AN UPPER BOUND FOR THE CSISZÁR $f-$DIVERGENCE IN TERMS OF THE VARIATIONAL DISTANCE AND APPLICATIONS

S.S. DRAGOMIR

Abstract. In this paper we point out an upper bound for the Csiszár $f-$divergence of two discrete random variables in terms of the generalized $r-$variational distance. Some particular cases for Kullback-Leibler distance, Hellinger and triangular discrimination, χ^2-distance, etc.. are considered.

1. Introduction

Given a convex function $f:[0,\infty)\to\mathbb{R}$, the *$f-$divergence functional*

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

was introduced by Csiszár [1]-[2] as a generalized measure of information, a *"distance function"* on the set of probability distribution $\mathbb{P}^n$. The restriction here to discrete distributions is only for convenience, similar results hold for general distributions. As in Csiszár [1]-[2], we interpret undefined expressions by

$$f(0)=\lim_{t\to0+}f(t),\quad 0f\left(\tfrac{0}{0}\right)=0,$$
$$0f\left(\tfrac{a}{0}\right)=\lim_{\varepsilon\to0+}\varepsilon f\left(\tfrac{a}{\varepsilon}\right)=a\lim_{t\to\infty}\tfrac{f(t)}{t},\ a>0.$$

The following results (Theorems 1 and 2, and Corollary 1) were essentially given by Csiszár and Körner [3].

Theorem 1. *(Joint Convexity) If $f:[0,\infty)\to\mathbb{R}$ is convex, then $I_f(p,q)$ is jointly convex in p and q.*

Theorem 2. *(Jensen's inequality) Let $f:[0,\infty)\to\mathbb{R}$ be convex. Then for any $p,q\in\mathbb{R}_+^n$ with $P_n:=\sum_{i=1}^{n}p_i>0$, $Q_n:=\sum_{i=1}^{n}q_i>0$, we have the inequality*

$$I_f(p,q)\geq Q_n f\left(\frac{P_n}{Q_n}\right). \tag{1.2}$$

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

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

It is natural to consider the following corollary.

Date: December 13, 1999.

1991 *Mathematics Subject Classification.* Primary 94Xxx, 26D15.

Key words and phrases. Csiszár $f-$divergence, Variational Distance, Divergence Measures in Information Theory, Jensen's inequality.

Corollary 1. *(Nonnegativity) Let $f:[0,\infty)\to\mathbb{R}$ be convex and normalised, i.e.,*

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

Then for any $p,q\in\mathbb{R}_+^n$ with $P_n=Q_n$, we have the inequality

$$I_f(p,q)\geq 0. \tag{1.5}$$

If f is strictly convex, equality holds in (1.5) iff

$$p_i=q_i \quad \text{for all} \quad i\in\{1,...,n\}. \tag{1.6}$$

In particular, if p,q are probability vectors, then Corollary 1 shows that, for strictly convex and normalized $f:[0,\infty)\to\mathbb{R}$ that

$$I_f(p,q)\geq 0 \;\text{ and } I_f(p,q)=0 \;\text{ iff } p=q. \tag{1.7}$$

We now give some more examples of divergence measures in Information Theory which are particular cases of Csiszár $f-$divergences.

1. **Kullback-Leibler distance** ([12]). The *Kullback-Leibler distance* $D(\cdot,\cdot)$ is defined by

$$D(p,q):=\sum_{i=1}^{n}p_i\log\left(\frac{p_i}{q_i}\right). \tag{1.8}$$

If we choose $f(t)=t\ln t$, $t>0$, then obviously

$$I_f(p,q)=D(p,q). \tag{1.9}$$

2. **Variational distance** (l_1-distance). The *variational distance* $V(\cdot,\cdot)$ is defined by

$$V(p,q):=\sum_{i=1}^{n}|p_i-q_i|. \tag{1.10}$$

If we choose $f(t)=|t-1|$, $t\in\mathbb{R}_+$, then we have

$$I_f(p,q)=V(p,q). \tag{1.11}$$

3. **Hellinger discrimination** ([13]). The *Hellinger discrimination* $h^2(\cdot,\cdot)$ is defined by

$$h^2(p,q):=\frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{p_i}-\sqrt{q_i}\right)^2. \tag{1.12}$$

It is obvious that if $f(t)=\frac{1}{2}\left(\sqrt{t}-1\right)^2$, then

$$I_f(p,q)=h^2(p,q). \tag{1.13}$$

4. **Triangular discrimination** ([24]). We define *triangular discrimination* between p and q by

$$\Delta(p,q)=\sum_{i=1}^{n}\frac{|p_i-q_i|^2}{p_i+q_i}. \tag{1.14}$$

It is obvious that if $f(t)=\frac{(t-1)^2}{t+1}$, $t\in(0,\infty)$, then

$$I_f(p,q)=\Delta(p,q). \tag{1.15}$$

5. χ^2-**distance.** We define the χ^2-*distance* (chi-square distance) by

$$D_{\chi^2}(p,q) := \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i}. \tag{1.16}$$

It is clear that if $f(t) = (t-1)^2$, $t \in [0,\infty)$, then

$$I_f(p,q) = D_{\chi^2}(p,q) \tag{1.17}$$

6. **Rényi** $\alpha-$**order entropy** ([14]). The $\alpha-$*order entropy* $(\alpha > 1)$ is defined by

$$R_\alpha(p,q) := \sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha}. \tag{1.18}$$

It is obvious that if $f(t) = t^\alpha$ $(t \in (0,\infty))$, then

$$I_f(p,q) = R_\alpha(p,q). \tag{1.19}$$

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

2. An Upper Bound for Csiszár $f-$Divergence

In the recent paper [28], the author proved the following inequality for Csiszár $f-$divergence:

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 (2.9) iff* $p = q$,

If we assume that $P_n = Q_n$ and Φ is normalised, then we obtain the simpler

$$0 \le I_\Phi(p,q) \le I_{\Phi'}\left(\frac{p^2}{q}, p\right) - I_{\Phi'}(p,q). \tag{2.2}$$

Applications for particular divergences which are instances of Csiszár $f-$divergence were also given.

Asimilar result of the above theorem has been presented in antoher paper by the author [29].

Theorem 4. *Let* Φ, p, q *be as in Theorem 3. Then we have the inequality*

$$0 \le I_\Phi(p,q) - Q_n \Phi\left(\frac{P_n}{Q_n}\right) \le I_{\Phi'}\left(\frac{p^2}{q}, p\right) - \frac{P_n}{Q_n} I_{\Phi'}(p,q). \tag{2.3}$$

If Φ *is strictly convex and* p_i, $q_i > 0$ $(i = 1, ..., n)$, *then the equality holds in (2.3) iff* $\frac{p_1}{q_1} = ... = \frac{p_n}{q_n}$.

Obviously, if $P_n = Q_n$ and Φ is normalised, then (2.3) becomes (2.2).

The following result concerning an upper and a lower bound for the Csiszár $f-$divergence in terms of the Kullback-Leibler distance $D(p,q)$ holds.

As in [30], we will say that the mapping $f : C \subset \mathbb{R} \to \mathbb{R}$, where C is an interval (in [30], the definition was considered in general normed spaces), is

(i) $\alpha-$*lower convex on* C if $f - \frac{\alpha}{2} \cdot |\cdot|^2$ is convex on C;

(ii) $\beta-$*upper convex* on C if $\frac{\beta}{2} \cdot |\cdot|^2 - f$ is convex on C;

(iii) $(m,M)-$*convex on* C (with $m \le M$) if it is both $m-$lower convex and $M-$iupper convex.

In [30], amongst others, the author has proved the following result for Csiszár $f-$divergence.

Theorem 5. *Let* $\Phi : \mathbb{R}_+ \to \mathbb{R}$ *and* $p,q \in \mathbb{R}_+^n$ *with* $P_n = Q_n$.

(i) *If* Φ *is* α*-lower convex on* $\mathbb{R}_+$*, then we have the inequality*

$$\frac{\alpha}{2} \cdot D_{\chi^2}(p,q) \le I_\Phi(p,q) - Q_n \Phi(1). \tag{2.4}$$

(ii) *If* Φ *is* $\beta-$*upper convex on* $\mathbb{R}_+$*, then we have the inequality*

$$I_\Phi(p,q) - Q_n \Phi(1) \le \frac{\beta}{2} \cdot D_{\chi^2}(p,q). \tag{2.5}$$

(iii) *If* Φ *is* $(m,M)-$*convex on* $\mathbb{R}_+$*, then we have the following sandwich inequality*

$$\frac{m}{2} \cdot D_{\chi^2}(p,q) \le I_\Phi(p,q) - Q_n \Phi(1) \le \frac{M}{2} \cdot D_{\chi^2}(p,q), \tag{2.6}$$

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

Of course, if Φ is normalised, i.e., $\Phi(1) = 0$ and p,q are probability distributions, then we get the simpler inequalities:

$$\frac{\alpha}{2} \cdot D_{\chi^2}(p,q) \le I_\Phi(p,q), \quad I_\Phi(p,q) \le \frac{\beta}{2} \cdot D_{\chi^2}(p,q) \tag{2.7}$$

and

$$\frac{m}{2} \cdot D_{\chi^2}(p,q) \le I_\Phi(p,q) \le \frac{M}{2} \cdot D_{\chi^2}(p,q). \tag{2.8}$$

In [30], some applications for particular instances of Csiszár $f-$divergences were also given.

Define the generalised $r-$*variational distance* by

$$V_r(p,q) := \sum_{i=1}^n q_i^{1-r} |p_i - q_i|^r, \tag{2.9}$$

where p,q are probability distributions and $r \in (0,1]$. Note that for $r = 1$, we recapture the usual variational distance (or l_1-distance).

The following theorem holds.

Theorem 6. *Let* $f : [0,\infty) \to \mathbb{R}$ *be a convex mapping and such that* $f(1) = 0$*, i.e.,* f *is normalised, and* f *is of* $r - H-$*Hölder type on* $[r,R]$*, i.e.,*

$$|f(x) - f(y)| \le H |x-y|^r \quad \textit{for all} \quad x,y \in [r,R]. \tag{2.10}$$

Then we have the inequality

$$0 \le I_f(p,q) \le H V_r(p,q). \tag{2.11}$$

Proof. We choose in (2.10) $x = \frac{p_i}{q_i}$, $y = 1$ $(i = 1, ..., n)$ to get

$$\left| f\left(\frac{p_i}{q_i}\right) - f(1) \right| \le H \left| \frac{p_i}{q_i} - 1 \right|^r, \tag{2.12}$$

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

If we multiply (2.12) by q_i, sum the obtained inequalities and use the generalised triangle inequality, we obtain

$$\begin{aligned} 0 &\le I_f(p,q) = \sum_{i=1}^n q_i f\left(\frac{p_i}{q_i}\right) = \left| \sum_{i=1}^n q_i f\left(\frac{p_i}{q_i}\right) - \sum_{i=1}^n q_i f(1) \right| \\ &\le H \sum_{i=1}^n q_i \left| \frac{p_i}{q_i} - 1 \right|^r = H V_r(p,q) \end{aligned}$$

and the inequality (2.11) is proved. ∎

Remark 1. *If we assume that f is convex, normalised and $L-$Lipschitzian on $[r, R]$, i.e., $r = 1$ and $H = L$, then we have the inequality*

$$0 \le I_f(p,q) \le L V_r(p,q), \tag{2.13}$$

where $V(p,q)$ is the usual variational distance.

The practical result is embodied in the following corollary.

Corollary 2. *If the mapping $f : [r, R] \to \mathbb{R}$ is convex, normalised, absolutely continuous on $[r, R]$ and $f' \in L_\infty[r, R]$, i.e., $\|f'\|_\infty := ess \sup_{t \in [r,R]} |f'(t)| < \infty$, then we have the inequality*

$$0 \le I_f(p,q) \le \|f'\|_\infty V_r(p,q). \tag{2.14}$$

The following theorem holds.

Theorem 7. *Assume that f is as in Theorem 6. If $p^{(j)}, q^{(j)}$ $(j = 1, 2)$ are probability distributions satisfying the condition*

$$r \le \frac{p_i^{(j)}}{q_i^{(j)}} \le R \quad \text{for all} \quad i \in \{1, ..., n\} \quad \text{and} \quad j \in \{1, 2\},$$

then obviously

$$r \le \frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}} \le R \quad \text{for all } i \in \{1, ..., n\} \quad \text{and} \quad \lambda \in [0, 1]$$

and we have the inequality

$$\begin{aligned} 0 &\le I_f\left(\lambda p^{(1)} + (1-\lambda) p^{(2)}, \lambda q^{(1)} + (1-\lambda) q^{(2)}\right) \\ &\quad -\lambda I_f\left(p^{(1)}, q^{(1)}\right) - (1-\lambda) I_f\left(p^{(2)}, q^{(2)}\right) \\ &\le H \lambda^r (1-\lambda)^r \sum_{i=1}^n \frac{\left| \det \begin{bmatrix} p_i^{(1)} & p_i^{(2)} \\ q_i^{(1)} & q_i^{(2)} \end{bmatrix} \right|^r}{\left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^r} \\ &\quad \times \left[\lambda^{1-r} \left[q_i^{(1)}\right]^{1-r} + (1-\lambda)^{1-r} \left[q_i^{(2)}\right]^{1-r} \right] \end{aligned} \tag{2.15}$$

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

Proof. If we choose in (2.10)

$$x = \frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}} \text{ and } y = \frac{p_i^{(1)}}{q_i^{(1)}} \ (i = 1, ..., n)$$

we get

$$\begin{aligned} &\left| f\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}}\right) - f\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right)\right| \\ \leq\ & H\left|\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}} - \frac{p_i^{(1)}}{q_i^{(1)}}\right|^r \\ =\ & H\frac{\left|\lambda q_i^{(1)} p_i^{(1)} + (1-\lambda) q_i^{(1)} p_i^{(2)} - \lambda p_i^{(1)} q_i^{(1)} - (1-\lambda) p_i^{(1)} q_i^{(2)}\right|^r}{\left[q_i^{(1)}\right]^r \left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^r} \\ =\ & \frac{H(1-\lambda)^r \left|q_i^{(1)} p_i^{(2)} - p_i^{(1)} q_i^{(2)}\right|^r}{\left[q_i^{(1)}\right]^r \left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^r}. \end{aligned} \tag{2.16}$$

If we multiply (2.16) by $\lambda q_i^{(1)}$, we obtain

$$\begin{aligned} &\left|\lambda q_i^{(1)} f\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}}\right) - \lambda q_i^{(1)} f\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right)\right| \\ \leq\ & \frac{H\lambda(1-\lambda)^r \left[q_i^{(1)}\right]^{1-r} \left|q_i^{(1)} p_i^{(2)} - p_i^{(1)} q_i^{(2)}\right|^r}{\left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^r} \end{aligned} \tag{2.17}$$

for all $i \in \{1, ..., n\}$ and $\lambda \in [0,1]$.

If in (2.10) we choose x as above but

$$y = \frac{p_i^{(2)}}{q_i^{(2)}} \ (i = 1, ..., n),$$

we get

$$\begin{aligned} &\left| f\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}}\right) - f\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right)\right| \\ \leq\ & H\left|\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}} - \frac{p_i^{(2)}}{q_i^{(2)}}\right|^r \\ =\ & H\frac{\left|\lambda p_i^{(1)} q_i^{(2)} + (1-\lambda) p_i^{(2)} q_i^{(2)} - \lambda p_i^{(2)} q_i^{(1)} - (1-\lambda) q_i^{(2)} p_i^{(2)}\right|^r}{\left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^r \left[q_i^{(2)}\right]^r} \\ =\ & \frac{H(1-\lambda)^r \left|p_i^{(1)} q_i^{(2)} - p_i^{(2)} q_i^{(1)}\right|^r}{\left[q_i^{(2)}\right]^r \left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^r}. \end{aligned} \tag{2.18}$$

If we multiply (2.18) by $(1-\lambda) q_i^{(2)}$, we obtain

$$\left|(1-\lambda) q_i^{(2)} f\left(\frac{\lambda p_i^{(1)}+(1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda) q_i^{(2)}}\right)-(1-\lambda) q_i^{(2)} f\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right)\right| \leq \frac{H \lambda^r (1-\lambda)\left[q_i^{(2)}\right]^{1-r}\left|q_i^{(1)} p_i^{(2)}-p_i^{(1)} q_i^{(2)}\right|^r}{\left[\lambda q_i^{(1)}+(1-\lambda) q_i^{(2)}\right]^r} \tag{2.19}$$

for all $i \in\{1, \ldots, n\}$ and $\lambda \in[0,1]$.

If we add (2.13) and (2.19) and use the triangle inequality, we get

$$\begin{aligned} &\left|\left[\lambda q_i^{(1)}+(1-\lambda) q_i^{(2)}\right] f\left(\frac{\lambda p_i^{(1)}+(1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda) q_i^{(2)}}\right)\right. \\ &\left.-\lambda q_i^{(1)} f\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right)-(1-\lambda) q_i^{(2)} f\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right)\right| \\ \leq\ & \frac{H \lambda^r (1-\lambda)^r\left|\det\left[\begin{array}{cc} p_i^{(1)} & p_i^{(2)} \\ q_i^{(1)} & q_i^{(2)} \end{array}\right]\right|^r}{\left[\lambda q_i^{(1)}+(1-\lambda) q_i^{(2)}\right]^r}\left[\lambda^{1-r}\left[q_i^{(1)}\right]^{1-r}+(1-\lambda)^{1-r}\left[q_i^{(2)}\right]^{1-r}\right] \end{aligned} \tag{2.20}$$

for all $i \in\{1, \ldots, n\}$ and $\lambda \in[0,1]$.

If we sum (2.20) over i from 1 to n and use the generalised triangle inequality, we obtain the desired inequality (2.15). ∎

Remark 2. *If we assume that f is $L-$Lipschitzian, then the inequality (2.15) becomes*

$$\begin{aligned} 0 \leq\ & I_f\left(\lambda p^{(1)}+(1-\lambda) p^{(2)}, \lambda q^{(1)}+(1-\lambda) q^{(2)}\right) \\ & -\lambda I_f\left(p^{(1)}, q^{(1)}\right)-(1-\lambda) I_f\left(p^{(2)}, q^{(2)}\right) \\ \leq\ & L \lambda(1-\lambda) \sum_{i=1}^n \frac{\left|\det\left[\begin{array}{cc} p_i^{(1)} & p_i^{(2)} \\ q_i^{(1)} & q_i^{(2)} \end{array}\right]\right|}{\lambda q_i^{(1)}+(1-\lambda) q_i^{(2)}}, \end{aligned} \tag{2.21}$$

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

3. Some Particular Cases

Using the inequality (2.14), i.e.,

$$0 \leq I_f(p, q) \leq\left\|f^{\prime}\right\|_{\infty} V(p, q), \tag{3.1}$$

provided that f is absolutely continuous on $[r, R]$ and $f^{\prime} \in L_{\infty}[r, R],\ \left\|f^{\prime}\right\|_{\infty}:= ess \sup_{t \in[r, R]}\left|f^{\prime}(t)\right|$, we shall be able to point out a number of interesting inequalities between different divergence measures in Information Theory.

Proposition 1. *Let p, q be two probability distributions with the property that*

$$0<r \leq \frac{p_i}{q_i}=: r_i<R<\infty \quad \textit{for all } i \in\{1, \ldots, n\}. \tag{3.2}$$

Then we have the inequality

$$0 \le D(p,q) \le \begin{cases} \left[\ln\sqrt{\frac{R}{r}} + \left|1+\ln\sqrt{rR}\right|\right] V(p,q) & \text{if} \quad 0 < r \le e^{-1} \\ (1+\ln R)\, V(p,q) & \text{if} \quad e^{-1} < r < 1 \end{cases}. \tag{3.3}$$

Proof. Consider the mapping $f:(0,\infty)\to\mathbb{R}$, $f(t)=t\ln t$. Then

$$f'(t) = \ln(et).$$

Therefore

$$\|f'\|_\infty = \sup_{t\in[r,R]} |f'(t)| = \max\{|\ln(er)|, \ln(eR)\}.$$

Now, observe that

1. If $0<r\le e^{-1}$, then $|\ln(er)| = \ln(er) = -1-\ln r$ and then

$$\begin{aligned}\max\{|\ln(er)|, \ln(eR)\} &= \frac{-1-\ln r+\ln R+1+|\ln r+1+1+\ln R|}{2} \\ &= \ln\left(\frac{R}{2}\right)^{\frac{1}{2}} + \left|1+\ln\sqrt{rR}\right|.\end{aligned}$$

2. If $e^{-1}<r<1$, then $|\ln(er)| = 1+\ln r$ and then

$$\max\{|\ln(er)|, \ln(eR)\} = \max\{\ln(er), \ln(eR)\} = 1+\ln R$$

and the proposition is proved.

∎

We also have the following proposition.

Proposition 2. *Let p,q be as in Proposition 1. Then we have the inequality:*

$$0 \le D(q,p) \le \frac{1}{r} V(p,q). \tag{3.4}$$

Proof. Consider the mapping $f(t) = -\ln t$, $t\in(0,\infty)$. Then

$$f'(t) = -\frac{1}{t} \quad \text{and} \quad \|f'\|_\infty = \frac{1}{r}.$$

Since

$$I_f(p,q) = -\sum_{i=1}^n q_i \ln\left(\frac{p_i}{q_i}\right) = \sum_{i=1}^n q_i \ln\left(\frac{q_i}{p_i}\right) = D(q,p),$$

the inequality (3.1) gives the desired result (3.4). ∎

We point out now a bound for the χ^2-divergence.

Proposition 3. *Let p,q be as above. Then*

$$0 \le D_{\chi^2}(p,q) \le 2(R-1)\, V(p,q). \tag{3.5}$$

Proof. Consider the mapping $f:(0,\infty)\to\mathbb{R}$, $f(t)=(t-1)^2$. As $f'(t) = 2(t-1)$, it follows that $\|f'\|_\infty = 2(R-1)$. Using (3.1), we obtain (3.5). ∎

The following result for Hellinger discrimination also holds.

Proposition 4. *Assume that the probability distributions p, q satisfy the condition (3.1). Then we have the inequality*

$$0 \leq h^2(p,q) \leq \frac{1}{4}\left[\frac{\sqrt{R}-\sqrt{r}}{\sqrt{rR}} + \left|\frac{\sqrt{R}+\sqrt{r}}{\sqrt{rR}} - 2\right|\right] V(p,q). \tag{3.6}$$

Proof. Consider the mapping $f : (0,\infty) \to \mathbb{R}$, $f(t) = \frac{1}{2}\left(\sqrt{t}-1\right)^2$. Then

$$f'(t) = \frac{1}{2} - \frac{1}{2\sqrt{t}} \quad \text{and} \quad f''(t) = \frac{1}{4\sqrt{t^3}},\ t \in (0,\infty).$$

Then

$$\begin{aligned}
\|f'\|_\infty &= \sup_{t\in[r,R]} |f'(t)| = \max\left\{\left|\frac{\sqrt{r}-1}{2\sqrt{r}}\right|, \left|\frac{\sqrt{R}-1}{2\sqrt{R}}\right|\right\} \\
&= \max\left\{\frac{1-\sqrt{r}}{2\sqrt{r}}, \frac{\sqrt{R}-1}{2\sqrt{R}}\right\} \\
&= \frac{1}{4}\left[\frac{\sqrt{R}-\sqrt{r}}{\sqrt{rR}} + \left|2 - \frac{\sqrt{R}+\sqrt{r}}{\sqrt{rR}}\right|\right].
\end{aligned}$$

Using (3.1), we deduce (3.6). ∎

Now, consider the Bhattacharyya distance which is defined by [27]

$$B(p,q) = \sum_{i=1}^{n} \sqrt{p_i q_i}. \tag{3.7}$$

The following proposition holds.

Proposition 5. *Assume that p, q are probability distributions. Then we have the inequality*

$$0 \leq 1 - B(p,q) \leq V_{\frac{1}{2}}(p,q), \tag{3.8}$$

where $V_{\frac{1}{2}}(p,q) = \sum_{i=1}^{n} \sqrt{q_i(p_i - q_i)}$ is the $\frac{1}{2}-$variational distance.

Proof. Consider the mapping $f : [0,\infty) \to [0,\infty)$ given by

$$f(t) = -\sqrt{t} + 1.$$

Obviously,

$$|f(x) - f(y)| = \left|\sqrt{x} - \sqrt{y}\right| \leq \sqrt{|x-y|}, \text{ for all } x, y \in [0,\infty),$$

which shows that f is of $\frac{1}{2}-$Hölder type with the constant $H = 1$.

Applying Theorem 6, we deduce (3.8). ∎

Another inequality for Bhattacharyya distance in terms of the variational distance V is embodied in the following proposition.

Proposition 6. *Assume that the probability distributions p, q satisfy the the condition:*

$$0 < r \leq \frac{p_i}{q_i} \text{ for all } i \in \{1, ..., n\}. \tag{3.9}$$

Then we have the inequality

$$0 \leq 1 - B(p,q) \leq \frac{1}{2\sqrt{r}} V(p,q). \tag{3.10}$$

Proof. For the mapping $f(t) = -\sqrt{t} + 1$, we have $f'(t) = -\frac{1}{2\sqrt{t}}$ and $\|f'\|_\infty = \sup_{t\in[0,\infty)} = \frac{1}{2\sqrt{r}}$. Applying (3.1), we deduce (3.11). ∎

Now, we consider another useful divergence measure in Information Theory which is known as the *Harmonic distance:*

$$M(p,q) := \sum_{i=1}^{n} \frac{2p_i q_i}{p_i + q_i}. \tag{3.11}$$

Proposition 7. *Assume that the probability distributions p, q satisfy the condition (3.9). Then we have the inequality*

$$0 \leq 1 - M(p,q) \leq \frac{2}{(r+1)^2} V(p,q). \tag{3.12}$$

Proof. Consider the mapping $f : [0,\infty) \to \mathbb{R}$, $f(t) = 1 - \frac{2t}{t+1}$. Obviously $f'(t) = -\frac{2}{(t+1)^2}$, $f''(t) = \frac{4}{(t+1)^3}$ and $\|f'\|_\infty = \sup_{t\in[r,\infty)} |f'(t)| = \frac{2}{(r+1)^2}$. As

$$I_f(p,q) = 1 - M(p,q),$$

then by (3.1) we deduce the desired inequality (3.12). ∎

Another useful divergence measure is the Jeffreys $J-$divergence defined by [26]

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

We can state the following proposition as well.

Proposition 8. *Assume that the probability distributions p, q satisfy the condition:*

$$\frac{p_i}{q_i} \leq R < \infty, \quad (i = 1, ..., n). \tag{3.13}$$

Then we have the inequality

$$0 \leq J(p,q) \leq \left(\ln R - \frac{1}{R} + 1\right) V(p,q). \tag{3.14}$$

Proof. Consider the mapping $f(t) = (t-1)\ln t$, $t > 0$. Then

$$\begin{aligned} f'(t) &= \ln t - \frac{1}{t} + 1; \quad t \in (0,\infty) \\ f''(t) &= \frac{t+1}{t^2}, \quad t \in (0,\infty) \end{aligned}$$

which shows that

$$\|f'\|_\infty = \sup_{t\in(0,R]} |f'(t)| = f'(R) = \ln R - \frac{1}{R} + 1.$$

As

$$I_f(p,q) = J(p,q),$$

then by (3.1) we deduce (3.14). ∎

Finally, the following result for the triangular discrimination holds (for the definition of triangular discrimination, see [24]).

Proposition 9. *If p, q are such that the condition (3.13) holds, then*

$$0 \leq \Delta(p,q) \leq \frac{(R-1)(R+3)}{(R+1)^2} V(p,q) \leq V(p,q). \tag{3.15}$$

The proof is obvious by (3.2) applied for the mapping $f(t) = \frac{(t-1)^2}{t+1}$, and we omit the details.

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

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

[25] D. DACUNHA-CASTELLE, Ecole d'ete de Probabilités de Saint-Fleour, III-1997, Berlin, Heidelberg: Springer 1978.

[26] H. JEFFREYS, An invariant form for the prior probability in estimation problems, *Proc. Roy. Soc. London*, *Ser. A*, **1986** (1946), 453-461.

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

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

[29] S.S. DRAGOMIR, A converse inequality for the Csiszár Φ-divergence, *submitted.*

[30] S.S. DRAGOMIR, Some inequalities for (m, M)-convex mappings and applications for the Csiszár $\Phi-$divergence in information theory, *submitted.*

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

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

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