondecreasing function. If f is differentiable convex on $\mathring{I}$, then $\partial f = \{f'\}$.

The following result is well known in literature as *Slater's inequality*. For the original proof due to Slater, see [**25**]. For related results, see Chapter I of the book [**21**] or Chapter 2 of the book [**22**].

We shall here follow the presentation in [**6**, pp. 129-130] where a slightly more general result for Slater's inequality is provided:

LEMMA 1. *Let $f : I \to \mathbb{R}$ be a nondecreasing (nonincreasing) convex function on I, $x_i \in I$, $p_i \ge 0$ with $P_n = \sum_{i=1}^{n} p_i > 0$ and for a given $\varphi \in \partial f$ assume that $\sum_{i=1}^{n} p_i \varphi(x_i) \ne 0$. Then one has the inequality*

$$\frac{1}{P_n} \sum_{i=1}^{n} p_i f(x_i) \le f\left(\frac{\sum_{i=1}^{n} p_i x_i \varphi(x_i)}{\sum_{i=1}^{n} p_i \varphi(x_i)} \right). \tag{2.3}$$

Proof. Let us give the proof for the case of nondecreasing functions only.

In this case $\varphi(x) \geq 0$ for any $x \in I$ and

$$\frac{\sum_{i=1}^{n} p_i x_i \varphi(x_i)}{\sum_{i=1}^{n} p_i \varphi(x_i)} \in I$$

being a convex combination of $x_i \in I$ with the nonnegative weights

$$\frac{p_i \varphi(x_i)}{\sum_{i=1}^{n} p_i \varphi(x_i)},\ i \in \{1, ..., n\}.$$

Now, if we use the inequality (2.2) we deduce

$$f(x) - f(x_i) \geq (x - x_i)\varphi(x_i) \quad \text{for any} \quad x, x_i \in I\ ,\ i \in \{1, ..., n\}. \tag{2.4}$$

Multiplying (2.4) by $p_i/P_n \geq 0$ and summing over $i \in \{1, ..., n\}$, we deduce

$$f(x) - \frac{1}{P_n}\sum_{i=1}^{n} p_i f(x_i) \geq x \cdot \frac{1}{P_n}\sum_{i=1}^{n} p_i \varphi(x_i) - \frac{1}{P_n}\sum_{i=1}^{n} p_i x_i \varphi(x_i) \tag{2.5}$$

for any $x \in I$. If in (2.5) we choose

$$x = \frac{\sum_{i=1}^{n} p_i x_i \varphi(x_i)}{\sum_{i=1}^{n} p_i \varphi(x_i)},$$

then we deduce the desired inequality (2.3). □

If we would like to drop the assumption of monotonicity for the function f, then we can state and prove in a similar way the following result (see also [**6**]):

Lemma 2. *Let* $f : I \to R$ *be a convex function,* $x_i \in I$, $p_i \geq 0$ *with* $P_n > 0$ *and* $\sum_{i=1}^{n} p_i \varphi(x_i) \neq 0$ *for a given* $\varphi \in \partial f$. *If*

$$\frac{\sum_{i=1}^{n} p_i x_i \varphi(x_i)}{\sum_{i=1}^{n} p_i \varphi(x_i)} \in I,$$

then the inequality (2.3) *holds true.*

Proof. Since

$$\frac{\sum_{i=1}^{n} p_i x_i \varphi(x_i)}{\sum_{i=1}^{n} p_i \varphi(x_i)} \in I,$$

hence we can use the inequality (2.4) and proceed as in the above Lemma 1. The details are omitted. □

The following inequality is well known in literature as Karamata's inequality, see [**21**, pp. 298] or [**22**, p. 212]:

Lemma 3. *Assume that* $0 < a \leq a_i \leq A < \infty$, $0 < b \leq b_i \leq B < \infty$ *for each* $i \in \{1, ..., n\}$. *Then for* $p_i > 0$, $\sum_{i=1}^{n} p_i = 1$, *one has the inequalities*

$$K^{-2} \leq \frac{\sum_{i=1}^{n} p_i a_i b_i}{\sum_{i=1}^{n} p_i a_i \sum_{i=1}^{n} p_i b_i} \leq K^2 \tag{2.6}$$

with $K = \frac{\sqrt{ab}+\sqrt{AB}}{\sqrt{aB}+\sqrt{bA}} > 1$.

Using Karamata's result, we may point out the following reverse of Jensen's inequality that may be useful in applications.

Lemma 4. *. Let $f:[0,\infty)\to\mathbb{R}$ be a monotonic nondecreasing convex function. Assume that $0<r\le x_i\le R<\infty$ for each $i\in\{1,...n\}$, $(p_i)_{i=1,...,n}$ is a probability distribution and for a given $\varphi\in\partial f$ consider*

$$K(r,R)=\frac{\sqrt{r\varphi(r)}+\sqrt{R\varphi(R)}}{\sqrt{r\varphi(R)}+\sqrt{R\varphi(r)}}.$$

Then we have the inequality

$$\sum_{i=1}^{n}p_i f(x_i)\le f\left(K^2(r,R)\sum_{i=1}^{n}p_i x_i\right). \tag{2.7}$$

Proof. From Lemma 3 we know that

$$\sum_{i=1}^{n}p_i f(x_i)\le f\left(\frac{\sum_{i=1}^{n}p_i x_i\varphi(x_i)}{\sum_{i=1}^{n}p_i\varphi(x_i)}\right). \tag{2.8}$$

If we apply Karamata's inequality for $a_i=x_i$, $b_i=\varphi(x_i)$, we get successively

$$\begin{aligned} f\left(\frac{\sum_{i=1}^{n}p_i x_i\varphi(x_i)}{\sum_{i=1}^{n}p_i\varphi(x_i)}\right) &= f\left(\frac{\sum_{i=1}^{n}p_i x_i\varphi(x_i)}{\sum_{i=1}^{n}p_i x_i\sum_{i=1}^{n}p_i\varphi(x_i)}\cdot\sum_{i=1}^{n}p_i x_i\right)\le \\ &\le f\left(K^2(r,R)\sum_{i=1}^{n}p_i x_i\right), \end{aligned}$$

since, obviously, $\varphi(x_i)\in[\varphi(r),\varphi(R)]$ being monotonic nondecreasing on $[r,R]$. The inequality (2.7) is thus proved. □

3. Some Inequalities for f-Divergences

The following result may be stated:

Theorem 3. *Let $f:[0,\infty]\to\mathbb{R}$ be a differentiable, convex and normalized function, i.e. $f(1)=0$ and $0\le r\le 1\le R\le\infty$. If there exists a real number m so that*

$$-\infty<m\le f'(x)\quad \text{for any}\quad x\in(r,R), \tag{3.1}$$

then for any probability distribution $p,q\in\mathcal{P}$ with

$$r\le\frac{p_i}{q_i}\le R\quad \text{for any}\quad i\in\{1,...,n\} \tag{3.2}$$

(if $r=0$ and $R=\infty$, the assumption (3.2) is always satisfied), one has the inequality

$$0\le D_f(p,q)\le f\left(\frac{D_{\Phi_*}(p,q)-m}{D_{f'}(p,q)-m}\right)-m\cdot\frac{D_{\Phi_\sharp}(p,q)}{D_{f'}(p,q)-m}, \tag{3.3}$$

where $\Phi_*(x):=xf'(x)$, $\Phi_\sharp(x):=(x-1)f'(x)$ for $x\in[0,\infty]$ and $D_{f'}(p,q)\ne m$.

Proof. Consider the auxiliary function $f_m(x)=f(x)-m(x-1)$, $x\in[0,\infty]$. Since $f_m'(x)=f'(x)-m,\ x\in(r,R)$, it follows that f_m is differentiable, convex and monotonic nondecreasing on (r,R) , and we may apply Lemma 1 to get

$$\sum_{i=1}^{n}q_i f_m\left(\frac{p_i}{q_i}\right)\le f_m\left(\frac{\sum_{i=1}^{n}q_i\frac{p_i}{q_i}f_m'\left(\frac{p_i}{q_i}\right)}{\sum_{i=1}^{n}q_i f_m'\left(\frac{p_i}{q_i}\right)}\right). \tag{3.4}$$

It is easy to see that

$$\sum_{i=1}^{n} q_i f_m\left(\frac{p_i}{q_i}\right) = \sum_{i=1}^{n} q_i \left[f\left(\frac{p_i}{q_i}\right) - m\left(\frac{p_i}{q_i} - 1\right)\right] = D_f(p,q)$$

and

$$\begin{aligned}\sum_{i=1}^{n} q_i \frac{p_i}{q_i} f_m'\left(\frac{p_i}{q_i}\right) &= \sum_{i=1}^{n} p_i \left[f'\left(\frac{p_i}{q_i}\right) - m\right] \\ &= \sum_{i=1}^{n} p_i f'\left(\frac{p_i}{q_i}\right) - m = D_{\Phi_*}(p,q) - m\end{aligned}$$

where $\Phi_*(x)$ has been defined above.

Also, one may observe that

$$\sum_{i=1}^{n} q_i f_m'\left(\frac{p_i}{q_i}\right) = \sum_{i=1}^{n} q_i \left[f'\left(\frac{p_i}{q_i}\right) - m\right] = D_{f'}(p,q) - m$$

and

$$\begin{aligned} f_m\left(\frac{D_{\Phi_*}(p,q)-m}{D_{f'}(p,q)-m}\right) &= f\left(\frac{D_{\Phi_*}(p,q)-m}{D_{f'}(p,q)-m}\right) - m\left(\frac{D_{\Phi_*}(p,q)-m}{D_{f'}(p,q)-m} - 1\right) \\ &= f\left(\frac{D_{\Phi_*}(p,q)-m}{D_{f'}(p,q)-m}\right) - m\cdot\frac{D_{\Phi_*}(p,q)-D_{f'}(p,q)}{D_{f'}(p,q)-m} \\ &= f\left(\frac{D_{\Phi_*}(p,q)-m}{D_{f'}(p,q)-m}\right) - m\cdot\frac{D_{\Phi_\sharp}(p,q)}{D_{f'}(p,q)-m}, \end{aligned}$$

which gives, by (3.4), the desired inequality (3.3). □

If one would like to drop the assumption of lower boundedness for the derivative f' (see (3.1)), one may need to impose another condition as described in the following theorem:

Theorem 4. *Let $f:[0,\infty)\to\mathbb{R}$ be a differentiable, convex and normalized function and $0\le r\le 1\le R\le\infty$. As above, consider $\Phi_*(x) = xf'(x)$ and assume that for two probabilities p and q satisfying (3.2) one has $D_{f'}(p,q)\neq 0$ and*

$$\frac{D_{\Phi_*}(p,q)}{D_{f'}(p,q)} \ge 0. \tag{3.5}$$

Then one has the inequality

$$0 \le D_f(p,q) \le f\left(\frac{D_{\Phi_*}(p,q)}{D_{f'}(p,q)}\right). \tag{3.6}$$

The proof follows in a similar way as the one in Theorem 3 by utilizing Lemma 2. We omit the details.

Now we can point out another result for f-divergences when bounds for the likelihood ratio $\frac{p}{q}$ are available:

Theorem 5. *Let $f:[0,\infty]\to\mathbb{R}$ be a differentiable convex and normalized function and $0\le r\le 1\le R\le\infty$ and let $K(r,R)$ be as stated in Lemma 4. If there exists a real number m so that*

$$-\infty < m \le f'(x) \quad \textit{for any} \quad x\in(r,R)$$

then for all probability distributions $p,q\in\mathcal{P}$ *satisfying*

$$r\le\frac{p_i}{q_i}\le R\ \ \text{for each}\ \ i\in\{1,...,n\}\ ,$$

one has the inequality

$$D_f(p,q)\le f(K^2(r,R))-m(K^2(r,R)-1).$$

PROOF. As in Theorem 3, the function $f_m(x)=f(x)-m(x-1)$ is differentiable, convex and monotonic nondecreasing on (r,R). If we apply Lemma 4 we get

$$\begin{aligned}\sum_{i=1}^n q_i f_m\left(\frac{p_i}{q_i}\right) &\le f_m\left(K^2(r,R)\cdot\sum_{i=1}^n q_i\frac{p_i}{q_i}\right)\\ &= f\left(K^2(r,R)\right)-m\left(K^2(r,R)-1\right),\end{aligned}$$

which completes the proof. □

4. Applications for Particular Divergences

We consider the Kullback-Leibler distance

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

that is the f-divergence for the convex function $f:(0,\infty)\to\mathbb{R},\ f\,(t)=t\log t$.

If we take the convex function $f\,(t)=-\log t$, then the corresponding f-divergence is

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

for all probability distributions $p,q\in\mathcal{P}$.

For the function $f\,(t)=-\log t$ we have

$$\Phi_*\,(t):=tf'\,(t)=-1\text{ and }\Phi_\sharp\,(t):=(t-1)\,f'\,(t)=\frac{1-t}{t},\ t>0.$$

Now for $0\le r\le 1\le R<\infty$ and $m=-\frac{1}{R}$ we have

$$m\le f'\,(t)=-\frac{1}{t}\ \text{for any}\ \ t\in(r,R)$$

and the condition (3.1) is satisfied.

We also have

$$D_{\Phi_*}\,(p,q)=-1\text{ and }D_{f'}\,(p,q)=-\sum_{i=1}^n\frac{q_i^2}{p_i}+1-1=-1-D_{\chi^2}\,(q,p)$$

where

$$D_{\chi^2}\,(p,q)=\sum_{i=1}^n\frac{(p_i-q_i)^2}{q_i}=\sum_{i=1}^n\frac{p_i^2}{q_i}-1$$

is the χ^2-distance between p and q.

We also have

$$D_{\Phi_\sharp}\,(p,q)=\sum_{i=1}^n q_i\frac{1-\frac{p_i}{q_i}}{\frac{p_i}{q_i}}=\sum_{i=1}^n\frac{q_i^2}{p_i}-1=D_{\chi^2}\,(q,p)\,.$$

Therefore, for any probability distribution $p, q \in \mathcal{P}$ with

$$r \leq \frac{p_i}{q_i} \leq R \quad \textit{for any} \quad i \in \{1, ..., n\}$$

we have by (3.3) the inequality

$$\begin{aligned} 0 \leq KL(q,p) &\leq -\ln\left(\frac{-1+\frac{1}{R}}{-1-D_{\chi^2}(q,p)+\frac{1}{R}}\right) \\ &+ \frac{1}{R}\cdot\frac{D_{\chi^2}(q,p)}{-1-D_{\chi^2}(q,p)+\frac{1}{R}}, \end{aligned}$$

which is equivalent to

$$0 \leq KL(q,p) \leq \ln\left(\frac{R\left(D_{\chi^2}(q,p)+1\right)-1}{R-1}\right) - \frac{D_{\chi^2}(q,p)}{R\left(D_{\chi^2}(q,p)+1\right)-1}. \tag{4.1}$$

Observe that

$$\frac{D_{\Phi_*}(p,q)}{D_{f'}(p,q)} = \frac{-1}{-1-D_{\chi^2}(q,p)} = \frac{1}{D_{\chi^2}(q,p)+1} > 0,$$

then by the inequality (3.6) we have

$$0 \leq KL(q,p) \leq \ln\left(D_{\chi^2}(q,p)+1\right) \tag{4.2}$$

for any $p, q \in \mathcal{P}$.

We notice that the inequality (4.2) can be obtained from (4.1) by letting $R \to \infty$.

Now, for the function $f(t) = t\log t$, we have

$$\Phi_*(t) := tf'(t) = t\log t + t.$$

Then

$$\begin{aligned} D_{\Phi_*}(p,q) &= \sum_{i=1}^n q_i\left(\frac{p_i}{q_i}\log\frac{p_i}{q_i}+\frac{p_i}{q_i}\right) = \sum_{i=1}^n p_i\log\frac{p_i}{q_i} + \sum_{i=1}^n p_i \\ &= KL(p,q)+1 \end{aligned}$$

and

$$D_{f'}(p,q) = \sum_{i=1}^n q_i\left(\log\frac{p_i}{q_i}+1\right) = 1 - KL(q,p).$$

Then, if we take $p, q \in \mathcal{P}$ with $1 > KL(q,p)$, by utilizing the inequality (3.6) we get

$$0 \leq KL(p,q) \leq \frac{1+KL(p,q)}{1-KL(q,p)}\ln\left(\frac{1+KL(p,q)}{1-KL(q,p)}\right). \tag{4.3}$$

For $\alpha > 1$ consider α-order entropy

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

which is an f-divergence for the convex function $f(t) = t^\alpha$.

We have

$$K(r,R) = \frac{\sqrt{rf'(r)}+\sqrt{Rf'(R)}}{\sqrt{rf'(R)}+\sqrt{Rf'(r)}} = \frac{r^{\frac{\alpha}{2}}+R^{\frac{\alpha}{2}}}{r^{\frac{1}{2}}R^{\frac{\alpha-1}{2}}+R^{\frac{1}{2}}r^{\frac{\alpha-1}{2}}}.$$

We have

$$f'(t) = \alpha t^{\alpha-1} \geq \alpha r^{\alpha-1}.$$

If we apply Theorem 5, then for all probability distributions $p, q \in \mathcal{P}$ satisfying

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

we have the inequality

$$D_\alpha(p,q) \leq \left(\frac{r^{\frac{\alpha}{2}} + R^{\frac{\alpha}{2}}}{r^{\frac{1}{2}} R^{\frac{\alpha-1}{2}} + R^{\frac{1}{2}} r^{\frac{\alpha-1}{2}}} \right)^{2\alpha} - \alpha r^{\alpha-1} \left[\left(\frac{r^{\frac{\alpha}{2}} + R^{\frac{\alpha}{2}}}{r^{\frac{1}{2}} R^{\frac{\alpha-1}{2}} + R^{\frac{1}{2}} r^{\frac{\alpha-1}{2}}} \right)^2 - 1 \right]. \tag{4.4}$$

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 functions and indirect observations, *Studia Sci. Math. Hungar,* **2** (1967), 299-318.

[5] I. CSISZÁR and J. KÖRNER, *Information Theory: Coding Theorem for Discrete Memoryless Systems,* Academic Press, New York, 1981. *Annual Conference of the Indian Society of Agricultural Statistics,* 1984, 1-44.

[6] S. S. DRAGOMIR, A survey on Cauchy-Buniakowski-Schwarz's type discrete inequality, *J. Ineq. Pure & Appl. Math.,* **4**(2003), Issue 3, Article 63, [ONLINE: http://jipam.vu.edu.au/article.php?sid=301].

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

[8] 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.

[9] 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.

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

[11] 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`

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

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

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

[19] 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.

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

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

[22] J. E. PEČARIĆ, F. PROSCHAN and Y. L. TONG, *Convex Functions, Partial Orderings and Statistical Applications*, Academic Press Inc., 1991.

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

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

[25] M. S. SLATER, A companion inequality to Jensen's inequality, *J. Approx. Theory,* **32** (1984), 160-166.

[1]Mathematics, College of Engineering & Science, Victoria University, PO Box 14428, Melbourne City, MC 8001, Australia.

E-mail address: `sever.dragomir@vu.edu.au`

URL: `http://rgmia.org/dragomir`

[2]School of Computational & Applied Mathematics, University of the Witwatersrand, Private Bag 3, Johannesburg 2050, South Africa