UPPER AND LOWER BOUNDS FOR CSISZÁR $f-$DIVERGENCE IN TERMS OF THE KULLBACK-LEIBLER DISTANCE AND APPLICATIONS

S.S. DRAGOMIR

Abstract. In this paper we point out an upper and a lower bound for the Csiszár $f-$divergence of two discrete random variables in terms of the Kullback-Leibler distance. Some particular cases for Hellinger and triangular discimination, χ^2 -distance and the Rényi $\alpha-$entropy, 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

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

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 10, 1999.

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

Key words and phrases. Csiszár $f-$divergence, Divergence measures in Information Theory, Jensen's Inequality, Hellinger and triangular discimination.

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

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

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 results are given.

2. Some Inequalities Between Csiszár $f-$Divergence and the Kullback-Leibler Distance

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 inequality

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

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

A similar 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 $m \leq M$) if it is both $m-$lower convex and $M-$upper 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*

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

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

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

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

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

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:

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

and

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

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

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. This result complements, in a sense, the results presented above in Theorem 5.

Theorem 6. *Assume that the generating mapping $f : (0,\infty) \to \mathbb{R}$ is normalised, i.e., $f(1) = 0$ and satisfies the assumptions*

(i) *f is twice differentiable on (r,R), where $0 \leq r \leq 1 \leq R \leq \infty$;*
(ii) *there exists the real constants m, M such that*

$$m \leq t f''(t) \leq M \quad \text{for all } t \in (r,R). \tag{2.9}$$

If p, q are discrete probability distributions satisfying the assumption

$$r \leq r_i := \frac{p_i}{q_i} \leq R \quad \text{for all } i \in \{1,...,n\}, \tag{2.10}$$

then we have the inequality

$$mD(p,q) \leq I_f(p,q) \leq MD(p,q). \tag{2.11}$$

Proof. Define the mapping $F_m : (0,\infty) \to \mathbb{R}$, $F_m(t) = f(t) - mt\ln t$. Then $F_m(\cdot)$ is normalised, twice differentiable and since

$$F_m''(t) = f''(t) - \frac{m}{t} = \frac{1}{t}(tf''(t) - m) \geq 0 \tag{2.12}$$

for all $t \in (r,R)$, it follows that $F_m(\cdot)$ is convex on (r,R). Applying the nonnegativity property of the Csiszár $f-$divergence functional for $F_m(\cdot)$ and the linearity property, we may state that

$$\begin{aligned} 0 &\leq I_{F_m}(p,q) = I_f(p,q) - mI_{(\cdot)\ln(\cdot)}(p,q) \\ &= I_f(p,q) - mD(p,q) \end{aligned} \tag{2.13}$$

from where results the first inequality in (2.11).

Define $F_M : (0,\infty) \to \mathbb{R}$, $F_M(t) := Mt\ln t - f(t)$, which is obviously normalised, twice differentiable and by (2.9), convex on (r,R). Applying the nonnegativity property of Csiszár $f-$divergence for F_M, we obtain the second part of (2.11). ∎

Remark 1. *If in (2.9) we have the strict inequality "$<$" for any $t \in (r,R)$, then the mappings F_m and F_M are strictly convex and the case of equality holds in (2.11) iff $p = q$.*

The following theorem concerning the convexity property of the Csiszár $f-$divergence also holds.

Theorem 7. *Assume that f satisfies the assumptions (i) and (ii) from Theorem 6. If $p^{(j)}$, $q^{(j)}$ $(j = 1,2)$ are probability distributions satisfying (2.10), i.e.,*

$$r \leq \frac{p_i^{(j)}}{q_i^{(j)}} \leq R \quad \text{for all} \quad i \in \{1,...,n\} \quad \text{and} \quad j \in \{1,2\}, \tag{2.14}$$

then

$$r \leq \frac{\lambda p_i^{(1)} + (1-\lambda)p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)q_i^{(2)}} \leq R \quad \text{for all } i \in \{1,...,n\} \quad \text{and} \quad \lambda \in [0,1] \tag{2.15}$$

and

$$\begin{aligned} & m\big[D\left(\lambda p^{(1)} + (1-\lambda)p^{(2)}, \lambda q^{(1)} + (1-\lambda)q^{(2)}\right) \\ & -\lambda D\left(p^{(1)}, q^{(1)}\right) - (1-\lambda)D\left(p^{(2)}, q^{(2)}\right)\big] \\ \leq\ & I_f\left(\lambda p^{(1)} + (1-\lambda)p^{(2)}, \lambda q^{(1)} + (1-\lambda)q^{(2)}\right) \\ & -\lambda I_f\left(p^{(1)}, q^{(1)}\right) - (1-\lambda)I_f\left(p^{(2)}, q^{(2)}\right) \\ \leq\ & M\big[D\left(\lambda p^{(1)} + (1-\lambda)p^{(2)}, \lambda q^{(1)} + (1-\lambda)q^{(2)}\right) \\ & -\lambda D\left(p^{(1)}, q^{(1)}\right) - (1-\lambda)D\left(p^{(2)}, q^{(2)}\right)\big] \end{aligned} \tag{2.16}$$

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

Proof. By (2.14), we have

$$r\lambda q_i^{(1)} \leq \lambda p_i^{(1)} \leq \lambda R q_i^{(1)} \text{ for all } i \in \{1, ..., n\} \tag{2.17}$$

and

$$r(1-\lambda) q_i^{(2)} \leq (1-\lambda) p_i^{(2)} \leq R(1-\lambda) q_i^{(2)} \text{ for all } i \in \{1, ..., n\}. \tag{2.18}$$

Summing the above (2.17) and (2.18), we obtain (2.15).

It is already known that the mappings F_m, F_M as defined in Theorem 6 are convex and normalised.

Applying the "Joint Convexity Principle" for $I_{F_m}(\cdot,\cdot)$, i.e.,

$$\begin{aligned} & I_{F_m}\left(\lambda\left(p^{(1)}, q^{(1)}\right) + (1-\lambda)\left(p^{(2)}, q^{(2)}\right)\right) \\ \leq\ & \lambda I_{F_m}\left(p^{(1)}, q^{(1)}\right) + (1-\lambda) I_{F_m}\left(p^{(2)}, q^{(2)}\right) \end{aligned} \tag{2.19}$$

and rearranging the terms, we may end up with the first inequality in (2.16).

The second inequality follows likewise if we apply the same property to the Csiszár $f-$divergence $I_{F_M}(\cdot,\cdot)$.

We omit the details. ∎

Remark 2. *If $m > 0$ in (2.9), then the inequality (2.11) is a better result than the positivity property of Csiszár divergence. The same will apply for the joint convexity of Csiszár divergence if $m > 0$.*

Using the inequality (2.2) which holds for Φ differentiable convex and normalised functions for p, q probability distributions, we can state the following theorem as well.

Theorem 8. *Let $f : [0,\infty) \to \mathbb{R}$ be a normalised mapping, i.e., $f(1) = 0$ and satisfies the assumptions:*

(i) f is twice differentiable on (r, R), where $0 \leq r \leq 1 \leq R \leq \infty$;
(ii) there exists the constants m, M such that

$$m \leq t f''(t) \leq M \quad \text{for all } t \in (r, R). \tag{2.20}$$

If p, q are discrete probability distributions satisfying the assumption

$$r \leq r_i = \frac{p_i}{q_i} \leq R \quad \text{for all } i \in \{1, ..., n\}, \tag{2.21}$$

then we have the inequality

$$\begin{aligned} & I_{f'}\left(\frac{p^2}{q}, p\right) - I_{f'}(p,q) - MD(q,p) \\ \leq\ & I_f(p,q) \\ \leq\ & I_{f'}\left(\frac{p^2}{q}, p\right) - I_{f'}(p,q) - mD(q,p). \end{aligned} \tag{2.22}$$

Proof. We know (see the proof of Theorem 6) that the mapping $F_m : (0,\infty) \to \mathbb{R}$, $F_m(t) = f(t) - mt\ln t$ is normalised, twice differentiable and convex on (r, R).

If we apply the second inequality from (2.2) for F_m, we may write:

$$I_{F_m}(p,q) \leq I_{F'_m}\left(\frac{p^2}{q}, p\right) - I_{F'_m}(p,q). \tag{2.23}$$

However,

$$\begin{aligned} I_{F_m}(p,q) &= I_f(p,q) - mD(q,p), \\ I_{F'_m}\left(\frac{p^2}{q},p\right) &= I_{f'(\cdot)-m[\ln(\cdot)+1]}\left(\frac{p^2}{q},p\right) \\ &= I_{f'}\left(\frac{p^2}{q},p\right) - mI_{\ln(\cdot)}\left(\frac{p^2}{q},p\right) - m \\ &= I_{f'}\left(\frac{p^2}{q},p\right) + mD\left(p,\frac{p^2}{q}\right) - m \end{aligned}$$

and

$$I_{F'_m}(p,q) = I_{f'}(p,q) + mD(q,p) - m.$$

Consequently, by (2.23), we have

$$\begin{aligned} & I_f(p,q) - mD(p,q) \\ \le\ & I_{f'}\left(\frac{p^2}{q},p\right) + mD\left(p,\frac{p^2}{q}\right) - m - I_{f'}(p,q) - mD(q,p) + m \\ =\ & I_{f'}\left(\frac{p^2}{q},p\right) + m\left(D\left(p,\frac{p^2}{q}\right) - D(q,p)\right) - I_{f'}(p,q). \end{aligned}$$

As a simple computation shows that $D\left(p,\frac{p^2}{q}\right) = -D(p,q)$, the second inequality in (2.22) is proved.

Consider $F_M(t) := Mt\ln t - f(t)$, which is obviously normalised, twice differentiable and convex on (r,R).

If we apply the second inequality from (2.2) for F_M, we may write:

$$I_{F_M}(p,q) \le I_{F'_M}\left(\frac{p^2}{q},p\right) - I_{F'_M}(p,q). \tag{2.24}$$

However,

$$\begin{aligned} I_{F_M}(p,q) &= MD(p,q) - I_f(p,q); \\ I_{F'_M}\left(\frac{p^2}{q},p\right) &= -MD\left(p,\frac{p^2}{q}\right) + M - I_{f'}\left(\frac{p^2}{q},p\right); \\ I_{F'_M}(p,q) &= -MD(q,p) + M - I_{f'}(p,q) \end{aligned}$$

and then, by (2.24), we get

$$\begin{aligned} & MD(p,q) - I_f(p,q) \\ \le\ & -MD\left(p,\frac{p^2}{q}\right) + M - I_{f'}\left(\frac{p^2}{q},p\right) + MD(q,p) - M + I_{f'}(p,q), \end{aligned}$$

which is equivalent with the first part of (2.22). ∎

Remark 3. *The inequality (2.22) is obviously equivalent with the following one:*

$$mD(q,p) \le I_{f'}\left(\frac{p^2}{q},p\right) - I_{f'}(p,q) - I_f(p,q) \le MD(q,p).$$

The above results have natural applications when the Kullback-Leibler distance is compared with a number of other divergence measures arising in Information Theory.

3. Some Particular Cases

Using Theorem 6, we shall be able to point out the following particular cases which are of interest in Information Theory.

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

$$0<r\leq \frac{p_i}{q_i}=r_i\leq R<\infty \quad \text{for all } i\in\{1,...,n\}. \tag{3.1}$$

Then we have the inequality

$$\frac{1}{R}D(p,q)\leq D(q,p)\leq \frac{1}{r}D(p,q). \tag{3.2}$$

Proof. Consider the mapping $f:[r,R]\to\mathbb{R}$, $f(t)=-\ln t$. Define $g(t)=tf''(t)=t\cdot\left(\frac{1}{t^2}\right)=\frac{1}{t}$. Then obviously

$$\sup_{t\in[r,R]} g(t)=\frac{1}{r} \text{ and } \inf_{t\in[r,R]} g(t)=\frac{1}{R}.$$

Also,

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

Now, using (2.11) with $m=\frac{1}{R}$ and $M=\frac{1}{r}$, we deduce the desired inequality. ∎

Corollary 2. *With the above assumptions for p and q, we have:*

$$r\leq \frac{D(p,q)}{D(q,p)}\leq R. \tag{3.3}$$

Corollary 3. *Assume that p,q satisfy the condition*

$$\left|\frac{p_i}{q_i}-1\right|\leq\varepsilon \quad \text{for all } i\in\{1,...,n\}. \tag{3.4}$$

Then we have the inequality

$$\left|\frac{D(p,q)}{D(q,p)}-1\right|\leq\varepsilon.$$

The following proposition connecting the χ^2-distance with the Kullback-Leibler distance holds.

Proposition 2. *Let p,q be two probability distributions satisfying the condition (3.1). Then we have the inequality:*

$$2r\leq\frac{D_{\chi^2}(p,q)}{D(p,q)}\leq 2R. \tag{3.5}$$

Proof. Consider the mapping $f:[r,R]\to\mathbb{R}$, $f(t)=(t-1)^2$. Define $g(t)=tf''(t)=2t$. Then, obviously,

$$\sup_{t\in[r,R]} g(t)=2R \text{ and } \inf_{t\in[r,R]} g(t)=2r.$$

Since

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

then, applying (2.11) for $m=2r$ and $M=2R$, we deduce the desired inequality. ∎

Remark 4. *The following inequality is well known in the literature*

$$D(p,q) \le D_{\chi^2}(p,q). \tag{3.6}$$

For a simple proof of this fact as well as for different applications in Information Theory, see [27].
Now, observe that from the first inequality in (3.5), we have

$$D(p,q) \le \frac{1}{2r} D_{\chi^2}(p,q). \tag{3.7}$$

We remark that if $\frac{1}{2r} \le 1$ *i.e.,* $r \ge \frac{1}{2}$*, the inequality (3.7) is better than (3.6).*

The following corollary is obvious.

Corollary 4. *Assume that the probability distributions* p,q *satisfy the condition (3.4). Then*

$$\frac{1}{2}\left|\frac{D_{\chi^2}(p,q)}{D(p,q)} - 2\right| \le \varepsilon. \tag{3.8}$$

The following inequality connecting the Kullback-Leibler distance with the Hellinger discrimination holds.

Proposition 3. *Assume that the probability distributions* p,q *satisfy the condition (3.1). Then we have the inequality:*

$$\frac{1}{4\sqrt{R}} D(p,q) \le h^2(p,q) \le \frac{1}{4\sqrt{r}} D(p,q). \tag{3.9}$$

Proof. Consider the mapping $f(t) = \frac{1}{2}\left(\sqrt{t}-1\right)^2$. Then $f'(t) = \frac{1}{2} - \frac{1}{2\sqrt{t}}$ and $f''(t) = \frac{1}{4\sqrt{t^3}}$. Define $g : [r,R] \to \mathbb{R}$ given by

$$g(t) = tf''(t) = \frac{1}{4\sqrt{t}}.$$

Then, obviously

$$\sup_{t\in[r,R]} g(t) = \frac{1}{4\sqrt{r}} \text{ and } \inf_{t\in[r,R]} g(t) = \frac{1}{4\sqrt{R}}.$$

Since

$$I_f(p,q) = h^2(p,q),$$

then by (2.11) for $m = \frac{1}{4\sqrt{R}}$ and $M = \frac{1}{4\sqrt{r}}$, we deduce the desired inequality (3.9). ∎

Remark 5. *The following inequality is well known in the literature (see for example* [25]*):*

$$D(p,q) \ge 2h^2(p,q) \tag{3.10}$$

for any p,q *probability distribution.*
From the second inequality in (3.9), we have

$$D(p,q) \ge 4\sqrt{r}h^2(p,q). \tag{3.11}$$

We remark that if $4\sqrt{r} \ge 2$*, i.e.,* $r \ge \frac{1}{4}$*, then the inequality in (3.11) is better than (3.10).*

The following result establishes a connection between the triangular discrimination Δ and the Kullback-Leibler distance.

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

(*i*) *If* $0<r\leq\frac{1}{2}$, *then we have*

$$8\min\left\{\frac{r}{(r+1)^3},\frac{R}{(R+1)^3}\right\}D(p,q)\leq\Delta(p,q)\leq\frac{32}{27}D(p,q). \tag{3.12}$$

(*ii*) *If* $\frac{1}{2}<r<1$, *then we have*

$$\frac{8R}{(R+1)^3}D(p,q)\leq\Delta(p,q)\leq\frac{8r}{(r+1)^3}D(p,q). \tag{3.13}$$

Proof. Consider the mapping $f(t)=\frac{(t-1)^2}{t+1}$. We have

$$f'(t)=1-\frac{4}{(t+1)^2}$$

and

$$f''(t)=\frac{8}{(t+1)^3}.$$

Define

$$g:[r,R]\to\mathbb{R},\ g(t)=tf''(t)=\frac{8t}{(t+1)^3},\ t\in[r,R].$$

We have

$$g'(t)=\frac{8(1-2t)}{(t+1)^4},$$

which shows that g has the maximum realized at $t_0=\frac{1}{2}$ and

$$\max_{t\in(0,\infty)}g(t)=g\left(\frac{1}{2}\right)=\frac{32}{27}.$$

We have the two cases:

1) If $0<r\leq\frac{1}{2}$, then

$$\begin{aligned}\sup_{t\in[r,R]}g(t)&=\frac{32}{27}\quad\text{and}\\ \inf_{t\in[r,R]}g(t)&=\min[g(r),g(R)]=\min\left\{\frac{8r}{(r+1)^3},\frac{8R}{(R+1)^3}\right\}.\end{aligned}$$

2) If $\frac{1}{2}<r<1$, then

$$\begin{aligned}\sup_{t\in[r,R]}g(t)&=g(r)=\frac{8r}{(r+1)^3}\quad\text{and}\\ \inf_{t\in[r,R]}g(t)&=g(R)=\frac{8R}{(R+1)^3}.\end{aligned}$$

Applying the inequality (2.11), we deduce (3.12) and (3.13). We omit the details. ∎

Remark 6. *It is clear, by the above arguments, that for every probability distribution we have the inequality*

$$\Delta\left(p,q\right)\leq\frac{32}{27}D\left(p,q\right). \tag{3.14}$$

We know that (see Topsoe [24]*)*

$$2h^{2}\left(p,q\right)\leq\Delta\left(p,q\right)\leq4h^{2}\left(p,q\right). \tag{3.15}$$

Now, as $D\left(p,q\right)\geq2h^{2}\left(p,q\right)$*, then we obtain*

$$\Delta\left(p,q\right)\leq2D\left(p,q\right), \tag{3.16}$$

which is not as good as our result (3.14).

Let us compare the *Rényi* $\alpha-$*entropy* with the Kullback-Leibler distance.
The following proposition holds.

Proposition 5. *Assume that the probability distributions* p,q *satisfy the condition (3.1). Then*

$$\alpha\left(\alpha-1\right)r^{\alpha-1}D\left(p,q\right)+1\leq R_{\alpha}\left(p,q\right)\leq\alpha\left(\alpha-1\right)R^{\alpha-1}D\left(p,q\right)+1 \tag{3.17}$$

for $\alpha>1$*.*

Proof. Consider the mapping $f:(0,\infty)\to\mathbb{R},\ f\left(t\right)=t^{\alpha}-1,\ \alpha>1.$ Then $f'\left(t\right)=\alpha t^{\alpha-1}$ and $f''\left(t\right)=\alpha\left(\alpha-1\right)t^{\alpha-2}$. Define $g:[r,R]\to\mathbb{R},\ g\left(t\right)=tf''\left(t\right)=\alpha\left(\alpha-1\right)t^{\alpha-1}$. It is obvious that

$$\sup_{t\in[r,R]}g\left(t\right)=\alpha\left(\alpha-1\right)R^{\alpha-1}\quad\text{and}\quad\inf_{t\in[r,R]}g\left(t\right)=\alpha\left(\alpha-1\right)r^{\alpha-1}.$$

Now, observe that $f\left(1\right)=0$, i.e., f is normalised and so we can apply the inequality (2.11) getting

$$\alpha\left(\alpha-1\right)r^{\alpha-1}D\left(p,q\right)\leq\sum_{i=1}^{n}q_{i}\left[\left(\frac{p_{i}}{q_{i}}\right)^{\alpha}-1\right]\leq\alpha\left(\alpha-1\right)R^{\alpha-1}D\left(p,q\right),$$

i.e.,

$$\alpha\left(\alpha-1\right)r^{\alpha-1}D\left(p,q\right)+1\leq R_{\alpha}\left(p,q\right)\leq\alpha\left(\alpha-1\right)R^{\alpha-1}D\left(p,q\right)+1$$

and the proposition is proved. ∎

We define the *Bhattacharyya distance* by (see [27])

$$B\left(p,q\right):=\sum_{i=1}^{n}\sqrt{p_{i}q_{i}}.$$

The following proposition holds.

Proposition 6. *Assume that the probability distributions* p,q *satisfy the condition (3.1). Then*

$$4\sqrt{r}\left[1-B\left(p,q\right)\right]\leq D\left(p,q\right)\leq4\sqrt{R}\left[1-B\left(p,q\right)\right]. \tag{3.18}$$

Proof. Consider the mapping $f:(0,\infty)\to\mathbb{R},\ f\left(t\right)=\sqrt{t}-1$. Then f is normalised, $f'\left(t\right)=\frac{1}{2}t^{-\frac{1}{2}},\ f''\left(t\right)=-\frac{1}{4}t^{-\frac{3}{2}}$. Define $g:[r,R]\to\mathbb{R},\ g\left(t\right)=tf''\left(t\right)=-\frac{1}{4}t^{-\frac{1}{2}}$. It is obvious that

$$\sup_{t\in[r,R]}g\left(t\right)=g\left(R\right)=-\frac{1}{4\sqrt{R}},\quad\inf_{t\in[r,R]}g\left(t\right)=g\left(r\right)=-\frac{1}{4\sqrt{r}}.$$

Applying the inequality (2.11), we have:

$$-\frac{1}{4\sqrt{r}}D(p,q) \le \sum_{i=1}^{n} q_i\left(\sqrt{\frac{p_i}{q_i}}-1\right) \le -\frac{1}{4\sqrt{R}}D(p,q),$$

i.e.,

$$1-\frac{1}{4\sqrt{r}}D(p,q) \le B(p,q) \le 1-\frac{1}{4\sqrt{R}}D(p,q),$$

which is equivalent to (3.18). ∎

We define the *harmonic* distance by

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

The following proposition holds.

Proposition 7. *Assume that p,q are two discrete probability distributions. Then*

$$1-\frac{16}{27}D(p,q) \le M(p,q) \le 1. \tag{3.19}$$

Proof. Consider the mapping $f:(0,\infty)\to\mathbb{R}$, $f(t)=\frac{2t}{t+1}-1$. Then f is normalised

$$f'(t)=\frac{2}{(t+1)^2},\quad f''(t)=\frac{-4}{(t+1)^3}.$$

Define $g:[r,R]\to\mathbb{R}$, $g(t)=tf''(t)=\frac{-4t}{(t+1)^3}$. Then

$$g'(t)=\frac{4(2t-1)}{(t+1)^4}.$$

It is clear that g is monotonic decreasing on $\left[0,\frac{1}{2}\right)$ and monotonic increasing on $\left(\frac{1}{2},\infty\right)$. We have

$$\begin{aligned}\inf_{t\in(0,\infty)} g(t) &= g\left(\frac{1}{2}\right) = -\frac{16}{27},\\ \sup_{t\in(0,\infty)} g(t) &= 0.\end{aligned}$$

Applying the inequality (2.11) for $m=-\frac{16}{27}$ and $M=0$, we deduce

$$-\frac{16}{27}D(p,q) \le \sum_{i=1}^{n} q_i\left\{\left[\frac{2\frac{p_i}{q_i}}{\frac{p_i}{q_i}+1}\right]-1\right\} \le 0,$$

which is equivalent to

$$-\frac{16}{27}D(p,q) \le M(p,q)-1 \le 0$$

and the inequality (3.19) is proved. ∎

The above result can be improved if we know more information about $r_i := \frac{p_i}{q_i}$ $(i=1,...,n)$. We can state the following proposition.

Proposition 8. *Assume that p,q satisfy the condition (2.10).*

(*i*) *If* $r \in \left(0, \frac{1}{2}\right)$, *then*

$$1 - \frac{16}{27} D(p,q) \leq M(p,q) \leq 1 - 4 \min\left\{\frac{r}{(r+1)^3}, \frac{R}{(R+1)^3}\right\} D(p,q). \tag{3.20}$$

(*ii*) *If* $r \in \left[\frac{1}{2}, 1\right)$, *then*

$$1 - \frac{4r}{(r+1)^3} D(p,q) \leq M(p,q) \leq 1 - \frac{4R}{(R+1)^3} D(p,q). \tag{3.21}$$

Proof. (*i*) If $r \in \left(0, \frac{1}{2}\right)$, then

$$\begin{aligned} -\frac{16}{27} &\leq g(t) \leq \max\{g(r), g(R)\} \\ &= \max\left\{-\frac{4r}{(r+1)^3}, -\frac{4R}{(R+1)^3}\right\} \\ &= -4\min\left\{\frac{r}{(r+1)^3}, \frac{R}{(R+1)^3}\right\}, \; t \in [r,R]. \end{aligned}$$

and then, applying (2.11), we may write

$$-\frac{16}{27} D(p,q) \leq M(p,q) - 1 \leq -4\min\left\{\frac{r}{(r+1)^3}, \frac{R}{(R+1)^3}\right\} D(p,q),$$

and the inequality (3.20) is proved.

(*ii*) If $r \in \left[\frac{1}{2}, 1\right)$, then

$$g(r) \leq g(t) \leq g(R) \quad \text{for all} \quad t \in [r,R],$$

that is,

$$-\frac{4r}{(r+1)^3} \leq g(t) \leq -\frac{4R}{(R+1)^3}, \; t \in [r,R].$$

Applying (2.11), we deduce (3.21).

∎

Let us consider the *J-divergence* defined by [26]

$$\begin{aligned} J(p,q) \quad : &= \sum_{i=1}^{n} (p_i - q_i) \log\left(\frac{p_i}{q_i}\right) \\ &= \sum_{i=1}^{n} q_i \left(\frac{p_i}{q_i} - 1\right) \log\left(\frac{p_i}{q_i}\right) \\ &= I_f(p,q), \end{aligned}$$

where $f : (0,\infty) \to \mathbb{R}$, $f(x) = (x-1)\ln x$.

The following proposition also holds.

Proposition 9. *Assume that* p, q *satisfy the condition (2.10). Then*

$$\frac{R+1}{R} D(p,q) \leq J(p,q) \leq \frac{r+1}{r} D(p,q). \tag{3.22}$$

Proof. Consider $f(t) = (t-1)\ln t$. Then $f'(t) = \ln t - \frac{1}{t} + 1$ and $f''(t) = \frac{t+1}{t^2}$. Define $g(t) = tf''(t) = 1 + \frac{1}{t}$. Then obviously

$$\sup_{t\in[r,R]} g(t) = 1 + \frac{1}{r}, \quad \inf_{t\in[r,R]} g(t) = 1 + \frac{1}{R}.$$

Now, using the inequality (2.11), for $M = \frac{r+1}{1}$, $m = \frac{R+1}{R}$, we obtain the desired result. ∎

Remark 7. *Similar results can be obtained by applying Theorem 8, but 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 Memory-less 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